



EUI WORKING PAPERS IN ECONOMICS

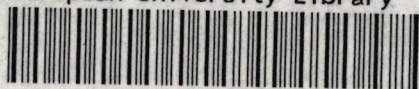
EUI Working Paper ECO No. 91/51

Preferred Point Geometry and Statistical Manifolds

FRANK CRITCHLEY, PAUL MARRIOTT
and
MARK SALMON

European University Institute, Florence

European University Library



3 0001 0016 8183 4

EUROPEAN UNIVERSITY INSTITUTE, FLORENCE
ECONOMICS DEPARTMENT

EUI Working Paper ECO No. 91/51

**Preferred Point Geometry
and Statistical Manifolds**

FRANK CRITCHLEY, PAUL MARRIOTT
and
MARK SALMON

BADIA FIESOLANA, SAN DOMENICO (FI)

All rights reserved.
No part of this paper may be reproduced in any form
without permission of the authors.

© Frank Critchley, Paul Marriott and Mark Salmon
Printed in Italy in October 1991
European University Institute
Badia Fiesolana
I-50016 San Domenico (FI)
Italy

Preferred Point Geometry and Statistical Manifolds.

by

Frank Critchley, Paul Marriott and Mark Salmon[†].

Departments of Statistics and Economics,
University of Warwick,

and

[†]Department of Economics,
European University Institute,
Florence.

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. all points θ in Θ are treated equally) and preferred point ones (e.g. when one point, is isolated i.e. if θ_0 is the true parameter). Although preferred point geometry is applicable generally in statistics, we focus here on its ability to encompass and extend the theoretical structure of Statistical Manifolds developed by Lauritzen(1987), in particular to Amari's expected geometry. A symmetric condition characterises when a preferred point geometry both subsumes a Statistical Manifold and simultaneously generalises it to arbitrary order; there are corresponding links to Barndorff-Nielsen's strings. The rather unnatural mixing of metric and non-metric 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.

American Mathematical Society 1980 subject classification.
Primary: 53B99; Secondary: 62F05, 62F12.

This work has been supported by an ESRC grant on 'Geodesic Inference, Encompassing and Preferred point Geometries in Econometrics'.

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 (1987). 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 parameterisation invariance within statistical inference. The relationship has already shown itself to be productive, for instance 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.

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 include string theory as introduced by Bandorff-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!

Secondly, 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, for example when all points θ in the parameter space Θ share the possibility of being the unknown θ_0 giving rise to the observed data. From other points of view it is not; frequently in statistics a single point has a special or 'preferred' status in Θ which could for instance correspond to the (hypothesised) true value or the 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 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 for 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 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 non-metric 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* ∇ 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 ∇ 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 all the connections used in Statistics are in fact metric connections and to show how the metrics which generate these connection are statistically very natural.

A connection ∇ 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 ∇ are Γ 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 ∂_k and ∂_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

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

when in fact the two formulas agree.

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 ∇ . 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 defined inductively.

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 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 (non-metric) connections, ∇^α , called α -connections. For each $\alpha \in \mathbf{R}$ its connection is defined by the Christoffel symbols given by

$$\Gamma_{ijk}^\alpha = \Gamma_{ijk}^0 - \frac{\alpha}{2} T_{ijk}$$

where Γ^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 ∇ and ∇^* are dual (or conjugate) with respect to a metric, $\langle \cdot, \cdot \rangle$, 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 ∇ if and only if it is flat with respect to ∇^* .

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

Definition. A Statistical manifold (M, g, ∇) is a manifold, M , with a metric g , and a torsion free connection ∇ whose *dual connection* ∇^* 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

$$\Gamma_{ijk}^0 - \frac{1}{2}T_{ijk} \quad \text{and} \quad \Gamma_{ijk}^0 + \frac{1}{2}T_{ijk}, \text{ respectively.}$$

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 (expected) Fisher information and T , the skewness, is defined by

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

Amari (1985) and (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 data generation process. However it does not reflect the fact that in any particular problem only one of them represents the true parameter.

(b) Statistical manifolds mix metric and nonmetric connections within a single structure. This is less than natural and unusual from a geometric perspective.

(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 enquire 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 inter alia White (1982), Gouriéroux,

Monfort and Trognon (1984)). In addition the preferred point structure enables us to consider problems of misspecification, and testing within the context of separate families following Cox (1961), in a proper geometric framework.

In this section we extend the existing notion of a Riemannian manifold by adapting its geometric structure to recognise the importance of a preferred point thereby bringing the geometric basis for inference into line with statistical methodology.

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

When considering a specific preferred point, θ_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 θ depends on the value of both θ and ϕ . However once the value of the preferred point is fixed at θ_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

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

This metric is just the second moment of the score vector with respect to the distribution labelled by ϕ . 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 θ is evaluated at the preferred point ϕ 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

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

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 a 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.

$$\tilde{g}_{ij}^\phi(\theta) = E_{p(x,\phi)} \left[\left(\frac{\partial}{\partial \theta_i} \ln p(x, \theta) - E_{p(x,\phi)} \left[\frac{\partial}{\partial \theta_i} \ln p(x, \theta) \right] \right) \cdot \left(\frac{\partial}{\partial \theta_j} \ln p(x, \theta) - E_{p(x,\phi)} \left[\frac{\partial}{\partial \theta_j} \ln p(x, \theta) \right] \right) \right]$$

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

$$\tilde{\Gamma}_{ijk}^\phi(\theta) = E_{p(x,\phi)} \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln p(x, \theta) \cdot \frac{\partial}{\partial \theta_k} \ln p(x, \theta) \right] - E_{p(x,\phi)} \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln p(x, \theta) \right] E_{p(x,\phi)} \left[\frac{\partial}{\partial \theta_k} \ln p(x, \theta) \right]$$

We can immediately see that by evaluating θ at the preferred point they agree with the +1-connection of Amari, as in the non-central 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, since away from the preferred point

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

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

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

We could however also generalise the second derivative form

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

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 $\theta = \phi$:

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

where $g^{\phi rs}$ is the inverse of g_{rs}^{ϕ} the preferred point metric from Example 1 and Γ^{ϕ} is its Christoffel symbol. In this case the Levi-Civita connection of h_{ij}^{ϕ} 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 (see White (1982) or Gourieroux, Monfort and Trognon (1984)). If θ lies outside our manifold and the maximum likelihood estimate on the manifold converges to θ^* , then the form of the information matrix at θ^* is

$$E_{\theta} \left[\frac{\partial^2}{\partial \theta_s \partial \theta_i} \ln p(x, \theta^*) \right] \left[E_{\theta} \left[\frac{\partial}{\partial \theta_s} \ln p(x, \theta^*) \right] \frac{\partial}{\partial \theta_i} \ln p(x, \theta^*) \right]^{-1} E_{\theta} \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln p(x, \theta^*) \right]$$

which may be compared with the corresponding preferred point metrics developed below.

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^{ϕ} and h^{ϕ} 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 $\theta = \phi$. Theorems 3 and 4 below show further that they are -1-compatible with Amari's expected geometry. 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.

In Section 3 we noted two different ways of viewing a Statistical manifold, either as (M, g, ∇) 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, ∇) .

5.1 Homogeneous structures.

Any preferred point geometry $(M, g^*(\theta))$ 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 θ and ϕ vary independently. We can recover a p -dimensional geometric structure on M by choosing some fixed value of the preferred point θ_0 and looking at the Riemannian manifold $(M, g^{\theta_0}(\theta))$. However this is not the only way of recovering a p -dimensional structure. Consider Figure 1, the diagram denotes the preferred point manifold as $M \times M$ with the vertical axis representing the range in the preferred point ϕ and the horizontal axis the range in the parameter θ . Thus the p -dimensional structure defined by fixing a value of the preferred point at θ_0 is represented by a horizontal line. Consider the p -dimensional structure defined by the diagonal $\theta = \phi$.

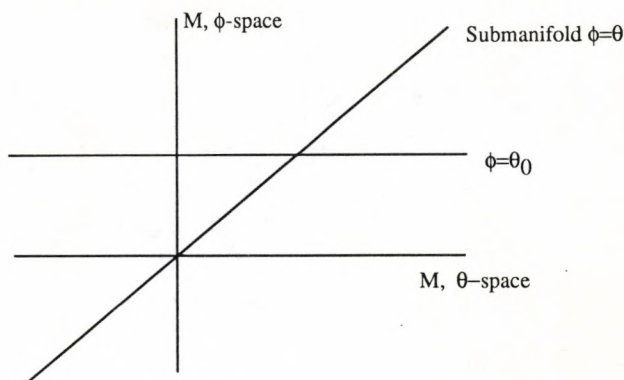


Figure 1.

Definition. A p -dimensional substructure of $(M, g^*(\theta))$ can be defined by taking geometric objects in the preferred point geometry and restricting to the diagonal $\phi = \theta$, as ϕ 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.

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

Example 5. 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 theorem 1 below, how a Statistical manifold can be seen as a 'homogeneous connection geometry' of a preferred point manifold.

Definition. Let us define $\nabla^\theta(\theta)$ to be the Levi-Civita connection of $g^\theta(\theta)$ evaluated at θ , and define $\bar{\nabla}(\theta)$ to be the Levi-Civita connection of the homogeneous Riemannian metric.

Theorem 1. (i) Let $(M, g^\theta(\theta))$ 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 iff

$$T(\theta) = 2(\bar{\nabla}(\theta) - \nabla^\theta(\theta))$$

which is a three 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 need to 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 θ_0 we know the transformation rule for the metric $g^{\theta_0}(\theta)$ is given by

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

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 θ . 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 θ_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) \diamond

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.

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 coordinate system ψ which is affine for $g^\phi(\theta)$, for all ϕ . Then for each ϕ the preferred point connection, $\nabla^\phi(\theta)$, agrees with the homogeneous connection, $\nabla^\theta(\theta)$, at all points θ .

Proof. Working in ψ -coordinates, the Christoffel symbols for the homogeneous connection at θ are those of the preferred point connection when θ is the preferred point. These are zero since the Christoffel symbols for $\nabla^\phi(\theta)$ are identically zero for all θ and ϕ in the ψ -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. \diamond

Example 6. 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 or connection 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

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

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, θ_0 the two metrics agree. They only differ at other points. It is therefore in their derivative at θ_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

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

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

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

Theorem 2. $\tilde{T}(\theta)$ is a homogeneous three 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, $(\theta \rightarrow \psi)$, p transforms as

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

thus $\frac{\partial}{\partial \theta} p_{ij}^{\theta_0}(\theta)$ 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 $g_\psi(\theta)$ and evaluating at $\theta = \theta_0$ we see that

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

where Γ 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. \diamond

Example 7. 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

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

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

$$\Gamma_{ijk}^0 - \frac{1}{2} (\tilde{T}_{kij} + \tilde{T}_{jki} - \tilde{T}_{ijk})$$

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) , i.e. 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, ∇) where ∇ 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, ∇) .

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^{ϕ}) generate the Statistical manifold (M, g, ∇) assuming the symmetry conditions stated above. We call ∇ , the homogeneous connection, the +1-connection of this structure, following Lauritzen. Let

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

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

Theorem 3. With the above definition $k_{ij}^{\theta_0}(\theta)$ is a preferred point metric whose connection at $\theta=\theta_0$ equals the dual connection to ∇ .

Proof. k^{θ_0} is a preferred point metric since it is an asymmetric two tensor which is positive definite when evaluated at the preferred point. Consider differentiating and evaluating at $\theta=\theta_0$.

$$\begin{aligned}\frac{\partial}{\partial \theta_k} k_{ij}^{\theta_0}|_{\theta_0=\theta} &= \frac{\partial}{\partial \theta_k} (g_{is}(\theta) g_{jt}(\theta) g^{\theta_0 st}(\theta))|_{\theta_0=\theta} \\ &= \frac{\partial}{\partial \theta_k} (g_{is}(\theta)) \delta_j^s + \frac{\partial}{\partial \theta_k} (g_{jt}(\theta)) \delta_i^t + g_{is}(\theta) g_{jt}(\theta) \frac{\partial}{\partial \theta_k} (g^{\theta_0 st}(\theta))\end{aligned}$$

Now

$$g_{rs}^{\theta_0}(\theta) g^{\theta_0 st}(\theta) = \delta_r^t$$

So differentiating gives

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

At $\theta=\theta_0$

$$\begin{aligned}\frac{\partial}{\partial \theta_k} (g^{\theta_0 st}(\theta)) &= -\frac{\partial}{\partial \theta_k} (g_{ij}^{\theta_0}(\theta)) g^{si}(\theta) g^{tj}(\theta) \\ &= -\left(\frac{\partial}{\partial \theta_k} (g_{ij}(\theta)) - T_{ijk}\right) g^{si}(\theta) g^{tj}(\theta)\end{aligned}$$

Hence substituting into the original equation we get

$$\frac{\partial}{\partial \theta_k} k_{ij}^{\theta_0} = \frac{\partial}{\partial \theta_k} (g_{ij}(\theta)) + T_{ijk}$$

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

$$\Gamma_{ijk}^0 + \frac{1}{2} T_{ijk}$$

in other words we have the dual connection to ∇ since its Christoffel symbols are $\Gamma_{ijk}^0 - \frac{1}{2} T_{ijk}$. \diamond

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 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^ϕ be a preferred point metric which generates Amari's expected Statistical manifold (M, g, ∇) where g is the Fisher information, and ∇ is the +1-connection. Then

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

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

Further,

$$l_{ij}^\phi = h_{is}^\phi (g^\phi)^{rs} h_{jr}^\phi$$

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

Proof. By calculation

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

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

At $\theta = \phi$ we see that

$$\begin{aligned} \frac{\partial}{\partial \theta_k} h_{ij}^\phi(\theta) \Big|_{\theta=\phi} &= -E_{p(x, \phi)} \left[\frac{\partial^3}{\partial \theta_i \partial \theta_j \partial \theta_k} \ln p(x, \phi) \right] + \Gamma_{ijs}^\phi(\phi) \cdot E_{p(x, \phi)} \left[\frac{\partial^2}{\partial \theta_r \partial \theta_k} \ln p(x, \phi) \right] \cdot g^{\phi rs} \\ &= -E_{p(x, \phi)} \left[\frac{\partial^3}{\partial \theta_i \partial \theta_j \partial \theta_k} \ln p(x, \phi) \right] - \Gamma_{ijk}^{+1}(\phi) \end{aligned}$$

$$\begin{aligned}
 &= -E_{p(x,\phi)} \left[\frac{\partial^3}{\partial \theta_i \partial \theta_j \partial \theta_k} \ln p(x,\phi) \right] - E_{p(x,\phi)} \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln p(x,\phi) \right] \frac{\partial}{\partial \theta_k} \ln p(x,\phi) \\
 &= -\frac{\partial}{\partial \theta_k} E_{p(x,\theta)} \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln p(x,\theta) \right] \Big|_{\theta=\phi} = \frac{\partial}{\partial \theta_k} g(\theta) \Big|_{\theta=\phi}
 \end{aligned}$$

Thus its homogeneous connection agrees with the 0-connection.

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

We note that if g^ϕ and h^ϕ are any two preferred point metrics generating the +1 and 0-connections, then l^ϕ constructed as above generates the -1-connection. Again, if g_1^ϕ and g_2^ϕ are preferred point metrics generating the α_1 and α_2 -connections then, for any real number λ , $\lambda g_1^\phi + (1-\lambda)g_2^\phi$ 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 Theorem 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^ϕ, h^ϕ, k^ϕ) and (g^ϕ, h^ϕ, l^ϕ) with g^ϕ as in Example 1, and the two further triples that result by replacing throughout g^ϕ by \tilde{g}^ϕ 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 2x2 classification of possibilities:

(a) g^ϕ or \tilde{g}^ϕ ? (We have already indicated that \tilde{g}^ϕ is preferred.)

(b) the k or l form of duality?

To make this second choice requires a clearer statistical grasp of what is involved. Before doing so in Section 8 we consider in more detail the preferred point metrics in the particular case of the full exponential family

7. Examples in the full exponential family case.

We consider the 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 θ is the *canonical* or *natural* parameter. Amari (1985) introduces the *expectation* parameter η with general element $\eta_i = E_{p(x,\theta)}[x_i]$. 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 μ^ϕ with general element

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

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

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

In particular, $\mu^\phi(\theta)$ vanishes identically in θ . When ϕ 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 θ -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:

- (1) In the θ -coordinate system, $\tilde{g}_{ij}^\phi(\theta) = g_{ij}(\phi) + \mu_i^\phi \mu_j^\phi$
- (2) In the θ -coordinate system, $\tilde{g}_{ij}^\phi(\theta) = g_{ij}(\phi)$, a constant, independent of θ .
- (3) In the μ -coordinate system, $\tilde{k}_{ij}^\phi(\mu) = g^{ij}(\phi)$, a constant, independent of μ .
- (4) In the θ -coordinate system, $\tilde{h}_{ij}^\phi(\theta) = g_{ij}(\theta)$
- (5) $\tilde{l}_{ij}^\phi(\theta) = \tilde{k}_{ij}^\phi(\theta)$, thus a constant in the μ -coordinates.

In particular for each θ_0 , Example 2, $(M, \tilde{g}_{ij}^{\theta_0}, \theta_0)$, is flat and the natural θ -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 μ -coordinates, are affine.

Proof. (1) and (2). These parts follow by a simple calculation.

(3) Amari (1985) notes that the change of basis matrix for the reparametrisation $\{\theta\} \rightarrow \{\eta\}$ is given by the inverse of the Fisher information matrix. Clearly, that for the reparametrisation $\{\eta\} \rightarrow \{\mu\}$ is minus the identity matrix. But, by definition $\tilde{k}_{ij}^{\phi}(\theta) = g_{is}(\theta) \tilde{g}^{\phi st}(\theta) g_{sj}(\theta)$. Thus in the μ -coordinate system (3) has the form $\tilde{g}^{\phi st}(\theta(\mu))$, i.e., the inverse of the form of (2) in the θ -coordinate system. The result now follows from (2).

(4) Since the metric in (2) is flat, the associated Christoffel symbols vanish in the θ -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,

$$\tilde{h}_{ij}^{\phi}(\theta) = -E_{p(x, \phi)}[-g_{ij}(\theta) - 0] = g_{ij}(\theta)$$

(5) Given (4), this is immediate. ◇

Amari has shown that the exponential family is +1-flat and that the θ -coordinates are +1-affine. We remark that this follows at once from part (2) of the above proposition, when we recall the corollary to Theorem 1, since the θ -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 (3) 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

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

Amari identifies this space with the tangent space to the manifold at the point θ . 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 θ_0 . Under these conditions the score vector for the sample \mathbf{x} will be the sum of the scores for each x_i . 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

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

where $\mu_i^{\theta_0}$ is defined in the previous section and \bar{g}^{θ_0} is the covariance 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 θ is the true parameter. Asymptotically the relationship between the maximum likelihood estimate and the score test is given by the approximation.

$$\sqrt{n}g(\theta)(\hat{\theta} - \theta)^i \approx \frac{1}{\sqrt{n}} \sum_s \frac{\partial}{\partial \theta_i} \ln p(x_s, \theta) \quad (1)$$

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

Denoting the true parameter value by θ_0 consider the distribution of $(\hat{\theta} - \theta)$ for all possible θ . 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 θ_0 .

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

or we could in fact use the score at any other point θ ,

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

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

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

Thus for *small* values of $(\hat{\theta} - \theta)$ we will 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 θ_0 is the true distribution. Here we are using a different condition, simply that if the m.l.e. lies in a small neighbourhood of θ then our approximation will be a good one. Note that if

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

it will be this second approach that provides a more accurate approximation to the distribution of the maximum likelihood estimator than the standard asymptotic approximation. A priori we would expect this condition to hold in a number of cases, for example in the case of misspecified models where $\hat{\theta} - \theta_0$ may never be small since the true data generation process does not lie on the manifold.

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

We shall first complete this calculation the exponential family case. We have, working in the natural coordinates θ , 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 $(\hat{\theta} - \theta)$

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

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 (2) from our preferred point analysis above we find that

$$n^{1/2}(\hat{\theta} - \theta) \div N\left(n^{1/2} g^{-1}(\theta), (\mu_i^{\theta_0}), (\bar{k}_{ij}^{\theta_0})^{-1}\right)$$

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

$$\left[g(\theta)(\tilde{g}^{\theta_0}(\theta))^{-1}g(\theta) \right]^{-1}$$

which is the inverse of the preferred point metric from Example 4, $\tilde{k}_{ij}^{\theta_0}$. Note that when θ is θ_0 this local approximation reduces to the standard one.

We can work in the better coordinates for this preferred point metric, $\tilde{k}_{ij}^{\theta_0}$, that is its affine coordinates, which are the expected μ coordinates. We have already seen that in this parametrisation the preferred point metric has the constant form $\tilde{g}^{ij}(\theta_0)$. Further the mean of the normal approximation will transform to be $n^{1/2}\mu$, since from Amari (1985) the change of basis formula 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, i.e. the local approximation will be

$$n^{1/2}(\hat{\mu} - \mu) \doteq N\left(-n^{1/2}(\mu_1^{\theta_0}), \tilde{g}^{ij}(\theta_0)\right)$$

We can also ask the question what formulation do we get outside the full exponential family case? The approximation (1) is based on the Taylor expansion

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

which is not a geometrically well behaved since it depends on the coordinates used, as is pointed out in Barndorff-Nielsen (1989) 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

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

For large values of n we can use the approximation

$$\frac{1}{n} \sum_s \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln p(x_s, \theta) - \tilde{g}^{\theta_0 rs} \tilde{\Gamma}_{ijs}^{\theta_0}(\theta) \frac{\partial}{\partial \theta_r} \ln p(x_s, \theta) \right] \approx$$

$$E_{p(x, \theta_0)} \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln p(x, \theta) - \tilde{g}^{\theta_0 rs} \tilde{\Gamma}_{ijs}^{\theta_0}(\theta) \frac{\partial}{\partial \theta_r} \ln p(x, \theta) \right] = \tilde{h}_{ij}^{\theta_0}(\theta)$$

Note that in the case of a full exponential family this approximation will reduce to (2) 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

$$\tilde{\Gamma}^\Phi = \tilde{h}_{is}^\Phi (\tilde{g}^\Phi)^{rs} \tilde{h}_{jr}^\Phi$$

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 θ denote any point in a parametric family $\{p(x, \theta)\}$ and let ϕ denote the point which is the true data generation process.

(i) The covariance of the (asymptotic) distribution of the score vector at θ is given by the preferred point metric

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

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 θ is

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

Its covariance is the inverse of the preferred point metric

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

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

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

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

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

where

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

and

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

$\tilde{\Gamma}_{ijr}^{\phi}$ 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, $\tilde{\Gamma}^{\theta_0}(\theta)$ is flat for each θ_0 and the μ -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, ∇) 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, ϕ , consider the higher order covariant derivatives at θ . We obtain the homogeneous structure by simply calculating these on the *diagonal* where θ is set to ϕ . Thus we could 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^{ϕ} by

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

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^{ϕ}) 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 non-metric α -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 θ working under the assumption that θ 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).

We have shown above that the first order correction term for the preferred point geometry h_{ij}^{ϕ} , 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 ϕ to be

$$E_{p(x,\phi)}[\partial_{ij}^2 \ln p(x,\phi) \partial_{rs}^2 \ln p(x,\phi) + \partial_{ij}^2 \ln p(x,\phi) \partial_r \ln p(x,\phi) \partial_s \ln p(x,\phi)] - g^{\phi \text{ lm}} \Gamma_{ijl}^{+1} \Gamma_{rsm}^{-1}$$

and this agrees with the second of Barndorff-Nielsen's fourth order tensors, which he denotes by $T_{ij;rs}$ 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 show how this may be developed from our preferred point structures. We have provided a clear theoretical basis for the non-metric 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.

One obvious development is to consider Barndorff-Nielsen's observed geometry rather than Amari's expected geometry. One possible way of doing this is to work with distributions which are conditional on the value of some ancillary statistic. The preferred point metrics may then be calculated via *conditional* expectations rather than the unconditional used in this paper.

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 Sun-ichi, (1985), *Differential geometrical methods in statistics*, Springer-Verlag Lecture notes in Statistics No.28.
- Amari Sun-ichi,(1987), *Differential Geometrical theory of statistics, Institute of Mathematical Statistics Lecture Notes- Monograph Series Volume 10.*
- Barndorff-Nielsen O., (1988), *Parametric Statistical Models and Likelihood*, Springer-Verlag, London
- Barndorff-Nielsen O., Blæsild P., Pace L. and Salvan A. (1990), *Formulas for Asymptotic Statistical Calculations*, Preprint, University of Århus.
- Barndorff-Nielsen O., Cox D.R. and Reid N., (1986), *The role of Differential Geometry in Statistical Theory*, *International Statistical Review*, 54, 1, 83-96.
- Chentsov N.N. (1972), *Statistical Decision Rules and Optimal Inference*, *Nauka*, Moscow; translated into English (1978), AMS, Rhode Island.

Preferred Point Geometry and Statistical Manifolds

Cox D.R. (1961), Tests of separate families of hypotheses, *Proceedings of the fourth Berkeley symposium in mathematical statistics and probability*, University of California Press, Berkeley CA. 105-123.

Cox D.R. and Hinkley D.V., (1974), *Theoretical Statistics*, London 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*, European University Institute, Florence.

Critchley F., Marriott P.K. and Salmon M (1991b), Geodesic inference, *in preparation*

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, Monfort and Trogon, (1984). Pseudo-maximum likelihood methods: Theory, *Econometrica*, 52, 681-700.

Kass R.E., (1987), Introduction to Differential Geometry in Statistical Inference, *Technical report* 397, Dept. of Statistics, Carnegie-Mellon University.

Lauritzen S.L., (1987), Statistical Manifolds, Differential Geometry in Statistical Inference, *Institute of Mathematical Statistics, Lecture Notes-Monograph Series Volume 10*.

Murray, M.K., (1988) Coordinate systems and Taylor series in statistics, *Proc. Royal Soc. London* A, 415, 445-452.

Murray, M.K. and Rice J.W., (1987) On differential geometry in statistics, *Research report: School of Maths. Sci.*, Flinders Univ. of South Australia.

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

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

Department of Statistics,
University of Warwick,
Coventry CV4 7AL,
United Kingdom.

Department of Economics,
European University Institute,
Badia Fiesolana,
I-50100, Firenze,
Italy.



EUI WORKING PAPERS

EUI Working Papers are published and distributed by the
European University Institute, Florence

Copies can be obtained free of charge – depending on the availability of
stocks – from:

The Publications Officer
European University Institute
Badia Fiesolana
I-50016 San Domenico di Fiesole (FI)
Italy

Please use order form overleaf

Publications of the European University Institute

To The Publications Officer
European University Institute
Badia Fiesolana
I-50016 San Domenico di Fiesole (FI)
Italy

From Name

 Address

- ☐ Please send me a complete list of EUI Working Papers
- ☐ Please send me a complete list of EUI book publications
- ☐ Please send me the EUI brochure Academic Year 1990/91

Please send me the following EUI Working Paper(s):

No, Author
Title:
No, Author
Title:
No, Author
Title:
No, Author
Title:

Date

Signature



Working Papers of the Department of Economics **Published since 1989**

89/370

B. BENSARD/R.J. GARY-BOBO/
 S. FEDERBUSCH
 The Strategic Aspects of Profit Sharing in the
 Industry

89/374

Francisco S. TORRES
 Small Countries and Exogenous Policy Shocks

89/375

Renzo DAVIDDI
 Rouble Convertibility: A Realistic Target

89/377

Elettra AGLIARDI
 On the Robustness of Contestability Theory

89/378

Stephen MARTIN
 The Welfare Consequences of Transaction Costs
 in Financial Markets

89/381

Susan SENIOR NELLO
 Recent Developments in Relations Between the
 EC and Eastern Europe

89/382

Jean GABSZEWICZ/ Paolo GARELLA/
 Charles NOLLET
 Spatial Price Competition With Uninformed
 Buyers

89/383

Benedetto GUI
 Beneficiary and Dominant Roles in
 Organizations: The Case of Nonprofits

89/384

Agustín MARAVALL/ Daniel PEÑA
 Missing Observations, Additive Outliers and
 Inverse Autocorrelation Function

89/385

Stephen MARTIN
 Product Differentiation and Market Performance
 in Oligopoly

89/386

Dalia MARIN
 Is the Export-Led Growth Hypothesis Valid for
 Industrialized Countries?

89/387

Stephen MARTIN
 Modeling Oligopolistic Interaction

89/388

Jean-Claude CHOURAQUI
 The Conduct of Monetary Policy: What have we
 Learned From Recent Experience

89/390

Corrado BENASSI
 Imperfect Information and Financial Markets: A
 General Equilibrium Model

89/394

Serge-Christophe KOLM
 Adequacy, Equity and Fundamental Dominance:
 Unanimous and Comparable Allocations in
 Rational Social Choice, with Applications to
 Marriage and Wages

89/395

Daniel HEYMANN/ Axel LEIJONHUFVUD
 On the Use of Currency Reform in Inflation
 Stabilization

89/400

Robert J. GARY-BOBO
 On the Existence of Equilibrium Configurations
 in a Class of Asymmetric Market Entry Games *

89/402

Stephen MARTIN
 Direct Foreign Investment in The United States

89/413

Francisco S. TORRES
 Portugal, the EMS and 1992: Stabilization and
 Liberalization

89/416

Joerg MAYER
 Reserve Switches and Exchange-Rate Variability:
 The Presumed Inherent Instability of the
 Multiple Reserve-Currency System

89/417

José P. ESPERANÇA/ Neil KAY
 Foreign Direct Investment and Competition in
 the Advertising Sector: The Italian Case

* Working Paper out of print

89/418
Luigi BRIGHI/ Mario FORNI
Aggregation Across Agents in Demand Systems

89/420
Corrado BENASSI
A Competitive Model of Credit Intermediation

89/422
Marcus MILLER/ Mark SALMON
When does Coordination pay?

89/423
Marcus MILLER/ Mark SALMON/
Alan SUTHERLAND
Time Consistency, Discounting and the Returns
to Cooperation

89/424
Frank CRITCHLEY/ Paul MARRIOTT/
Mark SALMON
On the Differential Geometry of the Wald Test
with Nonlinear Restrictions

89/425
Peter J. HAMMOND
On the Impossibility of Perfect Capital Markets

89/426
Peter J. HAMMOND
Perfected Option Markets in Economies with
Adverse Selection

89/427
Peter J. HAMMOND
Irreducibility, Resource Relatedness, and Survival
with Individual Non-Convexities

* * *

ECO No. 90/1**
Tamer BAŞAR and Mark SALMON
Credibility and the Value of Information
Transmission in a Model of Monetary Policy
and Inflation

ECO No. 90/2
Horst UNGERER
The EMS – The First Ten Years
Policies – Developments – Evolution

ECO No. 90/3
Peter J. HAMMOND
Interpersonal Comparisons of Utility: Why and
how they are and should be made

ECO No. 90/4
Peter J. HAMMOND
A Revelation Principle for (Boundedly) Bayesian
Rationalizable Strategies

ECO No. 90/5
Peter J. HAMMOND
Independence of Irrelevant Interpersonal
Comparisons

ECO No. 90/6
Hal R. VARIAN
A Solution to the Problem of Externalities and
Public Goods when Agents are Well-Informed

ECO No. 90/7
Hal R. VARIAN
Sequential Provision of Public Goods

ECO No. 90/8
T. BRIANZA, L. PHLIPS and J.F. RICHARD
Futures Markets, Speculation and Monopoly
Pricing

ECO No. 90/9
Anthony B. ATKINSON/ John
MICKLEWRIGHT
Unemployment Compensation and Labour
Market Transition: A Critical Review

ECO No. 90/10
Peter J. HAMMOND
The Role of Information in Economics

ECO No. 90/11
Nicos M. CHRISTODOULAKIS
Debt Dynamics in a Small Open Economy

ECO No. 90/12
Stephen C. SMITH
On the Economic Rationale for Codetermination
Law

ECO No. 90/13
Elettra AGLIARDI
Learning by Doing and Market Structures

ECO No. 90/14
Peter J. HAMMOND
Intertemporal Objectives

ECO No. 90/15
Andrew EVANS/Stephen MARTIN
Socially Acceptable Distortion of Competition:
EC Policy on State Aid

** Please note: As from January 1990, the EUI Working Papers Series is divided into six sub-series, each series will be numbered individually (e.g. EUI Working Paper LAW No. 90/1).

ECO No. 90/16

Stephen MARTIN
Fringe Size and Cartel Stability

ECO No. 90/17

John MICKLEWRIGHT
Why Do Less Than a Quarter of the
Unemployed in Britain Receive Unemployment
Insurance?

ECO No. 90/18

Mrudula A. PATEL
Optimal Life Cycle Saving
With Borrowing Constraints:
A Graphical Solution

ECO No. 90/19

Peter J. HAMMOND
Money Metric Measures of Individual and Social
Welfare Allowing for Environmental
Externalities

ECO No. 90/20

Louis PHILIPS/
Ronald M. HARSTAD
Oligopolistic Manipulation of Spot Markets and
the Timing of Futures Market Speculation

ECO No. 90/21

Christian DUSTMANN
Earnings Adjustment of Temporary Migrants

ECO No. 90/22

John MICKLEWRIGHT
The Reform of Unemployment Compensation:
Choices for East and West

ECO No. 90/23

Joerg MAYER
U. S. Dollar and Deutschmark as Reserve Assets

ECO No. 90/24

Sheila MARNIE
Labour Market Reform in the USSR:
Fact or Fiction?

ECO No. 90/25

Peter JENSEN/
Niels WESTERGÅRD-NIELSEN
Temporary Layoffs and the Duration of
Unemployment: An Empirical Analysis

ECO No. 90/26

Stephan L. KALB
Market-Led Approaches to European Monetary
Union in the Light of a Legal Restrictions
Theory of Money

ECO No. 90/27

Robert J. WALDMANN
Implausible Results or Implausible Data?
Anomalies in the Construction of Value Added
Data and Implications for Estimates of Price-
Cost Markups

ECO No. 90/28

Stephen MARTIN
Periodic Model Changes in Oligopoly

ECO No. 90/29

Nicos CHRISTODOULAKIS/
Martin WEALE
Imperfect Competition in an Open Economy

* * *

ECO No. 91/30

Steve ALPERN/Dennis J. SNOWER
Unemployment Through 'Learning From
Experience'

ECO No. 91/31

David M. PRESCOTT/Thanasis STENGOS
Testing for Forecastable Nonlinear Dependence
in Weekly Gold Rates of Return

ECO No. 91/32

Peter J. HAMMOND
Harsanyi's Utilitarian Theorem:
A Simpler Proof and Some Ethical
Connotations

ECO No. 91/33

Anthony B. ATKINSON/
John MICKLEWRIGHT
Economic Transformation in Eastern Europe
and the Distribution of Income

ECO No. 91/34

Svend ALBAEK
On Nash and Stackelberg Equilibria when Costs
are Private Information

ECO No. 91/35

Stephen MARTIN
Private and Social Incentives
to Form R & D Joint Ventures

ECO No. 91/36

Louis PHILIPS
Manipulation of Crude Oil Futures

ECO No. 91/37

Xavier CALSAMIGLIA/Alan KIRMAN
A Unique Informationally Efficient and
Decentralized Mechanism With Fair Outcomes

ECO No. 91/38

George S. ALOGOSKOUFIS/
Thanasis STENGOS
Testing for Nonlinear Dynamics in Historical
Unemployment Series

ECO No. 91/39

Peter J. HAMMOND
The Moral Status of Profits and Other Rewards:
A Perspective From Modern Welfare Economics

ECO No. 91/40

Vincent BROUSSEAU/Alan KIRMAN
The Dynamics of Learning
in Mis-Specified Models

ECO No. 91/41

Robert James WALDMANN
Assessing the Relative Sizes of Industry- and
Nation Specific Shocks to Output

ECO No. 91/42

Thorsten HENS/Alan KIRMAN/Louis PHILIPS
Exchange Rates and Oligopoly

ECO No. 91/43

Peter J. HAMMOND
Consequentialist Decision Theory and
Utilitarian Ethics

ECO No. 91/44

Stephen MARTIN
Endogenous Firm Efficiency in a Cournot
Principal-Agent Model

ECO No. 91/45

Svend ALBAEK
Upstream or Downstream Information Sharing?

ECO No. 91/46

Thomas H. McCURDY/
Thanasis STENGOS
A Comparison of Risk-Premium Forecasts
Implied by Parametric Versus Nonparametric
Conditional Mean Estimators

ECO No. 91/47

Christian DUSTMANN
Temporary Migration and the Investment into
Human Capital

ECO No. 91/48

Jean-Daniel GUIGOU
Should Bankruptcy Proceedings be Initiated by a
Mixed Creditor/Shareholder?

ECO No. 91/49

Nick VRIEND
Market-Making and Decentralized Trade

ECO No. 91/50

Jeffrey L. COLES/Peter J. HAMMOND
Walrasian Equilibrium without Survival:
Existence, Efficiency, and Remedial Policy

ECO No. 91/51

Frank CRITCHLEY/Paul MARRIOTT/
Mark SALMON
Preferred Point Geometry and Statistical
Manifolds

ECO No. 91/52

Costanza TORRICELLI
The Influence of Futures on Spot Price
Volatility in a Model for a Storable Commodity