Brit. J. Phil. Sci. **23** (1972), 123–190 *Printed in Great Britain*

Review Articles

THE LOGIC OF STATISTICAL INFERENCE[1]

To review a book seven years after its publication is unusual. The distribution of elapsed times between publication and review is probably multimodal, with a peak at a relatively short time, and subsidiary peaks at times corresponding to jubilees, centenaries, and so forth. It is a measure of the importance of Hacking's work that, in spite of the fact that the foundations of statistical inference have for ten years past been an area of very active controversy, a discussion restricted to his major theses still seems appropriate and up-to-date.

Re-reading the book one is again impressed with its easy-flowing style, full of felicitous phrases—such as 'cheerful concordat' to describe the current state of divided opinion on the foundations of set theory—but with careful attention to logical niceties. It will have been read by all who have been concerned with the foundations of statistical inference, and it is to be hoped that it will continue to be read by more and more, especially by mathematical statisticians who are all too prone to hare off into abstract mathematics without taking proper care to ensure that their mathematical model is relevant to the scientific or practical situation.

The simplest mathematical models for inferential processes are those which were first explicitly set forth by Neyman and Pearson. The elements are (i) a sample space S of possible results of the experiment in question; for instance, if we are tossing a penny ten times, S consists of the 2^{10} sequences like *HHTHTTTHTH* which could represent the results of the tosses, in the order in which they occurred; (ii) a parameter space Ω of possible values for an unknown parameter θ; for instance, Ω might consist of the points in the open unit interval $\{\theta : 0 < \theta < 1\}$, or the closed unit interval $\{\theta : 0 \leqslant \theta \leqslant 1\}$; (iii) a function $p(x,\theta)$ of two variables, x ranging over S and θ ranging over Ω, specifying the probability of getting the result x when the true value of the parameter is θ; for instance we may have

$$p(x,\theta) = \theta^{r(x)}(1-\theta)^{s(x)}$$

where $r(x)$ is the number of H's and $s(x)$ is the number of T's in x. Neyman and Pearson then characterised three types of inferential problem, (i) hypothesis testing, to 'accept' or 'reject', on the basis of x, the proposition that θ belongs to a specified subset Ω_0 of Ω; for instance, in the above example, Ω_0 might consist of the single point $\theta = \frac{1}{2}$, corresponding to the penny being unbiased; (ii) point estimation, to find a function $t(x)$, taking values in Ω which would in some sense or other be 'near' to θ; for instance we might take $r(x)/10$ as our estimate of θ in the example referred to, or perhaps $(r(x)+1)/12$; (*iii*) interval estimation, to find two functions $L(x)$ and $U(x)$, taking values in Ω, such that the probability is at least, say, 0·95 that θ lies between $L(x)$ and $U(x)$, and certain other conditions

[1] Review of Hacking, I. [1965]: *The Logic of Statistical Inference.* Cambridge: Cambridge University Press. £2.50. Pp. ix+232.

are satisfied, such as that the probability, for any point in Ω, that it lies between $L(x)$ and $U(x)$ is maximised when this point coincides with θ.

Neyman and Pearson's models were extended by Wald with the addition of two more elements, (iv) a decision space D of possible actions to be taken, and (v) a loss function $L(d,\theta)$ measuring the loss incurred by taking action d when the parameter value is in fact θ. Aside from welcoming Wald's extended model, neither Neyman nor Pearson ever insisted that these models exhausted all the possible inferential situations that could arise; indeed, Pearson especially went out of his way on more than one occasion to stress that statistical inference could be applied to a great variety of types of situation, for which any particular mathematical model might well have extremely limited validity. In spite of this, the scope offered by the Neyman-Pearson-Wald models for purely mathematical analysis was so rich that they dominated mathematical statistics for nearly a quarter of a century; and the dominance went so far that other types of model, such as the Bayesian model (which adds, as a fourth element to the three of Neyman and Pearson a probability distribution (the 'prior') on the parameter space Ω), or the model underlying Fisher's fiducial argument, or the compound decision model of Robbins, were rejected out of hand or, at best, ignored by many mathematical statisticians. Thus, for example, although it was pointed out in 1943 by Molina, and again in 1946 by the present writer, that the Bayesian model may often be more appropriate to problems of industrial sampling inspection than is the Neyman-Pearson hypothesis testing model, most writers on the subject adopted the latter model until about ten years ago. Or, again, although Robbins pointed out, in his [1952], the fascinating possibilities opened up by considering each Neyman-Pearson hypothesis testing problem as one of a group of such problems—so that one had a family of sample spaces, S_i, a family of parameters θ_i, and a family of probability functions p_i, with the suffix i running, say, from 1 to 50—it took more than ten years, and an article by Neyman himself to call the attention of the statistical world at large to the breakthrough (as Neyman described it) which Robbins had achieved. As to the fiducial argument there are, of course, genuine problems associated with defining its domain of validity, but to this day it is remarkable how many writers on the subject try to fit it into either the Bayesian or the Neyman-Pearson model, with which it manifestly is not consistent.

The faithfulness with which mathematicians adhere to a single model in spite of all the difficulties to which such an attitude gives rise tempts one to pun on the word and to call for more permissiveness. A philosophical discussion of fundamentals such as Hacking's book provides will, at the very least, inspire a more reflective point of view.

With the exception of the fiducial argument it is a common property of all the models for mathematical statistics referred to above, that they require no discussion of the semantics of the word 'probability'. For Neyman-Pearson-Wald, for the Bayesian model, and for the compound decision model we merely need to recognise in the situation to which the model is being applied some entity to which the Kolmogorov axioms apply; the nature of this entity may vary, indeed, from instance to instance—it may be a subjective state of mind in one case, an objective state of knowledge in another, or a real or hypothetical frequency in others. In many cases there may coexist a multiplicity of valid interpretations. The fact that statisticians who disagree widely in their attitudes to the foundations of their subject nonetheless tend to agree in their practical advice to

scientists and engineers suggests that such ambiguous cases form the majority of those which arise in practice. Such systematic ambiguity—to use Russell's phrase, though semi-systematic ambiguity would here be better—is, of course, characteristic of mathematics.

That Hacking is concerned to discuss at a deeper level, is made clear early in his argument where he says he is concerned to study a property of a *chance set-up* on which *trials* are conducted, called by some 'long run frequency', or 'chance', and often called 'probability'. A full account of possible meanings of 'probability' is *not* attempted. Hacking's 'chance' thus corresponds closely to what Jeffreys calls 'chance' as distinct from 'probability'. It is also closely related to my own notion of 'forward log-odds', or 'f-lods'.[1]

The differences between Hacking's 'chance' and my 'f-lods' arises from two causes (apart from the trivial logarithmic transformations). In the first place, I was concerned to *derive* the usual laws of addition and multiplication from more elementary notions; and in the second place I was concerned to preserve logical distinctions which Hacking does not bother with—in particular the distinction between an observable proposition and an hypothesis concerning chance distributions. Thus I borrowed Whitehead's term 'concrescence' (changing its meaning) to denote the conjunction of an observable proposition a and a statistical hypothesis α, to bring out the fact that my fields of observable propositions were closed under conjunction ($a \cdot b$ denoting the observable proposition 'a occurring in the first of a pair of trials and b occurring in the second'—so that $a \cdot b$ should be read as 'a and then b'), whereas putting together an observable proposition and a statistical hypothesis would not produce an observable proposition, but an entity of a different logical kind.[2] Another difference between my own account and Hacking's arises from my own leanings towards constructivism in relation to the foundations of mathematics, which lead me to avoid the use of negation wherever possible, and especially to avoid using the law of excluded middle. I am afraid all these rather irrelevant scruples made the main argument of my [1949] difficult to follow. Since it will be relevant to the discussion of Hacking's views on the fiducial argument, perhaps I may be permitted to outline my own development again here.

I start from the notion of a *simple proposition*, denoted by a symbol like $a, b, c, \ldots$, which is supposed to represent the *complete* description of the result of a trial. Trials are repeatable, and the result of a pair of trials, for instance, could be $a \,.\, b$ (read as 'a and then b'); such a pair of trials can be considered as again a (compound) trial. A statistical hypothesis about trials of a given kind then is taken to establish an ordering of simple propositions as to their plausibility: this ordering is supposed to satisfy certain simple axioms. It is then shown that it is possible to attach a number $l(a)$ to any simple proposition in such a way that $l(a) < l(b)$ if and only if a is less plausible than b.[3] This number obeys the multiplication rule, $l(a \,.\, b) = l(a) \,.\, l(b)$.

The field of simple propositions is not closed under disjunction—to say that

[1] Expounded in Barnard [1949].

[2] The set theoretical correlative of the conjunction $a.b$ is the cartesian product, not the intersection of sets a and b. Concrescence has no set-theoretical correlative.

[3] On the way I assumed an Archimedean axiom. I noted at the time that this axiom need not be assumed, and that to omit it would, while complicating the mathematics, enable the theory to cover some interesting possibilities. Such possibilities later received attention from Rényi, in connection with his theory of conditional probability spaces. (Cf. Rényi [1955].).

a trial resulted in '*a* or *b*' is not to describe the result completely. However, the composite proposition '*a* or *b*' is still observable, and the question arises whether the quantification of plausibility can be extended to cover disjunctions. It was shown later that this can indeed be done, and in essentially only one way. The resulting measure of plausibility then need not satisfy the addition rule, but it must (on certain 'smoothness' assumptions) satisfy a law of the form, for mutually exclusive *a* and *b*,

$$l(a \text{ or } b) = (l(a))P + (l(b))P^{1/p}$$

for some positive value of p. Choosing $p = 1$ can then be justified as being needed to keep the mathematical formulae simple—any other choice would lead to an exactly equivalent theory, looking mathematically more complicated.

The point of this development was twofold. The first aspect, which does not directly concern the discussion of Hacking's work, was to show how a quantitative theory of probability could be derived from weak assumptions about plausibility orderings and the notion of repeatable trials; the second aspect, relevant here, was to show the extent of the duality between the notion of plausibility for observable events (corresponding to Hacking's 'chance') and the notion of plausibility for statistical hypotheses—corresponding to Hacking's 'support'. In fact the development of the theory of plausibility for simple statistical hypotheses was exactly dual to that for simple observable propositions; the parallelism fails just at the transition from simple hypotheses to composite hypotheses. A simple hypothesis is a complete specification of the relative plausibilities of observable results of trials of a given kind; a composite hypothesis is just a disjunction of simple hypotheses. But the difference of logical status between a simple statistical hypothesis and a composite one seems much greater than that between a simple observable proposition and a disjunction of these. For we can always imagine a modification of the given experimental set-up which makes a disjunction of simple propositions itself a simple proposition relative to the modified set-up. In the coin tossing example discussed above we could imagine ourselves not being able to see the results of individual throws, but only the final score recorded on a counter which adds 1 for each *H*. The new sample space would then consist of the set {0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10} and the result $r = 1$ would correspond to the disjunction of the 10 alternatives for the original kind of trial such as *HTTTTTTTT*, *THTTTTTTT*, ... *etc.* With the modified set-up, '$r = 1$' would be a simple proposition.

We cannot indicate a corresponding general method for modifying the set-up to convert a composite hypothesis into a simple one. Such modifications are sometimes possible. For instance, the composite hypothesis which consists in saying that a pair (x, y) of measurements is normally distributed with correlation coefficient 0·5 (without specification of the means and variances) can be regarded as a simple hypothesis about the distribution of the estimated correlation coefficient r, calculated by the usual formula from a set of observations, together with a simple hypothesis about the sample configuration. This reduction is possible because the composite hypothesis here considered is invariant under changes of location and scale for x and changes of location and scale for y; we can imagine a change in the experimental set-up under which we become ignorant of the units and origin of measurement of x and of y—for example we could be given a set of pairs such as (2,−3), (1, 4), (9, 11), ... *etc.* in which the first figure is known to refer to the height of an individual, and the second to his

weight. The fact that we may not be told from what origin the height measurements are taken, nor whether they are in inches or in centimetres, does not in any way prevent us from assessing the plausibility of their being normally distributed, nor the plausibility of their correlation having any particular value. But in the absence of group invariance, or of some other special structural property having like effect, the reduction of a composite hypothesis to a simple one, by changing the experimental set-up, is not in general possible.

It follows that the quantitative measure of plausibility for simple hypotheses, unlike that for simple propositions, cannot be extended to cover disjunctions, so that while the multiplication rule operates, the addition rule does not. To quote Fisher: 'Whereas such a phrase as "the probability of A or B" has a simple meaning, where A and B are mutually exclusive possibilities, the phrase "the likelihood of A or B" is more parallel with "the income of Peter or Paul"—you cannot know what it is until you know which is meant'. In fact, the dual measures of plausibility, for observable propositions and for simple statistical hypotheses, correspond closely to the Fisherian concepts of probability and likelihood, respectively. But whereas for Fisher the likelihood is, by definition, proportional to the probability, in the theory as developed in my [1949] it was merely assumed that the ratio of relative plausibilities of two simple hypotheses on two different simple propositions was some function of the relative plausibilities of the two propositions on the two hypotheses. The form of the functional relationship was then deduced, and it turned out that the measure of plausibility of a hypothesis, on observation a, would have to equal the plausibility of a, on hypothesis α, multiplied by some function of α. This latter function could be regarded as the 'prior' likelihood of α, to use Edwards's term, though this possibility was not noted at the time. It could also refer to the 'acceptability' of α.

To return now to Hacking, he takes the laws of addition and multiplication for chances as given, and a specification of the numerical chances of possible outcomes of a trial of a given kind is a statistical hypothesis. He then introduces the idea of *support* for a proposition by data, and says (p. 28): 'If one hypothesis is better supported than another, it would usually be, I believe, right to call it the more reasonable. But of course it need not be reasonable positively to believe the best supported hypothesis, nor the most reasonable one.' He goes on to make clear that support is a concept independent of utility. This is, of course, unexceptionable; but the first sentence of the two quoted is open to objection. In a kind of trial in which we expect simple laws to operate we may have a hypothesis involving some weird mathematical function better supported than, say, a simple linear law of relationship. In such a case we might well say that the simple linear law was more reasonable, although the weird law was better supported. Such a discrepancy between support and reasonableness could be accounted for on the theory of my [1949] by the factor referred to in connection with 'prior likelihood', or 'acceptability'; there does not appear to be any corresponding possibility in Hacking's theory. To this extent Hacking's theory appears to be more strongly empirical, in that hypotheses are to be judged purely on the basis of observable data, not on any structural properties they may possess.

At this stage in his argument Hacking inserts a discussion of Wald's decision-theoretic approach to inferential problems. This section is beautifully argued and expressed and should be read with care by any who may still believe that decision theory provides all the answers to problems of inference, as also should be the following chapter, in which 'long run' justifications for rules of inference are

demolished. This then leads to the exposition of the 'Law of Likelihood'. Here it is important to note that Hacking's use of likelihood differs from Fisher's, and, to that extent, it fails to correspond to the 'b-lods' notion expounded in my [1949] quite apart from the possibility, already noted, of an 'acceptability factor'. For Hacking, likelihood is a predicate of an ordered pair of propositions, namely a statistical hypothesis and an outcome of a trial; the likelihood is the chance of the outcome if the hypothesis is true. He correctly says (p. 57) that 'Fisher generally uses likelihood as a predicate of hypotheses, in the light of data', but goes on to suggest that Fisher sometimes uses the term in the way Hacking does; I cannot find that this is so.

It should be added that in saying that it is important to note this point, I do not wish to imply that Hacking's argument is in any way weak—apart from the tacit commitment to empiricism already mentioned. The importance derives from the fact that statisticians who are accustomed to Fisher's usage may find it difficult to follow Hacking's reasoning.

Hacking's law of likelihood for discrete distributions involves the notion of a simple joint proposition, which is a proposition of the form 'the distribution of chances on trials of kind K on set-up X is D; outcome E occurs on trial T of kind K'. The likelihood of such a simple joint proposition is what the probability of E would be on trials of kind K if the distribution were D. The law of likelihood also involves the idea of a (composite) joint proposition, which differs from a simple joint proposition in that a class of distributions may be involved instead of only one, and that the kind of trial K' on which E occurs need not be that to which the distributions refer. The statement of the law of likelihood for discrete distributions then is: if h and i are simple joint propositions and e is a joint proposition, and e includes both h and i, then e supports h better than i if the likelihood of h exceeds that of i.

When h and i have the same distributional part, this law says that the occurrence of event E is better supported than the occurrence of F if the probability of E is greater than that of F; when h and i have the same observational part, the law says that the hypothesis H is better supported than the hypothesis H' if the likelihood (in Fisher's sense) of H is greater than the likelihood of H'. Apart from possible considerations of 'prior' 'likelihood', or 'acceptability', these principles seem unexceptionable.

Although Hacking shows that he has read the works of Fisher and of other statisticians with more care than have most statisticians, he fails to do justice to Fisher's conception of likelihood in suggesting that Fisher saw it only as a quantity to be maximised in connection with statistical estimation. The following quotation, from Fisher's introduction to the first edition of his [1925] clarifies the situation: 'This is not to say that we cannot draw, from knowledge of a sample, inferences respecting the population from which the sample was drawn, but that the mathematical concept of probability is inadequate to express our mental confidence or diffidence in making such inference, and that the mathematical quantity which appears to be appropriate for measuring our order of preference among different possible populations does not in fact obey the laws of probability. To distinguish it from probability, I have used the term "Likelihood" to designate this quantity, since both the words "likelihood" and "probability" are loosely used in common speech to cover both kinds of relationship.'

The next chapter of Hacking's book is concerned with 'tests of significance'. Here there seems to be a major flaw. He quotes, approvingly, W. S. Gossett as

saying that a test 'doesn't in itself necessarily prove that the sample is not drawn randomly from the population even if the chance is very small, say ·00001: what it does is to show that if there is any alternative hypothesis which will explain the occurrence of the sample with a more reasonable probability, say ·05 (such as that it belongs to a different population or that the sample wasn't random or whatever will do the trick) you will be very much more inclined to consider that the original hypothesis is not true'. He takes this passage by way of rebuttal of Fisher, whom he accuses of personal criticism of those who stressed the importance of alternatives to the hypothesis being tested, in connection with the theory of testing. In fact Fisher referred approvingly to the concept of the power curve of a test procedure and although he wrote: 'On the whole the ideas (*a*) that a test of significance must be regarded as one of a series of similar tests applied to a succession of similar bodies of data, and (*b*) that the purpose of the test is to discriminate or "decide" between two or more hypotheses, have greatly obscured their understanding', he was careful to go on and add 'when taken not as contingent possibilities but as elements essential to their logic'. It would take too long to give a full discussion of the logic of significance tests here; a reference to Anscombe's paper (Anscombe [1962]) must suffice. The difficulty with Hacking's account is that it leads him to say (p. 89): 'An hypothesis should be rejected if and only if there is some rival hypothesis much better supported than it is.' But there *always* is such a rival hypothesis, *viz.* that things just had to turn out the way they actually did.

The fact seems to be, that the domain of application of Hacking's theory, like that of the theory of likelihood, is confined to comparisons between statistical hypotheses. Situations in which only one hypothesis is in question, and the issue is whether agreement between this hypothesis and the data is so bad that we must begin to exercise our imaginations to try to think of an alternative that fits the facts better—tests of goodness of fit, in other words—are not satisfactorily dealt with by this theory.

Hacking's next chapter discusses the Neyman-Pearson approach to hypothesis testing and he reaches the conclusion, with which I agree, that this approach is best regarded as a theory applicable to the advance planning of experiments. On a purely personal point, I feel bound to comment on what is said on page 103: 'The idea that Neyman-Pearson tests could better serve before-trial betting than after-trial evaluation has lain fallow in a remark made by Jeffreys twenty years ago. Last year [*i.e.* in 1963] Dempster produced, for the first time in all that period, a similar distinction'. Hacking here seems to have overlooked my [1950] from which I quote: 'it is the view of the present writer that the arbitrary nature of the reference set involved, on the Neyman-Pearson theory, in a test of significance, is a decisive reason for rejecting that theory, as a theory of inference, in favour of using a theory of inference, such as that given by Fisher, where the idea of a reference set does not enter. It should be emphasised, however, that the Neyman-Pearson theory is an exceedingly valuable weapon in the advance planning of experimentation. To put the matter shortly, two kinds of quantity are involved in uncertain inference. The present writer has called them f-odds and b-odds, but Fisher has more aptly (though slightly less precisely) called them probability and likelihood. Probabilities are relevant before an experiment has been performed, when we are planning it. After the experiment has been performed, when we are drawing conclusions, likelihoods are relevant. As a theory

based on probabilities, the Neyman-Pearson theory is useful in planning, before the result is known; but after the result is known the theory of likelihood should be used.' Although I was conscious, when I wrote about likelihood in my [1947] that its use would be restricted to the comparison of one hypothesis with another, and was at that time and again in my [1949] careful to say so, in the passage quoted above I incautiously omitted to refer to this.[1]

Hacking's next chapter discusses randomness, in connection with random sequences and with random sampling. He refers to Church's suggestion to overcome the difficulties in von Mises's idea of a random sequence by using the fact that the set of computable functions is countable; but he does not appear to have noticed Ville's destructive criticism of this idea. Ville's point essentially was that any attempt to define mathematically the class of random sequences of 0's and 1's was doomed to failure, because, representing such sequences as binary fractions, one would be excluding as non-random any set of such fractions of Lebesgue measure zero. But however many such sets had been removed from the unit interval, the remaining set would, of course, have measure 1, and so would contain infinitely many more sets also of measure zero, and these sets ought also to be excluded. This argument persuaded many, including myself, that the search for a mathematical definition of randomness was doomed to failure, until Martin-Löf came along with an application of the idea of computability which went in a somewhat different direction from Church. Briefly, Martin-Löf's definition of a (finite) random sequence is, that it is a sequence which is computed by a Turing machine whose Gödel number (in the binary scale) is as long as the sequence itself. He overcomes Ville's difficulty by pointing to the fact that the set of *computable* sets of measure zero is countable, so that the union of all such sets also has measure zero. It is this union of all such sets which is removed from the unit interval.

Apart from brief references to random sequences, Hacking's main concern in this chapter is to discuss random sampling from finite populations in a way which does not seem to differ markedly from standard statistical doctrine. It is the next chapter, entitled 'The fiducial argument' which has perhaps drawn forth the most comment from reviewers, in that Hacking sets out a mode of reasoning leading effectively to probability statements about hypotheses (though Hacking sticks to the term 'support', rather than probability).

He draws a fruitful analogy with Frege's foundations for set theory and arithmetic, in which Russell found a contradiction. He says (p. 151) 'Just as (after Russell's criticism of Frege) we needed a more adequate characterisation of irrelevance. Each genius (Fisher and Frege) proferred a false conjecture. The task of future generations is not to confute the genius but to perfect the conjecture. Our principle of irrelevance is some way along that road.' Indeed it is, as also is the work of Sprott, and of Fraser, to which Hacking does not refer. But in his extended list of contents, Hacking says 'It is proved that any principle similar to the principle of irrelevance but stronger than it, must lead to contradiction. So the principle completes a theory of support.' In the chapter in question, however, Hacking makes this idea plausible, but could not be said to have produced a proof. My own conjecture here is, that the situation will continue to be analogous to that holding in the foundations of set theory, where, in a sense, the search for a definitive formulation has definitively been shown to

[1] Also, cf. Barnard [1951] and Barnard, Jenkins and Winsten [1962].

be fruitless. One is, in fact, tempted also to conjecture that, just as Hadamard said that his brain was so constructed that the axiom of choice appeared obvious to him, while to Borel it was far from obvious, and Cohen has shown that adding either the axiom of choice or its negation to the other axioms of set theory will not produce a contradiction, so we may find, when the rules of statistical inference are more fully formalised than now, that options remain.[1] Such a conjecture looks more plausible when we bear in mind that the appeal, for example in Hacking's theory, to the notion of a trial *of kind K* involves an act of class formation.

Hacking's next two chapters are concerned with estimation. He points to difficulties in the standard theories, but does not always go as far as he might have done. For example, on page 183 he says: 'Of course it might be true that the best of unbiased estimators were better than estimators of other kinds, even though inferior unbiased estimators are inferior to other estimators. But this has never been proved.' Indeed, there are counter examples. In inverse sampling to estimate a single probability θ it can be shown that the *only* unbiased estimator is restricted to taking the values 0 or 1; but if the sample space is the open unit interval, or, for instance, restricts θ to lie between $\frac{1}{3}$ and $\frac{2}{3}$, this means that the only values the unbiased estimate can take are known to be impossible ones. Hacking makes an interesting suggestion, that in asking for an estimate we should always specify a scale of measurement, and this leads him to reject the idea that in estimating θ we are necessarily at the same time estimating any function $f(\theta)$, and that our estimate of $f(\theta)$ must be $f(t)$ if our estimate of θ is t. This amounts to saying, in the Neyman-Pearson model, that the parameter space should be endowed with a metric, rather than being left as an unstructured set. Unfortunately Hacking does not develop this idea to any extent. My own guess would be that we should think of the rather weaker notion of a topology on the parameter space (a metric implies a topology, but not conversely), so that we can give a meaning to the idea of converging on the true value of θ when the sample size becomes larger and larger; in some cases a family of locally equivalent metrics—intermediate between a single metric and a topology—seems appropriate. Le Cam has used ideas of this kind, in connection with what Neyman termed 'Best Asymptotically Normal' estimates; his theory can be regarded as a further development and formalisation of the earlier part of Fisher's [1925*a*].

It is a pity that Hacking does not discuss the idea, eminently practicable now that we have computers with graphical output facilities, of regarding the graph of the likelihood function (or, equivalently, the graph of its logarithmic derivative) as answering the problem of estimation. It will indicate the value or values to which the data point most strongly (*i.e.* which are best supported, or most likely), it will indicate with what precision the data point to the neighbourhood of a particular value, and it will preserve all the relevant information in the data, in the sense that it is a minimal sufficient statistic.

The final two chapters are devoted to discussion of Bayes's theories, first in their 'objective' versions as represented by Bayes and by Jeffreys, then in the subjective versions associated with de Finetti, Ramsey and Savage. The objective versions are rejected 'for reasons which by now are entirely standard', while the subjective theory is explained 'sympathetically'. In so far as Hacking's theory is

[1] The possibility that *valid* theories of statistical inference could be genetically determined is worth meditating upon.

restricted to cases in which the notion of 'chance' is involved, its domain is narrower than that of the subjective Bayesians; at the same time it is more explicit in its application to problems of inference in the natural sciences.

The last chapter shows signs of having been written some time after the earlier ones, and it seems to shift the emphasis in places. For example, on page 222 Hacking comes near to discussing a 'goodness of fit' situation, and says 'My theory of statistical support does not attempt rigorous analysis of the reasoning here. . . . The theory of statistical support cannot judge the force with which an experiment counts against a simplifying assumption.' To this extent he appears to agree with the comment made above on his treatment of tests of significance. Again, on page 219, Hacking appears to entertain the possibility of something corresponding to the idea of 'prior likelihood' of 'acceptability' referred to above, and he explicitly refers to the point made by Fraser (and also by the present writer) that group invariance and other structural features of an experimental set-up may be relevant to its statistical interpretation.

It is clear that in this area there is much further exploration to be done. Hacking's book remains an invaluable guide book for anyone willing to join in this task.

G. A. BARNARD
University of Essex

REFERENCES

ANSCOMBE, F. J. [1962]: 'Tests of Goodness of Fit', *Journal of the Royal Statistical Society* (*B*), **25**, pp. 81–94.

BARNARD, G. A. [1947]: Review of Abraham Wald: 'Sequential Analysis', *Journal of the American Statistical Association*, **42**, pp. 658–64.

BARNARD, G. A. [1949]: 'Statistical Inference', *Journal of the Royal Statistical Society* (*B*), **11**, pp. 115–49.

BARNARD, G. A. [1950]: 'On the Fisher–Behrens Test', *Biometrika*, **37**, pp. 203–7.

BARNARD, G. A. [1951]: 'The Theory of Information', *Journal of the Royal Statistical Society* (*B*), **13**, pp. 46–64.

BARNARD, G. A., JENKINS, G. M. and WINSTEN, C. B. [1962]: 'Likelihood Inference and Time Series', *Journal of the Royal Statistical Society* (*A*), **125**, pp. 321–72.

FISHER, R. A. [1925*a*]: 'Theory of Statistical Estimation', *Proceedings of the Cambridge Philosophical Society*, **22**, pp. 700–25.

FISHER, R. A. [1925*b*]: *Statistical Methods for Research Workers.*

RÉNYI, A. [1955]: 'On a New Axiomatic Theory of Probability', *Acta Mathematica Academiae Scientiarum Hungaricae*, **6**, pp. 285–335.

ROBBINS, H. [1952]: 'Asymptotically Sub-Minimax Solutions of the Compound Decision Problem' in J. Neyman (*ed.*): *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pp. 131–48.

LIKELIHOOD

The fundamental question about statistical inference is philosophical: what primitive concepts are to be used? Only two answers are popular today. Edwards is the first scientist to write a systematic monograph advocating a third answer.[1]

[1] Edwards, A. F. [1972]: *Likelihood. An Account of the Statistical Concept of Likelihood and its Application to Scientific Inference.* Cambridge: Cambridge University Press. £3.80. Pp. xiii+235.