Austral. J. Statist., **20** (1), 1978, 43–59

STATISTICAL SOCIETY OF AUSTRALIA

THE KNIBBS LECTURE FOR 1977

(SEPTEMBER 5th, 1977)

FOUNDATIONS OF STATISTICAL INFERENCE: THE CASE FOR ECLECTICISM

D. R. COX

Department of Mathematics, Imperial College

Summary

First there is some general discussion of the place in statistics of work on the foundations of statistical inference. Three broad approaches, sampling theory, pure likelihood and Bayesian are distinguished and some general comments made on the advantages and limitations of each approach. It is concluded that all three have some role.

1. Introduction

I greatly appreciate the invitation to give this lecture, named in honour of a distinguished Australian.

It is, of course, impossible in one short paper to give a comprehensive review of the foundations of statistical inference, let alone to put forward new points. The object of the paper is partly to outline what seem to me a few key ideas and more specifically to put the case for eclecticism, the notion that all the main approaches have some merit, each in its appropriate circumstances. While I hope that many will be broadly sympathetic to this point of view, there are of course great advantages, aesthetic and severely practical, in a unified approach dealing with all inferential statistical problems; ultimately personal judgement and practical experience are involved in deciding whether the distortion and idealization inevitable in a superficially all-embracing theory are worthwhile. The central implied theme of this paper, like that of Cox and Hinkley (1974), is that it is fruitful to contemplate problems formulated to different depths of detail and to use different approaches accordingly. In many ways the most primitive formulation is that for a pure significance test, where only the null hypothesis under test need be explicitly formulated, and the richest formulation is that for Bayesian decision analysis where a full probability model, space of possible decisions, prior distribution and utility function are all assumed available.

The discussion deals throughout only with the more formal aspects of statistical inference and these are only a part, often a small part, of statistical work. The very important matter of the place in practical statistical work of formal theories of inference will not be discussed explicitly.

2. Some Preliminary Comments

Before discussing the central issues it is worthwhile commenting briefly on some preliminary questions.

(a) Role of Foundations

The aims of a study of foundations include

(i) qualitative clarification of the objectives of statistical work;

(ii) formal justification of, or even improvements to, procedures of analysis that seem at least sensible on general grounds;

(iii) the provision of a systematic basis for tackling new problems.

Of these (iii) is of the most obvious practical importance, (ii) is probably of most interest to teachers and beginning students and (i) is possibly of most interest to the non-specialist. We must to some extent judge the success of an approach by the variety and difficulty of the problems on which it is successful.

(b) Importance of disagreements

Differences between different people's analysis of the same data seem often to lie more in formulation of a model and of questions for study than in technical differences between different approaches to detailed statistical methods, although there is rather little specific evidence on this point. It does seem likely that the greatest impact of the differences between different approaches is on qualitative attitudes to what is being attempted, point (i) of (a) above. The disagreements that exist are, superficially at least, major but, I think, no more than those about the foundations of mathematics and the foundations of theoretical physics.

(c) Role of Decision Theory

The remainder of the paper deals hardly at all with decision analysis. The notion quite widely advanced 20–25 years ago that all statistical problems should be regarded as decision ones is now not often heard. Nevertheless the qualitative ideas of decision analysis do seem very valuable. The reasons that the detailed techniques seem of fairly limited applicability, even when a fairly clearcut decision element is involved, may be

(i) that, except in such fields as control theory and acceptance sampling, a major contribution of statistical technique is in presenting the evidence in incisive form for discussion, rather than in providing a mechanical prescription for the final decision. This is especially the case when a single major decision is involved;

(ii) the central difficulty may be in formulating the elements required for the quantitative analysis, rather than in combining these elements via a decision rule.

In the remainder of the paper we concentrate on the statistical analysis of observations in situations in which the uncertainty in the conclusions needs to be indicated. The explicit recognition that conclusions are uncertain is a characteristic feature of statistical inference in the narrow sense, and is the aspect most studied in formal theories. The recognition in a decision-making context that the evidence is uncertain does not in principle preclude vigorous decision-making, yet clearly at an individual level awkwardness may well be involved, whether for the official statistician, the technologist or the scientist. There are difficult issues here bearing on such questions as the role of enthusiasm and commitment, in all fields of applied statistical work, and more broadly still.

(d) Role of Probability and Parameters

Most technical discussion of statistical methods starts from a probability model and uses probability directly or indirectly in assessing uncertainty. Parameters, or sometimes future observations from the same random system, are regarded as the objective. Important and fruitful though this abstraction is, there is much statistical work in which intelligent description of the data can be considered entirely or largely independent of a probability model. In recent years there has been a welcome increased explicit recognition of the importance of descriptive statistics, i.e. statistical analysis in which the probabilistic element is absent or relatively minor. It seems very important that accounts of statistical technique should synthesize the descriptive and the probabilistic aspects and indeed the most valuable methods of analysis are frequently those with a dual role.

(e) Importance of Formulation

The critical importance of checking and improving formulation has already been mentioned. This applies particularly to primary aspects of a model, i.e. those defining the aspects of direct scientific or technological concern. All the following discussion concerns analysis in the light of a formulation assumed provisionally to be correct.

3. The Main Approaches

An initial classification of approaches to statistical inference is into sampling theory, in which some notion of judging procedures via their performance in repeated sampling is involved; pure likelihood, in which a likelihood function is taken as such to summarize what can be learnt; Bayesian, in which the parameter of interest is regarded, in the absence of data, as a random variable with a known distribution.

We comment on these in turn and then make some comparative remarks. Of course all three approaches come in various forms!

4. Sampling Theory Approach

The key notion is of judging what would happen in repeated applications and is of course central to significance tests, confidence intervals and so on. Both fiducial probability and Fraser's structural inference use the sampling idea, although less directly. The powerful appeal of sampling theory lies in its direct operational character.

Difficulties with the notion of behaviour in repeated sampling are as follows.

(i) The repetitions contemplated are nearly always hypothetical. This by itself seems no drawback; most of the operational definitions of science are based on experiments that are at least somewhat hypothetical. The hypothetical character is serious only when there is substantial doubt as to the hypothetical set of repetitions which should be considered. One aspect of this point is amplified in the next comment.

(ii) It is crucial that the probability calculations referring to a hypothetical series of repetitions are relevant to the particular situation under analysis, as was repeatedly emphasized by R. A. Fisher. This is achieved by matching the hypothetical repetitions to the particular situation and for this some form of conditioning, i.e. use of ancillary statistics, is needed. While in applications the point is often rather academic, it seems clear that in principle a satisfactory sampling theory must include appropriate conditioning. A satisfactory mathematical formulation does not seem yet to have been achieved. Fraser (1967) has in effect given an elegant discussion of some quite general linear models, but his interpretation is different. Conditioning enters in an important way into a careful discussion of the role of randomization in experimental design, and in particular of such aspects as the rejection of arrangements thrown up by the randomization and then judged unsatisfactory.

(iii) Some of the criteria used for choosing procedures are rather *ad hoc.*

(iv) Relatively very few problems have useful and elegant "exact" solutions and one rather often has to use the approximations of asymptotic theory.

(v) Combination of conclusions from the analysis of separate experiments is sometimes messy.

(vi) Most sampling-theory methods conflict with the so-called strong likelihood principle. This says that if Y and Z are two random variables with different sample spaces but with densities $f_Y(y;\theta)$ and $f_Z(z;\theta)$ depending on the same θ, and if y' and z' are such that for all θ

$$f_Y(y';\theta) \propto f_Z(z';\theta),$$

then the conclusions to be drawn from y' and from z' should be identical. The conflict is by itself of no importance, unless one regards the strong likelihood principle as justified by some further arguments, such as of Bayesian theory or those investigated by Birnbaum (1962); see also Birnbaum (1969) and Barnard, Jenkins and Winsten (1964). There has been a good deal of discussion of the principle; see, especially Durbin (1970), Basu (1975) and Kalbfleisch (1975). For further details, see the Appendix.

(viii) Except when the procedures correspond to a Bayesian analysis with a proper prior the procedures are incoherent in the technical Bayesian sense; for further discussion, see Section 6. There is the related matter of the existence in interval estimation of recognizable subsets (Pierce, 1973; Robinson,1975).

While these points, especially the existence of recognizable subsets, are worrying, and some radical reformulation may eventually be found that is more satisfactory, the great attractions of the sampling theory approach are the direct appeal of criteria based on physical behaviour, even if under idealized conditions, and the ability of the approach to give some answer to a very wide range of problems. Indeed the notion of calibrating statistical procedures via their performance under various circumstances seems of such direct appeal that it is difficult to see how some idea of the sort can be avoided; how would one feel about a method of analysis that under some plausible set of repetitions gave a qualitatively misleading answer most of the time?

5. Pure Likelihood Approach

The likelihood function, when it can be calculated, is central to all approaches to statistical inference. The pure likelihood approach is based on the notion that the likelihood function itself can be used for interpretation. Its advantages are simplicity and directness, that the strong likelihood principle is obeyed and that no contradictions with the Bayesian approach can arise. For very simple situations involving the comparison of two values of a parameter fixed in advance it does indeed seem best to consider just the likelihood ratio; a physical interpretation of the numerical value of the ratio is easily given, for instance via a hypothetical use of Bayes's theorem. Further for small numbers of parameters the notion of plotting likelihood curves or surfaces is qualitatively useful, in particular to clarify situations in which use of a single more or less symmetrical confidence interval would be misleading or where reparameterization is desirable.

Nevertheless, there seem strong arguments against the direct use of pure likelihood as a primary approach:

(i) in some problems it is not possible to find a likelihood function in useful form and in particular the pure test of significance is not covered;
(ii) it is easy to produce problems where a maximum of the likelihood is liable to arise at quite the wrong place;
(iii) multiparameter problems cause serious difficulties, both of finding convenient ways of presenting numerically or graphically functions of many variables, and also of interpretation, whenever a number of nuisance parameters are present.

To some extent these last difficulties can be by-passed by the consideration of modified likelihood functions, such as marginal, conditional and more generally partial likelihoods. Further study of such modified likelihoods is desirable on more general grounds; it would be good too, to study robust likelihoods and general techniques for deriving approximate likelihood functions in situations, such as some point process problems, where direct likelihood functions cannot be found in useful form.

Edwards (1972), in his account of pure likelihood methods, has introduced a notion of prior likelihood for introducing further information in a weaker form than a probability distribution. Plausibility theory (Barndorff-Nielsen, 1976) can be regarded as a parallel to pure likelihood theory in which, however, the likelihood for each parameter value θ is normalized by the largest likelihood that could have been achieved for that θ.

6. Bayesian Approach

We now consider approaches in which prior distributions are used for unknown parameters. The introduction of a prior distribution has two advantages. First, it is a way of injecting further information, which if relevant and sound, will usually sharpen the conclusions. Secondly, it allows a final statement of conclusions via a direct notion of probability. A more unified theory, freer of *ad hoc* notions, results.

It is reasonable to distinguish three broad types of prior distribution involving

(i) a prior frequency distribution;
(ii) some idea of "logical" probability or propensity;
(iii) personalistic probability.

The first is uncontroversial, probably still understudied, but relatively rarely applicable. Usually the prior distribution will have to be estimated, i.e. an empirical Bayes procedure is involved, and this may be constructed either by a fully Bayesian argument, in which case some form of (ii) and (iii) is still required, or via an appeal to sampling theory concepts. Cox (1975) gave a fairly general solution for approximate empirical Bayes confidence limits in such cases. A somewhat related way in which Bayesian and sampling theory arguments may be mixed is when a prior distribution is introduced only for the nuisance parameters of a problem. One important use is to avoid overperameterization, for instance when there are many sets of data with similar but unequal parametric values.

The best known development of statistical theory involving "logical" probability is that of Jeffreys (1961). It seems to me that while an approach along these lines is extremely appealing, present versions of the theory are not tenable. How is one to find numerically prior distributions in cases other than of initial ignorance? Without an answer to this, one of the advantages of the Bayesian approach is lost. In a state of initial ignorance, of course an idealization, the invariance rules suggested by Jeffreys for calculating prior densities are capable of leading to absurd conclusions. Note also the so-called marginalization paradoxes of Dawid, Stone, and Zidek (1973), illustrating the difficulties of improper priors.

The relatively recent resurgence of interest in the Bayesian approach centres on the third approach, the personalistic one, in which the prior probabilities measure what "you" feel about the various possibilities. de Finetti (1970, 1975), in particular, has advocated this as the single unifying notion of probability. The suggestion that a probability in say the kinetic theory of gases is best regarded as in essence about what "you" would bet on some statement concerning, say, the velocity of a molecule is intellectually interesting, but seems, to me at least, far fetched. For an excellent account of recent statistical work based on personalistic probability, see Lindley (1971). No attempt here will be made to review the theory.

Some of the central general issues seem to be the following.

(a) Role of Prior Information

There is no dispute that in many, if not all, problems prior information, i.e. information from sources other than the data under analysis, is available. Personalistic Bayesian methods insist on specification of this quantitatively and on an answer that is a synthesis of this information with that provided by the data. If it is desired to isolate the contribution of the data, this is done by examining the likelihood. Sampling theory methods leave prior information to be specified and incorporated qualitatively, unless, of course, the information comes from other "statistical" data, in which case a single composite formulation is in principle desirable. To reject personalistic probability as the primary notion in statistical inference is in no way to deny an important role to the personal judgement of the research worker.

(b) Need for a Personalistic Theory of Probability

The notion of a personalistic theory of probability is of considerable interest in giving a qualitative discussion of individual behaviour faced with uncertainty and, hopefully, in giving constructive guidelines. In some fields of major decision making, there may be little "hard" data and quantitative assessment of uncertainty may nevertheless be valuable, provided some consensus can be reached and the factual basis for it understood.

(c) Critique of the Usual Theory of Personalistic Probability

The emphasis on betting is not to everyone's taste, but this is presumably a question of presentation rather than substance. The usual presentation seems also to put excessive emphasis on self consistency rather than on explicit guidance on how to determine "your" probability, but it is not clear whether this too is not a question of presentation. The theory has the rather curious feature that while it is not overtly concerned with frequency, "your" probability that a frequency interpretation does hold, is by the weak law of large numbers, very close to one.

(d) Arguments for Coherency

It has been argued forcefully by advocates of the personalistic approach that the argument that the axioms for self-consistency imply Bayes's theorem, and hence in particular the strong likelihood principle, are so compelling that, in particular, the sampling approach should be abandoned. There are arguments, to my mind powerful, for not accepting this radical conclusion.

(i) Above all, if it is admitted that there are different kinds of uncertainty, the arguments for Bayesian coherency collapse.

(ii) The sampling theory approach is concerned with the relation between the data and the external world, however idealized the representation. The existence of an elegant theory of self consistent private behaviour seems no grounds for changing the whole focus of discussion.

(iii) The Bayesian argument assumes the consistency of prior information and that provided by the data, whereas, of course, any major inconsistency needs examination. It could be argued that a more careful formulation of the Bayesian analysis would overcome this, but an elaborate analysis in advance of all the types of inconsistency that might arise would be required and this is not feasible.

(iv) The arguments against the sampling theory approach need an assumption of temporal coherency (Vikers, 1965; Hacking, 1967), which is certainly not necessarily satisfied in applications. Briefly, the standard arguments about coherency apply at any one time. There is nothing in the theory as usually presented to stop the prior distribution after having got the data being quite different from that before the data were obtained. Temporal coherency requires that the prior does remain the same. Indeed Cox and Hinkley (1974, p. 396) show that when the prior changes with sample size in a particular way essentially the classical significance test is recovered. Now it is clear, possibly relatively rarely but certainly in some situations, that it may be good to revise one's assessment of other information in the light of the data, which may show entirely unanticipated features. Here temporal incoherency is positively to be encouraged. Of course this is related to the larger question of amending methods of analysis in the light of the data, which raises difficult issues from all points of view.

To summarize, it seems to me that at an immediate practical quantitative level there is little role for personalistic probability as a primary technique in the analysis of data. At a more theoretical level, however, much insight can and has been gained by examining problems from this point of view.

7. Concluding Remarks

This paper has ranged over a number of aspects of the formal theory of statistical inference; it must be stressed that these considerations are but a part and often a very small part, of statistical work. The broad conclusions have been that

(i) the general notion of interpretation via hypothetical repeated sampling, despite a number of difficulties, seems of central importance, not least because it gives a way of tackling relatively poorly specified problems;

(ii) "pure" study of the likelihood as such, without any sampling theory or Bayesian ideas is normally of limited usefulness, although examination of the shape of the likelihood function is of considerable importance in non-standard situations with small numbers of parameters. There is scope for further development of modified likelihood functions;

(iii) the Bayesian formalism is often the natural one for studying repetitive decision problems, such as those of stochastic control theory;

(iv) while study of personalistic probability is valuable, the circumstances under which it should be used directly in the interpretation of data seem very limited and the arguments that Bayesian coherency should be an overriding criterion seem unacceptable.

References

Barnard, G. A., Jenkins, G. M., and Winsten, C. B. (1962). "Likelihood inference and time series (with discussion)." *J. R. Statist. Soc. A*, 125, 321–372.

Barndorff-Nielsen, O. (1976). "Plausibility inference (with discussion)." *J. R. Statist. Soc. B*, 38, 103–131.

Basu, D. (1975). "Statistical information and likelihood (with discussion)." *Sankhyā A*, 37, 1–71.

Birnbaum, A. (1962). "On the foundations of statistical inference (with discussion)." *J. Amer. Statist. Assoc.*, 57, 269–326.

Birnbaum, A. (1969). "Concepts of statistical inference." In *Philosophy Science and Method; Essays in Honor of E. Nagel*, pp. 112–143, eds. Morgenbesser, S., Suppes, P. and White, M. St. Martin's Press, New York.

Cox, D. R. (1975). "Prediction intervals and empirical Bayes confidence intervals." In *Perspectives in Probability and Statistics*, pp. 47–55, ed. Gani, J. Academic Press, London.

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*. Chapman and Hall, London.

Dawid, A. P., Stone, M., and Zidek, J. (1973). "Marginalization paradoxes in Bayesian and structural inference (with discussion)." *J. R. Statist. Soc. B*. 35, 187–233.

Durbin, J. (1970)."On Birnbaum's theorem on the relation between sufficiency conditionality and likelihood." *J. Amer. Statist. Assoc.*, 65, 395–398.

de Finetti, B. (1970, 1975). *Theory of Probability*. 2 vols. Wiley, London.

Edwards, A. W. F. (1972). *Likelihood*. Cambridge University Press, London.

Fraser, D. A. S. (1967). *The Structure of Inference*. Wiley, New York.
Hacking, I. (1967). "Slightly more realistic personal probability." *Philos. Sci.*, 34, 311–325.
Jeffreys, H. (1961). *Theory of Probability*. 3rd. ed. Clarendon Press, Oxford.
Kalbfleisch, J. D. (1975). "Sufficiency and conditionality (with discussion)." *Biometrika*, 62, 251–268.
Lindley, D. V. (1971). *Bayesian Statistics: a Review*. SIAM, Philadelphia.
Pierce, D. A. (1973). "On some difficulties in a frequency theory of inference." *Ann. Statist.*, 1, 241–250.
Robinson, G. (1975). "Some counterexamples to the theory of confidence intervals." *Biometrika*, 62, 155–161.
Vickers, J. M. (1965). "Some remarks on coherency and subjective probability." *Philos. Sci.*, 32, 32–38.

Appendix

Note on the Strong Likelihood Principle

For a brief formulation and some key references, see Section 4. Two examples illustrating the principle are as follows:

(i) Let $Y=r$ successes be observed in n trials and let $Z=n$ trials be required to achieve r successes, where in both cases there are independent binary trials with probability of success θ. Take $y'=r$, $z'=n$.

(ii) Let Y be a fixed number of independent and identically distributed random variables normally distributed with mean θ and unit variance. Let Z represent a set of such variables, and the sample size, observations being taken sequentially until the sample mean first differs from zero by, say, at least four standard errors. Take y' and z' to be two sets of data with the same mean and the same sample size, and hence with identical likelihood functions.

According to the strong likelihood principle y' and z' should lead to identical inferences about θ. From a Bayesian viewpoint with a fixed prior this is obvious, from the viewpoint of pure likelihood it is essentially a matter of definition, whereas in the sampling theory approach most methods will lead to different answers for y' and for z'. The differences are numerically fairly small in the first example, and are major in the second example, at least as regards the possible truth of the hypothesis $\theta=0$.

Arguments to suggest that principles generally accepted in a sampling theory approach require the strong likelihood principle amount to the following. Consider an "enlarged" experiment in which with probability one half Y is observed and with probability one half Z is observed, and contemplate the outcomes

(α) Y is observed and $Y=y'$,

(β) Z is observed and $Z=z'$.

Now (α) and (β) have proportional likelihood functions and are thus easily shown to have the same value of the minimal sufficient statistic in the "enlarged" experiment. Hence by a sufficiency principle (α) and (β) ahould lead to the same inference. Now consider (α). The event "Y is observed" has fixed known probability, independent of θ, is uninformative about θ and hence by a principle of conditioning (use of ancillary statistic) the inference from (α) should be conditional on having observed Y, i.e. should be the same as that from the observations y' in the original component set-up. The analogous conclusion applies to z' and the strong likelihood principle follows.

There are two steps in the argument and hence to dissent from the conclusion one must dissent from one or both of the steps. First, we may argue (Kalbfleisch, 1975) that the "enlarged" experiment is an artefact and that before applying the sufficiency principle we should restrict attention to the physically meaningful experiments. This is to object to the sufficiency part of the argument. Note indeed that the "enlarged" experiment is peculiar to the particular y' and z' and that to apply the conclusion generally in the situation of the first example we need an "enlarged" experiment for every pair (r, n).

The second possibility is to object to the conditionality argument using in passing from the "enlarged" experiment back to a component experiment. Now while no definitive formulation of a conditionality principle has been reached, there are arguments for restricting ancillary statistics to be components of the minimal sufficient statistic for the problem; one argument for so doing is the admittedly rather *ad hoc* one of avoiding difficulties of nonuniqueness (Cox and Hinkley, 1974, p. 34), although the restriction has in any case some general appeal. Then in (α) and (β) the indicator of which random variable has been observed is not part of the minimal sufficient statistic, is not an ancillary statistic and therefore it is inappropriate to condition on it (Durbin, 1970).

Both these objections have some force, although in my opinion the first, being the less technical, is the more important. In addition the following points are relevant:

(a) in an incompletely specified situation, for example a significance testing problem with no explicitly formulated alternatives, the likelihood principle is inapplicable;

(b) it might be argued that the second example above, taken in the context of examining consistency with $\theta = 0$, is enough to refute the strong likelihood principle;

(c) acceptance of the strong likelihood principle virtually excludes an approach to statistical inference based on sample space calculations, except in incompletely specified problems. Now, while the desirability of such an approach is open for discussion, that such an approach should be totally excluded seems a very suspiciously strong conclusion to reach from an innocent-seeming start.

INVITED DISCUSSION OF THE PAPER BY PROFESSOR COX

Professor P. A. P. Moran (*Australian National University*): It is an honour to be asked to discuss this paper and I thank the Society for inviting me. Professor Cox has an unrivalled experience in both the uses of statistical inference and the study of its theory. Perhaps the best thing I can do is to comment on some of his remarks from my own personal viewpoint in the hope that this will help provoke discussion.

I do not understand the concept of a pure significance test. To me any significance test involves some idea of a class of alternatives. Even for the cases of portmanteau tests of, say, the Kolmogorov–Smirnov type, the alternative is thought of as a deviation from an assumed distribution rather than, say, a deviation in the direction of serial correlation in the sample values. To me the concept of power is of the greatest importance in practical statistics, and I repeatedly come across published works based on inefficient and misleading methods because the optimal or asymptotically optimal tests are not used. In spite of what Professor Cox says, I do not think that the fact that one has often to use asymptotic theory matters very much. Until recently the phrase "for large n' concealed a skeleton in the statistical cupboard in that it was never stated just how large n had to be. But nowadays with the use of computers, even desk computers, it is usually quite easy to get good estimates of the size and power of an asymptotic test simply by simulation.

What I feel about the Bayesians is that they are attempting to quantify statistical inference in the wrong place. All statistical inference involves some subjective judgment and we very often have a prior inference for one class of conclusions over another. But to turn this preference into numerical values seems to me misleading. Of course there are cases, such as in genetic counselling, where a Bayesian analysis makes sense, but these are more related to empirical Bayesian methods, to which no objection can be taken.

The situation is similar in part in Decision Theory. This is important and valuable when the loss, or utility, can be described numerically in empirical meaningful terms e.g. as a conventional cost for a wrong decision between alternatives, or as a quadratic loss function. But decision theory can be very misleading if the nature of the loss function is forgotten, and much of the recent writing on inadmissibility seems to me to be quite misguided.

Finally a remark on likelihood. I find the discussions on the strong and weak likelihood principles very confusing. But I do not believe that likelihood could be a universal principle in statistical inference. Likelihood is essentially a Radon–Nikodym derivative with respect to a base Borel-measure on a field which is invariant under all alternatives considered. But it is easy to construct simple, if not physically natural, examples where no such base measure can exist. Thus consider the singular distribution function which arises naturally from the Cantor ternary set constructed on the interval $(0, 1)$. Call this $F(x)$, and consider a sample of n independent values from the distribution $F((x-m)/\mu)$. It is desired to estimate the scale and location parameters m and μ. There is a perfectly sensible statistical problem, if not likely to arise in practice. But it is easy to see that no likelihood function can be defined.

Professor P. D. Finch (*Monash University*): It is to be expected that any one who has dealt with a broad range of practical problems will sympathise with the eclectic viewpoint put forward by Professor Cox, to the extent at least that, as a matter of fact, the present state of statistical practice is such that each of the main approaches has, in its proper place, added to rather than detracted from his understanding of the world. Nevertheless it is important to offset pre-emprive claims on behalf of this or that approach by emphasising, as Professor Cox has done today, the fruitfulness of analysing problems in different ways at different depths. But it would be unfruitful if acceptance of that thesis served, albeit unwittingly, to suppress curiosity about relationships *between* various approaches and suggested, by implication, that there was little to be gained from the development of a statistical framework on which the proper place of each of them could be manifestly exhibited in an unambiguous way. Thus whilst I endorse Professor Cox's emphasis on the *fact* of it being advantageous to choose one's approach according to context, curiosity forces me beyond that point and leads me to ask "why should this be so?"

The simplistic answer is, of course, that even as there are many different types of practical problem so too will there be diversity in the responses appropriate to their resolution. It is worthwhile, however, entertaining the possibility that the answer is more complex than this, that there is indeed a general framework of the sort envisaged but that, somewhere in the past, directions which might have revealed it were inadvertently neglected, perhaps by reason of their contexts going out of fashion or simply because realisation of their promise required, in fact, the prior development of other lines of enquiry.

Without any attempt to substantiate these suggestions here let me indicate their force by directing your attention to Bishop Butler's doctrine of probability as presented in his *Analogy of Religion* which appeared in 1736, two years before Daniel Bernoulli's introduction of "moral expectation" into the investigation of fairness in games of chance. Butler expounded the dictum "Probability is the very guide of life" and, had he been of Roman rather than Anglican persuasion, might well have become, thereby, the patron saint of statisticians. More importantly, however, his viewpoint, though firmly grounded in the tradition of British empiricism, was presented in a strictly theological context and, whilst its influence extended at least to 1870 with the publication of Cardinal Newman's *An Essay in Aid of a Grammar of Assent*, on cannot but suspect that this would have been more widely felt had it not originated in a setting which has since become so unfashionable. The relevance of Butler's thought for the matters at hand is best conveyed by his own words; consider, for example, the following passage:

"*That which chiefly constitutes Probability is expressed in the word Likely, i.e. like some truth, or true event; like it, in itself, in its evidence, in some more or fewer of its circumstances. For when we determine a thing to be probably true, suppose that an event has or will come to pass, it is from the mind's remarking in it a likeness to some other event, which we have observed has come to pass. And this observation forms, in numberless daily instances, a presumption, opinion, or full conviction, that such event has or will come to pass; according as the observation is, that the like event has sometimes, most commonly, or always so far as our observation reaches, come to pass at like distances of time, or place, or upon like occasions.*"

This passage shows very clearly that in Butler's day the term "likely" could be used in respect of a specific relationship *between* things, viz. the *likeness* between them, whereas nowadays, of course, "likely" is more usually predicated of a thing, as when we say that it is likely, and the former emphasis on likenesses has fallen into relative disuse. But emphasis on analogy, from one thing to another, focuses attention on the fact that, in practice, we are involved with describing one thing by another, the propriety of that description varying with the extent to which the describing object is like the thing being described. For Butler, therefore, probability would seem to be chiefly a measure of the extent to which one thing is like another and hence, of the propriety with which one of them may be used to describe the other. This view of probability underpins Butler's attitude towards what we would nowadays refer of as model building or formulating an appropriate family of probability measures as, for example, when he says:

"*Forming our notions of the constitution and government of the world upon reasoning, without foundation for the principles which we assume whether from the attributes of God or any thing else, is building a world upon hypothesis, like Des Cartes. Forming our notions upon reasoning* *from principles which are certain, but applied to cases to which we have no ground to apply them, (like those who explain the structure of the human body, and the nature of diseases and medicines from mere mathematics without sufficient data), is an error much akin to the former: since what is assumed in order to make the reasoning applicable is Hypothesis. But is must be allowed just, to adjoin abstract reasonings with the observation of facts, and argue from such facts as are known, to others that are like them; from that part... which comes under our view, to that larger and more general government over them which is beyond it;...*"

It would be foolish to suggest that Butler's remarks provide a complete blueprint for statistical practice but they do contain the germ of an idea which has since been neglected in favour of an emphasis more in accordance with recipes for building a model world on hypotheses à la Des Cartes. By way of contrast Butler's view of argument by analogy places the emphasis on description rather than inference about a model and to that extent runs counter to popular statistical thought. The view is, in fact, that the statistician is chiefly concerned with describing data, by saying of it that it is like to something else in this or that respect and to such a degree as is made evident by the data itself in the light of some designated gauge of the extent it is like the object used to describe it.

To formulate this idea in a more precise way we might consider two finite sets, the elements of one being objects to be described and those of the other being the describing objects, together with a measure or gauge of the degree of likeness between any two of them. In respect of a designated describing object δ and a designated described object ω one can compute the proportion of describing objects providing descriptions of ω which, according to the gauge in question, are no better than the designated one, this proportion I call the descriptive power of δ as a description of ω. If it be large, that is to say near unity, then δ is a relatively good description of ω inasmuch as relatively few of the descriptions then under consideration would be better than the one at hand. In like manner one can compute the proportion of described objects which are not better described by δ as a description of ω. If it be large then ω is relatively well characterised as a type δ object inasmuch as relatively few of the objects under concideration are better characterised in that way. In general one is looking for descriptions which have both high descriptive power and high characterising power. Unfortunately it turns out, as one might anticipate, that one often has to sacrifice something of the one to achieve what is desirable in the other.

Within this broad framework the various approaches of statistical theory re-appear in the way one chooses its three basic ingredients, viz.

what one takes to be the objects described, what one takes as possible descriptions and finally the way in which one gauges likeness between them. This freedom of choice accounts, in fact, for the way the various approaches are seen as competitors but it would be more sensible, and surely in the spirit of Professor Cox's eclectic thesis, to regard the *relativity* of description which results from those choices to be the chief object of statistical enquiry rather than the individual and possibly ideosyncratic description *per se.*

In practice, when one is involved with a designated pair (δ, ω) one is often interested in how descriptive and characterising power will vary under perturbations of the data ω and, in like manner, under perturbations of the description in question. It would, for instance, be without meaning to dialogue if we used a description δ of ω which turned out to have high descriptive power whatever the data in question. In this regard one might be inclined to think of a description δ as a possible "hypothesis" describing or explaining the data—for instance ω might be a set of ordered pairs in some finite region R of the real plane whereas δ is the set of ordered pairs in some designated subset D of that plane and the data is described as being like D. It turns out, however, that the ordinary power function calculations of statistics pertain not to alternative hypotheses in the sense of perturbations of the proposed description but rather to perturbations of the data itself. This fact enables one to recast some of the basic problems of statistical practice in a new and interesting way. But the exposition of that fact is not the point of today's meeting. The measure of how worthwhile one finds another's paper is the extent to which it forces one to think about its subject matter. I have clearly found Professor Cox's talk very worthwhile and I am sure you have too.

Reply by Professor Cox: I am very grateful for the constructive comments that have been made. I agree with Professor Moran that tests need consideration of alternatives, but these need not be formulated probabilistically, so that the only explicitly formulated part of the model may be the null hypothesis. Critical use of asymptotic theory is capable of resolving satisfactorily most "practical" problems, but it does seem disturbing that the rather large armoury of "exact" theory solves only such very special situations.

Professor Finch has made a number of interesting general points; certainly, as I indicated at the start of my paper, some more unified approach would be very attractive. His quotations are most interesting. Some of Butler's notions seem a little similar to H. Reichenbach's approach to probability in which $\mathrm{pr}(A)$ is assessed by finding a series of events judged to be about equally probable with A.