EUI WORKING PAPERS IN ECONOMICS

EUI Working Paper ECO No. 91/53

Preferred Point Geometry and the Local Differential Geometry of the Kullback-Leibler Divergence

FRANK CRITCHLEY, PAUL MARRIOTT
and
MARK SALMON

European University Institute, Florence

Please note

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

EUROPEAN UNIVERSITY INSTITUTE, FLORENCE

ECONOMICS DEPARTMENT

EUI Working Paper **ECO** No. 91/53

Preferred Point Geometry and the Local Differential Geometry of the Kullback-Leibler Divergence

FRANK CRITCHLEY, PAUL MARRIOTT
and
MARK SALMON

BADIA FIESOLANA, SAN DOMENICO (FI)

Printed in Italy in November 1991
European University Institute
Badia Fiesolana
I–50016 San Domenico (FI)
Italy

Preferred point geometry and the local differential geometry of the Kullback-Leibler divergence.

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

Departments of Statistics and Economics,
University of Warwick.
and
[†] Department of Economics,
European University Institute,
Florence.

Wed, Sep 18, 1991

The (asymmetric) Kullback-Leibler divergence function is rationalised as being geometrically a measure of preferred point geodesic distance based on path-length. This distance function is defined not on any particular parametric family but in an infinite dimensional function space in which all our finite dimensional parametric statistical families are embedded. In so doing we generalise results by Amari relating 'α-geodesic projection' to divergence functions. We are forced to consider concepts of flatness and embedding curvature. We develop a new total flatness condition under which our squared geodesic distance corresponds to twice the Kullback-Leibler divegence. We show that the space of densities itself possess a form of curvature which implies in particular that only a subset of the full exponential families may in fact be considered totally flat. We study the infinite dimensional function space of finite measures in which the Kullback-Leibler divergence is most naturally viewed as a metric based measure of distance. We also propose a global measure of curvature which may be compared with the pointwise measures of Efron and Amari.

American Mathematics Society Subject Classification. Primary: 53B99; Secondary 62F05, 62F12.

This work has been surported by ESRC grant 'Geodesic Inference, Ecompassing and Preferred Point Geometries in Econometrics'

The Geometry of the Kullback-Leibler Divergence

1. Introduction.

Ideas of distance in geometry have mostly been developments of the Euclidean axiom that the shortest path between two points is a straight line. The distance between these points is then defined as the length of this line. Following the developments which enable us to define what is meant by a straight line in spaces more complex than Euclid's plane, we find that we pass through most of the history of geometry itself. Indeed, many of the most important strides forward in mathematics occur in this journey and it involves the work of some of the greatest mathematicians: Euclid, Pythagoras's Theorem, Newton's calculus, Gauss's differential geometry, Euler's calculus of variations, through to Einstein's use of geometry in physics. Throughout this long history runs the central theme that we measure the separation of two points by finding the shortest path joining them. This use of minimum path lengths means that intuitive ideas of distance satisfy the following basic, now familiar, axioms. If D(m,n) is the distance from m to n then we would expect that

(1) Positivity: $D(m,n) \geq 0$ with equality if and only if n=m,

(2) Symmetry: $D(m,n)=D(n,m)$,

(3) The triangle inequality: $D(m,n) \leq D(m,p)+D(p,n)$.

Condition (2) follows from the intuitive idea that if there exists a path from m to n then there also exists a return path, from n to m. Condition (3) derives from the idea that if we take the shortest path from m to p followed by the shortest path from p to n then we have taken a path from m to n and since pathlengths are assumed additive we have gone at least as far as the shortest path joining them.

There has also been a natural interest in statistics on how to measure the separation of two density functions, see for instance Rao (1945, 1987), Burbea and Rao (1982), Jeffreys (1948), Bhattacharrya (1943) and Kullback and Leibler (1951). The fundamental role played by Fisher's information matrix in Rao's notion of distance may be contrasted with the Kullback-Leibler divergence function

$$d_{kl}(\theta,\theta') = \mathbf{E}_\theta[\ln p(x,\theta) - \ln p(x,\theta')].$$

which is not based on any explicit metric measure of distance. This function and many other proposed divergence or discrimination functions, are apparently quite different from the more geometric ideas of distance. For example, they do not satisfy conditions (2) and (3) above. These functions do however reflect the asymmetry which is fundamental to statistical inference given the isolation of some particular density as representing either the true data generation process or the maintained hypothesis.

The past fifteen years has seen a substantial development in the relationship between differential geometry and statistics. See for example, the review papers by Barndorff-Nielsen, Cox and Reid (1986) and by Kass (1989). In particular we note the work of Amari (1985) and his construction of an *expected geometry* on a parametric family of density functions. Using this geometry Amari was able to forge some links between the differential geometric concept of a geodesic and some common divergence functions from statistics, including the Kullback-Leibler measure. There also has been work on the 'Euclidean' geometry of the Kullback-Leibler divergence from a non differential geometric approach, see for example Čencov (1972), Csiszar (1975) and Loh (1983). Barndorff-Nielsen (1989) and Blæsild (1988, 1990) have used the concept of a *yoke*, one of which is minus the Kullback-Leibler measure, to generate very general geometric structures. Marriott (1989) defined and introduced a new differential geometric construction called a preferred point geometry, which was further developed in Critchley, Marriott and Salmon (1991), where in particular it was shown how preferred point geometry encompasses Amari's expected α-geometries. Amari's projection theorem and generalised Pythagorean theorem are highlights of the theory of dually flat manifolds, such as the expected α-geometries. However it must be emphasised that since α-geodesic paths on which these theorems are based are non-metic constructions and there is no concept of α-geodesic *distance* unless $\alpha = 0$.

In this paper we show how preferred point geometry gives rise to an asymmetric geometric structure which is particularly relevant to statistics. It can rationalise asymmetric statistical divergence functions, in particular the Kullback-Leibler divergence, as being geometrically natural (squared) distances based on pathlengths. Thus putting metric and non-metric distance concepts on the same theoretical footing enabling us to make rigorous comparisions between these apparently distinct concepts. In so doing we generalise two key theorems of Amari mentioned above and are able to gain a greater understanding of the particular information captured in these alternative approaches. The difference in information may be summarised using measures of curvature that extend Efron's (1975) definition of statistical curvature. Further a number of model selection procedures such as Akaike's information criterion (1973) are based on the Kullback-Leibler notion of distance and hence our development clarifies to some extent how the use of these discrimination measures may be related to formal statistical hypothesis tests and decision theory.

In Section 2 we review some preliminary material covering the required differential geometric and statistical background including Amari's results on α-flatness and α-diveregence. In Section 3 we show how any divergence function can be interpreted locally as a squared geodesic distance in a preferred point geometry and how the concept of a preferred point geometry is considerably more general. In Section 4 we look at different measures of intrinsic and embedding curvature. We study in particular some preferred point metrics and define a strong concept of flatness for parametric families. We show the geometric relationship between the preferred point metrics and the Kullback-Leibler divergence through the embedding of a parameter space in a higher dimensional manifold and in particular show how this 'total flatness' condition is sufficient to force our squared geodesic distances and twice the Kullback-Leibler divergence to agree. This strengthens Amari's result on Kullback-Leibler projection. We further show that the space of densities itself possesses a form of curvature which prevents the extension of Amari's result to its most natural geometric form. In Section 5 we consider some global measures of curvature derived as the difference between the different distances. We consider the inherent curvature of a full exponential family using these new measures demonstrating that only a restricted subset of this family can be considered 'totally flat'. This implies that the Kullback-Leibler divergence will only in very particular cases be able to be treated as a metric based measure of distance. We also look at examples of the curvature of some curved exponential families and compare the new measures of curvature with those of Efron and Amari and explore the relationship to the Kullback-Leibler divergence. In Section 6 we show how preferred point geometry is particularly appropriate to the analysis of the Pythagorean properties of the Kullback-Leibler divergence.

2. Geometric background.

Throughout, $M=\{p(x,\theta)\}$ denotes a finitely parametrised manifold of probability density functions obeying the regularity conditions listed in Amari (1985, page 16)

Divergence Functions.

Following Čencov (1972) and Amari (1985) we adopt the following definition.

Definition. A *divergence function* $d(\theta,\theta')$ is a smooth function on pairs of points in some parametric family $\{p(x,\theta)\}$ which satisfies the following conditions:

(i) $d(\theta,\theta') \geq 0$, and the equality holds when and only when $\theta=\theta'$.

(ii) $\partial_i d(\theta,\theta') = 0$ at $\theta = \theta'$, where $\partial_i = \partial/\partial\theta_i$.

(iii) $\partial_i\partial_j d(\theta,\theta')|_{\theta=\theta'} = g_{ij}(\theta)$, the Fisher information matrix.

These conditions state that a divergence function is locally quadratic in θ and θ', and in the local approximation, the hessian of the divergence function agrees with the Fisher information. Condition (ii) can be achieved with any distance function which satisfies (i) by squaring the function if necessary. Condition (iii) can be achieved by rescaling as long as the hessian of the divergence function is non-singular. There is a clear similarity in this definition to Barndorff-Nielsen's concept of a *normed yoke*. See Barndorff-Nielsen (1989), where condition (iii) is relaxed to be that the hessian must be non-singular.

Some well known examples of divergence functions include:

Kullback-Leibler:

$$d_{kl}(\theta,\theta') = \mathbf{E}_\theta[\ln p(x,\theta) - \ln p(x,\theta')]$$

Hellinger:

$$d_H(\theta,\theta') = 2\int\left(\sqrt{p(x,\theta)} - \sqrt{p(x,\theta')}\right)^2 dP$$

Renyi α-information:

$$d_R^\alpha(\theta,\theta') = \frac{1}{\alpha(\alpha-1)}\log\int\{p(x,\theta)\}^\alpha\{p(x,\theta')\}^{1-\alpha}dP \quad \text{where } 0<\alpha<1$$

A useful transformation of the Helliger divergence is the distance defined by

Bhattacharrya distance:

$$d_B(\theta,\theta') = 2\cos^{-1}\left[1-\frac{1}{4}d_H(\theta,\theta')\right]$$

This does not obey the axioms for a divergence, being a distance rather than a squared distance. However $\tilde{d}_B(\theta,\theta') = \frac{1}{2}d_B^2(\theta,\theta')$ is a divergence.

In all of the above P is some fixed dominating base measure.

As mentioned above one of the important features which distinguishes divergence functions from more geometric ideas of distance is that they do not necessarily have to be symmetric. From a purely geometric point of view of distance this might seem surprising but from a statistical point of view it is natural given, for example, the asymmetric role played in inference by a point null and a point alternative.

Riemannian Geometry.

Definition. A *Riemannian manifold*, (M, g_{ij}), is a manifold and a *metric tensor*, g_{ij}, which at each point θ is a symmetric and positive definite bilinear form on the tangent space at θ, TM_θ. This bilinear form is a smooth function of θ.

The idea of putting a Riemannian structure on a finite dimensional parametric family goes back to Rao (1945). He noticed that the Fisher information is a metric tensor, thus it defines a Riemannian structure.This idea has intuitive appeal through the Cramer-Rao Theorem viewing the

Fisher information as a local measure of distance at the true parameter. In a Riemannian geometry the concept of distance is based on measuring the lengths of curves in the manifold. Given a metric structure g_{ij} we can measure the length of any tangent vector with respect to the metric. Let $\gamma(t)$ be a curve on our manifold. We can then define the path length of $\gamma(t)$, from t=a to the point t=b, by

$$L(a,b) = \int_a^b \sqrt{g_{ij} \cdot \frac{d}{dt}\gamma^i(t)\frac{d}{dt}\gamma^j(t)}\, dt$$

A *geodesic* between two points is defined to be a curve joining them with the shortest pathlength, and this length is called the *geodesic distance*. We ignore the complications of this apparently simple construction for the moment: for details concerning these facts particularly the existence and uniqueness problem see Spivak (1970). However we note that all the geometry considered in this paper is local thus the above complications do not arise. Under the assumption of existence and uniqueness the geodesic distance between two points a and b, denoted by D(a,b), is symmetric and obeys the triangle inequality, i.e. conditions (2) and (3) above.

Given a coordinate system θ 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 the geodesics of the Levi-Civita connection are curves of shortest length with respect to the metric.

Statistical Manifolds.

One of the main developments in the application of differential geometry to statistics has been the realisation that a Riemannian structure is insufficient to contain all the statistical information in a parametric family. The first implicit use of a *connection* not defined from the Fisher metric was by Efron (1975) in his analysis of the statistical curvature of a curved exponential family. This was explicitly recognised by Čencov (1972) and by Dawid (1975) who introduced the use of a one parameter family of connections. Amari (1985) developed the application of these connections (α-connections) in different areas of statistical theory in some detail and in particular established a relationship between the α-connections and a certain class of divergence functions. The *observed geometry* of Barndorff-Nielsen (1989) and Blæsild (1988) also contains the α-connection structure and they are able to demonstrate the relationship to divergence functions via the observed yoke. This work was unified by Lauritzen (1987) who introduced the concept of a *Statistical manifold* which encompasses both expected and observed geometries.

Definition. A Statistical manifold (M,g,T) is a manifold, M, with a metric, g, and a symmetric three tensor T. Denoting by Γ^0 the Christoffel symbol for the Levi-Civita connection of the metric g then the α-connections are defined by their Christoffel symbols

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

Amari's geometric structure can then be viewed as a Statistical manifold (M, g, T) where g is the Fisher information and T is defined by

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

The manifold will be given by a parametric family $\{(p(x,\theta)\}$ which satisfies the regularity conditions of Amari (1985 page16). All manifolds in this paper will also be assumed to satisfy these conditions.

Curvature.

We also need the following basic differential geometric constructions.

Definition. The Riemann-Christoffel curvature tensor is defined by a metric, g, and a connection, with Christoffel symbols Γ, by the formula

$$R_{ijkm} = \left(\frac{\partial}{\partial\theta_i}\Gamma^s_{jk} - \frac{\partial}{\partial\theta_j}\Gamma^s_{ik}\right)g_{sm} + \left(\Gamma^r_{jk}\Gamma_{irm} - \Gamma^r_{ik}\Gamma_{jrm}\right)$$

If and only if this tensor is identically zero does there exists an *affine coordinate system* in a torsion free manifold (such as we consider). That is a coordinate system θ such that

$$\Gamma^k_{ij}(\theta) \equiv 0 \qquad \forall\theta$$

In this case the space is said to be *flat*. If the connection was the Levi-Civita connection for the metric g, then in the affine coordinate system the metric tensor $g_{ij}(\theta)$ is constant for all values of θ. Under these conditions the space and the metric are said to be *flat*.

Preferred Point Manifolds.

In view of the fundamentally asymmetric nature of statistical inference a natural question arises: does there exist an asymmetric differential geometric structure which reflects the statistical properties of the parametric family? In Marriott (1989) and Critchley, Marriott and Salmon (1991) a new geometric structure called a *preferred point geometry* was introduced and the relationship between it and Lauritzen's Statistical manifold structure was developed. In any statistical problem it is usually the case that some point in the parametric family is treated differently from the remaining points. A preferred point geometry explicitly recognises this asymmetry and conditions the geometry on the value of ϕ. In particular it defines a Riemannian geometry on the parametric family conditionally on ϕ.

Definition. A preferred point geometry is a pair, (M, $g^\phi(\theta)$), where M is a manifold and $g^\phi(\theta)$ is a metric which is defined as a smooth function of the preferred point ϕ, as well as θ. In fact we shall often use the weaker condition that $g^\phi(\theta)$ is only a metric for θ in an open neighbourhood of ϕ.

The following result can be found in Critchley, Marriott and Salmon (1991).

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

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

We see that when θ is evaluated at the preferred point ϕ the metric reduces to the Fisher information.

We can calculate the Levi-Civita connection for this metric and observe its relationship with Amari's expected geometry. The Christoffel symbols Γ^{ϕ}_{ijk} for the Levi-Civita connection for the preferred point metric $g^{\phi}(\theta)$ are given by

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

When θ equals ϕ the connection agrees with the +1-connection in Amari's Statistical manifold. We immediately see something of the power of the preferred point method as we can now rationalise the +1 connection as a metric, or Levi-Civita, connection in this preferred point geometry. The preferred point structure can in fact also rationalise as metric connections both the 0 and -1-connections of Amari as will be seen below and is formaly demonstrated in our earlier paper.

α-geodesics and divergence functions.

Amari (1985) considers the relationship between α-geodesics and a class of divergence functions, which he calls α-divergences, in the case of an α-flat family. We focus here on the -1-divergence which coincides with the Kullback-Liebler information d_{kl}.

The following theorem gives both the projection and Pythagorean result in the special case $\alpha = -1$.

Theorem 1.(I)[Amari (1985, page 90)]. If M is a full exponential family and N a submanifold of M, then for any point θ in M, the point θ' in N which minimises $d_{kl}(\theta,\theta')$ is joined to θ via a -1-geodesic which cuts N orthogonally in the Fisher metric at θ'.

(II)[Amari (1985, page 86)] Given three points θ, θ', and θ'' in an α-flat manifold S, let c be the -1-geodesic connecting θ and θ' and c' be the +1-geodesic joining θ' and θ''. If the angle between c and c' at θ' is $\pi/2$, measured in the Fisher metric, then

$$d_{kl}(\theta,\theta'') = d_{kl}(\theta,\theta') + d_{kl}(\theta',\theta'')$$

Thus we see that on a ± 1-flat family (the full exponential family) the point on N closest in the Kullback-Leibler sense to a point in M is given by the -1-geodesic projection. There is a strong analogy between this and the result in Riemannian geometry which states that the point on a submanifold which is (geodesically) closest to a fixed point is found by dropping a geodesic which cuts the submanifold orthogonally. Thus there is a parallel between geodesic distances and divergences. However it is important to notice that since the α-geodesics are non-metric whenever $\alpha \neq 0$ there is no concept of α-geodesic *distance* involved in Amari's results. In Theorem 3, by using a stronger flatness condition than ± 1-flatness, we show how preferred point metrics generalise this theorem to the substantially stronger result that twice the Kullback-Leibler divergence will *equal* a squared geodesic distance. Thus the projection which minimises the divergence will also minimise the geodesic distance. Part (II) of the theorem also implies that the divergence acts as a squared measure of distance, thus again it is interesting to see when the divergence is equal to a squared geodesic distance. Theorem 4 shows us that this stronger flatness condition only holds in a small subset of proper statistical densities due to an inherent curvature necessarily induced by the requirement that the density integrate to unity. In Section 6 we develop the Pythagorean result in a general preferred point context.

3. Geodesic Distances and Divergence Functions.

In view of Amari's results it is natural to consider the relationship between geodesic distance functions and those which come from statistically defined divergences. We now show how this can helpfully be seen in a preferred point context and how the concept of a preferred point geometry is general enough to encompass divergence function theory.

Definition. The *preferred point distance* between ϕ and θ is defined to be the geodesic distance from ϕ to θ using the g^{ϕ} metric, i.e. the metric which is conditioned on the preferred point ϕ. We denote the squared preferred point geodesic distance by $D(\phi,\theta)$.

The fundamental difference between Riemannian distances and preferred point distances can now be clearly seen -- preferred point distances need not be symmetric. Consider $D(\phi,\theta)$ and $D(\theta,\phi)$: the first is the minimum path length from ϕ to θ measured in the geometry defined by the metric g^{ϕ}, whereas the second is the minimum path length from θ to ϕ measured in the g^{θ} - geometry. In general both the geodesic paths and the pathlengths will be different in the different geometries.

We now show that given any divergence function there exists a preferred point metric locally compatible with it. This means that for all points in a neighbourhood of the preferred point the squared preferred point distance will agree with the divergence from the preferred point. The following construction generates a preferred point geometry from a divergence function. Note, however, this is not a canonical construction and the second part of the theorem shows that there are a great number of compatible preferred point geometries for each divergence function.

Theorem 2. (i) Let d(,) be a divergence function. Then there exists a preferred point metric $g^{\phi}(\ ,\ )$ which is compatible with d, i.e. if $D(\phi,\theta)$ is the squared ϕ-geodesic distance from ϕ to θ, then

$$D(\phi,\theta) = d(\phi,\theta)$$

for all θ in some neighbourhood of ϕ.

(ii)This preferred point metric can be chosen to be flat for each value of the preferred point. Further let g be an arbitrary metric on the manifold. Define

$$d_{\phi}(\theta)=d(\phi,\theta)$$

i.e., consider ϕ fixed. The gradient vector field of the function $d_{\phi}(\theta)$ with respect to the metric g, is given by

$$v_i := (\mathrm{grad}(d_\phi))_i(\theta) = g^{ij}(\theta)\frac{\partial}{\partial\theta^j}(d_\phi)$$

Then there exists a preferred point metric g^ϕ which is flat and whose geodesics from ϕ are the gradient flow lines of $d_\phi(\theta)$.

Proof. See Appendix.

This result states that there are many preferred point metrics compatible with a single divergence function. Our axiomatic definition of a divergence function in Section 2 is very general. None of the ideas have been particularly statistical in nature aside from the observation that statistics is fundamentally a preferred point subject and there exist statistically natural preferred point metrics like Example 1. In the rest of the paper we look at objects which are fundamentally statistical and move from the general theory of preferred point geometries and divergence functions to consider specific examples of both and their interrelationship. We concentrate on the Kullback-Leibler divergence function and three preferred point metrics which have a natural statistical intepretation which can be found in Critchley, Marriott and Salmon (1991).

There is a basic property of the Kullback-Leibler divergence which is not reflected in our axioms. The axioms consider a divergence function defined on a (finite dimensional) parametric family. In fact the Kullback-Leibler divergence is defined consistently on a much wider class of functions.

Definition. Let $S = \{p(x)\}$ be the set of all mutually absolutely continuous regular density functions on the sample space X with respect to a measure P on the sample space.

On S it is well known that the Kullback-Leibler divergence is well-defined, nonnegative and if $f,g \in S$ then, $d_{kl}(f,g)$ is equal to zero if and only if f=g almost everywhere. S is an infinite dimensional space in which the parametric families which we have been considering are embedded. This simple observation enables us to draw a clear distinction between the Kullback-Leibler divergence and the geodesic distance defined by the preferred point metric in Example 1. The geodesic distance between two points is not purely a function of these points but also of the particular, finite dimensional, manifold in which they are considered to lie. However the Kullback-Leibler divergence is purely a function of the points and independent of the manifold chosen. This distinction is demonstrated in the following example.

Example 2. Consider the example of a curved parametric family defined in Efron (1975). Let M be the 2-dimensional parametric family of bivariate normal distributions with covariance matrix I, the identity and parametrised by the mean value (η_1, η_2). Let N be the subfamily of models with mean vector given by

$$\eta = \left(\theta, \frac{\gamma_0}{2}\theta^2\right)$$

where $\theta \in (-\infty, \infty)$.

The Geometry of the Kullback-Leibler Divergence

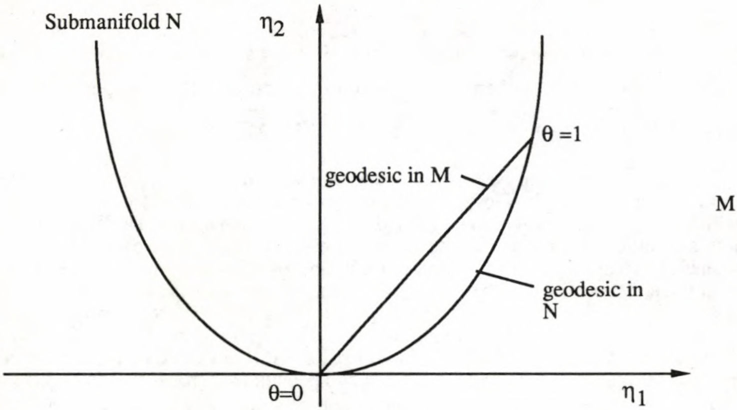

Figure 1

We can consider two possible geodesic distancse between two points, say $\theta = 0$ and 1, when they are seen either as elements of N or as elements of M respectively. In M the preferred point metric g^ϕ of Example 1 above is easily shown to be I when θ=0 is the preferred point. It is a constant in the mean parameters (η_1,η_2) and the geodesic is the straight line joining the two points. On the other hand the geodesic between the two points when regarded as elements of N is the relevant chord of the curve N itself and the geodesic distance reduces to the relevant arclength of N inside M. This is clearly a different distance. In contrast it is immediately clear that the Kullback-Leibler divergence between the two points is independent of the manifold in which they are considered to lie.

There is a fundamental trade off between these two approaches, one which exploits specific information about the assumed model and the other which is in some sense non-parametric in that it doesn't exploit this information. This trade off can be seen as reflecting the relative valuation of efficiency over robustness to mispecification which is present throughout statistical inference. Distances which are functions of the chosen manifold rather than just the end points clearly contain different information than non-parametric distance functions. Non-parametric functions may seem to be the more fundamental objects in the sense that they are defined independently from the non-canonical choice about the manifold in which statistical inference will be set. However it should be noted that many important statistical concepts, being defined relative to a finitely parameterised model, do depend on the choice of manifold, for example the distribution of both the score and maximum likelihood estimate.

4. Kullback-Leibler projection and total flatness.

The study of how one manifold can be embedded inside another is of fundamental importance to differential geometry and to ideas of curvature. From the above discussion it follows that in order to understand the geometry of the Kullback-Leibler divergence we need to consider this embedding. One method is through embedding curvature. In this and the following section we examine various measures of curvature associated with particular preferred point geometries and explore the corresponding notions of flatness. For illustration we look at the full exponential family case. Although exponential families can be considered flat in one sense we show that in general they do retain a form of curvature. This agrees with the remarks at the end of Section 3.3 of Amari (1985). We define this stronger form of curvature for a general family and show how it relates to the embedding of the parametric family inside an infinite dimensional function space. We are then able to define the geometric conditions under which the Kullback-Leibler divergence is consistent with our preferred point geodesic distance.

Since preferred point manifolds are generalisations of Riemannian manifolds we consider the metric, or Levi-Civita, connection as a measure of the curvature of the manifold. Denoting the preferred point metric by $g_{ij}^{\phi}(\theta)$ where ϕ is the preferred point, we define the preferred point Levi-Civita connection by its Christoffel symbols $\Gamma_{ijk}^{\phi}(\theta)$ as in Section 2.

Using this definition we have a measure of the curvature associated with the metric structure of a parametric family. However the family is naturally embedded in the larger function space S which has its own geometric structure given by the Kullback-Leibler divergence. Thus it is important to also look at the curvature defined by this embedding. In classical differential geometry there is a general concept of embedding curvature; if a manifold A with some geometric structure is embedded in B another manifold with its own geometric structure, then the *embedding curvature* of A in B is essentially determined by the 'difference' between two geometries. Any measure of this embedding curvature would ideally reflect the tension between the two geometric structures, the intrinsic geometry of A and secondly that induced by B on A. A desirable property of an embedding curvature would be that it characterised when these two geometries agreed. In Critchley, Marriott and Salmon (1991) we justified the preferred point geometry induced by g^{ϕ} on the parametric family $M=\{p(x,\theta)\}$. In this section we shall use preferred point geometry to define a measure of embedding curvature comparing the g^{ϕ}-geometry on M with the Kullback-Leibler on S.

Consider first how the Kullback-Leibler divergence changes in the parametric family

Definition. A preferred point metric is a 2-form. Consider the *preferred point 1-form* given in θ-coordinates by

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

This is a 1-form or covariant 1-tensor since under a change of coordinates it has the correct transformation properties and maps tangent vectors to the real numbers.

Lemma 1. If v is a tangent vector, the rate of change of the Kullback-Leibler divergence in the direction v is given by $\mu^{\phi}(v)$. That is the 1-form defined above applied to the tangent vector v.

Proof. In θ–coordinates we have

$$v = v^i\frac{\partial}{\partial\theta_i}$$

Thus the rate of change of the divergence is given by

$$\begin{aligned} v\left(E_{p(x,\phi)}[\ln p(x,\phi) - \ln p(x,\theta)]\right) &= v^i\frac{\partial}{\partial\theta_i}E_{p(x,\phi)}[\ln p(x,\phi) - \ln p(x,\theta)] \\ &= -v^iE_{p(x,\phi)}[\frac{\partial}{\partial\theta_i}\ln p(x,\theta)] \\ &= \mu^{\phi}(v) \end{aligned}$$

□

Since $d_{kl}(\phi,\phi) = 0$ the Kullback-Leibler geometry on N is determined by the one form μ^ϕ. Thus to compare the two geometries, and hence analyse the embedding curvature, we consider the rate of change of μ^ϕ with respect to the g^ϕ-geometry, in other words the g^ϕ-covariant derivative of μ^ϕ.

Definition. Let h^ϕ be the covariant derivative of μ^ϕ with respect to the g^ϕ-preferred point metric. It is the following preferred point metric which clearly reduces to the Fisher information when $\theta = \phi$:

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

where $g^{\phi\, rs}$ is the inverse of g^ϕ_{rs} the preferred point metric from Example 1 and Γ^ϕ is its Christoffel symbol.This preferred point metric was first defined in Critchley, Marriott and Salmon (1991) where it was shown that its Levi-Civita connection equals that of the 0-connection when $\theta = \phi$.

We will see that g^ϕ and h^ϕ provide via their Christoffel symbols the intrinsic and embedding geometry of N respectively. Following Critchley, Marriott and Salmon (1991) we now look at the full preferred point structure of a statistical space. To define this full structure we need to define the dual preferred point structure to g^ϕ.

Definition. Consider

$$k^\phi_{ij}(\theta) = h^\phi_{is}(\theta)g^{\phi\ st}(\theta)h^\phi_{tj}(\theta),$$

where g^ϕ and h^ϕ are defined above. Then $k^\phi_{ij}(\theta)$ is a preferred point metric which reduces to the Fisher information at $\theta = \phi$. Further its Levi-Civita connection agrees with the -1-connection of Amari's expected geometry as shown in Critchley, Marriott and Salmon (1991).

As an example of this complete structure let us consider the case of a full exponential family and compare it to Amari's expected Statistical manifold structure.

Example 3. Consider an exponential family given by

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

We can define two coordinate systems. The first given by the natural θ-coordinates and the second by

$$\mu^\phi_i(\theta) = -\mathbf{E}_{p(x,\phi)}[\frac{\partial}{\partial\theta_i}\ln p(x,\theta)]$$

There is a straightforward relationship between these coordinates and Amari's dual coordinate system (θ, η) on an exponential family where

$$\eta_i(\theta) = \mathbf{E}_{p(x,\theta)}[t_i(x)]$$

since in the full exponential family they differ simply by the translation

$$\mu^\phi_i(\theta) = \eta_i(\theta) - \eta_i(\phi).$$

The following is shown in Amari (1985) and Critchley, Marriott and Salmon (1991).

The full exponential family is both +1-flat and flat in the g^{ϕ}-metric, with the θ-coordinates affine in both cases. In these affine coordinates $g^{\phi}(\theta)$ will have a form which is independent of θ, and is given by

$$g^{\phi}_{ij}(\theta) = g_{ij}(\phi) = \frac{\partial^2 \psi}{\partial\theta_i \partial\theta_j}(\phi)$$

where g_{ij} is the Fisher information.

Dually we see that the family is both -1-flat and flat for the k^{ϕ}-metric, with the μ (and also η) parameters affine in both cases. Again in the affine coordinates the representation of k^{ϕ} will be independent of μ and is

$$k^{\phi}_{ij}(\theta) = \left[\frac{\partial^2 \psi}{\partial\theta_i \partial\theta_j}(\phi)\right]^{-1}$$

In an exponential family in the θ-coordinate system the Christoffel symbols of g^{ϕ}_{rs} are zero. Hence the metric h^{ϕ}_{rs} has the form

$$h^{\phi}_{ij}(\theta) = -\mathbf{E}_{p(x,\phi)}[\frac{\partial^2}{\partial\theta_i \partial\theta_j}\ln p(x,\theta)]$$

By a simple calculation this is minus the change of basis matrix for the change of coordinates from θ to μ. This result will hold in any g^{ϕ}_{rs}-flat manifold as long as the θ-coordinates are affine. Thus the embedding geometry alone determines this important change of basis between the pair of affine coordinates.

□

From the above example is clear that a general exponential family will not in fact be flat with respect to h^{ϕ}_{rs}, despite its g^{ϕ} and k^{ϕ} (or ± 1) flatness, and there will be no affine coordinate system for h^{ϕ}. However we are now in a position to define a strong form of flatness in which all three preferred point metrics have the same affine coordinate system. We can then show this flatness condition ensures that the three squared preferred point geodesic distances and the Kullback-Leibler divergence are all equivalent. Thus the embedding curvature of the manifold in S will be zero in the sense that two geometries induced on the manifold, from the metric and the divergence on S, agree.

Definition. A parametric family is *totally flat* if for each value of the preferred point ϕ there exists a *single* coordinate system which is affine for each of the three metrics g^{ϕ}, h^{ϕ} and k^{ϕ}. This coordinate system is called *coaffine*.

Total flatness is then implied whenever the statistical family has no intrinic (g^{ϕ}) curvature nor any embedding (h^{ϕ}) curvature relative to S.

Lemma 2. If g^ϕ is flat and the θ-coordinates are affine then k^ϕ is flat and the μ-coordinates are affine.

Proof. See appendix.

This lemma demonstrates that the key to the definition of total flatness is not that the two metrics g^ϕ and k^ϕ are flat on the same manifold since this always holds, rather it is that they have the same affine coordinates. The following lemmas shows that all that is needed is that the change of basis matrix, from θ to μ, is a constant in the correct coordinate system.

Lemma 3. If there exists a coordinate system θ which is affine for both g^ϕ and h^ϕ then the manifold is totally flat.

Proof. See appendix.

There is a dual result in which k^ϕ replaces g^ϕ in Lemma 3. We also have:

Lemma 4. If there exists a coordinate system θ which is affine for both g^ϕ and k^ϕ then the manifold is totally flat.

Proof. See appendix.

Example 4. Generalising the statistical manifold, M, of Efron's example above, consider the family of p-variate non-singular normal distributions with constant covariance matrix Σ parametrised by their mean values $\eta = (\eta_1, \dots, \eta_p)$. Apart from a constant, the log-likleihood is given by

$$l(\eta; x) = -\tfrac{1}{2}(x-\eta)^t \Sigma^{-1}(x-\eta)$$

Simple calculations then show that, whatever the preferred point ϕ,

$$g^\phi(\eta) = h^\phi(\eta) = \Sigma^{-1}$$

As this is independent of η, the η-coordinate system is both g^ϕ and h^ϕ affine. Then by Lemma 3, this coordinate system is coaffine and the space is totally flat

We now show that total flatness is a sufficient condition for each of the geodesic measures and the square root of the Kullback-Leibler divergence to be equal. This result can be considered an extension to Amari's results, described above, which relates the Kullback-Leibler divergence to the ± 1-connection in a ± 1-flat manifold. We have replaced the ± 1-flatness with the stronger total flatness condition and prove the stronger result that half the squared geodesic distance equals the divergence.

We first make the following remarks. Suppose that M is a totally flat family and that the θ-coordinates are coaffine. Then recalling that the metrics g^ϕ, h^ϕ and k^ϕ each agree with the Fisher information g at the preferred point ϕ, we have:

$$\forall\, \theta \in M, \quad g^\phi(\theta) = h^\phi(\theta) = k^\phi(\theta) = g(\phi)$$

Thus, in coaffine coordinates in a totally flat manifold, all three preferred point geometries reduce to the Euclidean geometry determined by $g(\phi)$. In particular their geodesics are the straight lines in θ-coordinates and the g^ϕ, h^ϕ and k^ϕ-geodesic squared distances from ϕ to θ are all simply

$$(\theta-\phi)^i g_{ij}(\phi)(\theta-\phi)^j$$

Theorem 3. Let a parametric family N be g^ϕ-flat, and let the θ-coordinates be g^ϕ-affine. Then the following are equivalent:

(i) N is (locally) totally flat.

(ii) The Kullback-Leibler divergence (locally) agrees with half the squared g^ϕ-geodesic distance

(iii) The Kullback-Leibler divergence is (locally) an exact quadratic function of the θ-coordinates given by

$$d^\phi_{kl}(\theta) = \tfrac{1}{2}(\theta^i - \phi^i) g_{ij}(\phi)(\theta^j - \phi^j)$$

Proof. Because the Christoffel symbols of the Levi-Civita connection of the g^ϕ-metric vanish in θ-coordinates,

$$h^\phi_{ij}(\theta) = \frac{\partial^2}{\partial\theta_i \partial\theta_j} d_{kl}(\phi,\theta) \qquad (1)$$

Now,

(i) $\Leftrightarrow$ $h^\phi(\theta)$ does not depend on θ in θ-coordinates, (using Lemma 3)
$\Leftrightarrow$ $d_{kl}(\phi,\theta)$ is quadratic in θ, (using (1))
$\Leftrightarrow$ (iii) (using the axioms for a divergence)

Thus, recalling the remarks before the theorem, (i)$\Rightarrow$(ii).

Finally (ii) $\Rightarrow$ (iii) at once as $g^\phi(\phi)=g(\phi)$ by hypothesis

□

From this result we can see that there are two distinct reasons for curvature occuring in this statistical problem. The first is that the family may have intrinsic curvature in the sense that g^ϕ may not be flat. The second more subtle form is due to the dual nature of statistical geometry and that even a family flat in the first sense, e.g. a full exponential family, can be curved in the second sense. That is it may be g^ϕ and k^ϕ flat without being totally flat.

The above theorem gives a characterisation of totally flat manifolds if the manifold is g^ϕ-flat. It is natural to see if a more direct characterisation of totally flat spaces can be found. Below we give a solution to this problem in the case of full exponential families. However the complete characterisation of totally flat spaces remains an interesting open question.

Let us take P_θ to be a p-dimensional full exponential family with canonical parameter θ and with representation

$$\frac{dP_\theta}{dP}(x) = B(x)\exp\{\theta^i t_i(x) - \psi(\theta)\}$$

relative to some dominating measure P. Then the metric h^ϕ is

$$h^\phi_{ij}(\theta) = \frac{\partial^2\psi}{\partial\theta_i\partial\theta_j}(\theta)$$

Note that the fact that h^ϕ does not depend upon the preferred point ϕ is a special property of full exponential families. In other words $h^\phi(\theta)$ equals $h^\theta(\theta)$ which is the θ-covariance of the canonical statistic t(x). Thus by Lemma 3 we have that P_θ is totally flat if and only if

$$\psi(\theta) = \tfrac{1}{2}A_{ij}\theta^i\theta^j + b_i\theta^i + c$$

where A is a symmetric positive definite matrix and A,b and c are all independent of θ. Thus we have shown:-

Lemma 5. Consider the above full exponential family P_θ. The following are equivalent:

(1) P_θ is totally flat.

(2) The covariance of the canonical statistic does not depend on the canonical parameter

(3) The log-likelihood is a quadratic function of the canonical parameter.

This last characterisation is of particular interest. Taking P to be Lesbegue measure $\lambda(x)$, we see that among the totally flat density functions are those of the form

$$\tilde{B}(x).N_p(A\theta + b, A)$$

where $N_p(A\theta+b,A)$ is the density of t(x), $\tilde{B}(x)$ has the same support as B(x) and A and b are as above. In the particular case $\tilde{B}(x)=1$ and t(x)=x we recover Example 4, the p-variate nonsingular normals with constant covariance matrix parametrised by mean value.

Theorem 3 gives the relationship between the intrinsic and embedding geometries of a parametric family in the most straightforward case. In general however a parametric family will not be totally flat and there will be a complex interrelationship between the various preferred point geodesics and divergence functions. Ideally the Kullback-Leibler divergence in a Euclidean embedding space would induce the metric g^ϕ in the submanifold of interest, in which case the flatness of g^ϕ would be sufficient for the two distance measures to agree. Theorem 3 shows us that in general g^ϕ-flatness will not be sufficient but indicates that to understand the general case we need to consider how the g^ϕ and k^ϕ -geometries are affected by the embedding of the statistical family in the space of proper density functions S. In fact it is necessary to consider a larger family than S in which g^ϕ may be seen as the metric induced by the Kullback-Leibler divergence. This construction makes clear the fundamental observation that the space of densities S is itself curved and the implications of this must follow through the analysis of curvature in all statistical models. This fundamental curvature is shown below to be essentially due to the preferred point nature of statistical inference.

Following Amari (1985) we make the definition.

Definition. Let $S = \{p(x)\}$ be the set of all mutually absolutely continuous regular density functions on the sample space X with respect to a measure P. Then let $\tilde{S}=\{m(x)\}$, where

$$m(x) = c.p(x), \quad c>0$$

be its extended set of finite measures. Further if M is a parametric family in S define $\tilde{M}$ by

$$\tilde{M} = \{m(x) \mid m(x) = c.p(x), c > 0 \text{ and } p(x) \in M\}$$

We call $\tilde{M}$ the *cone* defined by M.

By studying the preferred point geometry of such cones we can see the interrelationship between the g^ϕ, h^ϕ and k^ϕ preferred point metrics as a function of the embedding of M inside $\tilde{S}$. Further we will be able to see the general relationship between geodesics and divergences.

Lemma 6. Locally to a point in M, $\tilde{M}$ is a parametric family of finite measures. If θ is a parameterisation of M then the map

$$(\theta, K) \mapsto e^{K}.p(x,\theta)$$

gives a local parameterisation of $\tilde{M}$.

Proof. This follows from the regularity conditions on M given in Section 2.

Definition. Following the notation of fibre and tangent bundles we define a submanifold of $\tilde{M}$ which in $\{\theta,K\}$-coordinates has the form $\{\theta,K(\theta)\}$ to be a *section* of $\tilde{M}$ if $K(\theta)$ is differentiable.

The result below shows that the g^ϕ-geometry of a parametric family is determined by the cone it lies in.

Lemma 7. (i) g^ϕ is a positive semi-definite 2-tensor on $\tilde{M}$.

(ii) All sections of $\tilde{M}$ are g^ϕ-isomorphic.

Proof. See appendix.

Although each section of the cone has the same g^ϕ-geometry the same is not true of the h^ϕ-geometry. Since

$$\frac{\partial}{\partial\theta_i}\ln m(x,\theta,(K\theta)) = \frac{\partial}{\partial\theta_i}K(\theta) + \frac{\partial}{\partial\theta_i}\ln p(x,\theta)$$

the one form μ^ϕ is given by

$$-\mathbf{E}_{p(x,\phi)}\left[\frac{\partial}{\partial\theta_i}\ln m(x,\theta,K(\theta))\right] = -\frac{\partial}{\partial\theta_i}K(\theta) - \mathbf{E}_{p(x,\phi)}\left[\frac{\partial}{\partial\theta_i}\ln p(x,\theta)\right]$$

Thus h^ϕ is given by

$$\frac{\partial^2}{\partial\theta_i\partial\theta_j}K(\theta)-\Gamma_{ij}^{r\phi}(\theta)\frac{\partial}{\partial\theta_r}K(\theta)+\mathbf{E}_{p(x,\phi)}\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ln p(x,\theta)-\Gamma_{ij}^{r\phi}(\theta)\frac{\partial}{\partial\theta_r}\ln p(x,\theta)\right]$$

and this is not independent of the choice of function $K(\theta)$.

In other words, h^ϕ contains information about how M sits inside $\tilde{M}$. This remark leads to the following important insight. We can use the cone construction to prove a more general theorem than Theorem 3 which took place in S. Theorem 4 gives the conditions for the Kullback-Leibler divergence to be equal to half the squared g^ϕ-preferred point distance for embedding in the space $\tilde{S}$. To describe this we need to look at the Kulback-Leibler divergence in $\tilde{S}$. The following remarks show that the Kullback-Leibler acts in a linear way in $\tilde{S}$. It is the form of the restriction on S which turns it into a quadratic distance measure on the space of densities.

Proposition 1. If M is a parametric family with θ-parameterisation and $\tilde{M}$ is its cone with the (θ,K)-parametrisation then the Kullback-Leibler divergence between $\phi \in M$ and (θ,K) on $\tilde{M}$ is given by

$$d_{kl}^{\phi}(\theta,K)=-K+d_{kl}^{\phi}(\theta)$$

Proof. Immediate.

Note that this result shows that, contrary to its behaviour in S, d_{kl} can be negative in $\tilde{S}$. Further $\tilde{S}$ is closed under exponential mixing. In other words $\ln \tilde{S} := \{\ln f \mid f \in \tilde{S}\}$ is an affine space. Moreover, for each $f \in \tilde{S}$, d_{kl} maps $\ln \tilde{S}$ to **R** in such a way that it preserves the linear structure on the affine subspaces passing through f. That is:

Proposition 2. If $f,g \in \tilde{S}$ then $f^{1-\lambda}g^{\lambda} \in \tilde{S}$ and

$$d_{kl}^{f}(f^{1-\lambda}g^{\lambda})=\lambda.d_{kl}^{f}(g)$$

Proof. See appendix.

On S we have seen the Kullback-Leibler acts locally as a positive quadratic functional due to the axioms of a divergence. This follows from the fact the S is the set of densities and hence must fulfil the restriction that they integrate to unity. It is because the divergence is locally quadratic that it can be viewed as a geodesic quantity. We can use the greater flexiblity of the Kullback-Leibler on $\tilde{S}$ to prove the following result which shows that the Kullback-Leibler divergence can always be intepreted as a squared g^ϕ-geodesic distance on a submanifold of $\tilde{S}$.

Theorem 4. Let M be any g^ϕ-flat parametric family of densities and assume ϕ the preferred point lies in M. Then there exists M*, a parametric family in $\tilde{M}$ such that

(i) d_{kl}^{ϕ} is a divergence on M*,

(ii) M and M* have first order contact at ϕ. That is $\phi \in$ M* and $TM_\phi = TM^*_\phi$,

(iii) M* is g^ϕ-isomorphic to M, and

(iv) The Kullback-Leibler divergence on M* is equal to half the squared g^ϕ-geodesic distance.

(v) M* is totally flat.

Explicitely let M* be the section of the cone on N given by

$$\exp[\Lambda(\theta)]\, p(x,\theta)$$

where we choose $\Lambda(\theta)$ by

$$\Lambda(\theta) = d_{kl}^{\phi}(\theta) - \tfrac{1}{2} g_{ij}^{\phi}(\phi)(\theta-\phi)^i(\theta-\phi)^j$$

and θ are the g^ϕ affine coordinates.

Proof. Straightforward using Lemma 7.

□

From Theorem 4 we see another form of curvature in statistical geometry which affects the relationship between parametric family based distance functions, such as geodesics, and those on more general infinite dimensional embedding space. By this theorem for a g^ϕ -flat family in $\tilde{S}$ there will be agreement between the two forms of measurement if we have the condition that the preferred point 1-form is linear in the affine coordinates i.e.,

$$\mu_i^{\phi}(\theta) = \mathbf{E}_{p(x,\phi)}[\frac{\partial}{\partial\theta_i} \ln p(x,\theta)] = K_{ij}(\theta^j - \phi^j) \qquad (2)$$

This is a preferred point condition and varies with a different choice of ϕ. We can compare this to the condition that M* lies in the subspace of density functions so that each measure in M* integrates to one. In view of the fact that $\phi \in$ M* this is equivalent to

$$\mathbf{E}_{p(x,\theta)}[\frac{\partial}{\partial\theta_i} \ln p(x,\theta)] = 0 \qquad (3)$$

This is implied by (2) but is not equivalent to it since it is the restriction of (2) to the diagonal where $\theta=\phi$. This, in the language of preferred point geometry, is the homogeneous condition implied by (2). Thus the curvature comes from the preferred point geometry as the homogeneous condition (3) is not enough to fulfill the complete preferred point condition (2).

This curvature of S inside $\tilde{S}$, that is the fact that families obey (3) but not in general (2), is precisely the obstruction to Amari's projection theorem holding in its strongest form; i.e. that the Kullback-Leibler divergence agrees with the g^ϕ-squared geodesic distance in any g^ϕ-flat manifold.

5. Global measures of curvature.

The previous section has shown how we get different measures of distance depending on which space we consider the points to be embedded in. We can now reverse this idea to use this difference to produce some global measures of the different types of curvature reviewed above.

Definition. Let N be a preferred point manifold embedded inside M another preferred point manifold. Define d^N to be the N-geodesic distance and d^M the M-geodesic distance. We can then define the *geodesic difference function* of N in M from ϕ to θ to be

$$K(\phi,\theta) = d^N(\phi,\theta) - d^M(\phi,\theta)$$

In our statistical geometry have a function of this form corresponding to the g^ϕ, h^ϕ and k^ϕ-geodesics.

Definition. For a single family N we can define its *geodesic difference function* in S to be

$$K(\phi,\theta) = d^N(\phi,\theta) - \sqrt{d_{kl}(\phi,\theta)} \tag{4}$$

where d_{kl} is the Kullback-Leibler divergence. Again we have functions of this form corresponding to the g^ϕ, h^ϕ and k^ϕ-geometries. We see that if N is totally flat then all these functions will be zero.

In this section we shall look at these functions as 'curvatures' in full and curved exponential family examples and compare them to the standard pointwise measures of curvature. Let us consider first the full exponential families which are often thought of as the flat 'Euclidean spaces' of statistics. The results above on total flatness show that while they are g^ϕ-flat and k^ϕ-flat (and also ± 1-flat) they are not in general totally flat and do possess an embedding curvature which reflects the curvature of the dual geometry of a statistical manifold. In the language of differential geometry this curvature is an *obstruction* to the existence of a coaffine coordinate system. Thus recalling Theorem 3, it is convenient to use (4) with d^N being the g^ϕ-geodesic distance as a measure of the curvature induced by duality.

We can also apply these functions in curved exponential families.

Example 5. Consider again the model of Example 2. Let M be the 2-dimensional parametric family of bivariate normal models with covariance matrix I, the identity, and let N be the subfamily of models with mean vector given by

$$\eta_\theta = \left(\theta, \frac{\gamma_0}{2}\theta^2\right)$$

where $\theta \in (-\infty,\infty)$. Efron (1975) defined the curvature at θ to be

$$\gamma_\theta^2 = \frac{\gamma_0^2}{\left(1+\gamma_0^2\theta^2\right)^3}$$

and this is the embedding curvature of the submanifold N in the classical geometric sense. We can therefore extend this definition of curvature to the function

$$K(\phi,\theta) = d^N(\phi,\theta) - d^M(\phi,\theta)$$

where d is the g^ϕ-geodesic distance. The following lemma demonstrates that the two different methods of measuring curvature are consistent in the following sense.

Lemma 9. $d^N(\phi,\theta) - d^M(\phi,\theta) \equiv 0$ if and only if $\gamma_\theta \equiv 0$.

Proof. The geodesic distances will be equal if and only if N is totally geodesic in M. This holds if and only if the second fundamental form of N in M is zero and from a simple calculation this follows if and only if Efron's curvature is identically zero.

□

We also see that $d^M(\phi,\theta) = \sqrt{d_{kl}(\phi,\theta)}$ in this case which follows from Theorem 3 since M is totally flat.

In this one dimensional example we have an interesting difference between curvature in statistical geometry and in classical differential geometry. In classical differential geometry any one dimensional manifold will have no intrinsic curvature and we can always find an affine parametrisation. However as we can see from the above example in statistical geometry one dimensional families are not always totally flat due to the curvature of the dual structure. While we can find a g^ϕ-affine parametrisation and a k^ϕ-affine parameterisation, unless $d^N(\phi,\theta) - d^M(\phi,\theta)$ is identically zero these will not be the same. Lemma 9 shows that the amount of this curvature in this case is related to Efron's embedding curvature. However note that this case uses the property that M is a totally flat manifold and most curved exponential families will lie in full exponential families which do have curvature of their own. The following example shows that the more global measures of curvature can have useful applications in circumstances where pointwise measures of curvature have to be treated carefully.

Example 6. In the same full exponential family as above consider the curved exponential family given by the mean vector

$$\eta'_\theta = \left(\theta, \theta^4\right)$$

Efron's curvature at the point θ for a curve of the form $(\theta, f(\theta))$ is given by

$$\gamma_\theta = \left[\frac{(f''(\theta))^2}{\left[1+(f'(\theta))^2\right]^3}\right]^{\frac{1}{2}}$$

Hence in our example

$$\gamma_\theta = \left[\frac{144.\theta^4}{\left[1+16.\theta^6\right]^3}\right]^{\frac{1}{2}}$$

Thus the pointwise curvature at $\theta=0$ is $\gamma_0=0$ and this will not detect the non-linearity of the function. However since the geodesic distance embedding curvatures between $\theta=0$ and $\theta=1$ are a function of the curvature at each value of $\theta \in [0,1]$ the curvature inherent in the manifold will clearly be demonstrated.

6. The preferred point Pythagoras theorem.

Having looked at the projection theorem from the preferred point perspective we can now do the same for Amari's generalised Pythagoras Theorem. Here we show how the preferred point analysis is particularly appropriate to this problem in general, and produces a different proof of Amari's result in the case of exponential families.

For a parametric family our geometry is determined by two preferred point objects: μ^{ϕ} and g^{ϕ}. It is important, for any preferred point tensor, to understand how the geometry depends on the preferred point. In this section we look at this issue, mainly in the case of a full exponential family.

For a full exponential family the dependence of the g^{ϕ}-geometry on the preferred point is very simple. We note that such a family is flat and has the same affine coordinates, θ, for all values of ϕ. Thus for each choice of preferred point the geometry is Euclidean and all that changes with a different preferred point is the value of the quadratic form which determine the Euclidean distance. This quadratic form is determined by

$$g_{ij}^{\phi} = g_{ij}(\phi)$$

that is the Fisher information at ϕ.

The change in the preferred point 1-form is also simple in the full exponential family case. In θ-coordinates we have

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

If ϕ_1 and ϕ_2 are two choices of preferred point we have that the difference in the forms, which is also a 1-form, is

$$\begin{aligned}\mu_i^{\phi_1}(\theta) - \mu_i^{\phi_2}(\theta) &= \eta_i(\phi_2) - \eta_i(\phi_1) \\ &= \mathbf{E}_{p(x,\phi_2)}[x_i] - \mathbf{E}_{p(x,\phi_1)}[x_i]\end{aligned}$$

which is independent of θ, the point of evaluation. Thus the two 1-forms differ by a translation which is independent of θ.

We finally look at the dependence of h^{ϕ} on the preferred point.

Lemma 10. For a full exponential family the preferred point metric $h^{\phi}(\theta)$ is equal to the Fisher information, $g(\theta)$ at all points.

Proof. This follows since both equal the hessian

$$\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ln p(x,\theta)$$

in the natural θ-coordinates. □

One important consequence of Lemma 12 is that the curvature of h^{ϕ} is independent of the preferred point for a full exponential family.

We can now consider the generalised version to Pythagoras' Theorem for the Kullback-Leibler divergence proved in Amari (1985) and Čencov (1972). The result we shall prove follows from the considerations of how our preferred point geometries depend on the value of the preferred point and so in the particularly simple case of the full exponential family we get the reduction to Amari's neat result based on α-projections.

Definition. If m is a one form on M, a finite dimensional manifold, we define the *null path* of m through θ to be the solution of the differential equation

$$m_i\left(\frac{\partial}{\partial\theta_i}\right) = 0$$

which passes through θ. By the standard results on the existence and uniqueness of differential equations the null path will exist locally if m is non-singular at θ.

Lemma 11. Let ϕ_1, ϕ_2 and ϕ_3 be three points on a finite dimensional parametric family M. Define the 1-form m_i by

$$m_i(\theta) = \mu_i^{\phi_1}(\theta) - \mu_i^{\phi_2}(\theta)$$

Then if m is non-singular at ϕ_2 we have the Pythagorean relationship for the Kullback-Leibler divergence

$$d_{kl}(\phi_1,\phi_2) + d_{kl}(\phi_2,\phi_3) = d_{kl}(\phi_1,\phi_3)$$

if ϕ_3 lies on a null path of m through ϕ_2.

Proof. If ϕ_3 equals ϕ_2 then we clearly have the result. In general, write the above equation in the form

$$d_{kl}(\phi_1,\phi_2) = d_{kl}(\phi_1,\phi_3) - d_{kl}(\phi_2,\phi_3) \qquad (5)$$

The left hand side of this equation is constant with respect to ϕ_3 and the rate of change of the right hand side with respect to ϕ_3 is given by the one form

$$\mu_i^{\phi_1}(\phi_3) - \mu_i^{\phi_2}(\phi_3)$$

which is $m_i(\phi_3)$. Thus equation (5) holds as ϕ_3 moves along the null path of the 1-form m_i.

□

In the case of the full exponential family the null paths of the above 1-form are particularly simple. We shall give an alternative proof of Amari's generalised Pythagoras' Theorem in this case as a corollary to the previous Lemma.

Corollary. In the case of the full exponential family

$$d_{kl}(\phi_1,\phi_2) + d_{kl}(\phi_2,\phi_3) = d_{kl}(\phi_1,\phi_3)$$

if ϕ_3 lies on a +1-geodesic through ϕ_2 which is Fisher orthogonal to the -1-geodesic joining ϕ_1 and ϕ_2 .

Proof. The one form m_i is constant in θ-coordinates as shown above. Hence the null line is a θ-affine line or a +1-geodesic through ϕ_2. The direction is defined by the tangent vector v at ϕ_2 which satisfies

$$m(v)=0$$

or

$$(\mu_i^{\phi_1}(\phi_2) - \mu_i^{\phi_2}(\phi_2))[v^i] = (\mu_i^{\phi_1}(\phi_2))[v^i] = 0 \tag{6}$$

The tangent to the -1-geodesic joining ϕ_1 and ϕ_2 is calculated by using the change of basis matrix h^ϕ applied to the -1-geodesic μ^ϕ. If we denote this tangent by w^i in θ–coordinates then (6) reduces to

$$w^i g_{ij}(\phi_2) v^j = 0$$

□

7. Conclusions.

In this paper we have taken as our starting point Amari's projection and generalised Pythagoras' Theorems and their applications to the relationship between geodesics and divergences, in particular the ±1-geodesics and the Kullback-Leibler divergence. We have used the tools of preferred point geometry to throw new light on the relationship between statistically motivated measures of 'distance' and those based on more geometrical foundations. For the projection theorem we begin by noting that divergences, despite their inherent asymmetry, can be viewed (locally) as preferred point geodesic distances. In Theorem 2 we show this in general using simply the axioms of a divergence function. Moving to the Kullback-Leibler divergence in particular we note that this has particularly nice embedding properties and defines a distance function not on any one particular parametric family but on an infinite dimensional space in which all our regular families are embedded. Theorem 3 shows necessary conditions on the preferred point curvature for the squared preferred point geodesic distance (being a function of the family in which we are working) to equal the non-parametric Kullback-Liebler distance. Theorem 4 shows that the relationship between the two concepts is best expressed in the larger infinite dimensional space of finite measures since this avoids problems of the curvature induced by restricting attention just to density functions rather than general measures.

It is interesting to consider the above in the light of Yoke theory. Firstly we note that squared preferred point geodesic distances give new examples of (negative) normed Yokes. Again it will be interesting to compare the theory of these yokes, which are defined as functions of particular parametric family, with the non-parametric examples of expected and observed Yokes used by Barndorff-Nielsen and Blæsild. This latter will require an observed or, at least, a conditional version of preferred point geometry.

The discussion above also illustrates some different aspects of curvature in statistical theory. It is clear from the above that curvature can arise from separate although interconnected reasons;

(i) In the first place by restricting attention to some finite dimensional parametric family of densities. This induces a form of embedding curvature in S the space of densities.

(ii) We have seen that the space S itself is embedded in a larger space of finite measures. This embedding will cause curvature as Theorem 4 shows. This curvature follows from the difference between the preferred structure of the geometry and the definition of S which is not fully preferred point but homogeneous.

(iii) There is also curvature induced by the dual nature of statistical geometry. Good examples of this are the full exponential families which although in many ways are flat they in general will posses this dependence. The one case where this curvature and the others above disappears is the totally flat case described and partially classified above.

We have also seen how by considering the different ways of measuring distances between points via different embeddings we can define global measures of curvature. These measures often agree with pointwise measures but have the advantage of 'intergrating' these measures over the area of interest.

Finally we have seen how the preferred point approach is appropriate to the analysis of the generalised Pythagorean theorem of the Kullback-Leibler divergence.

Appendix.

Proof of Theorem 2. (i) By the positivity of the hessian of the divergence function at θ_0 we can apply Morse's lemma, see Poston and Stewart (1976). Let us fix θ_0 and treat the divergence function as a function of θ. We choose coordinates $\psi(\theta)$ such that locally

$$d(\theta_0,\theta) = \tfrac{1}{2} I_{ij}(\theta_0)\psi^i(\theta)\psi^j(\theta)$$

where I_{ij} Fisher information at θ_0.

Thus we can use this coordinate change to define a map from Θ to $\mathbf{R}^p$ by

$$\Psi: \theta \mapsto \psi(\theta)$$

We now define a metric on M by pulling back the standard metric on $\mathbf{R}^p$ via the map Ψ. Thus we define $g^{\theta_0}(\ ,\ )$ by

$$g^{\theta_0}(v_1,v_2) = \langle \Psi^* v_1, \Psi^* v_2 \rangle$$

where $\langle\ ,\ \rangle$ is the standard Euclidean inner product on $\mathbf{R}^p$, and $\Psi^*:TM \rightarrow \mathbf{R}^p$ is the lift of Ψ to the relevant tangent spaces. By construction the squared geodesic distance will equal the divergence locally.

(ii)We view $v_i(\theta)$ as a smooth function from $\mathbf{R}^p$ to $\mathbf{R}^p$. We show that the derivative of v is of full rank. By calculation the derivative at θ_0 equals

$$g^{ij}(\theta_0)\frac{\partial^2}{\partial\theta_i\partial\theta_j}(d_{\theta_0})(\theta_0) + \frac{\partial}{\partial\theta_j}g^{ij}.\frac{\partial}{\partial\theta_i}(d_{\theta_0})(\theta_0) = g^{ij}(\theta_0)\frac{\partial^2}{\partial\theta_i\partial\theta_j}(d_{\theta_0})(\theta_0)$$

using the vanishing of the first derivative of the divergence at θ_0. Thus since the metric is invertible at θ_0 and the hessian of the divergence function is invertible we see that the derivative of v is invertible at θ_0.

We apply Morse's lemma again. Choose coordinates $\psi(\theta)$ such that we can write

$$v_i(\theta) = \psi_i(\theta) - \psi_i(\theta_0)$$

In ψ coordinates the gradient flow lines of the divergence function are just the affine lines from the origin.

We shall show that under this map the Euclidean spheres in Ψ-space are the images of the level sets of the divergence function. We can see this since the value of the divergence function can be found intergrating the length of the divergence function along the integral curve of the gradient flow $\gamma(t)$. That is:

$$d_{\theta_o}(\theta) = \int_\gamma \left| v_i(\gamma(t)) \right| dt = \int_\gamma \left| \mathrm{grad}(d_{\theta_o}(\gamma(t))) \right| dt$$

Thus in ψ coordinates we see that the divergence function is constant on Euclidean spheres centred aorund θ_0.

We therefore define the metric in θ-space by pulling back the standard (flat) Euclidean metric on ψ-space via the map Ψ. □

Proof of Lemma 2. Since g^ϕ is flat, h^ϕ is the change of basis matrix for the change of coordinates from $\theta \to \mu$. Thus we shall calculate k^ϕ in the μ coordinates. By definition

$$k^\phi = h^\phi \left[g^\phi \right]^{-1} h^\phi$$

Let us denote a tensor T in θ-coordinate by T_{ij} and in μ-coordinates by T'_{ij} and suppress the dependence on the preferred point for clarity. The form of k^ϕ in μ-coordinates is

$$\begin{aligned} k'_{ij} &= h^{i\alpha} k_{\alpha\beta} h^{\beta j} \\ &= h^{i\alpha} h_{\alpha\gamma} g^{\gamma\delta} h_{\delta\beta} h^{\beta j} \\ &= g^{ij} \end{aligned}$$

Thus the form of k^ϕ in μ-coordinates equals the inverse of the form of g^ϕ in θ-coordinates. This form is independent of θ since the θ-coordinates are affine for g^ϕ. Thus the form of k^ϕ in μ-coordinates will be independent of μ. That shows that the μ coordinates are affine for k^ϕ and this metric is flat.

□

Proof of Lemma 3. From the above lemma we see that there exists a set of coordinates μ for which k^{ϕ} is constant, as well as θ in which both g^{ϕ} and h^{ϕ} are constant. Since g^{ϕ} is flat we again use the fact that h^{ϕ} is the change of basis matrix between θ and μ-coordinate. This matrix is independent of θ and as the μ-coordinates are affine for k^{ϕ} so are the θ coordinates. □

Proof of Lemma 4. Since g^{ϕ} is flat the h^{ϕ}-metric is also the change of basis for the θ and μ coordinate systems. However since θ is affine for both metrics the change of basis between must be a fixed affine map, and hence independent of θ. □

Proof of Lemma 7. (i) By the chain rule we see that g^{ϕ} is a 2-tensor on $\tilde{M}$. If θ is a coordinate system on M consider the coordinate system on $\tilde{M}$ given by (θ, K) where

$$m(x,\theta,K) = e^{K}p(x,\theta)$$

In this coordinate system we calculate g^{ϕ} to be

$$\left[\begin{array}{c|c} g^{\phi}_{ij}(\theta) & 0 \\ \hline 0 & 0 \end{array}\right]$$

Thus it is positive semi-definite.

(ii) Let us define a section of $\tilde{M}$ by using the (θ,K) coordinate system. Define this section to be $(\theta,K(\theta))$ for some real valued function $K(\theta)$. Calculating the g^{ϕ}-preferred point metric:

$$\frac{\partial}{\partial\theta_i}\ln m(x,\theta,K(\theta)) = \frac{\partial}{\partial\theta_i}\left[K(\theta)+\ln p(x,\theta)\right] = \frac{\partial}{\partial\theta_i}K(\theta)+\frac{\partial}{\partial\theta_i}\ln p(x,\theta),$$

hence

$$\mathrm{cov}_{\phi}\left[\frac{\partial}{\partial\theta_i}\ln m(x,\theta,K(\theta)).\frac{\partial}{\partial\theta_j}\ln m(x,\theta,K(\theta))\right] = \mathrm{cov}_{\phi}\left[\frac{\partial}{\partial\theta_i}\ln p(x,\theta).\frac{\partial}{\partial\theta_j}\ln p(x,\theta)\right]$$

or

$$g^{\phi}_{ij}(\theta,K(\theta)) = g^{\phi}_{ij}(\theta)$$

□

Proof of Proposition 2. If f and g are absolutely continous with respect to our base measure P then it is clear that $f^{1-\lambda}g^{\lambda}$ is also. It follows from Loh (1983) that $f^{1-\lambda}g^{\lambda}$ has finite measure and hence lies in $\tilde{S}$.

Further

$$d^{f}_{kl}(f^{1-\lambda}g^{\lambda}) = \mathbf{E}_{f(x)}\left[\ln f - \ln f^{1-\lambda}g^{\lambda}\right] = \mathbf{E}_{f(x)}\left[\ln f - (1-\lambda)\ln f - \lambda\ln g\right] = \lambda.d^{f}_{kl}(g)$$

□

References.

Amari S. (1985), *Differential-Geometric methods in statistics*, Lecture Notes in Statistics 28, Springer-Verlag.

Akaike (1973), Information theory and an extension of the maximum likelihood principle, *Second International Symposium on Information theory (ed. B.N. Petrov and F. Czaki)*, pp 267-281. Budapest: Akademiai Kiado.

Barndorff-Nielsen O.E. (1989), *Parametric statistical models and likelihood*, Lecture Notes in Statistics 50, Springer-Verlag.

Barndorff-Nielsen O.E., Cox R. D. and Reid N. (1986), The role of differential geometry in statistical theory, *Int. Statist. Rev. 54*, 83-96

Bhattacharrya A. (1943), On discrimination and divergence, *Proc. 29th Indian Sci. Cong.,* Part III, 13

Blæsild P. (1988), Yokes and tensors derived from yokes, *Research report 173*, Dept. Theor. Statist., Åhrus University, (to appear in *Ann. Inst. Statist. Math.*).

Blæsild P. (1990), Yokes: orthogonal and extended normal coordinates, *Research report 205*, Dept. Theor. Statist., Åhrus University

Burbea, J. and Rao C. R. (1982) , Entropy differential metrics distance and divergence measures in probability spaces: a unified approach. *J. Multi. Var. Analys. 12*, 575-596

Čencov N. N. (1972), *Statistical decision rules and optimal infenence*, Nuaka, Moscow; translated into English 1978).

Critchley F., Marriott P.K. and Salmon M. (1991), Preferred point geometry and statistical manifolds, *Preprint*, European University Institute, Florence.

Csiszar I. (1975), I-divergence geometry of probability distributions and minimisation problems, *Ann. Probability 3*, 146-158.

Dawid A. P. (1975), Discussion to Efron's paper, *Ann. Statist. 3.* 1231-1234.

Efron B., (1975), Defining the curvature of a Statistical problem (with discussion), *Ann. Statist. 3*, 1189-1242.

Jeffreys H., (1948), *Theory of Probability,* second edition. Clarendon press, Oxford.

Kass R.E., (1989), The geometry of asymptotic inference (with discussions). *Statistical Sciences 4*, 188-234.

Kullback S. L. and Leibler R.A., (1951), On information and sufficiency, *Ann. Math. Statist. 22*, 79-86.

Lauritzen S. L. (1987), Statistical Manifolds, *Differential geometry in statistical inference*, IMS Lecture Note Monograph Series, vol. 10, Hayward, California.

Loh Wei-Yin, (1983), A note on the geometry of the Kullback-Leibler information numbers, *Technical report No. 716*, Dept. of Statist., University of Wisconsin.

Marriott P.K., (1989), *Applications of differential geometry to statistics*, Ph. D. thesis, University of Warwick.

Poston T. and Stewart I.N., (1976), *Taylor expansions and catastrophes,* Research Notes in Mathematics 7, Pitman, London.

Rao, C. R., (1945), Information and accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Maths. Soc. 37*, 81-91.

Rao, C. R., (1987), Differential metrics in probability spaces, *Differential geometry in statistical inference*, IMS Lecture Note Monograph Series, vol. 10, Hayward, California.

Spivak, M (1970), *A comprehensive introduction to differential geometry*, Publish or Perish, Berkeley

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. BENSAID/ 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).

* Working Paper out of print

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 PHLIPS/
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 Forecastible 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 PHLIPS
Manipulation of Crude Oil Futures

ECO No. 91/37
Xavier CALSAMIGLIA/Alan KIRMAN
A Unique Informationally Efficient and Decentralized Mechanism With Fair Outcomes

* Working Paper out of print

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 PHLIPS
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

ECO No. 91/53
Frank CRITCHLEY/Paul MARRIOTT/
Mark SALMON
Preferred Point Geometry and the Local Differential Geometry of the Kullback-Leibler Divergence

* Working Paper out of print