Ann. Inst. Statist. Math.
Vol. 44, No. 1, 45–61 (1992)

STATISTICAL MORPHISMS AND RELATED INVARIANCE PROPERTIES

DOMINIQUE B. PICARD

Université Paris VII, U.F.R. de Mathematiques, Tour 45-55, 5eme etage. 2, Place Jussieu 75251 Paris, France
CNRS Unites associées 1321 et 1323 Statistique des Processus en Milieu Médical Université Paris V, 45 Rue des Saints Pères, 75005 Paris, France

(Received December 12, 1988; revised September 25, 1990)

Abstract. Our aim is to investigate a way to characterize the elements of a statistical manifold (the metric and the family of connections) using invariance properties suggested by Le Cam's theory of experiments. We distinguish the case where the statistical manifold is flat. Then, there naturally exists an entropy and it is proven that experiment invariance is equivalent to entropy invariance. If the statistical manifold is not flat, we introduce a notion of local invariance of selected order associated to the asymptotic (on n observations, n tending to infinity) expansion of the power of the Neymann Pearson test in a contiguous neighborough of some point. This invariance provides a substantial number of morphisms. This was not always true for the entropy invariance: particularly, the case of Gaussian experiments is investigated where it can be proven that entropy invariance does not characterize a metric or a family of connections.

Key words and phrases: Statistical manifold, Amari connections, comparison of experiments, likelihood expansions, asymptotic properties of tests.

1. Introduction

In the very last few years, many authors have pointed out the importance of the differential geometrical tools in Statistics (Chentsov (1972), Efron (1975), Amari (1982, 1983, 1985, 1987), Amari and Kumon (1983), Millar (1983), Barndorf-Nielsen *et al.* (1986) and Barndorf-Nielsen (1987)).

It seems that the natural frame in this context is the notion of statistical manifold (Lauritzen (1984) and Le Cam (1986)) i.e. a Riemannian manifold with a pair of dual torsion free connections. This is equivalent to consider a triple (E, g, T), where E is a manifold, g a Riemannian metric, and T a symmetric covariant tensor of order 3. Indeed, it is worthwhile to consider most of the introduced geometrical tools in this framework (Amari's connections, Barndorff-Nielsen's expected geometry, family of connections associated to a ψ-estimator

(Eguchi) or quasi-maximum likelihood estimators (Lauritzen (1984)).

Considering a particular triple (E, g, T) where E is a family of probabilities, a fundamental problem raising then, is what is remaining of a statistical problem after a reduction to the statistical manifold. This problem has obviously two faces: the first one consists in investigating the kind of morphisms that may be statistically important for a geometry to be invariant with. The second part is, of course, given a class of morphisms (statistically interesting), "what are the statistical manifolds that are invariant under this class and more precisely, when is a geometry uniquely determined, or, in other words, what class of morphisms is determining the geometry?" For the first part, our approach found its motivation in the work of Chentsov (1972). In a decision theory framework, Chentsov constructed a geometry where two experiments E_1 and E_2 are equivalent if you can pass from one to the other through two Markov kernels. This is perfectly natural from a decision point of view, since, in this case, for every chosen loss function, and every decision made upon E_1, there is a decision made upon E_2 with the same risk function, and conversely.

It is proven in Section 3 that, in linear exponential families, the Markov morphisms can be characterized in terms of invariance of the entropy function. Hence, the notion has a natural extension in a general flat manifold, where E needs not to be a probability measure space. Another aspect of this class of morphisms is that it can be relatively small: it is shown, in Section 5, that, although they happen to determine the geometry in simple cases (for instance when the entropy is quadratic (Proposition 5.1)), in the general case, this is not true (Proposition 5.2). This is one of the reasons why this class has been extended in Section 4, considering morphisms that are asymptotically preserving the Neyman-Pearson test power function respectively at the first and second orders. The essential result of this section is that those approximate Markov morphisms can be characterized as those leaving invariant respectively the geometries (E, F) (the classical Fisher Riemannian geometry) and (E, F, T) (the Amari's geometry). Here again, the notion has a natural extension in a general statistical monifold. But, this time, the manifold needs not to be flat.

2. Definitions and notations

2.1 *The experiment E*

We consider a measurable space (Ω, A) with a family of probabilities P_θ defined on (Ω, A), θ varying in an open connected subset Θ of $\mathbb{R}^k$. The family $E = \{P_\theta, \theta \in \Theta\}$ is assumed to satisfy the following conditions:

1) For every θ in Θ, P_θ is absolutely continuous with respect to a given σ-finite measure μ. We shall denote by $p(\cdot, \theta)$ the Radon-Nikodym derivative of P_θ with respect to μ:

$$p(x, \theta) = \frac{dP_\theta}{d\mu}(x).$$

2) $p(x, \cdot)$ is in $C^3(\Theta)$ for μ-almost all x in Ω.

3) If ∂_i denotes the differentiation $\partial/\partial\theta_i$, and $l(\cdot, \theta) = \log p(\cdot, \theta)$, then the functions $\partial_i l(\cdot, \theta)$ $(i = 1, \ldots, k)$ are in $L^2(dP_\theta)$ and are linearly independent. This

shows that $F_{ij}(\theta) = E_\theta[\partial_i l(x,\theta)\partial_j l(x,\theta)]$ is finite for every $i, j = 1, \dots, k$ and θ in Θ. In particular, the Fisher matrix $F(\theta) = (F_{ij}(\theta))_{1 \le i \le k, 1 \le j \le k}$ is (strictly) positive-definite for every θ in Θ.

4) The mapping $\theta \to F(\theta)$ is of class $C^1(\Theta)$ and the quantities:

$$
\begin{aligned}
&T_{ijk}(\theta) = E_\theta \partial_i l(x,\theta)\partial_j l(x,\theta)\partial_k l(x,\theta) \qquad \text{and} \\
&\Gamma^1_{ijk}(\theta) = E_\theta \partial_i \partial_j l(x,\theta)\partial_k l(x,\theta)
\end{aligned}
$$

are finite, with

$$
2E_\theta \partial_i \partial_j \partial_k l(x,\theta) = \partial_i F_{jk}(\theta) + \partial_j F_{ik}(\theta) + \partial_k F_{ij}(\theta) + T_{ijk}(\theta)
$$

for every $i, j = 1, \dots, k$ and θ in Θ.

5) If $R_\theta(\cdot, h)$ denotes the remainder of the Taylor expansion of $\log p(x, \theta + h) - \log p(x,\theta)$ up to order 3, we assume that $|h|^3 R_\theta(\cdot, h)$ is uniformly bounded by a function ψ which has moments up to order 3 for each θ in Θ.

Following Amari's formulation (1985), under the previous assumptions, the family E is sufficiently smooth to permit the introduction of a k-dimensional manifold structure on the statistical model, with $\theta \in \Theta$ playing the role of a coordinate system.

Rao (1945) has proven that the Fisher-Rao metric $F(\theta)$ defines a Riemannian structure on E letting $F_\theta(\partial_i, \partial_j) = F_{ij}(\theta)$ where ∂_i denotes $\partial/\partial\theta_i$ the natural basis vector of the tangent space at the point θ and $F_\theta(\partial_i, \partial_j)$ the inner product of the two vectors in the metric $F(\theta)$. Amari (1985) introduced the α-connections $\overset{\alpha}{\nabla}$, with $T(\partial_i, \partial_j, \partial_k)|_\theta = T_{ijk}(\theta)$, and if $T(E)$ is the set of vector fields of E,

$$
\forall X, Y, Z \in T(E), \qquad F(\overset{\alpha}{\nabla}_X Y, Z) = F(\overset{0}{\nabla}_X Y, Z) - \frac{\alpha}{2} T(X, Y, Z)
$$

where $\overset{0}{\nabla}$ is the Riemannian connection associated with F.

A particularly interesting example for the experiment is the following linear exponential family. Let $\boldsymbol{B}(\mathbb{R})$ be the Borelian σ-algebra on $\mathbb{R}$.

DEFINITION 1. A statistical experiment is a linear exponential family if and only if there exist k measurable mappings: $S_j : (\Omega, A) \to (\mathbb{R}, \boldsymbol{B}(\mathbb{R}))$, such that: $\forall\theta \in \Theta$, $p(x,\theta) = \exp\{(\theta, S(x)) - h(\theta)\}$ where $S = (S_1, \dots, S_k)$, $(\theta, S) = \sum \theta_i S_i$, and $\exp h(\theta) = \int_\Omega \exp(\theta, S(x)) d\mu(x)$.

We shall take Θ to be the interior of the domain of definition of h and assume that it is convex.

2.2 *Statistical manifolds*

Let us consider a statistical manifold (Lauritzen (1984)) e.g. a triple (E, g, R), where E is a Riemannian manifold, g is the metric and R is a covariant symmetric tensor of order 3. It is equivalent to consider $(E, g, \overset{\alpha}{\nabla}, \alpha \in \mathbb{R})$, where the $\overset{\alpha}{\nabla}$ are constituting the family of torsion free connections associated to (E, g, R) through the following formula:

$$g(\overset{\alpha}{\nabla}_X Y, Z) = g(\overset{0}{\nabla}_X Y, Z) - (\alpha/2)R(X, Y, Z), \qquad \forall X, Y, Z \in T(E).$$

$\overset{0}{\nabla}$ is the Riemannian connection associated with g.

At this stage, E does not need to be a family of probabilities, but if E is an experiment most of the geometrical tools may be relevantly considered in this framework (Amari's connections (1985), Barndorff-Nielsen's expected geometry (1987), family of connections associated with a ψ-estimator (Eguchi (1983)), quasi-likelihood (Lauritzen (1987))).

One says that $(E, g, \overset{\alpha}{\nabla}, \alpha \in \mathbb{R})$ is the likelihood statistical manifold associated to an experiment E if g is the Fisher metric (Rao (1945)) and $\overset{\alpha}{\nabla}$ are the Amari connections (1985). In the particular case of the linear exponential family experiment, the likelihood statistical manifold associated with the family E is such that, in θ-coordinate:

$$F_\theta(\partial_i, \partial_j) = F_{ij}(\theta) = \partial^2_{ij} h(\theta),$$
$$T(\partial_i, \partial_j, \partial_k)|_\theta = T_{ijk}(\theta) = \partial^3_{ijk} h(\theta).$$

2.3 *Invariances of metrics and connections in manifolds*

Let $E' = \{P_\theta, \theta \in \Theta'\}$ be a regular submanifold of E and ϕ be a mapping defined on E' such that ϕ is a diffeomorphism on its image. Let φ be the associated diffeomorphism defined on θ' by $\phi(P_\theta) = P_{\varphi(\theta)}$. Let us introduce the following notations: for X, vector field of $T(E')$: X^ϕ is the induced vector field i.e. the vector field such that for every C^∞ function k on $\phi(E')$:

$$X^\phi k(\cdot)|_{\phi(P_\theta)} = Xk(\phi(\cdot))|_{P_\theta}, \qquad \forall P_\theta \in E'.$$

g^ϕ is the induced metric i.e.:

$$g(X^\phi, Y^\phi)|_{\phi(P_\theta)} = g^\phi(X^\phi, Y^\phi)|_{P_\theta}, \qquad \forall P_\theta \in E'.$$

∇^ϕ is the induced affine connection i.e.:

$$\nabla_{X^\phi} Y^\phi|_{\phi(P_\theta)} = \nabla^\phi_{X^\phi} Y^\phi|_{P_\theta}, \qquad \forall P_\theta \in E'.$$

Let us recall the following definition (cf. for example Wolf (1967) and Spivak (1979)).

DEFINITION 2. Let ϕ defined as above. One says that ϕ is an isometry for the metric g (resp. an affine diffeomorphism for the connection ∇) or that g (resp. ∇) is invariant under ϕ if and only if, for every X and Y vector fields on E',

$$g(X,Y) = g^{\phi}(X^{\phi}, Y^{\phi}) \tag{2.1}$$

(resp.

$$\nabla_X Y = \nabla^{\phi}_{X^{\phi}} Y^{\phi}). \tag{2.2}$$

Remarks.

1) The essential meaning of this invariance is that ϕ preserves the curve length (resp. maps geodesics into geodesics).

2) In coordinate $(\theta_1, \ldots, \theta_k)$, choosing $X = \partial_i$, $Y = \partial_j$, $\varphi = (\varphi^1, \ldots, \varphi^k)$, then

$$X^{\phi} = \sum_l \frac{\partial \varphi^l}{\partial \theta_i} \frac{\partial}{\partial \varphi_l}, \qquad g^{\phi}(X^{\phi}, Y^{\phi})|_{P_\theta} = ({}^t D\varphi|_\theta \, g(-)|_{\varphi(\theta)} \, D\varphi|_\theta)_{ij}$$

and (2.1) is equivalent to:

$$\begin{aligned} &{}^t u g(-)|_\theta \, u = {}^t u {}^t D\varphi|_\theta \, g(-)|_{\varphi(\theta)} \, D\varphi|_\theta \, u \\ &\forall u \in \mathbb{R}^k / \sum u^i \partial_i \in T_\theta(E'), \qquad \forall \theta \in \Theta, \end{aligned} \tag{2.1'}$$

where $D\varphi$ is the Jacobian matrix of φ, ${}^t D\varphi$ is its transposed, and $g(-)$ denotes the matrix composed by the $g(\partial_l, \partial_m)|_{P_\theta}$ (also denoted $g_{lm}(\theta)$).

3) Let g' be any metric on the statistical experiment which is invariant with respect to ϕ, then (2.2) is equivalent to $g'(\nabla^{\phi}_{X^{\phi}} Y^{\phi}, Z^{\phi}) = g'(\nabla_X Y, Z)$ for every X and Y vector fields on E'. So that, if Γ_{ijl} denotes $g(\nabla_{\partial_i} \partial_j, \partial_l)$, (2.2) is equivalent to:

$$\begin{aligned} &u_1^i u_2^j u_3^l \left\{ \frac{\partial \varphi^s}{\partial \theta_i} \frac{\partial \varphi^r}{\partial \theta_j} \frac{\partial \varphi^m}{\partial \theta_l} \Gamma_{srm}(\varphi(\theta)) + \frac{\partial^2 \varphi^s}{\partial \theta_i \partial \theta_j} \frac{\partial \varphi^r}{\partial \theta_l} g_{sr}(\varphi(\theta)) \right\} \\ &\quad = u_1^i u_2^j u_3^l \Gamma_{ijl}(\theta) \\ &\forall u_1, u_2, u_3 \in \mathbb{R}^k / u^i \partial_i \in T_\theta(E'), \qquad \forall \theta \in \Theta. \end{aligned} \tag{2.2'}$$

DEFINITION 2'. A metric g (resp. an affine connection ∇) on a manifold E is said to be locally (at the point θ^0) invariant under a morphism ϕ defined as above on E' if ϕ is a local isometry (resp. a local affine diffeomorphism) i.e. if (2.1) (resp. (2.2)) is true at the point θ^0.

3. Markov morphisms in experiments and entropy morphisms in flat statistical manifolds

3.1 *Markov morphisms*

Let us recall some classical tools of the statistical decision theory (cf. Millar (1983), Le Cam (1986)):

i) Let (Ω, A), (Ω', A') be two measurable spaces. π: $\Omega \times \Omega' \to \mathbb{R}$ is a Markov kernel from (Ω, A) to (Ω', A') if:

$$\begin{aligned} &\forall \omega \in \Omega, && \pi(\omega, \cdot) \text{ is a probability measure on } (\Omega', A') \\ &\forall B \in A', && \pi(\cdot, B) \text{ is } A\text{-measurable.} \end{aligned}$$

If P is a probability measure on (Ω, A) and π a Markov kernel, let us denote by πP the probability measure on $(\Omega', A') : \int \pi(\omega, \cdot) dP(\omega)$.

ii) Let $E = \{P_\theta, \theta \in \Theta\}$ (resp. $\{Q_\theta, \theta \in \Theta\}$) be an experiment on (Ω, A) (resp. (Ω', A')) indexed by Θ, E and E' are said to be equivalent if there exist a Markov kernel π from (Ω, A) to (Ω', A'), and a Markov kernel π' from (Ω', A') to (Ω, A), such that $\forall \theta \in \Theta$, $P_\theta = \pi' Q_\theta$, $Q_\theta = \pi P_\theta$. This notion is natural from a decision point of view, since, in this case, for every chosen loss function and every finite decision made upon E, there is a finite decision made upon E' with the same risk, and conversely.

DEFINITION 3. Let $E_1 = \{P_\theta, \theta \in T_1\}$ be a regular submanifold of $E = \{P_\theta, \theta \in \Theta\}$ of dimension r $(r > 0)$. A diffeomorphism $\phi : P_\theta \to \phi(P_\theta) = P_{\varphi(\theta)} \in E$, from E_1 to $E_2 = \{P_{\varphi(\theta)}, \theta \in T_1\}$ is called a Markov morphism of the experiment E if the 2 sub-experiments E_1 and E_2 are equivalent.

3.2 *Entropy morphisms in flat statistical manifolds*

Let us recall that (E, g, R) is said to be flat if it is flat for one of the α-connections (say α_0). This notion is equivalent to the existence of a canonical system of coordinates τ (where the α_0-geodesics are straight lines) and an entropy function h such that g and R are derived from h through the formulas:

$$\begin{aligned} g_{ij} &= g(\partial_i, \partial_j) = \partial_{ij} h, \\ R_{ijk} &= R(\partial_i, \partial_j, \partial_k) = \partial_{ijk} h, \qquad (\partial_i = \partial / \partial \tau_i) \end{aligned} \qquad \text{Amari (1985).}$$

For instance, the likelihood statistical manifold associated to a linear exponential family is flat, with $\alpha_0 = +1$, $\tau = \theta$. In this case, the following morphisms appear to be particularly relevant: let E' be an r-dimensional regular submanifold of E $(r > 0)$. E' will be assumed to be α_0-convex (i.e. convex in τ-coordinate).

DEFINITION 4. Let ϕ be a morphism defined on E', ϕ is said to be an entropy morphism of the flat statistical manifold E if and only if:

1) ϕ is affine in τ coordinate.
2) The mapping $h \circ \phi - h$ is affine in τ coordinate over E'.

3.3 *Markov morphisms in linear exponential families*

The following proposition gives a characterization of Markov morphisms in linear exponential families.

PROPOSITION 3.1. *If E is a linear exponential family, ϕ a Markov morphism defined on an r-dimensional submanifold $E' = \{P_\theta, \theta \in T'\}$, then:*

1) ϕ can be extended as a Markov morphism to $C(E')$ corresponding to the convex hull of T' in $\Theta : C(E') = \{P_\theta, \theta \in C(T')\}$

2) ϕ is an entropy morphism of the likelihood statistical manifold associated with E, i.e.:

α) φ is an affine mapping:

$$\varphi(\theta) = A\theta + b, \qquad \forall\theta \in C(T')$$

β) the mapping $\theta \to h(\varphi(\theta)) - h(\theta)$ is affine i.e. $\forall 0 \leq \alpha \leq 1$, $\forall \theta_1, \theta_2 \in C(T')$,

$$\begin{aligned}(3.1) \qquad & h(\alpha\theta_1 + (1-\alpha)\theta_2) - \alpha h(\theta_1) - (1-\alpha)h(\theta_2) \\ & = h(\alpha\varphi(\theta_1) + (1-\alpha)\varphi(\theta_2)) - \alpha h(\varphi(\theta_1)) - (1-\alpha)h(\varphi(\theta_2)).\end{aligned}$$

The proof of Proposition 3.1 is based on the following lemma (Chentsov (1972), IV, Section 18, Lemma 18.7).

LEMMA 3.1. *If a pair of mutually absolutely continuous probability distributions on X, A, $\{P_0, P_1\}$ is Markov equivalent to a pair $\{R_0, R_1\}$ on Y, B, i.e. two Markov kernels π_{12} and π_{21} exist, such that,*

$$P_i = R_i \circ \pi_{21}, \qquad R_i = P_i \circ \pi_{12} \qquad (i = 0, 1)$$

then the equivalence may be extended to the two Hellinger arcs, i.e.: if P_α (resp. R_α) is the probability distribution with density with respect to the measure P_0 (resp. R_0),

$$\left(\frac{dP_1}{dP_0}\right)^{1-\alpha} \exp \bar{h}(\alpha) \qquad \left(\textit{resp.}\ \left(\frac{dR_1}{dR_0}\right)^{1-\alpha} \exp \bar{H}(\alpha)\right)$$

where

$$\exp \bar{h}(\alpha) = \left[\int \left(\frac{dP_1}{dP_0}\right)^{1-\alpha} dP_0\right]^{-1}$$

$$\left(\textit{resp.}\ \exp \bar{H}(\alpha) = \left[\int \left(\frac{dR_1}{dR_0}\right)^{1-\alpha} dR_0\right]^{-1}\right)$$

then, $P_\alpha = R_\alpha \circ \pi_{21}$, $R_\alpha = P_\alpha \circ \pi_{12}$, and

$$(3.2) \qquad \bar{h}(\alpha) = \bar{H}(\alpha), \qquad \textit{for every } \alpha, \qquad 0 \leq \alpha \leq 1.$$

PROOF OF PROPOSITION 2.1. We must have: $\forall 0 \le \alpha \le 1$, $\forall \theta_1, \theta_2 \in T'$: if $\alpha\theta_1 + (1-\alpha)\theta_2 \in T'$ then $\varphi(\alpha\theta_1 + (1-\alpha)\theta_2) = \alpha\varphi(\theta_1) + (1-\alpha)\varphi(\theta_2)$, using Lemma 3.1, otherwise, φ can be extended through the previous formula, i.e. φ becomes affine on $C(T')$: $\varphi(\theta) = A\theta + b$. It is not difficult to see that if φ was a diffeomorphism from T' to T'', it still is from $C(T')$ to $C(T'')$.

(3.1) is equivalent to (3.2) since $\bar{h}(\alpha) = h(\alpha\theta_1 + (1-\alpha)\theta_2) - \alpha h(\theta_1) - (1-\alpha)h(\theta_2)$, and the proof is complete. □

Let us take two familiar examples (Gaussian translation families and Gaussian families with unknown mean and unknown variance) and examine in both cases the Markov morphisms.

PROPOSITION 3.2. *For $E = \{N(\theta, F_0^{-1}), \theta \in \mathbb{R}^k\}$, where F_0 is a fixed positive definite matrix, then the set of Markov morphisms are the following*: $\varphi(\theta) = A\theta + b$, *where A is partially F_0 unitary i.e. A fixed subspace V of $\mathbb{R}^k$ exists such that*:

$$ {}^t x {}^t A F_0 A x = {}^t x F_0 x, \qquad \forall x \in V. $$

Remarks.

1) The expression "partially" unitary is referring to the space V where the equality of the two quadratic forms F_0 and ${}^t A F_0 A$ is true.

2) V is the vector space in $\mathbb{R}^k$ spanned by all the directions of segments contained in the convex hull $C(T')$.

3) The proof of Proposition 3.2 is easily performed using Proposition 3.1: let us observe that (3.1) is true if and only if, for any straight line $\gamma : [0,1] \to C(T')$, we have:

$$ \frac{d^2}{dt^2}[h(\varphi(\gamma(t))) - h(\gamma(t))] = 0 $$

which turns out to be the condition of Proposition 3.2

Let us now use the previous proposition to investigate an example:

PROPOSITION 3.3. *For $E = \{N(\mu, \sigma^2), (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_*^+\}$, the Markov morphisms consist in the following mappings*: (*in coordinate* (μ, σ^2)) $\varphi^{cd} : (\mu, \sigma^2) \to ((c\mu + d)/c^2, \sigma^2/c^2)$, *for any constants c, d, $c \neq 0$.*

PROOF OF PROPOSITION 3.3. In this proof, we shall always use the natural coordinate system:

$$ \theta_1 = \frac{\mu}{\sigma^2}, \qquad \theta_2 = \frac{1}{\sigma^2}, \qquad \theta = (\theta_1, \theta_2) $$

let us observe that ϕ^{cd} is θ linear since, in θ coordinate ϕ writes:

$$ \varphi_1^{cd}(\theta_1, \theta_2) = c\theta_1 + d\theta_2, \qquad \varphi_2^{cd}(\theta_1, \theta_2) = c^2\theta_2. $$

First we shall show that the morphisms φ considered in the proposition are actually Markov morphisms: this is of the upmost simplicity since it is clear that each φ^{cd}

can be associated with the following two Markov kernels: $x_{12}(x, dy)$ is the Dirac measure on $c^{-1}x + dc^{-2}$ and $\pi_{21}(x, dy)$ is the Dirac measure on $cx - dc^{-1}$. We shall, now, prove the converse.

By proposition 3.1, we have only two kinds of Markov morphisms to consider: those that are defined on a line segment in Θ and those defined on open sets of Θ. Only the first ones will cause some difficulties.

The proof will be made in three steps: First, it will be proven that, if a Markov morphism φ is defined on a set consisting of a line segment parallel to one of the axis, then φ is of the form described in the proposition. Secondly, it will be shown that, if φ is a Markov morphism defined only on a line segment in Θ, then there is another Markov morphism ζ, strongly connected with φ, which is defined on a line segment parallel to one of the axis. Thirdly we shall prove, because of the relationship between φ and ζ, that φ is the form described in the proposition.

i) Assume that $\varphi(\theta) = A\theta + \theta^0$, with

$$A = \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix}, \qquad \theta^0 = \begin{pmatrix} \theta_1^0 \\ \theta_2^0 \end{pmatrix}$$

is a Markov morphism defined on T' that contains a line segment parallel to the second axis (resp. the first), say the one connecting the points (θ_1, θ_2), (θ_1, θ_2') $((\theta_1, \theta_2)$, (θ_1', θ_2) resp.). We observe that (3.1) becomes: $\forall 0 \leq \alpha \leq 1$:

$$\begin{aligned}
(3.3) \qquad & -\frac{\theta_1^2}{(\theta_2' + \alpha(\theta_2 - \theta_2'))} + \log(\theta_2' + \alpha(\theta_2 - \theta_2')) + \alpha\frac{\theta_1^2}{\theta_2} \\
& - \alpha \log \theta_2 + (1-\alpha)\frac{\theta_1^2}{\theta_2'} - (1-\alpha)\log\theta_2' \\
& = -\frac{(a_{11}\theta_1 + a_{21}\theta_2' + \theta_1^0 + \alpha[a_{21}(\theta_2 - \theta_2')])^2}{(a_{12}\theta_1 + a_{22}\theta_2' + \theta_2^0 + \alpha[a_{22}(\theta_2 - \theta_2')])} \\
& \quad + \log(a_{12}\theta_1 + a_{22}\theta_2' + \theta_2^0 + \alpha[a_{22}(\theta_2 - \theta_2')]) \\
& \quad + \alpha\frac{(a_{11}\theta_1 + a_{21}\theta_2 + \theta_1^0)^2}{(a_{12}\theta_1 + a_{22}\theta_2 + \theta_2^0)} - \alpha\log(a_{12}\theta_1 + a_{22}\theta_2 + \theta_2^0) \\
& \quad + (1-\alpha)\frac{(a_{11}\theta_1 + a_{21}\theta_2' + \theta_1^0)^2}{(a_{12}\theta_1 + a_{22}\theta_2' + \theta_2^0)} \\
& \quad - (1-\alpha)\log(a_{12}\theta_1 + a_{22}\theta_2' + \theta_2^0).
\end{aligned}$$

First, let us remark that, because of the functions involved in (3.3) the equality is true not only for $0 \leq \alpha \leq 1$, but on the definition range of (3.3). If we let in (3.3) α tend to $-\theta_2'/(\theta_2 - \theta_2')$, then LHS tends to $-\infty$. This implies that it must do RHS i.e. $\theta_2^0 + a_{12}\theta_1 = 0$. Rewriting (3.3) and let again α tend to $-\theta_2'/(\theta_2 - \theta_2')$, we obtain $a_{22}\theta_1^2 = (a_{11}\theta_1 + \theta_1^0)^2$ i.e. on the segment $[(\theta_1, \theta_2), (\theta_1, \theta_2')]$, φ coincides with the morphism indicated in Proposition 3.3 associated with the following values:

$$c = a_{11} + \frac{\theta_1^0}{\theta_1}, \qquad d = a_{21}, \qquad \text{if } \theta_1 \neq 0,$$

$$c = a_{22}^{1/2}, \qquad d = a_{21}, \qquad \text{if } \theta_1 = 0.$$

An analogous evaluation (α tends to infinity), in the case of a segment parallel to the first axis proves that $a_{11}^2 = a_{22} + \theta_2^0/\theta_2$. This implies that also in this case φ coincide on the considered segment with a morphism of the form described in Proposition 3.3, with $c = a_{11}$, $d = a_{21} + \theta_1^0/t_2$. It is easy to see that if T' contains an open set in $\mathbb{R}^2$, the two previous statements agree together as well as with the form indicated in Proposition 3.3.

ii) Suppose now that an arbitrary Markov morphism φ is defined on a set T' containing a segment of line $[\theta, \theta']$ which is not parallel to the axis. It is then easy to establish that $\varphi \circ \varphi^{cd}$ is a Markov morphism defined at least on $(\varphi^{cd})^{-1}[\theta, \theta']$ for a chosen pair (c, d). As $(\varphi^{cd})^{-1}$ is easily found to be equal to $\varphi^{c^{-1}, -dc^{-3}}$, to obtain the second step of the proof we must show that every segment $[\theta, \theta']$ may be mapped into a segment parallel to one of the axis: this may be accomplished by choosing for example: $dc^{-2} = (\theta_1 - \theta_1')/(\theta_2 - \theta_2')$.

iii) By the first step, $\varphi \circ \varphi^{cd}$ is some $\varphi^{c'd'}$. If $\varphi(\theta) = A\theta + b$, then $\varphi \circ \varphi^{cd} = AC\theta + b$, where C is the matrix associated with φ^{cd}. Thus it follows that $b = 0$ and AC has the same form as C, therefore A has the same form as C i.e. φ is also some φ^{cd}. This concludes the third step. □

We will not go further on with examples. Let us only mention that Proposition 3.2 can also be used to prove the classical Chentsov's result e.g. in multinomial experiments the Markov morphisms are the morphisms obtained by permutations of the atoms (Chentsov (1972)).

Our conclusion of this section is that the Markov morphisms together with the entropy morphisms certainly are natural and genuine to consider but may eventually form a very sparse collection depending on the considered family E (even then reduced to one point—the identity—in Poisson or Inverse Gaussian distribution families for example).

4. Approximate Markov morphisms and morphisms leaving a statistical manifold locally invariant

4.1 *Approximate Markov morphisms*

As we saw in the previous section, one of the major interests of Markov equivalence of two experiments E and E' is that for every loss function and every finite decision on E there exists a finite decision on E' with the same risk (Le Cam (1986)). Another consequence is the following: let $\beta(\theta, \theta', \alpha)$ be the best test power one can obtain by testing θ against θ' at the level α. If ϕ is the Markov morphism associated with $E' = \{P_\theta, \theta \in T'\}$ and $E'' = \{P_{\varphi(\theta)}, \theta \in T'\}$

$$\beta(\theta, \theta', \alpha) = \beta(\varphi(\theta), \varphi(\theta'), \alpha), \qquad \forall \theta, \theta' \in T'$$

(Torgersen (1970)). As we previously indicated the differential geometrical arguments are especially powerful for the first and second order approximations. This suggests the introduction of the following asymptotic point of view: suppose that we have now n iid observations of our model available. Let $\beta_n(\theta, \theta', \alpha)$ be the power of likelihood-ratio test of θ against θ', based on the n iid observations at the level α.

DEFINITION 5. Let k be an integer greater than or equal to 1. φ is an approximate Markov morphism of order k (am$k$) of the experiment E at the point θ^0 if:

1) φ is defined on E', an r-dimensional regular submanifold of Θ $(r > 0)$, containing P_{θ^0}, $E' = \{P_\theta, \theta \in T'\}$ and sufficiently regular.

2) $\forall u$ such that $\theta_0 + u/n^{1/2}$ belongs to T',

$$\lim_{n\to\infty} n^{(k-1)/2}[\beta_n(\theta_0, \theta_0 + u/n^{1/2}, \alpha) - \beta_n(\varphi(\theta_0), \varphi(\theta_0 + u/n^{1/2}), \alpha)] = 0.$$

Remarks. The sequence of classes of approximate Markov morphisms of order k is a decreasing sequence. It is a consequence of Torgersen (1970) that Markov morphisms are am$k$ for every k. Amk could be useful, generalizing this approach, to determine what sort of tensors should be kept in mind at order k. But in this paper we restrict to $k = 1$ and 2.

4.2 *Morphisms leaving a statistical manifold locally invariant*

Let now (E, g, R) be a general statistical manifold.

DEFINITION 6. ϕ is a local statistical morphism of order 1 (ls1) of the statistical manifold (E, g, R) if:

1) ϕ is defined and sufficiently regular on E' an r-dimensional submanifold of E $(r > 0)$ containing θ^0.

2) g is invariant under ϕ at the point θ^0.

DEFINITION 7. ϕ is a local statistical morphism of order 2 (ls2) of the statistical manifold (E, g, R) if:

1) ϕ is an ls1.

2) $\overset{\alpha}{\nabla}$ is invariant under ϕ at the point θ^0, for every α.

Remarks. Obviously, condition 2) of Definition 7 may be replaced by the two following conditions:

1) The Riemannian connection $\overset{0}{\nabla}$ associated to the metric g is ϕ invariant at the point θ^0.

2) The tensor R is ϕ invariant at the point θ^0 i.e. $\forall X, Y, Z$ vector fields of $T(E')$,

$$R^\phi(X^\phi, Y^\phi, Z^\phi)|_{\theta^0} = R(X, Y, Z)|_{\theta^0}.$$

4.3 *Characterization and examples of approximate Markov morphisms*

Let us remind the reader that whether ϕ is an am$k$ or an ls$k$ its definition depends on a particular point θ^0 and a particular r-dimensional manifold containing this point. The following theorem makes the link between the two preceeding definitions.

THEOREM 4.1. *Let E be an experiment, and ϕ a sufficiently smooth mapping defined in a regular submanifold of E, E' on a neighborhood of θ_0,*

i) *ϕ is an* aml *if and only if ϕ is an* ls1 *associated to the likelihood statistical manifold of E.*

ii) *ϕ is an am2 if and only if ϕ is an ls2 associated to the likelihood statistical manifold of E.*

The results of Theorem 4.1 directly follow from the following lemma describing the behavior of β_n in terms of the previous geometric quantities.

LEMMA 4.1. *Under the hypothesis quoted above, for φ sufficiently smooth the following holds:*

$$\begin{aligned}
&\beta_n(\varphi(\theta^0), \varphi(\theta^0 + un^{-1/2})) \\
&\quad = N(N^{-1}(\alpha) + \|u.\varphi\|) \\
&\qquad - n^{-1/2}\eta(N^{-1}(\alpha) + \|u.\varphi\|)\{(\overset{1/3}{\nabla}_{u.\varphi}u.\varphi, u.\varphi)/2\|u.\varphi\| \\
&\qquad + (\overset{-1}{\nabla} - \overset{1}{\nabla}_{u.\varphi}u.\varphi, u.\varphi)N^{-1}(\alpha)/\|u.\varphi\|^2\} + o(n^{-1/2}).
\end{aligned}$$

Where $u = {}^t(u^1, \ldots, u^k)$ is a fixed point in $\mathbb{R}^k$, η and N denote respectively the density and the repartition functions of the normal law with zero mean and variance equal to 1. $u.\phi$ denotes the vector field $u^i\partial_i\phi^j\partial_j$ and the scalar product and norm are those associated with the Fisher metric, whereas $\overset{\alpha}{\nabla}$ is the α-connection introduced by Amari (1982). Both of them are taken at the point $\varphi(\theta^0)$.

PROOF OF LEMMA 4.1. We follow a classical way in Edgeworth expansions: cf., for example Akahira and Takeuchi (1981), where the same expansions are given in the one dimensional case with ϕ identity. Let $T_n = \sum_{i=1}^n[\log p(X_i, \varphi(\theta^0)) - \log p(X_i, \varphi(\theta^1))]$, $\theta' = \theta^0 + un^{-1/2}$. By expanding around θ^0, we obtain the following:

$$\begin{aligned}
E_{\varphi(\theta^0)}T_n &= \|u.\varphi\|^2/2 + (\overset{1/3}{\nabla}_{u.\varphi}u.\varphi, u.\varphi)/2n^{1/2} + o(n^{-1/2}), \\
E_{\varphi(\theta^0)}(T_n - E_{\varphi(\theta^0)}T_n)^2 &= \|u.\varphi\|^2 + (\overset{1}{\nabla}_{u.\varphi}u.\varphi, u.\varphi)/n^{1/2} + o(n^{-1/2}), \\
E_{\varphi(\theta^0)}(T_n - E_{\varphi(\theta^0)}T_n)^3 &= (\overset{1}{\nabla} - \overset{-1}{\nabla}_{u.\varphi}u.\varphi, u.\varphi)/2n^{1/2} + o(n^{-1/2}).
\end{aligned}$$

The same expansions can be obtained near θ'. By observing that:

$$\begin{aligned}
F_{jl}(\varphi(\theta')) &= F_{jl}(\varphi(\theta^0)) + \{(\overset{0}{\nabla}_{u.\varphi}\partial_i, \partial_j)|_{\varphi(\theta^0)} + (\overset{0}{\nabla}_{u.\varphi}\partial_j, \partial_i)|_{\varphi(\theta^0)}\}/n^{1/2} \\
&\quad + o(n^{-1/2}), \\
E_{\varphi(\theta')}T_n &= -\|u.\varphi\|\ |^2_{\varphi(\theta^0)}/2 - (\overset{-1/3}{\nabla}_{u.\varphi}u.\varphi, u.\varphi)|_{\varphi(\theta^0)}/2n^{1/2} + o(n^{-1/2})
\end{aligned}$$

and dropping the subscript θ^0 in RHS,

$$\begin{aligned}
E_{\varphi(\theta')}(T_n - E_{\varphi(\theta')}T_n)^2 &= \|u.\varphi\|^2 + (\overset{-1}{\nabla}_{u.\varphi}u.\varphi, u.\varphi)/n^{1/2} + o(n^{-1/2}), \\
E_{\varphi(\theta')}(T_n - E_{\varphi(\theta')}T_n)^3 &= (\overset{1}{\nabla} - \overset{-1}{\nabla}_{u.\varphi}u.\varphi, u.\varphi)/2n^{1/2} + o(n^{-1/2}).
\end{aligned}$$

It follows from Edgeworth expansion of T_n (cf. Akahira and Takeuchi (1981)) that

$$\begin{aligned} P_{\varphi(\theta^0)}(T_n \leq c_n) &= P_{\varphi(\theta^0)}(\{T_n - \|u.\varphi\|^2/2\}/\|u.\varphi\| \leq a_n) \\ &= N(a_n) - \eta(a_n)n^{-1/2}\{(\overset{1/3}{\nabla}_{u.\varphi}u.\varphi, u.\varphi)/2\|u.\varphi\| \\ &\quad + a_n(\overset{1}{\nabla}_{u.\varphi}u.\varphi, u.\varphi)/2\|u.\varphi\|^2 \\ &\quad + (a_n^2 - 1)(\overset{1}{\nabla} - \overset{-1}{\nabla}_{u.\varphi}u.\varphi, u.\varphi)/6\|u.\varphi\|^3\} + o(n^{-1/2}), \end{aligned}$$

where $a_n = \{c_n - \|u.\varphi\|^2/2\}/\|u.\varphi\|$. In order to obtain $P_{\varphi(\theta^0)}(T_n \leq c_n) = \alpha + o(n^{-1/2})$, a_n must be chosen such that:

$$\begin{aligned} a_n - N^{-1}(\alpha) = \{(\overset{1/3}{\nabla}_{u.\varphi}u.\varphi, u.\varphi)/2\|u.\varphi\| + N^{-1}(\alpha)(\overset{1}{\nabla}_{u.\varphi}u.\varphi, u.\varphi)/2\|u.\varphi\|^2 \\ + (N^{-1}(\alpha)^2 - 1)(\overset{1}{\nabla} - \overset{-1}{\nabla}_{u.\varphi}u.\varphi, u.\varphi)/6\|u.\varphi\|^3\} + o(n^{-1/2}). \end{aligned}$$

The same calculations applied to $P_{\varphi(\theta')}(T_n \leq c_n)$ conclude the proof. □

The first thing to observe is that ls1 are relatively easy to construct as can be seen in the following corollary:

COROLLARY 4.1. *Let ϕ be defined, in θ coordinate on a neighborhood of P_{θ^0} included in a submanifold of E, E' in E, by $\varphi(\theta) = A(\theta - \theta^0) + \theta^1$, then, if A is a fixed matrix in $\mathbb{R}^k$ partially unitary:*

$$^t x^t A g(-)|_{\theta^1} A x = {}^t x g(-)|_{\theta^0} x, \qquad \forall x \in T_{\theta^0}(E'),$$

ϕ is an ls1.

Notice that, here again, the word partially unitary is referring to the subspace where the two norms are equal. Let us observe that the ls2 also are forming a substantial set: let us give an example. For the sake of simplicity let us choose a system of coordinate θ such that $g(-)|_{\theta^0}$ is the identity matrix, and in this system, let us consider the following mapping:

$$\psi(\theta) = A(\theta - \theta^0) + \theta^0 + V(\theta_1 - \theta^0)^2/2$$

ψ is defined in a neighborhood of θ^0 in the following affine space:

$$T' = \left\{ \theta^0 + \lambda \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \lambda \in \mathbb{R} \right\}.$$

A is a matrix (a_{ij}) such that $a_{ij} = 0$ if $(i,j) \neq (1,1)$, $a_{11} = -1$, V is a vector of $\mathbb{R}^k$. ϕ is obviously an ls1. If $T_{111} = 0$, then ϕ is also an ls2 for $V^1 = -2\overset{0}{\Gamma}_{111}(\theta_0)$:

Relation (2.2′) with $u_1 = \lambda_1 \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$, $u_2 = \lambda_2 \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$, $u_3 = \lambda_3 \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$ writes:

$$\lambda_1\lambda_2\lambda_3\{(-\delta_1^s)(-\delta_1^s)(-\delta_1^m)\overset{\alpha}{\Gamma}_{smr}(\theta_0) + V^s(-\delta_1^r)g_{sr}(\theta_0)\} = \lambda_1\lambda_2\lambda_3\overset{\alpha}{\Gamma}_{111}(\theta_0)$$

(where δ^i_j is the Kronecker symbol) i.e. $\lambda_1\lambda_2\lambda_3\{-\overset{\alpha}{\Gamma}_{111}(\theta_0)-V^1\} = \lambda_1\lambda_2\lambda_3\overset{\alpha}{\Gamma}_{111}(\theta_0)$, which is verified for $T_{111} = 0$, and $V^1 = -2\Gamma_{111}(\theta_0)$.

5. Invariance under entropy morphisms in flat statistical manifolds

In this section, we consider (E, g, R) a flat statistical manifold (let h denote the associate entropy function) and investigate the invariance properties of metrics and connections under the associate entropy morphisms. It will be proven that the metric g and the connections $\overset{\alpha}{\nabla}$ are invariant under the entropy morphisms, but the converse is not true (i.e. even in simple cases there exists metrics and connections different from g and $\overset{\alpha}{\nabla}$ that are invariant).

Since in Section 3, we characterized the Markov morphisms of linear exponential families as the entropy morphisms of the associate likelihood statistical manifold, the statistical meaning of this section is important. Moreover, the examples and counterexamples we consider are always cases where (E, g, R) is the likelihood statistical manifold of a linear exponential family.

DEFINITION 8. A metric g' (resp. a connection ∇') on E is said to be entropy invariant if and only if it is invariant under every entropy morphism of the flat statistical manifold (E, g, R).

THEOREM 5.1. *In a flat statistical manifold (E, g, R), the metric g and every α-connection are entropy invariant.*

The proof is immediate, differentiating twice (resp. once again) the condition 2 of Definition 4. Moreover in certain cases the only entropy morphism is the identity. In those cases the result is not entirely surprising. We shall now investigate the converse: suppose that g' (resp. ∇') is some entropy invariant metric (resp. connection) on E, is it true that necessarily $g' = \lambda g$ for some constant $\lambda > 0$ (resp. $\nabla' = \overset{\alpha}{\nabla}$ for some constant α)? For this purpose, we investigate two examples: one where the property is true, a second where it is not.

Example 1. $(E = \mathbb{R}^k, g = F_0$, constant, $T = 0)$ is a flat manifold corresponding to the entropy function $h(\theta) = {}^t\theta F_0 \theta$. This example is, of course corresponding to the classical statistical experiment $E = \{N(\theta, F_0^{-1}), \theta \in \mathbb{R}^k\}$.

PROPOSITION 5.1. *If the entropy function is a quadratic form, then if g' (resp. ∇') is some entropy invariant metric (resp. connection) on E, necessarily $g' = \lambda g$ for some constant λ (resp. $\nabla = \overset{0}{\nabla}$).*

PROOF OF PROPOSITION 5.1. For the sake of simplicity, let us take $h(\theta) = \sum_{i=1}^k \theta_i^2$. In this case, it has been proven (Section 3) that the following φ defined on an affine space parallel to a subspace V of R^k by $\varphi(\theta) = A\theta + b$ with A such that $\forall x \in V\ {}^t x\, {}^t A A x = {}^t x x$ are Markov morphisms. Then choosing $\varphi(\theta) = A(\theta - \theta_0) + \theta_0$,

$V = \mathbb{R}^k$, A orthonormal leads to the fact that g' must verify, in θ coordinate:

$$ {}^tAg'_{\theta_0}(-)A = g'_{\theta_0}(-), \qquad \text{hence } g'_{\theta_0}(-) = \lambda(\theta_0)I_d, \qquad \text{for } \lambda(\theta_0) \in \mathbb{R}^+_*. $$

Choosing then $\varphi(\theta) = \theta - \theta_0 + \theta_1$ implies $\lambda(\theta_0) = \lambda(\theta_1)$. Let us now turn to the connection ∇. Let us observe that for those ϕ (2.2′) now writes:

$$ (2.2'') \qquad u_1^i u_2^j u_3^l a_i^s a_j^r a_l^m \Gamma_{srm}(\varphi(\theta)) = u_1^i u_2^j u_3^l \Gamma_{ijl}(\theta), \qquad \forall u_1, u_2, u_3 \in V. $$

For

$$ V = \left\{ \lambda \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \lambda \in \mathbb{R} \right\}, \qquad \varphi(\theta) = A(\theta - \theta_0) + \theta_0, \qquad a_1^1 = -1, \qquad a_i^j = 0 $$

for $(i,j) \neq (1,1)$. (2.2″) says that $\Gamma_{111}(\theta_0) = 0$. For

$$ V = \left\{ \lambda_1 \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \lambda_2 \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \lambda_1, \lambda_2 \in \mathbb{R} \right\}, \qquad \psi(\theta) = A(\theta - \theta_0) + \theta_0, $$
$$ a_1^2 = a_2^1 = \varepsilon, \qquad a_i^j = 0 $$

for $(i,j) \neq (1,2)$ or $(2,1)$ $(\varepsilon = \pm 1)$, (2.2″) with $u_1 = u_2 = u_3$ says that $\Gamma_{211}(\theta_0) = \varepsilon\Gamma_{221}(\theta_0) = 0$. For

$$ V = \left\{ \lambda \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \lambda \in \mathbb{R} \right\}, \qquad \varphi(\theta) = A(\theta - \theta_0) + \theta_0, $$
$$ a_1^1 = -1, \qquad a_2^2 = 1, \qquad a_3^3 = 1, \qquad a_i^j = 0 $$

for $(i,j) \neq (1,1), (2,2), (3,3)$, (2.2″) with $u_1 = u_2 = u_3$ says that $\Gamma_{123}(\theta_0) = 0$. □

Example 2. $(E = \mathbb{R} \times \mathbb{R}^+_*, h(\theta_1, \theta_2) = 1/2[-\log 2\pi\theta_2 + \theta_1^2/\theta_2])$ This example is corresponding to the statistical family:

$$ E = \{N(\mu, \sigma^2), (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+_*\}, \qquad \theta_1 = \mu/\sigma^2, \qquad \theta_2 = 1/\sigma^2. $$

PROPOSITION 5.2. *In the present case of statistical manifold, the set of entropy invariant metric is the following*: *(in $\theta = (\theta_1, \theta_2)$ coordinates)*

$$ g'_{11} = \gamma/\theta_2, \qquad g'_{12} = g'_{21} = -\gamma\theta_1/\theta_2^2, \qquad g'_{22} = \gamma\theta_1^2/\theta_2^3 + \delta/\theta_2^2 $$

$(g'_{ij} = g'(\partial_i, \partial_j)|_{(\theta_1,\theta_2)})$, *for arbitrary positive constants γ and δ.*

PROOF OF PROPOSITION 5.2. Using Proposition 3.3 of Section 3, one obtains that, in the present case, the entropy morphisms are forming the following class:

$$\{\varphi_1^{cd}(\theta_1,\theta_2) = c\theta_1 + d\theta_2, \varphi_2^{cd}(\theta_1,\theta_2) = c^2\theta_2, c \in \mathbb{R}_*, d \in \mathbb{R}\}.$$

Let us denote by $g_{ij}(\theta_1,\theta_2)$ the quantity $g'(\partial_i,\partial_j)|_{(\theta_1,\theta_2)}$. First, let us remark that each point $(0,\theta_2)$ is invariant under both φ^{+10} and φ^{-10}, this implies using (2.1′) (Section 2) that $g_{ij}(0,\theta_2) = 0$ for $i \neq j$. Since we also have that $\varphi^{c0}(0,1) = (0,c^2)$, we find that, if we denote by τ and δ the constants $g_{11}(0,1)$ and $g_{22}(0,1)$ respectively, then:

$$g_{11}(0,\theta_2) = \tau/\theta_2 \qquad \text{and} \qquad g_{22}(0,\theta_2) = \delta/\theta_2^2.$$

In addition, as the point (θ_1,θ_2) is invariant under φ^{-1d} for $d = 2\theta_1/\theta_2$, we have that $-g_{11}(\theta_1,\theta_2)(\theta_1/\theta_2) = g_{12}(\theta_1,\theta_2)$. Moreover, since for $c \neq 0$, $\varphi^{cd}(\theta_1,\theta_2) = (0,c^2\theta_2)$ if $d = -c\theta_1/\theta_2$, we must have $g_{12}(\theta_1,\theta_2) = -\tau\theta_1/\theta_2^2$ and $g_{22}(\theta_1,\theta_2) = \tau\theta_1^2/\theta_2^3 + \delta/\theta_2^2$. Therefore, it remains to be shown that the g chosen as indicated satisfies:

$${}^tD\varphi^{cd} g(-1)|_{\varphi^{cd}(\theta_1,\theta_2)}\, D\varphi^{cd} = g(-)|_{(\theta_1,\theta_2)}$$

for every admissible pair (c,d), (θ_1,θ_2). This, however, does not present any difficulty and the proof is complete. □

Remarks.

1) The Fisher metric g corresponds to $\tau = 1$, $\delta = 1/2$.

2) We did not investigate, the class of invariant connections in this case. Let us just mention that the Riemannian connections associated with the previous invariant metrics are invariant and it is not very difficult to see that these connections do not coincide with the Amari connections, which are invariant too.

3) Although we have no intuitive interpretation of this result, it could be useful to notice that the set of invariant metrics we obtain in this case is, exactly the set of Fisher metric g_F on the univariate elliptic model $E = \{F((\cdot - \mu)/\sigma), (\mu,\sigma) \in \mathbb{R} \times \mathbb{R}^+\}$, where F is varying in the set of all the probability distributions (cf. Mitchell (1986)).

Acknowledgements

The author would like to thank Professor S. I. Amari and S. L. Lauritzen for a lot of helpful discussions and very encouraging remarks and anonymous referees for their valuable comments on earlier versions.

REFERENCES

Akahira, M. and Takeuchi, K. (1981). Asymptotic efficiency of statistical estimators: concepts and higher order asymptotic efficiency, *Lecture Notes in Statist.*, **7**, Springer, New York.

Amari, S. I. (1982). Differential geometry of curved exponential families—curvature and information loss, *Ann. Statist.*, **10**, 357–385.

Amari, S. I. (1983). Comparison of asymptotically efficient tests in terms of geometry of statistical structures, *Bulletin of the International Statistical Institute (Proc. 44 Session)*, **50**, Book 2, 1.190–1.206.

Amari, S. I. (1985). Differential-geometrical methods in statistics, *Lecture Notes in Statist.*, **28**, Springer, New York.

Amari, S. I. (1987). Differential geometrical theory of statistics, *Differential Geometry in Statistical Inference*, IMS Lecture Notes-Monograph Series, Vol. 10, Hayward, California.

Amari, S. I. and Kumon, M. (1983). Differential geometry of Edgeworth expansions in curved exponential family, *Ann. Inst. Statist. Math.*, **35**, 1–24.

Barndorff-Nielsen, O. E. (1987). Differential and integral geometry in statistical inference, *Differential Geometry in Statistical Inference*, IMS Lecture Notes-Monograph Series, Vol. 10, Hayward, California.

Barndorff-Nielsen, O. E. and Jupp, P. (1984). Differential geometry profile likelihood and L. sufficiency, Research Report, 118, Department of Theoretical Statistics, Aarhus University.

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

Boothby, W. M. (1975). *An Introduction to Differentiable Manifolds and Riemannian Geometry*, Academic Press, New York.

Chentsov, N. N. (1972). *Statistical Decision Rules and Optimal Inference*, Nauka, Moscow (in Russian). (English translation (1982). *Translation of the Mathematical Monographs*, **53**, American Mathematical Society, Providence, Rhode Island.)

Efron, B. (1975). Defining the curvature of a statistical problem (with application to second order deficiency), *Ann. Statist.*, **3**, 1189–1242.

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

Lauritzen, S. L. (1984). Statistical manifolds, *IMS Monograph on "Differential Geometry in Statistical Inference"* (ed. A. Kass).

Lauritzen, S. L. (1987). Conjugate connections in statistical theory, *Proceedings of the GST Workshop*, Lancaster.

Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory*, Springer, New York.

Millar, L. (1983). Minimax principle in asymptotic theory, Ecole d'été de St Flour, *Lecture notes*, 976, 76–262, Springer, Berlin.

Mitchell, A. F. S. (1986). Statistical manifold of univariate elliptic distributions, Research Report, R 86.15, Institute of Electroniske Systemer, Aalborg.

Rao, C. R. (1945). On the distance between two populations, *Sankhyā*, **9**, 246–248.

Spivak, M. (1979). *A Comprehensive Introduction to Differential Geometry*, Publish or Perish, Houston.

Torgersen, E. N. (1970). Comparison of experiments when the parameter space is finite, *Z. Wahrsch. Verw. Gebiete*, **16**, 219–249.

Wolf (1967). *Spaces of Constant Curvature*, McGraw-Hill, New York.