Statistical Methods in Scientific Inference

by

A. W. F. EDWARDS

Gonville and Caius College, Cambridge

Examination of the conflicting statistical methods currently used in scientific inference reveals an increasing awareness of the utility of likelihood. The concept of prior likelihood is introduced as a means of completing a scheme of inference which does not share the logical disadvantages of other methods.

It is remarkable that, in a century which has seen such a large growth in the application of statistics to the natural sciences, the fundamental issues of statistical inference have not been resolved. There are not many more statisticians than opinions as to how to assess rival hypotheses in the light of data, and although what follows is not a review, a short account of the more influential points of view is necessary as an introduction to a new scheme of inference which, I believe, supplies precisely those elements which many hold to be essential without any of the features to which many object.

First, it is necessary to delimit the field of enquiry by defining "statistical hypothesis", which I take to be a specific probability model for the conceptual generation of observations, adopted so that it can be compared with other hypotheses in the light of a particular set of observations. The essential statistical feature is that it is possible to attach a probability to each of the possible outcomes, and its peculiarity is that it cannot be directly verified. I thus exclude from consideration a hypothesis such as "the next ball drawn from this bag will be black", for it does not, as specified, have statistical consequences, and is directly verifiable when the draw has been made. An event such as drawing a black ball may be called an "outcome", and it has long been recognized[1] that the probability of realizing a particular outcome is—when it can be calculated—a rational measure of belief that the outcome will occur. Much of what has been written about hypotheses concerns outcomes rather than statistical hypotheses. Occasionally a hypothesis is itself an outcome of another probability model, in which case the process of inference is much simplified.

The debate on statistical inference has been conducted with more than usual vigour, perhaps because of the unconscious uncertainty of each contributor about his own position. But the polemics of the past have been replaced by the tacit acceptance of a wide variety of usage and a widespread, if muted, concern that no single position is fully satisfactory.

At one end of the spectrum is Sir Harold Jeffreys[2], who holds that prior beliefs about hypotheses, held before any data are available, can be measured in terms of probability and so can be handled by the axioms of probability and, furthermore, that there are canonical values for such probabilities, about which ideal scientists would agree. The weighing of hypotheses, or of the different parameter-values of a model, is to be achieved by applying the axioms to the prior probabilities and to the probabilities of the data conditional on the hypotheses, thereby obtaining posterior probabilities for the hypotheses. The calculation may be conveniently made using Bayes's Theorem[3]:

$$\text{Posterior probability} \propto \text{Prior probability} \times \text{Likelihood}.$$

Here the likelihood, first formally defined by Fisher[4], is equal to the probability of the data given the hypothesis. The case of continuous data may be satisfactorily handled using the concepts of probability density and the likelihood ratio (see Hacking[5]). Jeffreys weighs hypotheses by forming the ratio of their posterior probabilities on the data. Lindley[6], a modern advocate of Jeffreys's fundamental position, lays a greater emphasis on significance-testing.

Next in the spectrum come the subjectivists, who hold that although prior beliefs may properly be measured by probability, there will be such disagreement about the values of the probabilities in any one instance that the only satisfactory method is to allow everyone his own subjective value. It is often suggested that this is to be assessed through conceptual bets. The modern use of subjective probability derives from the works of Keynes[7] and de Finetti[8], and the chief advocates of its use in statistical inference today are Good[9,10] and Savage[11]. Subjectivists hold that all beliefs may be interpreted by probability, and so Bayes's theorem is used, with the understanding that the posterior probabilities have an element of subjectivity, though how much is not always clear. This point has particularly concerned Smith[12], Good[13] and Dempster[14] who, though subjectivists, favour more complex schemes involving upper and lower probabilities. Good[10] has also advanced a scheme in which there are various levels, or "types", of probability. Thus the uncertainty about a prior probability of one type can be expressed by a prior probability distribution for the prior probabilities, and so on. Lindley[15], from his point of view, regards the "prior of a prior" as the orthodox solution to the problem of uncertainty in this context. Polya[16] is prepared to use probability only to make statements of inequality about degrees of belief.

It is customary to refer to all these positions as "Bayesian", but it should be remembered that the opinions vary from Jeffreys's counsel of perfection to Polya's inequalities. Good[10] writes that "The essential defining property of a Bayesian is that he regards it as meaningful to talk about the probability P(H/E) of a hypothesis H, given evidence E". Because this definition would almost certainly have excluded Thomas Bayes, many writers have preferred to continue with the older term "Inverse probability" for the resulting type of inference. Good continues: "An extreme Bayesian believes that every intuitive probability is precise, whereas less extreme Bayesians regard intuitive probabilities as only partially ordered so that each probability merely lies in some interval of values".

In the middle of the spectrum is Bartlett[17,18], who, following Carnap, holds that probabilities can be used to describe such beliefs, but that this type of probability must not be confused with classical frequency probability. Bartlett does not, however, develop any scheme for handling the two types jointly, but seems content to adopt the methods of significance-testing.

From this point of the spectrum onwards there is a denial that beliefs about hypotheses can be measured in terms of probability. No alternative measure is suggested, except where data alone are concerned, in which case some hold that relative likelihoods may be used. The use of likelihood for this purpose has been advocated by Fisher, particularly at the beginning[4] and end[19] of his career. (In the middle[20] he was more concerned to emphasize likelihood for its role in estimation.) Modern exponents of this view are Barnard[21-23], Birnbaum[24], Hacking[5] and also myself[25]. Hypotheses are to be weighed by likelihood ratios as far as data are concerned, but any prior opinions are to be included in the final assessment in the vague intuitive way in which opinions are combined. Publication of the likelihood surface (giving the likelihood at all parameter-values, or for all the rival hypotheses) is adopted as the best method of presenting results.

Next in the spectrum come those who view the weighing of hypotheses principally in terms of tests of significance and estimation. Fisher was much concerned with this approach, and contributed greatly to both fields. He was associated with many of the early tests, and argued strongly in favour of his fiducial probability, which seems to allow, in certain circumstances, probability statements about parameter-values to be made on the basis of data. The circumstances include an acceptance of Fisher's definition of probability and the complete absence of prior information. The statistical world is somewhat bemused by the concept. In practice, the argument has not been influential because of the difficulty or impossibility of phrasing many problems in its terms. By contrast, Fisher's influence on the theory of estimation has been enormous, largely through his advocacy[27] of the "method of maximum likelihood". Although the widespread use of the standard tests of significance and of the method of maximum likelihood is a measure of Fisher's influence, since the adoption of the method has often been largely empirical, the frequency with which they are used should not be taken as a measure of their logical validity.

Finally, at the opposite end of the spectrum from Jeffreys are the many advocates of the Neyman–Pearson[28] theory of hypothesis-testing. Hacking notes that this "is very nearly the received theory". The method is unexceptionable when it is necessary to decide absolutely between two hypotheses, for it establishes a critical value (in the case of a single variable) which divides the space of possible observations into two parts: if the observation falls in one part, one hypothesis is chosen, and if in the other part, the other hypothesis is chosen. Elegant theories aim at the choice of a critical value such that some combination of the probability of rejecting the "false" hypothesis and the probability of accepting the "true" one is optimized. This is the cornerstone of Wald's[29] sequential theory. Some Bayesian enthusiasts have added inverse probability as well, thus linking the two ends of the spectrum "round the back"[30].

Importance of Likelihood

In this literature there is abundant criticism of every point of view, for each protagonist has sought support for his own position by contrasting it with others. All but the most transparent criticisms (such as the Bayesians have sometimes suffered) seem to contain convincing elements, so that it is easy to be a sceptic who feels that there has been no firm progress towards the solution of the controversy. Yet necessity forces the practising scientist to adopt certain methods. Fortunately there is general agreement that in well behaved situations the several methods are highly correlated in effect, at least when the evidence from the data is near to overwhelming the prior opinions. Perhaps the chief dichotomy in the spectrum is not between those who measure belief in hypotheses by probability, and those who do not, but between those who seek to compare rival hypotheses by some relative measure, and those who apply tests based on the concept of repeated trials. From Jeffreys to the advocates of likelihood there is a measure of agreement that hypotheses may best be weighed against some scale of relative acceptability, rather than by significance-testing. They hold[2,5] that in this context the relevance of the significance-test, with its reliance on tail-areas, has never been satisfactorily established. Furthermore, there is general agreement that the likelihood, or likelihood function in the case of continuous parameters, is very important. It enters into all Bayesian methods in addition to providing a method in its own right. It possesses some attractive properties, which Birnbaum[24] in particular has analysed in depth.

Indeed, likelihood seems to be the one rock in a shifting scene, and the only criticisms that Bayesians have consistently levelled at the likelihood method are that it does not allow for the formal incorporation of prior beliefs, and that likelihood has no "meaning". On the latter point, the root objection seems to be that likelihood has no interpretation in terms of probability, which is precisely its merit. Birnbaum[24] and Barnard[22] have suggested various "evidential" interpretations, though it seems to me that none is needed. Lindley[31] complains that arguments based on the likelihood function alone do not allow nuisance parameters to be integrated out, a restriction which prevents logical distinctions being sacrificed to expediency.

Fisher[1,19,32–35] waged continuous war on inverse probability, echoing the doubts of Cournot[26], Boole[36] and Venn[1]. All his criticisms centred on the Bayesian's use of probability as a measure of belief. Its use, he argued, to specify prior belief in the absence of any "chance set-up" (Hacking[5]) led to the error of treating a problem as though one knew that of which one was in fact ignorant; it was not invariant to transformations of the parameters; and, in particular, the uniform prior distribution used to convey complete ignorance about a parameter led to apparently gratuitous information about any new parameter functionally related to the first. I agree with these criticisms, and developments in the logic of Bayesian inference since Fisher's death seem to have done little to meet them. As noted earlier, these developments have been designed to allow for the fact that there are differing degrees of confidence in subjective probabilities, but because these degrees are themselves measured in terms of probabilities, all of which, whatever their kind in Bayesian terms, obey the axioms of probability, a confusion of probabilities results. The confusion can only be removed by maintaining a rigorous distinction between frequency probabilities on the one hand and measures of degrees of belief in statistical hypotheses on the other, as well as between subjective and objective measures. This distinction was carefully drawn by Fisher[4,20] nearly fifty years ago in his definition of likelihood. In his last book[19] he was quite explicit:

"Apart from the simple test of significance, therefore, there are to be recognized and distinguished, between the levels of certain knowledge and of total nescience, two well-defined levels of logical status for parameters lying on a continuum of possible values, namely that in which the probability is known for the parameter to lie between any assigned values, and that in which no probability statements being possible, or only statements of inequality, the Mathematical Likelihood of all possible values can be determined from the body of observations available".

"The likelihood supplies a natural order of preference among the possibilities under consideration".

Venn[1], in 1876, had ventured the same idea: "To decide this question, what we have to do is to compare the relative frequency with which the two kinds of cause would produce such a result." Bernoulli had said as much a century before[37].

Fisher's long advocacy of likelihood was overshadowed for most of his life by his work on estimation, significance-testing, and the design of experiments, and the enthusiasm with which this was greeted. Current use of likelihood can more often be traced to a Bayesian argument (as in human genetics; Smith[38]) or to the analogy with information or entropy (as in physics), than to the direct influence of Fisher. Indeed, Barnard has been alone in his continued support for likelihood on theoretical grounds, though he has now been joined by Hacking.

Testing Hypotheses

In addition to contending with inverse probability, Fisher had repeatedly to stress the irrelevance of the Neyman–Pearson method of hypothesis-testing to the weighing of scientific hypotheses. Whatever its other merits, the concept of acceptance or rejection of alternative hypotheses is alien to their assessment in pure science. What comfort is it for a scientist, who has rejected a hypothesis in favour of its rival by the slenderest of margins, to know that if he applies the method in enough cases, the frequency with which he is wrong will in some sense be a minimum? His concern is for the inference he should draw from the data he has, not from the data he might have possessed. It seems to me that Fisher's[19] indictment of the Neyman–Pearson method as applied to scientific problems is completely convincing, and anything I add would be mere repetition. But anyone inclined to discount Fisher's writings may refer to Hacking.

Nor do I regard the classical methods of significance-testing, exemplified by the use of t, χ^2 and F, which Fisher did so much to promote, as entirely satisfactory from a logical point of view. No real explanation has yet been offered as to why one is allowed to add the clause "or greater value" so as to form a region of rejection; Fisher[19] skates round this when considering fiducial limits, and in connexion with confidence limits he admits that "This feature is indeed not very defensible save as an approximation". To what? His reaction to the point is to take refuge in the comparison of likelihoods, and it is interesting to note, as Hacking does, that Gosset ran for cover in the same direction.

And so, rejecting inverse probability as erroneous, the theory of hypothesis-testing as irrelevant, and significance-testing as no more than a valuable expedient, I am left with likelihood, coupled with a desire to incorporate my prior opinions into inference in a formal manner.

Fisher evidently never felt any such desire, or he would surely have made the proposal I now make. On the contrary, any tendency towards an overformalization of scientific induction, particularly in the direction of decision theory, made him reach for his pen[19]: " . . . the Natural Sciences can only be successfully conducted by responsible and independent thinkers applying their minds and their imaginations to the detailed interpretation of verifiable observations. The idea that this responsibility can be delegated to a giant computer programmed with Decision Functions belongs to the phantasy of circles rather remote from scientific research".

Nevertheless, the numbers of those who, while rejecting this notion, feel the need for a scheme by which prior belief can be formally incorporated into statistical inference may be gauged from the strength of the Bayesian school today, many of whose adherents seem to be prepared to smother their doubts about inverse probability for the sake of introducing prior belief. To adopt alternative means for achieving this is not to assert that the whole of statistical inference can be conducted with them, but it does allow some of the simpler problems to be handled more satisfactorily than before, which, in turn, may be expected to lead to a greater understanding of statistical inference as a whole.

Prior Likelihood

These means may be provided within the framework of likelihood, and without embracing the enigmas of inverse probability, by adopting the concept of prior likelihood. For if, as Fisher asserts, we can measure our relative belief in rival hypotheses or parameter values, contributed by a set of observations, by relative likelihood, where, in the absence of a chance set-up, probability is inappropriate, then it is natural to measure our relative belief, in the absence of specific data, by the same means. Because likelihoods are not probabilities, how else are data-based or "experimental" relative likelihoods to be interpreted other than by reference to a scale of relative belief which, though arbitrary, like temperature, acquires meaning through experience and example?

The formal treatment requires that we only contemplate relative likelihoods, or likelihood ratios, either for two contrasted hypotheses, or for a continuum of parameter-values, in which case the same end will be achieved by considering the likelihood function determined only down to a constant factor, which may be arbitrarily chosen by assigning unit value to the likelihood at some particular point (such as its maximum). Following Jeffreys's terminology[39], the natural logarithm of the likelihood ratio may be called the support for one hypothesis against another. When based exclusively on data, given the model, it may be referred to as experimental support. Loosely, the same terms may be adopted for the logarithm of the likelihood function, it being understood that true support can only be adduced for one parameter-value against another, by forming the difference of their log-likelihoods, in which case the arbitrary constant is cancelled.

I now define prior support, but only when the prior knowledge does not contain any element of a chance set-up. The prior support for one hypothesis or set of parameter-values against another hypothesis or set is S if, prior to any experiment, I support the one against the other to the same degree that I would if I had conducted an experiment which led to experimental support S in a situation in which, prior to this conceptual experiment, I was unable to express any preference for the one over the other. The defined quantity is, of course, subjective. It has, by virtue of its definition, the properties of experimental support. The conceptual experiment plays the same part as Laplace's "device of imaginary results" in Bayesian inference. Birnbaum[24] regards a "prior likelihood function" which is constant as "a natural formal representation . . . [of] the absence of prior information", but seems not to have pursued the matter. Barnard[21], similarly, has written of a "neutral result". The scheme of inference is now:

$$\text{posterior support} = \text{prior support} + \text{experimental support} \quad (1)$$

whereas in Bayesian thinking the analogous relation may be derived by forming relative probabilities (for two hypotheses) or odds, and taking natural logarithms:

$$\text{posterior log-odds} = \text{prior log-odds} + \text{experimental support} \quad (2)$$

for the logarithm of the ratio of the likelihoods is, of course, the experimental support, defined earlier.

Everyone agrees that, when a chance set-up exists, the application of Bayes's theorem is valid, for the prior probabilities are then determined by the set-up. Inference becomes an exercise in conditional probability, but may be formulated as equation (2). In the total absence of a chance set-up, I suggest that equation (1) is appropriate. Furthermore, I hold that any situation can be treated by one, the other, or a mixture, of these two schemes, as necessary, as long as throughout the inference log-odds and support, which are as probabilities to likelihoods, are not confused. In addition it will generally be prudent to maintain the distinction between prior and experimental support.

The differences between equations (1) and (2) are vital. If we start with a chance set-up and prior log-odds, we can state our results in probability terms, using equation (2) to give the posterior log-odds. But if we start with mere opinions, we are unable to make a terminal probability statement, but must adopt the weaker form of inference, using equation (1) to derive a posterior opinion.

Valid Bayesian prior probabilities or probability distributions can profitably be thought of as part of the model. All inferences of the kind here considered are conditional on specific statistical models, and an inference will only be generally accepted if the model on which it is based is generally accepted. The adoption of a prior chance set-up is an extension of the model, and good grounds are needed for accepting all aspects of a model before inferences based upon it are accepted. If the model is incomplete, so must the inference be. The adoption of further probabilities to specify this incompleteness, as Good[10] recommends, similarly requires a model, and is thus of no avail. The difficulty can only be overcome by adopting a measure of belief that is not subject to the axioms of probability. Likelihood, and hence support, supplies this.

The use of support avoids the inconsistencies of subjective probability. It enables situations with and without prior chance set-ups to be distinguished, so that we no longer need assume that of which we are, in fact, ignorant; it is invariant to transformations of the parameters, and, in particular, uniform prior support for a parameter implies uniform prior support for any single-valued transformation of that parameter.

It is not supposed that the weaker form of inference supplied by support can achieve, though in a more satisfactory way, as much as inverse probability, for it is precisely the "achievements" of inverse probability to which exception is taken. One important difference, mentioned by Barnard[22,23], is the inability of support to attach meaning to the disjunction of two hypotheses. Whereas in inverse probability "H_1 or H_2" is regarded as a hypothesis, provided H_1 and H_2 are mutually exclusive hypotheses, in terms of support this disjunction is meaningless. "The likelihood of A or B means no more than 'the stature of Jackson or Johnson'; you do not know what it is until you know which is meant" (Fisher[32]). And this is quite right, for although, when prior probabilities may be validly attached to H_1 and H_2 through the existence of a chance set-up, the distribution of outcomes of the hypothesis "H_1 or H_2" is defined as the weighted sum of the separate outcomes, in the absence of such a set-up no distribution of outcomes can be contemplated. The essence of a statistical hypothesis is a predicted distribution of outcomes, and so "H_1 or H_2" is not an admissible hypothesis when support is in use. Similarly, the suggestion that a continuous parameter lies in a particular interval, such as $0 < p < \frac{1}{2}$, is only a hypothesis if a prior distribution of p is admitted; otherwise it only admits of an inequality statement about the possible distributions of outcomes, which does not merit the title of hypothesis in the precise statistical sense in which the word is used here. Perhaps the lack of meaning of such statements is a clue to the obscurity of the fiducial argument.

Mendelian Example

A single illustration involving both probability and support must suffice. It is an extension of Fisher's[19] well known example. "In Mendelian theory there are black mice of two genetic kinds. Some, known as homozygotes (BB), when mated with brown yield exclusively black offspring; others, known as heterozygotes (Bb), while themselves also black, are expected to yield half black and half brown. The expectation from a mating between two heterozygotes is 1 homozygous black, to 2 heterozygotes, to 1 brown. A black mouse from such a mating has thus, prior to any test-mating in which it may be used, a known probability of 1/3 of being homozygous, and of 2/3 of being heterozygous. If, therefore, on testing with a brown mate it yields seven offspring, all being black, we have a situation perfectly analogous to that set out by Bayes in his proposition".

Fisher then computes the likelihoods of the "homozygote" and "heterozygote" hypotheses on the data, obtaining 1 and 1/128, and applies these to the prior probabilities, 1/3 and 2/3, by means of Bayes's theorem, obtaining posterior probabilities of 64/65 and 1/65. In the terms of equation (2), the prior log-odds for homozygosis against heterozygosis are $\log (1/2) = -0{\cdot}693$, the experimental support is $\log (128) = 4{\cdot}852$, and the posterior log-odds are thus 4·159, or log (64).

Now suppose there is another prior possibility: that the mouse in question is the offspring of a mating between two black mice, as before, but one is now homozygous BB, the other being Bb. The expectation amongst the offspring is one homozygous black to one heterozygous black, prior log-odds of zero. The experimental support is the same, and the posterior log-odds are therefore 4·852, or log (128).

One day we are given a black mouse to test, and are told that it is the offspring of one of these types of mating, with no prior evidence to favour one type over the other. The test-cross, as before, gave seven black mice. The experimental support is still 4·852. If the parental mating is of the first type (expectation 1 : 2) then, as we have seen, the posterior log-odds are 4·159, but if the mating is of the second type (expectation 1 : 1) the posterior log-odds are 4·852.

We must now consider the other two hypotheses, concerning the types of parental mating. We have no reason to believe in one type rather than the other, prior to the test-cross, and adopt prior support of zero. The likelihood of the first hypothesis is the probability of obtaining seven black mice from a test-cross when the tested mouse is homozygous with probability 1/3 and heterozygous with probability 2/3, or

$$\frac{1}{3} \times 1 + \frac{2}{3} \times \frac{1}{128} = \frac{65}{192}$$

and the likelihood of the second hypothesis is, similarly,

$$\frac{1}{2} \times 1 + \frac{1}{2} \times \frac{1}{128} = \frac{129}{256}$$

The likelihood ratio for the first hypothesis against the second is thus 260/387, a support of $-0{\cdot}398$. Because the prior support was zero, the posterior support is also $-0{\cdot}398$.

All knowledge about the hypotheses of primary interest can therefore be summarized by asserting that the probability of the mouse being homozygous is either 64/65 or 128/129, the support in favour of the second value being 0·398.

This is not a statement solely of probability or of support, but a hybrid of the two, as is to be expected from the nature of the problem. Such hybrid statements fill the continuum between support and probability. Inverse probability, by contrast, leads to the same answers for different problems. In this instance it would be argued that the prior probability of the mouse being homozygous was

$$\frac{1}{2} \times \frac{1}{3} + \frac{1}{2} \times \frac{1}{2} = \frac{5}{12}$$

and of being heterozygous, 7/12. The posterior probability of homozygosity would turn out to be 640/647. This is the answer we would obtain if we interpreted the posterior support of log (260/387) for the first type of mating against the second as if it were posterior log-odds; for we would then weight the posterior probabilities of the hypotheses of primary interest accordingly, obtaining

$$\left(260 \times \frac{64}{65} + 387 \times \frac{128}{129}\right)/(260 + 387) = 640/647$$

as before.

But this result is too strong, for it makes a probability statement where it is not valid. It draws a conclusion that should only be drawn if the mouse in question had had probability one-half of deriving from each type of mating. The earlier statement, involving support, supplies the weaker inference which is needed.

I thank Professor J. H. Edwards, Dr K. E. Machin and Dr J. H. Renwick for stimulating discussions. I also thank Professor D. G. Kendall for hospitality in his department.

Note added in revision. Through the courtesy of Mr A. D. McLaren and Dr D. Hudson I have recently seen a draft of Dr Hudson's paper which similarly advocates the use of prior likelihood.

1 Venn, J., *The Logic of Chance*, second ed. (Macmillan, London, 1876).
2 Jeffreys, H., *Theory of Probability*, second ed. (Oxford University Press, 1948).
3 Bayes, T., *Phil. Trans. Roy. Soc.*, **53**, 370 (1763) reprinted in *Biometrika*, **45**, 296 (1958).
4 Fisher, R. A., *Phil. Trans. Roy. Soc.*, A, **222**, 309 (1922).
5 Hacking, I., *Logic of Statistical Inference* (Cambridge University Press, 1965).
6 Lindley, D. V., *Introduction to Probability and Statistics from a Bayesian Viewpoint* (Cambridge University Press, 1965).
7 Keynes, J. M., *A Treatise on Probability* (Macmillan, London, 1921).
8 De Finetti, B., in *Studies in Subjective Probability* (Wiley, New York, 1964).
9 Good, I. J., *Probability and the Weighing of Evidence* (Griffin, London, 1950).
10 Good, I. J., *The Estimation of Probabilities* (MIT Press, Cambridge, Mass., 1965).
11 Savage, L. J., in *The Foundations of Statistical Inference* (Methuen, London, 1962).
12 Smith, C. A. B., *J. Roy. Statist. Soc.*, B, **23**, 1 (1961).
13 Good, I. J., *J. Roy. Statist. Soc.*, B, **23**, 1 (1961); in discussion.
14 Dempster, A. P., *J. Roy. Statist. Soc.*, B, **30**, 205 (1968).
15 Lindley, D. V., *J. Roy. Statist. Soc.*, B, **30**, 205 (1968); in discussion.
16 Polya, G., *Mathematics and Plausible Reasoning*, **2** (Princeton University Press, 1954).
17 Bartlett, M. S., *Essays on Probability and Statistics* (Methuen, London 1962).
18 Bartlett, M. S., *J. Roy. Statist. Soc.*, A, **130**, 457 (1967).
19 Fisher, R. A., *Statistical Methods and Scientific Inference* (Oliver and Boyd, Edinburgh, 1956).
20 Fisher, R. A., *Statistical Methods for Research Workers* (Oliver and Boyd, Edinburgh, 1925).
21 Barnard, G., *J. Roy. Statist. Soc.*, B, **11**, 115 (1949).
22 Barnard, G. A., *Proc. Fifth Berkeley Symp. on Math. Statist. and Probability*, **1**, 27 (1966).
23 Barnard, G., *J. Inst. Actuar.*, **93**, 229 (1967).
24 Birnbaum, A., *J. Amer. Statist. Assoc.*, **57**, 269 (1962).
25 Cavalli-Sforza, L. L., and Edwards, A. W. F., *Bull. Intern. Statist. Inst.*, **41**, 803 (1966).
26 Cournot, A. A., *Exposition de la Theorie des Chances et des Probabilités* (Hachette, Paris, 1843).
27 Fisher, R. A., *Contributions to Mathematical Statistics* (Wiley, New York, 1950).
28 Neyman, J., and Pearson, E. S., *Phil. Trans. Roy. Soc.*, A, **231**, 289 (1933). Reprinted in *Joint Statistical Papers* (Cambridge University Press, 1967).
29 Wald, A., *Sequential Analysis* (Wiley, New York, 1947).
30 Raiffa, H., and Schlaifer, R., *Applied Statistical Decision Theory* (Harvard University, 1961).
31 Lindley, D. V., *J. Amer. Statist. Assoc.*, **57**, 269 (1962); in discussion.
32 Fisher, R. A., *Proc. Camb. Phil. Soc.*, **26**, 528 (1930).
33 Fisher, R. A., *Proc. Roy. Soc.*, A, **144**, 285 (1934).
34 Fisher, R. A., *J. Roy. Statist. Soc.*, **98**, 39 (1935).
35 Fisher, R. A., *Proc. Amer. Acad. Arts and Sci.*, **71**, 245 (1936).
36 Boole, G., *An Investigation of the Laws of Thought* (Walton and Maberly, London, 1854).
37 Bernoulli, D. (1777), reprinted in *Biometrika*, **48**, 3 (1961).
38 Smith, C. A. B., *Amer. J. Human. Genet.*, **11**, 289 (1959).
39 Jeffreys, H., *Proc. Camb. Phil. Soc.*, **32**, 416 (1936).