BARNDORFF-NIELSEN, O. E., BLÆSILD, P. and MORA, M. (1989). Generalized higher order differentiation. *Acta Appl. Math.* To appear.
BARNDORFF-NIELSEN, O. E. and COX, D. R. (1989). *Asymptotic Techniques for Use in Statistics*. Chapman and Hall, London.
BARNDORFF-NIELSEN, O. E. and JUPP, P. E. (1989). Approximating exponential models: *Ann. Inst. Statist. Math.* To appear.
BARTLETT, M. S. (1953). Approximate confidence intervals. *Biometrika* **40** 12–19.
BLÆSILD, P. (1987a). Yokes: Elemental properties with statistical applications. In *Geometrization of Statistical Theory* (C. T. J. Dodson, ed.) 193–198. ULDM Publications, Lancaster.
BLÆSILD, P. (1987b). Further analogies between expected and observed geometries of statistical models. Research Report 160, Dept. Theoretical Statistics, Aarhus Univ.
BLÆSILD, P. (1988). Yokes and tensors derived from yokes. Research Report 173, Dept. Theoretical Statistics, Aarhus Univ.
BLÆSILD, P. and MORA, M. (1988). Differential strings: Basic theory. Research Report 183, Dept. Theoretical Statistics, Aarhus Univ.
FOSTER, B. L. (1958). Differentiation on manifolds without a connection. *Michigan Math. J.* **5** 183–190.
FOSTER, B. L. (1961). Some remarks on tensor differentiation. *Ann. Mat. Pura Appl.* **54** 143–146.
FOSTER, B. L. (1986). Would Leibniz lie to you? (Three aspects of the affine connection). *Math. Intelligencer* **8** 34–40, 57.
FOSTER, B. L. (1988). Higher derivatives in geometry and physics. Research Report 178, Dept. Theoretical Statistics, Aarhus Univ. *Proc. Roy. Soc. London Ser. A.* To appear.
MCCULLAGH, P. (1987). *Tensor Methods in Statistics.* Chapman and Hall, London.
MORA, M. (1988). Geometrical expansions for the distributions of the score vector and the maximum likelihood estimator. Research Report 172, Dept. Theoretical Statistics, Aarhus Univ.
SKOVGAARD, I. M. (1986). A note on the differentiation of cumulants of log likelihood derivatives. *Internat. Statist. Rev.* **54** 29–32.

Comment: On Multivariate Jeffreys' Priors

José M. Bernardo

Kass presents a lucid, well written description of the differential geometric foundations of such pervasive concepts in statistics as Fisher information, the Kullback-Leibler metric and information numbers, or the loglinear structure of exponential families. As the author points out, these topics directly relate to the role of reference priors in Bayesian Inference—an issue he regards as of "ongoing vital importance"—and one would expect a deeper understanding of such an issue from his wòrk. I will concentrate on this point.

JEFFREYS' PRIORS

Kass very clearly describes some of the more basic aspects of Jeffreys' priors. Specifically, I would like to draw your attention to four of those:

(i) Jeffreys' general rule is generated by the *natural* volume element of the information metric.

(ii) The main intuitive motivation for Jeffreys' priors is *not* their invariance, which is certainly a necessary, but in general far from sufficient, condition to determine a sensible reference prior; what makes Jeffreys' priors unique is that they are *uniform* measures in a particular metric which may be defended as the "natural" choice for statistical inference.

(iii) The existence of Jeffreys' priors requires rather strong, if fortunately frequent, regularity conditions.

(iv) Multivariate Jeffreys' priors are often inadequate to obtain marginal reference posterior distributions for its elements—as Jeffreys himself realized—and there does not seem to be an agreed *systematic* alternative; independent treatment of orthogonal parameters, when applicable, is only an *ad hoc* partial solution. Key references for the type of problems which may be encountered from routine use of Jeffreys' multivariate priors are Stein (1959) or Dawid, Stone and Zidek (1973).

While (i) and (ii) are possibly sufficient to be suspicious about any method for generating reference priors which does not reduce to Jeffreys' in *one-dimensional regular* problems, (iii) leaves room for improvement and (iv) clearly requires new work. When reading Kass' paper, I was hoping for some new hints about (iv) but I could not recognize any; I hope to see some comments in the rejoinder.

REFERENCE PRIORS

In my development of *reference* priors (Bernardo, 1979)—which reduce to Jeffreys' for one-dimensional regular problems—I explicitly recognized the importance of identifying *parameters of interest* and

Professor José M. Bernardo's mailing address is: Presidencia de la Generalitat, Departamento de Estadística, Palacio de Fuentehermosa, Caballeros 9, E-46001, Valencia, Spain.

nuisance parameters and tailoring the reference prior to this choice in order to avoid the problems mentioned above. The suggestion was to use a two-step reference prior. First find the conditional reference prior for the nuisance parameters given the parameters of interest, then find the reference prior for the parameter of interest in the marginal model formed by integrating out the nuisance parameters.

Two limitations of the method have however been observed. First, the conditional reference prior found in the first step is often improper, and yet it is subsequently used to form the marginal model for the parameter of interest; rigorously justifying this step requires a limiting operation on proper versions of the problem; Berger and Bernardo (1989) described and illustrated this approach. Second, it has been observed that merely grouping the parameters of a model into "parameters of interest" and "nuisance parameters" may not go far enough; recent work led Jim Berger and me to recommend providing a complete ordering of all parameters of a model, so that the reference prior is determined through a series of one-dimensional conditional steps. The ideas are well illustrated with the trinomial model that Kass uses throughout.

THE TRINOMIAL MODEL

Suppose that we are interested in the proportion θ_1 of individuals in the population that suffer a particular disease d_1, and have a random sample of n people classified into three categories; say we have y_1 individuals with disease d_1, y_2 with disease d_2 and $n - y_1 - y_2$ that are healthy. Thus, we have the trinomial model

$$p(y \mid \theta_1, \theta_2, n) = \frac{n!}{y_1!\, y_2!(n - y_1 - y_2)!}\, \theta_1^{y_1}\theta_2^{y_2}(1 - \theta_1 - \theta_2)^{n-y_1-y_2},$$

and θ_1 is our parameter of interest. Jeffreys' prior $\pi_J(\theta_1, \theta_2)$ is

$$\pi_J(\theta_1, \theta_2) = (2\pi)^{-1}\theta_1^{-1/2}\theta_2^{-1/2}(1 - \theta_1 - \theta_2)^{-1/2},$$

whose shape is that of Figure 1.

The corresponding marginal posterior distribution, given $\{y_1, y_2, n\}$, is the Beta distribution $Be(\theta_1 \mid y_1 + \frac{1}{2}, n - y_1 + 1)$; in particular, $E_J[\theta_1 \mid y_1, n] = (y_1 + \frac{1}{2})/(n + 3/2)$.

On the other hand, the *reference* prior $\pi_R(\theta_1, \theta_2)$ for the ordered sequence (θ_1, θ_2) is

$$\pi_R(\theta_1, \theta_2) = \pi(\theta_1)\pi(\theta_2 \mid \theta_1) = (\pi^{-2})\theta_1^{-1/2}(1 - \theta_1)^{-1/2}\theta_2^{-1/2}(1 - \theta_1 - \theta_2)^{-1/2},$$

whose shape is that of Figure 2, showing a higher concentration of probability mass than Jeffreys' around the point $(\theta_1, \theta_2) = (1, 0)$.

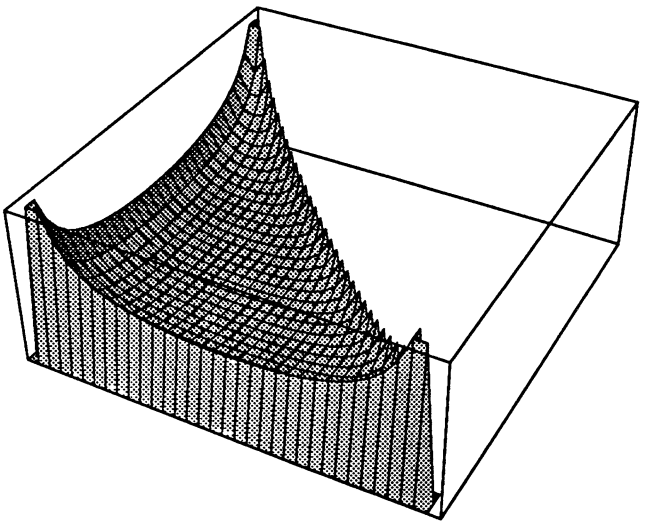

FIG. 1. *Multivariate Jeffreys' prior for the Trinomial model.*

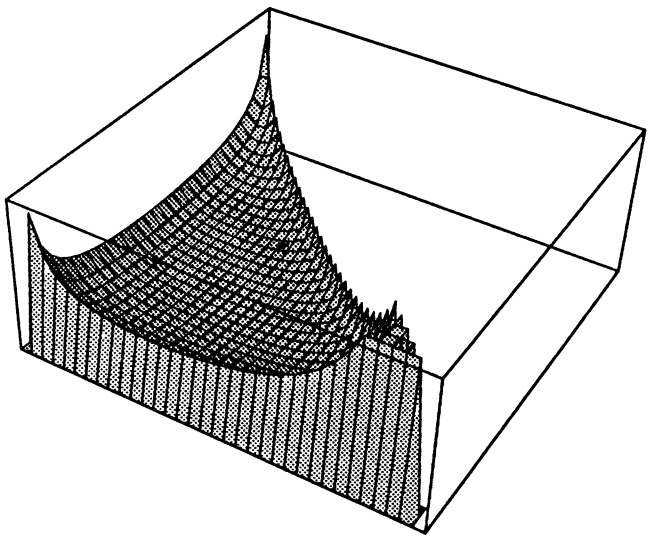

FIG. 2. *Reference prior for the sequence $\{\theta_1, \theta_2\}$ in the Trinomial model.*

The corresponding marginal posterior distribution of θ_1 is $Be(\theta_1 \mid y_1 + \frac{1}{2}, n - y_1 + \frac{1}{2})$ and $E_R[\theta_1 \mid y_1, n] = (y_1 + \frac{1}{2})/(n + 1)$

Suppose now that one realizes that the second disease d_2 has been subclassified into, say, k diseases and it is desired to use the whole data. If one used Jeffreys' multivariate prior for the corresponding multinomial model with $k + 2$ cells, the resulting marginal posterior distribution for θ_1 would be $Be(\theta_1 \mid y_1 + \frac{1}{2}, n - y_1 + (k + 2)\frac{1}{2})$ and, hence, $E_J[\theta_1 \mid y_1, n] = (y_1 + \frac{1}{2})/(n + (k + 3)\frac{1}{2})$. For large k this may be made arbitrarily small; thus, our conclusions on θ_1, the prevalence of disease d_1, would dramatically depend on the number of alternative diseases that one chooses to consider. This does not seem reasonable to me.

The reference prior is essentially immune to this

difficulty. Indeed the reference prior for the ordered sequence $(\theta_1, \cdots, \theta_m)$ is (see Berger and Bernardo, 1989 for details)

$$\pi_R(\theta_1, \cdots, \theta_m) = \pi(\theta_1)\pi(\theta_2 \mid \theta_1) \cdots \pi(\theta_m \mid \theta_1, \theta_2, \cdots, \theta_{m-1}) = (\pi^{-m}) \prod_{i=1}^{m} [\theta_i^{-1/2}(1 - \textstyle\sum_{j=1}^{i} \theta_j)^{-1/2}],$$

and the corresponding marginal reference distribution for θ_1 is

$$\pi_R(\theta_1 \mid y_1, \cdots, y_m, n) = \pi_R(\theta_1 \mid y_1, n) = Be(\theta_1 \mid y_1 + \tfrac{1}{2}, n - y_1 + \tfrac{1}{2}),$$

no matter how many cells are considered.

ADDITIONAL REFERENCES

BERGER, J. O. and BERNARDO, J. M. (1989). Estimating a product of means: Bayesian analysis with reference priors. *J. Amer. Statist. Assoc.* **84** 200–207.

BERGER, J. O. and BERNARDO, J. M. (1989). Ordered group reference priors with applications to multinomial and variance components problems. Technical Report 01/89, Dept. Estadística, Presidencia de la Generalidad Valenciana, Spain.

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

DAWID, A. P., STONE, M. and ZIDEK, J. V. (1973). Marginalization paradox in Bayesian and structural inference (with discussion). *J. Roy. Statist. Soc. Ser. B* **35** 189–223.

STEIN, C. M. (1959). An example of wide discrepancy between fiducial and confidence intervals. *Ann. Math. Statist.* **30** 877–880.

Comment

C. R. Rao

Geometric ideas do help in suggesting intuitive solutions to some complex problems and also in obtaining explicit solutions to specific problems through geometric methods. In his paper, "The geometry of asymptotic inference," Dr. Kass has demonstrated these two aspects by providing us with an excellent review of the past work and presenting some new ideas on the use of differential geometry in interpreting and developing statistical methodology. As Dr. Kass observed, differential geometry is a branch of mathematics "which is largely unfamiliar to most statisticians and may seem rather technical." I hope his paper will create some interest and encourage research in the differential geometric approach to statistical problems. However, I am tempted to share the caution expressed by Dr. D. J. Finney, in a similar situation, referring to some recent papers in multivariate analysis: "Amongst the many papers on statistical science published today, some appear to find outlets to mathematical theory without materially assisting scientific research." One may not fully subscribe to Dr. Finney's view, but the message is clear that enrichment of statistical methodology can take place only if its development is motivated by practical problems that are formulated in statistical terms. In this process, sophisticated mathematics could be used. I hope and believe as Dr. Kass does, that although "no claim can be made as yet that differential geometric research has made inroads into a large class of problems that is otherwise unreachable, the methods are so powerful, and the connections with statistics so plausible, that some further developments, of great methodological importance, might well occur."

In introducing differential geometric methods in statistics, I was motivated by the problem of discrimination between "populations" or "probability distributions " (p.d.'s), which naturally led to the need to introduce a metric in the space of p.d.'s. With a distance defined between two p.d.'s, it is possible to study the configuration of a given set of p.d.'s in terms of clusters and their hierarchical relationships.

In the case of a parametric family of p.d.'s characterized by a set of densities $\{f(x, \boldsymbol{\theta}): \boldsymbol{\theta} \in \Theta\}$, the metric was introduced by furnishing the parameter space Θ with a Riemannian quadratic differential metric (QDM)

$$\sum g_{ij} d\theta_i d\theta_j \tag{1}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots)'$, and (g_{ij}) is the Fisher information matrix (see Rao, 1945).

Using the QDM, one can compute the geodesic distance between any two p.d.'s represented by any two parameters θ and ϕ, which we denote by $D_g(\theta, \phi)$.

C. R. Rao is Eberly Professor of Statistics at Pennsylvania State University and Adjunct Professor at the University of Pittsburgh. His mailing address is: Department of Statistics, Pond Laboratory, Pennsylvania State University, University Park, Pennsylvania 16802.