becomes a key tool for connecting information theory and statistics.

Linear Programming Problem

It is interesting that the dual geometry is useful for some other problems. When a convex function $\Psi(\theta)$ is defined, we have a Legendre transformation from θ to η with a dual convex function $\Phi(\eta)$. We can introduce a dually flat geometry when it is equipped with a pair of convex functions. In the case of statistics, Φ is the negative of the entropy function and Ψ is the cumulant generating function. We have natural convex functions derived from linear and non-linear programming problems.

It is interesting to point out that a continuous version of the Karmarkar inner method is just to proceed along an m-geodesic in the space thus equipped with the dual connections. This method can easily be generalized to a nonlinear programming problem. This shows a wide applicability and universality of dual geometry.

ADDITIONAL REFERENCES

AMARI, S.-I. (1983). Comparisons of asymptotically efficient tests in terms of geometry of statistical structures. *Bull. Internat. Statist. Inst., Proc. 44th Session, Book 2* 1190–1206.

AMARI, S.-I. (1987b). Differential geometry of a parametric family of invertible linear systems—Riemannian metric, dual affine connections and divergence. *Math. Systems Theory* **20** 53–82.

AMARI, S.-I. (1987c). Dual connections on the Hilbert bundles of statistical models. In *Geometrization of Statistical Theory* (C. T. J. Dodson, ed.) 123–152. ULDM Publications, Lancaster.

AMARI, S.-I. (1989). Fisher information under restriction of Shannon information. *Ann. Inst. Statist. Math.* To appear.

AMARI, S.-I. and HAN, T. S. (1989). Statistical inference under multiterminal rate restrictions—a differential geometrical approach. *IEEE Trans. Inform. Theory.* To appear.

AMARI, S.-I. and KUMON, M. (1988). Estimation in the presence of infinitely many nuisance parameters—geometry of estimating functions. *Ann. Statist* **16** 1044–1068.

BARNDORFF-NIELSEN, O. E. (1988). *Parametric Statistical Models and Likelihood. Lecture Notes in Statist.* **50.** Springer, New York.

BARNDORFF-NIELSEN, O. E. and JUPP, P. E. (1989). Approximating exponential models. *Ann. Inst. Statist. Math.* To appear.

CAMPBELL, L. L. (1985). The relation between information theory and the differential geometric approach to statistics. *Inform. Sci.* **35** 199–210.

DODSON, C. T. J., ed. (1987). *Geometrization of Statistical Theory.* ULDM Publications, Lancaster.

KUROSE, T. (1988). Dual connections and affine geometry. Technical Report. UTYO-MATH 88-26.

LAURITZEN, S. L. (1987b). Conjugate connections in statistical theory. In *Geometrization of Statistical Theory* (C. T. J. Dodson, ed.) 33–51. ULDM Publications, Lancaster.

MITCHELL, A. F. S. (1989). The information matrix, skewness tensors and a-connections for the general multivariate elliptic distribution. *Ann. Inst. Statist. Math.* To appear.

NOMIZU, K. and PINKALL, U. (1986). On the geometry of affine immersions. Technical Report MPI 86-28, Max Planck Inst.

OKAMOTO, I., AMARI, S.-I. and TAKEUCHI, K. (1989). Asymptotic theory of sequential estimation procedures for curved exponential families. Unpublished manuscript.

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

Comment

O. E. Barndorff-Nielsen

Dr. Kass' fine account calls for little comment in itself. However, as he himself stresses, it leaves out parts of the subject, particularly of the more advanced aspects, and it may be useful here to outline briefly some of these parts so as to provide the interested reader with a fuller, though still far from comprehensive, picture of the scene. The discussion below relates mainly to work with which I have been to some degree associated, and it gives, in particular, virtually no impression of the important and extensive work of S.-I. Amari and his collaborators.

As will be indicated, the statistical problems have led to various developments and questions of a purely mathematical nature, and there are also interesting relations to theoretical physics.

INDEX NOTATION

The index notation of classical differential geometry and certain extensions thereof have turned out to be highly useful for many calculations in statistics, including some that are not of differential geometric nature (cf. McCullagh, 1987; Barndorff-Nielsen and Blæsild, 1988b; Barndorff-Nielsen and Cox, 1989, Chapter 5). The index notation makes many multivariate calculations just as easy as the corresponding

O. E. Barndorff-Nielsen is Professor in the Department of Theoretical Statistics at Aarhus University. His mailing address is: Matematisk Institut, Aarhus University, Ny Munkegade, DK-8000 Aarhus C, Denmark.

one-dimensional ones, and often throws light on the nature of the latter. The few simple rules of this type of notation can be quickly assimilated without studying differential geometry.

To illustrate this, and for use in the sequel, we let $r, s, t, \cdots$ and $a, b, c, \cdots$ denote indices, each varying over $1, 2, \cdots, d$, and we introduce the notation

$$\sum_{R_n/\nu} F(R_{n1}, \cdots, R_{n\nu}). \tag{1}$$

Here ν and n are fixed natural numbers with $\nu \le n$, $R_n = r_1 \cdots r_n$ is a fixed set of indices, and (1) indicates the sum over all partitions of R_n, into ν blocks $R_{n1}, \cdots, R_{n\nu}$, of the values of a numerical function F defined on such partitions.

Now, let $\omega = (\omega^1, \cdots, \omega^d)$ and $\psi = (\psi^1, \cdots, \psi^d)$ be alternative sets of coordinates on a d-dimensional differentiable manifold M. Then ω^r, ω^s, ω^t, $\cdots$ and ψ^a, ψ^b, ψ^c, $\cdots$ indicate generic coordinates. Applying (1), let

$$\omega^{R_n}_{/A_n} = \sum_{A_n/m} \omega^{r_1}_{/A_{n1}} \cdots \omega^{r_m}_{/A_{nm}}, \tag{2}$$

a sum of products of first- or higher-order derivatives of the coordinates ω^r with respect to the coordinates ψ^a. Generally we use / to indicate differentiation. Thus if f is a real function on M its first- and higher-order derivatives are denoted by $f_{/r}, f_{/rs}, f_{/rst}, \cdots$. In particular, $\omega^r_{/A_n} = \partial^n \omega^r / \partial a_1 \cdots \partial a_n$. Note that as a special case of (2) we have

$$\omega^{R_n}_{/A_n} = \omega^{r_1}_{/a_1} \cdots \omega^{r_n}_{/a_n}. \tag{3}$$

Now suppose that g is a real-valued function on M and f is a real-valued function on R; denote the νth order derivative of f by $f^{(\nu)}$ and write $y = g(\omega)$. Then the formulas for the first- and higher-order derivatives of the composite function $f \circ g$ in terms of those of f and g may be expressed compactly as

$$(f \circ g)_{/R_n}(\omega) = \sum_{\nu=1}^{n} f^{(\nu)}(y) \sum_{R_n/\nu} g_{/R_{n1}}(\omega) \cdots g_{/R_{n\nu}}(\omega). \tag{4}$$

This is a multivariate version of Faá di Bruno's formula.

The formulas for expressing multivariate moments in terms of multivariate cumulants and vice versa follow immediately as special cases, by taking $f(y)$ equal to exp y or log y and, correspondingly, g equal to the cumulant generating function or the moment generating function. For details, see Barndorff-Nielsen and Cox (1989), Chapter 5.

YOKES

Let Ω denote the domain of variation of ω, which may be thought of as a representation of the differentiable manifold M, and let Ω' be a copy of Ω. We shall consider functions g defined on $\Omega \times \Omega'$, and we let $g_r(\omega; \omega') = \partial g(\omega; \omega')/\partial\omega^r$, $g_{;r}(\omega; \omega') = \partial g(\omega, \omega')/\partial\omega'^r$ and generally

$$g_{R_m;S_n} = \frac{\partial}{\partial\omega^{r_1}} \cdots \frac{\partial}{\partial\omega^{r_m}} \cdots \frac{\partial}{\partial\omega'^{s_1}} \cdots \frac{\partial}{\partial\omega'^{s_n}} g(\omega; \omega'). \tag{5}$$

Furthermore, we use the notation

$$\not{g}_{R_m;S_n} = g_{R_m;S_n}(\omega; \omega), \tag{6}$$

i.e., a slash across a function on $\Omega \times \Omega'$ indicates the restriction of that function to the diagonal of $\Omega \times \Omega'$.

A *yoke* is a function $g(\omega; \omega')$, where $\omega \in \Omega$, $\omega' \in \Omega'$, having the properties that for all ω

(i) $g_r(\omega; \omega') = 0 \quad r = 1, \cdots, d$ if $\omega = \omega'$,
(ii) the matrix $[g_{rs}(\omega; \omega)]$ is nonsingular.

In the context of parametric statistics, the two most important yokes are the *observed likelihood yoke*

$$g(\omega; \omega') = l(\omega; \omega', a) - l(\omega'; \omega', a) \tag{7}$$

and the *expected likelihood yoke*

$$g(\omega; \omega') = E_{\omega'}\{l(\omega) - l(\omega')\}. \tag{8}$$

The first of these is the normed log likelihood $\bar{l}$, except that we have substituted the free variable ω' for $\hat{\omega}$, a being an ancillary statistic (for more explanation, see Barndorff-Nielsen, 1986a or 1988) and the latter is equal to $-I(\omega', \omega)$ where I indicates the Kullback–Leibler information.

Using the convention (6) we may reexpress (i) as the identity

$$\not{g}_r = 0 \tag{9}$$

which on repeated differentiation yields the sequence of relations

$$\not{g}_{rs} + \not{g}_{r;s} = 0, \tag{10}$$

$$\not{g}_{rst} + \not{g}_{rs;t} + \not{g}_{rt;s} + \not{g}_{r;st} = 0, \tag{11}$$

and so on, the general form being

$$\not{g}_{R_n} + \sum_{R_n/2} \not{g}_{R_{n1};R_{n2}} = 0, \tag{12}$$

where we have used the convention (1) with $\nu = 2$. Note in particular, from (10), that $\not{g}_{r;s}$ is a symmetric matrix although $g(\omega; \omega')$ is not assumed to be symmetric in ω and ω'.

It is also of interest to observe that on introducing the *normed yoke*

$$\bar{g}(\omega; \omega') = g(\omega; \omega') - g(\omega'; \omega'), \tag{13}$$

which is again a yoke, and defining $h(\omega; \omega')$ by

$$h(\omega; \omega') = \bar{g}(\omega'; \omega) \tag{14}$$

we have that h is also a yoke, constituting a kind of dual to g, and that

$$\not{h}_{r;s} = \not{g}_{r;s}. \tag{15}$$

Note that both of the likelihood yokes (7) and (8) are normed.

In the case of the expected likelihood yoke, the relation (12) may be rewritten as

$$\sum_{\nu=1}^{n} \sum_{R_n/\nu} E\{l_{R_{n1}} \cdots l_{R_{n\nu}}\} = 0, \tag{16}$$

which is the general form of the relations between the joint moments of log likelihood derivatives first indicated by Bartlett (1953). Incidentally, the joint cumulants of these derivatives satisfy the exactly analogous equations, i.e.,

$$\sum_{\nu=1}^{n} \sum_{R_n/\nu} K\{l_{R_{n1}}, \cdots, l_{R_{n\nu}}\} = 0, \tag{17}$$

as follows simply from an observation made by Skovgaard (1986).

Any yoke g induces a collection of geometrical objects on Ω, including a Riemannian (pseudo-) metric given by $\not{g}_{r;s}$ and a family of connections $\{\overset{\alpha}{\Gamma}: \alpha \in R\}$ given by

$$\overset{\alpha}{\Gamma}_{rst} = \frac{1+\alpha}{2} \not{g}_{rs;t} + \frac{1-\alpha}{2} \not{g}_{t;rs}. \tag{18}$$

Among the further objects derived from g are the two *connection strings*

$$\not{g}^{;r}_{t_1 \cdots t_n} = \not{g}^{s;r} \not{g}_{t_1 \cdots t_n;s}, \quad n = 1, 2, \cdots, \tag{19}$$

and

$$\not{g}^{r}_{;t_1 \cdots t_n} = \not{g}^{r;s} \not{g}_{s;t_1 \cdots t_n}, \quad n = 1, 2, \cdots, \tag{20}$$

where the Einstein summation convention is used. These are special cases of the concept of derivative strings discussed briefly below.

One of the applications of $\not{g}^{;r}_{t_1 \cdots t_n}$ and $\not{g}^{r}_{;t_1 \cdots t_n}$ is to the construction of tensors for use in asymptotic statistical analysis (cf. Barndorff-Nielsen, 1986b).

For further discussions of observed and expected likelihood geometries and of yokes and their statistical relevance, see Barndorff-Nielsen (1986a, 1987c, 1988), Blæsild (1987a, b, 1988), and below.

INVARIANT CALCULATIONS

It is generally desirable that the methods of parametric statistics should yield conclusions that are independent of the parametrization adopted, and this in itself makes differential geometry a natural tool for statistics, in line with its relevance for physics. (For some discussion of this point see Barndorff-Nielsen, 1988, Section 1.3.) While many statistical procedures are parametrization invariant in this sense, the way in which they are mathematically derived is very often not, and there is a need for developing an invariant calculus for statistical inference.

Such a calculus will, in particular, be helpful for the interpretation of the terms occurring in asymptotic expansions.

Most derivations in asymptotic statistical inference involve the use of Taylor's formula. However, Taylor's formula is not invariant under changes of coordinates, in the sense that for any n the actual value of an nth order Taylor approximation to a function f depends on which coordinate system is used on the domain of definition of f.

Let f be a scalar on M and, under the coordinate system ω, define the *symmetric covariant derivatives* $f_{//S_n}$ of f by the system of equations

$$f_{/R_n} = \sum_{\nu=1}^{n} f_{//S_\nu} \Gamma^{S_\nu}_{R_n} \tag{21}$$

which involves the Einstein summation convention and where (cf. (20))

$$\Gamma^{S_\nu}_{R_n} = \sum_{R_n/\nu} \not{g}^{s_1}_{;R_{n1}} \cdots \not{g}^{s_\nu}_{;R_{n\nu}}. \tag{22}$$

As another possibility, leading to a different kind of symmetric covariant derivatives $f_{//S_n}$, we might take

$$\Gamma^{S_\nu}_{R_n} = \sum_{R_n/\nu} \not{g}^{;s_1}_{R_{n1}} \cdots \not{g}^{;s_\nu}_{R_{n\nu}} \tag{23}$$

(cf. (19)). Then the multiarrays $f_{//S_n}$ are tensors, i.e., they obey the transform law

$$f_{//A_n} = f_{//S_n} \omega^{S_n}_{/A_n} \tag{24}$$

(cf. (3)).

An alternative method for deriving tensors from the ordinary derivatives $f_{/S_n}$ is, of course, by means of covariant differentiation using a connection, for instance one of the connections $\overset{\alpha}{\Gamma}$ defined above. However, the tensorial derivatives of f thus determined are not symmetric, i.e., invariant under permutation of the indices, and this appears unnatural in many, and in particular in statistical, contexts.

It is, incidentally, possible to develop a theory of generalized higher-order covariant differentiation that includes both of the above types as well as mixed forms of the two (Barndorff-Nielsen, Blæsild and Mora, 1988, 1989; see also Blæsild and Mora, 1988).

Note that

$$f_{//r} = f_{/r}$$

while, if we indicate covariant differentiation with respect to the connection $\overset{1}{\Gamma}{}^{t}_{rs} = \not{g}^{t}_{rs}$ by $///$,

$$f_{//rs} = f_{r///s}.$$

Furthermore, the symmetric covariant derivatives obey a Leibnitz type rule for scalars f and ϕ:

$$(25) \quad (f \cdot \phi)_{//R_n} = f_{//R_n} + \sum_{R_n/2} f_{//R_{n1}} \phi_{//R_{n2}} + \phi_{//Rn}.$$

Using the concept of a yoke, discussed in the previous section, it is now possible, in a convenient manner, to define *invariant Taylor-like expansions* (Barndorff-Nielsen, 1987b). Let g be a yoke and let f be a function on the parameter space. Furthermore, let $f_{//S_n}$ be the symmetric covariant derivative of f relative g, as defined by (21) and (22), and let

$$(26) \quad g^r = g^{r;s} g_s$$

and

$$g^{R_n} = g^{r_1} \cdots g^{r_n}.$$

Then f may be expanded around $\omega \in M$ as

$$(27) \quad f(\tilde{\omega}) = f(\omega) + \sum_{\nu=1}^{\infty} \frac{1}{\nu!} f_{//R_\nu}(\omega) g^{R_\nu}(\omega; \tilde{\omega}).$$

The quantity g^r behaves as a contravariant vector in ω and as a scalar in $\tilde{\omega}$, so that each of the terms

$$(28) \quad \frac{1}{\nu!} f_{//R_\nu}(\omega) g^{R_\nu}(\omega; \tilde{\omega})$$

is invariant under changes of coordinates in M. The approximation to f obtained by summing only up to $\nu = n$ in (27) will therefore be the same irrespective of the coordinate system employed. Moreover, for each ν, (28) is of the same order of magnitude in $\tilde{\omega} - \omega$ as the corresponding term in the Taylor expansion of f in ω-coordinates, i.e.,

$$\frac{1}{\nu!} f_{/R_\nu}(\omega)(\tilde{\omega} - \omega)^{R_\nu}.$$

We now indicate two applications of this idea. In both cases we use as yoke the observed likelihood yoke (7).

A highly accurate approximation to the distribution of the score vector $l = l_*(\omega) = (l_1, \cdots, l_d)$, conditionally on the ancillary a, is given by the probability density function

$$(29) \quad p^*(l_*; \omega \mid a) = c(\omega, a) |\hat{j}|^{1/2} |l_;|^{-1} e^{l - \hat{l}}$$

(Barndorff-Nielson, 1988). For the purposes of gaining geometric-statistical insight into the nature of this approximation and of performing approximate integrations of (29), it is of interest to expand $p^*(l_*; \omega \mid a)$ around ω with, of course, a normal density as the leading term.

To this end we rewrite (29) as

$$p^*(l_*; \omega \mid a) = (2\pi)^{-d/2} |j|^{-1/2} \bar{c}(\omega, a) \left\{ \frac{|j| \, |\hat{j}|}{|l_;|^2} \right\}^{1/2} e^{l - \hat{l}},$$

where $\bar{c}(\omega, a) = (2\pi)^{d/2} c(\omega, a)$. Next, we expand the two last factors invariantly, in the above-mentioned fashion, i.e.,

$$(30) \quad l - \hat{l} = -\tfrac{1}{2} j_{rs} l^r l^s + \tfrac{1}{6} t_{rst} l^r l^s l^t + \cdots$$

and

$$(31) \quad \left\{ \frac{|j| \, |\hat{j}|}{|l_;|^2} \right\}^{1/2} = 1 - t_{rst} j^{rs} l^t + \cdots$$

where t_{rst} is the observed skewness tensor given by

$$(32) \quad t_{rst} = -(t_{rst} + l_{rs;t}[3]),$$

where [3] indicates a sum of 3 similar terms, corresponding to suitable permutations of the indices.

On insertion in (29) this leads to the expansion

$$(33) \quad \begin{aligned} &p^*(l_*; \omega \mid a) \\ &= \phi_d(l_*; j^{-1}) \bar{c}(\omega, a) \{1 + \tfrac{1}{6} t_{rst} h^{rst}(l^*; j^{-1}) + \cdots\}, \end{aligned}$$

where $\phi_d(\cdot; \Sigma)$ denotes the d-dimensional normal density function with covariance matrix Σ and h^{rst} is a third order tensorial Hermite polynomial. Carrying the calculation a step further one obtains the asymptotic formula

$$(34) \quad p^*(l_*; \omega \mid a) \sim \phi_d(l_*; j^{-1}) \{1 + A_1 + A_2\}$$

where the invariant correction terms A_1 and A_2 are of orders $O(n^{-1/2})$ and $O(n^{-1})$, respectively, under ordinary repeated sampling with sample size n, and

$$(35) \quad A_1 = \tfrac{1}{6} t_{rst} h^{rst}$$

while

$$(36) \quad \begin{aligned} A_2 = {} & \tfrac{1}{24} t_{rstu} h^{rstu} - \tfrac{1}{4} t_{rs;tu}(2h^{rstu} + h^{rs} j^{tu}) \\ & + \tfrac{1}{72} t_{rst} t_{uvw} h^{rstuvw}, \end{aligned}$$

involving the contravariant tensorial Hermite polynomials of orders 2, 3, 4 and 6, based on l_* and j^{-1}. Furthermore, t_{rst}, t_{rstu} and $t_{rs;tu}$ are tensors, with t_{rst} defined by (32) and

$$(37) \quad t_{rstu} = l_{rstu} + l_{rs;tu}[4] + \tfrac{1}{2} l_{rs;tu}[6],$$

$$(38) \quad t_{rs;tu} = l_{rs;tu} + l_{rs;v} l_{w;tu} j^{vw}.$$

Letting, again, $///$ indicate covariant differentiation, here with respect to the connection $\overset{-1}{\Gamma}{}^t_{rs} = g^t_{;rs}$ we have

$$(39) \quad t_{rst} = j_{rs///t}$$

and

$$(40) \quad t_{rstu} = \tfrac{1}{4} t_{rst///u}[4].$$

Thus t_{rst} and t_{rstu} may be interpreted in classical differential geometric terms. However, the same appears not to be the case for $t_{rs;tu}$, although the Riemannian fourth-order curvature tensor can be expressed in terms of $t_{rs;tu}$ (Mora, 1988; Blæsild, 1988).

As the second application, we briefly consider the question of approximating a given model M by an exponential model, in the neighborhood of a fixed parameter point ω. This, as has often been suggested, can be done by writing, for ω' an arbitrary point in the neighborhood,

$$p(x;\,\omega') = p(x;\,\omega)e^{l(\omega')-l(\omega)} \tag{41}$$

and developing $l(\omega') - l(\omega)$ in a Taylor series up to any desired order. This procedure is, however, not parametrization invariant and instead one may proceed as above, using (27). If, for instance, we again choose (7) and (21)–(22) for the construction and develop to second order, we obtain the exponential approximation

$$\begin{aligned} p(x;\,\omega') \doteq a(\omega';\,\omega)p(x;\,\omega)\exp\{\theta_1(\omega')\cdot t_1(x) \\ + \theta_2(\omega')\cdot t_2(x)\} \end{aligned} \tag{42}$$

where $a(\omega';\,\omega)$ is the norming constant and

$$\begin{aligned} \theta_1(\omega')\cdot t_1(x) &= \theta_1(\omega';\,\omega)\cdot t_1(x;\,\omega) \\ &= l_r(\omega;\,\omega',\,a)\jmath^{rs}l_s(\omega;\,\hat{\omega},\,a) \end{aligned} \tag{43}$$

and

$$\begin{aligned} &\theta_2(\omega')\cdot t_2(x) \\ &\quad = \theta_2(\omega';\omega)\cdot t_2(x;\omega) \\ &\quad = \tfrac{1}{2}l_{rs}(\omega;\omega',a)l_s(\omega;\omega',a)\jmath^{rt}\jmath^{su} \\ &\qquad \cdot\{l_{tu}(\omega;\hat{\omega},a) - \not{l}_{v;tu}(\omega;\omega,a)\jmath^{vw}l_w(\omega;\hat{\omega},a)\}. \end{aligned} \tag{44}$$

For alternative approaches to the question of invariant approximating exponential models, see Amari (1987a) and Barndorff-Nielsen and Jupp (1989).

STRINGS

The concept and theory of strings as developed in a series of papers (Barndorff-Nielsen, 1986b; Barndorff-Nielsen and Blæsild, 1987a, b, 1988a; Barndorff-Nielsen, Blæsild and Mora, 1988, 1989; see also Murray, 1988) arose out of a seminal idea in a paper by McCullagh and Cox (1986). However, the core of the concept had been independently proposed earlier by Foster (1958, 1961); see also the illuminating article Foster, (1986).

It should be noted that strings in the present sense are quite different from a number of other concepts termed strings, in particular the 'superstrings' of theoretical physics which are currently attracting such exceptional interest. To distinguish them, the present type of strings are occasionally referred to as *derivative strings*.

The motivation for studying strings is partly their intrinsic mathematical interest and partly their actual and potential usefulness in statistics and other areas of applied mathematics. Some of their uses have been indicated above.

A derivative string is a parametrization-dependent multiarray H that obeys the following transformation law under change of coordinates

$$H^{A_kD_n}_{B_lC_m} = \sum_{\mu=1}^{m}\sum_{\nu=n}^{N} H^{R_kU_\nu}_{S_lT_\mu}\omega^{T_\mu}_{/C_m}\psi^{D_n}_{/U_\nu}\omega^{S_l}_{/B_l}\psi^{A_k}_{/R_k}. \tag{45}$$

Here we are again using the index notation discussed at the outset. The transformation law (45) generalizes those for tensors, for the Christoffel symbols of affine connections and for ordinary derivatives of real functions.

Derivative strings obey various useful rules of operation, one of the most important being that of *intertwining*, which produces tensors from pairs of strings (see, in particular, Barndorff-Nielsen and Blæsild, 1987b).

In differential geometry, differentials are treated as objects dual to derivatives. In close analogy hereto, there is a concept of *differential strings* which is dual to that of derivative strings. Differential strings may be characterized by a transformation law somewhat similar to (45) and there is a theory of differential strings that parallels that of derivative strings (see Blæsild and Mora, 1988).

PHYLA

Derivative strings and differential strings are, in consequence of the transformation laws that characterize them, instances of a broader concept that is presently being studied under the name of *phyla*. This concept would seem to encompass also the idea of "new tensors" indicated by Foster (1986, 1988). These two papers introduce quantities that may be seen as further instances of phyla. In Foster (1988), the quantities concerned are essential elements of the author's quest towards a unified field theory of physics.

ADDITIONAL REFERENCES

BARNDORFF-NIELSEN, O. E. (1987b). Differential geometry and statistics: Some mathematical aspects. *Indian J. Math.* **29** 335–350.

BARNDORFF-NIELSEN, O. E. (1987c). On some differential geometric concepts and their relations to statistics. In *Geometrization of Statistical Theory* (C. T. J. Dodson, ed.) 53–90. ULDM Publications, Lancaster.

BARNDORFF-NIELSEN, O. E. (1988). *Parametric Statistical Models and Likelihood. Lecture Notes in Statist.* **50**. Springer, New York.

BARNDORFF-NIELSEN, O. E. and BLÆSILD, P. (1987b). Strings: Contravariant aspect. *Proc. Roy. Soc. London Ser. A* **411** 421–444.

BARNDORFF-NIELSEN, O. E. and BLÆSILD, P. (1988a). Coordinate-free definition of structurally symmetric derivative strings. *Adv. in Appl. Math.* **9** 1–6.

BARNDORFF-NIELSEN, O. E. and BLÆSILD, P. (1988b). Inversion and index notation. Research Report 169, Dept. Theoretical Statistics, Aarhus Univ.

BARNDORFF-NIELSEN, O. E., BLÆSILD, P. and MORA, M. (1988). *Derivative Strings and Higher Order Differentiation.* Memoir 11, Dept. Theoretical Statistics, Aarhus Univ.

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 work. 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.