Vol.9.Issue.2.2021 (April-June)
©KY PUBLICATIONS

http://www.bomsr.com
Email:editorbomsr@gmail.com

RESEARCH ARTICLE

BULLETIN OF MATHEMATICS AND STATISTICS RESEARCH

A Peer Reviewed International Research Journal

INTERNATIONAL STANDARD SERIAL NUMBER 2348-0580

ON THE GEOMETRIC STRUCTURES OF THE STATISTICAL MANIFOLDS

ABOUBACAR NIBIRANTIZA[1], GILBERT NIBARUTA[2], DOMITIEN NDAYIRUKIYE[3], MENEDORE KARIMUMURYANGO[4]

[1]University of Burundi, Institute for Applied Pedagogy, Departement of Mathematics, Bujumbura-Burundi, Email:aboubacar.nibirantiza@ub.edu.bi

[2]Ecole Normale Supérieure, Departement des Sciences naturelles, section de Mathématiques, Bujumbura-Burundi, Email:nibarutagilbert@gmail.com

[3]Ecole Normale Supérieure, Departement des Sciences naturelles, section de Mathématiques, Bujumbura-Burundi, Email:domitienndayi@yahoo.fr

[4]University of Burundi, Institute for Applied statistics, Bujumbura-Burundi, Email:mene- dore.karimumuryango@ub.edu.bi

DOI:10.33329/bomsr.9.2.56

ABSTRACT

In this paper,wecompute the invariantgeometry of the statistical manifoldof the normal distributions by using the divergence approach. In order to give the isometries, we investigate the component functions of a Killing vector eld for the Fisher information metric on the statistical manifold of Gaussian distributions of probabilities. They are harmonic conjugate and then, they are both harmonic functions. Finally, we describe the form of its Killing vector elds.

Keywords and phrases:Statistical manifold, Fisher information metric,f -divergence, conjugate connections.

1 Introduction

Information geometry is the study of intrinsic properties of manifolds of probability distributions, called statistical manifold, by way of di erential geometry. In an another words, it provides a geometric approach to families of statistical models. Explicitly, many important structures in information theory and statistic can be treated as struc- tures in di erential geometry by regarding a space of probabilties as a di erential man- ifold endowed with a Riemannian metric and a familly of a ne connections. The key geometric structures are the Fisher quadratic form, called Ficher information metric and the Amari–Chentsov tensor. A statistical manifold is a generalization of a Riemannian manifold with the Levi-Civita connection. On a statistical manifold, duality of affine connections naturally arises [2]. A statistical manifold is simply a Riemannian manifold (M, g) with one additional structure given by a torsion-free affine connection $\bigtriangledown$ and its dual connection $\bigtriangledown^*$, which is also assumed to be torsion-free. We say $\bigtriangledown$ and $\bigtriangledown^*$ are mutually dual whenever

$$dg(X, Y) = g(\bigtriangledown X, Y) + g(X, \bigtriangledown^* Y)$$

holds for all vector fields X, Y on M. Thus the geometry of statistical manifold simply reduces to usual Riemannian geometry when $\bigtriangledown$ coincides with $\bigtriangledown^*$.

In statistics, the notion of sufficient statistic expresses the criterion for passing from one model to another without loss of information. This leads to the question how the geometric structures behave under such sufficient statistics.

In the present paper, we show that when we define the Kullback-Leibler divergence on the statistical manifold we can also computing the geometry of it and show that the metric, connections and curvature are invariant when we use the metric induced by the Kullback-Leibler divergence instead of the usual metric. Secondly, we compute that the Killing fields on the statistical manifold of normal distributions of probabilities are the infinitesimal generators of isometries; that is, flows generated by Killing fields are continuous isometries of the manifold. More simply, the flow generates a symmetry, in the sense that moving each point on an object the same distance in the direction of the Killing vector field will not distort distances on the object.

2 Notation and problem setting

2.1 Statistical manifold

Let $(X, \mathcal{B})$ be a measurable space, where X is a non-empty subset of $\mathbb{R}$ and $\mathcal{B}$ is the σ-field of subsets of X. Consider a n-dimensional statistical manifold

$$M = \{p(x; \xi) : \xi = [\xi_1, ..., \xi_n] \in E \subseteq \mathbb{R}^n\},$$

with coordinates $[\xi_1, ..., \xi_n]$, defined on X. Then M is a subset of $P(X)$, the set of all probability measures on X given by

$$P(X) := \{p : X \to \mathbb{R} : p(x) > 0 \ (x \in X); \int_X p(x)dx = 1\}. \tag{2.1}$$

In another words, M is parameterized by ξ and the set $M = \{p(x, \xi)\}$ forms a manifold with coordinate system ξ. Here, x may take discrete, continuous and vector values.

2.2 Divergence

Let us consider two points $p(x, \xi)$ and $Q(x, \xi')$ in a manifold S, of which coordinates are ξ and ξ'.

Definition 2.1. *A divergence $D(\xi : \xi')$ of two points $P(x, \xi)$ and $Q(x, \xi')$ is a differentiable function of ξ and ξ' which satisfies the following criteria:*

1. *$D(\xi : \xi') \geq 0$,*
2. *$D(\xi : \xi') = 0$ if and only if $\xi = \xi'$,*
3. *When P and Q are sufficiently close, by denoting their coordinates by ξ and $\xi' = \xi + d\xi$, the Taylor expansion of D is written as*

$$D(\xi : \xi + d\xi) = \frac{1}{2} \sum g_{ij}(\xi) d\xi_i d\xi_j + \mathcal{O}(|d\xi|^3) \tag{2.2}$$

and matrix $G = g_{ij}$ is positive-definite, depending on ξ .

We can also denote the divergence by $D(p : q)$ instead of $D(\xi : \xi')$.

A divergence represents a degree of separation of two points $P(x, \xi)$ and $Q(x, \xi')$, but it or its square root is not a distance. It does not necessarily satisfy the symmetry condition,so that in general $D(\xi : \xi') \neq D(\xi' : \xi)$. We may call $D(\xi : \xi')$ divergence from $P(x, \xi)$ to $Q(x, \xi')$. Moreover, the triangular inequality does not hold. It has the dimension of the square of distance, as is suggested by (2.2). It is possible to symmetrize a divergence by

$$D_S(\xi : \xi') = \frac{1}{2}(D(\xi : \xi') + D(\xi' : \xi)).$$

However, the asymmetry of divergence plays an important role in information geometry. When $P(x, \xi)$ and $Q(x, \xi')$ are sufficiently close, we define the square of an infinitesimal distance ds between them by using (2.2) as

$$ds^2 = 2D[\xi : \xi + d\xi] = g_{ij} d\xi_i d\xi_j. \tag{2.3}$$

A manifold M is said to be Riemannian when a positive-definite matrix $G(\xi)$ is defined on M and the square of the local distance between two nearby points ξ and $\xi + d\xi$ is given by (2.3). A divergence D provides M with a Riemannian structure.

2.3 Information monotonicity

We consider that the divergence $D(\xi : \xi')$ between $P(x,\xi)$ and $Q(x,\xi')$ changes to $\overline{D}(\xi : \xi')$ between $\overline{P}(y,\xi)$ and $\overline{Q}(y,\xi')$. Since the divergence $D(\xi : \xi')$ represents the dissimilarity of two distributions $P(x,\xi)$ and $Q(x,\xi')$, it is postulated that it decreases in general by this mapping,

$$\overline{D}(\xi : \xi') \leq D(\xi : \xi'). \tag{2.4}$$

The relation ((2.4)) is called information monotonicity.

Obviously, when the function k is one-to-one, that is invertible, there is no loss of information and the equality is required to hold in (2.4). However, there is a case when information is not lost even when k is not invertible. This is the case when x includes a redundant part, the distribution of which does not depend on ξ. We may abandon this part without losing information concerning ξ. The remaining part retains full information.

Definition 2.2. *A function*

$$s = k(x) \tag{2.5}$$

is called a sufficient statistic when the probability density function $P(x,\xi)$ is decomposed as

$$P(x,\xi) = \overline{P}(s,\xi)r(x). \tag{2.6}$$

This implies that the probability $P(x,\xi)$ is written as a function of s, except for a multiplicative term $r(x)$ which does not depend on ξ. The equality is required to hold in (2.4) when and only when y is a sufficient statistic. We give now the invariance criterion. As formulated by Amari and Nagaoka (2000), we have defintion.

Definition 2.3. *A geometrical structure of M is invariant when it satisfies the monotonicity* (2.4), *where the equality holds if and only if $y = k(x)$ is a sufficient statistic.*

2.4 Invariance of the geometric structures

In this section we consider the invariance properties of the geometric structures under suitable transformations of the variables in a statistical manifold. Here we can consider two kinds of invariance of the geometric structures; covariance under re-parametrization of the parameter of the manifold and the invariance under the transformations of the random variable.

2.4.1 Covariance under re-parametrization

Let $[\theta^i]$ and $[\eta_j]$ be two coordinate systems on the statistical model S, which are related by an invertible transformation $\eta = \eta(\theta)$. Let us denote $\partial_i = \frac{\partial}{\partial\theta^i}$ and $\partial^j = \frac{\partial}{\partial\eta_j}$. Let the coordinate expressions of the metric g be given by $g_{ij} =< \partial_i, \partial_j >$ and $\tilde{g}_{ij} =< \partial^i, \partial^j >$. Since $\partial^i = \sum_m \frac{\partial\theta^m}{\partial\eta_i}\frac{\partial}{\partial\eta_i}$ and $\partial^j = \sum_n \frac{\partial\theta^n}{\partial\eta_j}\frac{\partial}{\partial\eta_j}$, then the computation of the metric $\tilde{g}_{ij}$ induced by the re-parametrization implies the covariance formula.

Proposition 2.4. *The Fisher information metric g is covariant under re-parametrization and it is given by*

$$\tilde{g}_{ij} = \sum_m \sum_n \frac{\partial\theta^m}{\partial\eta_i}\frac{\partial\theta^n}{\partial\eta_j} g_{mn}.$$

.

Proof. The components of the Fisher information metric with respect to the coordinate system $[\theta^i]$ are given by

$$g_{ij} =< \partial_i, \partial_j >= E[\partial_i l(x,\theta)\partial_j l(x,\theta] = \int \partial_i l(x,\theta)\partial_j l(x,\theta)p(x,\theta)dx$$

Let $\bar{p}(x,\eta) = p(x,\theta(\eta))$. Since $\partial^i l(x,\eta) = \sum_m \frac{\partial\theta^m}{\partial\eta_i}\frac{\partial l(x,\theta)}{\partial\theta^m}$ we can compute that

$$\begin{aligned}
\tilde{g}_{ij}(\eta) =< \partial^i, \partial^j > &= \int \partial^i l(x,\theta)\partial^j l(x,\theta)\bar{p}(x,\theta)dx \\
&= \left[\sum_m \sum_n \frac{\partial\theta^m}{\partial\eta_i}\frac{\partial\theta^n}{\partial\eta_j}\right] \int \partial_m l(x,\theta(\eta))\partial_n l(x,\theta(\eta))p(x,\theta(\eta))dx \\
&= \left[\sum_m \sum_n \frac{\partial\theta^m}{\partial\eta_i}\frac{\partial\theta^n}{\partial\eta_j} g_{mn}(\theta)\right]_{\theta=\theta(\eta)}
\end{aligned}$$

□

2.4.2 Invariance under the transformation of the random variable

We define the invariance of Riemannian metric on a statistical manifold under a transformation of the random variable as follows. Let $S = \{\bar{q}(x;\xi) \mid \xi \in E \subseteq \mathbb{R}^n\}$ be a statistical manifold defined on a sample space X. Let x, y be random variables defined on sample spaces X, Y respectively and ϕ be a transformation of x to y. Assume that this transformation induces a model $S' = \{q(y;\xi) \mid \xi \in E \subseteq \mathbb{R}^n\}$ on Y. Then the density function $q(y;\xi)$ of the induced model S' takes the form [4],

$$q(y,\xi) = \bar{q}(\omega(y),\xi)\omega'(y) \tag{2.7}$$

where ω is a function such that $x = \omega(y)$ (i.e.,$y = \phi(x)$) and $\phi'(x) = \frac{1}{\omega'(\phi(x))}$.

Proposition 2.5. *The Fisher information metric is invariant under smooth one-to-one transformations of the random variable.*

Proof. Let us denote $\log q(x,\xi)$ by $l(q_y)$ and $\log p(x,\xi)$ by $l(p_x)$. From the equation (2.7) , we can write,

$$\bar{q}(x,\xi) = q(\phi(x),\xi)\phi'(x),$$
$$\partial_i l(q_y) = \partial_i l(\bar{q}_{\omega(y)}),$$
$$\partial_i l(q_{\phi(x)}) = \partial_i l(\bar{q}_x).$$

The Fisher information metric g' on the induced manifold S' is given by

$$\begin{aligned} g'_{ij}(q_\xi) &= \int_Y \partial_i l(q_y)\partial_j l(q_y) q(y,\xi)dy \\ &= \int_X \partial_i l(q_{\phi(x)})\partial_j l(q_{\phi(x)}) q(\phi(x),\xi)\phi'(x)dx \\ &= \int_X \partial_i l(\bar{q}_x)\partial_j l(\bar{q}_x)\bar{q}(x,\xi)dx \\ &= g_{ij}(\bar{q}_\xi) \end{aligned}$$

□

3 Geometry of statistical manifold

Let S be an n-dimensional statistical manifold. The tangent space to S at a point p_ξ is given by

$$T_\xi(S) = \{\sum_{i=1}^n \alpha^i \partial_i \mid \alpha^i \in \mathbb{R}\}.$$

Let define $l(x,\xi) = \log p(x;\xi)$ and consider the random variable $A(x)$ given by the partial derivatives $A(x) = \frac{\partial l(x,\xi)}{\partial \xi^i} =: \partial_i l(x,\xi), i = 1,\dots,n$. For the statistical manifold S, $\partial_i l(x,\xi)$'s are linearly independent functions in x for a fixed ξ. Let $T^1_\xi(S)$ be the n-dimensional vector space spanned by n functions $\{\partial_i l(x,\xi), i = 1,\dots,n\}$ in x. So

$$T^1_\xi(S) = \{A^i \partial_i l(x,\xi) \mid A^i \in \mathbb{R}\}.$$

There is a natural isomorphism [4] between these two vector spaces $T_\xi(S)$ and $T^1_\xi(S)$ given by

$$\partial_i \in T_\xi(S) \longleftrightarrow \partial_i l(x,\xi) \in T^1_\xi(S).$$

Obviously, a tangent vector $A = \sum_{i=1}^n A^i \partial_i \in T_\xi(S)$ corresponds to a random variable $A(x) = \sum_{i=1}^n A^i \partial_i l(x,\xi) \in T^1_\xi(S)$ having the same components A^i. Note that $T_\xi(S)$ is the differentiation operator representation of the tangent space, while $T^1_\xi(S)$ is the random variable representation of the same tangent space.The space $T^1_\xi(S)$ is called

[4] the 1-representation of the tangent space. Let A and B be two tangent vectors in $T_\xi(S)$ and $A(x)$ and $B(x)$ be the 1-representations of A and B respectively.

We define the inner product by

$$g_\xi(A,B) =< A,B >_\xi= E_\xi[A(x)B(x)] = \int A(x)B(x)p(x;\xi)dx. \tag{3.1}$$

Especially the inner product of the basis vectors ∂_i and ∂_j is given by

$$g_{ij}(\xi) =< \partial_i, \partial_j >= E_\xi[\partial_i l(x,\xi)\partial_j l(x,\xi)] = \int \partial_i l(x;\xi)\partial_j l(x;\xi)p(x;\xi)dx. \tag{3.2}$$

Note that $g =< .,. >$ defines a Riemannian metric on S called the Fisher information metric. On the Riemannian manifold $(S, g =< .,. >)$, we define n^3 functions Γ_{ijk} by

$$\Gamma_{ijk}(\xi) = E_\xi[(\partial_i\partial_j l(x;\xi))(\partial_k l(x;\xi))]. \tag{3.3}$$

and the tensor T given by

$$T_{ijk} = E[\partial_i l(x,\xi)\partial_j l(x,\xi)\partial_k l(x,\xi)]. \tag{3.4}$$

The functions Γ_{ijk} uniquely determine an affine connection ∇ on S by

$$\Gamma_{ijk}(\xi) =< \nabla_{\partial_i}\partial_j, \partial_k >_\xi . \tag{3.5}$$

Now, in order to compute the geometry on the statistical manifold of Gaussian distributions of probabilities

$$M = \{p(x,\mu,\sigma) := \frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{(x-\mu)^2}{2\sigma^2}), -\infty < \mu < +\infty,\ 0 < \sigma < +\infty\},$$

let give the following lemma.

Lemma 3.1. *If* $\Gamma(x) = \int_0^{+\infty} t^{x-1}e^{-t}dt, \forall x \in \mathbb{R}_0^+$ *and* $\mu \in \mathbb{R}$, *then*

$$E[(x-\mu)^k] = \begin{cases} (2^k)^{\frac{1}{2}}\frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{1}{2})}\sigma^k, & \textit{if} \quad k \textit{ is even} \\ 0 \textit{ if} \quad k \textit{ is odd.} \end{cases} \tag{3.6}$$

Proof. The proof of the lemma consists in computing the formula

$$E[(x-\mu)^k] = \int_{-\infty}^{+\infty} (x-\mu)^k \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx.$$

By setting $t = \frac{x-\mu}{\sqrt{2}\sigma}$ and $dx = \sqrt{2}\sigma dt$, we obtain

$$E[(x-\mu)^k] = 2\frac{\sqrt{2^k}\sigma^k}{\sqrt{\pi}} \int_0^{+\infty} t^k e^{-t^2} dt.$$

If k is odd, then the factor $\int_0^{+\infty} t^k e^{-t^2} dt$ is zero whereas if k is even and if we set $u = t^2$, this implies that $u^{\frac{k}{2}} = t^k$ and the computation shows that

$$\int_0^{+\infty} t^k e^{-t^2} dt = \frac{1}{2}\Gamma(\frac{k+1}{2}).$$

By replacing $\sqrt{\pi}$ by $\Gamma(\frac{1}{2})$ we get the result of the lemma 3.1. □

Proposition 3.2. *Let M be the statistical manifold of the normal distributions.*

1. *The Fisher metric is given by the matrix*

$$G = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{pmatrix}. \tag{3.7}$$

2. *The tensors T_{ijk} are given by*

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

and

$$T_{112} = T_{121} = T_{211} = \frac{2}{\sigma^3}, \quad \textit{whereas } T_{222} = \frac{8}{\sigma^3}.$$

Proof. To prove the first part of the proposition, we set

$$l(x,\xi) = \log p(x,\xi) = -\log\sqrt{2\pi} - \log\sigma - \frac{(x-\mu)^2}{2\sigma^2}.$$

We must compute the tensor g_{ij} by using the formulas $g_{11} = E[(\partial_\mu l(x,\xi)\partial_\mu l(x,\xi))]$ and $g_{22} = E[(\partial_\sigma l(x,\xi)\partial_\sigma l(x,\xi))]$ and $g_{12} = E[(\partial_\mu l(x,\xi)\partial_\sigma l(x,\xi))]$ and $g_{21} = E[(\partial_\sigma l(x,\xi)\partial_\mu l(x,\xi))]$. The lemma 3.1 leads us to get the results. To prove the second part, we use the formula $T_{ijk} = E[\partial_{\sigma_i} l(x,\xi)\partial_{\sigma_j} l(x,\xi)\partial_{\sigma_k} l(x,\xi)]$. Applying the lemma 3.1 we obtain the results. □

Using the previous results, we compute the geometry of statistical manifold of the normal distributions. We consider the metric induced by the Kullback-Leibler divergence $D_{KL}(p,q)$ and we prove the invariance of the geometry.

Proposition 3.3. *The Kullback-Leibler divergence from the density p of a normal random variable with mean μ_1 and variance σ_1^2 and the density q of a normal random variable with mean μ_2 and variance σ_2^2 is given by*

$$D_{KL}(p,q) = \log(\frac{\sigma_2}{\sigma_1}) + \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2} - \frac{1}{2}.$$

Proof. We must compute the integral

$$D_{KL}(p,q) := \int p(x) \log\left(\frac{p(x)}{q(x)}\right) dx.$$

It is obvious that this formula gives

$$D_{KL}(p,q) = \int (\log p(x) - \log q(x))\, p(x) dx.$$

Replacing the expression of $p(x)$ and $q(x)$ and computing, we obtain

$$D_{KL}(p,q) = \int \left(\log(\frac{\sigma_2}{\sigma_1}) + \frac{1}{2}\left(\frac{(x-\mu_2)^2}{\sigma_2^2} - \frac{(x-\mu_1)^2}{\sigma_1^2}\right)\right) \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{1}{2}\left(\frac{(x-\mu_1)^2}{\sigma_1^2}\right)\right) dx.$$

This expression equals to the following

$$E_1\left(\log(\frac{\sigma_2}{\sigma_1}) + \frac{1}{2}\left(\left(\frac{x-\mu_2}{\sigma_2}\right)^2 - \left(\frac{x-\mu_1}{\sigma_1}\right)^2\right)\right).$$

Using the fact that $E_1[(x-\mu_1)^2] = \sigma_1^2$, this expression can be written as follow

$$D_{KL}(p,q) = \log(\frac{\sigma_2}{\sigma_1}) + \frac{1}{2\sigma_2^2} E_1[(x-\mu_2)^2] - \frac{1}{2}. \tag{3.8}$$

Now, note that $(X-\mu_2)^2 = (x - \mu_1 + \mu_1 - \mu_2)^2$ and so

$$(X-\mu_2)^2 = (X-\mu_1)^2 + 2(X-\mu_1)(\mu_1-\mu_2) + (\mu_1-\mu_2)^2.$$

Putting this expression in the (3.8), we obtain

$$D_{KL}(p,q) = \log\left(\frac{\sigma_2}{\sigma_1}\right) + \frac{\sigma_1^2 + (\mu_1-\mu_2)^2}{2\sigma_2^2} - \frac{1}{2}.$$

□

Remark 3.4. *It is obvious that $D_{KL}(p,q) = 0$ if and only if $p = q$. In another words, when $\mu_1 = \mu_2$ and $\sigma_1 = \sigma_2$.*

Definition 3.5. *When the local coordinates $x = (x_1, \ldots, x_n)$ are defined for p and $y = (y_1, \ldots, y_n)$ for q on the n-dimensional statistical manifold, the following quantities are defined by derivating the divergence $D(x : y)$:*

$$D(\partial_i \partial_j : \partial_k, x) = \frac{\partial^3}{\partial x_i \partial x_j \partial y_k} D(x : y)|_{y=x}.$$

The following properties follow.

(*i*) $D(\partial_i : \cdot, x) = \frac{\partial}{\partial x_i} D(x : y)|_{y=x} = 0;$

(*ii*) $D(\cdot : \partial_j, x) = \frac{\partial}{\partial y_j} D(x : y)|_{y=x} = 0;$

(*iii*) $g^D = (g^D_{ij})$ where $g^D_{ij} = D(\partial_i \partial_j : \cdot, x) = D(\cdot : \partial_i \partial_j, x) = -D(\partial_i : \partial_j, x).$

To prove these properties, we consider that $y - x$ is smallest and considering that $D(x : x + dx) = \sum g_{ij}(x) dx^i dx^j + \mathrm{O}(|dx|^3)$, we have

$$D(x : y) = \sum g_{ij}(x)(x_i - y_i)(x_j - y_j) + \mathrm{O}(|x - y|^3) \tag{3.9}$$

Derivating the equation (3.9) with respect to x_i and (or) y_j and evaluating the result at $y = x$, we obtain the formulas.

Proposition 3.6. *The induced metric $g^D =<,>^D$ on the statistical manifold of normal distribution of probabilities S is the Fisher information metric $g =<,>$.*

Proof. The computation of the metric g^D_{ij} induced by the Kullback-Leibler divergence refers to the proposition 3.3 and the properties of the definition 3.5. We obtain

$$\begin{aligned} g^D_{11} &= \frac{\partial^2}{\partial \sigma_1 \partial \sigma_1} \left(\log\left(\frac{\sigma_2}{\sigma_1}\right) + \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2} - \frac{1}{2} \right) \\ &= \frac{1}{\sigma_1^2} \end{aligned}$$

and

$$\begin{aligned} g^D_{22} &= \frac{\partial^2}{\partial \sigma_2 \partial \sigma_2} \left(\log\left(\frac{\sigma_2}{\sigma_1}\right) + \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2} - \frac{1}{2} \right) \\ &= \frac{2}{\sigma_1^2} \end{aligned}$$

whereas g^D_{12} and g^D_{21} are zero. It is clear that these results are the same as in the first part of the proposition 3.2. □

Definition 3.7. *When one divergence D takes the form $D_f(p:q) = \int p_i f(\frac{q_i}{p_i})dx$ where f is a differentiable convex function which satisfies the condition $f(1) = 0$ then the divergence is called an f-divergence.*

It follows that the Kullback-Leibler divergence is an f-divergence because it is obvious that $f = -\log$.

4 α-geometry of statistical manifold

When a divergence D is defined in the statistical manifold, then two tensors g_{ij}^D and T_{ijk}^D are automatically induced from it.

4.1 Dual connections

Let us define the quantities $T_{ijk} = \Gamma^*_{ijk} - \Gamma_{ijk}$. Then the dual connections on the statistical manifold are written as

$$\Gamma_{ijk} = \Gamma^0_{ijk} - \frac{1}{2}T_{ijk},\ \Gamma^*_{ijk} = \Gamma^0_{ijk} + \frac{1}{2}T_{ijk}.$$

We can so define a parametric family of torsion-free connections $\triangledown^{(\alpha)}$ indexed by α ($\alpha \in \mathbb{R}$) by

$$\triangledown^{(\alpha)} = \frac{1+\alpha}{2}\triangledown + \frac{1-\alpha}{2}\triangledown^*,$$

where $\triangledown^{(0)}$ denoting the Levi-Civita connection associated with g.

The α-connection is defined by

$$\Gamma^{(\alpha)}_{ijk} = E\left[\left(\partial_i\partial_j l + \frac{1-\alpha}{2}\partial_i l \partial_j l\right)\partial_k l\right]. \tag{4.1}$$

This formula implies that $\Gamma^{(1)}_{ijk} = E[\partial_i\partial_j\partial_k l]$ and that

$$\Gamma^{(\alpha)}_{ijk} = \Gamma^{(1)}_{ijk} + \frac{1-\alpha}{2}T_{ijk}. \tag{4.2}$$

Let $\Gamma^{(\alpha)k} = g^{km}\Gamma^{(\alpha)}_{ijm}$. The α-curvature tensor is defined by

$$R^{(\alpha)}(\partial_i, \partial_j)\partial_k = \sum_l R^{(\alpha)}_{ijk}\partial_l$$

where

$$R^{(\alpha)l}_{ijk} = \partial_i\Gamma^{(\alpha)l}_{jk} - \partial_j\Gamma^{(\alpha)l}_{ik} + \sum_{m=1}^{n}\Gamma^{(\alpha)l}_{im}\Gamma^{(\alpha)m}_{jk} - \sum_{m=1}^{n}\Gamma^{(\alpha)l}_{jm}\Gamma^{(\alpha)m}_{ik}. \tag{4.3}$$

We define also that

$$R^{(\alpha)}_{ijkl} = \sum_{t=1}^{n} R^{(\alpha)t}_{ijk} g_{tl}.$$

Definition 4.1. *A statistical manifold is said to be α-flat if its α-curvature vanishes.*

Note also that the 0-geometry corresponds to the geometry of the Riemannian metric. The Levi-Civita connection is

$$\Gamma^{(0)}_{ijk} = \frac{1}{2}\left(\partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij}\right). \tag{4.4}$$

4.2 Dual connections derived from KL-divergence

The following is a key result connecting a divergence and dual geometry. In order to compute the affine dual connections induced by the Kullback-Leibler divergence on the statistical manifold of normal distributions of probabilities, let us begin with the following description.

Definition 4.2. *On statistical manifold, the two quantities $\Gamma^D_{ijk} = -D(\partial_i\partial_j : \partial_k)$ and $\Gamma^{D*}_{ijk} = -D(\partial_k : \partial_i\partial_j)$ are called dual affine connections with respect to the metric g^D induced by the Kullback-Leibler divergence.*

We define also [2] the tensor T^D_{ijk} and g^D_{ij} induced by the divergence D_{KL} by

$$T^D_{ijk} = -D(\partial_k : \partial_i\partial_j, x) + D(\partial_i\partial_j : \partial_k, x). \tag{4.5}$$

and

$$g^D_{ij} = D(\partial_i : \partial_j, x) \tag{4.6}$$

respectively. On the statistical manifold of the normal distributions of probabilities it is clear that to differentiate g^D_{ij} with respect σ and μ, we get the duality condition

$$\partial_k g^D_{ij} = \Gamma^D_{ijk} + \Gamma^{D*}_{ijk}.$$

We prove in following that the two tensors G^D and T^D derived from the KL-divergence are therefore invariant.

Theorem 4.3. *The invariant tensors derived from the Kullback-Leibler divergence in the manifold of normal distributions are given by*

$$g^D_{ij} = g_{ij} \text{ and } \quad T^D_{ijk} = -T_{ijk}.$$

Proof. The proof of the fisrt part of the theorem 4.3 consists in comparing the proposition 3.6 and the proposition 3.2. At the second part of the proof, we compute the tensors T^D_{ijk} by using the formula 4.5 of the dual affine connections with respect to the metric g^D and we obtain that

- $T^D_{111} = T^D_{122} = T^D_{212} = T^D_{221} = 0;$
- $T^D_{112} = T^D_{121} = T^D_{211} = -\frac{2}{\sigma^3};$
- $T^D_{222} = -\frac{8}{\sigma^3}.$

We conlude by comparing these results with the second part of the proposition 3.2. □

Remark 4.4. *Considering the results in theorem 4.3, we see that the α-geometry of the statistical manifold of normal distributions of probabilities is indexed by $\alpha = -1$.*

The remark 4.4 leads to the following result.

Proposition 4.5. *The statistical manifold of normal distributions of probabilities is -1-flat and its -1-Gaussian curvature vanishes.*

Proof. We have to compute the formula (4.3) of the α-curvature tensor and the formula of the α-Gaussian curvature defined by

$$K^{(\alpha)} = \frac{R^{(\alpha)}_{1212}}{\det(g_{ij})}.$$

The formulas (4.1) and (4.2) imply that $\Gamma^{(-1)}_{ijk} = \Gamma^{(1)}_{ijk} + T_{ijk}$. The computation of the Levi-Civita coefficients $\Gamma^{(0)}_{ijk}$ are easy and we obtain:

- $\Gamma^{(0)}_{111} = \Gamma^{(0)}_{212} = \Gamma^{(0)}_{221} = 0;$
- $\Gamma^{(0)}_{112} = \frac{1}{\sigma^3};$
- $\Gamma^{(0)}_{121} = \Gamma^{(0)}_{211} = -\frac{1}{\sigma^3};$
- $\Gamma^{(0)}_{122} = \Gamma^{(0)}_{222} = -\frac{2}{\sigma^3}.$

The computation of $\Gamma^{(-1)}_{ijk}$ coefficients are so obtained and we have

- $\Gamma^{(-1)}_{111} = \Gamma^{(-1)}_{112} = \Gamma^{(-1)}_{212} = \Gamma^{(-1)}_{221} = 0;$
- $\Gamma^{(-1)}_{121} = \Gamma^{(-1)}_{211} = \Gamma^{(-1)}_{122} = -\frac{2}{\sigma^3};$
- $\Gamma^{(-1)}_{222} = -\frac{6}{\sigma^3}.$

The -1-curvature $R_{ijk}^{(-1)l}$ is equal to zero and so the numerator of the formula

$$K^{(-1)} = \frac{R_{1212}^{(-1)}}{\det(g_{ij})}$$

can be computed and it is equal to zero. This induces that the statistical manifold of normal distributions of probabilities is -1-flat and the -1-Gaussian curvature vanishes. □

5 Isometries on the statistical manifold

We investigate that the component functions of a Killing vector field for the Fisher information metric on the statistical manifold of normal distributions of probabilities are harmonic conjugate. In particular, they are both harmonic functions.

Our goal is to describe the form of Killing vector fields in the manifold of normal distributions of probabilities and then to integrate them to obtain local isometries.

Definition 5.1. *Let (M, g) be a Riemann manifold. A Killing vector field is a vector field X on M satisfying the condition that the Lie derivative of g with respect to X vanishes, i.e.,*

$$\mathcal{L}_X g = 0. \tag{5.1}$$

For a Killing vector field X, for each openset $U \subset M$ and $p \in M$, the flow $\phi_t : U_p \to \phi_t(U_p)$ generated by X, satisfying $\phi_0(p) = p$ and $\frac{d}{dt}(\phi_t(p)) = X(\phi_t(p))$, is a familly of isometries, i.e,

$$\phi_t^* g = g, \tag{5.2}$$

for all t such that ϕ_t is defined. The Killing vector fields [5] are also known as infinitesimal isometries, a terminology that arises from the idea of integrating vectors fields to obtain isometries.

The following proposition [5] illustrates the Killing vector field condition in local coordinates, expressing it as a system of fisrt-order linear partial differential equations.

Proposition 5.2. *Let $(x_1, \ldots, x_n)$ be a system of coordiantes on a domain U with corresponding basis $\{(\partial_i)_p := \frac{\partial}{\partial x_i}|_p\}$ of $T_p(\mathbb{R}^n)$. Let g be a Riemannian metric tensor with components $[g_{ij}]$. Then $X = \sum X^i \partial_i$ is a Killing vector field if and only if the components X^i satisfy the $\frac{n(n+1)}{2}$ partial differential equations*

$$\sum_{k=1}^{n} \left(X^k \frac{\partial g_{ij}}{\partial x_k} + g_{jk} \frac{\partial X^k}{\partial x_i} + g_{ik} \frac{\partial X^k}{\partial x_j} \right) = 0, \quad i, j = 1, \ldots, n,\ i \leq j.$$

Proof. Let $X = \sum_{j=1}^{n} X^j \partial_j$ be smooth vector field on $\mathbb{R}^n$. For all $i = 1, \ldots, n$, by using the definition of the Lie derivative and by calculating the effet of the operator $\mathcal{L}_X \partial_i$ on an arbitrary smooth function $f : \mathbb{R}^n \to \mathbb{R}$ at a point $x \in \mathbb{R}^n$, we can show that

$$\mathcal{L}_X(\frac{\partial}{\partial_i}) = -\sum_{j=1}^{n} \frac{\partial X^j}{\partial x_i} \frac{\partial}{\partial_j}.$$

So, in the same way, we can show that

$$\mathcal{L}_X(dx_i) = \sum_{j=1}^{n} \frac{\partial X^i}{\partial x_j} dx_j = dX^i. \tag{5.3}$$

The formula (5.3) and the formula of the Lie derivative [5] of a k-form α with respect to a vector field X lead us to show that for every $j, k = 1, \ldots, n$,

$$\mathcal{L}_X(dx_j \otimes dx_k) = \sum_{r=1}^{n} \left(\frac{\partial X^j}{\partial x_r} dx_r \otimes dx_k + \frac{\partial X^k}{\partial x_r} dx_j \otimes dx_r \right). \tag{5.4}$$

In particular, if we compute the formula of the Lie derivative of $(0, 2)$-tensor $g = \sum_{i,j} g_{ij} dx_i \otimes dx_j$ with respect to a vector field X, then we obtain that

$$\mathcal{L}_X g = \sum_{i,j} h_{ij} dx_i \otimes dx_j,$$

where

$$h_{ij} = \sum_{k=1}^{n} \left(X^k \frac{\partial g_{ij}}{\partial x_k} + g_{kj} \frac{\partial X^k}{\partial x_i} + g_{ik} \frac{\partial X^k}{\partial x_j} \right).$$

It is obvious that $\mathcal{L}_X g = 0$ if the expression $h_{ij} = 0$. □

We have now sufficient ingredients to describe the form of Killing vector fields in the manifold of normal distributions of probabilities.

Proposition 5.3. *The Killing vector field for the Fisher information metric on the statistical manifold of normal distributions of probabilities is given by*

$$X(\mu, \sigma) = (\mu^2 - \sigma^2) \frac{\partial}{\partial \mu} + \mu\sigma \frac{\partial}{\partial \sigma}.$$

Proof. Let $X = X^1 \frac{\partial}{\partial \mu} + X^2 \frac{\partial}{\partial \sigma}$ be a vector field on the statistical manifold of Gaussian distributions of probabilities S. We must compute the components X^1 and X^2 of X.

If we apply the result in the proposition 5.2 for the Fisher information metric

$$(g_{ij}) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{pmatrix}$$

and after computations, we obtain

$$\frac{1}{\sigma^2}\left(\frac{\partial X^1}{\partial \sigma} + 2\frac{\partial X^2}{\partial \mu}\right) - \frac{3}{\sigma^3}\left(X^2 - \frac{2}{3}\sigma\frac{\partial X^1}{\partial \mu}\right) - \frac{3}{\sigma^3}\left(X^2 - \frac{4}{3}\sigma\frac{\partial X^2}{\partial \sigma}\right) = 0.$$

Noting that $\sigma \neq 0$, this relation implies the following system

$$\begin{cases} \frac{\partial X^1}{\partial \sigma} + 2\frac{\partial X^2}{\partial \mu} = 0 \\ \frac{\partial X^1}{\partial \mu} - 2\frac{\partial X^2}{\partial \sigma} = 0. \end{cases} \tag{5.5}$$

If we set $X'^2 = 2X^2$ then the system 5.5 can be written as follow

$$\begin{cases} \frac{\partial X^1}{\partial \sigma} + \frac{\partial X'^2}{\partial \mu} = 0 \\ \frac{\partial X^1}{\partial \mu} - \frac{\partial X'^2}{\partial \sigma} = 0. \end{cases} \tag{5.6}$$

This means that the components X^1 and X'^2 of the Killing vector field of the Fisher information metric are harmonic conjugates. In particular, they are harmonic functions, i.e., they satisfy Laplace's equation

$$\begin{cases} \frac{\partial^2 X^1}{\partial \mu^2} + \frac{\partial^2 X^1}{\partial \sigma^2} = 0 \\ \frac{\partial^2 X'^2}{\partial \mu^2} + \frac{\partial^2 X'^2}{\partial \sigma^2} = 0. \end{cases} \tag{5.7}$$

Then, the announced form of the Killing vector field follows. □

Integrating such vector fields to come up with a local isometry in closed form is impratical. However, one of a wide variety of computer software, for example Mathematica or MathLab can illustrate flow lines in these cases.

6 Acknowledgments

It is a pleasure to thank Carlos Ogouyandjou for his support by giving us the basic documents of references on the information geometry.

References

[1] S-I. Amari. Differential-Geometrical Methods in Statistics. *Lectures notes in Statistics 28*, Springer-Verlag, Berlin Heidelberg, 301p, 1985.

[2] S-I. Amari. Information Geometry and its Applications. *Applied Mathematical Sciences*, Volume 194, 376p, 2016.

[3] S-I. Amari, O.E. Barndorff-Nieisen, R.E. Kass, S.L. Lauritzen and C.R. Rao. Differential Geometry in Statistical Inference. *Lectures Notes-Monographes Series,* 241p, Volume 10, Shanti S.Gupta, Series Editor, 1987.

[4] K.V. Harsha and K. S Subrahamanian Moosath F-Geometry and Amari's α-Geometry on a Statistical Manifold. *Entropy 2014,* 16, 2472-2487; doi:10.3390/e16052472, www.mdpi.com/journal/entropy.

[5] A. Mclnerney. Fist Steps in Differential Geometry. Riemannian, Contact, Symplectic. *Undergraduate Texts in Mathematics, Springer*, 410p, 2013.

[6] S.I Amari and H. Nagaoka, Methods of Information geometry. *American Mathematical Society, Oxford University Press*, Translated from Japanese by Daishi Harada, Vol 191, 216p,1993.

[7] S. I Amari and A. Cichocki, Information geometry of divergence functions. *Bulletin of the Polish Academy of Sciences technical Sciences*, Vol 58, No 1, 2000, Doi:10.2478/v10175-010-0019-1.

This paper was produced by researchers from the geometry team in the Exact and Applied Sciences Laboratory (LSEA) of the University Center for Research and Applied Pedagogy to Sciences (CURPAS).