EUI WORKING PAPERS IN ECONOMICS

An Elementary Account of Amari's Expected Geometry

FRANK CRITCHLEY
PAUL MARRIOTT
MARK SALMON

European University Institute, Florence

EUROPEAN UNIVERSITY INSTITUTE, FLORENCE

ECONOMICS DEPARTMENT

EUI Working Paper **ECO** No. 94/12

**An Elementary Account of
Amari's Expected Geometry**

FRANK CRITCHLEY
PAUL MARRIOTT
MARK SALMON

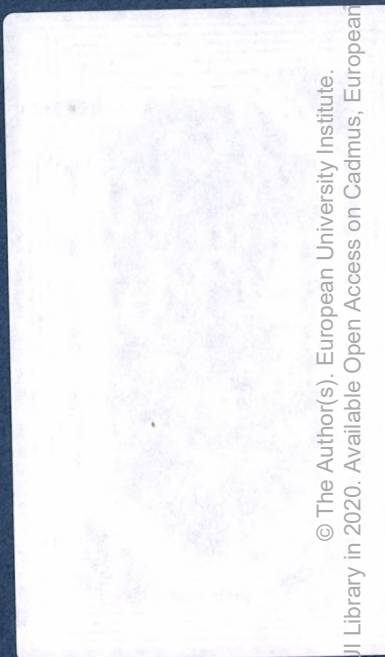**BADIA FIESOLANA, SAN DOMENICO (FI)**

An Elementary Account of Amari's Expected Geometry.

Frank Critchley
Birmingham University

Paul Marriott
Surrey University

Mark Salmon
European University Institute

6 September 1993

Abstract

An elementary and visual account of Amari's (1990) expected geometry is provided, focusing on the full exponential family case. Formal definitions of affine connections are not required. Rather, it is sufficient to consider the first two moments of the score function under the true distribution. Amari's fundamental non metric affine connection appears as the natural measure of the non constancy of the true covariance of the score. This covariance is constant in the natural parameters. Non linearity of the graph of the mean score in the natural parameter is seen to reflect a curvature present in nearly all parametric families.

The notion of ψ-duality is introduced. This is a natural duality between the score function in one parametrisation and the maximum likelihood estimate in another. It is seen to correspond to, and therefore provide a statistical interpretation of, the notion of duality in Amari's expected geometry.

1   Introduction.

Differential geometry has found fruitful application in statistical inference. In particular, Amari's (1990) *expected geometry* has been used in higher order asymptotic analysis, and in the study of sufficiency and ancillarity. However, we identify three drawbacks to the differential geometry approach. Firstly, the unfamiliar terms involved in such an approach can be difficult for the statistician to fully appreciate. Secondly, their statistical meaning can be less than completely clear, and finally the fact that, at its core, geometry is a visual subject can be obscured by the mathematical formalism required for a rigorous analysis, thereby hindering intuition. All three drawbacks apply particularly to the differential geometric concept of a non metric affine connection.

The primary objective of this paper is to mitigate these drawbacks in the case of Amari's expected geometric structure on a full exponential family. We aim to do this by providing an elementary account of this structure which is accessible geometrically, clearly based statistically and also visually presented.

Statistically, we use three natural tools, these are; the score function and its first two moments with respect to the true distribution. Geometrically, we are largely able to restrict attention to tensors. In particular, we are able to avoid the need to formally define an affine connection. It turns out that this account also sheds some new light on the choice of parametrisation as discussed by Amari (1990), extending earlier work by Bates and Watts (1980, 1981), Hougaard (1982) and Kass (1984).

A key feature of our account is that all expectations and induced distributions are taken with respect to one fixed distribution namely, that assumed to give rise to the data. This is the so called *preferred point* geometrical approach developed in Critchley, Marriott and Salmon (1992, 1993), on whose results we draw as appropriate.

To emphasise the visual foundation of geometric analysis we parallel the mathematical development with graphical illustrations using important examples of full exponential families. Although the analysis is not restricted to this case, we emphasise one dimensional examples so that simple pictures can be used to illustrate the underlying geometrical ideas and aid intuition.

There are also a number of points of contact between our account and Firth (1993).

Our hope is that this account will serve to broaden interest in an important and developing area. For a more formal but still readable treatment

1

of differential geometry, see Dodson and Poston (1977). For broader accounts of the application of differential geometry to statistics see the review papers or monographs by Barndorff-Nielsen, Cox and Reid (1986), Kass (1987, 1989), Amari (1990) and Murray and Rice (1993).

The paper is organised as follows. The elementary prerequisites are established in Section 2. The key elements of Amari's expected geometry of general families of distributions are briefly and intuitively reviewed in Section 3. In particular, his connections are discussed in terms of the characteristic statistical properties of their associated affine parametrisations. The final section contains our account of this geometry in the full exponential family case, as outlined above. Extending this account to the general case is necessarily more complicated and will be dealt with in a future paper.

2   Preliminaries.

2.1   The general framework.

Let

$$M = \{p(x, \theta) : \theta \in \Theta\}$$

be a p-dimensional parametric family of probability (density) functions. The available data $\mathbf{x} = (x_1, \ldots, x_n)^T$ is modelled as an i.i.d. random sample with some unknown true distribution $p(x, \phi) \in M$. Let the parameter space Θ be an open connected subset of $\mathbf{R}^p$. Formally, certain regularity conditions are entailed. These are detailed in Amari (1990, page 16).

2.2   The score function.

The score function

$$s(\theta, \mathbf{x}) = (\frac{\partial}{\partial \theta^1} \ln p(\mathbf{x}, \theta), \ldots, \frac{\partial}{\partial \theta^p} \ln p(\mathbf{x}, \theta))^T$$

is very natural to work with statistically as it contains precisely all the relevant information in the likelihood function; Integrating over Θ recovers the log likelihood function, l, up to an additive constant which is independent of θ. This is equivalent to the likelihood up to a multiplicative positive factor which may depend on $\mathbf{x}$ but not θ. Two different choices of the constant will contain the

2

same likelihood information (Cox and Hinkley 1974, page 12). We call this information the *shape* of the likelihood. Visually, the graph of the score function displays the shape of the likelihood in a natural and direct way. We use this to advantage later.

The score function is also a very natural tool to work with geometrically. An important concept of differential geometry is that of the *tangent space*. We can avoid the general abstract definition here as we have a concrete representation of this space in terms of the score function. Regarding now $\mathbf{x}$ as a random variable, following Amari (1990), we identify the tangent space TM_θ at each fixed $p(\mathbf{x}, \theta) \in M$ with the vector space of random variables spanned by

$$\{s_i(\theta, \mathbf{x}) = \frac{\partial}{\partial \theta^i} \ln p(\mathbf{x}, \theta) : i = 1, \ldots, p\}.$$

Under the regularity conditions noted in Section 2.1, this vector space has dimension p, the dimension of M.

2.3 Distribution of the score vector.

Naturally associated with each fixed tangent space TM_θ is the joint distribution ρ_θ^ϕ of the components of the score vector $s(\theta, \mathbf{x})$. This may be known analytically but can always, by the central limit theorem, be approximated asymptotically by the multivariate normal distribution $N_p(\mu^\phi(\theta), g^\phi(\theta))$ where

$$\mu^\phi(\theta) = \mathbf{E}_{p(x,\phi)}[s(\theta, \mathbf{x})] = n\mathbf{E}_{p(x,\phi)}[s(\theta, x)]$$

and

$$g^\phi(\theta) = \mathrm{Cov}_{p(x,\phi)}[s(\theta, \mathbf{x})] = n\mathrm{Cov}_{p(x,\phi)}[s(\theta, x)]$$

These last two quantities are the statistically natural tools that we shall employ in our account of Amari's geometry. The matrix $g^\phi(\theta)$ is assumed to be always positive definite.

Note that, for all ϕ,

$$\mu^\phi(\phi) = 0 \quad \text{and} \quad g^\phi(\phi) = I(\phi) = ni(\phi)$$

where I and i denote the Fisher information for the sample and for a single observation respectively.

For later use we define the random vector $\epsilon^\phi(\theta, \mathbf{x})$ by the decomposition

$$s(\theta, \mathbf{x}) = \mu^\phi(\theta) + \epsilon^\phi(\theta, \mathbf{x})$$

3

so that $\mathbf{E}_{p(x,\phi)}[\epsilon^\phi(\theta, \mathbf{x})]$ vanishes identically in θ and ϕ.

In the one dimensional case there is a particularly useful graphical representation of the three tools on which our account is based. For a particular realisation of the data $\mathbf{x}$ the plot of the graph of $s(\theta, \mathbf{x})$ against θ can give great insight into the shape of the observed likelihood function. We call this graph the *observed plot*. Together with this we use the *expected plot*. This is a graph of the true mean score together with an indication of variability. We make extensive use of this graphical method for particular important examples.

2.4 Reparametrisation.

So far, we have worked in a single parametrisation θ. It is important to consider what happens under a reparametrisation.

We consider reparametrisations $\theta \to \xi(\theta)$ that are smooth and invertible. Define,

$$B_i^\alpha(\theta) = \frac{\partial \xi^\alpha}{\partial \theta^i} \quad \text{and} \quad \bar{B}_\alpha^i(\xi) = \frac{\partial \theta^i}{\partial \xi^\alpha},$$

say. By the chain rule, the components of the score vector transforms as 1-tensors. That is:

$$s_\alpha(\xi(\theta), \mathbf{x}) := \frac{\partial l}{\partial \xi^\alpha} = \sum_{i=1}^p \bar{B}_\alpha^i(\xi(\theta))\frac{\partial l}{\partial \theta^i} := \sum_{i=1}^p \bar{B}_\alpha^i(\theta)s_i(\theta, \mathbf{x}) \tag{1}$$

for each fixed θ. This amounts to a change of basis for the vector space TM_θ. By linearity of expectations, the components of $\mu^\phi(\theta)$ are also 1-tensors. That is:

$$\mu_\alpha^{\xi(\phi)} = \sum_{i=1}^p \bar{B}_\alpha^i(\theta)\mu_i^\phi(\theta) \tag{2}$$

As covariance is a bilinear form, we see that $g^\phi(\theta)$ is a 2-tensor. That is, its components transform according to:

$$g_{\alpha\beta}^{\xi(\phi)}(\xi(\theta)) = \sum_{i=1}^p \sum_{j=1}^p \bar{B}_\alpha^i \bar{B}_\beta^j(\theta)g_{ij}^\phi(\theta) \tag{3}$$

By symmetry, the assumption of positive definiteness and since $g^\phi(\theta)$ varies smoothly with θ, $g^\phi(\theta)$ fulfils the requirements of a *metric tensor*, see Amari (1990, page 25). It follows at once, putting $\theta = \phi$, that the Fisher information also enjoys this property.

In parallel with this tensor analysis plotting the observed and expected plots for different parametrisations of the model can be extremely useful in

4

conveying the effects of reparametrisations on the shape of the likelihood and the statistical behaviour of important statistics such as the MLE.

3   Some elements of Amari's expected geometry.

3.1   Connections.

Formally, Amari's expected geometry is a triple (M, I, ∇^{+1}) in which M is a family of probability (density) functions and I the Fisher information metric tensor, as described above. The major difficulty in understanding revolves around the third component ∇^{+1} which is a certain non metric affine connection. In Section 4, we obtain a simple clear, statistical interpretation of it in the full exponential family case.

Here we note certain additional facts concerning connections and Amari's geometry, offering intuitive explanations and descriptions where possible. For a formal treatment see Amari (1990). We emphasise that such a treatment is not required here, as our later argument proceeds in terms of the elementary material already presented.

A connection allows us to (covariantly) differentiate tangent vectors and, more generally tensors, see Dodson and Poston (1977, Chapter 7). A connection therefore determines which curves in a manifold shall be called 'geodesic' or 'straight'. Generalising familiar Euclidean ideas, these are defined to be those curves along which the tangent vector does not change.

A metric tensor induces in a natural way an associated connection called the Levi-Civita or metric connection. In Amari's structure the Fisher information I induces the affine connection denoted by ∇^0. The Levi-Civita connection has the property that its geodesics are curves of minimum length joining their endpoints. No concept of length is associated with the geodesics corresponding to non metric connections.

Amari shows that the two connections ∇^0 and ∇^{+1} can be combined to produce an entire one parameter family $\{\nabla^\alpha : \alpha \in \mathbf{R}\}$ of connections, called the α-connections. The most important connections statistically correspond to $\alpha = 0, \pm\frac{1}{3}, \pm 1$, as we now explain.

5

3.2 Choice of parametrisation.

For each of Amari's connections it can happen that a parametrisation θ of M exists such that the geodesic joining the points labelled θ_1 and θ_2 simply comprises of the points labelled $\{(1-\lambda)\theta_1 + \lambda\theta_2 : 0 \le \lambda \le 1\}$. For example Cartesian coordinates define such a parametrisation in the Euclidean case. When this happens M is said to be *flat*, such a parametrisation is called *affine*, and the parameters are unique up to affine equivalence. That is, any two affine parametrisations are related by an affine transformation. In the important special case of a metric connection M is flat if and only if there exists a parametrisation θ in which the metric tensor is independent of θ.

For a connection to admit an affine parametrisation is a rather special circumstance. When it does, we may expect the affine parametrisation to have correspondingly special properties. This is indeed the case with Amari's expected geometry. When an α-connection has this property, the manifold is called α-flat and the associated parametrisations are called α-affine. Amari (1990, Theorem 5.12, page 152), to whom we refer for further details, established the following characteristic features of certain α-affine parameters. We note below the α-affine parametrisations which have the particular statistical property shown.

1. $\alpha = 1$, natural.

2. $\alpha = \frac{1}{3}$, normal likelihood.

3. $\alpha = 0$, Constant asymptotic covariance of the MLE.

4. $\alpha = -\frac{1}{3}$, zero asymptotic skewness of the MLE.

5. $\alpha = -1$, zero asymptotic bias of the MLE.

These correspond to the $\delta = 0, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, 1$ parametrisations respectively of Hougaard (1982), who studied the one dimensional curved exponential family case. In any one dimensional family an α-affine parameter exists for every α. A full exponential family, of any dimension, is always $+1$-flat and -1-flat, with the natural and mean value parameters respectively being affine. Amari (1990) also establishes the duality result that M is α-flat if and only if it is $-\alpha$-flat. This duality between ∇^α and $\nabla^{-\alpha}$ has nice mathematical properties but has not been well understood statistically.

The following section contains an elementary account of the above aspects of Amari's geometry.

4 The expected geometry of the full exponential family.

4.1 Introduction.

We restrict attention now to the full exponential family. In the natural parametrisation, θ, we have

$$p(x, \theta) = \exp\{\sum_{i=1}^{p} t_i(x)\theta^i - \psi(\theta)\}.$$

The mean value parametrisation is given by $\eta = (\eta^1, \ldots, \eta^p)$, where

$$\eta^i(\theta) = \mathbf{E}_{p(x,\theta)}[t_i(x)] = \frac{\partial \psi}{\partial \theta^i}(\theta).$$

These two parametrisations are therefore affinely equivalent if and only if ψ is a quadratic function of θ, as with the normal distributions with constant covariance. As we shall see this is a very special circumstance.

In natural parameters and in the one dimensional case, the score function is

$$s(\theta, \mathbf{x}) = n\{\bar{t}(\mathbf{x}) - \psi'(\theta)\} = n\{\bar{t}(\mathbf{x}) - \eta(\theta)\} \tag{4}$$

where $n\bar{t}(\mathbf{x}) = \sum_{r=1}^{n} t(x_r)$. From (4) we have the useful fact that $\hat{\eta} := \eta(\hat{\theta}) = \bar{t}$. Further the two moments of the score function under $p(x, \phi)$ are given by,

$$\mu^\phi(\theta) = n\{\psi'(\phi) - \psi'(\theta)\} = n\{\eta(\phi) - \eta(\theta)\} \tag{5}$$

$$g^\phi(\theta) = n\psi''(\phi) = I(\phi). \tag{6}$$

4.2 Examples.

The following one dimensional examples are used for illustrative purposes throughout: Poisson, Normal with constant (unit) variance, Exponential and Bernoulli.

Although, of course, the sample size affects the ϕ-distribution of $\bar{t}$, it only enters the above equations for the score and its first two moments as a multiplicative constant. Therefore our analysis, which is based solely on these quantities is essentially invariant under independent repetitions. Our third and fourth examples therefore implicitly cover the Gamma and Binomial families. Together then, these examples embrace most of the distributions widely used in generalised linear models (McCullagh and Nelder, 1989).

7

The examples are summarised algebraically, if not very communicatively, in Table 1. Better, they are displayed visually in Figures 1 to 4 respectively. For each example, for a chosen ϕ and n, shown in Table 1, we give *observed* and *expected* plots, both in the natural parametrisation θ and in a non-affinely equivalent parametrisation $\xi(\theta)$.

	Poisson(θ) (Figure 1)	Normal(θ,1) (Figure 2)	Exponential(θ) (Figure 3)	Bernoulli(θ) (Figure 4)
$t(x)$	x	x	$-x$	x
$\psi(\theta)$	$\exp(\theta)$	$\frac{1}{2}\theta^2$	$-\ln\theta$	$\ln(1+\exp(\theta))$
$s(\theta,\mathbf{x})$	$n(\bar{x}-\exp(\theta))$	$n(\bar{x}-\theta)$	$n(-\bar{x}+\theta^{-1})$	$n(\bar{x}-e^\theta(1+e^\theta)^{-1})$
$\mu^\phi(\theta)$	$n(e^\phi-e^\theta)$	$n(\phi-\theta)$	$n(-\phi^{-1}+\theta^{-1})$	$n\frac{e^\phi}{1+e^\phi}-n\frac{e^\theta}{1+e^\theta}$
$g^\phi(\theta)$	$n\exp(\phi)$	n	$n\phi^{-2}$	$ne^\phi(1+e^\phi)^{-2}$
$\xi(\theta)$	$\eta(\theta)=e^\theta$	$\theta^{\frac{1}{3}}$	$\eta(\theta)=-\theta^{-1}$	$\eta(\theta)=e^\theta(1+e^\theta)^{-1}$
$\bar{B}(\theta)$	ξ^{-1}	$3\xi^2$	ξ^{-2}	$(\xi(1-\xi))^{-1}$
$s(\xi,\mathbf{x})$	$n(\bar{x}-\xi)\xi^{-1}$	$3n(\bar{x}-\xi^3)\xi^2$	$-n(\bar{x}+\xi)\xi^{-2}$	$n(\bar{x}-\xi)(\xi(1-\xi))^{-1}$
$\mu^{\xi(\phi)}(\xi)$	$n(\xi(\phi)-\xi)\xi^{-1}$	$3n(\xi^3(\varphi)-\xi^3)\xi^2$	$-n(\xi(\phi)-\xi)\xi^{-2}$	$n\frac{(\xi(\phi)-\xi)}{(\xi(1-\xi))}$
$g^{\xi(\phi)}(\xi)$	$n\xi(\phi)\xi^{-2}$	$9n\xi^4$	$n\xi(\phi)^2\xi^{-4}$	$n\frac{\xi(\phi)(1-\xi(\phi))}{\xi(1-\xi)}$
ϕ	0	0	1	0
n	10	20	10	10

Table 1: Examples.

We take $\xi(\theta)$ to be the mean value parameter $\eta(\theta)$ except in the normal case where we take $\xi(\theta)=\theta^{\frac{1}{3}}$. In each case, ξ is an increasing function of θ. In the expected plots, we illustrate the first two moments of the score function under the true distribution (that is under $p(x,\phi)$) by plotting the mean ± 2 standard deviations. In the observed plots, to give some idea of sampling variability, we plot five observed score functions corresponding to the 5%, 25%, 50% 75% and 95% points of the true distribution of $\bar{t}$. Recall that these plots

precisely contain the shape of the observed and expected likelihood functions, thus are a direct and visual representation of important statistical information.

The observed score graphs do not cross since for each fixed parameter value, the observed score function is non decreasing affine function of $\bar{t}$. This holds in all parametrisations, using (1). From (1), (2), (4) and (5) it is clear that in any parametrisation the graph of the true mean score function coincides with that of the observed score for data where $\bar{t}(\mathbf{x})$ equals its true mean $\eta(\phi)$. In the examples the true distribution of $n\bar{t}$ is given by Poisson($\phi + \ln n$), Normal($n\phi, n$), -Gamma(ϕ, n), Binomial(n, ϕ), respectively.

The most striking feature of the plots is the constancy of the variance of the score across the natural parametrisation, and the fact that this property is lost in the alternative parametrisation. Also remarkable is the linearity of the normal plots in the natural parametrisation. A close inspection reveals that for each example, in the natural parametrisation, each of the observed plots differ only by a vertical translation, again this property will not hold in general. We use these and other features of the plots to better understand Amari's expected geometry.

Certain information is evident from the plots straight away. Under standard regularity conditions, the unique maximum likelihood estimate of a parameter for given data occurs when the graph of the corresponding observed score function crosses the axis from above. Thus as $\bar{t} = \hat{\eta}$, in our examples, these five crossing points are the 5%, 25%, 50%, 75% and 95% percentage points of the true distribution of the maximum likelihood estimate. The position of these five crossing points gives visual information about this distribution, in particular, about its position, variance and skewness.

Of more direct relevance to our present concern is the fact that, in these one dimensional cases there is a straightforward visual representation of the tangent space at each point. TM_θ can be identified with the vertical line through θ, and ρ_θ^ϕ (see Section 2.3) with the distribution of the intersection of this line with the graph of the observed score function. Identical remarks apply in any parametrisation. These tangent spaces are shown in both parametrisations, at the above five percentage points of the maximum likelihood estimate, as lines in the observed plots and as vertical bars in the expected plots.

In the observed plot, the five intersection points with any given tangent space TM_θ, are the five corresponding percentage points of ρ_θ^ϕ. The same is true in any increasing reparametrisation ξ. Thus, comparing the position of these five intersection points at corresponding parameter values in the two observed

9

plots gives direct visual information on the difference between ρ_θ^ϕ and $\rho_{\xi(\theta)}^{\xi(\phi)}$. In particular, on changes in skewness.

This brings to light a certain natural duality between the MLE and the score function. Consider the observed plots in the natural and mean value parametrisations. For any given θ consider the corresponding tangent space TM_θ and $TM_{\eta(\theta)}$ in the two plots. In each plot we have five horizontal and five vertical crossing points, as above, giving information about the distribution of the MLE and the score function respectively in the same parametrisation. Now, these two plots are far from independent. As $\hat{\eta}(\mathbf{x}) = \eta(\theta) + n^{-1} s(\theta, \mathbf{x})$, the horizontal crossing points in the mean parameter plot are just an affine transformation of the vertical crossing points in the natural parameter plot. The converse is true asymptotically. As we discuss below, this simple and natural duality between the MLE and the score function corresponds with the duality present in Amari's expected geometry.

4.3 Amari's +1-geometry

The above one dimensional plots have already indicated two senses in which the natural parametrisation is very special. We note here that this is so generally. Our analysis then provides a simple statistical interpretation of Amari's +1-connection.

From (4) we see that in the natural parametrisation the score function has the form of a stochastic part, independent of θ, plus a deterministic part, independent of the data. Recalling (1) and (4) we see that this property is lost in a non affine reparametrisation ξ, since $\bar{B}(\theta)$ $(:= \bar{B}_1^1(\theta))$ is independent of θ if and only if ξ is an affine transformation of θ. An equivalent way to describe this property is that the 'error term' $\epsilon^\phi(\theta, \mathbf{x})$ in the mean value decomposition of $s(\theta, \mathbf{x})$ defined at the end of Section 2.3 is independent of θ Or again, as $\mu^\phi(\phi)$ vanishes, that this decomposition has the form

$$s(\theta, \mathbf{x}) = \mu^\phi(\theta) + s(\phi, \mathbf{x}). \tag{7}$$

Note that ρ_θ^ϕ differs from $\rho_{\theta'}^\phi$ only by the translation $\mu^\phi(\theta) - \mu^\phi(\theta')$.

In this parametrisation, from one sample to the next, the whole graph of the observed score function just shifts vertically about its ϕ-expectation by the same amount $s(\phi, \mathbf{x})$.

As a consequence of (7), the ϕ-covariance of the score function is independent of θ, (and therefore coincides with $g^\phi(\phi) = I(\phi)$). But $g^\phi(\theta)$ is a metric

10

tensor (Section 2.4), and in this parametrisation, the metric is constant across all tangent spaces. Recalling Section 3.2 we note that if a metric is constant in a parametrisation then the parametrisation is affine for the metric connection. All tangent spaces thus have the same geometric structure and differ only by their choice of origin. For more details on this geometric idea of flatness, see Dodson and Poston (1977).

The metric connection is the natural geometric tool for measuring the variation of a metric tensor in an arbitrary parametrisation. But Critchley, Marriott and Salmon (1993) prove that, in the full exponential family, the metric connection induced by $g^\phi(\theta)$ coincides with Amari's +1-connection. Thus we have the simple statistical interpretation that ∇^{+1} is the natural geometric measure of the non constancy of the covariance of the score function in an arbitrary parametrisation. In the one dimensional case, the +1-connection measures the change of variability of the observed score across different points of M.

Looking again at Figures 1 to 4 we see a visual representation of this fact is that the ± 2 standard deviation bars on the expected plot are of a constant length for the θ-parametrisation, and this does not hold in the non affine ξ-parametrisation.

4.4 Amari's 0-geometry.

The fact that in the natural parametrisation all the observed score functions have the same shape invites interpretation. From (7) we see that the common information conveyed in all of them is that conveyed by their ϕ-mean. What is it?

The answer is precisely the Fisher information for the family. This is clear since μ^ϕ determines I via

$$I_{ij}(\theta) = -\frac{\partial \mu_j^\phi}{\partial \theta^i}(\theta)$$

while the converse is true by integration, noting that $\mu^\phi(\phi) = 0$. Thus, in natural parameters, knowing the Fisher information at all points is equivalent to knowing the true mean of the score function, (and hence all the observed score functions up to their stochastic shift term). In particular, in the one dimensional case, the Fisher information is conveyed by the graph of $\mu^\phi(\theta)$ as, for example, in the natural parameter expected plots of Figures 1 to 4.

Amari uses the Fisher information as his metric tensor. It is important to note that when endowed with the corresponding metric connection an exponen-

tial family is not in general flat. That is, there does not, in general, exist any parametrisation in which the Fisher information is constant. The multivariate normal distribution with constant covariance matrix and any one dimensional family are notable exceptions. In the former case, the natural parameters are affine. In the latter case, using (3) the affine parameters are obtained as solutions to the equation

$$\bar{B}(\xi(\theta))^2 \psi''(\xi(\theta)) = \text{constant}.$$

For example in the Poisson family where $\psi = \exp(\theta)$ one finds $\xi(\theta) = \exp(\frac{\theta}{2})$ as in Hougaard (1982).

Thus far we have seen that, in the case of the full exponential family, the fundamental components of Amari's geometry (M, I, ∇^{+1}) can be simply and naturally understood in terms of the first two moments of the score function under the distribution assumed to give rise to the data. I is defined by the true mean ∇^{+1} by I and the true covariance. Further, they can be understood visually in terms of the expected plots in our one dimensional examples. We now go on to comment on duality and choice of parametrisation.

4.5 Amari's -1-geometry and duality.

The one dimensional plots above have already indicated a natural duality between the score vector and the MLE. Also that there is a natural statistical curvature, even in the one dimensional case unless the manifold is *totally flat*. That is, unless the graph of the true mean score function is linear in the natural parametrisation. We develop these remarks here.

Amari (1990) shows that the mean value parameters

$$\eta(\theta) = \mathbf{E}_{p(x,\theta)}[t(x)] = \psi'(\theta)$$

are -1-affine and therefore, by his general theory, duality related to the natural $+1$-affine parameters θ. We offer the following simple and direct statistical interpretation of the duality. We have,

$$\hat{\eta} = \eta(\theta) + n^{-1}s(\theta, \mathbf{x}).$$

Expanding $\theta(\hat{\eta})$ to first order about η gives an asymptotic converse

$$\hat{\theta} \dot{=} \theta + n^{-1}\bar{B}(\theta)s(\theta, \mathbf{x}) = \theta + n^{-1}s(\eta, \mathbf{x}),$$

12

the right hand equality following from (1) and we use $\doteq$ to denote first order asymptotic equivalence. Thus the duality between the $+1$ and -1 connections can be seen as the above strong and natural asymptotic correspondence between the MLE in one parametrisation and the score function in another. In fact this simple statistical interpretation of Amari's duality is not restricted to the full exponential family, see Critchley, Marriott and Salmon (1993). It is established formally in a more general case here in Section 4.7.

4.6   Total flatness and choice of parametrisation.

The above approximation to $\hat{\theta}$ is exact when θ and η are affinely equivalent. In this case, $\hat{\theta}$ and $\hat{\eta}$ are in the same affine relationship and so their distributions have the same shape. In particular, as normality is preserved under affine transformations, these distributions are as close to normality as each other *whatever* the definition of closeness that is used. In the case where M is a constant covariance normal family $\hat{\theta}$ and $\hat{\eta}$ are both exactly normally distributed.

Affine equivalence of θ and η is a very strong property. When it holds much more is true. It is the equivalent in the full exponential family case of the general geometric notion of *total flatness* defined and studied in Critchley, Marriott and Salmon (1992). Recall that the natural parametrisation θ has already been characterised by the fact that the true covariance of the score function is constant in it. Total flatness entails this same parametrisation simultaneously has other nice properties. It is easy to show the following equivalences,

$$\theta \text{ and } \eta \text{ are affinely equivalent}$$
$$\Longleftrightarrow \quad \psi \text{ is a quadratic function of } \theta$$
$$\Longleftrightarrow \quad I(\theta) \text{ is constant in the natural parameters}$$
$$\Longleftrightarrow \quad \mu^\phi(\theta) \text{ is an affine function of } \theta$$
$$\Longleftrightarrow \quad \exists\, \alpha \neq \beta \text{ with } \nabla^\alpha = \nabla^\beta$$
$$\Longleftrightarrow \quad \forall \alpha, \forall \beta, \quad \nabla^\alpha = \nabla^\beta$$
$$\Longleftrightarrow \quad \text{the } \theta \text{ parametrisation is } \alpha-\text{affine for all } \alpha$$

In particular, the MLE of any α-affine parameters are all equally close in (any sense) to normality.

It is exceptional for a family M to be totally flat. Constant covariance multivariate normal families are a rare example. In totally flat manifolds the graph of $\mu^\phi(\theta)$ is linear in the natural parametrisation, as remarked upon in the one dimensional normal example of Figure 2. More usually, even in the one

dimensional case, a family M of probability (density) functions will exhibit a form of curvature evidenced by the non linearity of the graph of $\mu^{\phi}(\theta)$.

Recall that the graph of $\mu^{\phi}(\theta)$ enables us to connect the distribution of $\hat{\theta}$ and $\hat{\eta}$. Each observed graph is a vertical shift of the expected graph. This shift is an affine function of $\bar{t} = \hat{\eta}$. The intersection of the observed plot with the θ axis determines $\hat{\theta}$. When the expected plot is linear (the totally flat case) then $\hat{\theta}$ and $\hat{\eta}$ are affinely related and so their distributions have the same shape. When it is non linear they will not be affinely related. This opens up the possibility that, in a particular sense of 'closeness' one of them will be closer to normality.

In all cases, the 0-geometry plays a pivotal role between the ± 1-geometries. That is, the graph of $\mu^{\phi}(\theta)$ determines the relationship between the distributions of the MLE's $\hat{\theta}$ and $\hat{\eta}$ of the ± 1-affine parameters. We illustrate this for our examples in Figure 5. Both distributions are of course exactly normal when the parent distribution is. In the Poisson case the concavity of $\mu^{\phi}(\theta)$ means that $\hat{\theta}$ is less positively skewed than $\hat{\eta}$. The opposite relationship holds in the Exponential case as $\mu^{\phi}(\theta)$ is convex there.

4.7 Amari's $\pm\frac{1}{3}$-geometry and duality.

Amari's $\frac{1}{3}$-connection can be simply interpreted in terms of linearity of the graph of the true mean score function, at least in the one dimensional situation where the $\frac{1}{3}$-affine parameters are known to exist. If M is totally flat, the graph is linear in the natural parametrisation, as in the normal constant covariance family. It is therefore natural to pose the question: Can a parametrisation be found for a general M in which this graph is linear?

This question can be viewed in two ways. Firstly, for some given $p(x, \phi)$, is such a parametrisation possible? However in this case, any parametrisation found could be a function of the true distribution. In general, there will not be a single parametrisation that works for all ϕ. The second way is to look *locally* to ϕ. This is the more fruitful approach statistically. The question then becomes: Can a single parametrisation $\theta \rightarrow \xi$ be found such that, for all ϕ, the graph of the true mean score is linear locally to $\xi = \xi(\phi)$? In the one dimensional case, we seek ξ such that

$$\forall \phi, \quad \frac{\partial^2 \mu^{\xi(\phi)}(\xi)}{\partial \xi^2}\Big|_{\xi=\xi(\phi)} = 0$$

Such a local approach is sufficient asymptotically when the observed score function will be close to its expected value and the maximum likelihood estimate will

be close to the true parameter. Thus in such a parametrisation, *whatever* the true value, the observed log likelihood will asymptotically be close to quadratic near the MLE. Hence the name, *normal likelihood* parameter. Amari (1990) shows that such parameters always exist for a one dimensional full exponential family, and that they are the $\frac{1}{3}$-affine parameters.

The vanishing of the second derivative of the true expected score function in one parametrisation ξ finds a dual echo in the vanishing of the asymptotic skewness of the true distribution of the MLE in another parametrisation λ. This is called the $-\frac{1}{3}$-affine parametrisation as it is induced by Amari's $-\frac{1}{3}$-connection. Note again that the duality is between the score function and the MLE as in Section 4.5. This can be formalised as follows.

Consider any one dimensional full exponential family,

$$p(x, \theta) = \exp\{t(x)\theta - \psi(\theta)\}.$$

Let ξ and λ be any two reparametrisations. Extending the approach in Section 4.5, it is easy to show the following equivalences:

$$\hat{\xi} \dot{=} \xi + n^{-1}s(\lambda, \mathbf{x}) \iff \hat{\lambda} \dot{=} \lambda + n^{-1}s(\xi, \mathbf{x}) \iff \frac{\partial \lambda}{\partial \theta}\frac{\partial \xi}{\partial \theta} = \psi''(\theta).$$

In this case, we say that ξ and λ are ψ-dual. Clearly, the natural (+1-affine) and mean value (−1-affine) parameters are ψ-dual. A parameter ξ is called self ψ-dual if it is ψ-dual to itself. In this case we find again the differential equation for the 0-affine parameters given in Section 4.4. More generally, it can be shown straightforwardly that for any $\alpha \in \mathbf{R}$

$$\xi \text{ and } \lambda \text{ are } \psi-\text{dual} \Rightarrow [\xi \text{ is } \alpha-\text{affine} \iff \lambda \text{ is } -\alpha-\text{affine}]$$

Thus the duality between the score function and the MLE coincides quite generally with the duality in Amari's expected geometry.

Note that the simple notion of ψ-duality gives an easy way to find $-\alpha$-affine parameters once $+\alpha$-affine parameters are known. For example, given that $\xi = \theta^{\frac{1}{3}}$ is $\frac{1}{3}$-affine in the exponential family (Hougaard, 1982) where $\psi(\theta) = -\ln(\theta)$, one immediately has

$$\frac{\partial \lambda}{\partial \theta} = 3\theta^{-\frac{4}{3}}$$

whence $\theta^{-\frac{1}{3}}$ is $-\frac{1}{3}$-affine. Again, in the Poisson family, $\xi = \exp(\theta/3)$ is $\frac{1}{3}$-affine gives at once that $\exp(2\theta/3)$ is $-\frac{1}{3}$-affine.

15

The local linearity of the true score in $+\frac{1}{3}$-parameters suggests that asymptotically the distribution of the MLE of the $\pm\frac{1}{3}$-affine parameters will be relatively close compared, for example, to the those of the ±1-affine parameters. In particular, it suggests that both will show little skewness. Figure 6, which may be compared to Figure 5(c), conveys this information for our Exponential family example.

References.

Amari, S. (1990), *Differential-Geometrical methods in Statistics, second edition.* Springer-Verlag: Berlin. Lecture Notes in Statistics No. 28.

Barndorff-Nielson, O.E., Cox D.R. and Reid N. (1986), The Role of Differential Geometry in Statistical Theory, *International Statistical Review,* **54**:83-96

Bates, D.M. and Watts, D.G. (1980), Relative curvature measures of nonlinearity, *J. Roy. Statist. Soc.,* B **40**: 1-25.

Bates, D.M. and Watts, D.G. (1981), Parametric transforms for improving approximate confidence regions in non-linear least squares, *Ann. Statist.,* **9**:1152-1167.

Cox, D.R., and Hinkley, D.V., (1974), *Theoretical Statistics,* Chapman and Hall: London.

Critchley F., P.K. Marriott and M. Salmon (1992), On the local differential geometry of the Kullback-Liebler divergence. *Univ. of Surrey research report.*

Critchley F., P.K. Marriott and M. Salmon (1993) Preferred Point Geometry and Statistical Manifolds, *Annals of Stats,* (To appear).

Dodson, C.T.J. and T. Poston (1977), *Tensor geometry.* Pitman: London.

Firth, D. (1993), Bias reduction of maximum likelihood estimates, *Biometrika,* **80**: 27-38.

Hougaard, P. (1982), Parametrisation of nonlinear models, *J. Roy. Statist. Soc B,* **44**:244-252.

Kass, R.E. (1984), Canonical parametrisation and zero parameter effects curvature, *J. Roy. Statist. Soc B,* **46**:86-92.

Kass, R.E. (1987), Introduction, *Differential Geometry in Statistical Inference,* Institute of Mathematical Statistics: Hayward, California.

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

McCullagh, P. and Nelder, J.A. (1989), *Generalised Linear Models,* Chapman and Hall: London, second edition.

Murray M. K. and J.W. Rice (1993) *Differential Geometry and Statistics.* Chapman and Hall: London.

Expected Plot: Natural parameter

Expected Plot: ξ-parametrisation

Observed Plot: Natural parameter

Observed Plot: ξ-parametrisation

Figure

Expected Plot: Natural parameter

Expected Plot: ξ-parametrisation

Observed Plot: Natural parameter

Observed Plot: ξ-parametrisation

Figure 2.
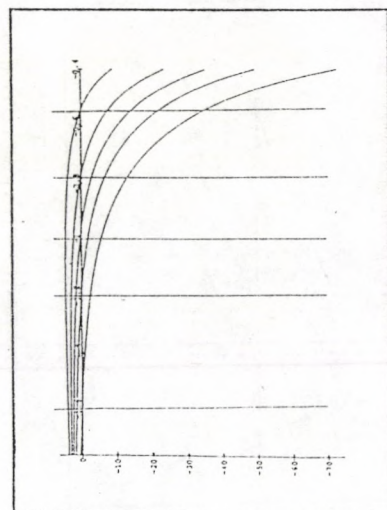
Expected Plot: Natural parameter

Expected Plot: ξ-parametrisation

Observed Plot: ξ-parametrisation

Observed Plot: Natural parameter

Figure 3.

Expected Plot: Natural parameter

Expected Plot: ξ-parametrisation

Observed Plot: Natural parameter

Observed Plot: ξ-parametrisation

Figure 4.

expected plot
Figure 5a. Poisson Case

distribution of $\hat{\eta}$

distribution of $\hat{\theta}$
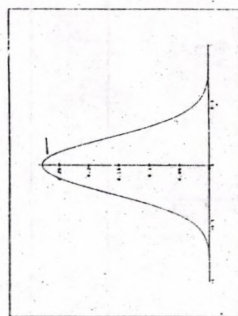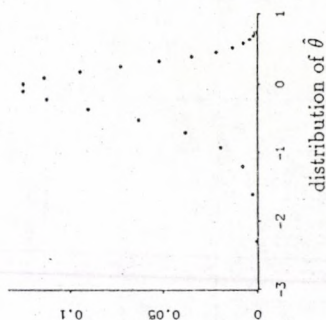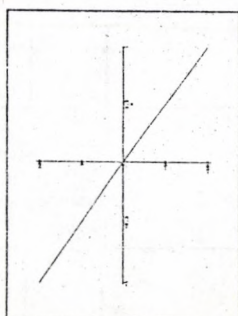
density of $\hat{\eta}$

expected plot
Figure 5b. Normal Case
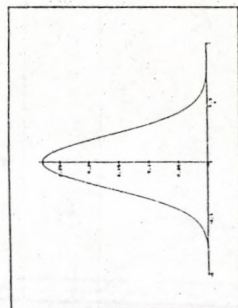
density of $\hat{\theta}$
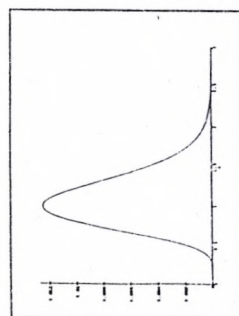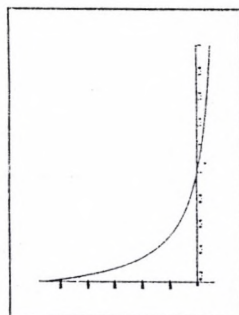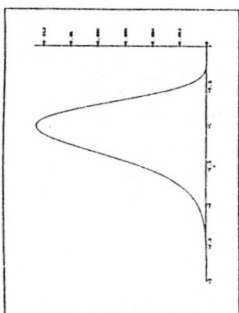
density of $\hat{\eta}$

expected plot
Figure 5c. Exponential Case

density of $\hat{\theta}$
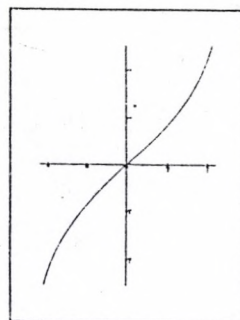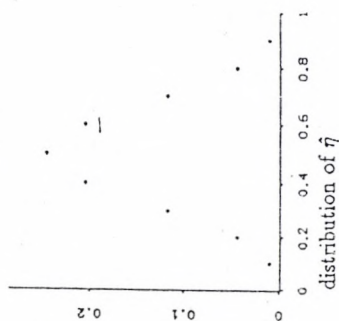
distribution of $\hat{\eta}$

expected plot
Figure 5d. Bernoulli Case

distribution of $\hat{\theta}$

23
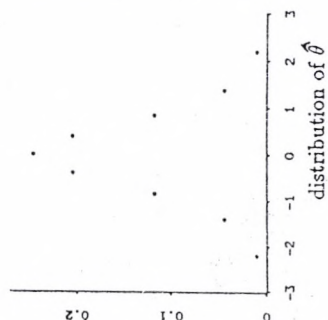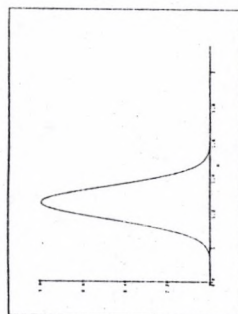
$-\frac{1}{3}$-parametrisation
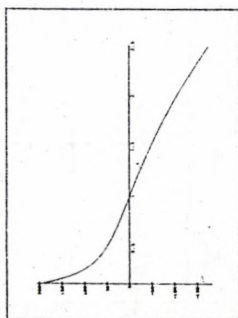
expected plot
Figure 6.
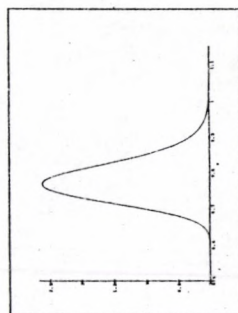
$+\frac{1}{3}$-parametrisation

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

Economics Department Working Paper Series

To   Department of Economics **WP**
European University Institute
Badia Fiesolana
I-50016 San Domenico di Fiesole (FI)
Italy

From   Name . . . .

Address. . . . .
. . . .
. . . .
. . . .
. . . .
(Please print)

❏ Please enter/confirm my name on EUI Economics Dept. Mailing List
❏ 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 1994/95

Please send me the following EUI **ECO** 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 1993

ECO No. 93/1
Carlo GRILLENZONI
Forecasting Unstable and Non-Stationary
Time Series

ECO No. 93/2
Carlo GRILLENZONI
Multilinear Models for Nonlinear Time
Series

ECO No. 93/3
Ronald M. HARSTAD/Louis PHLIPS
Futures Market Contracting When You
Don't Know Who the Optimists Are

ECO No. 93/4
Alan KIRMAN/Louis PHLIPS
Empirical Studies of Product Markets

ECO No. 93/5
Grayham E. MIZON
Empirical Analysis of Time Series:
Illustrations with Simulated Data

ECO No. 93/6
Tilman EHRBECK
Optimally Combining Individual
Forecasts From Panel Data

ECO NO. 93/7
Víctor GÓMEZ/Agustín MARAVALL
Initializing the Kalman Filter with
Incompletely Specified Initial Conditions

ECO No. 93/8
Frederic PALOMINO
Informed Speculation: Small Markets
Against Large Markets

ECO NO. 93/9
Stephen MARTIN
Beyond Prices Versus Quantities

ECO No. 93/10
José María LABEAGA/Angel LÓPEZ
A Flexible Demand System and VAT
Simulations from Spanish Microdata

ECO No. 93/11
Maozu LU/Grayham E. MIZON
The Encompassing Principle and
Specification Tests

ECO No. 93/12
Louis PHLIPS/Peter MØLLGAARD
Oil Stocks as a Squeeze Preventing
Mechanism: Is Self-Regulation Possible?

ECO No. 93/13
Pieter HASEKAMP
Disinflation Policy and Credibility: The
Role of Conventions

ECO No. 93/14
Louis PHLIPS
Price Leadership and Conscious
Parallelism: A Survey

ECO No. 93/15
Agustín MARAVALL
Short-Term Analysis of Macroeconomic
Time Series

ECO No. 93/16
Philip Hans FRANSES/Niels
HALDRUP
The Effects of Additive Outliers on Tests
for Unit Roots and Cointegration

ECO No. 93/17
Fabio CANOVA/Jane MARRINAN
Predicting Excess Returns in Financial
Markets

ECO No. 93/18
Iñigo HERGUERA
Exchange Rate Fluctuations, Market
Structure and the Pass-through
Relationship

ECO No. 93/19
Agustín MARAVALL
Use and Misuse of Unobserved
Components in Economic Forecasting

ECO No. 93/20
Torben HOLVAD/Jens Leth
HOUGAARD
Measuring Technical Input Efficiency for
Similar Production Units:
A Survey of the Non-Parametric
Approach

✳✳✳