JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

Number 310 JUNE, 1965 *Volume 60*

R. A. FISHER AND THE LAST FIFTY YEARS OF STATISTICAL METHODOLOGY*

M. S. Bartlett

University of Chicago

and

University College, London

1. OPENING REMARKS

Firstly may I express my appreciation of this invitation to address this conference. I feel particularly honoured at being the first to give this R. A. Fisher Memorial Lecture, perhaps partly because I have always tried to combine my profound admiration of his scientific achievements with a reluctance to be blinded by their brilliance; and it could be a matter of opinion how far such an attempt at detachment qualified me in particular to discuss his work. I understand that there was no obligation in this lecture to consider Fisher's work, but it seemed rather odd if in this first memorial lecture one did not take the opportunity to do so; and a convenient way of doing this, in view of Fisher's great influence on the development of statistics during the last fifty years, was to try to survey this period, with particular reference to the position Fisher occupied.

Let me make it clear that I am going to concentrate on Fisher's contributions to statistics. It is well-known that his work in genetics was of comparable status; it is largely represented by his book *The Genetical Theory of Natural Selection*, though in his subsequent work his further association with ecological and experimental studies in evolutionary genetics, and his share in the development of studies in the human blood groups, might especially be recalled. Let me also stress that, in the best and fullest sense of the phrase, I am thinking of Fisher as a working statistician. My reference in the title of this lecture to statistical methodology implies a deliberate emphasis on statistical method and technique (as distinct, say, from mathematical probability or the finer details of mathematical rigour), as the field where Fisher was for so long pre-eminent. Of course, as we shall see, we cannot separate statistical methodology from the theory of statistical inference: but it is sometimes advisable, when we find ourselves getting over-excited about the more controversial points of inductive

* Text of the first R. A. Fisher Memorial Lecture given in Chicago on December 29, 1964 at a joint session of the American Statistical Association, the Institute of Mathematical Statistics and the Biometrics Society. The lecture was supported by the Army Research Office, Office of Naval Research, and Air Force Office of Scientific Research by Contract No. Nonr-2121(23), NR 342-043.

logic, to remember the extent of the common and permanent body of statistical techniques now available to us, techniques which if they are to have proper scientific status should be as far as possible independent of the particular philosophy of the statistician practicing them.

2. FISHER AND STATISTICAL INFERENCE

Fisher's contributions to statistics began with a paper in 1912 advocating the method of maximum likelihood for fitting frequency curves, although the first paper of substance was his 1915 paper in *Biometrika* on the sampling distribution of the correlation coefficient. The stream of statistical papers which followed, especially after he had been appointed as statistician at Rothamsted, can perhaps be divided into three main lines, though of course all interconnected and exemplifying Fisher's general outlook, which I will comment on further in a moment. The first of these three lines consisted of the spate of solutions of exact sampling distribution problems, for which Fisher's geometrical approach proved so powerful. The second was the development of a more general and self-contained set of principles of statistical inference, especially for estimation problems. The third was the emergence of a precise technique of experimental design and analysis.

Now the second line did involve the systematic study of the large-sample or asymptotic properties of maximum likelihood estimates of unknown parameters, a study which is obviously classifiable under asymptotic or large-sample theory. Nevertheless, Fisher was also introducing even here concepts applicable to small samples such as information, likelihood and sufficiency; and by and large, his aim to provide a logically and mathematically precise theory of statistical inference in all its aspects seems fairly clear-cut. His ultimate degree of success I will come to later, but there is no question of the tremendous and immediate impact that so many of his results had because of their practical importance and value in statistical methodology. His consideration of small-sample theory, especially of exact small-sample distributions, was of course not new. The distribution for '*t*' had been correctly conjectured by W. S. Gosset ('Student') in 1908, and the problem of the correlation coefficient distribution was, while previously unsolved, already posed. In fact Fisher at the beginning of his 1915 *Biometrika* paper refers not only to Student's 1908 paper on the '*t*' distribution, but to a subsequent paper of Student's in the same volume on the distribution of the correlation coefficient where the exact result, in the null case, was also correctly conjectured. Nevertheless, it seems fair to say that statistical methods were mostly only available, at least in regard to the assessment of their accuracy, in the large-sample sense; and it will be recalled that the use of Karl Pearson's goodness of fit criterion χ^2 was necessarily rather fuzzy until the degrees of freedom controversy had been resolved by Fisher. This comprehensive tidying-up of exact distributions was, however, only a concomitant of his assault on the principles of statistical inference in general. Here of course he was on much more controversial ground. The theory of probability, on which any statistical principles must hinge, was still considered shaky both logically and mathematically. The mathematical foundations of probability theory were hardly satisfactorily formulated until 1933 by Kolmogorov. When

Fisher wrote his 1925 paper on estimation theory, he was obliged to insert an introductory paragraph on his definition of mathematical probability. This, in terms of infinite hypothetical populations, was a bit crude. Yet it was the minor issue, as no one doubted that the rules of probability could be successfully applied to statistical phenomena. The logical question was much more fundamental. How far could one isolate the inference problems in statistics from the inference problems in science, or indeed in everything?

It is to Fisher's credit that he succeeded in getting some way. He tried to pose problems of analysis as the reduction and simplification of statistical data. He put forward his well-known concept of amount of information* in estimation theory, such that information might be lost, but never gained (only extracted), by analysis. His concept has been of great practical value, especially in large-sample theory. His attempt to extend its relevance to small samples, by considering the case of an accumulation of small samples, was ingenious, but still strictly tied to large samples; it was rather curious that he did not notice the result later discovered by various workers and known as the Cramér-Rao inequality. But this result is still tied to mean square error, an arbitrary criterion in small samples, so that in any case small-sample theory is left more open. On the reduction of data, Fisher's concept of sufficiency was relevant both for estimation and testing purposes; though (i) we do not always have sufficiency (ii) even if we do, we still have to make an inference from our sufficient statistic. While Fisher was strongly critical of inverse probability methods (the so-called Bayesian approach), and rightly emphasized the relevance of the likelihood function as far as the data were concerned, he did not always make it clear exactly how much he was claiming—on the fundamental issue of induction itself I always found his writings extremely cryptic. Moreover if others attempted to expand or develop incomplete parts of the theory of statistical inference, such as Neyman and Pearson with their general theory of testing statistical hypotheses, he was downright rude! Yet the power of a test as introduced by these authors was a valuable tool in studying the statisticsl properties of tests in general; and often a very salutory reminder that a negative result in a significance test from a small sample might not imply as much as experimenters versed in Fisherian methods were sometimes inclined to believe.

It now seems convenient to group tests broadly into two classes; those that are merely 'goodness of fit' tests, involving perhaps many degrees of freedom, and not necessarily very sensitive to particular departures from the null hypothesis i.e. hypothesis under test; and those that refer specifically to one parameter, or at most a few parameters, and often better formulated as problems of confidence intervals (or regions) for the parameter (or parameters) concerned. Such intervals, if efficiently calculated, indicate automatically the accuracy with which a parameter is estimated, and include in effect a test of an entire range of possible values of the parameter. Unfortunately, this theory of confidence intervals, as developed by Neyman, is no longer synonymous with the theory of fiducial intervals, as developed by Fisher. I say 'unfortunately' because I regard this divergence as a regrettable red herring in the more per-

* $I(\theta) = E\{(dL/d\theta)^2\}$, where $L = \log p(S|\theta)$ is the logarithm of the probability of a random sample S when an unknown parameter has true value θ.

manent controversy between Bayesian and non-Bayesian methods. If I linger over it for a while, this is not to imply I want to over-emphasize its significance in the evolution of statistical methodology, but because Fisher, who felt strongly about it, did not argue fairly with his critics on this issue and as one of the founder members of this particular group I consider I have some right to comment. I suggested a moment ago, as indeed did John Tukey at the 1963 I.S.I. meeting at Ottawa,* that these arguments in the higher flights of statistical logic ought not to make much difference to our final conclusions in any particular statistical analysis (assuming of course that the analysis is based on a reasonable amount of statistical data). Why then do we bother with any of these more controversial issues? Are they not practically irrelevant to the development of statistical methodology and technique? Fortunately yes, to a large extent; but in a deeper sense, all of us tend to be affected in our more immediate tasks by our mental attitudes and general philosophies. We cannot therefore expect to divorce statistical methodology entirely from our philosophy of induction; I do hope, however, that we can keep a sense of perspective. I have no expectation of resolving the wider issues by anything I can say in this lecture, and do not propose to do more than remind you of them before I finish. First, however, I shall digress for a moment on the narrower issue of fiducial probability.

3. FIDUCIAL INFERENCE

Let me recall the situation in 1930. Fisher had just published a paper in the Proceedings of the Cambridge Philosophical Society entitled "Inverse probability." It was in this brief note that Fisher defined a 'fiducial interval' for an unknown parameter θ, with a known 'fiducial probability' that the true value of the parameter was covered by the interval. Discussing the case of an unknown correlation coefficient ρ, he said '. . . if we take a number of samples . . . from the same or different populations and for each calculate the fiducial 5 per cent value for ρ, then in 5 per cent of cases the true value of ρ will be less than the value we have found' (loc. cit., p. 535). Fisher's wording clearly implied, and it was the implication accepted at the time, that the fiducial interval calculated in terms of the sample statistic r was a random interval, and fiducial probability a statistical probability with the usual frequency interpretation. Such an interpretation was even current at the time outside professional statistical circles, for compare the remarks made by Eddington (who would have been in touch with Fisher's work) in his *New Pathways in Science* (1935, p. 126): 'We can never be sure of particular inferences; therefore we should aim at a system of inference that will give conclusions of which in the long run not more than a stated proportion, say $1/q$, will be wrong.'

At this date fiducial intervals were a particular class of 'confidence intervals,' particular because Fisher restricted his theory to exact intervals calculated from density functions, and to sufficient statistics. From the standpoint of inductive logic it must be stressed that the solutions can only be formulated in terms of a *hypothetical* statistical framework. Nevertheless, as a formal alternative to any Bayes solution it was a precise, ingenious and useful statistical technique.

* See p. 941–4, Linnik *et al.* (1964).

It might be noticed that Fisher's original example was not too happy, as nuisance parameters were strictly involved, and Fisher never himself examined in detail the problem of sufficiency in the case of more than one parameter. In the notorious extension to the difference in means of two samples from normal populations with different variances, I personally believe that Fisher made a straightforward mistake due to thinking of s^2 as a sufficient statistic for σ^2, and hence* s_1^2/s_2^2 as a sufficient statistic for σ_1^2/σ_2^2 (if this had been true, the Behrens-Fisher solution of the problem would have been acceptable on the orthodox frequency interpretation, and no controversy would have arisen). As circumstantial evidence I might point to Fisher's expression† for the information on a mean from a normal sample of size N with unknown variance viz. $N^2/[(N+2)s^2]$, an expression which has an infinite average in repeated samples from the same population for $N\leq 3$ and which, as an information formula, I regard as meaningless. In my original critical paper (1936a) on the Behrens-Fisher problem, I tentatively proposed the information formula

$$\frac{(N-2)^2}{(N-1)s^2+N(\bar{x}-m)^2},$$

with an average in repeated sampling from the same population of $(N-2)/\sigma^2$. The fireworks that followed the publication of my paper unfortunately annihilated, at least for some time, any further discussion of this separate problem. This to my mind is a pity as after nearly 30 years this problem is still to my knowledge not entirely resolved.

The trouble arises through the factorization of the probability function into two exact factors which nevertheless involve the unknown mean m. It follows that the usual dropping or cancellation of the differential element in the likelihood function or ratio is not immediate, and this creates some ambiguity in the proper definition of the appropriate likelihood function. The definition I adopted led to the splitting of the log likelihood equation for several samples $i=1 \cdots n$ from populations with common mean m but differing variances σ_i^2 into two portions as follows:

$$\frac{dL}{dm}=\frac{dL'}{dm}+\frac{dL''}{dm}=0$$

where

$$\left\{\begin{aligned}\frac{dL}{dm}&=\sum_{i=1}^{n}\frac{N_i(\bar{x}_i-m)}{\sigma_i^2},\\ \frac{dL'}{dm}&=\sum_{i=1}^{n}\frac{(N_i-2)N_i(\bar{x}_i-m)}{(N_i-1)s_i^2+N_i(\bar{x}_i-m)^2}.\end{aligned}\right.$$

These equations are consistent with the theory of information, but led to the paradox that samples of 2 or less contained no usable information. I do not think this is necessarily absurd when we remember that a very technical defi-

* This claim, which I consider an error, is repeated by Yates (1964, p. 347).

† See Fisher (1935b). §74.

nition of information is being used. The mean $\bar{x}_i$ still contains the amount of information N_i/σ_i^2, whether we know σ_i^2 or not; the question here is what fraction of this can be combined with that from other samples in an equation providing an estimate with precise and optimum accuracy. (For further discussion relating to this problem, see Neyman & Scott (1948) and James (1959).)

Coming next to the difference in means problem itself, I pointed out at the time that with two observations $x_{i(1)}$, $x_{i(2)}$ for each of two samples ($i=1, 2$) either of the statistics

$$t' = \frac{x_{1(1)} + x_{1(2)} - x_{2(1)} - x_{2(2)}}{|x_{1(1)} - x_{1(2)} + x_{2(1)} - x_{2(2)}|}$$

$$t'' = \frac{x_{1(1)} + x_{2(2)} - x_{2(1)} - x_{2(2)}}{|x_{1(1)} - x_{1(2)} - x_{2(1)} + x_{2(2)}|}$$

provided a t-quantity for testing $m_1 = m_2$ with one degree of freedom. Both Fisher and Yates have objected to the element of choice in t' and t''. I agree this is a weakness, but my purpose was not to put forward the best test, but a valid test serving to refute the Behrens-Fisher solutions (which takes in effect the less divergent of t' and t'' as a t-quantity with one degree of freedom). The test I actually proposed and used amounted in this case to using an estimate of $\sigma_1^2 + \sigma_2^2$ by taking the mean square of the denominators of t' and t'' and assigning it an unknown number of degrees of freedom between 1 and 2, and hence certainly as great as 1. (If $\sigma_1^2 = \sigma_2^2$, it would have 2 d.f., whereas if either σ_1^2 or σ_2^2 were zero it would have only 1.) This proposed test, being both valid and superior to the use of t' or t'' emphasized even more, in my opinion, the 'inefficiency' of the Behrens-Fisher solution.

Yates (in the paper cited) discusses the case

$$x_{1(1)} = 10\cdot 1, \qquad x_{1(2)} = 19\cdot 7, \qquad x_{2(1)} = 16\cdot 5, \qquad x_{2(2)} = 26\cdot 2.$$

This example gives

$$|t'| = 0\cdot 67, \qquad |t''| = 129, \qquad |T| = 0\cdot 95.$$

I do not know whether the use of T provides the optimum test in some sense, but I certainly believe it superior to the Behrens-Fisher test. For such a small sample (really too small for *statistical* inferences) it seems difficult to pin down further the uncertainty in the number of degrees of freedom for T by making use of the observed variance ratio s_1^2/s_2^2, though for larger samples Welch (1947) has shown how this may be done.

Incidentally, it has been shown, for example, by Jeffreys, that the Behrens-Fisher solution is the Bayes solution for uniform and independent prior distributions for m_1, m_2, $\log \sigma_1$ and $\log \sigma_2$. This is hardly a result to be used in support of *fiducial* probability; but in any case I would not regard such prior distributions as very sensible in this problem, as they are incompatible with any finite and non-zero observed value for s_1^2/s_2^2 to be used in the solution (cf. James *loc. cit.* p. 80).

Fisher had in his 1935 paper generalized rather tentatively to the concept of

simultaneous fiducial distributions, appearing to argue largely by analogy from the case of a single normal population with unknown mean and variance, a problem where no difficulty arises. In my own paper I noted that the single population case could be extended to any joint problem of location and scaling, in the sense that valid and consistent classes of confidence intervals for either parameter, when the other was known or unknown, could be generated from the two-parameter fiducial distribution. This result was independently noted by Pitman in 1938. When valid confidence intervals do not exist, I do not consider that a case for a separate class of fiducial intervals, derived independently either of the theory of confidence intervals or of inverse probability, has been established. A good deal more has been said or written since these early papers, but I cannot see that it has taken us much further forward. This applies, for example, to the discussion on fiducial probability at the I.S.I. meeting in 1953 at Ottawa,* where indeed support for my doubts on the value of fiducial inference as such will be found, for example from Lindley and Pitman, though admittedly not in every instance from the same side of the Bayesian fence. This view would not of course be supported by others, such as Barnard and Birnbaum, the latter writer's notion of 'intrinsic confidence intervals' being apparently close to Fisher's fiducial intervals. But I do not feel disposed to accept further concepts unless they have some demonstrable practical value. I can see some pros and cons in the case of the Bayesian, non-Bayesian, controversy, depending on one's attitude to the relative status of statistical probabilities and probabilities as degrees of belief. Both these distinct probability concepts have, whatever our personal predilections, considerable acceptance and recognition, both from statisticians and philosophers; yet more probability concepts, with no very clear interpretation, are hardly something we can welcome.

4. BAYESIAN INFERENCE

Having digressed for too long on fiducial inference, I hope I shall not do the same with Bayesian inference. However, some comment is unavoidable, as, in spite of Fisher's onslaught on inverse probability during his lifetime, an attack with which, with some reservations, I concur, the Bayesian approach is still very much with us. Indeed, from the efforts of Jeffreys, Savage, Good, Lindley and others, its status has been growing of late. This has partly been due to the development of decision theory and the introduction of loss functions. Nevertheless, as I see it, there are one or two fundamental issues in the use of Bayesian methods in statistical inference that are too often ignored or played down by one side or the other.

I have reminded you that there are two distinct probability concepts. One, the statistical, is narrower in scope but more precise within its proper context than the other, the notion of degree of belief.† The subject of statistics has not only an almost universal range of application, but its meaning has also tended to broaden from the orthodox one of being concerned with population or group

* See Linnik *et al.* 1964.

† In answer to a query by F. J. Anscombe after my lecture on what I mean by this remark, perhaps I could refer to my note (1936b), the last sentence of which concludes: '(Statistical) probabilities may be said to exist objectively in the usual and necessary sense—that they are theoretically measurable, and sufficiently well substantiated by experiment.'

phenomena. I think this last trend has been rather confusing. Statistical phenomena have their own properties associated with laws of large numbers and ergodic theory that are quite separate logically from our overriding and constant necessity to be making decisions and inductions in any field, whether it is statistical or not. If a comprehensive and unique theory of induction were acceptable for all phenomena, as some Bayesians would claim, then statistical phenomena would naturally be dealt with in the same way. There is plenty of evidence (which as good statisticians we must not suppress) that this ideal situation does not exist. As this is so, it is open to the statistician to analyse his material in a manner which he thinks useful and explicable both to himself and to others; and to this end he has developed various techniques and methods especially designed to reduce and simplify statistical data. Further than this I do not think we can go. It is not surprising that the statistician in the search for objective results has often formulated them in terms of statistical probabilities, for example, statistical confidence in conclusions—the so-called 'behaviouristic' approach, as Neyman has put it. Certainly, this type of procedure has its limitations and dangers, but there is a fundamental impasse here that should be clearly stated—I do not think, for example, that Fisher was ready to admit it, although, somewhat ironically, the small-sample theory which Fisher helped so much to develop rather brought it into the open. The statistical approach cannot by its nature deal with the unique sample—it must contemplate statistical variation which often from the Bayesian angle is irrelevant and a source of inefficiency. The statistician can attempt to reduce the effect of irrelevant variation by conditional inferences and the like, but he cannot eliminate all questions of sampling variability or he has no probability distribution to work with at all. But the Bayesian cannot deal with the unique sample either except by moving into a different field of discourse, the quantitative aspects of which are debatable. There are various distinguishable schools. The one including Good and Savage assign personal degrees of belief and utilities to relevant propositions. This seems to me of possible value to an individual, for example, in business, assessing his own different courses of action. I am not convinced, however, that it can claim scientific validity unless the degrees of belief can be generally accepted and hence idealized. This is the standpoint of Jeffreys and possibly of Lindley. It is, for example, not unreasonable to assign a uniform conventional prior distribution to denote prior ignorance in the case of a finite and discrete set of alternatives. This is in line with information theory in the communication theory sense (i.e. Shannon's concept, not Fisher's). This is, I believe, as far as the logician Rudolf Carnap got; and there are considerable difficulties with any more complicated but equally common prior distribution problems. It seems in any case meaningless to me to claim quantitative induction in a unique situation. Some common features with other similar problems are recognized. Indeed, I think Reichenbach went so far as to claim that no inductions could be formalized except on a frequency basis. Concentration on the unique sample can be carried too far. Yet once some behaviouristic or frequency justification of inductions is sought, the Bayesian has an additional problem to contend with. It is then necessary to distinguish between his assigned prior distribution and the true prior distribution, which certainly exists in some

cases in a frequency sense. Of course, in some sufficiently well-defined statistical situations a whole process can be studied as one problem, and the so-called empirical Bayes' procedures of Robbins seem to fit in here. But in a typical problem, say in decision theory, whereas most statisticians would grant that with the right prior distribution (in the frequency sense) and loss functions, the Bayes approach (including loss functions) would be the optimum one, and therefore certainly worth formulating, it is not always made clear what happens if, as is more likely, the wrong quantities are used.

The Behrens-Fisher test seems to afford an example where the use of a criticizable prior distribution leads to a solution inferior to other possible tests. Of course decision theorists have been aware of this problem ever since Abraham Wald developed decision theory and suggested the use of the minimax principle; but neither he nor later workers were very satisfied with this particular suggestion.

At present there is no final reconciliation of the Bayesian and non-Bayesian approaches; but at least the situation is hopeful in that there seem signs of less dogmatism, and more appreciation and tolerance of these respective viewpoints. I think even the Bayesians have appreciated some of Fisher's concepts such as sufficiency, likelihood and information, although they may be less relevant to their own approach (cf., however, Pratt, 1964) and although, like some of the rest of us, they may query the point of any of them, if put forward as concepts in their own right without the need for some acceptable probability interpretation. Some years ago (in my Inaugural Lecture at University College) I expressed the view that a final resolution of the inverse probability controversy must await a resolution of the question how much objectivity to assign to science and scientific theories. We may now admit Karl Pearson's thesis that scientific laws are man-made, but this does not prevent our proceeding on the lines that there is an objective world of phenomena to be described. A statistician who sets up a statistical model to be tested or whose parameters are to be estimated is proceeding along a well-known scientific path described by R. B. Braithwaite as following the hypothetico-deductive method. My own inclination is to follow this path. My impression is that it was Fisher's inclination also, at least in much that he did; but his claims for some kind of absolute validity and objectivity for some of his concepts such as fiducial inference seem to me rather to have fogged the issue.

There is some interest in one of the most recent (1962) notes of Fisher's in which he tried to show that inverse probability could in suitable cases be formulated in unequivocal terms. The type of example he took was one where particles are being emitted randomly at an unknown rate, and the investigation was arranged in two parts, the first part consisting of the random time to the first emission, and the second part containing some further statistical information on the same unknown parameter. Fisher formulated his solution in the terms of inverse probability, but as far as I can see he was merely combining two independent pieces of statistical information and no new situation has been engineered. His discussion merely seems to me to emphasize a failure in his later years to make it clear to others what his probability inferences for an unknown parameter meant. As he had by this time rejected the frequency interpretation the only other recognized interpretation is a degree of belief, but if so interpreted his formulation has to compete with more orthodox Bayesian approaches. In spite of some obvious advantages if an acceptable system of inference in terms of the statistical data alone could be formulated, I can see no evidence that Fisher's later point of view has really been helping the development of practical statistical methodology, and this to me at least is perhaps the most important criterion by which to judge.

I am afraid I have after all spent rather longer on these controversial issues than is perhaps justified, bearing in mind my expectation that they will remain with us for some time—certainly I have been longer than I originally intended in a discussion primarily on statistical methodology. To recapitulate, it has been suggested that the basic issue is on Bayesian versus non-Bayesian methods, and that Fisherian variants on the non-Bayesian approach should not obscure this. It has, however, also been pointed out that there are at least three Bayesian approaches, the individualistic or personal approach of Savage, the epistemological approach of Jeffreys, and the prior frequency approach often attributed to Karl Pearson. In earlier writings on this subject I have where clarity demanded it used different probability notation for different concepts, and I at least would find writings by others clearer if the same practice were followed, especially now that the different probability concepts have multiplied.

DIFFERENT PROBABILITY CONCEPTS WITH POSSIBLE NOTATIONS*

Concept	Bartlett (extended)	Carnap (extended)
1. Rational degree of belief (epistemological approach e.g. Jeffreys)	$\underline{P}$	$\underline{p_1}$
2. Statistical probability or chance (assumed)	$\underline{p}$	$\underline{p_2}$
3. Degree of belief (personal or individualistic approach e.g. Savage)	iP	ip_1
4. Estimated or guessed, e.g. prior, probability (frequency approach)	ep	ep_2
5. Fiducial probability (where equivalent confidence probability)†	$\underline{fp}$	fp_2
6. Fiducial probability (Fisher)	fP	fp_1

* Underlining denotes previous usage.

† In answer to a referee, confidence probabilities that are not fiducial probabilities might where necessary be denoted by cp (or cp_2).

It will be a relief now to turn to less debatable issues. Could I just note first, however, how our basic approach may unconsciously affect our attitude to particular branches of statistical methodology. There has, for example, been invaluable work done in recent years on 'non-parametric methods.' These methods are often much more 'robust' against the correctness of background assumptions about the model used, and therefore have an important rôle to play in modern statistical methodology; but I would be reluctant to see the statistician relinquish his responsibility to set up explicit statistical models which represent, as far as he can manage, the situation he is investigating and analysing. Indeed, with the coming, say, of operational research, or of mathematical biology, the use of models has been spreading, and their use by the statistician becoming more diverse and intricate.

5. THE DEVELOPMENT OF EXPERIMENTAL DESIGN AND ANALYSIS OF VARIANCE

Fisher's quest for precise methods of statistical inference led him to make tremendous advances in the technique and analysis of statistical experiments, his contributions in this general field being as important practically as anything else he did in statistics. From his day-to-day contact with agricultural experiments at Rothamsted, Fisher came to realise the essentials of good experimental design. He perceived the simultaneous simplicity and efficiency of balanced and orthogonal experimental designs. If any statistical assessment of error was to be possible, replication was of course necessary, and equal numbers of plots per treatment optimised the design and greatly simplified the 'analysis of variance' technique that was developed to go with it.

It was at this point that Fisher introduced one vital principle. When statistical data are collected as natural observations, the most sensible assumptions about the relevant statistical model have to be inserted. In controlled experimentation, however, randomness could be introduced deliberately into the design, so that any systematic variability other than due to the imposed treatments could be eliminated. This has been an invaluable device in practical experiments and sampling surveys of all kinds. Incidentally, it is a device that has been a source of some logical difficulty to the orthodox Bayesian (see Savage *et al.*, 1962, especially pp. 87–91).

The second principle Fisher introduced naturally went with the first. With the statistical analysis geared to the design, all variability not ascribed to the influence of the treatments did not have to inflate the random error. With equal numbers of replications for the treatments each replication could be contained in a distinct block, and only variability among plots in the same block were a source of error—that between blocks could be removed. This principle, like the first, is also of course of extreme importance in the design of sampling surveys.

The third principle Fisher introduced was in connection with treatment combinations of more basic factors, such as the testing of combinations of the three primary fertilizer ingredients, nitrogen, phosphate and potash, in agricultural trials. Fisher emphasized the gain by testing all combinations and breaking down the analysis into the separate degrees of freedom for main effects, first-order interactions and so on. This technique of factorial experimentation cut right across the current practice which Fisher criticized in his book on The Design of Experiments (1935b; p. 96) referring to the 'excessive stress laid on the importance of varying the essential conditions *only one at a time*.' He showed how the individual factors could be numerically assessed in the *absence of interaction* as if the other factors were absent, with, moreover, a wider inductive basis for the assessment. Moreover, if interaction were present, information was obtained on its effect that could not have been obtained by separate testing of the individual factors.

It will be noticed that the value of these methods was not dependent on the statistical analysis, but the simplicity and clarity of the relevant analysis, which Fisher emphasized was dictated by the design, greatly contributed to their rapid world-wide popularity. The analysis was basically classical least-squares analysis, but the orthogonality of the design rendered the estimation problem trivial. The associated estimation of error was systematized by the technique of 'analysis of variance,' perhaps a slightly unfortunate title as the analysis mainly consists of a numerical breakdown of the total sum of squares of the observations into its relevant additive parts. Once the technique had evolved (which did not happen overnight) and the appropriate significance tests were made available from statistical tables, based on Fisher's derivation and tabulation of the variance-ratio distribution, more complicated least-squares problems, such as non-orthogonal designs or multiple regression analysis, could also be dealt with.

It might be remarked that all this technical advance in experimental design had its dangers. The design and analysis were logically more tied to the null hypothesis of no treatment effects than to any alternative, and this gave somewhat undue importance to the rôle of the significance test. The *validity* of the methods for small samples was sometimes confused with their *sensitivity*. The complications that might ensue once the null hypothesis was false were not always followed through. The additive set of interactions in factorial experiments was always logically correct, but less practically relevant in some contexts than in others. Such limitations are, however, not dissimilar logically from those in other statistical fields of analysis. It is a tremendous practical gain to have simple, efficient routine methods of analysis, always provided that they are not elevated to a blind ritual.

6. MULTIVARIATE ANALYSIS

One of the useful extensions of analysis of variance technique for the analysis of experiments was the so-called analysis of covariance technique for adjusting final observations by initial ones made before treatments have been imposed. This technique is of one personal interest to me as being, back in 1933, the first instance I had that Fisher, like the rest of us, could err, in this case at least by implication. At that time he had merely given the adjustment without stressing that the ensuing non-orthogonality necessitated further analysis if an exact test of significance of treatment effects was required.

The simultaneous analysis of variance and covariance is of course something different, being an example of multivariate analysis as more usually defined. Fisher shares primarily with Hotelling in the United States and Mahalanobis in India the distinction of developing this technique, which has with the greater availability of computers been growing of late in importance. I have occasionally noticed a tendency to denigrate Fisher's astonishing power of geometrical reasoning, which I mentioned earlier, and which was to assist him in obtaining so many solutions of sampling distribution problems. It was the same approach that enabled Wishart to obtain with Fisher's guidance the so-called Wishart

distribution. It was the same approach too that enabled Fisher in 1928 to obtain the general distribution of the multiple correlation coefficient, a derivation that few people understood. I remember when lecturing in Cambridge before the war preferring Wilks' analytical derivation for this very reason. When I returned after the war, J. O. Irwin happened to raise the matter again in conversation, and I returned to Fisher's orginal derivation. This time I was delighted that I could follow it, and at once realized that his argument could be extended in principle to the general distribution of canonical correlations, a problem with which I was in consequence able to make some headway.

It was always something of a surprise to me that Fisher did not himself continue to make full use of geometrical argument in establishing the sampling theory for multivariate analysis. The sampling theory of his linear discriminant function was implicit in Hotelling's earlier work on the multivariate extension of the t-test, but the effect of eliminating a hypothetical discriminant function (or even a hypothetical first canonical variate in the more complex sampling problems of canonical correlation analysis) could also be studied by means of the reciprocal sampling relations between two sets of variables. Fisher, by attempting to proceed rather formally by pseudo-analysis of variance technique, was led at times into definite errors.*

7. TIME SERIES AND STOCHASTIC PROCESSES

One man, even of Fisher's calibre, cannot of course maintain the same level in all relevant areas. In retrospect, one can attempt to see Fisher's work against the general scientific background; and my own impression is of one rather serious omission in its coverage. His comparative neglect of the important Continental work on the foundations of mathematical probability I have already suggested was on the whole not practically very important. This neglect, however, extended to developments in the theory of random or stochastic processes, even when such developments were not of the more fundamental type associated with such contemporaries as Kolmogorov, Khintchine and Slutsky in the U.S.S.R., or Wiener, Feller and Doob in the U.S.A., but were associated with the more immediate problems of statistical theory and analysis such as the work in his own country of G. U. Yule on time series and A. G. McKendrick on stochastic models in biology.

This was particularly surprising in view of Fisher's active interest in stochastic processes in genetics. Thus Fisher's papers on the statistical analysis of data recorded in time included some interesting work on orthogonal polynomial fitting, but a rather rigid adherence to classical least-squares procedures. His 1929 paper on the 'Studentisation' of the classical significance test of a strict harmonic component (of unknown periodicity) came later than Yule's famous 1927 paper on autoregressive models for stationary time-series, but no discussion of the repercussion of Yule's work on the whole subject of periodicity in time-series, or indeed any reference at all to it, was made. It should not be necessary to remind you of the tremendous progress made with time-series analysis since then, progress which is making this subject one of the most rapidly developing branches of statistical methodology at the present time.

* *Cf.* for example, Fisher's discussion in *Statistical Methods for Research Worker's*, 10th Ed. (1946), Example 46.2 and my own (Bartlett, 1951-2, §4).

The work of McKendrick was less directly connected with statistical methodology, but fundamental in the development of stochastic models in biology and medicine. Moreover, it had indirect relevance to the study of 'contagious-type' distributions by Greenwood, Yule, Polya and Neyman. Here Fisher was on common ground with these other workers in the overlooking of McKendrick's work, which included, for example, the derivation in 1914 of the negative binomial distribution as a contagious-type distribution.

The problems of statistical analysis and statistical methodology arising in the general area of stochastic processes are very important and complex, as I have indicated already elsewhere (1959). They are certainly not confined to the problems of time-series analysis in the usual sense, and are rapidly growing as the use of stochastic process models develops in biology, economics, industry, medicine and psychology.

8. CONCLUDING REMARKS

It would be foolish of me to pretend I could adequately assess all the contributions to statistical methodology made during the last fifty years. I have not even mentioned yet the brilliant work done by Abraham Wald during the last war on the principles of sequential sampling. There are plenty of important developments on the more strictly practical level, such as the great development in the principles and practice of sampling surveys, or the development of cohort analysis in demographic studies. I am sure I shall be accused, especially on this side of the Atlantic, of underemphasizng the rôle of decision theory in modern statistics, my rather cursory references to it being associated with my view that (i) it is not properly classifiable as statistics, as I understand this term (I feel tempted to say it is more a way of life!), (ii) it involves concepts such as loss functions and prior probabilities which I find quantitatively to be of dubious practical value in the study of statistical phenomena, at least in the scientific field.

Obviously I have, while ranging to some extent outside Fisher's own activities, made these the starting point of this survey. With regard to his own work, my somewhat protracted preoccupation with the more debatable issues will be disliked by his more dedicated admirers. The trouble with great men, especially those with temperaments of comparable stature, is that they are liable to excite either allegiance or rebellion. This does not facilitate an objective judgment. However, let me recall Professor M. Fréchet's words: (1963, p. 169):

"Les statisticiens du monde entier savent quelle dette ils doivent à l'école statistique brittanique, et, en particulier, aux deux grands savants qui ont, l'un créé, l'autre transformé la statistique mathematique, Karl Pearson et Sir Ronald Fisher."

Fisher would no doubt have thought Fréchet only half-right; but for ourselves, we do not have to accept Fisher's complete infallibility in order to recognize his greatness as a scientist, and like Fréchet, acknowledge the permanent debt which we all as statisticians owe him.

REFERENCES

Bartlett, M. S. (1936a) "The information available in small samples." *Proc. Camb. Phil. Soc.* **32,** 560–6.

—— (1936b) Statistical probability. *J. Amer. Stat. Ass.* **31,** 553–5.

—— (1951–2) "The goodness of fit of a hypothetical discriminant function in the case of several groups." *Ann. Eugen.* **16,** 199–214.

—— (1959) The impact of stochastic process theory on statistics (published in *Probability and Statistics*, Stockholm, pp. 39–49).

—— (1961) "Probability, statistics and time." (Inaugural lecture, University College, London.)

Birnbaum, A. (1962) "On the foundations of statistical inference." *J. Amer. Stat. Ass.* **57,** 269–326.

Eddington, A. S. (1935) *New Pathways in Science.* (University Press, Cambridge.)

Fisher, R. A. (1912) "On an absolute criterion for fitting frequency curves." *Messenger of Mathematics.*

—— (1915) "Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population." *Biometrika* **10,** 507–21.

—— (1925) "Theory of statistical estimation." *Proc. Camb. Phil. Soc.* **22,** 700–25.

—— (1928) "The general sampling distribution of the multiple correlation coefficient." *Proc. Roy. Soc.* A **121,** 654–73.

—— (1929) "Tests of significance in harmonic analysis." *Proc. Roy. Soc. A.* **125,** 54–9.

—— (1930a) *The Genetical Theory of Natural Selection.* (University Press, Oxford.)

—— (1930b) "Inverse probability." *Proc. Camb. Phil. Soc.* **26,** 528–35.

—— (1935a) "The fiducial argument in statistical inference." *Ann. Eugen.* **9,** 174–80.

—— (1935b) *The Design of Experiments.* (Oliver and Boyd, Edinburgh.)

—— (1946). *Statistical Methods for Research Workers* (10th Ed. Oliver & Boyd, Edinborough.)

—— (1962) "Some examples of Bayes' method of the experimental determination of probabilities *a priori.*" *J. R. Statist. Soc.* B. **24,** 118–24.

Fréchet, M. (1963) See obituary of Sir Ronald Aylmer Fisher, 1890–1962. *J. R. Statist. Soc.* A. **126,** 159–78.

James, G. S. (1959) "The Behrens-Fisher distribution and weighted means." *J. R. Statist. Soc.* B. **21,** 73–90.

Kolmogorov, A. N. (1933) *Grundbegriffe der Wahrscheinlichkeitrechnung.* (Springer, Berlin).

Linnik, Yu. V. and other contributors. (1963) "Fiducial probability." *Bull. I.S.I.* **40** (vol. 2) 833–939.

McKendrick, A. G. (1914) "Studies on the theory of continuous probabilities with special reference to its bearing on natural phenomena of a progressive nature." *Proc. Lond. Math. Soc.* (2) **13,** 401–16.

—— (1926) "Applications of mathematics to medical problems." *Proc. Edin. Math. Soc.* **44,** 98–130.

Neyman, J. and Scott, E. L. (1948). "Consistent estimates based on partially consistent observations." *Econometrica.* **16,** 1–32.

Pitman, E. J. (1938) "The estimation of the location and scale parameters of any given form." *Biometrika,* **30,** 391–421.

Pratt, J. W. (1965) "Bayesian interpretation of standard inference statements." *J. R. Statist. Soc.* B, **27.**

Savage, L. J. and other contributors. (1962) *The foundations of statistical inference.* (Methuen, London.)

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

—— (1908b) "Probable error of a correlation coefficient." *Biometrika,* **6,** 302–10.

Welch, B. L. (1947) "The generalization of 'Student's' problem when several different population variances are involved." *Biometrika,* **34,** 35–8.

Yates, F. (1964) "Fiducial probability, recognisable sub-sets and Behrens' test." *Biometrics,* **20,** 343–60.

Yule, G. U. (1927) "On the method of investigating periodicities in disturbed series, with special reference to Wolfers's sunspot numbers." *Phil. Trans.* A. **226,** 267–98.