BASU, D. (1964). Recovery of ancillary information. *Sankhyā Ser. A* **26** 3–16.
BERGER, J. and WOLPERT, R. (1985). *The Likelihood Principle.* IMS, Hayward, Calif.
BIRNBAUM, A. (1962). On the foundations of statistical inference (with discussion). *J. Amer. Statist. Assoc.* **57** 269–326.
BIRNBAUM, A. (1972). More on concepts of statistical evidence. *J. Amer. Statist. Assoc.* **67** 858–886.
BUEHLER, R. J. (1982). Some ancillary statistics and their properties (with discussion). *J. Amer. Statist. Assoc.* **77** 581–594.
COX, D. R. (1958). Some problems connected with statistical inference. *Ann. Math. Statist.* **29** 357–372.
EVANS, M., FRASER, D. A. S. and MONETTE, G. (1986). On principles and arguments to likelihood. *Canad. J. Statist.* **14** 181–194.
FISHER, R. A. (1961). Sampling the reference set. *Sankhyā Ser. A* **23** 3–8.
FRASER, D. A. S. (1968). *The Structure of Inference.* Wiley, New York.
FRASER, D. A. S. (1979). *Inference in Linear Models.* McGraw-Hill, New York.
FRASER, D. A. S. and MASSAM, H. (1985). Conical tests: Observed level of significance and confidence regions. *Statist. Hefte (N.F.)* **26** 1–18.
FRASER, D. A. S. and REID, N. (1987). Fibre analysis and conditional inference. *Proc. 2nd Pacific Area Statistical Conference* 241–248.
FRASER, D. A. S. and REID, N. (1988). On conditional inference for a real parameter: A differential approach on the sample space. *Biometrika* **75** 251–264.

Rejoinder

Robert E. Kass

I am very grateful to the discussants for their comments, which have substantially enriched the material presented here. The remarks of Professors Amari, Barndorff-Nielsen, and Reid and Fraser require no reply. I do, however, wish to answer the specific queries raised by Professors Bernardo and Rao.

With regard to Rao's query (1), concerning characterizations of the information metric, I would refer interested readers to the original work of Centsov (1972) and the newer work of Picard (1989). I am not sure what Rao has in mind in his query (2) about the choice of affine connection. Part of the answer may come from the results of Centsov and Picard, but if Professor Rao is referring to the choice of α in the α-connection, perhaps helpful to the intuition is the observation in Kass (1984) that vanishing of the α-connection coefficients when $\alpha = -1, -1/3, 0, 1$ occurs for the bias-reducing, skewness-reducing, variance-stabilizing, and natural parameterizations, respectively, and when $\alpha = 1/3$, it occurs for the parameterization in which the expected values of the third derivatives of the loglikelihood vanish. These parameterizations were characterized in differential equation form, in the one-parameter case, by Hougaard (1982). There is also a very nice answer to part of Rao's query (3), due to Amari (1985, 1987a). In brief, Amari used higher derivatives of the imbedding $\eta(\cdot)$, defining a curved exponential family, to define both higher-order curvatures and appropriate statistics based on higher-order derivatives of the loglikelihood function. With these he obtained a complete decomposition of the information in the sample as an asymptotic expansion with geometrically-interpretable terms of decreasing order associated with the loglikelihood derivatives.

Finally, Professor Rao's point (4), and Professor Bernardo's request for comments in the rejoinder, concern Jeffreys' general rule for choosing a prior. I have a few things to say about this, though for the sake of brevity I will not try to argue my opinions in detail.

As a preliminary remark, I emphasize that by "reference prior" I mean a prior chosen according to any formal rule that may be applied without detailed consideration of the data-analytic context. Such a prior need not be considered "noninformative" in any well-defined sense. This is an important point, since it is dubious that the concepts of ignorance and lack of information can be given satisfactory definitions. I believe the idea of selecting a prior by *convention*, as a "standard of reference," analogous to choosing a standard of reference in other scientific settings, is due to Jeffreys (1955) page 277. This notion and terminology was adopted by Box and Tiao (1973) page 23. Unfortunately, Bernardo (1979) used the term "reference prior" for a specific rule, rather than the general concept, and this occasionally causes confusion.

There is great convenience in conventional choices, throughout statistics and throughout science. But convenience should not be confused with necessity: one might say that conventions are useful as long as they are not taken too seriously. Thus, I see the convenience in reference priors, just as I recognize the convenience in conventional levels of significance. In applications, however, such conveniences must be questioned. Sometimes they are justifiable timesavers, especially for communicating results, but often they are not. I consider reference priors to be "default" choices, but they are to be used only when their convenience outweighs any adverse effect on inferences they cause by misrepresenting knowledge.

From this perspective, the often-cited difficulties with Jeffreys' general rule mentioned by Bernardo and Rao (which are, I think, inevitable with any rule that is based only on the form of the sampling distribution), do not seem too worrisome. Design-dependence serves as an important reminder that the ability to represent widely varying states of knowledge is being traded away for ease of analysis and communication. There are many roads that lead to Jeffreys' general rule (see Jeffreys, 1946; 1961, pages 180–181, 192; Perks, 1947; Box and Tiao, 1973, pages 36–42; Bernardo, 1979; Kass, 1988). Though all of these are variants on a single asymptotic theme, in the very small sample case it is of little use to base inferences solely on a reference prior, without explicit sensitivity analysis, so large-sample motivation should not be disturbing. My view is that Jeffreys' general rule, with its modification to accommodate group-transformation structure, is the default among the defaults. I should perhaps say that I do not find myself using Jeffreys' general rule very often, but I continue to appreciate its availability in a pinch.

Bernardo has asked for a reply regarding the problem of determining a reference prior that would be appropriate for making inferences about a parameter of interest in the presence of nuisance parameters. I do not wish to stray far from the subject of my paper, but perhaps I can use this issue to elaborate a bit on my view of the inferential consequences of parameter-transformation invariance, which is so basic to the geometrical approach.

Suppose we have a parametric family with Θ being m-dimensional. It seems to me that "quantities of interest" are best interpreted as real-valued functions $\psi = g(\theta)$. These have a specific meaning regardless of the parameterization used, in the sense that they are automatically defined on any parameter space, e.g., $g(\gamma) = g(\theta(\gamma))$ (or they may be defined directly on the manifold of densities). There is no necessary relationship between the number of quantities of interest ψ and the dimension m of the model. It is nice to have available a reference prior for the vector θ because the inferences about θ entail inferences about *all possible* quantities of interest, and often the quantities considered to be of interest are, in fact, rather arbitrary. Invariance of such a reference procedure with respect to transformation of the parameter θ remains crucial.

On the other hand, sharply-defined inference problems may involve one or many quantities of interest that have been found to be interpretable. In such cases, one uses scales one likes to think about, which are determined by taste and context, and one usually has some idea of the a priori probable values of these quantities of interest. Thus, in my opinion, it would be rare that one would need to resort to an additional structural rule that would assign a marginal prior to ψ. The bigger worry involves the nuisance parameters, but use of a joint reference prior on θ (possibly modified to produce a specified marginal distribution on ψ) seems to me to continue to be a reasonable default in this often-difficult problem. Nonetheless, I will look forward to seeing the work of Bernardo and Berger, with the hope that it may at least be helpful in those situations I consider rare.

Again, I thank all the discussants for their contributions. I am sure their comments will greatly benefit readers of the paper.

ADDITIONAL REFERENCES

BERNARDO, J. M. (1979). Reference posterior distributions for Bayesian inference (with discussion). *J. Roy. Statist. Soc. Ser. B* **41** 113–147.

BOX, G. E. P. and TIAO, G. C. (1973). *Bayesian Inference in Statistical Analysis*. Addison-Wesley, Reading, Mass.

HOUGAARD, P. (1982). Parameterizations of non-linear models. *J. Roy. Statist. Soc. Ser. B* **44** 244–252.

PICARD, D. B. (1989). Statistical morphisms. Unpublished manuscript.