STATISTICAL MANIFOLDS

Steffen L. Lauritzen*

*Institute for Electronic Systems, Aalborg University Center, Aalborg, Denmark

1. INTRODUCTION

Euclidean geometry has served as the major tool in clarifying the structural problems in connection with statistical inference in linear normal models. A similar elegant geometric theory for other statistical problems does not exist yet.

One could hope that a more general geometric theory could get the same fundamental role in discussing structural and other problems in more general statistical models.

In the case of non linear regression it seems clear that the geometric framework is that of a Riemannian manifold, whereas in more general cases it seems as if a non-standard differential geometry has yet to be developed.

The emphasis in the present paper is to clarify the abstract nature of this differential geometric object.

In section 2 we give a brief introduction to the notions of modern differential geometry that we need to carry out our study. It is an extract from Boothby (1975) and Spivak (1970-75) and we are mainly using a coordinate-free setup.

Section 3 is an ultrashort summary of some previous developments. The core of the paper is contained in section 4 where we abstract the notion of a <u>statistical manifold</u> as a triple ($\underline{M}$,g,D) where $\underline{M}$ is a manifold, g is a metric and D is a symmetric trivalent tensor, called the <u>skewness</u> of the manifold. Section 4 is fully devoted to a study of this abstract notion.

Sections 5, 6, 7, and 8 are detailed studies of some examples of statistical manifolds of which some (the Gaussian, the inverse Gaussian and the Gamma) manifolds are of interest because of their leading role in statistical theory, whereas the examples in section 8 are mostly of interest because they to a large extent produce counterexamples to many optimistic conjectures. Through the examples we also try to indicate possibilities for discussing geometric estimation procedures.

In section 9 we have tried to collect some of the questions that naturally arise in connection with the developments here and in related pieces of work.

2. SOME DIFFERENTIAL GEOMETRIC BACKGROUND

A topological manifold $\underline{M}$ is a Hausdorff space with a countable base such that each point $p \varepsilon \underline{M}$ has a neighborhood that is homeomorphic to an open subset of IR^m. m is the dimension of $\underline{M}$ and is well-defined. A differentiable structure on $\underline{M}$ is a family

$$U = (U_\lambda, \phi_\lambda)_{\lambda \varepsilon \Lambda}$$

where U_λ is an open subset of $\underline{M}$ and ϕ_λ are homeomorphisms from U_λ onto an open subset of IR^m, satisfying the following:

(1) $\bigcup_\lambda U_\lambda = \underline{M}$

(2) for any $\lambda_1, \lambda_2 \varepsilon \Gamma$: $\phi_{\lambda_1} \circ \phi_{\lambda_2}^{-1}$ is a $C^\infty(IR^m)$ function wherever it is well defined

(3) if V is open, ψ: $V \to IR^m$ is a homeomorphism, and $\psi \circ \phi_\lambda^{-1}$, $\phi_\lambda \circ \psi^{-1}$ are C^∞ wherever they are well defined, then $(V, \psi) \varepsilon U$.

The condition (2) is expressed as ϕ_{λ_1} and ϕ_{λ_2} being compatible.

In very simple cases $\underline{M}$ is itself homeomorphic to an open subset of IR^m and the differentiable structure is just given by $(\underline{M}, \phi_0)$ and all sets $(U_\lambda, \phi_\lambda)$ where U_λ is an open subset of $\underline{M}$ and $\phi_\lambda \circ \phi_0^{-1}$ is a diffeomorphism.

The sets U_λ are called coordinate neighborhoods and ϕ_λ coordinates. The pair $(U_\lambda, \phi_\lambda)$ is called a local coordinate system.

$\underline{M}$, equipped with a differentiable structure is called a differentiable manifold or a C^∞-manifold.

A differentiable structure can be specified by any system satisfying (1) and (2). Then there is a unique structure $\underline{U}$ containing the specified local coordinate system.

The differentiable structure gives rise to a natural way of defining a <u>differentiable function</u>. We say that $f\colon \underline{M} \to \mathbb{R}$ is in $C^\infty(\underline{M})$ if it is a usual C^∞-function when composed with the coordinates:

$$f \in C^\infty(\underline{M}) \leftrightarrow f \circ \phi_\lambda^{-1} \in C^\infty(\phi_\lambda(U)) \text{ for all } \lambda.$$

Important is the notion of a <u>regular submanifold</u> $\underline{N} \subseteq \underline{M}$ of $\underline{M}$. A subset $\underline{N}$ of $\underline{M}$ is a regular submanifold if it is a topological manifold with the relative topology and if it has <u>preferred coordinate neighborhoods</u>, i.e. to each point $p \in \underline{N}$ there is a local coordinate system $(U_\lambda, \phi_\lambda)$ with $p \in U_\lambda$ such that

i) $\phi_\lambda(p) = (0,\dots,0)$; $\phi_\lambda(U_\lambda) =]-\varepsilon,\varepsilon[^m$

ii) $\phi_\lambda(U_\lambda \cap \underline{N}) = \{(x^1,\dots,x^n,0,\dots,0),\ |x^i| < \varepsilon\}$

$\underline{N}$ inherits then in a natural way the differentiable structure from $\underline{M}$ by $(V_\lambda, \tilde{\phi}_\lambda)$ where

$$V_\lambda = U_\lambda \cap \underline{N},\ \tilde{\phi}_\lambda = \phi_\lambda | V_\lambda,$$

where $(U_\lambda, \phi_\lambda)$ is a preferred coordinate system.

All $C^\infty(\underline{N})$-functions can then be obtained by restriction to $\underline{N}$ of $C^\infty(\underline{M})$-functions.

For $p \in \underline{M}$, $C^\infty(p)$ is the set of functions whose restriction to some open neighborhood U of p is in $C^\infty(U)$. We here identify f and $g \in C^\infty(p)$ if their restriction to some open neighborhood of p are identical.

The <u>tangent space</u> $T_p(\underline{M})$ to $\underline{M}$ at p is now defined as the set of all maps $X_p\colon\ C^\infty(p) \to \mathbb{R}$ satisfying

i) $X_p(\alpha f+\beta g) = \alpha X_p(f)+\beta X_p(g)$ $\qquad \alpha,\beta \in \mathbb{R}$

ii) $X_p(fg) = X_p(f)g(p)+f(p)X_p(g)$ $\qquad f,g \in C^\infty(p)$

One should think of X_p as a <u>directional derivative</u>. X_p is called a <u>tangent vector</u>.

$T_p(\underline{M})$ is in an obvious way a vector-space and one can show that $\dim(T_p(\underline{M})) = m$.

For each particular choice of a coordinate system, there corresponds a canonical basis for $T_p(\underline{M})$, with basis vectors being

$$E_{ip}(f) = \frac{\partial}{\partial x^i} f(\phi^{-1}(x))|_{x=\phi(p)}$$

A <u>vector field</u> is a smooth family of tangent vectors $X = (X_p, p\varepsilon\underline{M})$ where $X_p \varepsilon T_p(\underline{M})$. To define "smooth" in the right way, we demand a vector field X to be a map:

$$X: \quad C^\infty(\underline{M}) \to C^\infty(\underline{M})$$

i) $\quad X(\alpha f+\beta g) = \alpha X(f)+\beta X(g) \quad \alpha,\beta\varepsilon IR$

ii) $\quad X(fg) = X(f)g+fX(g) \quad f,g\varepsilon C^\infty(\underline{M})$

and now we write

$$X_p(f) = X(f)\ (p)$$

The vector fields on $\underline{M}$ are denoted as $\underline{X}(\underline{M})$. $\underline{X}(\underline{M})$ is a <u>module</u> over $C^\infty(\underline{M})$: if $f,g\varepsilon C^\infty(\underline{M})$, $X,Y\varepsilon X(\underline{M})$ then

$$(fX+gY)\ (h) = fX(h) + gY(h)$$

is also in $\underline{X}(\underline{M})$. $\underline{X}(\underline{M})$ is a <u>Lie-algebra</u> with the bracket operation defined as

$$[X,Y](f) = X(Y(f)) - Y(X(f)).$$

The Lie-bracket [ ] satisfies

$$[X,[Y,Z]] + [Y,[Z,X]] + [Z,[X,Y]] = 0 \qquad \text{(Jacobi identity)}$$

$$[X,Y] = -[Y,X] \qquad \text{(anticommutativity)}$$

$$\left.\begin{array}{l}[\alpha X_1+\beta X_2,Y] = \alpha[X_1,Y] + \beta[X_2,Y] \quad \alpha,\beta \varepsilon IR \\ [X,\alpha Y_1+\beta Y_2] = \alpha[X,Y_1] + \beta[X,Y_2] \quad \alpha,\beta \varepsilon IR\end{array}\right. \qquad \text{(bilinearity)}$$

Further one can easily show that

$$[X,fY] = f[X,Y] + (X(f))Y .$$

The locally defined vector fields E_i, representing differentiation w.r.t. local coordinates, constitute a natural basis for the module $\underline{X}(U)$, where U is a coordinate neighborhood.

A <u>covariant</u> <u>tensor</u> D <u>of order k</u> is a C^∞-k-linear map

$$D: \quad \underline{X}(\underline{M})\times\ldots\times\underline{X}(\underline{M}) \to C^\infty(\underline{M}),$$

i.e.

$$D(X_1,\ldots,X_k) \in C^\infty(\underline{M}),$$

$$D(X_1,\ldots,fX_i+gY_i,X_{i+1},\ldots,X_k)$$

$$= fD(X_1,\ldots,X_k) + gD(X_1,\ldots,Y_i,X_{i+1},\ldots,X_k).$$

A tensor is always pointwise defined in the sense that if $X_{pi} = Y_{pi}$, then

$$D(X_1,\ldots,X_k)(p) = D(Y_1,\ldots,Y_k)(p).$$

This means that any equations for tensors can be checked locally on a basis e.g. of the form E_i. These satisfy $[E_i,E_j] = 0$ and all tensorial equations hold if they hold for vector fields with mutual Lie-brackets equal to zero. This is a convenient tool for proving tensorial equations and we shall make use of it in section 3.

A Riemannian metric g is a positive symmetric tensor of order two:

$$g(X,X) \geqq 0 \quad g(X,Y) = g(Y,X)$$

Since tensors are pointwise, it can be thought of as a metric g_p on each of the tangent spaces $T_p(\underline{M})$.

A curve $\gamma = (\gamma(t), t\varepsilon[a,b])$ is a C^∞-map of $[a,b]$ into $\underline{M}$. Note that a curve is more than the set of points on it. It involves effectively the parametrization and is thus not a purely geometric object.

Let now $\dot\gamma$ denote any vector field such that

$$\dot\gamma(f)(\gamma(t)) = \frac{\partial}{\partial t} f\,(\gamma(t)) \quad \text{for all } t\varepsilon[a,b], f\varepsilon C^\infty(\underline{M})$$

The length of the curve γ is now given as

$$|\gamma| = \int_a^b \sqrt{g(\dot\gamma,\dot\gamma)_{\gamma(t)}}\,dt.$$

Curve length can be shown to be geometric.

An important notion is that of an affine connection on a manifold. We define an affine connection as an operator ∇

$$\nabla: \ \underline{X}(\underline{M}) \times \underline{X}(\underline{M}) \to \underline{X}(\underline{M})$$

satisfying (where we write $\nabla_X Y$ for the value)

i) $$\nabla_X(\alpha Y+\beta Z) = \alpha\nabla_X Y + \beta\nabla_X Z, \quad \alpha,\beta \in \mathbb{R}$$

ii) $$\nabla_X(fY) = X(f)Y + f\nabla_X Y$$

iii) $$\nabla_{fX+gY}Z = f\nabla_X Z + g\nabla_Y Z \quad .$$

An affine connection can be thought of as a directional derivation of vector fields, i.e. $\nabla_X Y$ is the "change" of the vector field Y in X's direction.

An affine connection can be defined in many ways, the basic reason being, that "change" of Y is not well defined without giving a rule for comparing vectors in $T_{p_1}(\underline{M})$ with vectors in $T_{p_2}(\underline{M})$, since they generally are different spaces.

An affine connection is exactly defining such a rule via the notion of parallel transport, to be explained in the following. We first say that a vector field X is parallel along the curve γ if

$$\nabla_{\dot\gamma} X = 0 \text{ on } \gamma,$$

where again $\dot\gamma$ is any vector field representing $\frac{\partial}{\partial t}$.

Now for any vector $X_{\gamma(a)} \in T_{\gamma(a)}(\underline{M})$ there is a unique curve of vectors

$$X_{\gamma(t)}, t\in[a,b], \quad X_{\gamma(t)} \in T_{\gamma(t)}(\underline{M})$$

such that $\nabla_{\dot\gamma} X = 0$ on γ, i.e. such that these are all parallel, and such that $X_{\gamma(a)}$ is equal to the given one. We then write

$$X_{\gamma(b)} = \Pi_\gamma(X_{\gamma(a)})$$

and say that Π_γ defines parallel transport along γ. Π_γ is in general an affine map.

Note that Π_γ depends effectively on the curve in general.

An affine connection can be specified by choosing a local basis for the vector-fields (E_i, i=1,...,m) and defining the symbols (C^∞-functions)

$$\Gamma^k_{ij}, \; i,j,k=1,\dots,m$$

by

$$\nabla_{E_i} E_j = \Gamma^k_{ij} E_k \; (= \sum_{k=1}^m \Gamma^k_{ij} E_k),$$

where we adopt the summation convention that whenever an index appears in an expression as upper and lower, we sum over that index. Using the properties of an affine connection we thus have for an arbitrary pair of vector-fields

$$X = f^i E_i, \; Y = g^i E_i$$

$$\nabla_X Y = f^i E_i(g^j) E_j + f^i g^j \Gamma_{ij}^k E_k.$$

A <u>geodesic</u> is a curve with a parallel tangent vector field, i.e. where

$$\nabla_{\dot{\gamma}} \dot{\gamma} = 0 \text{ on } \gamma.$$

Associated with the notion of a <u>geodesic</u> is the <u>exponential map</u> induced by the connection.

For all $p \varepsilon \underline{M}$, $X_p \varepsilon T_p(\underline{M})$ there is a unique geodesic γ_{X_p}, such that

$$\gamma_{X_p}(0) = p \quad \dot{\gamma}_{X_p}(0) = X_p \qquad (**)$$

This is determined in coordinates by the differential equations below together with the initial conditions (**)

$$\ddot{x}^k(t) + \dot{x}^i(t)\dot{x}^j(t)\Gamma_{ij}^k(\underline{x}(t)) = 0$$

where $\gamma_{X_p}(t) = (x^1(t),\ldots,x^m(t))$ in coordinates.

Defining now for $X_p \varepsilon T_p(\underline{M})$

$$\exp\{X_p\} = \gamma_{X_p}(1)$$

we have $\exp\{tX_p\} = \gamma_{X_p}(t)$.

The exponential map is in general well defined at least in a neighborhood of zero in $T_p(\underline{M})$ and can only in special cases be defined globally.

In general, geodesics have no properties of "minimizing" curve length. However, on any Riemannian manifold, (i.e. a manifold with a metric tensor g), there is a <u>unique</u> affine connection $\bar{\nabla}$ satisfying

i) $$\bar{\nabla}_X Y - \bar{\nabla}_Y X - [X,Y] \equiv 0$$

ii) $$Xg(Y,Z) = g(\bar{\nabla}_X Y,Z) + g(Y,\bar{\nabla}_X Z).$$

This connection is called the <u>Riemannian connection</u> or the <u>Levi-Civita connection</u>.

Property i) is called <u>torsion-freeness</u> and property ii) means that <u>the parallel</u> transport $\bar{\Pi}$ <u>is isometric</u>, which is seen by the argument.

$$\dot{\gamma}g(Y,Z) = g(\bar{\nabla}_{\dot{\gamma}}Y,Z) + g(Y,\bar{\nabla}_{\dot{\gamma}}Z) = 0 \quad \text{if } \bar{\nabla}_{\dot{\gamma}}Y = \bar{\nabla}_{\dot{\gamma}}Z = 0.$$

We can then write $g(\bar{\Pi}_\gamma Y,\bar{\Pi}_\gamma Z)_{\gamma(b)} = g(Y,Z)_{\gamma(a)}$ or just $g(\bar{\Pi}_\gamma Y,\bar{\Pi}_\gamma Z) = g(Y,Z)$.

If $\bar{\nabla}$ is <u>Riemannian</u>, its geodesics <u>will</u> locally minimize curve length.

To all connections ∇ there is a torsion free connection $\tilde{\nabla}$ such that this has the same geodesics. <u>All connections in the present paper are torsion free</u>, whereas not all of them are Riemannian.

When the manifold is equipped with a Riemannian metric, it is often convenient to specify the connection through the symbols (C^∞-functions) Γ_{ijk}, where

$$\Gamma_{ijk} = g(\nabla_{E_i}E_j,E_k).$$

Defining the matrix of the metric tensor and its inverse as

$$g_{ij} = g(E_i,E_j) \quad (g^{ij}) = (g_{ij})^{-1},$$

the symbols are related to those previously defined as

$$\Gamma_{ij}^k = g^{kl}\Gamma_{ijl}.$$

The Riemannian connection is given by

$$\bar{\Gamma}_{ijk} = \tfrac{1}{2}(E_i(g_{jk}) + E_j(g_{ik}) - E_k(g_{ij})).$$

A connection defines in a canonical way the <u>covariant derivative</u> of a tensor D as

$$(\nabla_X D)(X_1,\dots,X_k) = XD(X_1,\dots,X_k) - \sum_{i=1}^{k} D(X_1,\dots,\nabla_X X_i,\dots,X_k).$$

$(\nabla_X D)$ is again a covariant tensor of order k and the map

$$S(X,X_1,\dots,X_k) = (\nabla_X D)(X_1,\dots,X_k)$$

becomes a tensor of order k+1. The fact that the Riemannian connection preserves inner product under parallel translation can then be written as

$$(\bar{\nabla}_X g)(Y,Z) \equiv 0.$$

Similarly, if D is a multilinear map from $\underline{X}(\underline{M})\times\dots\times\underline{X}(\underline{M})$ into $\underline{X}(\underline{M})$ its covariant derivative is defined as

$$(\nabla_X D)(X_1,\ldots,X_k) = \nabla_X D(X_1,\ldots,X_k) - \sum_{i=1}^{k} D(X_1,\ldots,\nabla_X X_i,\ldots,X_k).$$

Such multilinear maps are called tensor fields.

An important tensor field associated with a space with an affine connection is the curvature field, R: $\underline{X}(\underline{M}) \times \underline{X}(\underline{M}) \times \underline{X}(\underline{M}) \to \underline{X}(\underline{M})$

$$R(X,Y)Z = \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z.$$

A manifold with a connection satisfying $R \equiv 0$ is said to be flat. If the connection is torsion free, the curvature satisfies the following identities:

a) $$R(X,Y)Z = -R(Y,X)Z$$

b) $$R(X,Y)Z + R(Y,Z)X + R(Z,X)Y = 0$$

(Bianchi's 1st identity)

c) $$(\nabla_X R)(Y,Z,W) + (\nabla_Y R)(Z,X,W) + (\nabla_Z R)(X,Y,W) = 0$$

(Bianchi's 2nd identity).

Strictly speaking, a) does not need torsion freeness.

On a Riemannian manifold, we also define the curvature tensor $\bar{R}$ as

$$\bar{R}(X,Y,Z,W) = g(\bar{R}(X,Y)Z,W)$$

where $\bar{R}$ is used in two meanings, both referring to the Riemannian connection.

The Riemannian curvature tensor satisfies

i) $$\bar{R}(X,Y,Z,W) = -\bar{R}(Y,X,Z,W)$$

ii) $$\bar{R}(X,Y,Z,W) + \bar{R}(Y,Z,X,W) + \bar{R}(Z,X,Y,W) = 0$$

iii) $$\bar{R}(X,Y,Z,W) = -\bar{R}(X,Y,W,Z)$$

iv) $$\bar{R}(X,Y,Z,W) = \bar{R}(Z,W,X,Y)$$

We shall use the symbol R also for the curvature tensor

$$R(X,Y,Z,W) = g(R(X,Y)Z,W),$$

when $\underline{M}$ has a Riemannian metric g and a torsion-free but not necessarily Riemannian connection ∇. Then i) and ii) are satisfied, but not necessarily iii) and iv).

If (E_{ip}) is a local basis for $T_p(\underline{M})$, the curvature tensor can be calculated as

$$R_{ijkm} = R(E_i,E_j,E_k,E_m)$$
$$= (E_i(\Gamma^s_{jk}) - E_j(\Gamma^s_{ik}))g_{sm} + (\Gamma_{irm}\Gamma^r_{jk} - \Gamma_{jrm}\Gamma^r_{ik}).$$

The <u>sectional curvature</u> is given as

$$K(\sigma_{X,Y}) = \frac{g(R(X,Y)Y,X)}{g(X,X)g(Y,Y)-g(X,Y)^2}$$

and determines in a <u>Riemannian</u> manifold also the curvature. If the curvature satisfies i) to iv) the sectional curvature also determines R.

Two other contractions of the curvature tensor are of interest: The <u>Ricci-curvature</u>

$$c_1R(X,X) = \sum_{i=1}^{m-1} g(R(u_i,X)X,u_i)$$
$$= g(X,X)\sum_{i=1}^{m-1} K(\sigma_{X,u_i})$$

where $(X/g(X,X),u_1,\ldots,u_{m-1})$ is an orthonormal system for $T_p(\underline{M})$.

Finally the <u>scalar curvature</u> is

$$S(p) = \sum_{i=1}^{m} c_1R(u_i,u_i)$$

where $u_1,\ldots,u_m$ is an orthonormal system in $T_p(\underline{M})$. We then have the identity

$$S(p) = \sum_{i,j} K(\sigma_{u_i},u_j).$$

If $\underline{N}$ is a regular submanifold of $\underline{M}$, the tangent space of $\underline{N}$ can in a natural way be identified with the subspace of $\underline{X}(\underline{M})$ determined by

$$X \in \underline{X}(\underline{N}) \subseteq \underline{X}(\underline{M}) \leftrightarrow [f=g \text{ on } \underline{N} \rightarrow X(f) = X(g) \text{ on } \underline{N}].$$

In that way all tensors etc. can be inherited to $\underline{N}$ by restriction. If $\underline{M}$ has a Riemannian metric, $\underline{N}$ inherits it in an obvious way, and this preserves curve length, in the sense that the length of a curve in $\underline{N}$ w.r.t. the metric inherited, is equal to that when the curve is considered as a curve in M.

An affine connection is inherited in a more complicated way: We define

$$({}_{\underline{N}}\nabla_X Y)(p) = P_p(\nabla_X Y)(p)$$

where P_p is the projection w.r.t. g onto the tangent space $T_p(\underline{N}) \subseteq T_p(\underline{M})$ of the vector $(\nabla_X Y)_p$ which is not necessarily in $T_p(\underline{N})$. In fact we define the embedding curvature of $\underline{N}$ relative to $\underline{M}$ as the tensor field $\underline{X}(\underline{N}) \times \underline{X}(\underline{N}) \to \underline{X}(\underline{M})$

$$H_{\underline{N}}(X,Y) = \nabla_X Y - {}_{\underline{N}}\nabla_X Y$$

or, equivalently, as

$$H_{\underline{N}}(X,Y,Z) = g(H_{\underline{N}}(X,Y),Z)$$

where $X,Y \in \underline{X}(\underline{N})$, $Z_p \in \underline{X}(\underline{N})_p^{\perp}$ (or $Z \in \underline{X}(\underline{M})$).

If $H_{\underline{N}} \equiv 0$ we say that $\underline{N}$ is a totally geodesic submanifold of $\underline{M}$. A totally geodesic submanifold has the property that any curve in $\underline{N}$ which is a geodesic w.r.t. the connection on $\underline{N}$, also is a geodesic in $\underline{M}$.

3. THE DIFFERENTIAL GEOMETRY OF STATISTICAL MODELS

A family of probability measures $\underline{P}$ on a topological space $\underline{X}$ inherits its topological structure from the weak topology. Most statistical models are <u>parametrized</u> at least locally by maps (homeomorphisms)

$$\psi: U \to \Theta \subseteq \mathbb{R}^m$$

where U is an open subset of $\underline{P}$ and Θ an open subset of $\mathbb{R}^m$. From this parametrization we get $\underline{P}$ equipped with a differentiable structure, provided the various local parametrizations are <u>compatible</u>. Considering for a while only local aspects, we can think of $\underline{P}$ as $\{P_\theta, \theta\varepsilon\Theta\}$. We let now $f(x,\theta)$ denote the density of P_θ w.r.t. a dominating measure μ and assume these to be C^∞-functions of θ. Under suitable regularity assumptions we can now equip $\underline{P}$ with a <u>Riemannian metric</u> by defining $l(x,\theta) = \log f(x,\theta)$ and

$$g_{ij}(\theta) = g(E_i,E_j)_{P_\theta} = E_\theta(E_i(l)E_j(l)).$$

The metric is the <u>Fisher information</u> and different parametrizations define the same metric on $\underline{P}$. Similarly we can define a family of affine connections (the α-connections) on $\underline{P}$ by the expressions

$$\overset{\alpha}{\Gamma}_{ijk} = \bar{\Gamma}_{ijk} - \frac{\alpha}{2}T_{ijk}, \ \alpha\varepsilon\mathbb{R}, \text{ where}$$

$$T_{ijk}(P_\theta) = E_\theta\{E_i(l)E_j(l)E_k(l)\}, \text{ and}$$

$\bar{\Gamma}_{ijk}$ is the Riemannian connection.

The Fisher information as a metric was first studied by Rao (1945) and the α-connections in the case of finite and discrete sample spaces by Chentsov (1972). Later the α-connections were introduced and investigated independently and in full generality by Amari (1982).

For a more fair description of the history of the subject (the above is indecently short), see e.g. the introduction by Kass in the present monograph, Amari (1985) and/or Barndorff-Nielsen, Cox and Reid (1986).

Two of these connections play a special role:

The exponential connection (for $\alpha=1$) and

the mixture connection (for $\alpha=-1$).

The exponential connection has $\overset{1}{\Gamma}_{ijk} \equiv 0$ when expressed in the <u>canonical parameter</u> in an exponential family, and similarly when we express $\overset{-1}{\Gamma}_{ijk}$ (the mixture connection) in the <u>mean value coordinates</u> of an exponential family $\overset{-1}{\Gamma}_{ijk} \equiv 0$. Further we have the formulae

$$\overset{1}{\Gamma}_{ijk} = \mathbb{E}(E_i E_j(1) E_k(1)) \text{ and}$$

$$T_{ijk} = 2(\bar{\Gamma}_{ijk} - \overset{1}{\Gamma}_{ijk})$$

which often are useful for computations.

These structures are in a certain sense canonical on a statistical manifold. Chentsov (1972) showed in the case of discrete sample spaces that the α-connections were the only invariant connections satisfying certain invariance properties related to a decision-theoretic approach. Similarly, the Fisher information metric is the only invariant Riemannian metric. These results have recently been generalized to exponential families by Picard (1985).

On the other hand, similar geometric structures have recently appeared such as minimum-contrast geometries (Eguchi, 1983) and the observed geometries introduced by Barndorff-Nielsen in this monograph.

The common structure that seems to appear again and again in current statistical literature is not standard in modern geometry since it involves study of the interplay between a Riemannian metric and a non-Riemannian connection or even a whole family of such connections.

It seems thus worthwhile to spend some time on studying this structure from a purely mathematical point of view. This has already been done to some extent by Amari (1985). In the following section we shall outline the mathematical structures.

4. STATISTICAL MANIFOLDS

A statistical manifold is a Riemannian manifold with a symmetric and covariant tensor D or order 3. In other words a triple $(\underline{M},g,D)$ where $\underline{M}$ is an m-dimensional C^∞-manifold, g is a metric tensor and D: $\underline{X}(\underline{M}) \times \underline{X}(\underline{M}) \times \underline{X}(\underline{M}) \to C^\infty(\underline{M})$ a trilinear map satisfying

$$D(X,Y,Z) = D(Y,X,Z) = D(Y,Z,X)$$
$$(=D(X,Z,Y) = D(Z,X,Y) = D(Z,Y,X))$$

D is going to play the role T_{ijk} in the previous section. We use D to distinguish the tensor from the torsion field. D is called the skewness of the manifold.

Instead of D we shall sometimes consider the tensor field $\tilde{D}$ defined as

$$g(\tilde{D}(X,Y),Z) = D(X,Y,Z).$$

We have here used that the value of a tensor field is fully determined when the inner product with an arbitrary vector field Z is known for all Z.

The above defined notion could seem a bit more general than necessary, in the sense that some Riemannian manifolds with a symmetric trivalent tensor D might not correspond to a particular statistical model.

On the other hand the notion is general enough to cover all known examples, including the observed geometries studied by Barndorff-Nielsen and the minimum contrast geometries studied by Eguchi (1983).

Further, all known results of geometric nature for statistical manifolds as studied by Amari and others can be shown in this generality and it seems difficult to restrict the geometric structure further if all known examples should be covered by the general notion.

Let now $(\underline{M},g,D)$ (or$(\underline{M},g,\tilde{D})$) be a statistical manifold. We now define a family of connections as follows:

$$\overset{\alpha}{\nabla}_X Y = \bar{\nabla}_X Y - \frac{\alpha}{2}\tilde{D}(X,Y) \tag{3.1}$$

where $\bar{\nabla}$ is the Riemannian connection. We then have

<u>3.1 Proposition</u> $\overset{\alpha}{\nabla}$ <u>as defined by</u> (3.1) <u>is a torsion free connection</u>. <u>It is the unique connection that is torsion free and satisfies</u>

$$(\overset{\alpha}{\nabla}_X g)(Y,Z) = \alpha D(X,Y,Z) \tag{3.2}$$

<u>Proof</u>: That $\overset{\alpha}{\nabla}$ is a connection: Linearity in X is obvious. Scalar linearity in Y as well. We have

$$\overset{\alpha}{\nabla}_X(fY) = \bar{\nabla}_X(fY) - \frac{\alpha}{2}\tilde{D}(X,fY) = X(f)Y + f\overset{\alpha}{\nabla}_X Y.$$

Torsion freeness follows from symmetry of $\tilde{D}$:

$$\overset{\alpha}{\nabla}_X Y - \overset{\alpha}{\nabla}_Y X - [X,Y] = \bar{\nabla}_X Y - \bar{\nabla}_Y X - [X,Y]$$

$$-\frac{\alpha}{2}[\tilde{D}(X,Y) - \tilde{D}(Y,X)] = 0.$$

That $\overset{\alpha}{\nabla}$ satisfies (3.2) follows from

$$(\overset{\alpha}{\nabla}_X g)(Y,Z) = Xg(Y,Z) - g(\overset{\alpha}{\nabla}_X Y,Z) - g(Y,\overset{\alpha}{\nabla}_X Z)$$

$$= (\bar{\nabla}_X g)(Y,Z) + \alpha D(X,Y,Z) = 0 + \alpha D(X,Y,Z).$$

If $\tilde{\nabla}$ is torsion free and satisfies (3.2) we obtain:

i) $$Xg(Y,Z) = g(\tilde{\nabla}_X Y,Z) + g(Y,\tilde{\nabla}_X Z) + \alpha D(X,Y,Z)$$

ii) $$Zg(X,Y) = g(\tilde{\nabla}_X Z,Y) + g(\tilde{\nabla}_Y Z,X) + \alpha D(X,Y,Z) + g([Z,X],Y) + g([Z,Y],X)$$

iii) $$Yg(Z,X) = g(\tilde{\nabla}_Y Z,X) + g(\tilde{\nabla}_X Y,Z) + \alpha D(X,Y,Z) - g([X,Y],Z)$$

Calculating now i) - ii) + iii) we get

$$Xg(Y,Z) - Zg(X,Y) + Yg(Z,X) = \alpha D(X,Y,Z)$$
$$-g([Z,X],Y) - g([Z,Y],X) - g([X,Y],Z) + 2g(\tilde{\nabla}_X Y,Z).$$

Since this equation also is fulfilled for $\overset{\alpha}{\nabla}$ we get

$$g(\overset{\sim}{\nabla}_X Y,Z) = g(\overset{\alpha}{\nabla}_X Y,Z), \text{ whereby } \overset{\sim}{\nabla} = \overset{\alpha}{\nabla}.$$

Obviously $\overset{0}{\nabla} = \bar{\nabla}$, the Riemannian connection.

To check what happens when we make a parallel translation we first consider the notion of a conjugate connection (Amari, 1983).

Let ($\underline{M}$,g) be a Riemannian manifold and ∇ an affine connection. The conjugate connection ∇^* is defined as

$$g(\nabla^*_X Y,Z) = Xg(Y,Z) - g(Y,\nabla_X Z) \tag{3.3}$$

3.2 Lemma ∇^* is a connection. $(\nabla^*)^* = \nabla$.

Proof: Linearity in X is obvious. So is linearity in Y w.r.t. scalars. We have

$$g(\nabla^*_X(fY),Z) = Xg(fY,Z) - g(fY,\nabla_X Z)$$

$$= X(f)g(Y,Z) + fXg(Y,Z) - fg(Y,\nabla_X Z)$$

$$= g(X(f)Y + f\nabla^*_X Y,Z).$$

And further

$$g((\nabla^*)^*_X Y,Z) = Xg(Y,Z) - g(\nabla^*_X Z,Y)$$

$$= Xg(Y,Z) - \{Xg(Z,Y) - g(\nabla_X Y,Z)\} = g(\nabla_X Y,Z).$$

If we now let Π_γ, Π^*_γ denote parallel transport along the curve γ we obtain:

3.3 Proposition

$$g(\Pi_\gamma X,\Pi^*_\gamma Y) = g(X,Y)$$

Proof: Let X be ∇-parallel along γ and Y ∇^*-parallel. Then we have

$$\dot{\gamma}g(X,Y) = g(\nabla_{\dot{\gamma}}X,Y) + g(X,\nabla^*_{\dot{\gamma}}Y) = 0.$$

In words Proposition 3.3 says that parallel transport of pairs of vectors w.r.t. a pair of conjugate connections is "isometric" in the sense that inner product is preserved.

Finally we have for the α-connections, defined by (3.1):

3.4 Proposition $(\overset{\alpha}{\nabla})^* = \overset{-\alpha}{\nabla}$.

Proof:

$$g(\overset{\alpha}{\nabla}_X Y,Z) = g(\bar{\nabla}_X Y,Z) - \frac{\alpha}{2} D(X,Y,Z)$$

$$g(Y,\overset{-\alpha}{\nabla}_X Z) = g(Y,\bar{\nabla}_X Z) + \frac{\alpha}{2} D(X,Z,Y)$$

Adding and using the symmetry of D together with the defining property of the Riemannian connection we get

$$g(\overset{\alpha}{\nabla}_X Y,Z) + g(Y,\overset{-\alpha}{\nabla}_X Z) = Xg(Y,Z) \tag{3.4}$$

The relation (3.4) is important and was also obtained by Amari (1983). If we now consider the curvature tensors R and R* corresponding to ∇ and ∇^* we obtain the following identity:

3.5 Proposition

$$R(X,Y,Z,W) = -R^*(X,Y,W,Z) \tag{3.5}$$

Proof: Since we shall show a tensorial identity, we can assume $[X,Y] = 0$ as discussed in section 1. Then we get

$$\begin{aligned} XYg(Z,W) &= X(g(\nabla_Y Z,W) + g(Z,\nabla^*_Y W)) \\ &= g(\nabla_X \nabla_Y Z,W) + g(\nabla_Y Z,\nabla^*_X W) \\ &\quad + g(\nabla_X Z,\nabla^*_Y W) + g(Z,\nabla^*_X \nabla^*_Y W). \end{aligned}$$

By alternation we obtain

$$\begin{aligned} 0 = [X,Y]g(Z,W) &= XYg(Z,W) - YXg(Z,W) \\ &= R(X,Y,Z,W) + R^*(X,Y,W,Z). \end{aligned}$$

Note that the Riemannian connection is self-conjugate which gives the well known identity for the Riemannian curvature tensor, see section 1.

Consequently we obtain

3.6 Corollary The following conditions are equivalent

i) $R = R^*$

ii) $R(X,Y,Z,W) = -R(X,Y,W,Z)$

Proof: It follows directly from (3.5).

And, also as a direct consequence:

3.7 Corollary ∇ is flat if and only if ∇^* is.

The identity ii) is not without interest and we shall shortly investigate for which classes of statistical manifolds this is true. Before we get to that point we shall investigate the relation between a statistical manifold and a manifold with a pair of conjugate connections.

We define the tensor field $\tilde{D}_1$, and the tensor D_1 in a manifold with a pair (∇,∇^*) of conjugate connections by

$$\tilde{D}_1(X,Y) = \nabla^*_X Y - \nabla_X Y$$

$$D_1(X,Y,Z) = g(\tilde{D}_1(X,Y),Z).$$

We then have the following

<u>3.8 Proposition</u> <u>If</u> ∇ <u>is torsion free, the following are equivalent</u>

i) ∇^* <u>is torsion free</u>

ii) D_1 <u>is symmetric</u>

iii) $\bar{\nabla} = \frac{1}{2}(\nabla+\nabla^*)$

<u>Proof</u>: That D_1 is symmetric in the last two arguments follows from the calculation

$$\begin{aligned} D_1(X,Y,Z) &= g(\nabla^*_X Y,Z) - g(\nabla_X Y,Z) \\ &= Xg(Y,Z) - g(Y,\nabla_X Z) - [Xg(Y,Z)-g(Y,\nabla^*_X Z)] \\ &= D_1(X,Z,Y) \end{aligned}$$

The difference between two connections is always a tensor field. i) $\leftrightarrow$ ii) follows from the calculation

$$\begin{aligned} g(\nabla^*_X Y-\nabla^*_Y X-[X,Y],Z) &= g(\nabla_X Y-\nabla_Y X-[X,Y],Z) \\ &\quad + D_1(X,Y,Z) - D_1(Y,X,Z). \end{aligned}$$

That iii) $\rightarrow$ i) is obvious since then $\nabla^*=2\bar{\nabla}-\nabla$.

To show that i) $\rightarrow$ iii) we use the uniqueness of the Riemannian connection. We define

$$\tilde{\nabla} = \frac{1}{2}(\nabla+\nabla^*)$$

and see that this is torsion free, when ∇ and ∇^* both are. But

$$g(\tilde{\nabla}_X Y,Z) + g(Y,\tilde{\nabla}_X Z) = \tfrac{1}{2}g(\nabla_X Y,Z) + \tfrac{1}{2}g(\nabla^*_X Y,Z)$$
$$+ \tfrac{1}{2}g(Y,\nabla^*_X Z) + \tfrac{1}{2}g(Y,\nabla_X Z) = Xg(Y,Z)$$

showing that $\tilde{\nabla}$ is Riemannian and thus equal to $\bar{\nabla}$.

Suppose now that ∇ is given with ∇^* being torsion free. We can then define a family of connections as

$$\overset{\alpha}{\tilde{\nabla}}_X Y = \bar{\nabla}_X Y - \frac{\alpha}{2}\tilde{D}_1(X,Y)$$

and we obtain

<u>3.9 Corollary</u> $\overset{\alpha}{\tilde{\nabla}}{}^* = \overset{-\alpha}{\tilde{\nabla}}$, $\overset{1}{\tilde{\nabla}} = \nabla$, $\overset{-1}{\tilde{\nabla}} = \nabla^*$.

<u>Proof</u>: It is enough to show $\overset{1}{\tilde{\nabla}} = \nabla$. But

$$\overset{1}{\tilde{\nabla}}_X Y = \tfrac{1}{2}(\nabla_X Y + \nabla^*_X Y) - \tfrac{1}{2}(\nabla^*_X Y - \nabla_X Y) = \nabla_X Y.$$

We have thus established a one-to-one correspondence between a statistical manifold ($\underline{M}$,g,D) and a Riemannian manifold with a connection ∇ whose conjugate ∇^* is torsion free, the relation being given as

$$D(X,Y) = \nabla^*_X Y - \nabla_X Y$$
$$\nabla_X Y = \bar{\nabla}_X Y - \tfrac{1}{2}D(X,Y).$$

In some ways it is natural to think of the statistical manifolds as being induced by the metric (Fisher information) and <u>one</u> connection (∇) (the exponential), but the representation ($\underline{M}$,g,D) is practical for mathematical purposes, because D has simpler transformational properties than ∇.

By direct calculation we further obtain the following identity for a statistical manifold and its α-connections

<u>3.10 Proposition</u>

$$g(\overset{\alpha}{\nabla}_X Y,Z) - g(\overset{\alpha}{\nabla}_X Z,Y) = g(\bar{\nabla}_X Y,Z) - g(\bar{\nabla}_X Z,Y) \tag{3.6}$$

<u>Proof</u>: The result follows from

$$g(\overset{\alpha}{\nabla}_X Y,Z) - g(\overset{\alpha}{\nabla}_X Z,Y) = g(\bar{\nabla}_X Y,Z) - g(\bar{\nabla}_X Z,Y)$$
$$- \frac{\alpha}{2}D(X,Y,Z) + \frac{\alpha}{2}D(X,Z,Y)$$

and the symmetry of D.

We shall now return to studying the question of identities for the curvature tensor of a statistical manifold. We define the tensor

$$F(X,Y,Z,W) = (\bar{\nabla}_X D)(Y,Z,W)$$

where D is the skewness of the manifold, and $\bar{\nabla}$ is the Riemannian connection. We then have

<u>3.11 Proposition</u> <u>The following are equivalent</u>

i) $\overset{\alpha}{R} = \overset{-\alpha}{R}$ <u>for all</u> $\alpha \in \mathbb{R}$

ii) F <u>is symmetric</u>

<u>Proof</u>: The proof reminds a bit of bookkeeping. We are simply going to establish the identity

$$\overset{-\alpha}{R}(X,Y,Z,W) - \overset{\alpha}{R}(X,Y,Z,W) = \alpha\{F(X,Y,Z,W) - F(Y,X,Z,W)\} \tag{3.7}$$

by brute force.

Symmetry of F in the last three variables follows from the symmetry of D. We have

$$2\alpha F(X,Y,Z,W) = 2\alpha XD(Y,Z,W)$$
$$-2\alpha(D(\bar{\nabla}_X Y,Z,W) + D(Y,\bar{\nabla}_X Z,W) + D(Y,Z,\bar{\nabla}_X W))$$

Since $\bar{\nabla} = \frac{1}{2}(\overset{\alpha}{\nabla} + \overset{-\alpha}{\nabla})$ and $\alpha D(X,Y,Z) = g(\overset{-\alpha}{\nabla}_X Y,Z) - g(\overset{\alpha}{\nabla}_X Y,Z)$ we further get

$$\begin{aligned}2\alpha D(\bar{\nabla}_X Y,Z,W) &= 2g(\overset{-\alpha}{\nabla}_Z W,\bar{\nabla}_X Y) - 2g(\overset{\alpha}{\nabla}_Z W,\bar{\nabla}_X Y)\\ &= g(\overset{-\alpha}{\nabla}_Z W,\overset{\alpha}{\nabla}_X Y) + g(\overset{-\alpha}{\nabla}_Z W,\overset{-\alpha}{\nabla}_X Y)\\ &- g(\overset{\alpha}{\nabla}_Z W,\overset{\alpha}{\nabla}_X Y) - g(\overset{\alpha}{\nabla}_Z W,\overset{-\alpha}{\nabla}_X Y),\end{aligned}$$

and similarly for the two other terms. Further we get

$$\begin{aligned}2\alpha XD(Y,Z,W) &= 2X(g(\overset{-\alpha}{\nabla}_Y Z,W) - g(\overset{\alpha}{\nabla}_Y Z,W))\\ &= 2g(\overset{-\alpha}{\nabla}_X\overset{-\alpha}{\nabla}_Y Z,W) - 2g(\overset{\alpha}{\nabla}_X\overset{\alpha}{\nabla}_Y Z,W)\\ &+ 2g(\overset{-\alpha}{\nabla}_Y Z,\overset{\alpha}{\nabla}_X W) - 2g(\overset{\alpha}{\nabla}_Y Z,\overset{-\alpha}{\nabla}_X W)\end{aligned}$$

Collecting terms we get the following table of terms in $2\alpha F(X,Y,Z,W)$, where lines 1-3 are from $2\alpha XD(Y,Z,W)$, 4 and 5 from $2\alpha D(\bar{\nabla}_X Y,Z,W)$ 6 and 7 from $2\alpha D(Y,\bar{\nabla}_X Z,W)$ and 8 and 9 from $2\alpha D(Y,Z,\bar{\nabla}_X W)$.

Table of terms of $2\alpha F(X,Y,Z,W)$

	with + sign	with - sign
1.	$2g(\overset{-\alpha}{\nabla}_X \overset{-\alpha}{\nabla}_Y Z,W)$	$2g(\overset{\alpha}{\nabla}_X\overset{\alpha}{\nabla}_Y Z,W)$
2.	$g(\overset{-\alpha}{\nabla}_X Y,\overset{\alpha}{\nabla}_X W)$	$g(\overset{\alpha}{\nabla}_Y Z,\overset{-\alpha}{\nabla}_X W)$
3.	$g(\overset{-\alpha}{\nabla}_X Y,\overset{\alpha}{\nabla}_X W)$	$g(\overset{\alpha}{\nabla}_Y Z,\overset{-\alpha}{\nabla}_X W)$
4.	$g(\overset{\alpha}{\nabla}_Z W,\overset{\alpha}{\nabla}_X Y)$	$g(\overset{-\alpha}{\nabla}_Z W,\overset{\alpha}{\nabla}_X Y)$
5.	$g(\overset{\alpha}{\nabla}_Z W,\overset{-\alpha}{\nabla}_X Y)$	$g(\overset{-\alpha}{\nabla}_Z W,\overset{-\alpha}{\nabla}_X Y)$
6.	$g(\overset{\alpha}{\nabla}_Y W,\overset{\alpha}{\nabla}_X Z)$	$g(\overset{-\alpha}{\nabla}_Y W,\overset{-\alpha}{\nabla}_X Z)$
7.	$g(\overset{\alpha}{\nabla}_Y W,\overset{-\alpha}{\nabla}_X Z)$	$g(\overset{-\alpha}{\nabla}_Y W,\overset{\alpha}{\nabla}_X Z)$
8.	$g(\overset{\alpha}{\nabla}_Y Z,\overset{\alpha}{\nabla}_X W)$	$g(\overset{-\alpha}{\nabla}_Y Z,\overset{-\alpha}{\nabla}_X W)$
9.	$g(\overset{\alpha}{\nabla}_Y Z,\overset{-\alpha}{\nabla}_X W)$	$g(\overset{-\alpha}{\nabla}_Y Z,\overset{\alpha}{\nabla}_X W)$

Lines 4 and 5 disappear by torsion freeness and alternation. Lines 2 + 9 add up to zero. Lines 3 + 7 disappear by alternation. Lines 6 + 8 also. What is left over are only terms from line 1 whereby

$$2\alpha F(X,Y,Z,W) - 2\alpha F(Y,X,Z,W)$$

$$= 2\overset{-\alpha}{R}(X,Y,Z,W) - 2\overset{\alpha}{R}(X,Y,Z,W)$$

and the result and (3.7) follows.

A statistical manifold satisfying this kind of symmetry shall be called conjugate symmetric. We get then immediately

3.12 Corollary The following is sufficient for a statistical manifold to be conjugate symmetric

$$\exists\, \alpha_0 \;\text{ such that }\; \overset{\alpha_0}{R} \equiv 0,$$

i.e. that the manifold is α_0-flat.

As shown e.g. in Amari (1985), exponential families are ± 1 - flat and therefore always conjugate symmetric.

In a conjugate symmetric space, the curvature tensor thus satisfies all the identities of the Riemannian curvature tensor, i.e. also

$$\left.\begin{aligned}\overset{\alpha}{R}(X,Y,Z,W) &= -\overset{\alpha}{R}(X,Y,W,Z)\\ \overset{\alpha}{R}(X,Y,Z,W) &= \overset{\alpha}{R}(Z,W,X,Y)\end{aligned}\right\} \tag{3.8}$$

This implies as mentioned earlier that the sectional curvature determines the curvature tensor.

We shall later see examples of statistical manifolds actually generated by a statistical model that are <u>not</u> conjugate symmetric.

It also follows that the condition

$$\exists\ \alpha_0 \neq 0 \ \underline{\text{such that}}\ \overset{\alpha_0}{R} = \overset{-\alpha_0}{R} \tag{3.9}$$

<u>is sufficient for conjugate symmetry</u>.

Amari (1985) investigated the case when the statistical manifold was α_0 (and thus $-\alpha_0$) flat in detail, showing the existence of local conjugate coordinates (θ^i) and (η_i) such that $\overset{\alpha_0}{\Gamma}_{ijk} = 0$ in the θ-coordinates and its conjugate $\overset{-\alpha_0}{\Gamma}_{ijk} = 0$ in the η-coordinates.

Further that <u>potential functions</u> $\psi(\theta)$ and $\phi(\eta)$ then exist such that

$$g_{ij}(\theta) = E_i E_j \psi(\theta) \quad g^{ij}(\eta) = E_i E_j(\phi(\eta))$$

and the θ- and η-coordinates then are related by the <u>Legendre</u> transform:

$$\theta^i = E_i(\phi(\eta)) \quad \eta_i = E_i(\psi(\theta))$$

$$\psi(\theta) + \phi(\eta) - \theta^i\eta_i = 0.$$

In a sense α_0-flat families are geometrically equivalent to exponential families.

If $\underline{N}$ is a regular submanifold of $(\underline{M},g,D)$, the tensors g and D are inherited in a simple way (by restriction). The α-connections are inherited by orthogonal projections on to the space of tangent vectors to $\underline{N}$, i.e. by the equation

$$g(\overset{\alpha}{\tilde{\nabla}}_X Y,Z) = g(\overset{\alpha}{\nabla}_X Y,Z) \text{ for } X,Y,Z \in \underline{X}(\underline{N}). \tag{3.10}$$

It follows from (3.10) that <u>the α-connections induced by the restriction of g and D to $\underline{X}(\underline{N})$ are equal to those obtained by projection</u> (3.10). This consistency condition is rather important although it is so easily verified.

A submanifold is <u>totally α-geodesic</u> (or just α-geodesic) if

$$\overset{\alpha}{\nabla}_X Y \in \underline{X}(\underline{N}) \text{ for all } X,Y \in \underline{X}(\underline{N}).$$

If the submanifold is α-geodesic for all α we say that it is geodesic. We then note the following

3.12 Proposition A regular submanifold $\underline{N}$ is geodesic if and only if there exist $\alpha_1 \neq \alpha_2$ such that $\underline{N}$ is α_1-geodesic and α_2-geodesic.

Proof: Let $X,Y \in \underline{X}(\underline{N})$ and $Z_p \in T_p(\underline{N})^{\perp}$ $p \in \underline{N}$.

Then $\underline{N}$ is α_i-geodesic, i=1, 2 iff

$$g(\overset{\alpha_1}{\nabla}_X Y,Z)_p = g(\overset{\alpha_2}{\nabla}_X Y,Z) = 0$$

for all such X,Y,Z. But since

$$g(\overset{\alpha}{\nabla}_X Y,Z) = g(\bar{\nabla}_X Y,Z) - \frac{\alpha}{2} D(X,Y,Z)$$

this happens if and only if $D(X,Y,Z)_p = 0$ for all such X,Y,Z, whereby $\underline{N}$ is geodesic iff it is α_i-geodesic, i=1,2.

In statistical language, geodesic (α-geodesic) submanifolds will be called geodesic (α-geodesic) hypotheses. A central issue is the problem of existence and construction of α-geodesic and geodesic foliations of a statistical manifold.

A foliation of $(\underline{M},g,D)$ is a partitioning

$$\underline{M} = \bigcup_{\xi \in \Xi} \underline{N}_{\xi}$$

of the manifold into submanifolds $\underline{N}_{\xi}$ of fixed dimension n(<m). $\underline{N}_{\xi}$ are called the leaves of the foliation.

The foliation is said to be geodesic (or α-geodesic) if the leaves are all geodesic (or α-geodesic).

It follows from Proposition 3.12 that geodesic foliations of full exponential families (and of α-flat families) are those that are affine both in the canonical and in the mean value parameters, in other words precisely the affine dual foliations studied by Barndorff-Nielsen and Blaesild (1983). In the paper cited it is shown that existence of such foliations are intimately tied to basic statistical properties related to independence of estimates and ancillarity. Proposition 3.12 shows that the concept itself is entirely geometric in its nature.

It seems reasonable to believe that the existence (locally as well as globally) of foliations of statistical models could be quite informative. It plays at least a role when discussing procedures to obtain estimates and ancillary statistics on a geometric basis.

Let $\underline{N}$ be a submanifold of $\underline{M}$ and suppose that $\hat{p}\varepsilon\underline{M}$ is an estimate of p, obtained assuming the model $\underline{M}$. Amari (1982, 1985) discusses the α-estimate of p assuming $\underline{N}$ as follows.

To each point p of $\underline{N}$ we associate an ancillary manifold $A_\alpha(p)$

$$A_\alpha(p) = \overset{\alpha}{\exp}\,(T_p(\underline{N})^\perp)$$

where $\overset{\alpha}{\exp}$ is the exponential map associated with the α-connection and $T_p(\underline{N})^\perp$ is the set of tangent vectors orthogonal to $\underline{N}$ at p. In general the exponential map might not be defined on all $T_p(\underline{N})^\perp$, but then let it be maximally defined.

$\hat{\hat{p}}$ is then an α-estimate of p, assuming $\underline{N}$ if

$$\hat{p} \in A_\alpha(\hat{\hat{p}}).$$

Amari (1985) shows that if $\underline{M}$ is α-flat and $\underline{N}$ is $-\alpha$-geodesic, then the α-estimate is uniquely determined and it minimizes a certain divergence function.

This suggest that it might be worthwhile studying procedures that use the $-\alpha$-estimate for α-geodesic hypotheses $\underline{N}$, and call such a procedure <u>geometric estimation</u>. In general it seems that one should study the decomposition of the tangent spaces at $p\varepsilon\underline{N}$ as

$$T_p(\underline{M}) = T_p(\underline{N}) \oplus T_p(\underline{N})^\perp$$

and especially the maps of these spaces onto itself induced by α-parallel transport of vectors in $T_p(\underline{N})$, $-\alpha$ parallel transport of vectors in the complement, both along closed curves in $\underline{N}$.

It should also be possible to define a teststatistic in geometric terms by a suitable lifting of the manifold $\underline{N}$, see also Amari (1985). Things are especially simple in the case where $\underline{M}$ has dimension 2 and $\underline{N}$ has dimension 1 and we shall try to play a bit with the above loose ideas in some of the examples to come.

5. THE UNIVARIATE GAUSSIAN MANIFOLD

Let us consider the family of normal distributions $N(\mu,\sigma^2)$, i.e. the family with densities

$$f(x;\mu,\sigma) = \sqrt{2\pi\sigma^2}^{\,-1} \exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\}, \mu\in\mathbb{R}, \sigma>0$$

w.r.t. Lebesgue measure on IR. This manifold has been studied as a Riemannian manifold by Atkinson and Mitchell (1981), Skovgaard (1984) and, as a statistical manifold in some detail by Amari (1982, 1985). Working in the (μ,σ) parametrization we obtain the following expressions for the metric, the α-connections and the D-tensor (skewness) expressed as T_{ijk} (cf. Amari, 1985).

$$g = \frac{1}{\sigma^2}\begin{pmatrix}1 & 0\\ 0 & 2\end{pmatrix}$$

$$\overset{\alpha}{\Gamma}_{111} = \overset{\alpha}{\Gamma}_{122} = \overset{\alpha}{\Gamma}_{212} = \overset{\alpha}{\Gamma}_{221} = 0$$

$$\overset{\alpha}{\Gamma}{}^{1}_{11} = \overset{\alpha}{\Gamma}{}^{2}_{12} = \overset{\alpha}{\Gamma}{}^{2}_{21} = \overset{\alpha}{\Gamma}{}^{1}_{22} = 0$$

$$\overset{\alpha}{\Gamma}_{112} = (1-\alpha)/\sigma^3 \qquad \overset{\alpha}{\Gamma}{}^{2}_{11} = (1-\alpha)/(2\sigma)$$

$$\overset{\alpha}{\Gamma}_{121} = \overset{\alpha}{\Gamma}_{211} = -(1+\alpha)/\sigma^3 \qquad \overset{\alpha}{\Gamma}{}^{1}_{12} = \overset{\alpha}{\Gamma}{}^{1}_{21} = -(1+\alpha)/\sigma$$

$$\overset{\alpha}{\Gamma}_{222} = -2(1+2\alpha)/\sigma^3 \qquad \overset{\alpha}{\Gamma}{}^{2}_{22} = -(1+2\alpha)/\sigma$$

$$T_{111} = T_{122} = T_{212} = T_{221} = 0$$

$$T_{112} = T_{121} = T_{211} = 2/\sigma^3 \qquad T_{222} = 8/\sigma^3$$

The α-curvature tensor is given by

$$\overset{\alpha}{R}_{1212} = (1-\alpha^2)/\sigma^4,$$

so the manifold is conjugate symmetric, and the scalar (sectional) curvature by

$$K^{\alpha}(\sigma_{12}) = \overset{\alpha}{R}_{1221}/(g_{11}g_{22}) = -(1-\alpha^2)/2$$

For $\alpha = 0$ (the Riemannian case) we have $K(\sigma_{12}) = -1/2$ and the manifold is the space of constant negative curvature (<u>Poincaré's halfplane</u> or <u>hyperbolic space</u>). Note that it also has constant α-curvature for all α although nobody knows what that implies, since such objects have never been studied previously.

To find all α-geodesic submanifolds of dimension 1 we proceed as follows. Let (e,E) denote the tangent vector fields

$$e = \frac{\partial}{\partial\mu} \qquad E = \frac{\partial}{\partial\sigma}\,.$$

If we have $\mu = \mu_0$ constant on $\underline{N}$, $\underline{X}(\underline{N})$ is spanned by E. Since $\overset{\alpha}{\Gamma}{}^{1}_{22} = 0$ we have

$$\overset{\alpha}{\nabla}_E E = f_{\alpha} E \text{ for all } \alpha,$$

and thus that the submanifolds

$$\underline{N}_{\mu_0} = \{(\mu,\sigma)\,|\,\mu=\mu_0\},\mu_0\in\mathbb{R}$$

are <u>geodesic</u> submanifolds and the family

$$(\underline{N}_{\mu},\mu\in\mathbb{R}) \tag{4.1}$$

constitutes a <u>geodesic foliation</u> of the Gaussian manifold.

If μ is non-constant on $\underline{N}$, we must be able to parametrize $\underline{N}$ locally as

$$(t,\sigma(t)),\ t\in I \subseteq \mathbb{R}.$$

The tangent space to $\underline{N}$ is then spanned by

$$N = e + \dot{\sigma}\,E$$

where we have let $\dot{\sigma}(t) = \frac{\partial}{\partial t}\sigma(t)$ and extended σ to a function defined on all of the manifold by $\sigma(x,y) := \sigma(x)$.

$$\overset{\alpha}{\nabla}_N N = \overset{\alpha}{\nabla}_{e+\dot{\sigma}E}(e+\dot{\sigma}E) = \overset{\alpha}{\nabla}_e e + 2\dot{\sigma}\overset{\alpha}{\nabla}_e E + \dot{\sigma}^2\overset{\alpha}{\nabla}_E E + \ddot{\sigma}E \tag{4.2}$$

where we have used torsion freeness and the fact that $e(\dot{\sigma}) = \ddot{\sigma}$, $E(\dot{\sigma}) = 0$. Using now the expressions for $\overset{\alpha}{\Gamma}{}^{k}_{ij}$, we get

$$\overset{\alpha}{\nabla}_N N = \frac{-(1+\alpha)}{\sigma}\,2\dot{\sigma}e + \left[\frac{1-\alpha}{2\sigma} + \ddot{\sigma} - \frac{1+2\alpha}{\sigma}\,\dot{\sigma}^2\right]E.$$

If this again has to be in the direction of N, we must have

$$-\frac{1+\alpha}{\sigma}2\dot{\sigma}^2 = \frac{1-\alpha}{2\sigma} + \ddot{\sigma}\frac{1+2\alpha}{\sigma}\dot{\sigma}^2$$

which by multiplication with 2σ reduces to the differential equation

$$2\ddot{\sigma}\sigma + 2\dot{\sigma}^2 = (\alpha-1)$$

This is most conveniently solved by letting $u = \sigma^2$, whereby $\ddot{u} = 2\ddot{\sigma}\sigma + 2\dot{\sigma}^2$ and the equation becomes as simple as

$$\ddot{u} = \alpha-1 \leftrightarrow u(t) = \tfrac{1}{2}(\alpha-1)t^2 + Bt + C, \tag{4.3}$$

such that the α-geodesic submanifolds are either straight lines ($\alpha = 1$) or parabolas in the (μ,σ^2)-parametrisation.

The special case $\alpha = 1$, $B = 0$ corresponds to the manifolds

$$\underline{\tilde{N}}_{\sigma_0} = \{(\mu,\sigma)|\sigma=\sigma_0\},\ \sigma_0\varepsilon IR_+$$

that give a <u>1-geodesic foliation</u>.

Another special case is the submanifolds of <u>constant variation coefficient</u>

$$\underline{V}_\gamma = \{(\mu,\sigma)|\sigma=\gamma\mu\}, \gamma\varepsilon IR_+$$

that we now see <u>are α-geodesic if and only if $\alpha = 1+2\gamma^2$</u> by inserting into (4.3). $\underline{V}_\gamma$ are now connected submanifolds but is composed by two non-connected submanifolds $\underline{V}_\gamma^+$ and $\underline{V}_\gamma^-$

$$\underline{V}_\gamma^+ = \{(\mu,\sigma)|\mu>0\}\cap\underline{V}_\gamma,\ \underline{V}_\gamma^- = \{(\mu,\sigma)|\mu>0\}\cap\underline{V}_\gamma.$$

The $(\underline{V}_\gamma^+,\underline{V}_\gamma^-)$ manifolds do <u>not</u> represent α-geodesic foliations since they are not α-geodesic for the <u>same</u> value of α. For $\alpha = 0$ we see that the geodesic submanifolds are parabola's in (μ,σ^2) with coefficient $-\tfrac{1}{2}$ to μ^2, a result also obtained by Atkinson and Mitchell (1981) and Skovgaard (1984).

Consider now the hypothesis $(\mu,\sigma)\ \varepsilon\underline{V}_\gamma$, i.e. that of constant variation coefficient. We shall illustrate the idea of geodesic estimation in this example as described at the end of section 3.

$\underline{V}_\gamma$ is $\alpha=1+2\gamma^2$ geodesic. The ancillary manifolds to be considered are then $-\alpha$-geodesic manifolds orthogonal to $\underline{V}_\gamma$.

An arbitrary $-\alpha$-submanifold is the "parabola"

$$\sigma = (-(1+\gamma^2)\mu^2+B\mu+C)^{\frac{1}{2}}$$

which follows from (4.3) with $\alpha = -(1+2\gamma^2)$. Its tangent vector is equal to

$$e+\dot{\sigma}E = \frac{1}{2\sigma}[-2(1+\gamma^2)\mu+B]E+e.$$

The tangent vector of the hypothesis is

$$e+\gamma E.$$

They are at right angles at $(\mu_0,\gamma\mu_0)$ if and only if

$$1+\frac{1}{\sigma}[-2(1+\gamma^2)\mu_0+B]=0 \leftrightarrow B=(1+2\gamma^2)\mu_0.$$

The ancillary manifold intersects at $(\mu_0,\gamma\mu_0)$ if and only if

$$-(1+\gamma^2)\mu_0^2+(1+2\gamma^2)\mu_0^2+C=\gamma^2\mu_0^2 \leftrightarrow C=0.$$

The $-(1+2\gamma^2)$-geodesic ancillary manifolds are thus given as

$$\underline{W}_\mu = \{(t,\sigma_\mu(t))\,|\,t\varepsilon I_\mu\},\ \mu\varepsilon IR\setminus\{0\}$$

$$(\underline{W}_0 = \{(0,\sigma)\,|\,\sigma\varepsilon IR_+\})$$

where $\sigma_\mu{}^2(t) = -(1+\gamma^2)t^2 + (1+2\gamma^2)\mu t$ and

$$I_\mu = \begin{cases}]0,\ \frac{1+2\gamma^2}{1+\gamma^2}\mu[& \text{if } \mu>0 \\ ]\ \frac{1+2\gamma^2}{1+\gamma^2}\mu,0\ [& \text{if } \mu<0. \end{cases}$$

The manifolds $\underline{W}_\mu$, $\mu\varepsilon IR$ <u>actually constitute a</u> $-(1+2\gamma^2)$ -<u>foliation of the Gaussian manifold</u>. To see this, let $(\bar{x},s)$ be an arbitrary point in $\underline{M}$. If we try to solve the equation

$$(\bar{x},s^2) = (t,-(1+\gamma^2)t^2+(1+2\gamma^2)\mu t)$$

we obtain exactly one solution $\tilde{\mu}$ for $x\neq 0$, given as

$$\tilde{\mu} = \frac{(1+\gamma^2)\bar{x}^2+s^2}{(1+2\gamma^2)\bar{x}} = \frac{(1+\gamma^2)\bar{x}+\gamma^2\cdot\frac{s^2}{\gamma^2\bar{x}}}{(1+2\gamma^2)}, \tag{4.4}$$

i.e. a linear combination of $\bar{x}$ and $\frac{s^2}{\gamma^2\bar{x}}$.

$\tilde{\mu}$, as determined by (4.4) is <u>the geometric estimate</u> of μ, when $\bar{x}$ and s denote the empirical mean and standard deviation of a sample $x_1,\ldots,x_n$. It is by construction (see Amari (1982)) consistent and first-order efficient.

A picture of the situation is given below in three different parametrizations: (μ,σ), (μ,σ^2), and (μ,σ^{-2}):

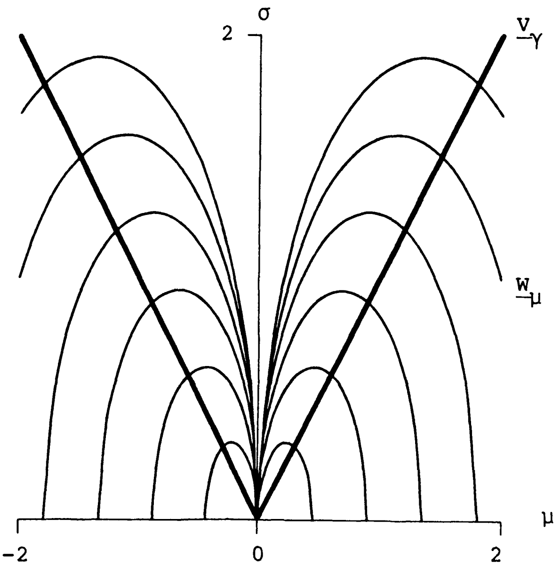

Fig. 1: Geometric estimation with constant coefficient of variation, (μ,σ)-param.

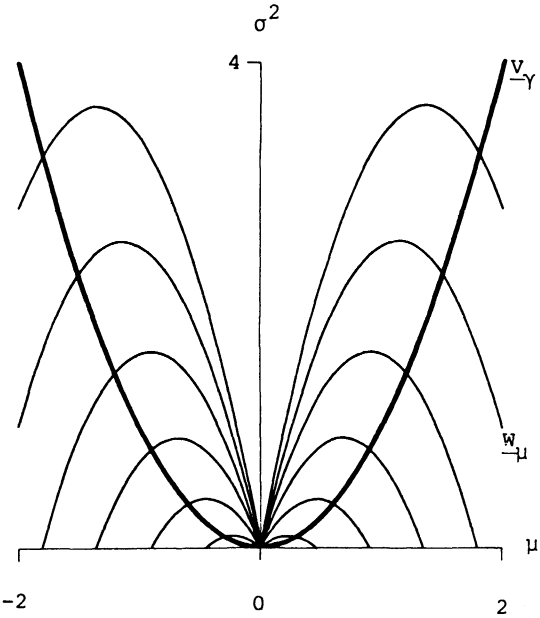

Fig. 2: Geometric estimation, (μ,σ^2)-param.

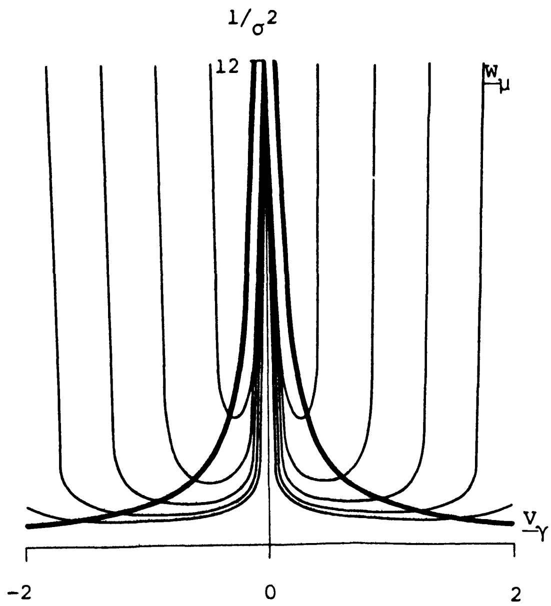

Fig. 3: Geometric estimation, $(\mu, \frac{1}{\sigma^2})$ param.

To obtain a geometric ancillary and test-statistic we proceed as follows:

We take a system of vectors on the hypotheses whose directions are $-(1+2\gamma^2)$ -parallel and whose lengths are equal to one. Further they are to be orthogonal to the hypothesis (and thus tangent to the estimation manifolds). The directions should thus be given as

$$v = (v_1, v_2) = -e + \frac{1}{2\gamma} E.$$

To obtain unit length, we get $||v|| = \frac{1}{\sigma}\sqrt{\frac{2\gamma^2+1}{2\gamma^2}} = \sqrt{\frac{2\gamma^2+1}{2\gamma^4\mu^2}}$

when $\sigma=\gamma\mu$, and our orthogonal field is thus

$$\underline{v}(\mu) = [v_1(\mu), v_2(\mu)] = a[-\mu, \frac{\mu}{2\gamma}]$$

where $a = (2\gamma^4/(2\gamma^2+1))^{\frac{1}{2}}$. To find the exponential map

$$\overset{-(1+2\gamma^2)}{\exp}\ \{t\underline{v}(\mu)\} = (f(t,\mu), \sigma(t,\mu))$$

we shall solve the equations

$$\sigma^2(t,\mu) = -(1+\gamma^2)f(t,\mu)^2 + (1+2\gamma^2)f(t,\mu)\mu \tag{4.5}$$

$$\frac{d\, f\,(0,\mu)}{d\, t} = -a\mu \text{ and } f(0,\mu) = \mu$$

$$\ddot{f}=2\dot{f}\,\frac{\dot{\sigma}}{\sigma}\,(1+\alpha) \leftrightarrow \ddot{f} = -2/\gamma^2\dot{f}\,\frac{\dot{y}}{y} \tag{4.6}$$

since only the <u>speed</u> of the geodesic has to be determined. (4.6) is easily seen to be equivalent to

$$\dot{f} = K\sigma^{-4\gamma^2} \text{ for some } K\neq 0. \tag{4.7}$$

Inserting (4.5) into this we obtain

$$\dot{f} = K(-(1+\gamma^2)f^2 + (1+2\gamma^2)\mu f)^{-2\gamma^2}$$

and separation of variables yield

$$\int_0^f[-(1+\gamma^2)u^2 + (1+2\gamma^2)\mu u]^{2\gamma^2}du = Kt+C$$

Substituting $v=u/\mu$ we get

$$\mu^{4\gamma^2+1}G(\frac{f(t,\mu)}{\mu}) = Kt+C \tag{4.8}$$

where $G(x) = \int_0^x [-(1+\gamma^2)v^2+(1+2\gamma^2)v]^{2\gamma^2}dv.$

Using the initial condition $f(0,\mu)=\mu$ we get

$$C = \mu^{4\gamma^2+1}\, G(1)$$

and the condition $\dot{f}(0,\mu) = -a\mu$ yields together with (4.7)

$$K = \sigma^{4\gamma^2}(0,\mu)(-a\mu) = -a\gamma^{4\gamma^2}\mu^{4\gamma^2+1},$$

whereby

$$\mu^{4\gamma^2+1}\, G(\frac{f(t,\mu)}{\mu}) = -a\gamma^{4\gamma^2}\mu^{4\gamma^2+1}t + \mu^{4\gamma^2+1}G(1),$$

and dividing by $\mu^{4\gamma^2+1}$ yields thus

$$G(\frac{f(t,\mu)}{\mu}) = -a\gamma^{4\gamma^2}t + G(1)$$

and therefore $f(t,\mu) = \mu h(t)$ where

$$h(t) = G^{-1}(-a\gamma^{4\gamma^2}t + G(1)).$$

Inserting this into (4.5) yields

$$\sigma(t,\mu) = \mu\,\sqrt{-(1+\gamma^2)h(t)^2+(1+2\gamma^2)h(t)}$$

which is <u>linear</u> in μ. If we now interpret points of same "distance" from the hypothesis as those where t is fixed and only μ varying, we see that $s/\bar{x}$ is in one-to-one correspondence with t. We shall therefore say that $s/\bar{x}$ is the geometric ancillary and this it also is the geometric test statistic for the hypothesis $\sigma=\gamma\mu$.

It is of course interesting, although not surprising, that this test statistic (ancillary) is obtained solely by geometric arguments but still equal to the "natural" when considering the transformation structure of the model.

6. THE INVERSE GAUSSIAN MANIFOLD

Consider the family of inverse Gaussian densities

$$f(x;\chi,\psi) = \sqrt{\frac{\chi}{2\pi}}e^{\sqrt{\chi\psi} - \frac{1}{2}(\chi x^{-1} + \psi x)}x^{-3/2}, \quad \chi,\psi>0$$

w.r.t. Lebesgue measure on IR_+. We choose to study this manifold in the parametrization (η,θ), where

$$\eta = \chi^{-1} \qquad \theta = \sqrt{\frac{\psi}{\chi}}, \text{ i.e.}$$

$$f(x;\eta,\theta) = h(x)\eta^{-\frac{1}{2}}e^{\frac{\theta}{\eta} - \frac{1}{2\eta}(x^{-1}+\theta^2 x)}.$$

The metric tensor and the skewness tensor can now be calculated either by using their definition directly or by calculating these in the (χ,ψ) coordinates and using transformation rules of tensors. We get

$$g = \begin{pmatrix} \frac{1}{2\eta^2} & 0 \\ 0 & \frac{1}{\theta\eta} \end{pmatrix}$$

and $T_{112}=0$, $T_{111}=\eta^{-3}$, $T_{122}=\theta^{-1}\eta^{-2}$, $T_{222}=-\frac{3}{\theta^2\eta}$.

The Riemannian connection is now determined by

$$\bar{\Gamma}_{ijk} = \tfrac{1}{2}[\partial_i g_{jk}+\partial_j g_{ik}-\partial_k g_{ij}], \text{ such that}$$

$$\overset{\alpha}{\Gamma}_{111} = -(1+\alpha)/(2\eta^3), \ \overset{\alpha}{\Gamma}_{112} = \overset{\alpha}{\Gamma}_{211} = \overset{\alpha}{\Gamma}_{121} = 0$$

$$\overset{\alpha}{\Gamma}_{221} = (1-\alpha)/(2\theta\eta^2), \ \overset{\alpha}{\Gamma}_{122} = \overset{\alpha}{\Gamma}_{212} = -(1+\alpha)/(2\theta\eta^2)$$

$$\overset{\alpha}{\Gamma}_{222} = (3\alpha-1)/(2\theta^2\eta)$$

Multiplying with the inverse metric we get

$$\overset{\alpha}{\Gamma}{}^1_{11} = -(1+\alpha)/\eta \qquad \overset{\alpha}{\Gamma}{}^2_{11} = \overset{\alpha}{\Gamma}{}^1_{12} = \overset{\alpha}{\Gamma}{}^1_{21} = 0$$

$$\overset{\alpha}{\Gamma}{}^2_{12} = \overset{\alpha}{\Gamma}{}^2_{21} = -(1+\alpha)/(2\eta) \qquad \overset{\alpha}{\Gamma}{}^1_{22} = (1-\alpha)/\theta$$

$$\overset{\alpha}{\Gamma}{}^2_{22} = (3\alpha-1)/2\theta.$$

To find all geodesic submanifolds of dimension one we first notice that since $\overset{\alpha}{\Gamma}{}^2_{11} \equiv 0$, the manifolds

$$\underline{N}_{\theta_0} = \{(\eta,\theta)|\theta=\theta_0\}$$

are α-geodesic for all α, i.e. <u>geodesic</u> and they constitute a <u>geodesic foliation</u> of the inverse Gaussian manifold. Because

$$E\, X = \theta^{-1}$$

they correspond to hypotheses of constant expectation.

Consider now a submanifold of the form $(\eta(t),t)$, i.e. with tangent N given as

$$N = \dot{\eta}\, e + E, \text{ where } e = \frac{\partial}{\partial\eta},\ E = \frac{\partial}{\partial\theta}.$$

We extend η by letting $\eta(x,y) := \eta(y)$, i.e. such that $e(\dot{\eta}) = 0$, $E(\dot{\eta}) = \ddot{\eta}$. Then

$$\overset{\alpha}{\nabla}_N N = \dot{\eta}^2 \overset{\alpha}{\nabla}_e e + 2\dot{\eta}\overset{\alpha}{\nabla}_e E + \ddot{\eta} e + \overset{\alpha}{\nabla}_E E$$

$$= (\ddot{\eta} - \frac{1+\alpha}{\eta}\dot{\eta}^2 + \frac{1-\alpha}{t})e + (-\frac{1+\alpha}{\eta}\dot{\eta} + \frac{3\alpha-1}{2t})E$$

We now have $\overset{\alpha}{\nabla}_N N = hN$ iff

$$\dot{\eta}[-\frac{1+\alpha}{\eta}\dot{\eta} + \frac{3\alpha-1}{2t}] = [\ddot{\eta} - \frac{1+\alpha}{\eta}\dot{\eta}^2 + \frac{1-\alpha}{t}]$$

which reduces to the differential equation

$$\ddot{\eta} - \frac{3\alpha-1}{2t}\dot{\eta} = \frac{\alpha-1}{t}.$$

This is first solved for $\alpha = \frac{1}{3}$:

$$\ddot{\eta} = -\frac{2}{3t} \leftrightarrow \dot{\eta} = -\frac{2}{3}\log t + C \leftrightarrow$$

$$\eta(t) = -\frac{2}{3}t\log t + C_1 t + C_2.$$

For $\alpha \neq \frac{1}{3}$ we get by letting $u = \dot{\eta} t^{\frac{1-3\alpha}{2}}$ that u satisfies the differential equation

$$\dot{u} = (\alpha-1)t^{-\frac{1+3\alpha}{2}} \leftrightarrow u(t) = \frac{2(\alpha-1)}{1-3\alpha} t^{\frac{1-3\alpha}{2}} + C_1$$

Whereby

$$\eta(t) = \frac{2(\alpha-1)}{1-3\alpha} t + Bt^{\frac{3\alpha+1}{2}} + C, \ \alpha \neq \frac{1}{3}$$

For $\alpha=1$ (the exponential connections) we get the parabolas:

$$\eta(t) = Bt^2 + C$$

and for $\alpha=-1$ (the mixture connection) we get the curves:

$$\eta(t) = -t + B/t + C.$$

In the Riemannian case ($\alpha=0$) we get

$$\eta(t) = -2t + B\sqrt{t} + C$$

that are parabolas in $(\sqrt{\theta},\eta)$.

The <u>curvature tensor</u> is given by

$$\overset{\alpha}{R}_{1212} = (E_1(\overset{\alpha}{\Gamma}{}^s_{21}) - E_2(\overset{\alpha}{\Gamma}{}^s_{11}))g_{2s} + (\overset{\alpha}{\Gamma}_{1r2}\overset{\alpha}{\Gamma}{}^r_{21} - \overset{\alpha}{\Gamma}_{2r2}\overset{\alpha}{\Gamma}{}^r_{11}) = \frac{1-\alpha^2}{4\theta\eta^3}$$

The manifold is thus <u>conjugate symmetric</u> (we already know, since it is an exponential family) and the sectional curvature is

$$K^{\alpha}(\sigma_{12}) = -\overset{\alpha}{R}_{1212}/(g_{11}g_{22}) = -(1-\alpha^2)/2.$$

Note that the Riemannian curvature ($\alpha=0$) is again constant equal to $-\frac{1}{2}$, as in the Gaussian case. In fact <u>the α-curvature is exactly as in the Gaussian case</u>. We can map the inverse Gaussian manifold to the Gaussian by letting

$$\mu = \sqrt{2\theta} \qquad \sigma^2 = \eta/2$$

and this map is a Riemannian isometry. However, <u>it does not preserve the skewness tensor</u> and thus the Gaussian and inverse Gaussian manifolds do not seem to be isomorphic as <u>statistical manifolds</u>, although they are as Riemannian manifolds.

Corresponding to the hypothesis of constant coefficient of variation, we shall investigate the submanifold corresponding to the exponential transformation model $\sqrt{\chi\psi} = \frac{1}{\gamma}$, γ fixed, i.e.

$$h(x)\sqrt{\frac{\sigma}{\gamma}}\, e^{\frac{1}{\gamma} - \frac{1}{2\gamma}(\sigma x^{-1} + \sigma^{-1}x)} \qquad \sigma>0$$

which in the (η,θ)-parametrization is a straight line through the origin (as const. coeff. of var.)

$$\{\eta = \gamma\theta\} = \underline{V}_{\gamma}$$

This submanifold is α-geodesic if and only if

$$\gamma = \frac{2(\alpha-1)}{1-3\alpha} \leftrightarrow \alpha = \frac{2+\gamma}{2+3\gamma}\,.$$

The tangent space to $\underline{V}_{\gamma}$ is spanned by $\gamma\,e + E$, and the orthogonal $-\alpha$-geodesic submanifolds are given by solving the equations

$$\frac{2(\alpha-1)}{1-3\alpha}\,\tilde{\theta} = -\frac{2(1+\alpha)}{1+3\alpha}\,\tilde{\theta} + B\,\tilde{\theta}^{\frac{1-3\alpha}{2}} + C \tag{5.1}$$

to get the intersecting point and orthogonality at $(\frac{2(\alpha-1)}{1-3\alpha}\,\tilde{\theta},\tilde{\theta})$ gives

$$B = -\frac{8\alpha\,\tilde{\theta}^{\frac{3\alpha+1}{2}}}{1-9\alpha^2}\,.$$

Combining this with (5.1) we get C=0, i.e. the estimation manifolds are given as

$$\eta_{\tilde{\theta}}(t) = -\frac{2(1+\alpha)}{1+3\alpha}\,t - \frac{8\alpha\,\tilde{\theta}^{\frac{3\alpha+1}{2}}}{1-9\alpha^2}\,t^{\frac{1-3\alpha}{2}}$$

The manifolds $\underline{W}_{\tilde{\theta}}$, $\tilde{\theta}>0$ again constitute a $-\alpha$-foliation of the inverse Gaussian manifold as is seen by solving the equations

$$(\eta_0,\theta_0) = (\eta_{\tilde{\theta}}(t),t)$$

which gives $t=\theta_0$, and

$$\tilde{\theta} = \left[\frac{(3\alpha-1)(1+\alpha)}{4\alpha}\,\theta_0^{\frac{3\alpha+1}{2}} + \frac{9\alpha^2-1}{8\alpha}\,\eta_0\theta_0^{\frac{3\alpha-1}{2}}\right]^{\frac{2}{3\alpha+1}}$$

$$= \theta_0\left[\frac{(3\alpha-1)(\alpha+1)}{4\alpha} + \frac{9\alpha^2-1}{8\alpha}\,\frac{\eta_0}{\theta_0}\right]^{\frac{2}{3\alpha+1}}$$

This again determines a geometric estimate $\tilde{\theta}$ of θ from a sample $x_1,\dots,x_n$ from the inverse Gaussian distribution, and this is obtained by letting

$$\eta_0 = 1/\bar{x} \qquad \theta_0 = \frac{1}{n}\sum x_i^{-1} - 1/\bar{x} ,$$

and inserting $\alpha = (2+\gamma)/(2+3\gamma)$ into the expression given above.

7. THE GAMMA MANIFOLD

Consider the family of gamma densities

$$f(x;\mu,\beta) = (\beta/\mu)^{\beta}\, x^{\beta-1}/\Gamma(\beta)\, \exp\{-\frac{x\beta}{\mu}\} \qquad \mu>0,\ \beta>0$$

w.r.t. Lebesgue measure on IR_+. The metric tensor is obtained by direct calculation in the (μ,β)-parametrization as

$$g = \begin{pmatrix} \frac{\beta}{\mu^2} & 0 \\ 0 & \phi(\beta) \end{pmatrix}$$

where $\phi(\beta) = D^2 \log \Gamma(\beta) - 1/\beta$.

The Riemannian connection is now obtained by

$$\bar{\Gamma}_{ijk} = \tfrac{1}{2}[\partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij}] \text{ to be}$$

$$\bar{\Gamma}_{111} = -\beta/\mu^3;\ \bar{\Gamma}_{112} = -\ 1/(2\mu^2);\ \bar{\Gamma}_{121} = \bar{\Gamma}_{211} = 1/(2\mu^2)$$

$$\bar{\Gamma}_{222} = \tfrac{1}{2}\phi'(\beta),\ \bar{\Gamma}_{221} = \bar{\Gamma}_{122} = \bar{\Gamma}_{212} = 0.$$

Similarly we calculate $\overset{1}{\Gamma}_{ijk}$ by the formula

$$\overset{1}{\Gamma}_{ijk} = \mathbb{E}(E_iE_j(1)E_k(1)) \text{ to be}$$

$$\overset{1}{\Gamma}_{111} = -2\beta/\mu^3 \qquad \overset{1}{\Gamma}_{121} = 1/\mu^2$$

$$\overset{1}{\Gamma}_{122} = \overset{1}{\Gamma}_{112} = \overset{1}{\Gamma}_{212} = \overset{1}{\Gamma}_{222} = \overset{1}{\Gamma}_{221} = 0$$

and the skewness tensor $T_{ijk} = 2(\bar{\Gamma}_{ijk} - \overset{1}{\Gamma}_{ijk})$

$$T_{111} = 2\beta/\mu^3 \qquad T_{112} = T_{121} = T_{211} = -1/\mu^2 \qquad T_{222} = \phi'(\beta)$$

$$T_{221} = T_{122} = T_{212} = 0,$$

whereby the α-connections are determined to be

$$\overset{\alpha}{\Gamma}_{111} = \frac{(1+\alpha)\beta}{\mu^3} \qquad \overset{\alpha}{\Gamma}_{112} = \frac{\alpha-1}{2\mu^2}$$

$$\overset{\alpha}{\Gamma}_{121} = \overset{\alpha}{\Gamma}_{211} = \frac{1+\alpha}{2\mu^2} \qquad \overset{\alpha}{\Gamma}_{222} = \frac{1-\alpha}{2}\phi'(\beta)$$

$$\overset{\alpha}{\Gamma}_{122} = \overset{\alpha}{\Gamma}_{212} = \overset{\alpha}{\Gamma}_{221} = 0.$$

Multiplying by the inverse metric we get

$$\overset{\alpha}{\Gamma}{}^1_{11} = -\frac{1+\alpha}{\mu} \qquad \overset{\alpha}{\Gamma}{}^2_{11} = \frac{\alpha-1}{2\mu^2\phi(\beta)}$$

$$\overset{\alpha}{\Gamma}{}^1_{12} = \overset{\alpha}{\Gamma}{}^1_{21} = \frac{1+\alpha}{2\beta} \qquad \overset{\alpha}{\Gamma}{}^2_{22} = \frac{1-\alpha}{2}\frac{\phi'(\beta)}{\phi(\beta)}$$

and all other symbols equal to zero.

The <u>curvature</u> is by direct calculation found to be

$$\overset{\alpha}{R}_{1212} = \frac{(\alpha^2-1)[\phi(\beta)+\beta\phi'(\beta)]}{4\mu^2\beta\phi(\beta)}$$

The space is conjugate symmetric and therefore the curvature tensor is fully determined by the <u>sectional (scalar) curvature</u> which is

$$K^{(\alpha)} = -\overset{\alpha}{R}_{1212}g^{11}g^{22} = \frac{1-\alpha^2}{4}\frac{[\phi(\beta)+\beta\phi'(\beta)]}{\beta^2\phi(\beta)}$$

Note that this is even for $\alpha=0$ different from the two previous examples in that the curvature is non-constant and truly dependent on the shape parameter β.

To find all geodesic submanifolds we proceed as follows:

If $\mu=\mu_0$ is constant on $\underline{N}$, $\underline{X}(\underline{N})$ is spanned by the tangent vector E corresponding to differentiation w.r.t. the second coordinate. Since

$$\overset{\alpha}{\nabla}_E E = \frac{1-\alpha}{2}\frac{\phi'(\beta)}{\phi(\beta)}E,$$

these submanifolds are geodesic for all values of α and constitute a <u>geodesic foliation</u> of the gamma manifold.

Considering the manifold given by $\beta=\beta_0$, its tangent space is spanned by e and since

$$\overset{\alpha}{\nabla}_e e = -\frac{1+\alpha}{\mu}e + \frac{\alpha-1}{2\mu^2\phi(\beta)}E$$

these are <u>α-geodesic if and only if α=1</u>.

In general let us consider a hypothesis (submanifold) of the type $(f(t),t)$. Its tangent vector is

$$\dot{f}\, e + E \text{ and } e(\dot{f}) = 0,\ E(\dot{f}) = \ddot{f}$$

we have

$$\overset{\alpha}{\nabla}_{\dot{f} e + E}(\dot{f}\, e + E) = \dot{f}^2\overset{\alpha}{\nabla}_e e = 2\dot{f}\overset{\alpha}{\nabla}_e E + \ddot{f}\, e + \overset{\alpha}{\nabla}_E E$$

$$= [-\dot{f}^2\,\frac{1+\alpha}{\mu} + \dot{f}\,\frac{1+\alpha}{\beta} + \ddot{f}]e + [\dot{f}^2\,\frac{\alpha-1}{2\mu^2\phi(\beta)} + \frac{1-\alpha}{2}\,\frac{\phi'(\beta)}{\phi(\beta)}]E$$

If we now let $\beta=t$ $\mu=f$ and multiply the coefficient to E by $\dot{f}$ we obtain the equation

$$-(1+\alpha)\,\frac{\dot{f}^2}{f} + (1+\alpha)\,\frac{\dot{f}}{t} + \ddot{f} = \frac{\alpha-1}{2\phi(t)}\,\frac{\dot{f}^3}{f^2} + \frac{1-\alpha}{2\phi(t)}\,\phi'(t)\dot{f}$$

which unfortunately does not seem soluble in general. For α=1 the solutions are $f(t) = t/(At+B)$.

8. TWO SPECIAL MANIFOLDS

In the present section we shall see that things are not always as simple as the previous examples suggest, but even then we seem to be able to get some understanding from geometric considerations.

First we should like to notice that when we combine two experiments independently with the same parameter space, both the Fisher information metric and the skewness tensors are additive. Let $X \sim P_\theta$ $Y \sim P_\theta$ and let A_i, B_i denote the derivative of the two log-likelihood functions

$$A_i = \frac{\partial}{\partial\theta_i} \log f\,(x;\theta) \qquad B_i = \frac{\partial}{\partial\theta_i} \log g(y;\theta).$$

Then the skewness tensor is to be calculated as

$$\begin{aligned} T_{ijk} &= E(A_i+B_i)(A_j+B_j)(A_k+B_k) \\ &= EA_jA_jA_k + EB_iB_jB_k \end{aligned}$$

since all terms containing both A's and B's vanish due to the independence and the fact that $EA_i = EB_j = 0$.

If we now let $X \sim N(\mu,\sigma^2)$, $Y \sim N(\sigma,1)$ and X and Y independent we get by adding the information and skewness tensors that in the (μ,σ)-parametrization

$$g = \frac{1}{\sigma^2}\begin{pmatrix} 1 & 0 \\ 0 & 2+\sigma^2 \end{pmatrix} \qquad g^{-1} = \sigma^2 \begin{pmatrix} 1 & 0 \\ 0 & \dfrac{1}{2+\sigma^2} \end{pmatrix}$$

and that, as in the Gaussian manifold, we have

$$T_{111} = T_{122} = T_{212} = T_{221} = 0 \qquad T_{112} = 2/\sigma^3 \qquad T_{222} = 8/\sigma^3.$$

Since derivatives of the metric are as in the Gaussian case, so are the $\overset{\alpha}{\Gamma}_{ijk}$-symbols:

$$\overset{\alpha}{\Gamma}_{111} = \overset{\alpha}{\Gamma}_{122} = \overset{\alpha}{\Gamma}_{212} = \overset{\alpha}{\Gamma}_{221} = 0$$

$$\overset{\alpha}{\Gamma}_{121} = \overset{\alpha}{\Gamma}_{211} = -(1+\alpha)/\sigma^3 \qquad \overset{\alpha}{\Gamma}_{222} = -2(1+2\alpha)/\sigma^3.$$

But the α-connections are truly different which is seen by looking at the $\overset{\alpha}{\Gamma}{}^k_{ij}$-symbols:

$$\overset{\alpha}{\Gamma}{}^2_{11} = (1-\alpha)/(2\sigma+\sigma^3) \qquad \overset{\alpha}{\Gamma}{}^1_{12} = \overset{\alpha}{\Gamma}{}^1_{21} = -(1+\alpha)/\sigma$$

$$\overset{\alpha}{\Gamma}{}^2_{22} = -2(1+2\alpha)/(2\sigma+\sigma^3)$$

and all others equal to zero. Considering now the <u>curvature tensor</u> we get

$$\overset{\alpha}{R}_{1212} = (1-\alpha)\,\frac{[2(1+\alpha)+\alpha^2(2-\alpha)]}{\sigma^4(2+\sigma^2)} = -\overset{\alpha}{R}_{2112}$$

and this is clearly <u>different</u> from $\overset{-\alpha}{R}_{1212}$ whereby this space is <u>not</u> conjugate symmetric. The sectional curvature is not determining the curvature tensor because e.g. $\overset{1}{R}_{1212} \equiv 0$ but the space is not 1-flat since

$$\overset{\alpha}{R}_{1221} = -\overset{-\alpha}{R}_{1212} = -(1+\alpha)\,\frac{[2(1-\alpha)+\sigma^2(2+\alpha)]}{\sigma^4(2+\sigma^2)} = -\overset{\alpha}{R}_{2121}$$

From standard properties of the curvature tensor we have $\overset{\alpha}{R}_{iikm} = 0$, but we obtain by direct calculation that

$$\overset{\alpha}{R}_{1211} = \overset{\alpha}{R}_{2111} = \overset{\alpha}{R}_{1222} = \overset{\alpha}{R}_{2122} = 0,$$

such that the above components are the only ones that are not vanishing.

If we try to find the geodesic submanifolds we first observe that because $\overset{\alpha}{\Gamma}{}^1_{22} = 0$ for all α, the submanifolds

$$\underline{N}_{\mu_0} = \{(\mu,\sigma)\,|\,\mu=\mu_0\}$$

are totally geodesic for all α, and thus constitute <u>a geodesic foliation</u> of the manifold. Following the remarks at the end of section 4, relating geodesic foliations to the affine dual foliations of Barndorff-Nielsen and Blaesild (1983), it is of interest to know that also in this example, the maximum likelihood estimates of σ^2 and μ are independent as expected from the foliation. We shall now proceed to find the remaining geodesic manifolds.

If we consider manifolds of the type $(t,f(t))$ with tangent vector

$e + \dot{f}\,E$ we get

$$\overset{\alpha}{\nabla}_{e+\dot{f}\,E}\,(e+\dot{f}\,E) = \overset{\alpha}{\nabla}_e e + 2\dot{f}\overset{\alpha}{\nabla}_e E + \dot{f}^2\overset{\alpha}{\nabla}_E E + \ddot{f}\,E$$

$$= -\frac{2\dot{f}}{f}(1+\alpha)e + \left(\frac{1-\alpha}{2\sigma+\sigma^3} - \frac{2(1+2\alpha)}{2\sigma+\sigma^3}\dot{f}^2\right)E$$

Multiplying the coefficient to e with $\dot{f}$ and inserting $\sigma=f$ we get the equation

$$\frac{2\dot{f}^2}{f}(1+\alpha) = \dot{f}^2\,\frac{2(1+2\alpha)}{f(2+f^2)} + \frac{\alpha-1}{f(2+f^2)} - \ddot{f}$$

Multiplying on both sides with $f(2+f^2)$ and collecting terms gives

$$2\dot{f}^2f^2(1+\alpha) + 2\ddot{f}f + \ddot{f}f^3 + 2\dot{f}^2 = \alpha-1$$

and this does not seem to have a particularly nice solution.

Note that $f(t)$ and γt is <u>not</u> a solution since then $\dot{f}=\gamma$ $\ddot{f}=0$ and we obtain the equation for α:

$$2\gamma^4t^2(1+\alpha) + 2\gamma^2 = \alpha-1$$

which can only hold when $\alpha = -1$ and then we get

$$2\gamma^2 = -2$$

which is impossible.

In this example the "constant coefficient of variation" does also not have any simple group transformational properties.

It seems then of interest to see what happens if we consider the model with $X\sim N(\mu,\sigma^2)$, $Y\sim N(\log\sigma,1)$ which is related to the example just considered but where the "constant coefficient of variation" <u>is</u> transformational. The model is also transformational itself (the affine group). By the same argument as before the skewness tensor becomes identical to that of the univariate Gaussian manifold. The metric, however, becomes

$$g = \frac{1}{\sigma^2}\begin{pmatrix}1 & 0\\ 0 & 3\end{pmatrix} \qquad g^{-1} = \sigma^2\begin{pmatrix}1 & 0\\ 0 & \frac{1}{3}\end{pmatrix}$$

whereby we calculate the Riemannian connection to be

$$\bar{\Gamma}_{112} = 1/\sigma^3 \qquad \bar{\Gamma}_{211} = \bar{\Gamma}_{121} = -1/\sigma^3$$

$$\bar{\Gamma}_{222} = -3/\sigma^3 \qquad \bar{\Gamma}_{111} = \bar{\Gamma}_{122} = \bar{\Gamma}_{212} = \bar{\Gamma}_{221} = 0.$$

The α-connections are

$$\overset{\alpha}{\Gamma}_{122} = (1-\alpha)/\sigma^3 \qquad \overset{\alpha}{\Gamma}_{121} = \overset{\alpha}{\Gamma}_{211} = -(1+\alpha)/\sigma^3$$

$$\overset{\alpha}{\Gamma}_{222} = (-3-4\alpha)/\sigma^3 \qquad \overset{\alpha}{\Gamma}_{111} = \overset{\alpha}{\Gamma}_{122} = \overset{\alpha}{\Gamma}_{212} = \overset{\alpha}{\Gamma}_{221} = 0,$$

or in the Γ^k_{ij}-symbols:

$$\overset{\alpha}{\Gamma}{}^2_{11} = (1-\alpha)/3\sigma \qquad \overset{\alpha}{\Gamma}{}^1_{21} = \overset{\alpha}{\Gamma}{}^1_{21} = -(1+\alpha)/\sigma$$

$$\overset{\alpha}{\Gamma}{}^2_{22} = -(3+4\alpha)/3\sigma.$$

The curvature tensor can be calculated to be

$$\overset{\alpha}{R}_{1212} = \frac{(1-\alpha)(3+\alpha)}{\sigma^4} \qquad \overset{\alpha}{R}_{1221} = -\frac{(1+\alpha)(3-\alpha)}{\sigma^4}.$$

So we do indeed again have a manifold that is <u>not conjugate symmetric</u>. All other components are again vanishing apart from R_{2112}, R_{2121}. The space is <u>not</u> flat for any value of α.

Considering the problem of finding all geodesic submanifolds we have the same situation as earlier in that

$$\underline{N}_{\mu_0} = \{(\mu,\sigma)\,|\,\mu=\mu_0\}$$

together constitute a foliation that is geodesic for all values of α, again in accordance with the independence of $\hat{\mu}$ and $\hat{\sigma}$.

Consider now a submanifold of the type $[t,f(t)]$ with tangent $e + \dot{f}E$. We get

$$\overset{\alpha}{\nabla}_{e+\dot{f}E}(e+\dot{f}E) = \overset{\alpha}{\nabla}_e e + 2\dot{f}\overset{\alpha}{\nabla}_e E + \dot{f}^2\overset{\alpha}{\nabla}_E E + \ddot{f}E$$

$$= -\frac{2\dot{f}}{f}(1+\alpha)e + \left[\frac{1-\alpha}{3f} - \frac{3+4\alpha}{3f}\dot{f}^2 + \ddot{f}\right]E$$

Multiplying the coefficient to e by $\dot{f}$ and everything by 3f and reducing, we obtain the following differential equation:

$$(3+2\alpha)\dot{f}^2 + 3f\ddot{f} = \alpha-1\ .$$

For $\alpha=0$ (the Riemannian case), we get

$$3\dot{f}^2 + 3f\ddot{f} = -1.$$

Letting $f = \sqrt{u}$, $u = f^2$ we obtain as in the Gaussian case the equation

$$\ddot{u} = -\frac{2}{3} \leftrightarrow u = -\frac{1}{3}t^2 + At + B,$$

i.e. again parabolas in the (μ,σ^2) parametrization but with a different coefficient to t^2.

Note that, in fact, considered as a <u>Riemannian</u> manifold there is no essential difference between this and the univariate Gaussian manifold, since we have constant scalar Riemannian curvature equal to

$$-\frac{3}{\sigma^4} \cdot \frac{\sigma^4}{3} = -1,$$

i.e. again a hyperbolic space.

If $\alpha \notin \{1,-\frac{3}{2}\}$ the following special parabolas are solutions:

$$\sigma^2 = \frac{\alpha-2}{2\alpha+3}\mu^2 + B\mu + B^2\,\frac{2\alpha+3}{1-\alpha}, \text{ B arbitrary.}$$

$\sigma^2 = \sigma_0^2$ is 1-geodesic. For $\alpha = -3/2$ <u>no</u> parabolas are geodesic. The equation then reduces to

$$f\ddot{f} = \frac{\alpha-1}{3},$$

the general solution to which cannot be obtained in a closed form.

If we consider the transformation submodel of "constant coefficient of variation" $\sigma=\gamma\mu$ corresponding to $f(t)= t$, we get the equation

$$(3+2\alpha)\gamma^2 + 0 = \alpha-1.$$

Solving this for α we find the following peculiarity:

$$\alpha = (3\gamma^2+1)/(1-2\gamma^2) \quad \text{if } \gamma^2 \neq \tfrac{1}{2}$$

<u>but if</u> $\gamma = \sqrt{2}/2$, <u>the equation has no solution</u>!! In other words, all "constant variation coefficient submanifolds" of the manifold studies are α-geodesic for suitably chosen α <u>except one</u> $(\gamma^2 = \tfrac{1}{2})$.

A reasonable explanation for this is at present beyond my imagination. Is there a missing connection $(\alpha=\infty)$? Have I made a mistake in the calculations? Or is it just due to the fact that the phenomenon is related to how this model is a submodel of the strange two-dimensional model. In any case, there is a remarkable disharmony between the group structure and the geometry.

To go a bit further we consider the three-dimensional manifold ((μ,σ,ξ)-parametrized) obtained from considering $X \sim N(\mu,\sigma^2)$, $Y \sim N(\xi,1)$. The metric for this becomes

$$g = \begin{pmatrix} \frac{1}{\sigma^2} & 0 & 0 \\ 0 & \frac{2}{\sigma^2} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and the skewness-tensor and the α-connections are identical to the Gaussian case when only indices 1 and 2 appear and all involving the third coordinate are equal to zero. Letting (e,E,F) denote the basis vectors for the tangent space determined by coordinatewise differentiation, we consider now the "constant coefficient of variation" submanifold:

$$\{(t,\gamma t, \log \gamma\, t),\ t \in \mathbb{R}^+\}$$

with tangent-vector $N = e + \gamma E + \frac{1}{t}F$, and we get

$$\overset{\alpha}{\nabla}_N N = \overset{\alpha}{\nabla}_{e+\gamma E}(e+\gamma E) + (-\frac{1}{t^2})F$$

$$= \overset{\alpha}{\nabla}_e e + 2\gamma\overset{\alpha}{\nabla}_e E + \gamma^2\overset{\alpha}{\nabla}_E E - \frac{1}{t^2}F$$

Inserting the expressions for the α-connections we obtain

$$\overset{\alpha}{\nabla}_N N = -2\,\frac{1+\alpha}{t}e - (\frac{\alpha-1}{2\gamma t} + \gamma\frac{1+2\alpha}{t})E - \frac{1}{t^2}F$$

$$= -\frac{1}{t}[2(1+\alpha)e + (\frac{\alpha-1}{2\gamma} + \gamma(1+2\alpha))E + \frac{1}{t}F\,].$$

If this derivative shall be in N's direction we must have

$$2(1+\alpha) = 1 \rightarrow \alpha=-\tfrac{1}{2},$$

but also

$$\gamma = \frac{\alpha-1}{2\gamma} + \gamma(1+2\alpha) \rightarrow 2\gamma^2 = -\frac{3}{2},$$

which is impossible. We conclude thereby that this transformational model is <u>not α-geodesic for any α</u>, considered as a submodel of the full exponential model.

9. DISCUSSION AND UNSOLVED PROBLEMS

The present paper seems to raise more questions than it answers. We want to conclude by pointing out some of these, thereby hoping to stimulate research in the area.

1. How much structure of a statistical model is captured by its "statistical manifold", the manifold being defined through expected geometries as by Amari, minimum contrast geometries as by Eguchi or observed geometries as by Barndorff-Nielsen? On the surface it looks as if only structures up to third order are there and as if one should include symmetric tensors of higher order to capture more.

2. Some statistical manifolds $(\underline{M}_1, g_1, D_1)$ and $(\underline{M}_2, g_2, D_2)$ are "alike", locally as well as globally. Various types of alikeness seems to be of some interest. Of course the full isomorphism, i.e. maps from $\underline{M}_1$ to $\underline{M}_2$ that preserves both the Riemannian metric and the skewness tensor. But also maps that preserve some structure, but not all could be of interest, in analogy with the notion of a conformal map in Riemannian geometry (maps that preserve angles, i.e. the metric up to multiplication with a function). There are several possibilities here. Isometries that preserve the skewness tensor up to a scalar or up to a function. Maps that preserve the metric up to scalars and/or functions and do and do not preserve skewness etc. etc.

3. In connection with the above there remains to be done a lot of work on classification of statistical manifolds in a pure mathematical sense, i.e. characterize manifolds up to various type of "conformal" equivalence, "conformal" here taken in the senses described above. A classic result is that

two Riemannian manifolds are locally isomorphic if they have identical curvature tensors. Do similar things hold for statistical manifolds and their α-curvatures? Note that the inverse Gaussian and Gaussian manifolds seem to be alike but not fully isomorphic. Results of Amari (1985) seem to indicate that α-flat families are very similar to exponential families. Are they in some sense equivalent? There might be many interesting things to be seen in this direction.

4. Some statistical manifolds seem to have special properties. As mentioned above we have e.g. $\underline{\alpha\text{-flat families}}$, but also manifolds that are <u>conjugate symmetric</u> or manifolds with $\underline{\text{constant } \alpha\text{-curvature}}$ both for a particular α and for all α at the same time. Which maps preserve these properties? Can they in some sense be classified?

5. How does the geometric structures behave when we form <u>marginal</u> and <u>conditional</u> experiments? Some work has been done on this by Barndorff-Nielsen and Jupp (1984, 1985).

6. Is there a <u>decomposition theory</u> for statistical manifolds. We have seen that there might be a connection between the existence of geodesic foliations and independence of estimates. There might be a de Rham-like theory to be discovered by studying parallel transports along closed curves in flat manifolds?

7. Chentsov (1972) showed that the expected geometries were the only ones that obeyed the axioms of a decision theoretic view of statistics, in the case of finite sample spaces. It seems of interest to investigate generalizations of this result, both to more general spaces and to other foundational frameworks. Picard (1985) has generalized the result to the case of exponential families and has some results pertaining to the general case.

8. What insight can be gained by studying the difference between observed and expected geometries?

9. How is the relation between the geometric structure of a Lie-transformation group and the geometric structure of its transformational statis-

tical models?

Other questions and problems are raised by Barndorff-Nielsen, Cox, and Reid (1986) and in the book by Amari (1985).

Acknowledgements

The author is grateful to Ole Barndorff-Nielsen, Preben Blaesild, and Erik Jørgensen for discussions relevant to this manuscript at various stages.

REFERENCES

Amari, S.-I. (1982). Differential geometry of curved exponential families - curvatures and information loss. Ann. Statist. 10, 357-385.

Amari, S.-I. (1985). Differential-Geometrical Methods in Statistics. Lecture Notes in Statistics Vol. 28, Springer Verlag. Berlin, Heidelberg.

Atkinson, C. and Mitchell, A. F. S. (1981). Rao's distance measure. Sankhya A 43 345-365.

Barndorff-Nielsen, O. E. and Blaesild, P. (1983). Exponential models with affine dual foliations. Ann. Statist. 11 753-769.

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

Barndorff-Nielsen, O. E. and Jupp, P. E. (1984). Differential geometry, profile likelihood and L-sufficiency. Res. Rep. 118. Dept. Theor. Stat., Aarhus University.

Barndorff-Nielsen, O. E. and Jupp, P. E. (1985). Profile likelihood, marginal likelihood and differential geometry of composite transformation models. Res. Rep. 122. Dept. Theor. Stat., Aarhus University.

Boothby, W. S. (1975). An Introduction to Differentiable Manifolds and Riemannian Geometry, Academic Press.

Chentsov, N. N. (1972). Statistical Decision Rules and Optimal Conclusions (in Russian) Nauka, Moscow. Translation in English (1982) by Amer. Math. Soc. Rhode Island.

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

Eguchi, S. (1983). Second order efficiency of minimum contrast estimators in a curved exponential family. Ann. Statist. 11 793-803.

Picard, D. (1985). Invariance properties of the Fisher-Rao metric and Chentsov-Amari connections using le Cam deficiency. Manuscript. Orsay, France.

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

Skovgaard, L. T. (1984). A Riemannian geometry of the multivariate normal model. Scand. J. Statist. 11 211-223.

Spivak, M. (1970-75). Differential Geometry Vol. I-V. Publish or Perish.