*The Annals of Statistics*
1993, Vol. 21, No. 3, 1197–1224

# PREFERRED POINT GEOMETRY AND STATISTICAL MANIFOLDS[1]

BY FRANK CRITCHLEY, PAUL MARRIOTT AND MARK SALMON

*University of Warwick, University of Surrey and European University Institute*

A new mathematical object called a preferred point geometry is introduced in order to (a) provide a *natural* geometric framework in which to do statistical inference and (b) reflect the distinction between homogeneous aspects (e.g., any point $\theta$ may be the true parameter) and preferred point ones (e.g., when $\theta_0$ is the true parameter). Although preferred point geometry is applicable generally in statistics, we focus here on its relationship to statistical manifolds, in particular to Amari's expected geometry. A symmetry condition characterises when a preferred point geometry both subsumes a statistical manifold and, simultaneously, generalises it to arbitrary order. There are corresponding links with Barndorff-Nielsen's strings. The rather unnatural mixing of metric and nonmetric connections in statistical manifolds is avoided since all connections used are shown to be metric. An interpretation of duality of statistical manifolds is given in terms of the relation between the score vector and the maximum likelihood estimate.

**1. Introduction.** A great deal of recent research has been concentrated on the interface between differential geometry and statistics, see, for example, the review papers by Barndorff-Nielsen, Cox and Reid (1986) and by Kass (1989). One goal of this activity is to establish a natural and productive relationship between these two disciplines and hence to deepen our understanding of statistical methods. Some of the attractions are obvious, at least from the statistical side. The language, insight and intuition that geometry has to offer is invaluable in certain complex statistical issues. Moreover the two disciplines are clearly compatible in certain important senses. For example, the coordinate free approach to geometry mirrors the parameterisation invariance approach to inference. The relationship has already shown itself to be productive. Instances of this include the higher-order asymptotic theory of statistical inference as studied, for example, in part II of the monograph by Amari (1985), and the invariant asymptotic expansions being developed by Barndorff-Nielsen (1988) and co-workers as well as in many journal publications.

This developing relationship has not been all one-sided. In particular, it has produced new mathematical objects worthy of study in their own right. These

Received April 1991; revised June 1992.

[1]Research supported by an ESRC grant on "Geodesic Inference, Encompassing and Preferred Point Geometries in Econometrics."

*AMS* 1991 *subject classifications*. Primary 53B99; secondary 62F05, 62F12.

*Key words and phrases*. Preferred point geometry, statistical manifolds, Amari's expected geometry.

include string theory as introduced by Barndorff-Nielsen (1988) in connection with invariant asymptotic expansions, and also statistical manifolds as formalised in Lauritzen (1987). These manifolds include as a special case the expected geometry of Amari (1985) as well as the observed geometry of Barndorff-Nielsen (1988) and the minimum contrast geometry of Eguchi (1983).

Notwithstanding the progress already made a number of fundamental issues remain. Two in particular stand out. First, Barndorff-Nielsen, Cox and Reid (1986) end their review of the role of differential geometry in statistical theory with the following sentence (the italics are ours): "While the introduction of more specifically geometrical notions has considerable potential, it remains a challenging task to introduce such ideas in a way that is statistically *wholly natural*." Equally, whatever their elegant properties, Lauritzen's statistical manifolds are not wholly natural *geometrically*, at least in the tautological sense that no pure geometer had ever dreamed them up before.

Second, in the traditional geometric approach, all points in a manifold are treated equally. No point is singled out for special treatment, in which case we say that the geometry is *homogeneous*. Current geometrisations of statistics follow this homogeneous approach. From some points of view this is natural statistically. When, for example, all points $\theta$ in the parameter space $\Theta$ share the possibility of being the unknown $\theta_0$ giving rise to the data. From other points of view it is not. Frequently in statistics a single point has a special or "preferred" status in $\Theta$. This could be the (hypothesised) true value or the (constrained) maximum likelihood estimate. In such cases it is natural to consider defining a geometry on the whole manifold which reflects the status of the preferred point.

With these considerations in mind, we introduce in this paper a new mathematical object called a *preferred point geometry*. Its definition and use are guided by two principles: to be as natural and simple as possible from both the statistical and the geometric viewpoint and, where appropriate, to reflect in a natural way the special status of the preferred point. The expectation is that by providing a natural geometric framework for statistical inference, preferred point geometries will provide a productive way to conduct statistical inference in practice.

A preferred point geometry essentially is a Riemannian structure. Riemannian geometries are generalisations of Euclidean geometry to curved spaces in which the metric tensor determines the whole geometry. The metric enables us to define the length of any curve in the manifold. In particular, it is natural to use the metric or Levi–Civita connection whose geodesics are curves of minimum length. The idea behind preferred point geometry is to follow this natural path but to use a metric which depends upon the choice of preferred point. One statistically natural such choice of inner product or metric is the covariance of the score vector, or the inverse of the covariance of the maximum likelihood estimate, taken with respect to the true distribution. In such a case the true distribution will play the role of preferred point.

In this paper we develop the relationship between preferred point geometry and statistical manifolds, often focussing for illustration on the expected geometry of Amari (1985). Wider uses of preferred point geometry in statistical inference are outlined in the final section. Any preferred point geometry has a homogeneous geometry associated with it. We show that, if and only if a certain symmetry condition holds, this homogeneous geometry subsumes a statistical manifold structure while at the same time naturally generalising it to arbitrary order. There are corresponding links with Barndorff-Nielsen's strings. The geometrically rather unnatural mixing of metric and nonmetric connections that occurs in statistical manifolds is avoided by following a preferred point approach as all connections are shown to be metric based. Statistically natural preferred point metrics are provided for the expected geometry case, where the duality between connections can be interpreted as reflecting a certain duality between the score vector and the maximum likelihood estimate. Duality theorems for arbitrary preferred point manifolds are also given.

The plan of the paper is as follows. Section 2 briefly reviews some necessary differential geometrical background. Section 3 gives a short summary of statistical manifold theory and identifies aspects of it that can be seen as shortcomings. Section 4 introduces preferred point geometries formally and gives examples. These are used as running examples throughout the paper. Section 5 discusses the derivation of statistical manifolds from preferred point geometries. Section 6 deals with duality in preferred point geometry. Section 7 follows through the earlier examples paying particular attention to the full exponential family case. Section 8 shows how asymptotic links between the distribution of the score vector and of local approximations to the distribution of the maximum likelihood estimate throw light on the nature of duality in Amari's expected geometry. The natural extension of statistical manifolds to higher order is discussed in Section 9. The final section reviews extensions and further work.

**2. Some differential geometric background.** We briefly review the differential geometric constructions used in this paper. Amari (1985) and Barndorff-Nielsen, Cox and Reid (1986) are two sources which cover the differential geometry used in the current statistical literature. Murray and Rice (1987) and Murray (1988) give a more mathematical treatment and cover the higher order covariant derivatives and covariant Taylor series which are used in this paper.

We shall assume familiarity with the concept of a *tensor*, a *manifold*, a *vector field*, a *metric*, a *Riemannian manifold*, $(M, g)$ where $M$ is the manifold and $g$ any metric defined on this manifold and a *connection* $\nabla$ defined on the *tangent bundle* $TM$ of the manifold. Also the *Riemann–Christoffel curvature tensor* and the concept of a *flat* metric and *affine coordinate system* are used. Definitions of these can be found in Barndorff-Nielsen, Cox and Reid (1986). Also we shall use the concept of an *n-form*, and a *covariant derivative*, definitions of which will be found in Murray and Rice (1987).

Given a coordinate system $\{\theta\}$ on a manifold $M$ and a metric $g_{ij}(\theta)$, we define the *Levi–Civita* or *metric* connection by its Christoffel symbols,

$$\Gamma_{ijk} = \frac{1}{2}\left(\frac{\partial g_{jk}(\theta)}{\partial \theta_i} + \frac{\partial g_{ik}(\theta)}{\partial \theta_j} - \frac{\partial g_{ij}(\theta)}{\partial \theta_k}\right).$$

For a Riemannian manifold this is the natural connection and it is characterised by the property that, if we denote the metric by $\langle\ ,\ \rangle$, then

$$X\langle Y, Z\rangle = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle,$$

where $X, Y, Z$ are vector fields and $\nabla$ is the Levi–Civita connection. It also has the property that the geodesics of the Levi–Civita connection are curves of minimum length among those paths that lie in the manifold.

Of course not all connections are metric or derived from metrics in this way. It is one main purpose of this paper to show that in fact all the connections used in statistics are metric connections and to show how the metrics which generate these connections are statistically very natural.

A connection $\nabla$ allows us to differentiate tangent vectors and covariant tensors. Such a derivative is known as a *covariant derivative*. It also induces the covariant derivative of differential forms and so allows the covariant version of Taylor's theorem. Following Murray and Rice (1987) we denote this induced connection by $\tilde{\nabla}$. Although the mathematics is similar this must not be confused with the dual connection of a statistical manifold which is defined below and is used extensively in this paper. $\tilde{\nabla}$ is defined by its Christoffel symbols $\tilde{\Gamma}$ where if the Christoffel symbols of $\nabla$ are $\Gamma$ we have

$$\tilde{\Gamma}_{ij}^{k} = -\Gamma_{ij}^{k}.$$

Thus given a function $f$ on a manifold its "first derivative" is a one form $df$, and its covariant second derivative is the two form $\tilde{\nabla} df$ where

$$\tilde{\nabla} df = \sum_{k,l=1}^{n} \frac{\partial^2 f}{\partial \theta^k \partial \theta^l} d\theta^k \otimes d\theta^l + \sum_{k,l=1}^{n} \frac{\partial f}{\partial \theta^k} \tilde{\Gamma}_{jl}^{k}\, d\theta^j \otimes d\theta^k$$

evaluating on the tangent vectors $\partial_k$ and $\partial_l$ will give the formula

$$\tilde{\nabla} df(\partial_k, \partial_l) = \frac{\partial^2 f}{\partial \theta^k \partial \theta^l} - \frac{\partial f}{\partial \theta^k}\Gamma_{jl}^{k},$$

which we shall call the *covariant Hessian* of the function $f$. It has the property that it is a two tensor unlike the standard Hessian which is a tensor if and only if

$$\frac{\partial f}{\partial \theta^i} = 0 \qquad \forall\, i,$$

when in fact the two formulas agree. Thus the Hessian will be a tensor if $f$ is a constant or in any tangent space corresponding to a turning point of $f$.

We finish this short review by stating Murray and Rice's formula for a covariant version of Taylor's theorem, which has the advantage that each term in the series is a tensor, thus invariant to reparametrisation, unlike the standard Taylor's series expansion. Let $f$ be a function on a manifold with connection $\nabla$. Let $\gamma(t)$ be a geodesic, then if $p = \gamma(0)$ and $v = \gamma'(0)$ we have the following expansion,

$$f(\gamma(t)) = f(\gamma(0)) + \sum_{k=1}^{\infty} \frac{t^k}{k!} \tilde{\nabla}^{k-1} df(v, \dots, v),$$

where the higher order covariant derivatives are dèfined inductively. Such covariant Taylor expansions can be used in statistical applications due to the invariant nature of each term which then have a specific statistical and geometric interpretation, see McCullagh and Cox (1986) and Barndorff-Nielsen (1986).

**3. Statistical manifold theory.** We briefly recall here the definition of a statistical manifold and some basic properties. Lauritzen (1987) proposed the following mathematical structure for a *statistical manifold* as a unification of earlier work notably by Amari, Barndorff-Nielsen and Eguchi.

DEFINITION. A statistical manifold $(M, g, T)$ is a manifold $M$ with a metric $g$ and a covariant 3-tensor $T$, symmetric in all indices, called the skewness.

From the mathematical point of view the new element of this geometric structure is its one parameter family of (nonmetric) connections $\nabla^{\alpha}$ called *α-connections*. For each $\alpha(\in \mathbf{R})$ a connection is defined by the Christoffel symbols given by

$$\Gamma^{\alpha}_{ijk} = \Gamma^{0}_{ijk} - \frac{\alpha}{2} T_{ijk},$$

where $\Gamma^0$ are the Christoffel symbols for the Levi–Civita connection of the metric $g$.

Amari (1985), Barndorff-Nielsen (1988) and Chentsov (1972) have all shown the fundamental importance of these connections in statistical theory. Also the statistical curvature in an exponential family defined by Efron (1975) can be seen as the curvature associated to Amari's +1-connection. A number of significant results about these connections do exist without a complete theoretical framework. For example, one important property is that of *duality* or *conjugacy*.

DEFINITION. Two connections $\nabla$ and $\nabla^*$ are dual (or conjugate) with respect to a metric $\langle\ ,\ \rangle$ if for all vector fields $X, Y, Z$ we have

$$\langle \nabla_X^* Y, Z \rangle = X\langle Y, Z\rangle - \langle Y, \nabla_X Z\rangle.$$

See Lauritzen [(1987), page 181]. A remarkable result here is that a statistical manifold is flat with respect to $\nabla$ if and only if it is flat with respect to $\nabla^*$.

This concept of duality gives an equivalent way of defining a statistical manifold.

DEFINITION. A statistical manifold $(M, g, \nabla)$ is a manifold $M$ with a metric $g$, and a torsion free connection $\nabla$ whose *dual connection* $\nabla^*$ is also torsion free.

The relationship between the two definitions of a statistical manifold is given by the Christoffel symbols of the connection and its dual. These are

$$\Gamma^0_{ijk} - \tfrac{1}{2}T_{ijk} \quad \text{and} \quad \Gamma^0_{ijk} + \tfrac{1}{2}T_{ijk}, \quad \text{respectively}.$$

Lauritzen noted that the dual connection is torsion free if and only if the tensor $T$ is symmetric.

DEFINITION. In this paper our examples are particularly concerned with the *expected geometry* version of a statistical manifold, where the metric $g$ is the Fisher information and $T$, the skewness, is defined by

$$T_{ijk}(\theta) = \mathbf{E}_\theta\left[\frac{\partial}{\partial\theta_i}\ln p(x,\theta)\cdot\frac{\partial}{\partial\theta_j}\ln p(x,\theta)\cdot\frac{\partial}{\partial\theta_k}\ln p(x,\theta)\right].$$

Amari (1985, 1987) has considered this example of a statistical manifold where the manifolds are parametric families of distributions satisfying certain regularity conditions stated in Amari [(1985), page 16]. We shall assume all parametric families in this paper satisfy these same conditions.

We offer the following remarks about certain aspects of statistical manifolds which the present paper sets out to clarify by adopting a preferred point perspective:

(a) Statistical manifolds are homogeneous geometries. No point is singled out for special treatment. For example the expected geometry reflects the fact that all parameter values share the property that they could represent the true one. However, it does not reflect the fact that in any particular problem only one of them is the true parameter.

(b) Statistical manifolds mix metric and nonmetric connections within a single structure. This is a novel construction in pure geometric theory and thus invites further investigation.

(c) The statistical meaning and implications of duality, and of the results concerning it, are far from clear.

(d) As Lauritzen (1987) points out, statistical manifolds are limited to third order structures. For example in the expected geometry case, they deal with the second and third moments of the score vector only. This limits the information which they contain. It is therefore natural to inquire about higher order extensions.

**4. Preferred point manifolds.** The application of statistical methods naturally requires some assumption regarding the true distribution. This induces an asymmetry between the true model and other points in the manifold which represent alternative models. Existing statistical manifold theory does not adequately capture the formal structure required to properly analyse statistical methods since the geometry does not reflect this asymmetry. The new geometric structure which we introduce takes account of this asymmetry by treating one point as being different or *preferred*. In fact the preferred point need not correspond to the true distribution since this may not lie in the manifold of models being currently considered. This is particularly important since we rarely ever know the true data generation process. Taking this view of the preferred point enables us to consider the properties of estimators when the true data generation process does not necessarily lie in the assumed model set [see, e.g., White (1982) or Gourieroux, Monfort and Trognon (1984)]. In addition the preferred structure enables us to consider problems of misspecification, and testing for separate families following Cox (1961), within a proper geometric framework.

DEFINITION. A preferred point manifold (or geometry) is a pair $(M, g^{\phi}(\theta))$, where $M$ is a manifold and $g^{\phi}(\theta)$ a symmetric covariant 2-tensor on $M$ which is positive definite in a neighbourhood of $\theta = \phi$ and defined smoothly as a function of the (preferred) point, $\phi \in M$.

Thus, in a neighbourhood of $\phi$, $g^{\phi}(\theta)$ is a metric whose value is defined as a smooth function of the preferred point $\phi$ as well as $\theta$. Clearly it is enough for the preferred point metric to be positive definite at the preferred point, due to the smooth dependence on $\phi$.

When considering a specific preferred point $\theta_0$ we shall use the notation $(M, g^{\theta_0}(\theta), \theta_0)$. If the manifold $M$ is a $p$-dimensional object, then the full preferred point manifold $(M, g^{\phi}(\theta))$ is essentially $2p$-dimensional, since the geometric structure at a point $\theta$ depends on the value of both $\theta$ and $\phi$. However once the value of the preferred point is fixed at $\theta_0$, $(M, g^{\theta_0}(\theta), \theta_0)$ is again $p$-dimensional.

We show in Section 5 that the relationship between preferred point manifolds and statistical manifolds can be seen by examining what happens on the diagonal; that is, when $\theta = \phi$. We shall see that on the diagonal the preferred point metric can agree with the metric in the statistical manifold, while its Levi-Civita connection will agree with the $+1$-connection. We call such a preferred point manifold $+1$-compatible. There will in fact be infinitely many preferred point manifolds compatible with any statistical manifold. It is an important question, therefore, whether there are compatible preferred point geometries which are natural and compelling *in their own right*. The following examples, which we develop throughout the paper, show that the answer is in the affirmative for Amari's expected geometry.

Reference to the context will avoid confusion in using the same symbol $g^{\phi}(\theta)$ for a general preferred point metric and also for that used in Example 1:

EXAMPLE 1. Consider the preferred point geometric structure $(M, g^{\phi}(\theta))$, where $M$ is a (regular) parametric family of densities $\{p(x, \theta)\}$, $\phi \in M$ is the preferred point, and the preferred point metric is given by

$$g_{ij}^{\phi}(\theta) = \mathbf{E}_{p(x,\phi)}\left[\frac{\partial}{\partial\theta_i}\ln p(x,\theta) \cdot \frac{\partial}{\partial\theta_j}\ln p(x,\theta)\right].$$

This metric is just the second moment of the score vector with respect to the distribution labelled by $\phi$. This differs from the Fisher information metric simply by the fact that all expectations are taken consistently with respect to the preferred point. We see that when $\theta$ is evaluated at the preferred point $\phi$ the metric is just the Fisher information at that point. The Christoffel symbols for the Levi–Civita connection for the preferred point metric $g^{\phi}(\theta)$ are given by

$$\Gamma_{ijk}^{\phi}(\theta) = \mathbf{E}_{p(x,\phi)}\left[\frac{\partial^2}{\partial\theta_i\,\partial\theta_j}\ln p(x,\theta) \cdot \frac{\partial}{\partial\theta_k}\ln p(x,\theta)\right].$$

At $\theta = \phi$ the connection agrees with the $+1$-connection in Amari's statistical manifold. See Amari [(1985), page 39]. Thus $(M, g^{\phi}(\theta))$ is $+1$-compatible with Amari's statistical manifold. We immediately see the power of the preferred point method as we can now rationalise the $+1$-connection as metric, or Levi–Civita, connection for this (preferred point) metric. Thus we can bring the concept of metric connection back into the core of statistical geometry.

EXAMPLE 2. Consider a related preferred point metric which is the $p(x, \phi)$ covariance of the score vector:

$$\begin{aligned}\tilde{g}_{ij}^{\phi}(\theta) = \mathbf{E}_{p(x,\phi)}\Bigg[&\left(\frac{\partial}{\partial\theta_i}\ln p(x,\theta) - \mathbf{E}_{p(x,\phi)}\left[\frac{\partial}{\partial\theta_i}\ln p(x,\theta)\right]\right)\\ &\cdot\left(\frac{\partial}{\partial\theta_j}\ln p(x,\theta) - \mathbf{E}_{p(x,\phi)}\left[\frac{\partial}{\partial\theta_j}\ln p(x,\theta)\right]\right)\Bigg].\end{aligned}$$

Clearly this reduces to the Fisher information when $\theta = \phi$. The Christoffel symbols of the Levi–Civita connection in this case are

$$\begin{aligned}\tilde{\Gamma}_{ijk}^{\phi}(\theta) = {}& \mathbf{E}_{p(x,\phi)}\left[\frac{\partial^2}{\partial\theta_i\,\partial\theta_j}\ln p(x,\theta)\frac{\partial}{\partial\theta_k}\ln p(x,\theta)\right]\\ &- \mathbf{E}_{p(x,\phi)}\left[\frac{\partial^2}{\partial\theta_i\,\partial\theta_j}\ln p(x,\theta)\right]\mathbf{E}_{p(x,\phi)}\left[\frac{\partial}{\partial\theta_k}\ln p(x,\theta)\right].\end{aligned}$$

We can immediately see that by evaluating $\theta$ at the preferred point they agree with the $+1$-connection of Amari, as in the noncentral moment case. Thus $(M, \tilde{g}^{\phi}(\theta))$ is also $+1$-compatible with Amari's expected geometry. It is, perhaps, more natural to consider the central moments, as in Example 2, since away from the preferred point

$$\mathbf{E}_{p(x,\phi)}\left[\frac{\partial}{\partial \theta_i} \ln p(x,\theta)\right] \neq 0.$$

The previous two preferred point geometries can be seen as generalisations of the Fisher information when it is written in the form

$$\mathbf{E}_{p(x,\theta)}\left[\frac{\partial}{\partial \theta_i} \ln p(x,\theta) \cdot \frac{\partial}{\partial \theta_j} \ln p(x,\theta)\right].$$

We could however also generalise the second derivative form

$$-\mathbf{E}_{p(x,\theta)}\left[\frac{\partial^2}{\partial \theta_i \, \partial \theta_j} \ln p(x,\theta)\right].$$

EXAMPLE 3. If we took the expectation of the Hessian about some fixed preferred point we would not get a tensor since it does not transform correctly under a change of coordinates. Rather we have to look at the covariant version of the Hessian which was defined in the introduction. We obtain the following preferred point metric which clearly reduces to the Fisher information when $\theta = \phi$:

$$h_{ij}^{\phi}(\theta) = -\mathbf{E}_{p(x,\phi)}\left[\frac{\partial^2}{\partial \theta_i \, \partial \theta_j} \ln p(x,\theta) - \left(g^{\phi}\right)^{sr} \Gamma_{ijs}^{\phi}(\theta) \frac{\partial}{\partial \theta_r} \ln p(x,\theta)\right],$$

where $g^{\phi rs}$ is the inverse of $g_{rs}^{\phi}$ the preferred point metric from Example 1 and $\Gamma^{\phi}$ is its Christoffel symbol. In this case the Levi-Civita connection of $h_{ij}^{\phi}$ equals that of the 0-connection when $\theta = \phi$ (see Theorem 4). We note we get the same compatibility if, in the formula, we replace the metric of Example 1 with that of Example 2, as long as we change to the corresponding Christoffel symbol.

Examples 1, 2 and 3 generate the $+1$- and 0-connections of Amari when evaluated at the preferred point. We give below preferred point metrics that are $-1$-compatible with Amari's expected geometry. Indeed we will show later that these are dual to their $+1$-compatible counterparts in a sense that reflects a certain duality between the score vector and the maximum likelihood estimate.

We have seen the preferred point generalisation of two different forms of the Fisher information matrix. We can also consider the form of this matrix in the case of misspecification. If $\theta$ lies outside our manifold and the maximum likelihood estimate on the manifold converges to $\theta^*$, then the form of the information matrix at $\theta^*$ is

$$\mathbf{E}_\theta\left[\frac{\partial^2}{\partial\theta_s\,\partial\theta_i}\ln p(x,\theta^*)\right]\left[\mathbf{E}_\theta\left[\frac{\partial}{\partial\theta_s}\ln p(x,\theta^*)\frac{\partial}{\partial\theta_t}\ln p(x,\theta^*)\right]\right]^{-1}$$
$$\times\mathbf{E}_\theta\left[\frac{\partial^2}{\partial\theta_t\,\partial\theta_j}\ln p(x,\theta^*)\right].$$

For details of this form of the information matrix see White (1982) or Gourieroux, Monfort and Trognon (1984).

Example 4. Consider the formulas

$$k_{ij}^{\phi}(\theta) = g_{is}(\theta)g^{\phi st}(\theta)g_{jt}(\theta) \quad\text{and}\quad l_{ij}^{\phi}(\theta) = h_{is}^{\phi}(\theta)g^{\phi st}(\theta)h_{jt}^{\phi}(\theta),$$

where $g$ is the Fisher information and $g^{\phi}$ and $h^{\phi}$ are defined in Examples 1 and 3. Then $k_{ij}^{\phi}(\theta)$ and $l_{ij}^{\phi}(\theta)$ are both preferred point metrics which reduce to the Fisher information at $\theta = \phi$. Further they are $-1$-compatible with Amari's expected geometry. This is proved in Theorems 3 and 4. Again as in the case of Example 3, we get the same results if we replace the metric of Example 1 everywhere with that of Example 2.

**5. The derivation of statistical manifolds from preferred point geometry.** Having observed these links with the particular case of Amari's expected geometry, this section considers the general relationship between statistical manifolds and preferred point geometries. It shows how the statistical manifold structure can be explained in the setting of preferred point manifolds. The new geometry enables us to extend the definition of statistical manifolds from purely a third order structure to any arbitrary order. In particular we can provide a geometric interpretation for the fourth order extensions proposed by Barndorff-Nielsen. This is taken up in Section 9. In Section 6 we see the general relationship between a preferred point geometry and the dual structure of statistical manifolds. This is given a direct statistical interpretation in Section 8.

In Section 3 we noted two different ways of viewing a statistical manifold, either as $(M, g, \nabla)$ or as $(M, g, T)$. We shall look at both of these interpretations in turn and see how they can be understood in terms of preferred point geometry. We start with $(M, g, \nabla)$.

5.1. *Homogeneous structures.* Any preferred point geometry $(M, g^{\phi}(\theta))$ has twice the dimension of its underlying parameter space. Let $M$ be $p$-dimensional. We can then view the preferred point geometry as a geometric structure on $M \times M$ as we let both $\theta$ and $\phi$ vary independently. We can recover a $p$-dimensional geometric structure on $M$ by choosing some fixed value of the preferred point $\theta_0$ and looking at the Riemannian manifold $(M, g^{\theta_0}(\theta))$. This is natural statistically when, for example, we regard $\theta_0$ as the (hypothesised) true value. However, this is not the only way of recovering a $p$-dimensional

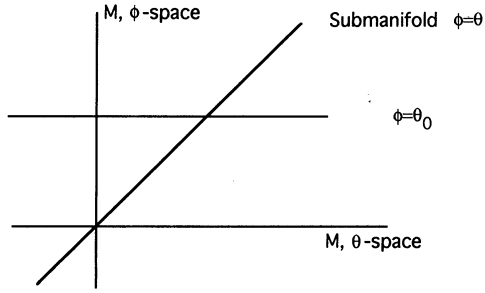


FIG. 1.

structure. Consider Figure 1, the diagram denotes the preferred point manifold as $M \times M$ with the vertical axis representing the variation in the preferred point $\phi$ and the horizontal axis the variation in the parameter $\theta$. Thus the $p$-dimensional structure defined by fixing a value of the preferred point at $\theta_0$ is represented by a horizontal line. Consider the $p$-dimensional structure defined by the diagonal $\theta = \phi$.

DEFINITION. A $p$-dimensional substructure of $(M, g^{\phi}(\theta))$ can be defined by taking geometric objects in the preferred point geometry and restricting to the diagonal $\phi = \theta$, as $\phi$ varies across $M$. Such a structure is called *homogeneous*. This name is used since in these geometric structures all points are given the same importance and no one point is singled out to be preferred.

Theorem 1 shows there is a Riemannian structure induced by a preferred point geometry $(M, g^{\phi}(\theta))$. This is the one given by $(M, g^{\theta}(\theta))$, called the *homogeneous Riemannian* structure.

EXAMPLE. If we are working in any of the preferred point geometries given by Examples 1, 2, 3 or 4, then the homogeneous Riemannian structures are all given by the Fisher metric. This follows at once from properties noted above.

In a preferred point manifold we can define other homogeneous structures apart from the Riemannian one. We shall now show, in the theorem below, how a statistical manifold can be seen as a *homogeneous connection geometry* of a preferred point manifold.

DEFINITION. Let us define $\nabla^{\phi}(\theta)$ to be the Levi–Civita connection of $g^{\phi}(\theta)$ evaluated at $\theta$, and define $\overline{\nabla}(\theta)$ to be the Levi–Civita connection of the homogeneous Riemannian metric.

THEOREM 1. (i) *Let $(M, g^{\phi}(\theta))$ be a preferred point manifold. The homogeneous structure $(M, g^{\theta}(\theta))$ defines a Riemannian manifold, which we will denote by $(M, g)$.*

(ii) $\nabla^{\theta}(\theta)$ *is a homogeneous connection structure. In particular* $\nabla^{\theta}(\theta)$ *defines a torsion free connection on* $M$.

(iii) *Then* $(M, g, \nabla^{\theta}(\theta))$ *is a statistical manifold if and only if*

$$T(\theta) = 2\big(\overline{\nabla}(\theta) - \nabla^{\theta}(\theta)\big),$$

*which is a* 3*-tensor, is symmetric. In this case we call* $T$ *the skewness of the preferred point manifold, following Lauritzen.*

PROOF. (i) Defining $g(\theta) = g^{\theta}(\theta)$, we show that $g$ is a metric on $M$. Clearly it is a symmetric, positive definite, bilinear form. Thus we need to show that it transforms as a tensor. Since for each fixed preferred point $\theta_0$ we know the transformation rule for the metric $g^{\theta_0}(\theta)$ is given by

$$g_{ij}^{\theta_0}(\theta) \mapsto \frac{\partial \theta_r}{\partial \psi_i}(\theta) \cdot \frac{\partial \theta_t}{\partial \psi_j}(\theta) g_{rt}^{\theta_0}(\theta)$$

setting $\theta_0 = \theta$ we get the result. Thus $(M, g) = (M, g^{\theta}(\theta))$ is a Riemannian manifold.

(ii) Denote the Christoffel symbol for the Levi–Civita connection of the preferred point metric $g^{\theta_0}(\theta)$ by $\Gamma_{ijk}^{\theta_0}(\theta)$. The homogeneous connection information is defined by the Christoffel symbols

$$\Gamma_{ijk}^{\theta}(\theta)$$

for each $\theta$. As for the metric case we can easily check that $\Gamma_{ijk}^{\theta}(\theta)$ transforms as a Christoffel symbol for a connection. Furthermore for all $\theta_0$, $\Gamma_{ijk}^{\theta_0}(\theta)$ is a Levi–Civita connection and hence torsion free. That is, its Christoffel symbols are symmetric in $i$ and $j$. Therefore the Christoffel symbols $\Gamma_{ijk}^{\theta}(\theta)$ are also symmetric in $i$ and $j$, and so represent a torsion free connection.

(iii) The final part of this theorem follows from a result by Lauritzen (1987, page 183). □

We give as a corollary some conditions under which the preferred point connection and the homogeneous connection of a preferred point geometry agree everywhere, not just at the preferred point. This result is used in Section 7 in the exponential family case. We recall that an affine coordinate system for the metric $g_{ij}(\theta)$ is one in which the metric has a constant representation, independent of $\theta$, and a metric for which an affine coordinate system exists is called a *flat* metric.

COROLLARY. *Let* $(M, g^{\phi}(\theta))$ *be a preferred point manifold such that* $g^{\phi}(\theta)$ *is a flat metric for each value of the preferred point. Suppose, further, that there exists a coordinates system* $\psi$ *which is affine for* $g^{\phi}(\theta)$, *for all* $\phi$. *Then for each* $\phi$ *the preferred point connection* $\nabla^{\phi}(\theta)$ *agrees with the homogeneous connection* $\nabla^{\theta}(\theta)$ *at all points* $\theta$.

PROOF. Working in $\psi$-coordinates, the Christoffel symbols for the homogeneous connection at $\theta$ are those of the preferred point connection when $\theta$ is the preferred point. These are zero since the Christoffel symbols for $\nabla^{\phi}(\theta)$ are identically zero for all $\theta$ and $\phi$ in the $\psi$-coordinate system. Consider the two connections $\nabla^{\phi}(\theta)$ and $\nabla^{\theta}(\theta)$. As both their Christoffel symbols are zero everywhere and they are both torsion free they must be identical. □

EXAMPLE. The preferred point geometry of Example 2 generates Amari's expected statistical manifold. We have already seen that the homogeneous Riemannian structure is given by the Fisher metric. We therefore look at the second order homogeneous structure. We have already seen that this is given by the $+1$-connection of Amari. Therefore we simply need to check that the symmetry condition on $T$ holds. The 3-tensor in this case is given by the tensor

$$\mathbf{E}_{\theta_0}\left[\frac{\partial}{\partial\theta_i}\ln p(x,\theta)\frac{\partial}{\partial\theta_j}\ln p(x,\theta)\frac{\partial}{\partial\theta_k}\ln p(x,\theta)\right]\Bigg|_{\theta_0=\theta},$$

which is clearly symmetric.

Example 1 will also generate Amari's expected statistical manifold in exactly the same way. Thus this demonstrates that there is not a one to one correspondence between preferred point geometries and statistical manifolds. Furthermore, it is easy to construct preferred point geometries for which the skewness is not symmetric. Thus for both these reasons preferred point geometries should be considered as a more general construction.

5.2. *Correction terms.* In this section we look at the relationship between the second form of the statistical manifold structure $(M, g, T)$ and that of preferred point geometries. We show that there is a direct geometric interpretation of Amari's skewness tensor in the preferred point geometry.

As we saw in the previous section a statistical manifold is a homogeneous structure. It is important to understand how a homogeneous structure differs from a preferred point one at the preferred point. Consider, as an example, the contrast between the homogeneous Riemannian and preferred point metric structures. At the preferred point $\theta_0$ the two metrics agree. They only differ at other points. It is therefore in their derivative at $\theta_0$ that they could disagree. To demonstrate this consider the preferred point metric as a perturbed version of the homogeneous metric.

We define a perturbation of a preferred point metric $g_{ij}^{\theta_0}(\theta)$ to be $p_{ij}^{\theta_0}(\theta)$, where

$$g_{ij}(\theta) = g_{ij}^{\theta_0}(\theta) + p_{ij}^{\theta_0}(\theta), \qquad p_{ij}^{\theta_0}(\theta_0) = 0$$

and $g$ is the homogeneous metric. Thus the perturbation is a symmetric 2-tensor which is zero at the preferred point. The difference in the derivatives of the homogeneous structure and the preferred point metric are given by the derivative of the perturbation. We therefore define the *first order correction to the homogeneous Riemannian structure* to be

$$\tilde{T}_{ijk}(\theta) = \frac{\partial}{\partial \theta_k} p_{ij}^{\theta_0}(\theta)\bigg|_{\theta_0=\theta}.$$

THEOREM 2. *$\tilde{T}(\theta)$ is a homogeneous 3-tensor. Further if $\tilde{T}(\theta)$ is symmetric in all its indices, then we have that $\tilde{T}(\theta) = T(\theta)$, defined above, and so the symmetry condition of Theorem 1 holds.*

PROOF. Since we are evaluating on $\theta_0 = \theta$ we see that $\tilde{T}(\theta)$ is homogeneous. Under a change of variable, $(\theta \to \psi)$, $p$ transforms as

$$p_{ij}^{\theta_0}(\theta) \mapsto \frac{\partial \theta_r}{\partial \psi_i}(\theta) \cdot \frac{\partial \theta_t}{\partial \psi_t}(\theta) p_{rt}^{\theta_0}(\theta)$$

thus $(\partial/\partial\theta)p_{ij}^{\theta_0}(\theta)$ transforms as

$$\frac{\partial}{\partial \theta_k} p_{ij}^{\theta_0}(\theta) \mapsto \frac{\partial \theta_r}{\partial \psi_i}(\theta) \cdot \frac{\partial \theta_t}{\partial \psi_j}(\theta) \frac{\partial \theta_s}{\partial \psi_k}(\theta) \frac{\partial}{\partial \theta_s} p_{rt}^{\theta_0}(\theta)$$
$$+ p_{rt}^{\theta_0}(\theta) \frac{\partial \theta_s}{\partial \psi_k}(\theta) \frac{\partial}{\partial \theta_s}\left(\frac{\partial \theta_r}{\partial \psi_i}(\theta) \cdot \frac{\partial \theta_t}{\partial \psi_j}(\theta)\right).$$

Hence evaluating on the diagonal $\theta = \theta_0$ and using $p_{ij}^{\theta_0}(\theta_0) = 0$ we have the transformation rule

$$\tilde{T}_{ijk}(\theta) \mapsto \frac{\partial \theta_r}{\partial \psi_i}(\theta) \cdot \frac{\partial \theta_s}{\partial \psi_j}(\theta) \frac{\partial \theta_t}{\partial \psi_k}(\theta) \tilde{T}_{rst}(\theta).$$

Thus it is a tensor.

Calculating the Christoffel symbols of the Levi–Civita connection of $g_{ij}(\theta)$ and evaluating at $\theta = \theta_0$ we see that

$$\Gamma_{ijk}^{\theta_0}|_{\theta=\theta_0} = \frac{1}{2}\left(\frac{\partial g_{jk}}{\partial \theta_i} + \frac{\partial g_{ik}}{\partial \theta_j} - \frac{\partial g_{ij}}{\partial \theta_k}\right) + \frac{1}{2}\left(\frac{\partial p_{jk}^{\theta_0}}{\partial \theta_i} + \frac{\partial p_{ik}^{\theta_0}}{\partial \theta_j} - \frac{\partial p_{ij}^{\theta_0}}{\partial \theta_k}\right)$$
$$= \Gamma_{ijk} + \frac{1}{2}\left(\tilde{T}_{kij} + \tilde{T}_{jki} - \tilde{T}_{ijk}\right),$$

where $\Gamma$ is the Christoffel symbol for the homogeneous metric. Thus if $\tilde{T}(\theta)$ is symmetric we see that it must equal $T(\theta)$ and so $T$ itself must be symmetric. □

EXAMPLE. Consider the preferred point metrics of Examples 1 and 2. Both these examples generate Amari's geometry. We calculate that in both cases $\tilde{T}(\theta)$ equals

$$\mathbf{E}_\theta\left[\frac{\partial}{\partial \theta_i} \ln p(x,\theta) \frac{\partial}{\partial \theta_j} \ln p(x,\theta) \frac{\partial}{\partial \theta_k} \ln p(x,\theta)\right].$$

In other words, the skewness as defined by Amari. (We have the symmetry condition required and $\tilde{T}(\theta) = T(\theta)$ as calculated above.)

We need to use this first order correction if we are calculating the homogeneous connection structure from the homogeneous Riemannian one. The homogeneous metric gives us (Lauritzen's) 0-connection, but by adding the first order correction we get

$$\Gamma^0_{ijk} - \tfrac{1}{2}\left(\tilde{T}_{kij} + \tilde{T}_{jki} - \tilde{T}_{ijk}\right)$$

which, under the right symmetry conditions, is a $+1$-connection.

Thus we have seen that the homogeneous structure of a preferred point geometry generates a triple of the form $(M, g, T)$, that is, the manifold, its homogeneous metric and the first order correction to the derivative of the homogeneous metric. Lauritzen (1987) shows that if we have such a structure and $T$ is symmetric, then we have a statistical manifold and it can be written as $(M, g, \nabla)$ where $\nabla$ is the $+1$-connection generated by $g$ and $T$. We have shown that if we have Lauritzen's symmetry condition then the preferred point geometry which generates $(M, g, T)$ generates the same statistical manifold $(M, g, \nabla)$.

**6. Duality in preferred point manifolds.** Thus a preferred point geometry can generate the basic structure of a statistical manifold in both its forms. The internal mathematical structure of a statistical manifold also has an interpretation in preferred point geometry theory. In particular the significance of the dual connection can be understood very clearly using our new techniques. We first develop the mathematical framework of duality in preferred point geometries. In Section 8 statistical interpretations are given for our constructions. Recall that Examples 1 and 2 generate Amari's $+1$-connection in a natural way, Example 3 generates the 0-connection and Example 4 generates the $-1$-connection. This section shows how this behaviour can be generalised.

DEFINITION. Let the preferred point geometry $(M, g^{\phi})$ generate the statistical manifold $(M, g, \nabla)$ assuming the symmetry conditions stated above. We call $\nabla$, the homogeneous connection, the $+1$-connection of this structure, following Lauritzen. Let

$$k^{\phi}_{ij}(\theta) = g_{is}(\theta) g^{\phi st}(\theta) g_{jt}(\theta),$$

where $g^{\phi ij}(\theta)$ is the inverse of $g^{\phi}_{ij}(\theta)$ and $g$ is the homogeneous metric for the preferred point geometry.

THEOREM 3. *With the above definition $k^{\theta_0}_{ij}(\theta)$ is a preferred point metric whose connection at $\theta = \theta_0$ equals the dual connection to $\nabla$.*

PROOF. $k^{\phi}$ is a preferred point metric since it is a symmetric 2-tensor which is positive definite when evaluated at the preferred point. Consider differentiating and evaluating at $\theta = \theta_0$:

$$\frac{\partial}{\partial \theta_k} k^{\theta_0}_{ij}\bigg|_{\theta_0=\theta} = \frac{\partial}{\partial \theta_k}\big(g_{is}(\theta) g_{jt}(\theta) g^{\theta_0 st}(\theta)\big)\bigg|_{\theta_0=\theta}$$

$$= \frac{\partial}{\partial \theta_k}\big(g_{is}(\theta)\big)\delta^s_j + \frac{\partial}{\partial \theta_k}\big(g_{jt}(\theta)\big)\delta^t_i + g_{is}(\theta) g_{jt}(\theta)\frac{\partial}{\partial \theta_k}\big(g^{\theta_0 st}(\theta)\big).$$

Now

$$g^{\theta_0}_{rs}(\theta) g^{\theta_0 st}(\theta) = \delta^t_r.$$

So differentiating gives

$$\frac{\partial}{\partial \theta_k}\big(g^{\theta_0}_{rs}(\theta)\big) g^{\theta_0 st}(\theta) + g^{\theta_0}_{rs}(\theta)\frac{\partial}{\partial \theta_k}\big(g^{\theta_0 st}(\theta)\big) = 0.$$

At $\theta = \theta_0$,

$$\frac{\partial}{\partial \theta_k}\big(g^{\theta_0 st}(\theta)\big) = -\frac{\partial}{\partial \theta_k}\big(g^{\theta_0}_{ij}(\theta)\big) g^{si}(\theta) g^{tj}(\theta)$$

$$= -\left(\frac{\partial}{\partial \theta_k}\big(g_{ij}(\theta)\big) - T_{ijk}\right) g^{si}(\theta) g^{tj}(\theta).$$

Hence substituting into the original equation we get

$$\frac{\partial}{\partial \theta_k} k^{\theta_0}_{ij} = \frac{\partial}{\partial \theta_k}\big(g_{ij}(\theta)\big) + T_{ijk}.$$

Therefore the Christoffel symbols for this metric are, using the symmetry condition,

$$\Gamma^0_{ijk} + \tfrac{1}{2} T_{ijk};$$

in other words we have the dual connection to $\nabla$ since its Christoffel symbols are $\Gamma^0_{ijk} - (1/2)T_{ijk}$. □

Thus if a preferred point metric generates a $+1$-connection there exists a related preferred point metric which generates the $-1$-connection. This method of constructing the dual connection is however, not entirely consistent with the preferred point philosophy since the dual preferred point metric is constructed using a homogeneous object. The following construction which is appropriate to expected geometries will also produce a preferred point metric which generates the dual connection. In Section 8 we shall see how the first can be viewed as simply a special case of the following method.

As recalled in Section 2, Murray and Rice (1987) give a clear exposition of the way connections can be used to calculate invariant Taylor series using the higher order covariant derivatives of a function. This gives us an interpretation of the metric in Example 3. It is simply the expected value of the covariant Hessian using the connection defined by Example 1. We generalise this construction in the following theorem which shows that for the expected geometry case any preferred point geometry whose homogeneous connection is the $+1$-connection will also generate a preferred point metric which is the generalisation of the Hessian form of the Fisher information. This metric will generate the 0-connection as its homogeneous connection structure. Furthermore, from these two a third preferred point metric can be constructed which generates the $-1$-connection.

THEOREM 4. *Let $g^{\phi}$ be a preferred point metric which generates Amari's expected statistical manifold $(M, g, \nabla)$ where $g$ is the Fisher information, and $\nabla$ is the $+1$-connection. Then*

$$h^{\phi}_{ij}(\theta) = -\mathbf{E}_{p(x,\phi)}\left[\frac{\partial^2}{\partial\theta_i\,\partial\theta_j}\ln p(x,\theta) - \left(g^{\phi}\right)^{sr}\Gamma^{\phi}_{ijs}\frac{\partial}{\partial\theta_r}\ln p(x,\theta)\right]$$

*is a preferred point metric whose homogeneous connection is the 0-connection. Further,*

$$l^{\phi}_{ij} = h^{\phi}_{is}\left(g^{\phi}\right)^{rs}h^{\phi}_{ir}$$

*is also a preferred point metric and its homogeneous connection is the $-1$-connection.*

PROOF. By calculation

$$\frac{\partial^2}{\partial\theta_i\,\partial\theta_j}\ln p(x,\theta) - \left(g^{\phi}\right)^{sr}\Gamma^{\phi}_{ijs}\frac{\partial}{\partial\theta_r}\ln p(x,\theta)$$

is a 2-tensor and hence so is $h^{\phi}_{ij}$. Further at $\phi$ we see that $h^{\phi}_{ij}$ is positive definite since at $\phi$ it equals the Fisher information. Therefore $h^{\phi}_{ij}$ is positive definite in a region of $\phi$ thus is a preferred point metric.

At $\theta = \phi$ we see that

$$\begin{aligned}
\left.\frac{\partial}{\partial\theta_k}h^{\phi}_{ij}(\theta)\right|_{\theta=\phi} &= -\mathbf{E}_{p(x,\phi)}\left[\frac{\partial^3}{\partial\theta_i\,\partial\theta_j\,\partial\theta_k}\ln p(x,\phi)\right] + \Gamma^{\phi}_{ijs}(\phi) \\
&\qquad \times \mathbf{E}_{p(x,\phi)}\left[\frac{\partial^2}{\partial\theta_r\,\partial\theta_k}\ln p(x,\phi)\right]\cdot g^{\phi rs} \\
&= -\mathbf{E}_{p(x,\phi)}\left[\frac{\partial^3}{\partial\theta_i\,\partial\theta_j\,\partial\theta_k}\ln p(x,\phi)\right] - \Gamma^{+1}_{ijk}(\phi) \\
&= -\mathbf{E}_{p(x,\phi)}\left[\frac{\partial^3}{\partial\theta_i\,\partial\theta_j\,\partial\theta_k}\ln p(x,\phi)\right] \\
&\qquad - \mathbf{E}_{p(x,\phi)}\left[\frac{\partial^2}{\partial\theta_i\,\partial\theta_j}\ln p(x,\phi)\frac{\partial}{\partial\theta_k}\ln p(x,\phi)\right] \\
&= -\left.\frac{\partial}{\partial\theta_k}\mathbf{E}_{p(x,\theta)}\left[\frac{\partial^2}{\partial\theta_i\,\partial\theta_j}\ln p(x,\theta)\right]\right|_{\theta=\phi} = \left.\frac{\partial}{\partial\theta_k}g(\theta)\right|_{\theta=\phi}.
\end{aligned}$$

Thus its homogeneous connection agrees with the 0-connection.

The proof that $h^{\phi}_{is}(g^{\phi})^{rs}h^{\phi}_{ir}$ is a preferred point metric with homogeneous metric $g$ and connection the $-1$-connection exactly mirrors that of Theorem 3. □

We note that if $g^{\phi}$ and $h^{\phi}$ are any two preferred point metrics generating the $+1$- and 0-connections, then $l^{\phi}$ constructed as above generates the $-1$-connection. Again, if $g_1^{\phi}$ and $g_2^{\phi}$ are preferred point metrics generating the $\alpha_1$- and $\alpha_2$-connections, then, for any real number $\lambda$, $\lambda g_1^{\phi} + (1-\lambda)g_2^{\phi}$ is another preferred point metric that generates the $\lambda\alpha_1 + (1-\lambda)\alpha_2$ connection. However rather than pursue such purely formal matchings of preferred point metrics, we prefer to concentrate on results such as Theorems 3 and 4 for which a clear statistical interpretation is available.

We can now summarise the content of these theorems as they relate to Amari's expected geometry. There are four triples of natural preferred point metrics which generate the $(+1, 0, -1)$ connections, respectively. They are $(g^{\phi}, h^{\phi}, k^{\phi})$ and $(g^{\phi}, h^{\phi}, l^{\phi})$ with $g^{\phi}$ as in Example 1, and the two further triples that result by replacing throughout $g^{\phi}$ by $\tilde{g}^{\phi}$ of Example 2 to obtain $(\tilde{g}^{\phi}, \tilde{h}^{\phi}, \tilde{k}^{\phi})$ and $(\tilde{g}^{\phi}, \tilde{h}^{\phi}, \tilde{l}^{\phi})$. Thus there are two choices to make in this $2 \times 2$ classification of possibilities:

(a) $g^{\phi}$ or $\tilde{g}^{\phi}$? (We have already indicated that $\tilde{g}^{\phi}$ is preferred.)
(b) the $k$ or $l$ form of duality?

To make this second choice requires a clearer statistical grasp of what is involved. Again we refer forward to Section 8.

**7. Examples in the full exponential family case.** In this section we look in more detail at our examples of preferred point metrics in the case of a full exponential family.

We consider a full exponential family whose density with respect to some carrier measure can be written

$$p(x,\theta) = \exp\{x^i\theta_i - \psi(\theta)\}.$$

Here $\theta$ is the *canonical* or *natural* parameter. Amari (1985) introduces the *expectation* parameter $\eta$ with general element $\eta_i = \mathbf{E}_{p(x,\theta)}[x_i]$. It is immediate that in this case

$$\eta_i = \frac{\partial\psi}{\partial\theta_i}(\theta).$$

Because of our preferred point geometry context, and in order to be able to generalise to arbitrary likelihoods, we introduce the preferred point expectation parameter $\mu^{\phi}$ with general element

$$\mu_i^{\phi}(\theta) = \mathbf{E}_{p(x,\phi)}\left[\frac{\partial}{\partial\theta_i}\ln p(x,\theta)\right].$$

In the exponential family case, these two parametrisations are related by:

$$\mu_i^{\phi}(\theta) = \eta_i(\phi) - \eta_i(\theta).$$

In particular, $\mu^{\theta}(\theta)$ vanishes identically in $\theta$. When $\phi$ is evident from the context, we write $\mu_i(\theta)$ for $\mu_i^{\phi}(\theta)$.

PROPOSITION. *Let $g_{ij}(\theta)$ denote the form of the Fisher information in the $\theta$-coordinate system and $g^{ij}(\theta)$ its inverse, in the same coordinates. Consider the above exponential family. We recall that*

$$g_{ij}(\theta) = \frac{\partial^2 \psi}{\partial\theta_i\,\partial\theta_j}(\theta).$$

*Then we have the following results*:

(a) *In the $\theta$-coordinates system, $g_{ij}^{\phi}(\theta) = g_{ij}(\phi) + \mu_i^{\phi}\mu_j^{\phi}$.*

(b) *In the $\theta$-coordinate system, $\tilde{g}_{ij}^{\phi}(\theta) = g_{ij}(\theta)$, a constant, independent of $\theta$.*

(c) *In the $\mu$-coordinate system, $\tilde{k}_{ij}^{\phi}(\mu) = g^{ij}(\phi)$, a constant, independent of $\mu$.*

(d) *In the $\theta$-coordinate system, $\tilde{h}_{ij}^{\phi}(\theta) = g_{ij}(\theta)$.*

(e) *$\tilde{l}_{ij}^{\phi}(\theta) = \tilde{k}_{ij}^{\phi}(\theta)$, thus a constant in the $\mu$-coordinates.*

*In particular for each $\theta_0$, Example 2, $(M, \tilde{g}_{ij}^{\theta_0}, \theta_0)$, is flat and the natural $\theta$-coordinates are affine, and its dual preferred point manifold $(M, \tilde{k}_{ij}^{\theta_0}, \theta_0)$, which in this case equals $(M, \tilde{l}_{ij}^{\theta_0}, \theta_0)$, is also flat and the expected $\mu$-coordinates, are affine.*

PROOF. (a) and (b). These parts follow by a simple calculation.

(c) Amari (1985) notes that the change of basis matrix for the reparametrisation $\{\theta\} \to \{\eta\}$ is given by the inverse of the Fisher information matrix. Clearly, that for the reparametrisation $\{\eta\} \to \{\mu\}$ is minus the identity matrix. But, by definition $\tilde{k}_{ij}^{\phi}(\theta) = g_{is}(\theta)\tilde{g}^{\phi st}(\theta)g_{tj}(\theta)$. Thus in the $\mu$-coordinate system (c) has the form $\tilde{g}^{\phi st}(\theta(\mu))$, that is, the inverse of the form of (b) in the $\theta$-coordinate system. The result now follows from (b).

(d) Since the metric in (b) is flat, the associated Christoffel symbols vanish in the $\theta$-coordinate system. So the covariant Hessian equals the standard Hessian in this coordinate system. Moreover in the present exponential family

$$\frac{\partial^2}{\partial\theta_i\,\partial\theta_j}\ln p(x,\theta) = -\frac{\partial^2\psi}{\partial\theta_i\,\partial\theta_j}(\theta) = -g_{ij}(\theta).$$

Hence,

$$\tilde{h}_{ij}^{\phi}(\theta) = -\mathbf{E}_{p(x,\phi)}\left[-g_{ij}(\theta) - 0\right] = g_{ij}(\theta).$$

(e) Given (d), this is immediate. □

Amari has shown that the exponential family is $+1$-flat and that the $\theta$-coordinates are $+1$-affine. We remark that this follows at once from part (b) of the above proposition, when we recall the corollary to Theorem 1, since the $\theta$-coordinates are affine for all values of the preferred point, and the fact that the homogeneous connection for $(M, \tilde{g}_{ij}^{\phi}(\theta))$ is the $+1$-connection. By part (c) of this proposition, similar remarks apply to Amari's result that the exponential family is $-1$-flat and that the expected coordinates are $-1$-affine.

Finally we note that the preferred choice $\tilde{g}$ gives neater results than $g$. Explicit formulae for $h$, $k$ and $l$ are not given here.

**8. Duality and asymptotic statistics.** In this section we take an asymptotic statistical look at duality in Amari's expected geometry.

The preferred point metric in Example 2 has a direct statistical interpretation in the space of random variables spanned by

$$\left\{ \frac{\partial}{\partial \theta_i} \ln p(x, \theta) \right\}_{i=1 \cdots p}.$$

Amari identifies this space with the tangent space to the manifold at the point $\theta$. Consider a random sample $\mathbf{x} = \{x_1, \dots, x_n\}$ and let the data generation process lie in our manifold of distributions with parameter $\theta_0$. Under these conditions the score vector for the sample $\mathbf{x}$ will be the sum of the scores for each $x_s$. So we can consider it in the vector space spanned by these vectors or in the representation of the tangent space. We can apply the central limit theorem immediately and see that the score for $\mathbf{x}$ has an asymptotic normal distribution

$$\sum_{s=1}^{n} \frac{\partial}{\partial \theta} \ln p(x_s, \theta) \ \dot{\sim}\ N\left(n\left(\mu_i^{\theta_0}\right), n\tilde{g}_{ij}^{\theta_0}\right),$$

where $\mu_i^{\theta_0}$ is defined in the previous section and $\tilde{g}_{ij}^{\theta_0}$ is the covariance which is given by the preferred point metric from Example 2.

Having found a statistical interpretation for the metric of Example 2 in terms of the score vector we now show a direct statistical interpretation of the duality of a statistical manifold in both the special case of an exponential family, and then in generality. We shall show that the duality in fact corresponds to the relationship between the asymptotic distributions of two random variables, the score and the maximum likelihood estimate.

It is helpful to first recall the derivation of the asymptotic normality of the maximum likelihood estimate from Cox and Hinkley [(1974), page 294]. The asymptotic distribution of the random variable $(\hat{\theta} - \theta)$ is derived under the assumption that $\theta$ is the true parameter. Asymptotically the relationship between the maximum likelihood estimate and the score test is given by the approximation

$$(*) \qquad \sqrt{n}\, g(\theta)(\hat{\theta} - \theta) \approx \frac{1}{\sqrt{n}} \sum_{s} \frac{\partial}{\partial \theta_i} \ln p(x_s, \theta)$$

where we have an i.i.d. sample vector $(x_1, x_2, \ldots, x_n)$ and $g(\theta)$ is the Fisher information at $\theta$. The key to the calculation is the approximation of the maximum likelihood random variable with a sum of i.i.d. random variables which are score vectors. Under the assumption that $\theta$ is the true parameter value the asymptotic distribution of each of the score vectors is $N(0, g(\theta))$. Thus $(*)$ is used to calculate the asymptotic distribution of $n^{1/2}(\hat{\theta} - \theta)$ as $N(0, g^{-1}(\theta))$.

Denoting the true parameter value by $\theta_0$ consider the distribution of $(\hat{\theta} - \theta)$ for all possible $\theta$. We have a choice over which set of linear i.i.d. random variables to approximate $(\hat{\theta} - \theta)$. We could use, as in the standard derivation, the score at $\theta_0$

$$\sum_s \frac{\partial}{\partial \theta_i} \ln p(x_s, \theta_0)$$

or we could use the score at $\theta$

$$\sum_s \frac{\partial}{\partial \theta_i} \ln p(x_s, \theta).$$

The first course will give the same derivation as the classical case except for a trivial translation. Following the second choice and expanding the score at the m.l.e. in a Taylor expansion we get

$$0 = \sum_s \frac{\partial}{\partial \theta_i} \ln p(x_s, \theta) + (\hat{\theta} - \theta)^j \sum_s \frac{\partial^2}{\partial \theta_i \, \partial \theta_j} \ln p(x_s, \theta) + O\Big((\hat{\theta} - \theta)^2\Big).$$

Thus for *small* values of $(\hat{\theta} - \theta)$ we have a good approximation

$$-(\hat{\theta} - \theta)^j \sum_s \frac{\partial^2}{\partial \theta_i \, \partial \theta_j} \ln p(x_s, \theta) \approx \sum_s \frac{\partial}{\partial \theta_i} \ln p(x_s, \theta).$$

Note the difference in the approximation arguments between this and the standard case. There the approximation between the two random variables is due to the fact that asymptotically $(\hat{\theta} - \theta_0)$ converges to zero since we have assumed that $\theta_0$ is the true distribution. Here we are using a different condition. If the m.l.e. lies in a small neighbourhood of $\theta$, then our approximation will be a good one. Note that if

$$|(\hat{\theta} - \theta)| \ll |(\hat{\theta} - \theta_0)|$$

it will be this second approximation which is the better one. This condition will hold in a number of cases. For example, in the case of misspecified models where the distance between the maximum likelihood estimate and the true data generation process will never get small since the true data generation process does not lie on the manifold.

We use this different approach to give us an approximation to the distribution of the maximum likelihood estimate in a small region of the point $\theta$, whether or not this is the true parameter. We call this method the *local approximation to the distribution*.

We shall first complete this calculation in the exponential family case. We have, working in the natural coordinates $\theta$, the formula that

$$-\frac{1}{n}\sum_s \frac{\partial^2}{\partial\theta_i\,\partial\theta_j}\ln p(x_s,\theta) = g_{ij}(\theta)$$

the Fisher information. Hence we get the approximation for small $(\hat\theta - \theta)$

$$(**) \qquad n^{1/2}g(\theta)(\hat\theta-\theta) \approx n^{-1/2}\sum_s \frac{\partial}{\partial\theta_i}\ln p(x_s,\theta).$$

This approximation is then used to locally approximate the asymptotic distribution of the m.l.e. since the asymptotic distribution of the score is known. Thus using $(**)$ we find that

$$n^{1/2}(\hat\theta-\theta) \mathrel{\dot\sim} N\Big(n^{1/2}g^{-1}(\theta)\cdot\big(\mu_i^{\theta_0}\big), \big(\tilde k_{ij}^{\theta_0}\big)^{-1}\Big).$$

In particular the covariance of the asymptotic distribution if $n^{1/2}(\hat\theta-\theta)$ is

$$\Big[g(\theta)\big(\tilde g^{\theta_0}(\theta)\big)^{-1}g(\theta)\Big]^{-1},$$

which is the inverse of the preferred point metric from Example 4, $\tilde k_{ij}^{\theta_0}$. Note that when $\theta$ is $\theta_0$ this local approximation reduces to the standard one.

We can work in better coordinates for this preferred point metric $\tilde k_{ij}^{\theta_0}$, that is, its affine coordinates which are the expected $\mu$ coordinates. We have already seen that in this parametrisation the preferred point metric has the constant form $g^{ij}(\theta_0)$. Further the mean of the normal approximation will transform to be $n^{1/2}\mu$, since from Amari (1985) the change of basis matrix is given by the Fisher information matrix. Thus in this coordinate system the variance will be a constant and the mean will simply be a translation from the preferred point for each different point of evaluation, that is, the local approximation will be

$$n^{1/2}(\hat\mu-\mu) \mathrel{\dot\sim} N\big(-n^{1/2}\big(\mu_i^{\theta_0}\big), g^{ij}(\theta_0)\big).$$

We can also ask the question what formulation do we get outside the full exponential family case? The approximation $(*)$ is based on the Taylor expansion

$$0 = \sum_s \frac{\partial}{\partial\theta_i}\ln p(x_s,\theta) + (\hat\theta-\theta)^j\sum_s \frac{\partial^2}{\partial\theta_i\,\partial\theta_j}\ln p(x_s,\theta) + O\Big(\big(\hat\theta-\theta\big)^2\Big),$$

which is not geometrically well behaved since it depends on the coordinates used, as is pointed out in Barndorff-Nielsen (1987) or Murray and Rice (1987). It is more natural to use the covariant version of Taylor's theorem with respect to the metric we have on the space of the scores. This would give the formula

$$0 = \sum_s \frac{\partial}{\partial\theta_i} \ln p(x_s, \theta)$$
$$+ (\hat\theta - \theta)^j \sum_s \left[ \frac{\partial^2}{\partial\theta_i\,\partial\theta_j} \ln p(x, \theta) - \tilde g^{\theta_0 rs} \tilde\Gamma^{\theta_0}_{ijs} \frac{\partial}{\partial\theta_r} \ln p(x, \theta) \right]$$
$$+ O\Big((\hat\theta - \theta)^2\Big).$$

For large values of $n$ we can use the approximation

$$\frac{1}{n} \sum_s \left[ \frac{\partial^2}{\partial\theta_i\,\partial\theta_j} \ln p(x_s, \theta) - \tilde g^{\theta_0 rt} \tilde\Gamma^{\theta_0}_{ijt}(\theta) \frac{\partial}{\partial\theta_r} \ln p(x_s, \theta) \right]$$
$$\approx \mathbf{E}_{p(x,\theta_0)} \left[ \frac{\partial^2}{\partial\theta_i\,\partial\theta_j} \ln p(x, \theta) - \tilde g^{\theta_0 rt} \tilde\Gamma^{\theta_0}_{ijt}(\theta) \frac{\partial}{\partial\theta_r} \ln p(x, \theta) \right] = \tilde h^{\theta_0}_{ij}(\theta).$$

Note that in the case of a full exponential family this approximation will reduce to $(**)$ and the same analysis follows through.

Thus for large $n$ and small $(\hat\theta - \theta)$ we would find the inverse of the covariance of this distribution is given by the metric,

$$\tilde l^{\phi} = \tilde h^{\phi}_{is} (\tilde g^{\phi})^{rs} \tilde h^{\phi}_{jr}$$

as in Theorem 4. We have shown that in this general case we have the duality between the preferred point metric for the score and the metric for the local approximation of the distribution of the m.l.e. This duality corresponds exactly to that between the $+1$- and $-1$-connections when evaluated at the preferred point. We sum up the key points in the above developments in the following theorem.

THEOREM 5. *Let $\theta$ denote any point in a parametric family $\{p(x, \theta)\}$ and let $\phi = \theta_0 \in \Theta$ denote the point which is the true data generation process.*

(i) *The covariance of the (asymptotic) distribution of the score vector at $\theta$ is given by the preferred point metric*

$$n \cdot \tilde g^{\theta_0}_{ij}(\theta) = n \cdot \mathbf{E}_{p(x,\theta_0)} \left[ \left( \frac{\partial}{\partial\theta_i} \ln p(x, \theta) - \mu_i \right) \left( \frac{\partial}{\partial\theta_j} \ln p(x, \theta) - \mu_j \right) \right]$$

*whose homogeneous connection is Amari's $+1$-connection.*

(ii) *If the family is a full exponential family, then the local approximation to the distribution of $n^{1/2}(\hat\theta - \theta)$ around $\theta$ is*

$$n^{1/2}(\hat\theta - \theta) \mathrel{\dot\sim} N\Big( n^{1/2} g^{-1}(\theta) \cdot (\mu^{\theta_0}_i), \big(\tilde k^{\theta_0}_{ij}\big)^{-1} \Big).$$

*It covariance is the inverse of the preferred point metric*

$$\tilde{k}^{\theta_0}(\theta) = g(\theta)\left(\tilde{g}^{\theta_0}(\theta)\right)^{-1} g(\theta),$$

*where g is the Fisher information. The homogeneous connection for this metric is Amari's* $-1$*-connection.*

(iii) *In the full exponential family both these metrics are flat and their affine parametrisations are the natural and expected parameters, respectively.*

(iv) *In the expected coordinate system the local approximation of the distribution of* $n^{1/2}(\hat{\theta} - \theta)$ *has the constant variance form*

$$n^{1/2}(\hat{\mu} - \mu) \dot{\sim} N\left(-n^{1/2}\left(\mu_i^{\theta_0}\right), g^{ij}(\theta_0)\right).$$

(v) *In a general parametric family the distribution of* $n^{1/2}(\hat{\theta} - \theta)$ *which is constructed using the score at* $\theta$ *is given by*

$$n^{1/2}(\hat{\theta} - \theta) \dot{\sim} N\left(n^{1/2} g^{-1}(\theta) \cdot \left(\mu_i^{\theta_0}\right), \left(\tilde{l}_{ij}^{\theta_0}\right)^{-1}\right),$$

*where*

$$\tilde{l}_{ij}^{\theta_0}(\theta) = \tilde{h}_{is}^{\theta_0}\left(\tilde{g}^{\theta_0}\right)^{rs} \tilde{h}_{rj}^{\theta_0}$$

*and*

$$\tilde{h}_{ij}^{\phi}(\theta) = \mathbf{E}_{p(x,\phi)}\left[\frac{\partial^2}{\partial\theta_i\,\partial\theta_j}\ln p(x,\theta) - \tilde{g}^{\phi rs}\tilde{\Gamma}_{ijr}^{\phi}\frac{\partial}{\partial\theta_s}\ln p(x,\theta)\right],$$

$\tilde{\Gamma}_{ijr}^{\phi}$ *are the Christoffel symbols for* $\tilde{g}^{\phi}(\theta)$ *the preferred point metric in* (i). *Again the homogeneous connection for this preferred point metric is Amari's* $-1$*-connection.*

(vi) *In the full exponential family,* $\tilde{l}^{\theta_0}(\theta)$ *is flat for each* $\theta_0$ *and the* $\mu$*-coordinates are affine. The homogeneous connection for this preferred point geometry is Amari's* $-1$*-connection.*

**9. Higher order extensions of statistical manifolds.** In this section we demonstrate a natural way to extend the definition of a statistical manifold to higher order via preferred point theory and note the consequent connections with string theory as developed by Barndorff-Nielsen (1988) and co-workers.

We consider first the $(M, g, \nabla)$ form of a statistical manifold. A natural way of extending the definition of a statistical manifold structure beyond third order is to take the higher order homogeneous structures given by a preferred point geometry. For each preferred point $\phi$ consider the higher order covariant derivatives at $\theta$. We obtain the homogeneous structure by simply calculating these on the *diagonal* where $\theta$ is set to $\phi$. Thus we can propose the following definition for the generalisation of a statistical manifold.

DEFINITION. Denote the $k$th order covariant derivative induced by the Levi–Civita connection of the preferred point metric $g^{\phi}$ by

$$\left(\nabla^{\phi}\right)^{(k)}(\theta)$$

and the corresponding homogeneous object by

$$\nabla^{(k)}(\theta) = (\nabla^{\theta})^{(k)}(\theta).$$

DEFINITION. We define a $k$th order statistical manifold generated by the preferred point geometry $(M, g^{\phi})$ to be the $k$-tuple

$$(M, g, \nabla, \dots, \nabla^{(k-2)}(\theta)).$$

One of the most important uses of differential geometry in statistics has been in its application of methods which are invariant under a change of coordinates or reparametrisation. This property can be seen as the defining property of geometry. A particular consequence of using these invariant methods has been highly developed in Barndorff-Nielsen (1988) with his use of *strings*. For example, one use of these new geometric objects is the construction of invariant Taylor series and the connected problem of choosing coordinates around a point in the manifold. Murray and Rice (1987) and Murray (1987) give a good account of this application, and demonstrate the importance of the nonmetric $\alpha$-connection in this theory. It is therefore clear that the $k$th order statistical manifold will contain enough information to calculate the first $k$ terms of a covariant Taylor expansion around $\theta$ working under the assumption that $\theta$ is the true parameter.

We can also consider extending the alternative $(M, g, T)$ statistical manifold structure to higher order. We recall that the skewness was defined as the covariant derivative of the difference between the homogeneous metric and the preferred point one. Thus the higher order generalisation would include tensors which give the higher order covariant derivatives of this difference. Also it would include the covariant derivatives of the difference between all order homogeneous structures and their corresponding preferred point versions. We do not give an explicit form for the $k$th-order version here since we feel the definition above is more natural. However Barndorff-Nielsen (private correspondence) has proposed that two tensors are required to extend the statistical manifold structure to fourth order via (expected) yoke theory. These tensors do not have a classical interpretation. We now show how to generate both these tensors from preferred point geometry and how it gives them direct geometric interpretations and also allows a coordinate free derivation.

In Amari's statistical manifold the next important correction term will be the covariant derivative of the difference between the Christoffel symbols of the homogeneous connection and the preferred point connection. This is simply twice the covariant derivative of the skewness. Thus this correction term is

$$\nabla^{\phi} T(\theta)|_{\theta=\phi} = \nabla^{+1} T(\theta)|_{\theta=\phi}$$

since on the $\phi = \theta$ diagonal we have the identification

$$\nabla^{\phi} f(\theta)|_{\theta=\phi} = \nabla^{+1} f(\theta)|_{\theta=\phi}.$$

The right-hand side is one of the tensors which Barndorff-Nielsen identifies, see Barndorff-Nielsen, Blæsild, Pace and Salvan (1990) or Barndorff-Nielsen (1989).

We have shown above that the first order correction term for the preferred point geometry, $h^{\phi}_{ij}$, is zero. Therefore it is natural to look at the second derivative of the difference between the homogeneous metric and the preferred point one. We calculate this evaluated at $\phi$ to be

$$\mathbf{E}_{p(x,\phi)}\left[\partial^2_{ij} \ln p(x,\phi)\partial^2_{rs} \ln p(x,\phi) + \partial^2_{ij} \ln p(x,\phi)\, \partial_r \ln p(x,\phi)\, \partial_s \ln p(x,\phi)\right] - g^{\phi lm}\Gamma^{+1}_{ijl}\Gamma^{-1}_{rsm}$$

and this agrees with the second of Barndorff-Nielsen's fourth order tensors, which he denotes by $\mathbf{T}_{ij;rs}$ see Barndorff-Nielsen, Blæsild, Pace and Salvan (1990).

It will be interesting to develop invariant asymptotic expansions following the higher order preferred point route and compare these with the work of Barndorff-Nielsen and others.

**10. Conclusion and further work.** In this paper we have developed the theory of preferred point geometry and its application to statistical inference. In doing so we have extended the existing notions of a statistical manifold and shown how this may be developed from our preferred point structures. We have provided a clear theoretical basis for the nonmetric connections used previously by Amari which may now be seen as particular preferred point metric connections. In doing so we have provided a formal basis for a statistical methodology which rests on the need to condition inference on some particular point in the parameter space. We have explored the duality inherent in the statistical manifolds structure and shown that in the expected geometry case it corresponds to a duality between the maximum likelihood estimate and the score vector.

The statistical interest of preferred point manifolds is by no means limited to the light they throw on statistical manifolds and Amari's expected geometry. In this final section we briefly indicate further work in progress.

An obvious question to ask is the following. Are there situations in which natural preferred point manifolds exist which approximate to (ideally, coincide with) Barndorff-Nielsen's observed geometries? The answer is not entirely straightforward. One approach which we are pursuing is, following Barndorff-Nielsen, to condition on an ancillary in defining a preferred point metric but, unlike him, to then take expectations with respect to the corresponding conditional distribution obtained under $\phi$. See also subsection 7.2.2 of McCullagh (1987).

Turning to more general applications of preferred point geometry, consider first the fundamental asymmetry between null and alternative hypotheses. This asymmetry finds no expression in homogeneous geometries, whereas preferred point geometries are ideally suited for this purpose. This aspect is developed in another paper on asymmetry and the differential geometry of parameter spaces, Critchley, Marriott and Salmon (1991a). In this paper we also examine the role of divergence functions such as the Kullback Leibler distance which can be shown to be compatible in a certain natural sense with particular preferred point metrics. There are understandably clear strong links here with Barndorff-Nielsen's yoke theory. There is the possibility of developing a preferred point geometry for nonparametric or "distribution free" statistics when the particular manifold is embedded in a higher dimensional function space.

A major feature of preferred point geometries is their Riemannian nature. This allows us to speak of *geodesic distances* and not just projections along geodesics curves. Indeed, given that we can establish a preferred point manifold in a way that is statistically natural, there are grounds to believe that the geometrically natural quantity of geodesic distance will serve as a useful test statistic, see Critchley, Marriott and Salmon (1991b).

## REFERENCES

AMARI, S.-I. (1985). *Differential Geometrical Methods in Statistics. Lecture Notes in Statist.* **28**. Springer, Berlin.

AMARI, S.-I. (1987). Differential geometrical theory of statistics. In *Differential Geometry in Statistical Inference* 19–94. IMS, Hayward, CA.

BARNDORFF-NIELSEN, O. (1986). Strings, tensorial combinants and Bartlett adjustments. *Proc. Roy. Soc. London Ser. A* **406** 127–137.

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

BARNDORFF-NIELSEN, O. (1988). *Parametric Statistical Models and Likelihood*. Springer, London.

BARNDORFF-NIELSEN, O. (1989). Invited contribution to the discussion of R. E. Kass: The geometry of asymptotic inference. *Statist. Sci.* **4** 222–227. [Corrigendum: *Statist. Sci.* **5** (1990) 370.]

BARNDORFF-NIELSEN, O., BLÆSILD P., PACE, L. and SALVAN, A. (1990). Formulas for asymptotic statistical calculations. Preprint, Univ. Århus.

BARNDORFF-NIELSEN, O., COX, D. R. and REID, N. (1986). The role of differential geometry in statistical theory. *Internat. Statist. Rev.* **54** 1 83–96.

CHENTSOV, N. N. (1972). *Statistical Decision Rules and Optimal Inference*. Nuaka, Moscow. [In Russian; English translation (1978) Amer. Math. Soc., Providence, RI.]

COX, D. R. (1961). Tests of separate families of hypotheses. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* 105–123. Univ. California Press, Berkeley.

COX, D. R. and HINKLEY, D. V. (1974). *Theoretical Statistics*. Chapman and Hall, London.

CRITCHLEY, F., MARRIOTT, P. K. and SALMON, M. (1991a). Preferred point geometry and the local differential geometry of the Kullback-Leibler divergence. Preprint. Univ. Warwick.

CRITCHLEY, F., MARRIOTT, P. K. and SALMON, M. (1991b). The geometry of the maximum likelihood estimate. Preprint, Univ. Warwick.

EFRON, B. (1975). Defining the curvature of a statistical problem. *Ann. Statist.* **3** 1189–1242.

EGUCHI, S. (1983). Second order efficiency of minimum contrast estimators in a curved exponential family. *Ann. Inst. Statist. Math.* **36A** 199–206.

GOURIEROUX, C., MONFORT, A. and TROGNON, A. (1984). Pseudo-maximum likelihood methods: Theory. *Econometrica* **52** 681–700.

KASS, R. E. (1989). The geometry of asymptotic inference (with discussion). *Statist. Sci.* **4** 188–234.

LAURITZEN, S. L. (1987). Statistical manifolds. In *Differential Geometry in Statistical Inference* 163–216. IMS, Hayward, CA.

MCCULLAGH, P. (1987). *Tensor Methods in Statistics*. Chapman and Hall, London.

MCCULLAGH, P. and COX, D. R. (1986). Invariants and likelihood ratio statistics. *Ann. Statist.* **14** 1419–1430.

MURRAY, M. K. (1988). Coordinate systems and Taylor series in statistics. *Proc. Roy. Soc. London Ser. A* **415** 445–452.

MURRAY, M. K. and RICE, J. W. (1987). On differential geometry in statistics. Research report, School Maths. Sci., Flinders Univ., South Australia.

SPIVAK, M. (1970). *A Comprehensive Introduction to Differential Geometry*. Publish or Perish, Berkeley, CA.

WHITE, H. (1982). Maximum likelihood estimation in misspecified models. *Econometrica* **50** 1–26.

FRANK CRITCHLEY
DEPARTMENT OF STATISTICS
UNIVERSITY OF WARWICK
COVENTRY CV4 7AL
UNITED KINGDOM

PAUL MARRIOTT
DEPARTMENT OF MATHEMATICAL AND COMPUTING SCIENCES
UNIVERSITY OF SURREY
GUILDFORD GU2 5XH
UNITED KINGDOM

MARK SALMON
DEPARTMENT OF ECONOMICS
EUROPEAN UNIVERSITY INSTITUTE
BADIA FIESOLANA
I-50100, FIRENZE
ITALY