DISCUSSION

G.A. BARNARD (*University of Waterloo, Canada*):

I welcome very much Professor Akaike's presence at this conference; not only because I have not before had the opportunity to discuss issues with one whose contributions to statistics have been so constructive and important, and I welcome this now; but most of all because I have been coming to think that the apparent division of statisticians into doctrinally opposing 'schools' of Bayesian and anti-Bayesians was doing great harm to our subject, and would do more if it was allowed to continue. The purity of doctrine of the organising Committee of this Conference is beyond question; also beyond question is the fact that Professor Akaike is not a subscriber to the pure doctrine. The fact that he was invited, and accepted, to speak here is therefore specially welcome.

To deal first with a minor point. I cannot go along with Professor Akaike's criticism of Savage's axioms, any more that I think Wolfowitz's criticisms were justified. Surely the young boy's difficulty arises from his regarding the difference of ranks by 1 as marginal. Unless he moves himself and his friends to a Muslim country he will have to decide eventually; and his eventual choice may always be supposed to arise from a perception that one, at least, of the 'marginal' differences is, in fact, more important than the others.

Wolfowitz's criticisms seemed to me misdirected, because if we have a *general* rule for choosing one from among any set of decision rules, we must be able to rank the set.

However, I would go along with a slight re-formation of Professor Akaike's argument. For the 'inconsistent triad' to which his argument leads arises also in the Arrow-Condorcet Theorem concerning the impossibility of collective 'democratic' choice. And I agree that statistics is nothing if it is not concerned with *objective* analyses of data, in some sense -- whatever we may care to say about the 'nature' of the probabilities with which we deal these probabilities must be *agreed* between several people. And it follows that Savage's argument does not serve to demonstrate the necessary existence of such *agreed* propabilities for any proposition we care to think of. We cannot therefore take for granted the existence of agreed prior probabilities for all the parameters involved in a model of an experiment. But such agreed priors are necessary for the universal applicability of Bayes'Theorem.

Professor Akaike appears to accept what I call the Likelihood Model (LM) as typical of the logical structure of an experiment. This specifies the sample space $S = [x]$ of possible results, the parameter space $\Omega = [\theta]$ of possible parameter values, and the probability function $f(x,\theta)$ giving the probability of x when the parameter value is θ. Given the three elements $[S,\Omega,f]$ we can deduce the distribution of the likelihood function and from it derive, at least in some cases, a 'prior' distribution for θ in accordance with Professor Akaike's principles. If we then represent the inference by the posterior distribution of θ relative to this prior and the observed likelihood function we shall obtain an inference which has a clear frequency interpretation. I assume that Professor Akaike would accept that such an inference is appropriate only when we really have no observational basis for any statement about the parameter values other than that which is implicit in the design of the experiment.

We should not, I think, be over-ready to assume we are in this state of ignorance. For example, as Akaike shows, his rule can be made to yield the rules of ridge regres-

sion. But this, in spite of its recent vogue, by no means always gives an improvement on standard least squares. A group of statisticians in a large chemical company were persuaded to review a sample of experiments they had analysed in the past, in cases where the true values of the parameters had become essentially known. It turned out that ridge regression was, on the whole, worse than ordinary least squares for these cases. A reason for this could be found in the fact that ridge regression can be seen to be equivalent to an assumption that the parameters being estimated themselves follow a spherical normal distribution centred on the origin. Thus if all parameters in an experiment are roughly of the same order of magnitude, and their signs are randomly distributed, we can expect ridge regression to improve over ordinary least squares. But if, as with the experiments reviewed, one or two of the parameters were very large, while the remainder were quite small, ridge regression would not do so well. Thus background knowledge of the kinds of parameter value likely to be met with should be used in addition to such information about the prior as may be deduced from the experimental design. This last, in fact, can only be supposed to convey information about the prior in so far as it reflects the knowledge of the experimenter.

Thus I believe that when we use a Bayesian model for the analysis of an experiment we should regard the prior distribution in much the same way as we regard the form assumed for the probability function as something which may perhaps be important to our inference, so that we should be careful to check how far this is so; and as something which is capable of objective verification, at least in a long run and which, of course, we should so verify. This long run verifiability is, I think, the source of the objectivity, such as it is, which may be claimed for a Bayesian analysis.

In appealing to long run verifiability, of course, we need to specify the long run we consider relevant, the class of experiments to which we judge the current one belongs. A chemical engineer will find little difficulty in viewing the current chemical reaction rate constant which he is measuring as one of a set of such rates; and for the given equipment which he has available he will have had to set his temperature and other features of his design so that the reaction rate is neither too fast nor too slow to be measurable. He will then not go far wrong in using a prior distribution which is reasonably uniform over the range of measurable values. Similar considerations apply to an econometrician measuring elasticities, etc. But a physicist who is measuring the velocity of light, or some fundamental natural constant, would find it hard to regard his parameter as just one of a class of such, following a distribution concerning which he has any knowledge at all. This is why, it seems to me, Bayesian models are appropriate for experiments in chemical engineering, or in econometrics —provided, of course, the conclusions are understood as being subject to the correctness, to sufficient approximation, of the prior assumptions— but they are less appropriate for fundamental work in physics.

Professor Dawid has, as usual, presented us with a paper which stimulates us to further examination of foundations. But he omits, I think, to question a presupposition which ought to be questioned: Do nuisance parameters, as now commonly understood, exist? How often can we, or should we, 'ignore' parameter that enters into the specification of our experimental model?

When Hotelling first introduced the term, 'nuisance parameter' meant just what it

said: a parameter that one would have preferred to omit from one's model. But now the term has come to have an unfortunate use among the adherents of what (at Barndorff-Nielsen's suggestion) may be called the 'prespecification school' of mathematical statisticians. By this is meant the school which, given a model for an experimental situation, lays down *in advance of the data* the *kind* of conclusion that is to be reached. In relation to nuisance parameters, a typical requirement is that a 'test' of prespecified size should be provided that is 'unbiased' or 'similar'. Such prespecified requirements can easily lead to absurdity. The 2x2 table:

	A	*not-A*	*Total*
Population I	*a*	*b*	*m*
Population II	*c*	*d*	*n*
Total	*r*	*s*	*N*

provides a simple example. If p_1, p_2 are the probabilities of A in populations I and II respectively, we often are interested in the crossratio parameter $\theta = p_1q_2/p_2q_1$, where $q_i = 1\text{-}p_i$, $i = 1,2$, and less interested in the 'nuisance parameter' $(p_1+p_2)/2$ which we may denoted ϕ. If we now prespecify (as is done, for example, in Lehmann's book) a test of $\theta = 1$ of size (say) 0.05, which is to be similar, or unbiased, against alternatives $\theta \neq 1$, we must reject the hypothesis tested with probability not less that 0.05 when $p_1 = 2 \times 10^{-10}$ and $p_2 = 10^{-10}$. But when this is the case we will, with practical certainty, get the result $a=0$, $b=m$, $c=0$, $d=n$, and so we must reject, given this result, with probability 0.05. But to reject at all with such a result is clearly absurd.

One of the biggest advantages of the Bayesian approach over that of the prespecification school is that the Bayesian model gives a primary inference in the form of a posterior distribution for all the parameters involved in the experimental model. *If*, for example, the posterior distribution is very nearly normal, then it may be judged reasonable to express the conclusion in terms of an 'estimate' with a standard error; but whether this will be so, or not, may well depend on the data as well as on the model and the prior. And it may turn out that we can express the posterior in terms of two parameters θ and ϕ such that, a posteriori, these two variables are, to a sufficient approximation, independent. In such a case we can treat ϕ as a 'nuisance parameter' in relation to θ; but to do this in other cases can be dangerous. Certainly the mere fact that we would *like* to make an inference about θ without referring to ϕ by no means implies that we can. If we insist on doing so we are guilty of adopting the 'prespecification' approach.

Thus Dawid's statement (p.5) that 'From the point of view of the single Bayesian B, the marginal likelihood is as good as any ordinary likelihood. . .' needs qualification. If θ and ϕ, given x, are far from independent, then further information which may well come to hand concerning ϕ will affect the conclusions we draw concerning θ; and the mere statement of the marginal distribution of θ will give no expression to this fact. By contrast, if we have an 'ordinary' likelihood for θ — that is, one from an experiment in which θ alone is involved — then no further information about another parameter ϕ alone will affect our conclusion, provided θ and ϕ are independent a priori. I am, of course, assuming here something which I regard as fundamental to natural science — the possibility of *knowing* that two distinct experiments are independent of each other.

I am led to wonder whether the term 'nuisance parameter' should not be left to the exclusive use of the prespecification school. I have long thought it unfortunate that Student's t statistic should have such beautiful properties that we are too often tempted to make inferences (posteriors or confidence distributions) which relate to location μ only, when in fact we almost always ought to make simultaneous inferences about location μ and scale σ together. Terms like 'primary' and 'secondary', to indicate a *ranking* of our interest, rather than a total lack of interest, would usually be more appropriate. George Box has introduced the term 'discrepancy parameter' to describe the kind of parameter in which our interest is minimal. The concept which I would like now to discuss is a little different, I think, and I shall use the term 'model adjustment parameter'. I hope I will not be drummed out of the Conference if I describe the idea in connection with a 'classical', not-necessarily-Bayesian problem:

We are given two samples, of size m, n respectively, from normal populations with means μ_1, μ_2, and standard deviations σ_1, σ_2, all unknown. We want to test whether $\mu_1=\mu_2$ or not. In the text books we are told that if it is known that $\sigma_1 = \sigma_2$, then we can reduce the problem to a t-test; but if it is now known that $\sigma_1 \neq \sigma_2$ then, typically, we are told the problem is 'difficult'. We may, or may not, be referred to Fisher's tables, or to Welch or to Gurland. What never occurs, so far as I can tell, is an invitation to set $\lambda = \sigma_2/\sigma_1$ and then put

$$t(\lambda) = (\bar{y}\text{-}\bar{x}) \;/\; \sqrt{\left[\left((1/m) + (\lambda^2/n)\right)\left((m\text{-}1)s_x^2 + (n\text{-}1)s_y^2/\lambda^2\right)/(m+n\text{-}2)\right]}$$

In many cases, if we plot $t(\lambda)$ over the plausible range of λ we shall find that it varies only trivially. In this case, we can make our inference about the difference between μ_1 and μ_2 knowing that it will be unaffected by the possible difference in the variances. If things turn out otherwise, then, but only then, we must either obtain more information about λ, or we must resort to Behrens-Fisher, or some such type of argument. I myself cannot recall a practical case where the inference was seriously affected by λ.

The parameter λ here plays the role of what I propose to call a 'model adjustement' *(MA)* parameter. This is a parameter which needs to be specified in order to define the distributions involved in our experiment, but which varies over a relatively narrow range. We have reason to suppose that our inference will, with high probability, turn out not to depend on that value of the *MA* parameter. Should we be disappointed in this hope, then we must either record the fact that our inference depends on the value taken for this parameter, or we must find out more precisely what value this parameter really takes. Finally we may use some special form of argument to derive an inference which *explicitly* depends on ignorance of the *MA* parameter value.

Typical of the arguments 'from ignorance' here referred to is that involved in the derivation of the Behrens-Fisher test, where we have a pivotal quantity $s_x^2\lambda^2/s_y^2$ for the parameter concerning which we wish to express our ignorance. We condition on the observed ratio s_x^2/s_y^2 and *conventionally* retain the distribution of the pivotal by a conceptual distribution of the parameter λ. Statements we then make concerning the other parameters must be interpreted as referring to the reference set thus conceptually generated. Whether or not such modes of reasoning come to be generally understood and accepted

will largely depend on whether the results they give appear 'reasonable' to the scientific community at large; in this respect the conventions as to the interpretation of 'ignorance' thus introduced may be compared with such conventions as the interpretation of the terms 'set', 'class', etc., in the foundations of mathematics. Most mathematicians seem to agree that the results derivable using the 'axiom of choice' correspond to their 'intuition' of the sort of structure that mathematics ought to be. It is known that the axiom of choice can be negated without thereby introducing a contradiction into set theory;but systems built on such negation seem in some sense 'pathological' to most mathematicians. To put the matter loosely, when we ask about the difference of means when both location parameters are unknown, we are asking a slightly silly question; and we must be content with a slightly silly answer. Certainly the answers thus arrived at, subject to marginalization paradoxes though they be, seem to me to have as much Bayesian justification as the answers obtained by simple marginalization from a posterior involving the secondary, MA, or nuisance parameter if we then forget that our conclusions concerning the parameter of primary interest are subject to modification should further data become available concerning the parameter we have integrated out.

Thus I agree with Zellner's comment concerning 'marginalization paradoxes'. We must remember that statistics is intended for application to the advancement of science. Fanatical insistence on freedom from 'incoherence' can lead to such complicatedly interrelated analyses of data as to go well beyond the capacity of our understanding. Judicious simplifications are an essential component of scientific advance.

P.R. FREEMAN (*Leicester University*):

When I first read Professor Akaike's paper I thought "If he goes to Spain and reads that, he'll be a brave man indeed". Well, he has - and he is. How can I react? I could fill all the discussion time with an uptight, strict Bayesian reply but this would be too negative. I must first, though, say that I can see no force in the counterexample to Savage's axiom of choice and that only very rarely (as in weather forcasting) am I at all interested in the expected performance of a Bayesian procedure. I can't therefore see any sense in the argument of section 2.3 and would happily condemn to the statistical mental asylum anyone who needed to know whether sampling was going to be direct or inverse before stating his prior for θ. Similarly, after many close readings of section 3 I am still not clear exactly what "objectivity" is claimed for the likelihood function and prefer to stick to the viewpoint of that great statistician Shakespeare (1598) who said

> But (by your leave) it never yet did hurt,
> To lay down likelihoods and forms of hope.

Likelihoods are, to me, just as much "forms of hope" as any other ingredients in the inference mixture.

To be more positive, let me turn to matters on which we can agree wholeheartedly. I take Professor Akaike's point to be that there are more things in real analysis than are dreamed of in any of our statistical philosophies. There must always

be a rather messy interplay between the data and the choice of model, of parameters and of priors on those parameters if our analyses are to be of any value at all. This paper presents some very ingenious ways in which this can happen and they all show great promise in the applications we see. But we can all do quite well (well, nearly all, my own paper being one exception) when we generate an artificial set of data with known parameter values, know we are using the correct model and furthermore reuse the data to choose the best prior for us. Figs. 3 and 4 are the only ones relating to real data, so I should like to see several more real examples before judging the results.

To me, the two fundamental questions raised by this paper are:

i) Do these ideas give us any more insight or flexibility than could be obtained by keeping to Bayesian orthodoxy? Is there any reason to suppose, for example, that choosing d to maximise $L(d,\sigma_d^2)$ is any better than letting it be a hyperparameter of the prior distribution for a, itself having a suitably woolly distribution? The latter gives you all the advantages of coherence and allows the data to dictate automatically what are the likely values of d and to give a suitably weighted posterior distribution for a.

ii) Does the gain in common sense outweigh the ad-hockery that is immediately needed as soon as coherence is abandoned? Why, for example, do we take $c_0 = 0$ in example (a), why the particular choice of D in examples (b) and (c), and so on? If we are not very careful we shall find ourselves in just as muddled a state as the poor frequentists.

Finally, I am puzzled by the last example on polynomial fitting where no mention is made of the purpose. Do we just want a good fit or a good prediction, or do we really want to know the "true" order of the polynomial and to estimate its coefficients? Without any context I can't judge the meaning of the results presented.

Professor Dawid disarms criticism of his paper by openly admitting that much of it is not of direct interest to Bayesians. Here at least is one statement I can broadly agree with. The paper does give me one way of telling when a frequentist is being incoherent, but frequentists are so seldom coherent that this is somewhat superfluous. Those of us who enjoy explicitly exposing the incoherence of frequentist methods might find some of the results here useful, however.

In the components of variance example (5.6), it seems essential to allow $\tau^2 < 0$ in order to get all information about σ^2 concentrated in S_3. This is not as crazy as it seems and has indeed already been advocated by Nelder (1977). Since τ^2 is the *excess* of variance between rows over variance within rows, a negative value is possible but has strange implications. The correlation between a pair of observations in different rows (value of i) has to be *greater* than that between a pair in the same row. It is hard to imagine real datasets where this would happen.

I should like to ask if any of the results in this paper throw any more light on that undefined concept of "no available information about Θ in the absence of knowledge of Φ" introduced by Kalbfleisch and Sprott (1970). I remember the concept coming under heavy attack at that time, and the authors trying hard to make it rigorous, but I cannot recall seeing any further published work.

Finally, the distinction between parameters of interest and nuisance parameters is not always at all clear. In model discrimination problems, for example, we do not know

which parameters will be of interest until we have decided which model is most likely to be true. Perhaps we need to introduce the idea of nuisance models here.

D. PEÑA (*Escuela de Organización Industrial, Madrid*):

My comments on the papers for this session will be limited to the paper by Professor Akaike, because it appears to me to be the most ambitious and most polemical of the two papers, at least within the context of this conference, and because it touches areas that are more related to my particular interests and competence.

Briefly, the paper by Professor Dawid appears to me to confirm what Bayesian Statisticians already know: namely, that the treatment of nuisance parameters within the Bayesian framework is general and coherent, in contrast with the many partial solutions adopted by classical statisticians.

My criticisms of the paper by Akaike fall into three categories: (1) I do not agree with a number of the general methodological comments made in the paper; (2) I am not convinced that the goodness-of-fit criteria, based on the Kullback-Leibler measure of information, suggested by Akaike, provide a significant improvement over previously existing criteria; (3) it appears to me that the general linear model, developed by Akaike in this paper, is mainly designed to solve the problem of fitting many parameters to few observations, and therefore focuses on the solution of problems in practical statistical analysis that are, initially, so ill-defined that the investigator, no matter what methodology he uses, can learn little from the data.

Beginning with the first point, general methodological questions, I do not share the opinion, expressed by Akaike (Section 1), that Bayes procedures represent only "one possible way of utilizing the information provided by the likelihood function". I would agree, with Jeffreys and others, that Bayes methodology embodies the scientific principle of "learning from experience" in an essencially non-deterministic world. The justification of Bayesian methodology is, in my view, that it provides a unified and internally consistent approach to dealing with uncertainty, both in the context of statistical inference and decision.

Professor Akaike presents two objections to the subjective interpretation of Bayesian procedures. First, he objects to the postulate of linear ordering of preferences in Savage's axiom system, and offers an example of a preference structure that appears, at first sight, to be sensible, but in fact is not transitive. It seems clear to me that the transitivity axiom is needed in any coherent theory of decision that is to be applied to real life problems with any degree of success. Raiffa (1968, pp. 75-86) has shown, in a very convincing way as far as I can see, how it is always possible to build a "money-pump" against the intransitive subject.

The second objection, in Akaike's words (Section 2.2) is:

> "To take the parameters (as) something prespecified and assume that the prior distribution can or should be determined independently of the data distribution constitutes a serious misconception about the inferential use of the Bayes procedure".

I certainly agree that, in principle, the data distribution should be taken into account in specifying the prior distribution in the non-informative situations typical of much of statistical inference. However, this is not a new point and, in the concrete example offered by Akaike (Bernoulli versus Pascal sampling), it is not of much practical importance; see Box and Tiao (1973, pp. 45-46). The dependence of the prior distribution on the data distribution is also present in the maximal-data-information prior distributions suggested by Zellner (1977).

In summary, with respect to general methodological questions, the "Conceptual difficulties of the subjective approach" suggested by Akaike do not seem convincing to me, and therefore, it does not seem to me to be necessary to look for new foundations for Bayesian Inference.

I now move on to a second class of comments, those related to the new goodness-of-fit criteria developed by Professor Akaike. This paper introduces a new information criterion, the *ABIC*, to select the optimal value of the constant d in his mathematically elegant, general linear model. In essence, this new criterion, the *ABIC*, is simply the older criterion, the *AIC*, also developed by Akaike, applied to the general linear model of this paper. These statistical criteria are based primarily on the Kullback-Leibler measure of information, but their justification, as far as statistical optimality is concerned, has remained heuristic. I am sure that a strengthening of the tie between information theory and statistics is a useful research objective, but I suspect that the particular criteria presented in this paper are equivalent, in most cases, to classical statistical test criteria. To support this view, let us consider a problem frequently treated by Professor Akaike (1974, 1976, 1978) in which the minimum *AIC* is applied: The selection of the order of a stationary normal autorregressive stochastic process. In this case a model with $p + k$ parameters is chosen over a model with only p if:

$$AIC\,(p + k) < AIC\,(p).$$

The above inequality is equivalent to:

$$N \operatorname{Ln} \hat{\sigma}^2\,(p + k) + 2\,(p + k) < N\,Ln\,\hat{\sigma}^2\,(p) + 2p,$$

where $\hat{\sigma}^2\,(p + k)$ and $\hat{\sigma}^2\,(p)$ are the estimated residual variances of the two models, and N is the number of observations. This implies:

$$\mathrm{N\,Ln}\,\frac{\hat{\sigma}^2(p)}{\hat{\sigma}^2(p+k)} > 2k$$

and, using the fact that $Ln\,(1 + x) \approx x$ when x is small, this reduces to:

$$\frac{N\,\{\hat{\sigma}^2\,(p) - \hat{\sigma}^2\,(p + k)\}}{\hat{\sigma}^2\,(p + k)} > 2k$$

which is equivalent to:

$$F_{K,N} = \frac{N\{\hat{\sigma}^2(p) - \hat{\sigma}^2(p+k)\}}{K\hat{\sigma}^2(p+k)} > 2$$

where $F_{K,N}$ is the statistic F with K and N degrees of freedom. In this calculation I have used the assumption that N is large enough so that $N - p \approx N$. Asymptotically, we obtain the classical likelihood ratio test based on the χ^2 with k degrees of freedom (Bartlett (1978), pp. 306-307):

To summarize, if N is large, the AIC is equivalent to a likelihood ratio test based in the χ^2 with k degrees of freedom and critical value of $2k$. The fact that this critical value remains equal to $2k$ explains the observed behaviour of the AIC and, in particular, its bias toward overparametrization pointed out by Shibata (1976).

My third category of comments referrs to the importance of choosing a parametrization that facilitates the process of learning from the data. To illustrate the usefulness of his general linear model, Professor Akaike considers the decomposition of a time series into trend, seasonal and irregular components, a problem that I find particularly important for those of us who are working in practical time series analysis. In this application the formulation by Akaike involves $2N$ parameters for the decomposition of N time series observations. The determination of this very large number of parameters is based mainly on a priori restrictions. This procedure can be regarded, in Bayesian terms, as equivalent to the application of a highly informative prior distribution about the structure of the decomposition.

We would expect this procedure to yield reasonable results in those cases in which seasonal structure is very clear at the outset, as it is in the cases presented in the paper. However, in cases in which the seasonal structure is not at all clear from the outset, we will face one of the following two unpromising choices: (1) to apply the same kind of restriction used by Akaike in the cases of the paper, which may well permit us to learn little from the data; or (2) to formulate a new set of restrictions with no guidelines for this selection.

For these reasons, I feel uncomfortable with this solution to the decomposition problem. I believe that learning from experience means, among other things, to allow the data to correct our a priori beliefs. To achieve this end, I would prefer procedures more in the spirit of Box-Jenkins (1970), that is, the use of a well-designed system of diagnostic checks, together with an iterative process of model-building which, of course, must place major emphasis on parsimonious parametrizations. In this sense, it seems to me that work along the lines of Box, Hillmer and Tiao (1976) is more promising than that presented in this paper for the time series decomposition problem.

In closing, I would like to thank both authors for their contributions to this session. To Professor Dawid I would like to express my regrets that my fields of interest and competence have not permited me to pay more attention to his paper, and I would like to thank Professor Akaike for the very stimulating and polemical paper that he has offered us on this occasion.

J.M. DICKEY (*University College Wales*):

The emphasis in the paper by Professor Akaike seems wrong to me. He writes, "It is almost trivial to see that no practically useful Bayes procedure is defined without the use of the likelihood function". This may be true in the narrow technical sense of the word "procedure", namely for an act-valued function defined on a sample space. However, such Bayesian methods can also be viewed as forming a mere subarea of subjective probability modeling in which expert opinion is quantified in its various complexities, joint dependencies, and conditioning on concomitant variables and on experimental data. It is then a rather special case to have statistical data on which to condition, and an imbedded statistical model, by which Bayes' theorem would be the form taken for the probability conditioning. This seems to be the view taken by De Finetti in his work, and it also describes the standpoint of my own paper in these Proceedings.

Think of the problem of probabilistically quantifying a physician's opinion of a cancer patient's survival under treatment with a combination of radiotherapy and a particular new drug. Suppose no proper statistical data is yet available. Probabilistic predictions (previsions) are needed for various types of patients and various treatment schedules. Perhaps, an experiment needs planning. How shall the expert's opinion be used now in planning the experiment and treating those patients who cannot wait for the definitive data?

Should it be used in the form of a subjective probability model fitted to elicited aspects of his opinion? Or should chaos reign in the deliberate rejection of any theory? Does no data mean nothing can yet be done? Perhaps one prefers to use subjective probability rather than have chaos. It is easy to say "yes" here to subjective probability modeling in an absence of statistical data, because there are no competing methods.

But now, if one says, "Yes, I shall quantify opinion probabilistically", what I say is, "Suppose one has a *little* bit of statistical data; does one now use some entirely different approach not based on subjective probability?" Suppose there is not enough data for maximum likelihood or for the use of an ignorance prior to yield sensible probabilistic previsions. And now I ask "What if one has a little larger amount of statistical data?"

You see what I am driving at. At what point does one throw away the notion of quantifying opinion by probability? At what point does one say, "I am no longer willing to specify a prior distribution as an expression of opinion"?.

S. GEISSER *(University of Minnesota):*

I am highly sympathetic to the view advocated by Professor Akaike and others that in certain contexts the prior distribution of a parameter need not be determined independently of the data distribution (likelihood). Whenever (as often is the case) the parameter is a hypothetical construct, unobservable, and artificially devised to promote a convenient model and useful only inasmuch as predictive distributions can be calculated, there seems to me no grave difficulty in taking this view. Professor Akaike, however, has really taken the bull by the horns when he chooses a coin tossing

experiment to illustrate his view that it is irrational to adopt one and the same prior for the two sampling plans that led to the same likelihood for the parameter θ. In my view it is very difficult, if not impossible, to argue that it is rational not to adopt the same prior in this particular situation.

In this situation, if anywhere, θ comes closer to being a physical property of the coin than in most other experiments statisticians deal with. The sampling plan can in no way affect this property. Hence one can rightly argue that the two different sampling rules invoked are irrelevant towards inferring about this "physical entity". If one takes the view, as I do, that even in this case the predictive distribution of a future observation is paramount rather than the posterior distribution of θ - neither should be affected by the sampling rule once the sample is in hand. If one takes this from the ususal parametric framework (for a predictivist it is always more confortable to be able to frame the problem in terms of observables) and one can do so to a degree in this case, we can sharpen the divergence of opinion on what is rational. To my mind, there is always fuzziness in frameworks involving hypothetical unobservables. Jeffreys (1939) discusses the case where there are N binary trials with an unknown number R of one type and N-R of the other. A sample of n is drawn and T of one type observed and the predictive distribution of R obtained,assuming all possibilities are, a priori, equally likely for R. Here the sampling is hypergeometric. One could have also sampled until T was observed and hence obtained a negative hypergeometric sampling distribution for the total sampled. The "likelihood" (actually in either case it is a probability conditional on the potential observable, R) of R is unaltered as in the parametrized negative binomial - binomial situation.

It appears that here in the completely observable situation, Professor Akaike would be on very precarious ground in sustaining his view that it is irrational for the same statistician to have a single prior for R given only that the sampling plan was at issue.

With respect to Professor Dawid's paper, if one restricts one's attention to the prediction of observables or potential observables, then the problem of nuisance parameters, with its imposing glossary of terms, completely vanishes. Although this is my philosophical stance, I admit to harboring some genuine regret as to having my view universally adopted since it would preclude the appearance of much elegant research such as Professor Dawid's and many of those listed in his references.

D.V. LINDLEY (*University College London*):

The criticism of the axioms offered by Professor Akaike fails to distinguish between the descriptive and the prescriptive views. A person who has preferences like the young boy would lose money for sure and, although it may be an accurate description, it is hardly a prescription for sensible behaviour. My description of Akaike is closely related to that of a prescriptive person: he obtains sound answers for wrong reasons.

An alternative approach to polynomial fitting is available by Young (1977). He fits polynomials of very high degrees using a prior that reflects scientific opinion that low-

degree polynomials are more reasonable than those of high degree. This approach finishes up with a low-degree polynomial and avoids the difficulties of choice between models. Generally, it often seems sensible for a Bayesian to fit the largest model he can. Model choice is really a decision problem of what variables to observe in a future experiment.

A. O'HAGAN (*University of Warwick*):

I find myself in disagreement with some of the things Professor Akaike has to say, for instance the whole of sections 2 and 3. But Professor Akaike has too much experience with data to produce silly analysis however misguided his philosophy might be, so I was not surprised that the technique he advocates in section 5 for estimating the variance parameters σ^2 and d is perfectly sensible. In fact, in O'Hagan (1976) I reached a similar conclusion, that one should (*a*) estimate variance parameters by the mode of their marginal distribution (after integrating out the other parameters), then (*b*) estimate the other parameters by the mode of their conditional distribution given that the variance parameters have the values obtained in (*a*). Professor Akaike does not put priors on σ^2 and d, so his step (*a*) is a maximization of "marginal likelihood".

A.F.M. SMITH (*University of Nottingham*):

The examples presented in Section 4 of Professor Akaike's paper are interesting examples of what *I* would call, in constrast to the opening paragraph of that section, "the common-sense approach to constrained least squares". If the author is interested in "the common-sense approach to Bayesian Statistics" he might try Lindley and Smith (1972).

REPLY TO THE DISCUSSION

AKAIKE, H. (*The Institute of Statistical Mathematics, Tokyo*):

Just before the presentation of my paper I felt that I was rather out of place. After receiving the comments I recognized that my participation in the meeting was extremely rewarding. I must express my sincere thanks to the organizing committee and those who contributed to the discussion for providing me such an enjoyable intellectual experience.

Professor Barnard disagrees with my critical view of Savage's postulate on linear ordering of preference. Nevertheless, by the recent review article of Professor Good (1979), it seems that Savage himself considered his system of subjective probability incomplete, as it rejects the concept of randomization. To accept the concept of randomization is equivalent to accepting the impossibility of uniquely specifying a prior distribution.

Professor Barnard's warning against assuming ignorance without sufficient analysis of a particular situation is extremely valuable. My recent experience on developing a smoothness prior for the distributed lag model treated by Shiller shows that a Bayesian model can produce a significantly distorted image of the reality (Akaike, 1979). It seems that the only sensible way out of this difficulty is to develop several alternative Bayesian models and evaluate their likelihoods with respect to the available data.

Professor Barnard's general opinion on the use of Bayesian models is so close to mine that it is almost impossible for me to point out any significant differences. The basic idea here is to base the justification of the use of a Bayesian model on the following identity

$$\text{objective} = \text{social} = \text{long run}.$$

We consider that the information expressed in terms of a prior distribution must at least be communicable. This communicability can only be gained by placing the prior distribution within the context of its particular application. This observation, I think, is the gist of Professor Barnard's comments.

Finally, I wholeheartedly support Professor Barnard's view on the danger of the excessive separation of doctrines of statistics. Each doctrine tends to suppress activities outside of it. At one point this tendency begins to act against the progress of human knowledge. A real innovation can never be placed properly within an existing doctrine and there should be no end of the progress of human knowledge.

Professor Freeman surprises me by rejecting the basic Bayesian principle of rationality, the maximization of expected utility. He then violates the teaching of subjective probability by ignoring, without reason, the information of whether the sampling is direct or inverse in the case of a binomial experiment.

Professor Freeman is particularly sensitive to "objectivity", as a sensible statistician should always be. Statistics always deals with data which represent the outside world. Even if the choice of a data distribution is subjective, the likelihood determined by data is an objective evaluation of the assumed data distribution. Even Shakespeare cannot fight against the objectivity of data.

The prudence shown by Professor Freeman against the numerical results reported in my paper is impressive. Particularly his preference of real examples to artificial ones reveals his position to consider statistics as something related with the outside world.

To the two questions raised by Professor Freeman I answer as follows: (i) The idea stressed by the examples discussed in the paper is the importance of the technical understandability of prior distributions. The examples also suggest the utility of defining an objective procedure of the choice of a prior distribution. Any subjectively chosen proper prior distribution, however wooly it may be, cannot be free from a possible gross misspecification. (ii) There should be no problem in choosing c_0 and D, if their technical meanings are clearly understood.

As to the predicament of Professor Freeman about the last example on polynomial fitting my explanation is that I am only interested in getting a good predictive distribution. There is no meaning in talking about the "true" order, as this is infinite.

Dr. Peña considers that the conceptual difficulties of the subjective approach is

not substantial. His conclusion is based on two observations. The first is that there are Bayesians, like Jeffreys, Box, Tiao and Zellner, who treat the problem of inference to Dr. Peña's satisfaction. But these people are not subjective Bayesians. They all accept the use of improper prior distributions, which is unacceptable to strictly subjective Bayesians. Dr. Peña's second observation is that my criticism of Savage's postulate of linear ordering of preference is already sufficienty disproved by Raiffa's "money-pump" argument. Raiffa's explanation starts by assuming that a person with incoherent preference has made a decision. What I am insisting with the example of the boy with the preference described in the text of my paper is that he is trapped in a state of indecision. Thus Raiffa's "money-pump" argument does not constitute any disproof of my criticism of the difficulty of Savage's axiom.

As to the criticism of Dr. Peña of the information criterion I must say that the classical tests are often disguised realizations of estimations when there are several possible models. The optimality of the minimum *AIC* procedure is discussed by Akaike (1978b) and Shibata (1980), but what I am interested in here is the use of the concept of likelihood or entropy in Bayesian modeling rather than the use of minimum *AIC* type procedure.

Dr. Peña's criticism of the use of the general linear model for seasonal adjustment suprises me. The whole procedure is objectively defined. It is simple and can be tested by anyone who is interested in it. The procedure is completely free from the *ad hoc* manipulations of data, at the beginning and end of the time series, by Census Methods of seasonal adjustment. I do not deny the possibility of other procedures, but I must mention that there is nothing like a canonical form for a system varying with time and that this makes the ordinary parametric approach to the seasonal adjustment problem very difficult. The main point of the introduction of the present general linear model is the clarification of the importance of technical understandability of a prior distribution. I hope that Dr. Peña would agree with me to consider the fact that a computer program is already in existence and is producing useful outputs without much human intervention as a clear demonstration of the power of this approach.

Professor Lindley considers my criticism of Savage's axiom to be due to the confusion of descriptive and prescriptive views. My criticism of subjective Bayesians is that their prescriptive attitude looks very much like the attitude of a physician who gives a huge collection of precriptions of drugs to a patient and leaves the burden of identifying the proper choice to the patient. The "money-pump" argument tells the patient that he must take a drug described by the physician but does not help him in making his choice.

Young's (1977) paper on polynomial fitting is not free from the basic difficulty. The prior distribution contains two hyperparameters. Apparently Young did not propose any systematic approach to the choice of the hyperparameters.

Dr. O'Hagan tells me that I am producing sensible result with the help of a misguided philosophy. In his 1976 paper, Dr. O'Hagan makes use of an improper prior distribution. The result mentioned in his comment is then obtained by adjusting the Bayesian model so as to produce a result consistent with the result obtained by conventional statistics. These observations show that he himself is subscribing to the "misguided" philosophy, the common-sense approach to statistics.

Professor Smith reminds me that the paper by Lindley and Smith (1972) is a pioneering work on the common-sense approach to Bayesian statistics. Actually Lindley and Smith accept the use of an improper prior distribution, which is not acceptable to strict Bayesians. The paper demonstrates the point that the technical understandability of the prior distribution is the key to the successful application of a Bayesian model. Certainly, this is one of the themes of my present paper, but my main emphasis is on the use of likelihood as an objective measure of the goodness of a model. Even the goodness of a Bayesian model can be checked by comparing the likelihoods of competing models.

Professor Geisser is sympathetic to my common-sense approach but he fears, with Professor Hill at the time of the meeting, that I am touching on a too delicate subject when I referred to the direct and inverse binomial experiments. It looks to me that he is too much influenced by the so-called objective theory of probability. Within the statistical context, it must be accepted, every probability is conditional on available information. If we knew whether the experiment was direct or inverse, this constitutes a part of our prior information. Thus the assumption of prior independence of the probability of head in a coin tossing with the information of the type of experiment is acceptable only under certain specific circumstances.

Consider the situation where you are served a piece of pie. When you know that the pie was prepared by a cook who is notorious for poisoning your attitude towards the pie will be different from that when you know that the cook had a perfect record. Dr. Peña drew my attention to Box and Tiao (1973, pp. 45-46) who accepted the difference of the ignorance priors for the two sampling schemes. Thus I am not alone here.

Professor Dickey points out that my emphasis on likelihood is wrong and reminds me of the importance of interpreting a prior distribution as an expression of a personal opinion. In Professor Dickey's argument I sense, as in almost every argument by subjectivist Bayesians, a rash inclination towards the assumption of the state of ignorance, or of no information. I consider this a dangerous sign. Particularly, when Professor Dickey forcefully puts forward the dichotomy between subjective probability and chaos, I see a curious analogy between his position and that of epistemological traditionalism observed by Popper (1965, p.6) who states 'we can interpret tradicionalism as the belief that, in the absence of an objective and discernible truth, we are faced with the choice between accepting the authority of tradition, and chaos'.

We notice that Professor Dickey's argument gains weight only when he uses the word "expert opinion" instead of an arbitrary "opinion". What discriminates an expert's opinion from a layman's is that the former is backed by experiences, either of the expert's own or someone else's. The experience are appreciated only when they constitute objective information. In constructing his prior distribution, the expert will evaluate, at least informally, the likelihoods of various conditional statements with respect to this information. It is the objectivity thus obtained that makes an expert's prior distribution respectable.

Now we come to the discussion of the state of ignorance. For a person who tries to collect information to establish a hard prior opinion, it is a rule rather than exception that he faces the lack of information which prevents him from determining a unique prior distribution. The impossibility of uniquely determining his prior distribution, typ-

ycally represented by the introduction of hyperparameters, is the representation of the lack of information and however hard he may try to elicit the details of his opinion he cannot produce information out of nothing. Nevertheless, due to the limitation of time, he has to make a decision, and this requires a unique choice of a prior distribution. How should he act in such a situation? The answer seems clear. The effort in defining a prior distribution is mainly directed towards delineating relatively important possibilities. When the effort comes to a halt due to the lack of information, we come to the phase of making a decision. The situation is typically represented by that of planning the experiment in Professor Dickey's comment. Here the emphasis is on paying attention to every possibility. Who will favor a physician's whim to a carefully designed experiment which takes into account every possible course of patient's condition? Thus, at the point where the collection of further relevant information becomes impossible, the emphasis is switched from restraining to dispersing the distribution of the prior probability. This may sometimes lead to the use of improper prior distributions. The effect of this dispersing is evaluated by its effect on the resulting predictive distribution. Here the recognition of the necessity of switching the point of view, during the process of developing a prior distribution, seems crucial.

The above is an amplified version of the procedure for the construction of a prior distribution discussed in my paper. A simple but concrete example of application of this procedure is discussed in Akaike (1980).

Thus in our approach the subjective elements are always exposed to some objective tests through the use of prior experiences or data, and the somewhat obscure concept "opinion", required to complete a prior distribution, is replaced by a description of a strategy for making a decision. This strategy and its design principle are described objectively and can be tested in the long run through the accumulation of experiences of its use by a scientific community. Thus, contrary to the suggestion of Professor Dickey, we do not put much emphasis on the interpretation of a prior distribution as an expression of "opinion".

DAWID, A.P. (*The City University, London*):

Should the Bayesian be interested in concepts springing from a frequentist or "prespecification" approach to inference, or can he afford to dismiss them cursorily as "incoherent"? Although I am fully committed to the Bayesian position, I can't accept that the only good ideas are those had by Bayesians. Consequently I regard it as practically important, as well as theoretically amusing, to investigate non-Bayesian ideas, and find out how they relate to Bayesian ones. So perhaps I should revoke my suggestion that some of the definitions of my paper are of no interest to Bayesians, for if we are to be good statisticians (which is surely more important than being coherent) we must not dismiss such concepts out of hand — at any rate, not before a thorough investigation of the type I have attempted.

While agreeing with Professor Barnard that we could well drop the term "nuisance parameter", I am puzzled by his suggestion that we are guilty of some sort of sin if we lay down, before getting data, that we are only interested in what we can learn about

the parameter Θ. If this is an example of "prespecification", I can only conclude that there must be a large overlap between that approach and Bayesian ideas. If we are faced with a decision problem in which the parameter enters the loss function only through Θ, why should we not make an inference about Θ alone, whatever the data may turn out to be, and whatever the dependence between Θ and some nuisance parameter Φ?

I do, however, accept Professor Geisser's point that the formulation of the problem in terms of parameters at all may be mistaken. A reformulation involving the unknown values of future observations would involve quite different theory, at least for the frequentist. Such an approach might perhaps be used in the model discrimination context mentioned by Professor Freeman, since, while we may not know what parameters are of interest, we will surely be able to pinpoint what it is that we should like to be able to predict. Nevertheless, for the Bayesian, an emphasis on prediction is not a pre-requisite for the problem of nuisance parameters to disappear - he never had a problem in the first place. It is classical ideas which present problems. If we take a predictive standpoint, then it becomes appropriate to compare the straightforward Bayesian approach to prediction with classical counterparts (see, for example, Section 6 of Dawid, 1979a). But that is another story.

Barnard's discussion of a "model-adjustment" parameter is important. It attacks the problem of the robustness of an inference about the parameter of interest. His assumption is that the likelihood, while not being a function of θ only, is nevertheless approximately so, for most data. If the data suggests that this approximation is good, then we can pretty well ignore the fact that there are really some nuisance parameters around. If, however, we have exceptional data, we may have to be more careful. This suggests an interesting line of research; in particular, how would the Bayesian formalize the property that, to a good approximation, his model involves only the parameter of interest? This kind of problem, in which approximations may be valid for some data values, but not for others, is of great general importance to a sensible Bayesian approach. In particular, the marginalization paradox does not rule against using the paradoxical posterior distribution for the data at hand, but warns that it cannot be a good approximation to a coherent posterior for *all* possible data values. In concerning ourselves with these things, we are, of course, leaving the pre-specification approach squarely behind, and rightly so.

The incoherence in Example (5.6) is not really concerned with the question whether or not we can have $\tau^2 < 0$, as suggested by Professor Freeman. As I point out, we could, for example, get variation-independence between σ^2 and (μ, σ_0^2) if σ^2 and τ^2 are subject to $\sigma^2 \leq 1$, $\tau^2 \geq (1-\sigma^2)/J$. The real difficulty is the dependence on J. Thus, while one might argue that it is coherent to use only S for inference about σ^2 for the experiment performed, one could not allow this same argument simultaneously for another such experiment, with different J. This is analogous to the discussion at the end of the previous paragraph, with the difference that, there, we had to worry about inferences from different data in one experiment, while here we must worry about different experiments. But the comparison of different experiments is a valid and important concern of the theory of coherence.

As for the concepts of "no available information about Θ in the absence of

ycally represented by the introduction of hyperparameters, is the representation of the lack of information and however hard he may try to elicit the details of his opinion he cannot produce information out of nothing. Nevertheless, due to the limitation of time, he has to make a decision, and this requires a unique choice of a prior distribution. How should he act in such a situation? The answer seems clear. The effort in defining a prior distribution is mainly directed towards delineating relatively important possibilities. When the effort comes to a halt due to the lack of information, we come to the phase of making a decision. The situation is typically represented by that of planning the experiment in Professor Dickey's comment. Here the emphasis is on paying attention to every possibility. Who will favor a physician's whim to a carefully designed experiment which takes into account every possible course of patient's condition? Thus, at the point where the collection of further relevant information becomes impossible, the emphasis is switched from restraining to dispersing the distribution of the prior probability. This may sometimes lead to the use of improper prior distributions. The effect of this dispersing is evaluated by its effect on the resulting predictive distribution. Here the recognition of the necessity of switching the point of view, during the process of developing a prior distribution, seems crucial.

The above is an amplified version of the procedure for the construction of a prior distribution discussed in my paper. A simple but concrete example of application of this procedure is discussed in Akaike (1980).

Thus in our approach the subjective elements are always exposed to some objective tests through the use of prior experiences or data, and the somewhat obscure concept "opinion", required to complete a prior distribution, is replaced by a description of a strategy for making a decision. This strategy and its design principle are described objectively and can be tested in the long run through the accumulation of experiences of its use by a scientific community. Thus, contrary to the suggestion of Professor Dickey, we do not put much emphasis on the interpretation of a prior distribution as an expression of "opinion".

DAWID, A.P. (*The City University, London*):

Should the Bayesian be interested in concepts springing from a frequentist or "prespecification" approach to inference, or can he afford to dismiss them cursorily as "incoherent"? Although I am fully committed to the Bayesian position, I can't accept that the only good ideas are those had by Bayesians. Consequently I regard it as practically important, as well as theoretically amusing, to investigate non-Bayesian ideas, and find out how they relate to Bayesian ones. So perhaps I should revoke my suggestion that some of the definitions of my paper are of no interest to Bayesians, for if we are to be good statisticians (which is surely more important than being coherent) we must not dismiss such concepts out of hand — at any rate, not before a thorough investigation of the type I have attempted.

While agreeing with Professor Barnard that we could well drop the term "nuisance parameter", I am puzzled by his suggestion that we are guilty of some sort of sin if we lay down, before getting data, that we are only interested in what we can learn about

the parameter Θ. If this is an example of "prespecification", I can only conclude that there must be a large overlap between that approach and Bayesian ideas. If we are faced with a decision problem in which the parameter enters the loss function only through Θ, why should we not make an inference about Θ alone, whatever the data may turn out to be, and whatever the dependence between Θ and some nuisance parameter Φ?

I do, however, accept Professor Geisser's point that the formulation of the problem in terms of parameters at all may be mistaken. A reformulation involving the unknown values of future observations would involve quite different theory, at least for the frequentist. Such an approach might perhaps be used in the model discrimination context mentioned by Professor Freeman, since, while we may not know what parameters are of interest, we will surely be able to pinpoint what it is that we should like to be able to predict. Nevertheless, for the Bayesian, an emphasis on prediction is not a pre-requisite for the problem of nuisance parameters to disappear - he never had a problem in the first place. It is classical ideas which present problems. If we take a predictive standpoint, then it becomes appropriate to compare the straightforward Bayesian approach to prediction with classical counterparts (see, for example, Section 6 of Dawid, 1979a). But that is another story.

Barnard's discussion of a "model-adjustment" parameter is important. It attacks the problem of the robustness of an inference about the parameter of interest. His assumption is that the likelihood, while not being a function of θ only, is nevertheless approximately so, for most data. If the data suggests that this approximation is good, then we can pretty well ignore the fact that there are really some nuisance parameters around. If, however, we have exceptional data, we may have to be more careful. This suggests an interesting line of research; in particular, how would the Bayesian formalize the property that, to a good approximation, his model involves only the parameter of interest? This kind of problem, in which approximations may be valid for some data values, but not for others, is of great general importance to a sensible Bayesian approach. In particular, the marginalization paradox does not rule against using the paradoxical posterior distribution for the data at hand, but warns that it cannot be a good approximation to a coherent posterior for *all* possible data values. In concerning ourselves with these things, we are, of course, leaving the pre-specification approach squarely behind, and rightly so.

The incoherence in Example (5.6) is not really concerned with the question whether or not we can have $\tau^2 < 0$, as suggested by Professor Freeman. As I point out, we could, for example, get variation-independence between σ^2 and (μ, σ_0^2) if σ^2 and τ^2 are subject to $\sigma^2 \leq 1$, $\tau^2 \geq (1-\sigma^2)/J$. The real difficulty is the dependence on J. Thus, while one might argue that it is coherent to use only S for inference about σ^2 for the experiment performed, one could not allow this same argument simultaneously for another such experiment, with different J. This is analogous to the discussion at the end of the previous paragraph, with the difference that, there, we had to worry about inferences from different data in one experiment, while here we must worry about different experiments. But the comparison of different experiments is a valid and important concern of the theory of coherence.

As for the concepts of "no available information about Θ in the absence of

knowledge of Φ", the various ideas of *S, G, M*-ancillarity etc., all express classical attempts to capture this notion: I refer Professor Freeman to Barndorff-Nielsen's book. I don't think there is a full-blooded Bayesian interpretation, because of the difficulty of defining "absence of knowledge". Marginal ancillarity is not really appropriate, depending as it does very much on the form of prior knowledge about Φ. But if we once again drop a pre-specification approach, it may be that the concept can be given some meaning in terms of robustness or approximation, relevant only for certain data and classes of prior distributions.

REFERENCES IN THE DISCUSSION

AKAIKE, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**, 716-722.

— (1976). Canonical correlation analysis of time series and the use of information criterion. In *System Identification: Advances and Case Studies.* (Mehra and Lainiotis eds.) New York: Academic Press.

— (1978). On the likelihood of a time series model. *The Statistician,* **27**, 217-235.

— (1979). Smoothness priors and the distributed lag estimator. *Tech. Report No. 40,* Stanford University.

BARTLETT, M.S. (1978). *An introduction to Stochastic Processes.* Cambridge: University Press.

BOX, G.E.P. and JENKINS, G.M. (1970). *Time Series Analysis, Forecasting and Control.* New York: Holden-Day.

BOX, G.E.P. and TIAO, G.C. (1973). *Bayesian Inference in Statistical Analysis.* New York: Addison Wesley.

BOX, G.E.P., HILLMER, G.C. and TIAO, G.C. (1976). Analysis and modelling of seasonal Time Series. Presented at *NBER/Bureau of the Census Conference on Seasonal Analysis of Economic Time Series.* Washington, D.C.

GOOD, I.J. (1979). Book review of *Logic, Law and Life: Some Philosophical Complications,* (R.G. Colodomy ed.) *J. Amer. Statist. Assoc.* **74**, 501-502.

KALBFLEISCH, J.D. and SPROTT, D.A. (1970). Application of likelihood method to models involving large numbers of parameters (with discussion). *J. Roy. Statist. Soc. B.* **32**, 175-208.

LINDLEY, D.V. and SMITH A.F.M. (1972). Bayes estimates for the linear model (with discussion). *J. Roy Statist. Soc. B.* **34**, 1-41.

NELDER, J.A. (1977). A reformulation of linear models (with discussion). *J. Roy. Statist. Soc. A* **140**, 48-77.

O'HAGAN, A. (1976). On posterior joint and marginal modes. *Biometrika* **63**, 329-333.

POPPER, K.R. (1965). *Conjectures and Refutations: The Growth of Scientific Knowledge.* New York: Basic Books. Also in Harper and Row, (1968) New York.

RAIFFA, H. (1968). *Decision Analysis.* New York: Addison-Wesley.

SHAKESPEARE, W. (1598). *The Second Part of the History of Henry the Fourth, I. 3,* 35-36. London: Wise and Aspley.

SHIBATA, R. (1976). Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika* **63**, 117-126.

— (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Ann. Statist.* **8**, 147-164.

YOUNG, A.S. (1977). A Bayesian approach to prediction using polynomials. *Biometrika* **64**, 309--317.

ZELLNER, A. (1977). Maximal data information prior distributions. In *New Developments in the Applications of Bayesian Methods.* (A. Aykas & C. Brumat eds.) Amsterdam: North-Holland.