

THE BAYESIAN CONTROVERSY IN STATISTICAL INFERENCE

BY PROF. G. A. BARNARD, M.A., F.I.M.A., F.I.S., F.S.S. HON., A.R.C.S.
(University of Essex)

[Submitted to the Institute, 27 February 1967]

WHEN I was invited to read this paper I felt not only honoured but particularly pleased to be given the opportunity to set forth for your criticism the views I have come to hold concerning the complex of problems centred round the use of Bayes's theorem. For what body of people has for longer been engaged in the application of the mathematical doctrine of probabilities to the affairs of life? And so, what body could be better fitted to judge the merits and the faults of any attempt to clarify the principles of the subject? It is no accident that the original publication of Bayes's famous paper was brought about by the author of the Northampton Life Table; and in the modern period, in dealing with the criticisms of Bayes's postulate stemming from Boole, the credit for a major advance is shared between Sir Harold Jeffreys and Mr Wilfred Perks, independent originators of the theory of invariant prior distributions. And, to anticipate a point I shall develop in more detail later on, it appears to me that the experience of actuaries in the formation of categories as, for instance, by occupational group, as abstainers or non-abstainers, and so on, can be highly relevant to the effective use of Bayes's theorem in many wider contexts; and an examination of the principles underlying the formation of categories should improve our insight into problems of statistical inference in general.

2. Another reason for expecting specially valuable comment is that we are less likely here than elsewhere to be tempted to see the issues under the false dichotomy of the interpretation of probability either as a 'rational degree of belief' or as a 'hypothetical limiting frequency'. For while perhaps there may be measures of credibility which do not obey the addition law and the multiplication law for independent events, we are surely all agreed that mathematical probability does obey these laws; and James Bernoulli showed, before 1700, that if the probability of an event is p , then its relative frequency in an indefinitely long series of independent trials must converge to p in the sense that, by taking the number of trials n as sufficiently large, we can secure that the probability is arbitrarily small that the relative frequency r/n should differ from p by more than any given amount. And if we are prepared to admit the notion of an actual infinity of independent trials, we have the Borel-Cantelli theorem which says that the probability is 1 that r/n converges to p in the

ordinary mathematical sense. Thus, if we accept the addition and the multiplication laws, we can refuse to identify probability with such a *hypothetical* limiting frequency only if either

- (i) we refuse to associate arbitrarily small probability with impossibility,
or
- (ii) we refuse to allow that any one case can be regarded as one of a hypothetical series of independent similar cases.

As to (i), it is well known that we have to be careful in associating arbitrarily small probability with impossibility—to avoid, for example, taking ‘impossibility’ in the strong sense of logical impossibility—but such points as this are largely technical; while as to (ii) it surely is possible to regard our universe as one of an indefinitely large number of possible universes, even though some of us may feel such a notion is not very helpful. For my part, at any rate, probability is *both* a ‘rational degree of belief’ *and* a ‘hypothetical limiting frequency’. Of course, it is not any *actual* frequency—since any number of actual trials must necessarily be finite, actual frequencies can, at best only be *estimates* of probabilities. I conceive that the actuary is fully accustomed to the notion that the probabilities he uses can be thought of both as the rational degrees of belief in connexion with any individual case to which they are applied, and as approximating the relative frequencies with which events may be expected to happen in long runs of such cases.

3. There is an analogy with two different definitions of temperature—as the function of state variables which is the same for all bodies in thermal equilibrium, and as the energy of motion of molecules of an ideal gas; the equivalence of the definitions is a theorem of the subject of statistical mechanics.

HISTORICAL REVIEW

4. When Bayes’s article was published, in 1763, Bernoulli’s theorem had been known for over fifty years, and its application, by way of converse, to the estimation of probabilities was understood to some extent, as the following quotation from De Moivre’s *Doctrine of Chances* (3rd Ed., London, 1756, p. 252) shows:

... so, *conversely*, if from numberless Observations we find the Ratio of the Events to converge to a determinate quantity, as to the Ratio of P to Q ; then we conclude that this Ratio expresses the determinate Law according to which the Event is to happen.

For let that Law be expressed not by the Ratio $P : Q$ but by some other, as $R : S$; then would the Ratio of the Events converge to this last, not to the former: which contradicts our *Hypothesis*. And the like, or greater, Absurdity follows, if we should suppose the Event not to happen according to any Law, but in a manner altogether desultory and uncertain; for then the events would converge to no fixt Ratio at all.

5. This passage occurs at the end of a section in which De Moivre demonstrates, in the case $p = \frac{1}{2}$, and asserts in general, that the standardized deviation of relative frequency r/n from the probability p ,

$$D = \sqrt{n/p(1-p)} \cdot ((r/n) - p)$$

has a distribution which converges to the standard 'normal' form as n tends to infinity. Thus Richard Price was not altogether accurate in the introduction to Bayes's paper, when he said, of De Moivre's method for estimating probabilities from frequencies, 'it is not obvious how large the number of trials must be in order to make them (the estimates of probability) exact enough to be depended on in practice'. Price was correct, of course, in asserting that the problem as posed by Bayes:

Given the number of times in which an unknown event has happened and failed;
Required the chance that the probability of its happening in a single trial lies somewhere between any two degrees of probability that can be named.

had never before been solved. By 'an unknown event' Bayes makes clear he means 'any event concerning the probability of which nothing at all is known antecedently to any trials made or observed concerning it'.

6. As is well known, Bayes wrote an introduction to his essay which Price evidently saw but did not, most unfortunately, reproduce.* In his introduction, Bayes made it clear that he soon saw how to solve his problem 'provided some rule could be found according to which we ought to estimate the chance that the probability for the happening of an event perfectly unknown, should lie between any two named degrees of probability, antecedently to any experiments made about it; and that it appeared to him that the rule must be to suppose the chance the same that it should lie between any two equidifferent degrees; . . . But he afterwards considered, that the *postulate* on which he had argued might not perhaps be looked upon by all as reasonable . . .' He therefore restricted his mathematical argument to a case where the event in question, the coming to rest of one billiard ball O to the left of another ball W , when O is thrown upon the table so that its position is uniformly distributed over the table, can be guaranteed to have a probability uniformly distributed antecedently. He secures this by supposing the ball W also to be thrown upon the table in the same way as O . His mathematical argument led him to the expression

$$\int_a^b x^r (1-x)^s dx \bigg/ \int_0^1 x^r (1-x)^s dx$$

for the probability that the chance p of O falling to the left of W , when r successes and s failures have been observed, lies between a and b . And only

* The fact that Bayes's papers may have been preserved by a member of the Cotton family, or by the Vicar of Speldhurst, is worth mentioning, in the faint hope that, in spite of the failure of Augustus de Morgan's attempt to recover them, they may yet come to light.

after he had this result, rigorously derived, did he add his Scholium, in which he suggested that the same expression could be used in the case of an unknown event which had been observed to succeed r times and to fail s times.

7. Whatever doubts Bayes had concerning the suggestion made in his Scholium, they do not seem to have troubled later writers on probability for nearly a century. Laplace, it is true, allows in his general account of the theory for the possibility that the values of his parameters are not, *a priori*, equally possible; but after giving the more general expression corresponding to

$$\int_a^b x^r (1-x)^s w(x) dx \bigg/ \int_0^1 x^r (1-x)^s w(x) dx$$

for the case when the values of x occur with probabilities proportional to $w(x)$ *a priori*, he remarks that we can imagine $w(x)$ to be proportional to the density *a posteriori*, after an experiment which gives the likelihood function $w(x)$, starting from a uniform prior. Thus, he says, the more general case can be reduced to the case of a uniform prior. However, in his applications, for example where x is the probability of a birth being a male birth, he takes his prior as uniform, without further discussion.

8. It was not until the decade 1830–40 that serious doubt seems to have been cast on the appropriateness of the assumption of a uniform prior. A detailed study of the discussions which went on at this time would be well repaid. S. D. Poisson, A. A. Cournot, A. Quetelet, G. Boole, A. De Morgan, Sir John Lubbock, C. Babbage, J. S. Mill, and several others were keenly interested in the possibility of using the doctrine of chances in the elucidation of the ‘laws’ of ‘social science’ (recently adumbrated by Auguste Comte), and with such applications in mind they attempted to clarify the foundations of the subject. An indication of the ebb and flow of discussion can be obtained from reading the passages on the subject in Mill’s *Logic*. In the early editions he adopts a strictly frequentist interpretation of probability (would he have had this from Quetelet?); but in the later editions he adopts a much more subjective view, in particular accepting the notion that two possibilities which appear to us as ‘equally possible’ are to be taken as having the same probability. He seems to have perceived that taking ‘cases’ as equally likely *a priori* would be possible provided the ‘cases’ were finite in number, but that difficulties would arise as soon as the number became infinite. On the other hand, his argument that, if the probability of B given A is $2/3$, and the probability of B given C is $3/4$, then the probability of B given both A and C must be $11/12$, illustrates the degree of sheer mathematical confusion liable to arise before the publication of Boole’s *General Method in the Calculus of Probabilities*.

9. As has been indicated, a fair analysis of the ideas current in the 1830s would require much deeper study than I have made, and the following outline must be taken as liable to revision. The dominant fact would seem

to have been that, for the first time, reasonably reliable social statistics were becoming available. These enabled statisticians—in the original sense of the word—to estimate the probabilities of social phenomena by the method originating with Johann Bernoulli, and expressed above in the quotation from De Moivre. Thus, as Professor Anscombe has remarked, although we do not know for certain who invented the histogram, it seems likely that this, perhaps the most useful of all statistical devices, was invented by Quetelet. It leads directly to the notion of using an observed frequency to estimate a probability. And from this it is a short step to the identification of probability with some kind of limiting frequency. As already noted, Mill's discussion of the subject shows him oscillating between a subjective interpretation of probability and an objective one in terms of observable frequency, without, apparently, recognizing any incompatibility between the two views. For Poisson also there are the two aspects of probability, but in his case their difference is clearly appreciated; in fact he comes near to anticipating the ideas of 'probability-1' and 'probability-2' put forward by Carnap. Cournot insists most strongly on the frequency interpretation of probability, and, in making a careful restatement of the argument of De Moivre, quoted above, Cournot comes near to a confidence-interval type of statement about an unknown probability from an observed frequency. Cournot and Boole both explicitly reject Bayes's postulate. Perhaps the most powerful argument that can be brought against the postulate is that if p is an unknown quantity lying between 0 and 1, so is p^3 ; so if, on these grounds, p is to be taken as uniformly distributed *a priori*, so also must p^3 be. But p and p^3 cannot simultaneously be uniformly distributed between 0 and 1. This argument, or rather a similar one to the effect that Bayes's postulate is self-contradictory, is to be found in Boole's *Laws of Thought*. Also to be found there is the argument that, if the probability of E on hypothesis H is p , then, when E has been observed, the posterior probability of H is

$$P = ap/(ap + c(1-a))$$

where a and c are *arbitrary constants*, the former representing the *a priori* probability of the hypothesis H , the latter the probability that, if the hypothesis were false, the event E would present itself. Boole points out that, in regarding a and c as arbitrary constants, he differs from De Morgan who had effectively taken a to be $\frac{1}{2}$ and c to be 1, and proceeds to write: '... it is with diffidence that I express my dissent on these points from mathematicians generally, and more especially from one who, of English writers, has most fully entered into the spirit and methods of Laplace; and I venture to hope, that a question, second to none other in the Theory of Probabilities in importance, will receive the careful attention which it deserves'. If Boole could have seen how much the question he raised has been discussed subsequently, and is still being discussed, he would surely have felt that his wish was granted.

10. It is perhaps worth while to draw attention to one aspect of Boole's work. Namely, that he was one of the first mathematicians to treat systematically of numerical problems not possessing uniquely defined numerical solutions. That arbitrary constants, and arbitrary functions, may enter into the general solutions of differential equations had, of course, been known for nearly a century; but it was still customary to suppose that a numerical problem must, if it had an answer at all, have a unique answer. It may have been his familiarity with other situations in which solutions lacked uniqueness that led Boole to perceive, and to be ready to accept, the notion that such lack of uniqueness might exist in apparently well-specified problems of probability. How intrusive the notion of uniqueness is may be illustrated by reference to Sir Harold Jeffreys's *Theory of Probability*, in which he says (p. 33, 3rd Ed.): 'no practical man will refuse to decide on a course of action merely because we are not quite sure which is the best way to lay the foundations of the theory. He assumes that the course of action that he actually chooses is the best . . .' But if one picks up a builder's hammer, finding it lying to hand, to drive in a small nail, one does not thereby imply that one has ruled out the possibility that a second's look would reveal another hammer, of more appropriate weight, also lying close at hand; we may say to ourselves: 'This will serve the purpose', without implying that it necessarily will *best* serve the purpose. Or, to take another case, if a conductor takes a passage of music at a certain speed on a given occasion, this does not imply that he thinks it the best speed—he may well think that different speeds bring out different aspects of the work, and has on this occasion really made an arbitrary choice. Or again, in a more important context, if we were to insist that, in any given state of knowledge, there was always a 'best' indicated method of treatment of a diseased patient, we would pose extremely difficult ethical problems in connexion with clinical trials. Finally, it has been suggested by Prof. E. S. Pearson that his 'theory' of testing hypotheses represents *one* way of seeing how probability considerations enter into scientific judgments, and that there may well be other ways of looking at these matters, all of which have something to contribute; and it must be admitted that, as of now, there is a great deal to be said for this view.

11. Interest in the theory of probability continued to be widespread for the twenty years after 1840, though the emphasis shifted somewhat towards more technical matters, as is exemplified by Whitworth's *Choice and Chance*, and the brilliant analytical work of the Rev. Leslie Ellis concerning random walks. After 1860 the whole field of science came to bear a deterministic look, with the triumphant development of natural philosophy, of Darwinism, and of deterministic theories of social development; and it may have been this, rather than the disquiet, commonly referred to, concerning the foundations of the subject, which led to the decline of interest in probability theory. We may note, at any rate, that

Mendel's work on genetics was carried out at this time and, as is well known, his essentially probabilistic theory failed to attract the attention it deserved for nearly fifty years.

12. The end of the century brought a return of interest, with the work of Francis Galton and his followers, Karl Pearson, Weldon and Sheppard. These men were primarily empiricists, with rather little feeling for logical subtleties. It is a great pity that Edgeworth, whose analytical powers were of a very high order, was regarded by Galton as 'too mathematical' to work on the problems he and Karl Pearson were raising, so that Edgeworth's main attention was deflected towards economics at a time when that field of study was not so much in need of his talents. However, Edgeworth devoted several papers to the theory of statistical estimation, including the famous one of 1908 in which, with the help of A. E. H. Love, he gives the analysis involved in the asymptotic version of the information limit to the variance, without, however, giving a clear interpretation of the result. He early made it clear that a distinction was to be made between the theory of small and the theory of large samples; in the case of the former, the prior distribution for the parameter being estimated would bring an element of arbitrariness into the solution, but such would not be the case with large samples. Whether the lack of dependence on the prior distribution, in the case of large samples, was due to the fact that any smooth prior could be taken as effectively uniform over the small range involved, or whether it was due to the possibility of applying an argument of De Moivre's type, is never really made clear, in those papers which I have been able to read.

A similar ambiguity appears in Karl Pearson's work. With his χ^2 measurement of discrepancy he adopts an argument of the De Moivre type; but in *Tables for Statisticians and Biometricians*, the table (computed by Major Greenwood) on 2×2 tables is based on an assumption of Bayes's postulate, tacitly accepted. He did indicate an interest in the statistical distribution of the values of constants of nature, and this has been taken to indicate an 'empirical Bayes' attitude, in which Bayes's theorem is applied to a prior distribution which is itself based on experience. But the evidence is again not clear.

THE MODERN PERIOD

13. It was, indeed, one of the major services which R. A. Fisher rendered to statistics, that he insisted on much tauter standards of reasoning in the subject. Although his first paper on the method of maximum likelihood (1912) is in line with views then generally accepted, in making use of inverse probability arguments (though of a restricted kind), he soon began to concentrate on the sampling distributions of statistics, and in fact he first obtained almost all the standard results now in common use in the field, with the sole exception of the distribution of Student's t , in

which Gosset was the pioneer. And in his basic 1922 paper he set out the reasoning ascribed above to Boole, showing the self-contradictory character of Bayes's postulate, and went on to suggest that inferences from samples about parameters, except in the 'trivial' cases where the population is known to be itself a sample from a known super-population, are to be expressed in terms of likelihood rather than probability. Likelihood differs not only from the 'frequency probability' involved in the specification of the distribution of observations, but also, for example, from the 'logical probability' of Keynes, in not obeying the addition law. If A and B are mutually exclusive, the probability of ' A or B ' is the sum of the probability of A plus the probability of B ; but the likelihood of ' A or B ' is undefined—it is like the height of 'Peter or Paul'—it means nothing unless we specify which is meant. Yet likelihood resembles probability in that, in an important class of cases, it may be held to measure the degree of our rational belief in a conclusion.

14. Fisher's first reference in this paper to inverse probability is worth quoting. He refers to 'the fundamental paradox of inverse probability, which like an impenetrable jungle arrests progress towards precision of statistical concepts. The criticisms of Boole, Venn and Chrystal have done something towards banishing the method, at least from the elementary textbooks of Algebra; but though we may agree with Chrystal that inverse probability is a mistake (perhaps the only mistake to which the mathematical world has so deeply committed itself), there yet remains the feeling that such a mistake would not have captivated the minds of Laplace and Poisson if there had been nothing in it but error'. The truth, embodied along with error, in the Bayesian postulate, was, according to Fisher, the fact that we really can learn from experience, and that the knowledge we thus gain is affected with uncertainty. The prime error consisted in the assumption that *all* uncertainty is measurable in terms of probability.

15. An argument which Fisher might have developed, but did not, in support of the distinction he was drawing between uncertainty as it affects statistical hypotheses and uncertainty as it affects observable propositions, is the purely logical one that the logical disjunction of two observable propositions is itself an observable proposition of essentially the same logical status; but the logical disjunction of two statistical hypotheses, each of which enables us to calculate the probability of any observable proposition, does not itself enable us to calculate any probabilities. At the time when Fisher was developing his ideas on likelihood, the Dutch mathematician L. E. J. Brouwer was elaborating his critique of the logic of set theory, putting special emphasis on the dangers of over-facile treatment of logical disjunction, in particular in connexion with the law of excluded middle. But at that time interest in these aspects of the foundations of mathematics was largely confined to the Continent, and Fisher never seems to have come across the work.

16. The evolution of Fisher's views could readily be traced from a

'variorum edition' of *Statistical Methods for Research Workers*—something we may hope may one day be made available. In the first (1925) edition, no qualifications (other than the 'trivial case' mentioned above) are placed on the statement that inferences from sample to population are to be framed in terms of likelihood, not probability. But later, after Fisher had discovered the fiducial argument, his statement is much modified. Indeed, until near the end of his life Fisher seems to have thought that the fiducial argument would enable us to make probability statements about parameters in most cases of practical importance. He does eventually, however, seem to have realized that situations where the fiducial argument may apply are the exception rather than the rule, so that his original position concerning likelihood was nearer the truth than he later came to think.

17. We shall not enter here into a discussion of the fiducial argument, since it represents only a fascinating byway in statistical inference. It is important that this byway should be explored, and in particular the conditions for avoiding contradictions in the application of the argument to more than one parameter need to be specified more exactly than hitherto. Such studies must involve the examination of structures on the parameter space, related to structures on the space of observations, which are clearly relevant also to problems where the fiducial argument is inapplicable. Another place where such structures arise is in connexion with the invariance theory of Jeffreys and Perks; a weakness of this theory seems to be that it derives the structure of the parameters exclusively from the probability structure of the observations, without regard to other structures (such as group structures) which the observations may possess. In his discussion of joint estimates of location and scale, Jeffreys himself seems to suggest that group structures as well as probability-induced structures are relevant. But a good deal remains to be clarified here. The simplicity of the case of one parameter is associated with the fact that one here has a simple total ordering structure.

18. Jeffreys's earliest work on probability was nearly contemporary with Fisher's early work. The two outstanding characteristics of his approach have seemed to me to be his embedding of probability into a wider theory of scientific inference—in connexion, for example, with his 'simplicity postulate'—and his modification of Bayes's postulate in the direction of allowing non-uniform prior distributions to represent ignorance in certain situations. He has rightly complained of critics of his approach who have failed to notice the novelty of his theory in this respect; so much was the application of Bayes's theorem in statistical inference traditionally connected with the assumption of a uniform prior, that when he abandoned the universal assumption of uniformity the fact was overlooked. As he says, there is no more need for the idea that the prior distribution must always be uniform than 'there is to say that an oven that has once cooked roast beef can never cook anything but roast beef'. However, he admits that the principles he adduces for making a choice

of prior distribution only 'sometimes indicate a unique choice, but in many problems some latitude is permissible, so far as we know at present. In such cases, and in a different world, the matter would be one for decision by the International Research Council. Meanwhile we need only remark that the choice in practice, within the range permitted, makes very little difference to the results'. Looked at from a distance, this position does not seem to be very different from that of, for example, Edgeworth. Another man whose views were closely related to Jeffreys was Haldane, who proposed, for example, the prior density $1/p(1-p)$ for an unknown probability instead of Bayes's uniform assessment and who proposed a prior having a 'lump' of probability at the null hypothesis with the rest spread out, in connexion with tests of significance. Haldane used to say that statistics was primarily concerned with 'medium sized samples' where the number of observations is not so large that efficiency of estimation is unimportant, and yet not so small that the prior distribution has an appreciable effect.

19. Ten years after Jeffreys and Fisher wrote their first papers on the foundations of statistical inference a radically new departure was made by Ramsey and De Finetti though their work did not receive the attention it merited from statisticians until some twenty years later, when Savage gave their views his weighty support and further developed them. Ramsey took as his starting-point the idea of probability as a guide in choice of a course of action, and showed that an individual whose actions satisfied certain criteria of consistency would have to act *as if* he possessed a utility function describing his preferences between the possible outcomes of his acts, and a prior distribution over the various possible states of nature which would be modified in the light of his observations in accordance with Bayes's theorem. His actions would be such as to maximize his expected utility.

20. Although the school of statisticians taking their inspiration from Savage and De Finetti is often described as the 'Bayesian School', the distinguishing feature of their approach is not the use of Bayes's theorem, which they share, as has been seen, with many other statisticians; nor is it, of course, their use of Bayes's postulate, which they explicitly deny. For them, there is no general way of expressing ignorance about a parameter, such as Bayes's postulate purports to supply; each individual will, in any given case, be in a state of partial knowledge about the parameter he is considering (he must, for example, know enough about it to be able to identify it) and such knowledge as he has will be reflected in a prior distribution which is in principle peculiar to him. All probabilities are therefore subjective, and it is this emphasis which gives the school its special character. But it must not be thought that the subjectivity of probabilities implies any disregard for objective facts; on the contrary, since any individual's distribution for a parameter must be modified, in accordance with Bayes's theorem, on the basis of observations, it is possible clearly to identify

the part of the distribution which does come from the observations, as consisting in the likelihood function. And since the task of a statistician is not to provide his clients with their prior distributions—these they must provide for themselves—the statistician's task reduces, in principle, to the description of the likelihood function, once the observations have been made. And before the observations have been made, in planning experiments, the statistician needs to consider the sampling distribution of the set of likelihood functions which may arise. Thus there is, or should be, little in practice which divides a follower of Fisher, using likelihood, from a subjective Bayesian; the difference, such as it is, concerns the view taken about the thought processes which go on, or should go on, in the minds of scientists, not with their overt behaviour in communicating with their fellow scientists.

21. Neyman is another who has emphasized the connexion of statistical inference with action. In fact he, perhaps alone of all the theorists considered, explicitly denies the possibility of learning from experience, and instead has introduced the concept of 'inductive behaviour', in which, for example, we 'assert' that a parameter lies within a confidence interval, and the interval is derived from the observations in such a way that the probability that it covers the true value of the parameter is not less than some preassigned value, the confidence coefficient. As a theory of inference, of course, such a view is easily made fun of—no one has been found making the 'assertions' which are the subject of the theory; what, in practice, is done is to assert, or rather specify, the confidence interval, *together with* its confidence coefficient, and in practice the latter is interpreted as a sort of probability (or degree of confidence, or rational belief) attached to the proposition that the parameter lies inside the interval. The individual who makes a practice of baldly asserting, in accordance with the theoretical prescription, cannot be thought of as corresponding to a real person; but as a metaphorical picture the idea is not without its uses. In the same way, the Neyman-Pearson account of significance tests, in which one is supposed to reject the hypothesis tested whenever the sample points fall within the critical region, and not otherwise, cannot be thought to correspond with what actually happens—indeed, Prof. Pearson has made it clear that its purpose is only to draw attention to some aspects of the processes of inference. What is essential in the Neyman-Pearson approach is the idea of the operating characteristic of a statistical procedure, which describes the performance, in probability terms, of the procedure, for each possible parameter value. This notion received its fullest development at the hands of Wald, in his theory of statistical decision functions.

22. When Wald first put forward his theory, he associated with it a suggested principle for choice among alternative decision functions—the famous minimax principle. This corresponds to the heresy of Manicheism, according to which this world is the worst of all possible—in so far

as the performance of a statistical procedure depends (as it does, in general) on what the state of nature is, it is to be judged by its performance in the worst possible state of nature. In this sense, it would, I think, be true to say that no one now adopts the minimax principle as a normative one. The fact that some particular procedure is minimax is often of interest—since, for example, it often implies that the expected loss associated with its use is independent of the true value of the unknown parameter; but it seems to be generally agreed that the only normative principle operating in the theory is that of admissibility—that one should restrict one's choice of statistical procedure to the class of admissible procedures. A procedure is admissible if no other procedure exists which gives sometimes a smaller expected loss, and never gives a larger loss. The central result of the theory is the complete class theorem, according to which, *modulo* minor details, the class of all admissible procedures is coextensive with the class of all Bayes's procedures—that is, procedures which would rationally be followed by someone believing the parameter in question to have been chosen at random from a population of values following some prior distribution.

23. Now the class of Bayes's procedures is characterized by the fact that the observations enter into such procedures only via the likelihood function. Thus here again we appear to have further confirmation of the importance of likelihood. The position of the likelihood function in the decision-theory approach is, however, rather different from what it is in the subjective Bayesian theory. Dr Henry Scheffe has coined the term 'restricted likelihood principle' for the idea that, in any given experimental situation, two possible results which yield the same likelihood function should yield the same inference. The 'unrestricted likelihood principle' asserts that any two results which give the same likelihood function should yield the same inference, regardless of whether they arise in the same experimental situation or not. The central theorem of the Wald theory, together with the principle of admissibility, implies the restricted likelihood principle; but not the unrestricted principle—though it should perhaps be emphasized that the unrestricted principle is not incompatible with decision theory. The subjective Bayesian approach, and Jeffreys's approach, imply the unrestricted likelihood principle.

24. The latest in our catalogue of approaches to the principles of statistical inference is the 'Empirical Bayes' approach originated, in its mathematical form, by H. E. Robbins. What Robbins does, briefly, may be described by saying that he groups together sets of inference or decision problems, and uses the collective data from all of them to estimate a prior distribution from which the individual parameter values may be regarded as having been sampled. Then, using this estimated prior combined with the data for a given problem by means of Bayes's theorem in the usual way, he arrived at individual inferences or decisions. These individual decisions can then be shown to be, on the average, usually better than would have been arrived at treating each case separately. In estimating

the prior distribution from the observations themselves, we seem to be pulling ourselves up with our own bootstraps; and the remarkable thing is that we do lift ourselves off the ground in this way. This is true even if the problems thus collected together stand in no relation to each other; we may gain even more if we can collect together problems in which the parameter values lie close to one another.

THE PRESENT POSITION

25. Apart, perhaps, from the aim of encouraging others to make a fuller study of the history of the subject—the best excuse I can make for its inadequacies—the above account makes clear that to hope for any final resolution of the Bayesian controversy is like hoping for a final resolution of the fundamental problems of philosophy. Indeed, the controversy is closely connected with those that have occupied philosophers for centuries, in so far as it centres around the principle of sufficient reason and related matters. And like the problems of philosophy, as understood today, it appears to have become highly fragmented, to be dealt with by examination of multitudes of varying instances rather than by means of stating a few principles of universal application, in terms of which all problems can be solved.

26. This is not, of course, to say that we cannot extract some general principles. One of those which would, I think, be universally accepted, is the ‘sufficiency principle’. This relates to those problems in which we can observe a random variable x whose distribution depends on an unknown parameter θ , but is of known form. The sufficiency principle says that, if we can by a one-to-one transformation transform the observable x (which, in general, will be presented by a vector, or set of numbers) to a pair of observables (y, z) , in such a way that the probability distribution of y depends on θ , but the distribution of z , given y , does not depend on θ , then inferences or decisions based on y alone, ignoring the value of z , are as good as inferences or decisions based on x itself. The term ‘sufficient’, and the first explicit use of the principle, are due to Fisher who showed that if x is a vector of errors e_1, e_2, \dots, e_n , known to be Normally distributed with zero mean and unknown variance, $\theta = \sigma^2$, we may take y to be the sum of squares

$$S = e_1^2 + e_2^2 + \dots + e_n^2$$

and z to be, for example, the vector whose i^{th} component is $u_i = e_i/S$. Then as x ranges over the whole n -dimensional space, y ranges over the interval from 0 to ∞ , and z ranges over the unit sphere in n dimensions. The distribution of y contains θ as a scale parameter, but the distribution of z is the uniform distribution over the surface of the sphere, whatever the value of θ . Thus any decision about θ may just as well be based on the

single number y as on the set of n numbers x . On the other hand if instead of taking y to be S we take it to be

$$T = |e_1| + |e_2| + \dots + |e_n|$$

then as x ranges over n -dimensional space, y ranges from 0 to ∞ , but if $u_i = e_i/T$, z now ranges over the surface of an n -dimensional polyhedron over which its distribution is not uniform, but concentrated either near the vertices, or away from the vertices, depending on the value of θ . Thus S is sufficient for θ but T is not.

27. Fisher's own use of the sufficiency principle was somewhat limited by the fact that he almost always thought of the transformed variables (y, z) as ranging over sets of numbers. The domain of applicability of the principle is very much widened if we allow y to range over a set of functions. The limitation to numbers was natural for Fisher, because he was working in the era of desk calculators when the statistician would typically calculate other numbers from the numbers representing his data. But now that we have computers capable of generating graphical forms of output it is natural to think in terms of calculating (the graphs of) functions from the numerical data. And whereas the property of sufficiency can be possessed by a single numerical y only when the distribution of the original observations belongs to a particular family (the exponential family) of distributions, we can *always* find a y ranging over a set of functions having this sufficiency property. In fact if we take y to be the likelihood function—defined as proportional to the probability function for the observations, which becomes a function of θ when we insert the numerical values of the observations, conveniently normalized, usually, to have its maximum value 1—then not only does y have the sufficiency property, it also has the property of *minimal* sufficiency. This means that if y' is any other numerical- or functional-valued statistic having the sufficiency property, then the value of y (i.e. the likelihood function) can be computed from y' , though the converse is not necessarily true. The information needed to calculate y must be contained in y' , but the information needed to calculate y' need not be contained in y . Thus y contains the minimal amount of information which there must be in any statistic having the sufficiency property.

28. Thus, provided we are sure about the relationship between the parameter θ we are interested in and our observations—as happens, perhaps most often, in the application of statistics in nuclear physics and in genetics—then we can be sure we will have summed up all the information our observations have to tell us about θ if we draw the likelihood function; and this is, not surprisingly, becoming common practice amongst nuclear physicists and geneticists (for physics see, for example, F. Solmitz, *Ann. Rev. Nuclear Science* 1964). We must be careful, of course, not to interpret the likelihood function as if it were a probability function—its *ordinates* have meaning, but not the areas under its curve, as Fisher emphasized in the passage quoted above. If the ordinate of the likelihood at

$\theta = \theta_1$, $L(\theta_1)$, is higher than the ordinate $L(\theta_2)$ at $\theta = \theta_2$, we can say that the data point towards the value θ_1 rather than towards θ_2 , or that, on this evidence, θ_1 is more plausible than θ_2 —though of course we may have other evidence which points the other way. And the interpretable relationships between likelihoods are not merely the ordinal ones; ratios of likelihoods have a definite meaning. If, for example we have $L(\theta_1)/L(\theta_2) = 2$, and we denote the likelihood function for θ from another body of independent evidence by $L'(\theta)$, then if, between θ_1 and θ_2 this other evidence points the other way, this other evidence will outweigh the former evidence if, and only if, $L'(\theta_2)/L'(\theta_1)$ is greater than 2.

29. The likelihood function can also be interpreted in terms of the Neyman-Pearson concept of an operating characteristic, though the interpretation is rather complicated to state in a general way. To take a very simple example first, suppose we have a logistic dosage-response relationship in which the probability p of positive response is given by

$$\log(p/(1-p)) = x - \theta$$

so that, in the usual terminology, the slope is known to be unity, and the 50% effective dose is the unknown parameter θ . Now suppose we have only one observation, for which the response was in fact positive, at dose $x = 2$. Then since it follows from the above relationship that the probability of a positive response,

$$p = \exp(x - \theta)/(1 + \exp(x - \theta))$$

at dose x , it follows that the likelihood function, given our observation, is

$$\begin{aligned} L(\theta) &= \exp(2 - \theta)/(1 + \exp(2 - \theta)) \\ &= 7.3891/(7.3891 + \exp \theta). \end{aligned}$$

(Note: no special normalization is necessary, since the maximum value is attained at $-\infty$, and is there 1.) Now suppose we were testing the hypothesis that $\theta = 5$, against alternatives $\theta < 5$. We might take positive response as being in the critical region, and negative response as not in the critical region. For this test, the 'size' of the critical region (i.e. the 'level of significance') is the probability of positive response, in the hypothesis tested, namely

$$7.3891/(7.3891 + \exp 5) = 1/20.08$$

while the power of the test for any value of θ is seen to be $L(\theta)$. Generally, we have, for any test for which the critical region consists of a single point and for which the maximum of the power function is 1, the power curve of the test and the curve of likelihood, given that the point in the critical region has been observed, are one and the same. For most tests the maximum power will in fact be 1, but in case the maximum is P , less than 1, all we would have to do to make the power curve coincide with the specified likelihood function would be to scale it up in the ratio $P : 1$ —or to re-normalize the likelihood to have maximum value P instead of 1.

30. If, now, the critical region consists of (say) four points, c_1, c_2, c_3, c_4 , and $L(\theta|c_i)$ denotes the likelihood for θ , given that the point c_i has been observed, we shall obtain the power function by rescaling each likelihood to a maximum P_i equal to the maximum, for variable θ , of the probability of observing c_i , and then adding and then rescaling again to make the maximum 1,

$$C(\theta) = \sum_i P_i L(\theta|c_i) / \text{Max}_\theta \sum_i P_i L(\theta|c_i)$$

Thus any power curve can be regarded as obtained by a weighted summation of likelihood functions.

31. But as we remarked above, the Neyman-Pearson theory of hypothesis testing is not to be taken literally, as giving a procedure which any real experimenter is likely to follow. One respect in which the metaphorical account is likely to over-simplify what happens arises from the fact that people do not choose a significance level for a test and then mechanically stick to it, whatever result is observed. Rather, they take the result, and find on what level it is significant. If, for instance, the four-point critical region discussed above had significance level 0.05, this significance probability might be made up of 0.02 for c_1 , 0.015 for c_2 , 0.01 for c_3 , and 0.005 for c_4 ; and if so, had c_2 in fact been observed, the quoted significance level would have been 0.03, not 0.05. And the power curve would have been obtained by combining only three likelihood functions, not four.

32. One direction in which this line of thought has been pursued by adherents of the Neyman school has been in the development of 'multiple-decision procedures'. Instead of considering the simple two-way choice, to reject the hypothesis tested or to accept it, these writers have considered many-way choices—for example, to 'provisionally reject', 'firmly reject', etc. The logical conclusion of this direction of development is to have as many choices as there are distinct likelihood functions; not more, because making different choices for points giving the same likelihood function would involve making the choice depend on the z component of the pair (y, z) (in which y represents the likelihood function); that is to say, it would be made to depend on something whose distribution does not involve the parameter in question. Of course, we do have occasion to consider randomized decision rules in the theory of multiple-decision procedures; but these correspond either to situations where the evidence is really evenly balanced, and we can, without loss, make a non-randomized decision, or to cases where there is an artificial restriction on the choices considered. Thus, if we allow ourselves the maximum multiplicity of choice consistent with due regard for differences in possible data, we are led to consider multiple-choice procedures for which the components of the operating characteristic are just the likelihood functions, rescaled perhaps to make the maxima equal to the appropriate maxima of probabilities.

33. Another way of relating power curves and likelihoods is to take the choices as fixed—for example as two-way, in hypothesis testing. Then the division of the sample space into a critical region and its complement can be thought of as a decision to observe a new random variable, T , which takes the value 0 if the sample points fall in the critical region and 1 if it falls outside. The power curve of the test procedure will then be the likelihood, given $T = 0$, perhaps rescaled.

34. Thus any likelihood can be regarded as a power function, and any power function can be regarded as a likelihood, except only for the fact that in the case of the power function the ordinates of the curve represent actual probabilities, while with the likelihood function the ordinates will, in general, only be proportional to probabilities. The principal advantage of power functions over the corresponding likelihoods lies in the fact that, if the maximum of the power function is low, we may be inclined to doubt the adequacy of the mathematical model, whereas if we use likelihood we are taking the model as gospel. We have already stated that for the present we are assuming the validity of the model, so that the only question at issue is the value of the parameter; in real life, of course, we shall not have such faith. But the adequacy of the model must in any case be tested by another mode of argument, the simple test of significance, based on our disbelief that miracles can be reproduced. Briefly, if we find that our observation has very tiny probability, for any value of the parameter, we shall be suspicious of the model; and we shall regard our suspicions as confirmed if, on repeating our observations, we find again a similar result.

35. Returning to our position of unquestioning faith in the model—or, rather, of leaving this issue on one side—it would seem that the maximum of a power curve will have little relevance in relation to a result already obtained—though it could, of course, be very relevant in advance planning of an experiment. If the result appears at best improbable, we simply have to say that improbable events must occur with no more, but also with no less, than their appropriate frequency, no matter how surprised we may be that they should happen to us. And in this case, we can argue that in looking at the power curve we shall be looking at the relative sizes of its ordinate; viewing it, in fact, as we would a likelihood curve.

36. Reasoning along these lines, I would suggest, it appears that the central and most valuable concept of the Neyman-Pearson approach to statistical inference, that of the power curve or operating characteristic, can be regarded as essentially equivalent to that of a likelihood function. Such differences as there are arise from the use of the concept of power in the advance planning of experiments, with which we are not at present concerned.

37. Turning now to the Bayesians, these range from the strict adherents of the doctrine of personal probability, such as Savage and De Finetti, through those like Jeffreys and Lindley who think there is some way of

expressing, in prior probability terms, a state of ignorance, to those, like Robbins and Esther Samuel, who advocate an 'empirical' Bayes approach. I have mentioned only those who have identified themselves rather strongly with a particular point in the spectrum of opinions. Many practical statisticians, including myself, have advocated the use of the Bayesian approach in connexion with decision problems, with a prior distribution arrived at by making use, in a necessarily rather informal way, of whatever information is available about the parameter externally to the experiment in question. The majority of papers now being published in the field of sampling inspection, for example, seem to be written from this point of view. While the existence of this body of opinion should not be forgotten—especially if, as happens sometimes, the impression gets abroad that statisticians are hopelessly entangled in internecine disputes over foundations—when we are discussing principles it is better to focus on the more radical points of view.

38. We may notice first that all Bayesians agree in taking the information from the experiment to be embodied in the likelihood function $L(\theta)$. For them, the basis of whatever action we may take, or opinions we may form, is the posterior distribution whose ordinates are proportional to the product of those of the prior distribution and the likelihood ordinates. Thus any of the Bayesian positions effectively amounts to a set of prescriptions as to how to use the likelihood function from an experiment. So that in pursuit of the œcumenical aims we have in mind, we can say that the Bayesians can be brought under the likelihood umbrella, as can the adherents of the Neyman-Pearson school.

39. In view of the universal importance which attaches to the likelihood function, it is remarkable how little attention is paid, in textbooks of statistics, to discussing the forms the likelihood function may take. If any of us were asked to sketch the general shape of the Normal curve, or a typical Poisson or binomial distribution, we could all do so without having to stop and think. But I wonder if the same could be said, for instance, of the likelihood function for a Poisson parameter, on the basis of a sample with given total? We do not have terms with which to classify or describe the various shapes of likelihood functions, except those, such as multimodal, which can be taken over straight from terms describing statistical distributions. This is surely something that should be corrected. But meanwhile, it is worth asking, why this neglect?

40. One reason is the general neglect of estimation problems in favour of tests of significance which has been a regrettable feature of textbooks for so long. This is being remedied in the more recently published books. Another is the unwillingness, which seems to be widespread, to consider any formulations of the problem of estimation intermediate between those which are referred to as 'point estimation' and 'interval estimation'. We do sometimes—very rarely, I think—need to choose a single number to represent the best guess we can make as to the value of a parameter; and

we do, more often, need to see what values of a parameter are not ruled out by the data as too implausible. But surely we usually need to know more—towards what sets of values do the data point, and how strongly? And although the art of the combination of observations, and of the reduction of data, tend to be regarded as somewhat old-fashioned, they still remain important parts of the work of statisticians; and for these purposes, when it is applicable, the likelihood function is pre-eminent.

41. Another reason, perhaps, for the relative neglect of the study of the likelihood function is the undue concentration there has been on the method of *maximum* likelihood. When data storage facilities were extremely mediocre it was sensible to use the fact that the logarithmic derivative of the likelihood function is often nearly linear to summarize the description of it by specifying where its graph crossed the θ -axis (the maximum likelihood estimate), and the slope at that point (the standard error); if only two numbers could be stored, these two would often be the best choice in order to permit an approximate reconstruction of the likelihood function from them. But such need for economy is less typical now, and our habits should change to correspond.

42. Returning to the discussion of Bayesian approaches to inference, they have in common the view that something is needed to complement the likelihood function in the interpretation of experimental data. The empirical Bayesians suggest that the data should be complemented by other information derived from experiments involving similar parameters—not the same parameter, since if we have other experiments on the same parameter these will provide a likelihood function which can be combined with that from the current experiment to produce a resultant likelihood embodying all the experimental information. With the empirical Bayes approach, therefore, we have to face the problem of deciding what other cases of parameter estimation can be regarded as similar to the one we have in hand at any given time. It is here, it seems to me, that the special experience of actuaries should be highly relevant.

43. To illustrate the issue with a special case, consider the problem of estimating the accident rate, per 10,000 vehicle-miles, to 1500 c.c. cars of make X and year Y . Information is likely to be available about the accident rate for 1500 c.c. cars generally, about the accident rate for cars generally of that make, and perhaps to cars made in that year. If θ is the accident rate we want to estimate, we can say that θ belongs to population A , of all 1500 c.c. cars, and so has prior distribution $p(\theta, A)$; also, θ belongs to population B , of all cars of the given make, and so has prior distribution $p(\theta, B)$; and again, θ belongs to the population C of all cars produced in the given year, and so has prior distribution $p(\theta, C)$. We will very rarely have any information about the joint behaviour of makes, years and engine capacities, which would enable us to use the fact that θ belongs to all three of A , B and C , and so arrive at a resultant prior $p(\theta, A, B, C)$. Which prior, then, are we to use? As I understand it, this problem must be similar

to that which must be faced in deciding into what categories a set of insurance risks should be divided.

44. Another aspect of the problem presents itself when we have data only, for example, on 1500 c.c. cars. We know that the car of make X and year Y which we are concerned with is a 1500 c.c. car; but it was not one of the set on which the data about 1500 c.c. cars was based. It is a new type of car. Is the information on other 1500 c.c. cars relevant to the present one, or not?

45. The subjective Bayesian approach has won wide favour in applications to business problems, through the powerful advocacy of Raiffa and Schlaifer. Rigorously interpreted, there is little that one who thinks in terms of likelihood would find to disagree with in it, since so far as objective statistical behaviour is concerned it says that the likelihood function is precisely what needs to be communicated to anyone concerned to assess experimental data. It is then for each individual who has the information in the likelihood function to combine it with his personal prior distribution, and act or infer accordingly; but no one individual should try to foist his prior on to anyone else. Thus the practice sometimes adopted by adherents of this school, of adopting some particular prior as in some obscure sense 'natural' and reporting the posterior distribution resulting from combining the likelihood with this chosen prior, is really inconsistent with the viewpoint adopted.

46. Another possibly dangerous misinterpretation of the subjective Bayesian approach arises sometimes in connexion with risk analysis, venture analysis, and such techniques for the assessment of capital investment proposals. Instead of attempting to make single-figure predictions of future production, sales, prices, etc., and applying the usual discounted cash flow analysis to these returns, the forecasters are asked to set values L , U , say, such, for example, that they are willing to bet 10 to 1 that the actual figure will be above L and also bet 10 to 1 that the actual figure will be below U . They are asked also to give a value M such that they think the actual figure is as likely to lie above M as to lie below it. Then on the basis of L , M and U as percentiles, a probability distribution is estimated for the figure in question. Combining this distribution with others involved in the problem, arrived at in a similar way, it is possible to arrive at a distribution of values on net present worth, instead of a single-figure assessment. It is most important to realize that all that is achieved here by the application of the rules of probability is internal consistency of behaviour; there is no guarantee whatever that procedures based on this method will in the long run prove most profitable. For this to happen, the actual long-run frequency with which actual figures lie below L must be about $1/11$, the long-run frequency with which they are below M must be around $1/2$, and so on; and, perhaps most important and most difficult to achieve, the quantities which are treated as statistically independent in the calculations must in fact be approximately independent in the actual long run

of experience. It is most important that feedback provisions should be incorporated into any system of venture analysis to secure that such agreement with long-run frequency should (in the not-so-long run!) be achieved. Above all, those who construct the component distributions need to be quite clear that agreement with long-run frequency should be the aim.

47. The 'logical Bayesian' approach of Jeffreys and Lindley, according to which there are ways of expressing ignorance of a parameter in terms of a suitable prior distribution, is perhaps the most difficult one to discuss briefly. Perhaps I may be allowed simply to state my view, that *in general* such a description of total ignorance is impossible. But this does not rule out the possibility that in special cases—where, for example, the parameter θ in question may properly be regarded as an element of a group—we may be able to express ignorance. In the case of a single parameter, it is natural to use a prior measure which is equivalent to the Haar invariant measure for the group—with scale parameters, which form a group under multiplication, for instance, the measure with element $d\theta/\theta$ is invariant, since the integral of this from any point a to any point b is the same as the integral from ca to cb , for any positive c —both integrals being $\log b/a$. The uniqueness of the Haar measure guarantees the avoidance of the paradox that Boole arrived at from Bayes's postulate of uniform prior distribution. But when the parameter is not an element of a group—and sometimes when it is an element of a multidimensional group, difficulties arise. Dr Novick has used some order considerations, instead of group structure considerations, to arrive at natural priors in some instances. Both group approach and the ordering approach of Dr Novick, lead to results formally similar to those reached by the fiducial argument, in many cases; and my impression is that the connexion is not purely superficial. But just as the fiducial argument cannot always be applied, so I think we cannot always find a prior distribution which expresses ignorance. Jeffreys's suggestion that an international science commission should lay down priors in difficult cases is really no more than a way of suggesting that the likelihood function should be quoted in these cases—no one will really *believe* the conventional priors, and, since they are specified, it will be possible to deduce the likelihood from knowledge of the posterior distributions. Indeed, since in continuous cases (as Laplace remarked long ago) any smooth prior distribution for θ is equivalent to a uniform prior distribution for some function of θ ; and since with a uniform prior the likelihood and the posterior density are proportional to one another, the suggestion of an international commission to lay down standard priors amounts to proposing an international agreement to express likelihood functions in terms of certain types of parameter—and such an agreement would clearly have some merit, though it would not by itself justify integration of the resulting likelihoods. The attempt of Jeffreys and Perks to relate the prior to the distribution of the observations

meets and, I think, fails to resolve, the difficulty that one and the same parameter may enter into the distribution of observations in two different types of experiment, and so might have to be given two different prior distributions.

CONCLUSION

48. Instead of emphasizing the differences between schools of thought on statistical inference, I have tried above to emphasize their points of agreement. In agreeing, one way or another, on the importance of the likelihood function, the various schools are in a measure of agreement which goes a good deal further than is commonly supposed—particularly if we are prepared, as I am, to agree that the likelihood function may be interpreted sometimes with reference to a prior distribution when this latter is a suitable way of expressing information about the parameter external to the experiment in question.

49. Another point that should be borne in mind, relating to agreement between statisticians, is that I know of no one who does not in fact accept the simple significance test type of argument in those cases where it really falls to be used—indicated above, where the alternatives to the hypothesis being tested cannot be specified in parametric terms. And, perhaps most important of all, we must not forget that a great part of statistics is, and always will be, concerned with the ways in which data can be re-presented so as to suggest hypotheses and structure, rather than to test them in any formal way, or to estimate any formally specified parameters. This work, of 'data analysis' as it is coming to be called, is most important, and is having a tremendous phase of development thanks to the more powerful methods of data presentation that have come to us as a result of computers. And it owes little or nothing to the Bayesian, and nothing at all to the likelihood approach. It generates its own controversies, like any other lively subject—recent discussions about smoothing the periodogram may be cited as instances—but these are not excessively protracted. Thus we statisticians do really agree on major issues—and I include actuaries along with statisticians. Above all, we are united, I hope, in our view of the importance of our subject!

ABSTRACT OF THE DISCUSSION

Prof. G. A. Barnard, in introducing the paper, said that he felt not only honoured but particularly pleased to be given the opportunity of presenting the paper to a body of people who, probably for longer than anyone else, had been engaged in the application of mathematical doctrine and probabilities in the affairs of life.

In § 2 he had indicated his view that the controversy between the so-called subjective degree of belief theory of probability and the frequency theory of probability was false and that the one entailed the other.

Starting in § 4 he had made an historical review, pointing out, first of all, the interpretation of Bernoulli's theorem about numbers to give estimates of probabilities in terms of observed frequencies which had been current in the time of De Moivre, who shortly preceded Bayes who attempted to go somewhat further in giving a precise estimate of the probability limits for a probability. Bayes had postulated that an unknown event probability should be taken to be equally distributed between any two main degrees, but Bayes himself had stated that he had doubts as to the general applicability of that argument.

In § 8, he had drawn attention to the fact that the period 1830–40 had been one in which a great deal of discussion of the principles of the subject went on and he had named some of those active at the time. It was the time of the formation of societies such as the Statistical Society. His opinion was that a great many of the notions and ideas which had become current again within the past ten or twenty years had been current, possibly in a somewhat less precise form, around that time. In particular, Boole seemed to have had the clearest notions on the subject, and it was a remarkable feature of his work that he had been particularly clear on the fact that there might well be a problem which was posed quite correctly mathematically but to which the solution was not unique. Some had a habit of thinking that if a problem were correctly posed, there should be a unique answer. When contemplated in that bare form, the fallacy was obvious, but in a covered up version it affected a great deal of thinking.

After Boole, the subject had declined until the turn of the century, and in particular until Fisher had come on to the scene. The central thought of Fisher's work was rested in the notion of sufficiency, the idea that it was possible to take a great mass of data and to sum it up in a relatively small number of statistics which estimated the parameters. The aim was to secure that as much of the relevant information as possible passed from the original mass of data to the summary of statistics, the small number of statistics. In particular, when that was possible without the loss of any information—when the statistics summarized embodied all the relevant information in the data—the property of sufficiency was obtained. Fisher had been bothered with that for a long time, and in his day he had been restricted to thinking of summarization of data in simple numerical terms. Currently, however, with the graphical approach and computers, the summarization of data in the form of a function which could be drawn by a computer could be envisaged. Fisher had early thought that it might be possible always to find sufficient numerical statistics, but after fifteen years he had finally concluded that it would not be possible. It was remarkable that subsequently it had turned out that it was always possible to get sufficient statistics in the form of a graph of any likelihood function. That likelihood function always embodied all the relevant information in the data available.

He was discussing circumstances in which they were wholly satisfied that the probability model being used was appropriate to the data. Other modes of inference were

applicable when that was not the case. But when they were happy that their model of Normal distribution applied to the data, then the likelihood function seemed to sum up the information which the data had to convey.

In § 28 he had said that the physicists had already adopted the practice of quoting their results in the form of a likelihood function and that he understood that the geneticists were doing the same. He had indicated there and later how it was possible to think about likelihood functions and to interpret them in relation to data. That had not hitherto been common practice and some experience was needed if full use were to be made of the ideas.

The essential notion of likelihood and the idea of sufficiency derived from Fisher had been considered before, but there had also been the very important contribution of Jeffreys who had got away from the narrow concept of prior uniform distribution. Perks had made a similar suggestion at about the same time. There had, however, been a great deal of publicity lately in connexion with the subjective Bayesian school of inference who differed from Jeffreys and Perks. For example, Savage and De Finetti denied the possibility of expressing ignorance in terms of a uniform or of any other form of prior distribution but considered that it was the job of each individual to provide his own prior distribution.

He had drawn attention to the fact that, provided that that point of view was interpreted quite strictly, what it was reduced to was that what they should indicate to individuals about data was the likelihood function, because it was that which was multiplied by prior distribution to obtain the posterior basis for action.

In §§ 42 and 43 he had pointed out that it was an important feature of the likelihood approach that when an individual needed to use it to interpret the likelihood function in some particular way, if he wanted to take action on the basis of data, he could do it in various ways; he could supply the prior from his subjective imagination or from some empirical data about what he considered to be related information. An illustration of car accident rates was given in § 43. In fact, there was often some doubt which would be the most relevant information on which to base the prior and he suggested that the advice of actuaries, who were sure to have had problems of that kind to handle, would be extremely valuable.

In § 46 he had drawn attention to the fact that, in risk analysis or the subjective Bayesian approach to investment decision, it was sometimes overlooked that it was not enough that the probabilities entering into those calculations should reflect or know information about the subject matter. If the investments were to be profitable, it was important that there should be a close relationship between the probability and the actual frequency of the events in question. That was sometimes overlooked and it could lead to serious miscalculations.

Throughout the paper he had drawn attention to the fact that whilst statisticians, perhaps by nature, tended to disagree, as did any other group of people who were lively intellectually, they might be creating an impression of a great deal more disagreement than in fact existed. There was universal agreement with the principle of sufficiency, at least in its narrow interpretation, and from that followed the importance of the notion of likelihood. There were many problems which were described as data analysis where in any case the formalization of the problem was not sufficiently rigid to allow the application of any of the ideas which he had been discussing. In fact, while there might be disagreements on those matters, statisticians agreed on major issues—and among the statisticians he included actuaries—and above all they agreed on the importance of the subject.

Mr L. M. Eagles, in opening the discussion, said that in the paper the author had provided a clear and concise summary of the views of leading statisticians and philosophers on Bayes's theorem and Bayes's postulate since they were formulated midway through the eighteenth century. The latter part was concerned with an examination of the various schools of thought then current among statisticians concerning the nature of statistical inference.

The controversies as to the way to draw conclusions about the underlying parameters from data sprang from the deeper question of the definition of probability. Skill in the practical application of probabilities was considered one of the marks of the actuary and thus actuaries should have a definite contribution to make.

It was twenty years since the Institute had last considered the Bayesian controversy at a Sessional Meeting when discussing Perks's interesting paper *Some Observations on Inverse Probability including a New Indifference Rule* (J.I.A. 73, 285), so that a further consideration of the subject was timely.

There was no question that Bayes's theorem could be applied to derive the posterior distribution, where the prior distribution could be completely specified, for example the trivial case where a known density was sampled, but where full information on the prior distribution was seldom available. Should an attempt be made to incorporate general evidence about the possible nature of the prior in the statistical model, that could be regarded as admissible only if 'probability' could be regarded as distinct from 'limiting frequency'. For if those were identical, then it was clear that no probability statement could be made incorporating the investigator's feeling about the nature of an unknown prior. The statistician would necessarily proceed on the basis solely of the data with which he was presented—although it should be noted that he would have to make an assumption as to the form of the underlying density. As actuaries, however, they were familiar with time rates, such as mortality rates, which were certainly probabilities but could not, in his opinion, be entirely regarded as limiting frequencies. Therefore, he agreed with the author that to define 'probability' as 'limiting frequency' was not satisfactory, for to assert that was to assert that rational measures of probability which were not limiting frequencies—as irrational numbers—and yet obeyed the additive and multiplicative rules, were not probabilities.

It seemed to him preferable to define probabilities as measures of rational belief which obeyed the mathematical rules of probability, and then to examine how that mathematical model could be applied to the real world.

That approach might seem to attempt to define probability by ignoring the problems which philosophers had found. However, if actuaries were to be concerned with attempting solutions to practical problems, their chief concern was to know whether the mathematical theory of probability gave useful solutions to their problems; they wanted to know how they could apply those mathematical tools.

They should not, then, regard the Bayesian controversy as between truth and error. There were three major approaches to the problems of estimation and inference mentioned in the paper. Two were Bayesian; the use of an invariant transformation to derive the prior given the sample data, advocated by Jeffreys and Perks was one; the other was that of the subjectivists and empiricists both of whom adopted considerations extraneous to the data to determine the prior. Both of those approaches therefore derived a posterior distribution by combining the likelihood and a prior density. The third approach was that commonly attributed to Neyman and Pearson; that was to seek to divide the sample space into two regions according to the hypothesis it was desired to test, in such a way that for a given level of significance they maximized the power of the test. He thought

that the investigator should choose among those conceptual models according to the problem which he had in hand, keeping in mind the limitations of the model that he decided to utilize.

The paper demonstrated that the power functions of the classical theory were directly proportional to corresponding likelihood functions. Thence it should be noted that whichever model was adopted, the best summary of the data was the likelihood, and so it seemed to him that one of the most important suggestions was that advanced in § 39—that they should become familiar with the shapes of likelihood functions, which they had neglected by limiting attention to maximum likelihood. The Institute syllabus for statistics did not give any prominence to a description of likelihood functions. It was a subject that he had never studied in detail and he felt that omission ought to be remedied.

Another point which ought to be made was that not only did the likelihood afford a complete summary of the data, but it also exercised a greater influence over the posterior distribution than did the prior distribution, and that explained why statisticians on either side of the Bayesian controversy would draw the same inference from a body of data.

As actuaries, the problem of statistical inference—or, more strictly, estimation—with which they were most commonly confronted was that of determining the rates of mortality applicable to a given group of lives. The application of Bayesian methods to the problems of time rates was considered by Perks in his paper to the Institute. Perks obtained as the best estimate of the rate the ratio of the number of deaths to the exposed to risk multiplied by the number of unit time-intervals involved. However, that was a problem where an empirical Bayesian approach might be of interest.

It was frequently the case that, for a given group of lives, they would be prepared to lay very heavy odds that the mortality rate for a certain age lay in a certain narrow range. For example, in examining a life office experience, the mortality rates might be expected to compare closely with those derived from C.M.I. data. In that situation it seemed to him that it was quite legitimate to make an assumption about the prior distribution of q_x , and that useful results could be obtained even if the shape of the prior adopted were very elementary.

To give an illustration, suppose it were assumed that the underlying mortality followed a mortality table for which $q_x = 0.1$, that at age x , 1,000 lives were exposed to risk, and that there had been 80 deaths. Was there evidence of a real reduction in the rate of mortality? Suppose they assumed a 3-point distribution for the prior, the probability $q_x = 0.08$ being 0.1, the probability $q_x = 0.1$ being 0.8, and the probability $q_x = 0.12$ being 0.1, so that its mean and mode were at 0.1. Assuming that the number of deaths were binomially distributed, the posterior distribution could be derived by applying Bayes's theorem. That would also be a 3-point distribution with probability $q_x = 0.08$ being 0.5742, the probability $q_x = 0.1$ being 0.4257, and the probability $q_x = 0.12$ being 0.0001. Thus there was clear evidence of a real change in the parameter, and the best estimate of q_x might then be considered to be the mean of the posterior distribution, or 0.0885.

The same result could have been obtained in another way. It might have been assumed that the standard deviation of the distribution of deaths was about 10, and thence that 80 deaths was about two standard deviations from 100 deaths. Assuming the normal approximation to the binomial there was clear evidence at a high percentage level of significance that there had been a change in the level of mortality. The advantage of the method of using a prior distribution was that it showed the actuary how far he could go in making a cautious estimate of allowing an improvement in mortality.

Dr I. J. Good (a visitor) welcomed the opportunity of contributing to the discussion since subjective probability had been one of his ruling passions for over 30 years and he had made much practical use of it since 1940. He agreed with most of the paper but he thought that it would be more useful if he commented on points of disagreement and on extensions, rather than on points of agreement.

In § 1 there was a reference to the invariance theories of Jeffreys and Perks. For a t -category multinomial sample, their theories give rise to initial densities proportional respectively to $(p_1 p_2 \dots p_t)^{-\frac{1}{2}}$ and $(p_1 p_2 \dots p_t)^{-1/t}$, which could not both be right. They were concerned with the estimation of the physical probabilities, p_i . But it seemed to him that, if their arguments were sound, they should be useful for significance testing also—that was, for testing whether the equiprobable hypothesis $p_1 = \dots = p_t = 1/t$ was true or approximately true. He had found, however, that neither of those densities was adequate for that purpose; and the density proportional to $(p_1 \dots p_t)^{-1}$, as proposed by A. D. Roy and D. V. Lindley, was actually disastrous, since it forced the acceptance of the null hypothesis irrespective of the sample, except in the extreme case in which only one of the t categories was represented in the sample. (See *The Estimation of Probabilities*, M.I.T. Press, 1965, pp. 28 and 38; 'How to Estimate Probabilities'. *J. Inst. Math. Applics.* 2, 1966, 364–83, esp. p. 374; and 'A Bayesian significance test for Multinomial Distributions', *J.R.S.S. Series B*, forthcoming). On 22 March he would be giving a detailed paper on significance tests for the multinomial distribution from the Bayesian point of view at the Royal Statistical Society, so he would not say more about it until then.

He agreed with the authors about the importance of an examination of the principles underlying the formation of categories, so much so that he had advocated a new name for it—'Botryology'—based on the English prefix 'botry' from the Greek meaning a cluster of grapes. (*The Scientist Speculates*, Heinemann, London, 1962, pp. 120–32.)

In § 2 there was an implication that the limiting frequency definition had to be accepted if the addition and product laws were accepted. But that was misleading since the Venn limit, as it was sometimes called, applied to some infinite sequences that would not be regarded as random sequences and therefore was unacceptable as a definition of probability. It needed to be elaborated, as was done by von Mises, if it were to serve as a definition. But then it became complicated.

In § 9, the histogram was described as 'perhaps the most useful of all statistical devices'. He thought that counting was even more useful, and it led at once to binomial and multinomial sampling of which the histogram was virtually a special case.

In § 13, it was said that likelihood might be held to measure the degree of rational belief in a conclusion in an important class of cases. He would like to know what that class of cases was since he could not see how the class could be described without reference to subjective or logical probability. The likelihood was all right if either the initial distribution was uniform or else it was smooth and the sample was large.

In § 16 there was a reference to the fiducial argument. In his opinion that was an argument based on a simple fallacy except when it was interpreted in terms of a Bayesian argument, as was done by Jeffreys. It was about time that the fiducial argument was quietly forgotten.

In § 18 there should be a reference to G. F. Hardy's use of the beta distribution as an initial distribution, as an improvement on Bayes's postulate (1889, in correspondence in *Insurance Record*, reprinted in *T.F.A.* 8, in 1920).

At the end of § 20, there was a suggestion that initial distributions should not be used in the communication between scientists. That seemed to him to be wrong. He thought it useful to mention the implications of various assumptions concerning the initial

distributions, provided that the assumptions were stated clearly. In that respect the initial probabilities and distributions had much the same status as a statistical model of any other kind.

Regarding § 24, it was not true that H. E. Robbins originated the empirical Bayes method in its *mathematical* form. Turing had done so in 1941, in connexion with the estimation of the probabilities of species, and he (Dr Good) had published that work with extensions in 1953 (*Biometrika*, 40, 1953, 237–64; see also I. J. Good and G. H. Toulmin, *Biometrika*, 43, 1956, 45–63). When he (Dr Good) had lectured on probability estimation at Columbia University a few years previously, Prof. Robbins himself had pointed out that, in the species problem, the method which he had described was an example of the empirical Bayes method. He had felt like the man who discovered that he had been writing prose all his life.

In § 28, it was said that θ_1 was more plausible than θ_2 when its likelihood was higher. If ‘plausible’ meant ‘probable’ then Bayes’s postulate was implicit. If it did not mean ‘probable’, then he would like to know what it did mean.

In § 34, it was said that if their observation had very tiny probability they would be suspicious of their model. But observations of continuous variables nearly always had very tiny probability. He thought that that difficulty could be resolved by making use of the ‘surprise index’ (see, for example, *Annals of Math. Statist.* 27 (1956), 1130–5 and 28 (1957) for corrections).

Regarding the problem considered in § 43, some relevant work on the estimation of small probabilities in contingency tables was given in *J.R.S.S. Series B*, 18 (1956), 113–24, in *Annals of Math. Statist.* 34 (1963), 911–34, and in his own M.I.T. monograph mentioned earlier.

In § 48 there seemed to be an implication that there either was or was not an adequate initial distribution, that everything was black or white. The partial ordering of probabilities, which had been emphasized by Keynes, Koopman, Good and C. A. B. Smith, showed how various shades of grey could be introduced. Black and white were merely extreme forms of grey.

Again, to say that data analysis, or ‘datum analysis’ as it should be called, owed little to the Bayesian, was to say that it had little logic behind it, that it was an art or technique rather than a science. It used mathematics and computers as tools, but Picasso had suggested the use of a painting machine as an adjunct to the artist. When datum analysis became more scientific it would probably be based on two things, first, the psychology of perception applied to the presentation of data in easily assimilable form, and, secondly, the ordering of hypotheses according to their simplicity, where the simplicity of a hypothesis would be related to its subjective or logical probability. Since the psychology of perception depended on information theory and subjective probability, it seemed to him that scientific datum analysis would lean heavily on Bayesian methods. To say that a statistical technique did not depend on subjective probability was tantamount to saying that it did not depend on rationality. To the Bayesian, all things were Bayesian.

Prof. M. S. Bartlett (a visitor), expressed his appreciation of the author’s survey and congratulated the Institute on having arranged it. All agreed that the questions were fundamental and the attempt to reconcile various views was to be praised. All he wanted to do was to underline some of the important points which the author had made.

The first was to emphasize that there was a division between probability as a belief and probability as something founded on frequency. In some sense frequency had to be

seriously considered as a basis for any probability which was to be of practical use. There was not one Bayesian School but at least three. Two of those used subjective probability in one sense or another. There was the subjective or personal probability of Savage, Good and others, and the slightly different view of Jeffreys, where a certain amount of convention was used in assessing prior probabilities. The author noted that that meant that an appeal would have to be made to an international commission to have those approved, and that would be a very dangerous situation. The other view of prior probabilities was certainly not applicable in all cases, but it was important to consider when it was applicable. It considered prior probability in some sense to be based on frequency.

That view tied in with the empirical approach of Robbins or Turing. In the company of actuaries it was especially important to stress that approach because if they were considering the application of probability methods in insurance and actuarial work, it would be no good, if an insurance firm went bankrupt, to plead that the assessment of probability on which various risks and approximations had been based, in fact had been entirely consistent in the actuary's judgment with his own various probabilities! That would not be received very favourably by the people who had lost their money!

The various issues should be judged by their success or failure. To a certain extent that was basing them in the long run on a frequency basis. It was important to stress the difference between an empirical attitude of some kind, where various methods were based on results, and just pleading internal consistency.

He did not entirely cavil with the emphasis in the paper on the importance of the likelihood function, but he had some niggling doubts about the value of the study of likelihood functions in general. He was not so keen on the comment that there was now no difficulty in having many tapes from computers containing all sorts of results. There was the situation in which a summarization of the data was demanded—not more and more results—whether they were likelihood functions or functions of an unknown parameter or anything else!

As for the summarization, it could be argued that the original data were equivalent to the complete likelihood function, perhaps better because it did not depend on any specific probability theory, which might be inaccurate.

In § 3, the author said that there was an analogy with two different definitions of temperature—as the function of state variables which was the same for all bodies in thermal equilibrium, and as the energy of motion of molecules of an ideal gas. In fact, the definitions of terms of mechanics demonstrated more insight into the subject than the thermodynamic view. And even the thermodynamic view was rather more concrete than the analogy of probability as a subjective probability. He remembered Prof. van Dantzig comparing the subjective assessment of probability with the attempt to assess temperature by measuring their own feelings, as to whether it gave them the shivers or not! That would not be a method of dealing scientifically with temperature!

Prof. J. Durbin (a visitor), expressed pleasure in being invited to attend the meeting and being invited to participate in the discussion. He added his congratulations to the author for his excellent review of a basic problem. He also coupled with that a personal tribute to him in gratitude for all he had learnt over the years from his many public discussions of statistical inference.

Indeed, his first serious doubt about the basis of the sampling theory approach to statistical inference had arisen after the paper presented by the author before the Royal Statistical Society with two colleagues, four or five years previously. In his contribution

to the discussion on that occasion he, (Prof. Durbin), had expressed concern about the justification of conditional tests. Thinking about that further had led him more and more to doubt the basis of conventional statistical inference.

Discussions of inverse probability versus the sampling theory approach had of course taken place many times over the previous 100 to 150 years. He did not agree, however, that that was a quasi-religious question on which the same things were said time and time again, because a great deal of theoretical work had been done as a result of which genuine progress had been made over the past 60 years. Many avenues of hope had been followed up and many good ideas had been demonstrated to lead nowhere.

In his view the key to the understanding of the problem lay in a study of the development of Fisher's ideas. Early in his career Fisher had taken up the firm position symbolized by that graphic phrase in his *Statistical Methods* 'The theory of Inverse Probability is founded upon an error and must be wholly rejected.' Fisher had been one of the first people, however, to recognize some of the difficulties in the sampling theory approach and he had introduced the ideas of conditional tests and fiducial theory as a way of meeting them. In doing so he appeared to the speaker to have been groping for a theory not too far from a Bayesian system but without the subjectivist associations of the latter. He recalled the keen expectation with which many had looked forward to the appearance of Fisher's book in 1956, hoping that it would present a coherent system. It had been a disappointment to find that he was unable to complete the theories which he had started so brilliantly.

Since then, many other outstanding people had worked on problems of conditional tests, fiducial theory and the likelihood principle as providing a basis for inference without bringing in a prior distribution. It was becoming generally agreed among professionals that those inquiries had led to disappointment, at least as far as developing a coherent system of inference was concerned.

Personally he took a somewhat intermediate position between Dr Good and Prof. Bartlett and wished to say a word about the implications of such a position for the teaching of statistics in universities. Many were moving to a position where they could see some merit in approaching some inference problems from a Bayesian point of view but would not want to base their teaching primarily on that standpoint.

During the previous 50 years conventional mathematical statistics had developed into a subject of great intellectual distinction. While teaching it he had felt that its educational value was enormous, both in the nature of the problems posed and in the methods of solution. An advanced course on statistical inference based on the sampling theory approach had a very substantial content. If they were limited entirely to a Bayesian standpoint based on personal probability, almost all that was cut out in one stroke. Essentially, all they currently needed to teach was how to set up likelihood functions together with a few general rules on point, and perhaps interval, estimation. The intellectual content would be very small compared with that of the previous theory. If they wished to be less personal and to develop teaching from the more objective approach of Jeffreys and Lindley—he had tried that in the past year—it was found to be more difficult to convince a class of students of the credibility of the approach as compared with that based on sampling theory. He concluded that a switch from the sampling theory approach to a Bayesian approach would have very substantial implications for the content of courses on statistics at universities, the consequence of which could not yet be fully foreseen.

In § 48 the author had mentioned that the differences between schools of statisticians were not great. Too much should not be made of that point because in the speaker's

view the similarities were largely accidental. That arose in part because of the remarkable properties of the Normal distribution and in part because most of the problems which Fisher had used as examples had a group structure which made them amenable to the Jeffreys type of treatment so that both methods gave the same answer. He wondered whether that was to some extent a mathematical accident and not because of any essential similarity of views.

The author's work on statistical inference had had a considerable impact on the speaker and he had therefore scrutinized the paper carefully to see whether he could discern any move in the author's views since the paper before the Royal Statistical Society in 1962. At the end of § 48 the author had said that 'the likelihood function might be interpreted sometimes with reference to a prior distribution when this latter was a suitable way of expressing information about the parameter external to the experiment in question'. He would like to ask the author the crucial question, whether in any circumstances he would be prepared to regard subjective knowledge as information for that purpose or whether he was speaking only about some kind of objective distribution of the parameters.

Mr F. M. Redington said that he might be misunderstood if he said that he was not very well acquainted with absolute ignorance: better to say that he had little right to be speaking in such distinguished company because his detailed knowledge of the subject was small and rusty. There was, however, something which he had wanted to say for many years and which he had tried to say in a paper on *Ideas and Statistics* to the Students' Society twelve years earlier. Reading the author's paper, he realized that the battle was being fought and won by better men than himself, including the author. But he would like to say it in his own way.

From the numbers present at the meeting, it did not seem that many members felt that prior probabilities were of much concern to them. They were, however, the very fibre of their everyday life. He would take a simple example.

A did something. B said, 'You know why A did that'—and proceeded to analyse A's motives, usually to his discredit. Given that those were indeed A's motives, then the posterior probability of A's action was certainty. But the explanation was known to be wrong because it was entirely at conflict with the known prior view of A and his make-up. Human beings seldom differed about the posterior probabilities. It was always in the prior probabilities that the real differences were to be found.

The particular thought which he wanted to express was that they could never finally judge a population from a sample. Returning to his example, suppose they wanted to feed into a computer all the facts they could gather about A in order to judge the prior probabilities of his action. It would be necessary to put in A's history, which included his environment, which included his parents and the society in which he lived, which would include their parents and their society and so on. There was no stopping until the whole of the universe, past and present, and, some philosophers would add, future, had been fed into the computer.

The fact was that the universe was indivisible, and a good deal of the attempts at statistical inference were essentially attempts to sub-divide it—indeed, to detach science from life. He did not think that that could be done, other than in examination papers and other fairy tales. He was sorry that mathematicians did not quietly accept that fact, because it was not only helpful but comforting.

It was helpful because science was always at its most arid when sequestered. It reached

greater and greater heights of refinement on greater and greater trivialities. The life blood of science came from outside itself.

He found it comforting because when he came to assess prior probabilities there came a point at which he ought to say to himself, 'Here, I am at the limit of pure reason. Beyond this point lies the universe. I cannot pass, but nor can a computer. Nor for that matter can Big Brother.' He thought that those remarks were true, and obviously true for the real world of *A's* motives. In the final analysis they were true of every statistical judgment.

There was—however tiny an element it might be for large random samples—a residual uncertainty about the prior distribution which he might call the universal uncertainty. In a way, that could be called a subjective approach, because the way out could only be subjective. But on those grounds a theatre might as well be described as an exit house. The concept needed a positive, almost aggressive, description. It was an indivisible or unitary concept.

There was, and had been for 4,000 years, a certain suicidal tendency in science, to cut itself off from life and live in a closed aseptic world of its own. He had felt that deeply in his younger days when significance tests had been rife and much brilliant and profound nonsense had been talked. It was consoling to him to read in many passages of the paper that more vital views were returning. There was much in the author's own expression of view with which he had sympathy, but the author had not said it in the way in which the speaker wanted it said!

Mr L. V. Martin suggested that the whole concept of probability was an expression of lack of knowledge. If the answer was known there would be no use for probability theory at all. He thought it reasonable to say that the probability that the top card of a shuffled pack of cards was a spade was $\frac{1}{2}$, but that was the probability only if they did not know what that top card was. If they had had a surreptitious peep beforehand, probability ceased to enter into it. If an event had happened but they did not know how it had turned out, or if it were about to happen and they did not know how it would result, the probability of success would be calculated on the basis of all the knowledge possessed at the time. That estimate might be changed by later information, until doubt was swallowed up in certainty. The difference between the estimate of probability at any time and the eventual unity or nothing, was due solely to ignorance. With inverse probability the same consideration applied. Imagine a condition in which they had no knowledge of prior probability whatever; then a sample could reveal next to nothing about the posterior probability. If they picked a ball out of a bag n times, replacing it each time, and each ball was red, they could make no logical deduction about the bag's contents, except the trivial one that there was at least one red ball in the bag. Someone might have filled the bag with white balls and added a single red ball to them.

The problem of scientific inference was also like that. Unless they made some assumption about the nature of the world, they could make no progress. Because things behaved in a regular way, they came to believe in the repeatability of experiments. If a particular chemical experiment gave the same result half a dozen times, they were prepared to believe that it would always give the same result. That was because they had assumed a prior probability distribution largely concentrated at nought and one; this was assumed because experience showed that there were many 'nought or one' cases in the universe, but the argument was not logically watertight. It therefore seemed to him that any search for a prior probability distribution to represent complete ignorance was rather like looking for the end of a rainbow; they would never get there. If they

approached a problem in complete ignorance, they could not assume any one possible prior probability distribution to be any more likely than any other. All that a sample would do would be to eliminate some of the extreme cases. The experiment could tell something useful only if they implicitly assumed a prior probability distribution, either an objective one based on their knowledge of how the universe was made up or a subjective one based on their whole experience of life. In fact, that was what they did in practice time and again without realizing it. Unless they accepted, as they had to, that they knew something of the universe being sampled, they could not obtain more than the most trivial information from a sample. If scientific inference were to work, it could not be based on a sound logical basis; they had to make an act of faith, and to base their calculations on a subjective estimate of the prior probabilities involved. There seemed to be nothing which could represent complete ignorance other than all the possible prior probability distributions that could exist combined in unknown proportions.

Prof. P. G. Moore congratulated the author on his clear and detailed exposition of the Bayesian approach to the problems of inference. Many practising statisticians treated those problems of inference in a manner which tended to suggest that there was no difference in practice between Bayesians and others. Put in a nutshell, the Bayesian approach sought to provide a unified mathematical model of the way the mind worked in all the varied situations where inference was required. Such an ideal was intellectually attractive but, to put it in its setting, it was wise to consider, first, the differences between the Bayesians and the scheme of inference commonly associated with Neyman and Pearson—although in fact that scheme commonly included threads gathered from many areas—and secondly, the effect that the prior distribution had upon the decisions made.

To highlight those two facets of the procedure and thus enable attention to be concentrated more firmly on the important issues, he would base his remarks on an illustration that had been used before, that of King Hiero's legend. That had been first expounded by L. J. Savage at a conference in London in 1959, and subsequently published in 1962 as part of a monograph entitled *Foundations of Statistical Inference*. The discussion of the legend was later considerably extended in a talk by Prof. E. S. Pearson given at Cornell in April 1961, subsequently published in its turn in the *Annals of Mathematical Statistics* in 1962.

Briefly, the legend ran as follows: King Hiero had ordered a new crown and believed that the goldsmiths might have adulterated the gold, either with lead or with silver. Secondly, Archimedes had hit on the idea, presumably unknown to the goldsmiths, of determining the density of the crown by weighing it in air and in water and also weighing a specimen of known gold in both air and water. Thirdly, by that test Archimedes was assessing some quantity λ by means of a measure Y , which in turn might itself be the mean of n independent test results x_i . Fourthly, for pure gold, the procedure could be arranged so that the $\lambda = 0$, whilst for lead $\lambda > 0$ and for silver $\lambda < 0$. Fifthly, Archimedes knew from earlier experiments that Y was Normally distributed about λ with known standard error σ .

On the 'standard' significance test approach the King selected some small value α in the range $0 < \alpha < 1$ at his discretion. Values such as 0.05 or 0.01 were commonly chosen. He then computed the probability that the quotient Y/σ , where σ was the standard deviation of Y , would be at least as large as the observed value, given that the true value λ were 0. If that probability were less than α , the King should reject the null hypothesis that $\lambda = 0$, and hang the goldsmiths, otherwise he should accept it.

The King, it was assumed, attached some credence to the possibility that there had

been no cheating—i.e. that $\lambda = 0$, and also believed that there was some defined prior distribution of λ , conditional on cheating. He wanted to hang the goldsmiths if at all guilty, otherwise not. The Bayesian approach sought to bring that prior assessment of the situation into consideration with the experimental results, to form a posterior assessment of the situation.

Calculation along Bayesian lines showed that in that situation the final or posterior odds in favour of innocence ($\lambda = 0$) were a multiple of the initial or prior odds, the multiplier being dependent on the initial probability density for λ (given $\lambda \neq 0$).

To obtain some representative results, assume that the standard deviation of the observed measure Y is 0.25, and that the King only hanged when the posterior odds on guilt were at least 10 to 1. Then consider two cases: in the first the prior probability density for λ was $\pi(\lambda) = 0.20$ for $-1 \leq \lambda \leq 2.5$ and zero elsewhere. The critical value for Y then fell at $Y = 0.61$. That cut-off point for Y corresponded to a conventional significance level of 0.015.

In the second instance, the prior probability density for λ was $\pi(\lambda) = 0.05$ for $-1 \leq \lambda \leq 4$ and zero elsewhere. The critical value for Y was now 0.85 which corresponded to a conventional significance level of 0.0007.

Thus, keeping the posterior odds situation fixed, the critical level seemed, as ought surely to be expected, to be sensitive to the prior distribution adopted. That could mean that a model which was clarifying in theory might be difficult to apply in practice because it called for the introduction of parameters whose values did not, in many instances, really exist. Alternatively, the lesson to be learnt from that analysis might be that the method of approach was of value just because it forced the King to face up to issues which he would otherwise have failed to appreciate fully.

The analogy could be readily transferred into the area of business decisions. For example, a capital investment decision regarding the size of a plant to install would instinctively be judged by the notion of a prior distribution of the likely market. That instinct could, and often would, be modified as the result of a market survey and the prior distribution turned into a posterior distribution, from which the necessary assessment and decision was made.

The fact should, of course, be borne in mind that much of statistics was concerned with estimation rather than decision-making. As the author had remarked, textbooks customarily emphasized the latter at the expense of the former, yet practising statisticians spent a large part of their time advising on estimation situations. An appropriate sense of perspective had, therefore, to be maintained when considering the role of Bayesian statistics within the spectrum of statistical activity.

Mr T. H. Beech said that he was afraid that in their actuarial training many of them had reached perhaps too easily the conclusion that the practical effects of the deep philosophical differences of principle tended to be minimal in most circumstances. Although they were probably wrong to let that view sway them into some neglect of those conceptual problems, nonetheless it was reassuring to find no less a statistician than the author confirming the essential validity of that conclusion.

However, the return of Bayes's rule to a more central position through the revival in respectability of subjective probability not only had an influence on existing procedures but opened up possible treatments of situations previously thought intractable. In § 46 the author warned of the dangers of a too free use of subjective probabilities assessed by asking questions of the betting odds type, although admitting that that procedure at least led to consistent behaviour, a situation by no means typical in the business

context. He felt that the dangers of that approach to capital project analysis should not be exaggerated. Even if still somewhat tentative it appeared to point the way forward, and results should improve as refinements were developed. One area in particular need of research lay in the psychometric field. It was necessary to establish more efficient means of bringing into quantitative form the subjective probability pattern lying in the manager's subconscious. Asking for a figure on which he would require odds of 10 to 1 was a big step forward from asking questions in terms of a probability of x per cent, but there was still an area of unreality remaining. Managers did not customarily think of their projects as betting operations. He had a feeling that there was still just beyond present grasp a more realistic way of framing the question. Of course, in years to come they might see a breed of managers who thought in probability terms, but that was a long way off, and in the meantime they had to do the best they could. Incidentally, while feed-back of actual results versus probability forecasts was desirable, the manner of the feedback needed careful consideration. Still dealing with a manager unused to probability thinking, there might be much to be said in principle for not disclosing to him his long-run results, the analysts instead adjusting the probability levels appropriately. In practice, however, that approach might be impracticable owing to its giving rise to a suspicion of chicanery, an impression which should at all costs be avoided. However, the difficulty about disclosing results to the manager and letting him do his own corrections was that he would probably tend to over-correct and a damped oscillation might be set up. In any case, the relationship between actual and expected results would be unstable for longer than it need be. His recent remarks (*J.I.A.* 93, 211) on a similar difficulty in the context of subjective forecasting of Stock Exchange investment parameters were relevant.

Probably the only real solution, albeit a long-term one, was not only to have managers trained in the probability approach, but to encourage them to make subjective probability estimates as a matter of habit on day-to-day matters so as to develop their abilities in that direction to a high level. It was suggested that continual practice, with the necessary feed-back of results, would go a long way to solving the key problem of bringing to the surface the subconscious probabilities. That in itself would be a great step forward. More important still, each manager would, through the feed-back of results, be enabled gradually to improve his own probability thinking.

On a more pedestrian level, they might hope in any continuing programme of estimation for a gradual crystallization of prior probability patterns through the feed-back of data from relevant previous exercises. The author mentioned in that field work by Robbins. Would he please give a detailed reference?

One difficulty which he had always found in accepting the utilities approach was the fact that the entire preference pattern with regard to uncertain profits and losses of amounts below a reference level, of, say, £100,000, was determined on the basis of the certainty equivalents corresponding to different probabilities of receiving or losing the full £100,000. That result was difficult to accept from the common-sense point of view. He found it difficult to believe that a person's attitudes to various opportunities involving relatively small gains and losses could be predicted purely from a knowledge of his attitude to gains and losses of a single reference amount chosen as being larger than any of the possible outcomes. While admiring the virtuosity of the argument, with its repeated jumps backwards and forwards between mathematical and subjective probability, he could never put away an uneasy feeling that one of those jumps was, somehow, not legitimate. He had a feeling that there should be a price to pay for the reduction in dimensionality produced by the use of a single pair of reference amounts. If the price

paid was that the results obtained were purely normative and not realistic, it could be that the price paid for the convenience of a single pair of reference levels was too high, particularly when those levels were chosen to be outside the range of levels of interest. While that choice avoided the dangers of extrapolation, it probably reduced the accuracy of the certainty equivalents which the manager quoted and on which the whole exercise was then based. Perhaps the author would comment.

A pleasing feature of the paper was the number of matters touched on in brief, but potent, asides. It was interesting to find, for example, the minimax principle cut down to size. Until recently, at any rate, the attractively simple logic involved appeared to have led to the use of the principle in many contexts where such use could lead to over-conservatism. In the field of military strategy, for example, the assumption that any real or potential enemy could or would act in the optimum manner as seen by the minimax principle seemed quite unrealistic when the tremendous inertia embodied in any major government-military machinery was considered. Neglect of the bounds set by such inertial features could lead to expensive protection against possibilities that were quite unrealistic, to the detriment of the pursuit of other more rewarding lines of approach.

Finally, he would comment on the example given by the opener of the use of prior probabilities in connexion with mortality rates. The great objection to the 'significance' approach was the essential discontinuity of the process. Followed strictly, it led to a straight choice between accepting the null hypothesis (which was almost certainly not strictly true) and jumping to the average of a sample as a new assessment. Such discontinuity seemed to go against the pattern of life. Using the Bayesian approach, particularly when based on a continuous prior distribution, each new experiment permitted, by feed-back, a change of opinion, which seemed both realistic and rational. It might be that in practice the adjustment to the mortality table turned out to be slight and it could be ignored, but it was better that that should be a conscious choice on grounds of practical expediency, rather than an automatic result arising out of a probably arbitrarily chosen significance level.

He had been a little worried when the example was introduced in terms of a prior distribution concentrated at three points, although the final step of averaging according to the posterior probabilities eventually showed the 'continuous adjustment' feature. However, that feature was a particular virtue of the Bayesian treatment and as such it might have been demonstrated more naturally if the example had been based on a prior distribution which was continuous, such as the Normal, rather than on one concentrated at three points.

Mr W. Perks said that when he had been invited to close the discussion he had felt grave doubts about his ability to do justice to the paper and to the discussion—although at that time he had not seen the paper. While parts of the paper, which he had enjoyed reading, were over the top of his head, he had kept in touch with most of it. He had also enjoyed listening to the discussion. The Institute was fortunate indeed that the author had produced such an interesting paper which had provoked such an interesting and valuable discussion. Fortunately, it was not his but the author's responsibility to reply.

He had been glad to hear the opener come down on the side of those who believed that the limit of relative frequency was not enough for all probability situations. He had committed himself to that position in 1947 and was glad that since then so many others had come round to the same point of view. He had not been entirely alone in that position but many actuaries and statisticians had not been prepared to accept that anything

more than the limit of relative frequency was required. He had been treated as something of a heretic in 1947 and now found that he had become almost orthodox.

He had been glad also to hear the opener support the author in the suggestion for a study of the shapes of likelihood functions. That was important. It arose in many statistical problems and played by far the greatest part in the posterior distributions. The opener had referred to using past experience in combination with current data for time rates such as mortality rates. He was unhappy about that, because he would not know how to fix the relative proportions to be used for the past experience and the new experience respectively. He might do something like it for his own purposes but would not like to make public any such subjective combination.

That kind of procedure had, however, long been studied in America by the Casualty Actuarial Society under the name of 'credibility procedures'. It was important to appreciate that the word 'credibility' was used there in a different sense from that in which Dr Good, for example, referred to certain Bayesians as 'credibilists'.

He looked forward with keen interest to Dr Good's forthcoming paper on significance tests for the multinomial using Bayesian theory. His own 1947 paper did not deal with significance. He was still not quite sure whether indifference priors had any relevance to significance tests. He rather thought that Jeffreys was right in giving a finite prior probability to the null hypothesis. It did not stand a chance otherwise. If a null hypothesis was being tested, there should be a good reason for posing it.

Dr Good had referred to G. F. Hardy's work on Bayes in relation to his suggestion of the beta distribution to express prior information regarding the binomial parameter. He paid tribute to Dr Good for the punctilious way in which he always gave his references and acknowledgments. Dr Good appeared to say that a prior distribution was in effect an assumption rather in the same way as the adoption of the sampling model was an assumption. If that was what he meant, then he agreed with him.

Prof. Bartlett had referred to the fact that Jeffreys had suggested that in cases where there were difficulties in fixing a suitable indifference prior, they might be settled by international agreement—and he had added that that would be dangerous. In fact, he did not think that it would be very dangerous, because any prior rules that had a chance of acceptance would have very little effect in practice on the posterior distribution as anybody who had done any arithmetic in that field would appreciate.

Prof. Bartlett was right that in insurance matters and when quoting premiums relative frequencies (and, of course, average claims) were of the utmost importance and it might be that the neglect of objective statistics was a major cause of recent events in motor insurance. He (the speaker) stressed that he was not personally interested in prior distributions in practical affairs. The place of prior distributions was at the foundations of the subject.

Prof. Durbin had spoken interestingly about the advantages and disadvantages of basing a teaching course on Bayesian methods. The advantages of the complete unification that became possible in statistical inference seemed overwhelming.

An enormous amount of work had been done by mathematicians and philosophers on the Bayesian theme in the past 10 years but it was not possible for a busy actuary, trying to keep abreast of practical affairs affecting the responsibilities of his job, to read everything on such a wide-ranging, theoretical subject, even if, like the speaker, he was keenly interested in it. Something of importance could easily be missed. But he was comforted by the thought that leading statisticians had also been guilty of that.

For example, he could not understand how anyone could still support the Haldane prior for the binomial parameter (i.e., $x^{-1}(1-x)^{-1}$) since it had long ago been shown

to produce quite unacceptable answers in cases where the sample results were all successes or all failures. Similar objections also applied to the corresponding multinomial rule. He believed that any prior rule purporting to express total ignorance ought to stand or fall by being tested to destruction. As far as he knew, the Jeffreys-Perks binomial prior had stood up to that test and he knew of no other that had done so. Until recently he had thought that his multinomial prior similarly stood up to that test, but he had to confess—as had already appeared in print—that Dr Good had effectively shown cases in which there was trouble, although he felt that it was possible to preserve something of the rule by distinguishing between prior and posterior groupings of categories.

He felt that the author had rendered a valuable service by showing the important place taken by the likelihood as a common factor in the many different approaches to statistical inference. The likelihood clearly expressed the essential content of the observations but of course it did more than that. It was also essentially dependent upon the sampling model assumed for good or perhaps indifferent reasons. As the author said in § 34, they were taking the model as gospel.

The sampling model played such an important part in the likelihood that he found it difficult to understand why there was such an objection by some people—he was not sure whether the author was one of them—to going a tiny step further and using properties of the sampling distribution to determine the prior distribution expressing total ignorance. After all, Jeffreys had shown long ago that any normal change in the prior distribution had no more effect on the posterior distribution than just one observation more or less which was a much smaller effect than would be produced by quite moderate changes in the sampling model itself.

It was a little odd to see Prof. Moore reaching a similar conclusion in his recent paper in the *Journal* (*J.I.A.* 92, 326), without once mentioning Jeffreys. Prof. Moore seemed to imply that whether or not there was any real doubt about the mean of the prior distribution, its standard deviation and shape could vary quite considerably without making much difference to the posterior distribution.

Why did he stress the need for indifference priors? The author had mentioned several times that the likelihoods were not additive. He (the speaker) suggested that they needed to deal with them in such a way that they effectively become additive and integrable. That was the real function of the indifference priors. Without additivity he did not see how they could make interval estimates or embody utilities in their estimates. Even point estimation of continuous parameters, apparently based on the likelihoods alone, implied an arbitrary metric for the likelihoods because they inevitably used a limited number of decimal places. Estimating to one decimal really involved interval estimating with a range of plus or minus 0.05.

The subjective Bayesians, with whom he had much sympathy, used prior distributions to express prior information. He would do the same willingly, but he claimed that whenever he did that he was implicitly using an origin for nil information because without an origin the expression of the actual prior information was on shifting sands. In a particular situation there might be a certain amount of prior information expressed by a prior distribution. If they took some information away, and then some more, each information situation should provide the basis for a subjective prior. When all was taken away, what was left? Zero information. To be a Bayesian, it was essential to have some expression for zero information, otherwise there was no starting point.

It was true that usually prior information could be expressed only approximately and so a vague origin was good enough in practice. But those were not really practical questions at all. They were matters of fundamental principle and the expression of an

origin for nil information based on the properties of the sampling distribution seemed a necessary and perfectly proper ingredient in a sound system of statistical inference.

His position about using prior information was clearly indicated in his 1947 paper when he rejected the invariant prior for time rates such as mortality rates. By using dx/x he had suggested in effect lumping up the prior probabilities towards zero. In fact, actuaries were usually concerned about a set of probability or mortality rates which all influenced inferences about each other. That was what their graduation processes were about. Dr Good had shown that multinomial sampling also often involved pattern between the individual categories.

At the end of his paper, in § 47, the author counted it as a criticism of the invariant prior rules that they might require different prior distributions for the same parameter when the sampling procedures were different. On the contrary, he (the speaker), regarded that as a point in their favour. He had always found it rather odd that statisticians should be content that statistical inferences should not be affected by the nature of the stopping rule in sequential sampling. That did not seem reasonable to him. It was true that the same observations might yield the same likelihoods, and the only source from which different inferences could come was the prior distributions. The stopping rules ought, normally, to have an effect on the posterior distributions. It seemed perfectly sound to him that the prior distribution should depend on the sampling procedure which determined the sampling distribution. For example, for the binomial parameter they usually thought in terms of fixing the sample size n and observed the number m of successes. If, instead, they went on selecting until they had a fixed number m of successes then the random variable was n and the sampling distribution was different. The likelihoods were in the same form, but he would not be at all surprised if the invariant prior rules were found to be different.

He agreed with the author that all the difficulties of estimating two or more parameters at a time or of joint posterior distributions had not been resolved, but he believed that those problems arose when there was dependence between the posterior distributions. Rather than reject the whole business of indifference priors, he would like to see the mathematicians trying to resolve the remaining difficulties. It would be helpful if the author would explain more precisely what he had in mind in his references in § 17 to structures. Was he right in thinking that the patterns in Dr Good's multinomial distributions and a set of mortality rates analysed by age would be examples of such structures? If so, he suggested that provision for such structures ought to be allowed to influence the sampling model.

He was glad to see that the author did not hold it against the invariance theory that it sometimes used what were often called 'improper' distributions. It was a pity that in an otherwise remarkably interesting recent book on the *Logic of Statistical Inference* the author Ian Hacking should hold 'improper' distributions as a conclusive reason for rejecting the invariant rules. He thought that he had disposed of that argument twenty years before in his discussion with Joseph.

In several places the author suggested that the experience of actuaries might be helpful in relation to various aspects of the Bayes problems. He was afraid that the author would be disappointed, however, because actuaries very rarely dealt with small samples in isolation and the question of the prior distribution was so insignificant that they never bothered about it at all. Moreover, they always sought to introduce in the estimates a margin for contingencies and, might he say, profit.

A number of actuaries over the years had written significantly on the Bayes theory, including G. F. Hardy, Makeham, Calderon and Lidstone. There had been a paper at the

Faculty before the war by Anderson and Reid in which they combined past experience with new data through the Bayes theorem.

He had found the author's reference to the current work on data analysis of considerable interest, particularly the reference to 'smoothing'. It would be ironic if statisticians became interested in the traditional actuarial graduation processes at a time when the interest of actuaries had waned.

He could not conclude without expressing the pleasure which he had derived from the paper and at finding that his own position in the subject had become almost orthodox. He also offered sincere thanks to the author for his generous references to his own modest contribution to the subject 20 years previously.

The President (Dr B. Benjamin), in proposing a vote of thanks to the author, said that the author was an eminent mathematician and the Institute, which did not often enough listen to the leaders of those disciplines from which actuarial practices derived their theoretical justification, owed him a great debt of gratitude for the paper. He had long had a reputation of being a brilliant lecturer, and that reputation had been shown to be amply justified. One definition of hell would be the place to which was taken the soul of an expired mathematician who had given his name to a school of thought and particularly to a concept of probability. Think what torments he would suffer! If there were an atmosphere in that place, it would ring with statements attributed to him but which he would very rarely recognize as his own. His name would be invoked in disputes from which he might have wished to dissociate himself. Blows might even be struck. Compared with that, it seemed to the President that the torments which Mr Perks now suffered from being orthodox were very light! Certainly the author had shown it all to be rather pointless by his masterly historical review. To quote his own words, by the 'extraction of general principles' he had eroded the area of disagreement. All, it seemed, could now shelter under the likelihood umbrella.

The author, as befitted a man who has made a notable contribution to the development of quality control and who had been President of the Operational Research Society, was not obsessed with abstraction but had reminded the meeting of the practical job of statistics—data analysis—in which the controversy about inference was overshadowed by other problems. It was interesting, too, that just as the computer had severely modified actuarial views on graduation and deterministic approaches to mortality, so on a higher plane it had removed the need for economy in exploring the likelihood function. It was no longer necessary to employ simple models which were capable of description by a limited number of parameters. A very large and complex mathematical model could be constructed and left to describe itself by simulation exercises.

He would have liked § 41 to have been developed further, but there was a limit to the material that a paper could cover and something had to be left to the imagination. He knew that all would join in expressing sincere thanks for a thoroughly stimulating evening.

The author, in acknowledging the vote of thanks, said that his relationship with Mr Perks could be illustrated, perhaps not to his own credit, by recalling that Mr Perks had written to him 18 years earlier about invariant priors and had had no reply. He admitted that his record in maintaining his correspondence was not normally good, but in that instance the failure to reply had been due to the fact that he was uncertain how to reply. He had been thinking about the subject since receiving the letter, and had not yet finished thinking about it.

He was satisfied that additional principles were involved in the formation of those notions, but preferred not to speak of them as probabilities, but as some other kind of notion associated with credibility. When quoting the value of a parameter to one place of decimals, the suggestion was that there was a natural measure for it. That was what he had in mind when referring to structures.

He was grateful to the opener for picking up what he had regarded as his main point for emphasis—the need to form a habit of providing the likelihood function as often as possible.

While all had suffered from large masses of computer output which tended to come from those machines, he had had in mind that the computer should enable them, by film or visual display equipment, to see likelihood functions in a reasonably compact form. Although theoretically a graph in a certain sense conveyed an infinite amount of information, in fact it was more easily understood than a short column of figures. He had had that kind of graphical output in mind in advocating the use of likelihood functions in practical work.