Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales
49(193):844-871, octubre-diciembre de 2025. doi: https://doi.org/10.18257/raccefyn.3260

Mathematics

Original article

A survey on information geometry: statistical manifolds and statistical models

Un estudio sobre geometría de la información: variedades estadísticas y modelos estadísticos

Nicolás Martínez-Alba[1], Olga Garatejo-Escobar[2]

[1] Departamento de Matemáticas, Universidad Nacional de Colombia, Bogotá, Colombia
[2] Ciencias Básicas, Universidad Distrital Francisco José de Caldas

Abstract

This survey presents an exposition of the foundational concepts of information geometry from a probabilistic viewpoint, as well as recent developments in differential geometry related to this area of information science. We conclude with a non-exhaustive but motivating list of applications of the theory in probability.

Keywords: Fisher metric; Statistical model; Dual connection; Statistical manifold; Exponential family; Mixture family.

Resumen

En este estudio se exponen los conceptos fundamentales de la geometría de la información desde una perspectiva probabilística y se presentan los avances recientes en geometría diferencial relacionados con esta área de las ciencias de la información. Concluimos con una lista, no exhaustiva, pero sí motivadora, de aplicaciones de la teoría en probabilidad.

Palabras clave: Métrica de Fisher; Modelo estadístico; Conexión dual; Variedad estadística; Familia exponencial; Familia mezcla.

Citation: Martínez-Alba & Garatejo-Escobar. A survey on information geometry: statistical manifolds and statistical models. Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales. 49(193):844-871, octubre-diciembre de 2025. doi: https://doi.org/10.18257/raccefyn.3260

Editor: Santiago Vargas Domínguez

***Corresponding autor:**
Nicolás Martínez-Alba;
nmartineza@unal.edu.co

Received: July 19, 2025
Accepted: September 12, 2025
Published on line: October 9, 2025

This is an open access article distributed under the terms of the Creative Commons Attribution License.

Introducción

Information science is a term used to describe the interdisciplinary studies to explore and to analyze information theory and data scenarios (**Kolp, Snoeck, Vanderdonckt &Wautelet,** 2019). These type of studies rely on the interaction of several fields, such as statistical inference, signal processing, machine learning, or neural networks. A fundamental part of this interaction is the role of statistics, with probability theory playing a key part in interpreting data and modeling uncertainty. However, the scope of science information extends beyond statistics, and new developments have emerged in other branches of mathematics. One notable example of this interaction is the so called *geometric science information*. Information geometry is a specialized field within geometry science information that applies concepts from differential geometry to the study of statistical models. Its foundation goes back to the independent works of H. Hotelling (1930) and R. Rao (1945), who proposed a mathematical framework to give a new point of view to families of probability distributions. A central concept in this field is the Fisher information metric, which provides a differentiable structure on the space of probability densities. This metric allows for the exploration of various geometric properties within statistical models, such as distances between probability distributions and their curvature. The Fisher metric is deeply connected to key statistical concepts, including expected value, entropy, and divergence, making it a powerful tool for understanding the geometry of information theory. For a more detailed discussion of this topic, we refer to Nielsen's work on information geometry (**Nielsen,** 2022) and (**Amari & Nagaoka,** 2000) comprehensive insights.

In the probability setting, the exponential and mixture families are two fundamental statistical concepts that describe two types of probability densities, each with its own statistical and algebraic characteristics. These families gain importance in information geometry, where they not only provide statistical properties but also induce geometric structures, such as dual connections, torsion, geodesics, curvatures, divergences, and symmetric 3-tensors. In this work, we focus our results to dual connection, Amari-Chentsov tensor and divergence functions. *Divergence functions* are generating functions that quantify the distance between two probability distributions, playing a crucial role in understanding of their geometric relations. The notion of *dual structure*, introduced by Amari and Chentsov, builds on the relationship between the exponential and mixture families, where each family is associated with a specific affine connection. This duality allows the analysis of statistical models from two complementary perspectives, providing deeper insights into their geometry. The *Amari-Chentsov tensor* together with divergences, offers a comprehensive framework for studying both, the statistical and geometric, properties of probability distributions, bridging the gap between statistical inference and geometry. The relationship between divergence, dual structure, and the Amari-Chentsov tensor lies at the heart of information geometry.

The aim of this manuscript is to provide an introduction to the field of information geometry, relating the probabilistic framework with tools from differential geometry. Along the way, several key concepts will be introduced, and some recent applications will be discussed.

Organization of the paper: Section geometrical setting, is a brief summary of the explicit description of the Riemannian geometry used in the manuscript, including the main geometric notion of *affine connection* and the special case of *Levi-Civita* connection for Riemannian metrics. Section statistical model and its structure, contains the main notions of the geometry of statistical model, including the Fisher metric, exponential and mixture families and torsion-free dual connections with explicit examples. In section geometric concepts for information geometry, we abstract the notion of statistical model to any Riemannian manifold summarizing classical results. In particular, we present new proofs for already known facts on information geometry (Theorem 0.19 and Theorem 0.26) avoiding the use of Christoffel symbols (which are the usual proofs in the literature) and include key relations between nearly-statistical and dual structures and also solitons in geometry (see Lemma 0.33, Proposition 0.34 and Proposition0.35). We close this section by giving new results in isostatistical inmersion in Proposition 0.37. Section applications deals with some recent applications of the previous geometrical construction on statistical manifolds, in particular we present relation with Bayes' theorem, Monte Carlo method, student t-distribution, MLE (maximum likelihood estimation) and clustering patterns

Geometrical setting

Let us begin by presenting some basic concepts of Riemannian manifolds. We refer the reader to (**Jost & Jost**, 2008) for a more detailed description.

A manifold M of dimension n is a connected paracompact Hausdorff space for which every point $x \in M$ has a neighborhood U_x that is homeomorphic to an open subset Ω_x of $\mathbb{R}^n$. Such a homeomorphism $\phi_x : U_x \to \Omega_x$ is called a *coordinate chart*. If for two charts the function $\phi_x \circ \phi_y^{-1} : \Omega_y \to \Omega_x$ is a C^r-diffeomorphism we say that the manifold has a C^r-differentiable structure. The collection $\{(U_x, \phi_x)\}$ is called *differentiable structure* or *smooth structure* for M. We denote T_xM to the vector space which consists of all tangent vectors to curves in M on the point x. It is called the *tangent space* of M at the point x. The disjoint union of all tangent spaces, $TM = \sqcup_{x\in M} T_xM$ is known as *tangent bundle* and is also equipped with differentiable structure. In this way, we can construct smooth functions $X : M \to TM$ so that $X(m) \in T_mM$ which are known as *vector fields*, and the space of vector fields (or sections of TM) is denoted by $\Gamma(TM)$. By construction, tangent vectors (with fixed $m \in M$) can also

Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales
49(193):844-871, octubre-diciembre de 2025. doi: https://doi.org/10.18257/raccefyn.3260

be seen as linear function $X_m : C^\infty(M) \to \mathbb{R}$ satisfying Leibniz rule, in other words they are derivations of the algebra $C^\infty(M)$. When we do not fix points in M, i.e we have a vector field, we get $X : C^\infty(M) \to C^\infty(M)$ linear and satisfying a Leibniz rule, again is a derivation, but in addition we have a natural operation $[X,Y] := X \circ Y - Y \circ X$, named *Lie bracket*.

A *Riemannian metric* on a differentiable manifold M is given by an inner product g on each tangent space T_xM which depends smoothly on the base point x. A *Riemannian manifold* is a differentiable manifold equipped with a Riemannian metric. In any system of local coordinates $(x_1,\ldots,x_n)$ from coordinates charts, the Riemannian metric is represented by a positive definite, symmetric matrix $(g_{ij}(x))_{1\leq i,j\leq n}$ where the coefficients depend smoothly on x.

Let $\gamma : [a,b] \to M$ be a smooth curve. The length of γ is defined as:

$$L(\gamma) = \int_a^b \|\gamma'(t)\| dt,$$

where the norm of the tangent vector γ' is given by the Riemannian metric as $\|\gamma'(t)\|^2 = g_{\gamma(t)}(\gamma'(t),\gamma'(t))$. This value is invariant under re-parametrization of the curve. Taking the infimum of the values $L(\gamma)$ among all the curves γ joining two points $p,q \in M$ we can define a distance function on M and the topology of this distance coincides with the topology of the manifold structure of M.

The metric tensor g also allows us to define a natural differential operation on vector fields that extends the notion of directional derivatives in the Euclidean case. This is known as *Levi-Civita connection* $\nabla^{(0)} : \Gamma(TM) \times \Gamma(TM) \to \Gamma(TM)$ that is $\mathbb{R}$-bilinear but for the algebra of $C^\infty(M)$ it is tensorial in the first variable and satisfies Leibniz rule for the second variable, i.e.,

$$\nabla^{(0)}_{fX} Y = f\nabla^{(0)}_X Y \text{ and } \nabla^{(0)}_X fY = (Xf)Y + f\nabla^{(0)}_X Y,$$

where $f \in C^\infty(M)$. A fundamental theorem of Riemannian geometry states that this is the *unique* connection that is torsion free and metric, that is:

$$\nabla^{(0)}_X Y - \nabla^{(0)}_Y X = [X,Y] \quad \text{and} \quad Zg(X,Y) = g(\nabla^{(0)}_Z X, Y) + g(X, \nabla^{(0)}_Z Y),$$

for any three vector fields X,Y,Z in M, and $[\cdot,\cdot]$ is the commutator of vector fields as first-order differential operator, i.e., $[X,Y] = X \circ Y - Y \circ X$.

Remark 0.1. *Levi-civita connection is a particular class of* **affine connection** *over a manifold M. An affine connection is a $\mathbb{R}$-linear map $\nabla : \Gamma(TM) \times \Gamma(TM) \to \Gamma(TM)$ so that for all smooth function f and any pair of vector field X,Y it holds:*

$$\nabla_{fX} Y = f\nabla_X Y \text{ and } \nabla_X fY = (Xf)Y + f\nabla_X Y. \tag{1}$$

Affine connections are the natural extension of directional derivative when we change the metric on the configuration space. This structure can also be extended to any vector bundle in order to define a right notion of directional derivative. In addition, this is the geometric notion suitable to define the curvature (as Riemann tensor and scalar curvature) that are also used in the Einstein field equation in cosmology

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

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

This operation also can be used for the tangent of a curve (also interpreted as the vector field along a curve) and gives us the following situation: given a smooth curve $\gamma : (a,b) \to M$,

the curve is called *geodesic* if it satisfies $\nabla_{\gamma'}\gamma' = 0$. In a local coordinates $(x_1,\dots,x_m)$, the geodesics can be written for each $i = 1,\dots,m$ as the second-order ODE:

$$x_i''(t) + \sum_{j,k} \Gamma^i_{jk}(\gamma(t)) x_j'(t) x_k'(t) = 0,$$

where the functions Γ^i_{jk} are known as the *Christoffel symbols* of ∇.

Theorem 0.2. ***(Jost & Jost**, 2008, Theorem 1.4.2) Let M be a Riemannian manifold, $x \in M$ and $v \in T_xM$. Then there exist $\varepsilon > 0$ and precisely one geodesic $c : [0,\varepsilon] \to M$ with $c(0) = x, c'(0) = v$. In addition, c depends smoothly on x and v.*

Statistical model and its structure

The purpose of this section is to define the statistical model and its structure, based on differential manifolds whose points are probability distributions. First, we construct manifolds as finite sets of signed measurements, which statistically correspond to the sample space of any particular event. Second, we define the Fisher metric on tangent vectors of the manifold as an inner product. Considering the above, we define a statistical model and special classes of families to study its geometric structure. For more details, associated concepts, and application we refer to (**Amari & Nagaoka**, 2000) and (**Ay, Jost, Vân Lê, & Schwachhöfer**, 2017) and reference therein.

0.1 Finite sample space

From a particular geometrical perspective (**Ay et al.**, 2017), we will determine the probability space for finite sample spaces, i.e., $\Omega = I$ where $I = \{1,2,...,n\}$. From linear algebra, it is known that $S(I)$ is a vector space (as the space of real-valued function from I, and with dual space $S(I)^*$), that can be identified with the linear forms $\mu : S(I)^* \longrightarrow \mathbb{R}$, which in canonical dual basis $\{\delta^1,\cdots\delta^n\}$, is written as:

$$\mu = \sum_{i\in I} \mu_i \delta^i, \tag{2}$$

where each element δ^i is the dual covector, that is $\delta^i(j) = \delta_{ij}$. Indeed, the election of this basis allows us to define the identification (as smooth manifold) $\psi : S(I) \longrightarrow \mathbb{R}^n$ by the coordinate map:

$$\psi(\mu) = \psi(\sum_{i\in I} \mu_i \delta^i) = (\mu_1,\cdots,\mu_n). \tag{3}$$

In other words, this is the coordinate function of $S(I)$ which endows a differentiable structure for $S(I)$ with local model $\mathbb{R}^n$.

A first goal is to endow with a geometry the space of probability. Let us begin with the open submanifold $\mathscr{M}(I)$ of $S(I)$, as the *positive measures* on I:

$$\mathscr{M}_+(I) := \{\mu \in S(I) : \mu_i > 0, \forall i \in I\}, \tag{4}$$

whose topological closure is given by the *non-negative measures*:

$$\mathscr{M}(I) := \{\mu \in S(I) : \mu_i \geq 0, \forall i \in I\}, \tag{5}$$

as *manifold with corners*. Now consider the map $\phi : S(I) \longrightarrow \mathbb{R}$ (defined by $\phi(\mu) = \sum_{i=1}^n \mu_i$), from the regular value theorem we get the submanifold

$$\mathscr{P}_+(I) := \mathscr{M}_+(I) \cap \phi^{-1}(1) = \{\mu \in \mathscr{P}(I) : \mu_i > 0, \forall i \in I, \sum_{i\in I} \mu_i = 1\}, \tag{6}$$

Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales
49(193):844-871, octubre-diciembre de 2025. doi: https://doi.org/10.18257/raccefyn.3260

as level set on the open submanifold $\mathscr{M}_+(I)$. And also, we have the closure of $\mathscr{P}_+(I)$, as:

$$\mathscr{P}(I) := \{\mu \in \mathscr{M}(I) : \mu_i \geq 0, \forall i \in I, \sum_{i\in I} \mu_i = 1\}. \tag{7}$$

It worth to mention that, as $\phi^{-1}(1)$ is a level set, then $\mathscr{P}_+(I)$ decrease in 1 the dimension of $\mathscr{M}_+(I)$.

Example 0.3. *For the set of natural numbers $I = \{1, \cdots, n, n+1\}$ let us consider:*

$$\boldsymbol{U} := \{x = (x_1, \cdots, x_n) \in \mathbb{R}^n : x_i > 0, \forall i \in I, \sum_{i=1}^{n} x_i < 1\}, \tag{8}$$

and the application $\varphi(x) : \boldsymbol{U} \longrightarrow \mathscr{P}_+(I)$, by:

$$\varphi(x) = \sum_{i=1}^{n} x_i \delta^i + (1 - \sum_{i=1}^{n} x_i)\delta^{n+1}, \tag{9}$$

that defines the smooth structure for $\mathscr{P}_+(I)$, and gives us explicit coordinates system. $\diamond$

0.2 Fisher metric and statistical model

We want to introduce an inner product that considers the point-wise data in $\mathscr{M}_+(I)$. This idea will be promoted to a metric on the whole manifold $\mathscr{M}_+(I)$. For this, we will begin with an inner product that will depend on each element μ of the set of positive measures on I. Let us fix $\mu \in \mathscr{M}_+(I)$ and define the inner product $\langle \cdot, \cdot \rangle_\mu$ on $S(I)^*$ as follows:

$$\langle f, g \rangle_\mu = \mu \cdot (fg) = \sum_{i=1}^{n} \mu_i f_i g_i \tag{10}$$

for any $f, g \in S(I)^*$.

It is well known that $S(I)$ and $S(I)^*$ are canonical isomorphic, but we also can obtain a family of isomorphisms parametrized by $\mathscr{M}_+(I)$. Just note that when consider the basis $\{e_i\}$ and $\{\delta^i\}$ on $S(I)^*$ and $S(I)$ respectively, the element $\mu \in \mathscr{M}_+(I)$ (with representation $\mu = \sum_{i\in I} \mu_i \delta^i$) induces an isomorphism between $S(I)$and $S(I)^*$ by $\frac{da}{d\mu} := \sum_{i\in I} \frac{a_i}{\mu_i} e_i$ in $S(I)^*$. Thus, we rewrite the relation (10) in $S(I)$ by:

$$\langle a, b \rangle_\mu = \langle \frac{da}{d\mu}, \frac{db}{d\mu} \rangle_\mu = \sum_i \frac{1}{\mu_i} a_i b_i \tag{11}$$

with $a, b \in S(I)$.
In order to promote the inner product (11) to a Riemannian metric on $\mathscr{M}_+(I)$ we first define its tangent space. It is a straightforward computation to verify that tangent space of a vector space (seen as a manifold) is again the same vector space. In our scenario, we get the following:

$$T_\mu S(I) \cong \{\mu\} \times S(I), \quad \text{and} \quad T_\mu \mathscr{M}_+(I) \cong \{\mu\} \times S(I). \tag{12}$$

Definition 0.4. *(**Ay et al.**, 2017, Definition 2.1) **The Fisher metric** (or metric tensor* $\mathbf{g}$*) on $\mathscr{M}_+(I)$ is defined on each $\mu \in \mathscr{M}_+(I)$ by $\mathbf{g}_\mu : T_\mu \mathscr{M}_+(I) \times T_\mu \mathscr{M}_+(I) \longrightarrow \mathbb{R}$ such that, for two tangent vectors $A \sim (\mu, a), B \sim (\mu, b) \in T_\mu \mathscr{M}_+(I)$*

$$\mathbf{g}_\mu(A, B) := \langle a, b \rangle_\mu. \tag{13}$$

*A **statistical model** for a n-dimensional manifold M is a pair (g, p), with g a Riemannian metric in M and an embedding $p : M \hookrightarrow \mathscr{M}_+(I)$ ($\xi \in M \mapsto p(\xi) = \sum_{i\in I} p_i(\xi)\delta^i$), such that the pull-back of the Fisher metric coincides with g, i.e. for $X, Y \in T_\xi M$*

$$g_\xi(X, Y) = \mathbf{g}_{p(\xi)}(dp_\xi X, dp_\xi Y). \tag{14}$$

The main example of statistical model is the space $M = \mathscr{P}_+(I)$ with the natural embedding on $\mathscr{M}_+(I)$.

In particular, if this manifold is composed of points of probability, it gives rise in statistics to Fisher information matrix. To illustrate the above, we take the manifold $\mathscr{P}_+(I)$ of example 0.3, obtaining the respective information matrix associated with Fisher metric:

Example 0.5. *Consider the coordinate system 9 of $\mathscr{P}_+(I)$. The coefficients of the information matrix associated with Fisher's metric are given by:*

$$g_{ij}(\mu) = \sum_{k=1}^{n} \frac{1}{\mu_k}\delta_{ki}\delta_{kj} + \frac{1}{\mu_{n+1}} = \begin{cases} \frac{1}{\mu_i} + \frac{1}{\mu_{n+1}} & \text{si } i=j \\ \frac{1}{\mu_{n+1}} & \text{in another case} \end{cases} \tag{15}$$

Explicitly we have:

$$G(\mu) := (g_{ij}(\mu)) = \frac{1}{\mu_{n+1}} \begin{pmatrix} \frac{\mu_{n+1}}{\mu_1}+1 & 1 & \dots & 1 \\ 1 & \frac{\mu_{n+1}}{\mu_2}+1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & \frac{\mu_{n+1}}{\mu_n}+1 \end{pmatrix}$$

The inverse matrix of $G(\mu)$ is the probability covariance matrix of μ, each element $i \in \{1,\dots n\}$ has a probability μ_i, the coefficients are:

$$g^{ij}(\mu) = \begin{cases} \mu_i(1-\mu_i) & \text{si } i=j \\ -\mu_i\mu_j & \text{in another case} \end{cases}$$

*In statistics, the diagonal of the matrix $G^{-1}(\mu)$, are the values of the variances of the variables and the remaining coefficients give the value of the correlation between the variables. In fact, this is the statistical origin of the Fisher metric as a covariance matrix (**Rao**, 1992).*$\diamond$

Example 0.6. *In the case of $I = \{1,2\}$, the space $\mathscr{P}_+(I)$ is the positive part of the plane $x+y+z=1$. For convenience of the reader, we choose two points $\nu = (1/3,1/3,1/3)$ and $\mu = (0.12,0.08,0.80)$ for which the Figure 1. shows the centered balls (for the Fisher metric) with the same radio in $\mathscr{P}_+(I)$.*

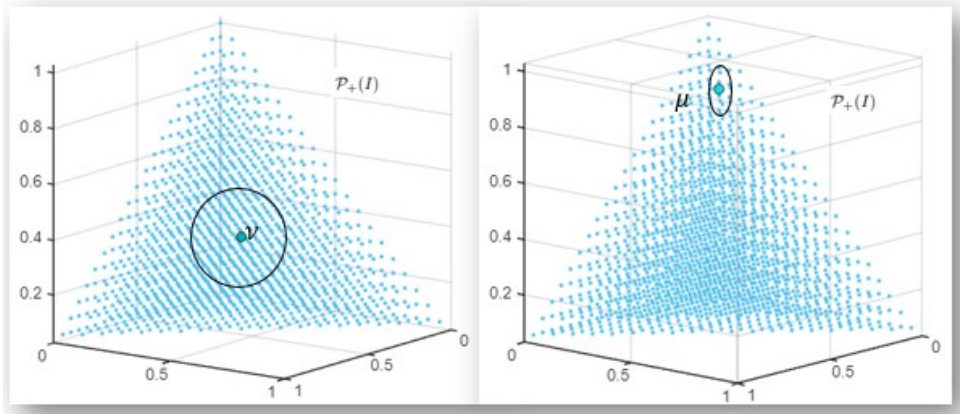

Figura 1. Balls in Fisher metric centered at ν and μ with same radii.

It is worth to note that the meaning of the size of the ball is the quantity of information at each point. To clarify this claim, note that the third coordinate in μ is close to 1, the other

Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales
49(193):844-871, octubre-diciembre de 2025. doi: https://doi.org/10.18257/raccefyn.3260

two coordinates must be closed to 0, but for ν this data is homogeneous. For this reason the ball centered at μ will have more restriction that the one centered at ν, i.e, we get more information on μ. $\diamond$

In order to find the expression of g, it is sufficient to take the canonical directions of tangent $X = \frac{\partial p_i}{\partial \xi_l}, Y = \frac{\partial p_i}{\partial \xi_j}$ in $\mathscr{M}_+(I)$, to obtain:

$$\mathbf{g}_\xi(X,Y) = \sum_{i\in I} \frac{1}{p_i(\xi)} \frac{\partial p_i}{\partial \xi_l}(\xi) \frac{\partial p_i}{\partial \xi_j}(\xi).$$

Now, applying the logarithmic derivative the Fisher metric is rewritten as:

$$\mathbf{g}_\xi(X,Y) = \sum_{i\in I} p_i(\xi) \frac{\partial \log p_i}{\partial \xi_l}(\xi) \frac{\partial \log p_i}{\partial \xi_j}(\xi).$$

This representation of the Fisher metric is more familiar in the standard treatment of information geometry, using the definition of expected value:

$$\mathbf{g}_\xi(X,Y) := E[\frac{\partial \log p_i}{\partial \xi_l}(\xi) \frac{\partial \log p_i}{\partial \xi_j}(\xi)] \tag{16}$$

Using again the notion of logarithmic derivative but on the statistical model $\mathscr{P}_+(I)$, we obtain:

$$\sum_{i\in I} p_i \frac{\partial \log p_i}{\partial \xi_l} = \sum_{i\in I} \frac{\partial p_i}{\partial \xi_l} = \frac{\partial}{\partial \xi_l} \sum_{i\in I} p_i = \frac{\partial}{\partial \xi_l} 1 = 0,$$

which implies

$$\begin{aligned} 0 = \frac{\partial}{\partial \xi_l} E[\frac{\partial}{\partial \xi_j} \log p_i] &= \sum_i \frac{\partial}{\partial \xi_l}(p_i \frac{\partial}{\partial \xi_j} \log p_i) = \sum_i ((\frac{\partial}{\partial \xi_l} p_i) \frac{\partial}{\partial \xi_j} \log p_i) + (p_i \frac{\partial}{\partial \xi_l} \frac{\partial}{\partial \xi_j} \log p_i) \\ &= \sum_i (p_i (\frac{\partial \log p_i}{\partial \xi_l}) \frac{\partial \log p_i}{\partial \xi_j}) + (p_i \frac{\partial}{\partial \xi_l} \frac{\partial}{\partial \xi_j} \log p_i) = g_\xi(X,Y) + \sum (p_i \frac{\partial}{\partial \xi_l} \frac{\partial}{\partial \xi_j} \log p_i \end{aligned}$$

Finally, this yields another equivalent way to define the Fisher metric on $\mathscr{P}_+(I)$:

$$g_\xi(X,Y) := -E[\frac{\partial^2 \log p_i}{\partial \xi_l \partial \xi_j}(\xi)]. \tag{17}$$

Remark 0.7. *Everything that has been done so far with finite I can be extended in a similar way to measurable spaces equipped with probability measures, where I is not necessarily finite but is endowed with a σ-algebra and forms a measurable space. A more detailed treatment of this theory can be found in reference (**Ay et al.**, 2017, Section 3.2.1), however we will give a brief introduction to an equivalent way to tackle this situation in Section 0.3.*

In the following example, we consider a space of distributions parameterized by two terms, where we change the finite sum by integration in the definition of the Fisher metric:

Example 0.8. *Consider the normal distribution x as real random variable with $\xi = (\mu, \sigma)$ where the parameters are the mean μ and the standard deviation σ, that is:*

$$\mathscr{N}(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-\mu)^2}{2\sigma^2}). \tag{18}$$

In order to apply the formulae for the Fisher metric (20) *in this case with, we should note that*

$$\ln \mathscr{N}(x;\mu,\sigma) = -\frac{(x-\mu)^2}{2\sigma^2} - \frac{1}{2}\ln(2\pi\sigma^2).$$

When we take partial derivatives and integration with respect to the distribution we obtain

$$g_{(\mu,\sigma)} = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{pmatrix}. \tag{19}$$

*For a detailed computation for this metric, we refer to (**Garatejo Escobar**, 2024, section 1.3).* $\diamond$

Note that, the previous example says that the statistical model for normal distribution identifies *isometrically* with the hyperbolic space:

$$(H := \{(\mu,\sigma) : \mu \in \mathbb{R}, \sigma > 0\}, \frac{d\mu^2 + 2d\sigma^2}{\sigma^2}).$$

A remarkable geometrical consequence of this identification lies in the distance-minimizing curves in this geometry. Although, in the usual euclidean space the minimizing distance between two normal distributions is a segment of a line in the plane, the Fisher metric is not. Indeed, a geodesic is a section of a semicircle because these are the geodesic in the hyperbolic plane, as shown in Figure 2:

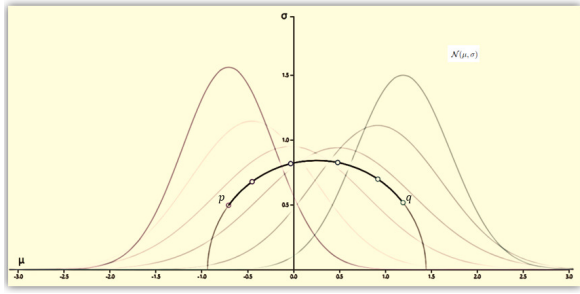

Figura 2. Shortest path between the normal distributions $p = \mathscr{N}(-0.75, 0.5)$ and $q = \mathscr{N}(1.25, 0.5)$ in the Fisher metric .

0.3 Statistical models: another way to understand them

For this brief section, we will adopt the introduction given in (**Amari & Nagaoka**, 2000) for the space of probability distribution and its adaptation to finite positive measures. This approach has the advantage that it it also works on non-finite sample space.

Let us begin with a σ-algebra on a manifold Ω and $\mathscr{M}_+(\Omega,\Theta)$ as the space of finite positive measures on Ω parametrized by an open set $\Theta \subset \mathbb{R}^N$. Indeed, we will assume that the parametrization $p(x,\cdot)$ of finite positive measure on Ω can be identified with a point in Θ in such a way that the assignment $x \mapsto p(x,\cdot)$ is smooth. In this case, the space of probability measures can be seen as $\mathscr{P}_+(\Omega,\Theta) := \{\mu \in \mathscr{M}_+(\Omega,\Theta) : \int_\Omega \mu = 1\}$ which is a 1 corank submanifold of $\mathscr{M}_+(\Omega,\Theta)$. In the space $\mathscr{M}_+(\Omega,\Theta)$, we define the **Fisher** metric tensor as

$$\mathbf{g}_\xi(X,Y) = \int_\Omega p(\cdot,\xi) X(\log p(\cdot,\xi)) Y(\partial \log p(\cdot,\xi)),$$

for each pair of tangent vectors X, Y at $\xi \in \mathscr{M}_+(\Omega,\Theta)$ where are seen as derivation of real functions.

Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales
49(193):844-871, octubre-diciembre de 2025. doi: https://doi.org/10.18257/raccefyn.3260

An again, we can use the definition 0.4 for a statistical model $p : M \hookrightarrow \mathscr{M}_+(\Omega,\Theta)$ for a n-dimensional manifold M.

Example 0.9. *By a similar computation as in equation* (16)*, but for the general case of* $M = \mathscr{P}_+(\Omega,\Theta)$ *with the natural embedding, the Fisher metric takes the following form*

$$g_\xi(X,Y) := -E[\frac{\partial^2 \log p_i}{\partial\xi_l \partial\xi_j}(\xi)]. \tag{20}$$

◇

Example 0.10. *The following examples show some specific probability distributions in terms of statistical model, for which there is possible to use* (20) *to compute the Fisher metric.*

- *Normal Distribution:* $x \in \mathbb{R}$, $\Theta = \{(\mu_1,\mu_2)| -\infty < \mu_1 < \infty, 0 < \mu_2 < \infty\}$

$$p(x;\Theta) = \frac{1}{\sqrt{2\pi}\mu_2} \exp\Big[-\frac{(x-\mu_1)^2}{2\mu_2^2}\Big] \tag{21}$$

  The matrix associated with the Fisher metric is ***Garatejo Escobar****, 2024, section 1.3:*

$$G(\mu_1,\mu_2) = \begin{pmatrix} \frac{1}{\mu_2^2} & 0 \\ 0 & \frac{2}{\mu_2^2} \end{pmatrix}. \tag{22}$$

- *Multivariate Normal Distribution:* $x \in \mathbb{R}^k$*, with* $k > 1$, $\Theta = \{(\mu,\bar{\mu})|\mu \in \mathbb{R}^k, \bar{\mu} \in \mathbb{R}^{k\times k} : positive\ definite\}$

$$p(x;\Theta) = (2\pi)^{\frac{-k}{2}} (det(\bar{\mu}))^{\frac{-1}{2}} \exp\{-\frac{1}{2}(x-\mu)^t \bar{\mu}^{-1}(x-\mu)\}$$

◇

0.4 Cramer-Rao inequality

The Cramér–Rao theorem is a fundamental result in statistical estimation theory that provides a lower bound on the variance of unbiased estimators. Specifically, it states that under regularity conditions, the variance of any unbiased estimator of a parameter cannot be smaller than the reciprocal of the Fisher information. An estimator is a map: $\hat{\xi} : \Omega \longrightarrow \Theta$, that associates to every data $x \in \Omega$ a probability distribution. We say that $\hat{\xi}$ is an *unbiased estimator* if: $E_\xi[\hat{\xi}(x)] = \xi$ for $\forall\xi \in \Theta$.

The mean squared error of an unbiased estimator $\hat{\xi}$ may be expressed as the *variance-covariance* matrix $V_\xi[\hat{\xi}] = [v_\xi^{ij}]$ where:

$$v_\xi^{ij} := E_\xi[(\hat{\xi}^i(x) - \xi^i)(\hat{\xi}^j(x) - \xi^j]$$

Theorem 0.11 (Cramér-Rao inequality)**.** ***(Amari & Nagaoka****, 2000, Theorem 2.2) The variance-covariance matrix* $V_\xi[\hat{\xi}]$ *of an unbiased estimator* $\hat{\xi}$ *satisfies*

$$V_\xi[\hat{\xi}] \geq g_\xi^{-1} \tag{23}$$

in the sense that $V_\xi[\hat{\xi}] - g_\xi^{-1}$ *is positive semidefinite.*

An unbiased estimator $\hat{\xi}$ that achieves the equality $V_\xi[\hat{\xi}] = g_\xi^{-1}$ for all ξ is called an **efficient estimator**.

Exponential and mixture families in geometry

In probability and statistics, exponential families and mixture families are two fundamental classes of probability distributions with rich geometric and statistical structures. Exponential families are important because they admit efficient statistical inference, possess conjugate priors in Bayesian analysis. In contrast, mixture families arise when distributions are combined using convex combinations. Together, exponential and mixture families provide the core of information geometry in the for of *dually flat geometry* enabling a deep geometric understanding of statistical models and inference.

Definition 0.12. *(**Ay et al.**, 2017, Definition 3.1) An **exponential family** is a family of probability distributions $p(\cdot;\vartheta)$ with embedding $p : M \hookrightarrow \mathscr{P}_+(\Omega,\Theta)$, of the form:*

$$p(x;\vartheta) = \exp[\gamma(x) + \sum_{i=1}^{n} f_i(x)\vartheta^i - \psi(\vartheta)], \tag{24}$$

where, x is a real random variable, $\vartheta = (\vartheta^1,\ldots,\vartheta^n)$ is a n-dimensional parameter with function $\gamma(x)$, $f_1(x),\ldots f_n(x)$ and $\psi(\vartheta)$, under the normalized condition $\int_M p(x;\vartheta)dx = 1$.

From the defining relation (24) of exponential family, we have $\log p(x;\vartheta) = \gamma(x) + \sum_{i=1}^{n} f_i(x)\vartheta^i - \psi(\vartheta)$, which gives us:

$$\frac{\partial^2 \log p(x;\vartheta)}{\partial\vartheta^i\partial\vartheta^j} = -\frac{\partial^2}{\partial\vartheta^i\partial\vartheta^j}\psi(\vartheta).$$

As the right hand side is independent of the variable x, we get that:

$$-\int_M \frac{\partial^2 \log p(x;\vartheta)}{\partial\vartheta^i\partial\vartheta^j} p(x;\vartheta)dx = \frac{\partial^2}{\partial\vartheta^i\partial\vartheta^j}\psi(\vartheta)\int_M p(x;\vartheta)dx,$$

which finally gives us an equivalent version of (20) as:

$$g_{ij}(p) = -E\Big[\frac{\partial^2 \log p(x;\vartheta)}{\partial\vartheta^i\partial\vartheta^j}\Big] = \frac{\partial^2}{\partial\vartheta^i\partial\vartheta^j}\psi(\vartheta), \tag{25}$$

thus, the function $\psi(\vartheta) = \log\int \exp[\gamma(x) + \sum_{i=1}^{n} f_i(x)\vartheta^i]dx$ allows us to give an explicit expression for the components of the information matrix associated to the Fisher metric. Therefore, the exponential family is a statistical model whose Fisher metric is defined by the coefficients in (25). The example 0.8 can be revisited in terms of the exponential family as follows:

Example 0.13. *Rewriting the expression* (18)*, we get:*

$$\mathscr{N}(x;\mu,\sigma^2) = \exp(\ln(\frac{1}{\sqrt{2\pi\sigma^2}}))\exp(-\frac{x^2-2\mu x+\mu^2}{2\sigma^2}) = \exp(\frac{\mu x}{\sigma^2} - \frac{x^2}{2\sigma^2} - \frac{\mu^2}{2\sigma^2} - \ln(\sqrt{2\pi\sigma^2})),$$

and use a new inner product to obtain

$$\mathscr{N}(x;\mu,\sigma^2) = \exp((x,x^2)\cdot(\frac{\mu}{\sigma^2},-\frac{1}{2\sigma^2}) - \frac{\mu^2}{2\sigma^2} - \frac{1}{2}\ln(2\pi\sigma^2)), \tag{26}$$

expressed in new variables with the following change of variables $\vartheta^{\mathbf{1}} = \frac{\mu}{\sigma^2}$ and $\vartheta^{\mathbf{2}} = -\frac{1}{2\sigma^2}$. The normal distribution takes the form (24) *with parameters $\vartheta = (\vartheta^{\mathbf{1}},\vartheta^{\mathbf{2}})$ and we note that $\frac{\mu^2}{2\sigma^2} = -\frac{(\vartheta^{\mathbf{1}})^2}{4\vartheta^{\mathbf{2}}}$ y $\sigma^2 = -\frac{1}{2\vartheta^{\mathbf{2}}}$, which directly implies that:*

Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales
49(193):844-871, octubre-diciembre de 2025. doi: https://doi.org/10.18257/raccefyn.3260

$$\mathscr{N}(x;\vartheta^{\boldsymbol{1}},\vartheta^{\boldsymbol{2}}) = \exp\Big((x,x^2)\cdot(\vartheta^{\boldsymbol{1}},\vartheta^{\boldsymbol{2}}) - \Big(-\frac{(\vartheta^{\boldsymbol{1}})^2}{4\vartheta^{\boldsymbol{2}}} + \frac{1}{2}\ln(\pi) - \frac{1}{2}\ln(-\vartheta^{\boldsymbol{2}})\Big)\Big),$$

where the function $\psi(\vartheta)$ *is:*

$$\psi(\vartheta) = -\frac{(\vartheta^{\boldsymbol{1}})^2}{4\vartheta^{\boldsymbol{2}}} + \frac{1}{2}\ln(\pi) - \frac{1}{2}\ln(-\vartheta^{\boldsymbol{2}}).$$

According to (25)*, the components of the Fisher information matrix for the exponential family are calculated with the second partial derivatives in each parameter of* $\psi(\vartheta)$ *yielding again* (19)*.* $\diamond$

Definition 0.14. *(**Ay et al.**, 2017, section 4.2) An* ***mixture family*** *is a family of probability distributions* $p(\cdot;\eta)$ *with embedding* $p : M \hookrightarrow \mathscr{P}_+(\Omega,\Theta)$*, of the form:*

$$p(x;\eta) = c(x) + \sum_{i=1}^{d} h^i(x)\eta_i, \tag{27}$$

where x *is a real random variable,* $\eta = \eta_1,\ldots,\eta_d$ *is a parameter* d*-dimensional,* $c(x)$ *y* $h^1(x),\ldots,h^d(x)$ *are integrable functions under normalization property* $\int_M c(x)dx = 1$ *and* $\int_M h^i(x)dx = 0$

From the expression (27) we have $\frac{\partial^2}{\partial\eta_i\partial\eta_j}\log p(x;\eta) = -\frac{h^i(x)h^j(x)}{[p(x;\eta)]^2}$, which yields (cf. (**Nielsen**, 2020, expression (94))):

$$g_{ij}(p) = \int_M \frac{h^i(x)h^j(x)}{p(x;\eta)}dx, \tag{28}$$

which corresponds to the coefficients of the matrix associated with the Fisher metric.

An example of probability distribution grouped in this family, is the collection of probability functions for a finite sample, as was presented in the example 0.3 and coefficients in (15).

0.5 Geometric notions associated to exponential and mixture families

Once we fix these two families, we proceed to obtain more geometrical data. One of the possible geometrical notions, we can apply is the *parallel transport*. This construction is based on the exposition in (**Ay et al.**, 2017, section 2.4). Let μ and ν be two points in $\mathscr{M}_+(\Omega,\Theta)$, and $A \sim (\mu,a)$ in $T_\mu\mathscr{M}_+(\Omega,\Theta)$ a tangent vector. Denote $\Pi^{(e)}_{\mu,\nu}$ as a parallel transport in $T\mathscr{M}_+(\Omega,\Theta)$ with coordinate system (e), given by:

$$\begin{aligned}\Pi^{(e)}_{\mu,\nu} : T_\mu\mathscr{M}_+(\Omega,\Theta) &\longrightarrow T_\nu\mathscr{M}_+(\Omega,\Theta)\\ a &\mapsto (\nu,(\tilde{\phi}_\nu^{-1}\circ\tilde{\phi}_\mu)(a)) = \sum_i \nu_i\frac{a_i}{\mu_i}\delta^i.\end{aligned} \tag{29}$$

Denote $\Pi^{(m)}_{\mu,\nu}$ as a parallel transport in $T\mathscr{M}_+(\Omega,\Theta)$ with coordinate system (m), given by:

$$\begin{aligned}\Pi^{(m)}_{\mu,\nu} : T_\mu\mathscr{M}_+(\Omega,\Theta) &\longrightarrow T_\nu\mathscr{M}_+(\Omega,\Theta)\\ a &\longmapsto \sum_i a_i\delta^i = a.\end{aligned} \tag{30}$$

The remarkable behavior of metric concerning parallel transport in $A \sim (\mu,a)$ and $B \sim (\nu,b)$ is:

$$g_\nu(\Pi^{(e)}_{\mu,\nu}A,\Pi^{(m)}_{\mu,\nu}B) = \sum_i \frac{1}{\nu_i}(\nu_i\frac{a_i}{\mu_i})b_i = \sum_i \frac{1}{\mu_i}a_ib_i = g_\mu(A,B), \tag{31}$$

indicating the invariance of the Fisher metric, for the two parallel transports.

It is a well known fact in differential geometry, that any parallel transport defines an *affine connection* which is the global notion associated to derivations, torsion, curvature, and geodesics. In this case, the parallel transports $\Pi^{(e)}_{\mu,\nu}$ and $\Pi^{(m)}_{\mu,\nu}$ will determine two types of affine connections, (**Ay et al.**, 2017, Proposition 2.4) as follows: given a curve $\gamma : (-\varepsilon,\varepsilon) \longrightarrow \mathscr{M}_+(\Omega,\Theta)$ with $\gamma(0)=\mu$ and $\dot\gamma(0)=A$, we define **affine e-connection** $\widetilde\nabla^{(e)}_A B|_\mu$ in $\mathscr{M}_+(\Omega,\Theta)$ as:

$$\widetilde\nabla^{(e)}_A B|_\mu = \lim_{t\to 0}\frac{1}{t}(\Pi^{(e)}_{\gamma(t),\mu}(B_{\gamma(t)})-B)\in T_\mu\mathscr{M}_+(\Omega,\Theta),$$

applying (29) on $B=\sum_i b_{\mu,i}\delta^i$, we rewrite:

$$\begin{aligned}\widetilde\nabla^{(e)}_A B|_\mu &= (\mu, \lim_{t\to 0}\frac{1}{t}(\sum_i \mu_i \frac{b_{\gamma(t),i}}{\gamma_i(t)}\delta^i - \sum_i b_{\mu,i}\delta^i))\\ &= (\mu, \sum_i \mu_i\{\frac{\frac{\partial b_i}{\partial a_\mu}(\mu)\gamma_i(t) - b_{\gamma(t),i}\dot\gamma_i(t)}{\gamma^2(t)}\}_{t=0}\delta^i),\end{aligned}$$

evaluating at $t=0$, with $\gamma_i(0)=\mu_i$ y $\dot\gamma_i(0)=a_{\mu,i}$:

$$\begin{aligned}\widetilde\nabla^{(e)}_A B|_\mu =&(\mu, \sum_i \mu_i\{\frac{\frac{\partial b_i}{\partial a_\mu}(\mu)\gamma_i(0) - b_{\gamma(0),i}\dot\gamma_i(0)}{\gamma^2(0)}\}\delta^i)\\ =&(\mu, \sum_i \frac{\partial b_i}{\partial a_\mu}(\mu)\delta^i - \sum_i \frac{1}{\mu_i}b_{\mu,i}a_{\mu,i}\delta^i),\end{aligned}$$

by expressions (2) y (13), we have:

$$\widetilde\nabla^{(e)}_A B|_\mu =(\mu, \frac{\partial b}{\partial a_\mu}(\mu) - \mu(\frac{da_\mu}{d\mu}\cdot\frac{db_\mu}{d\mu})) = (\mu, \frac{\partial b}{\partial a_\mu}(\mu) - g_\mu(A,B)). \tag{32}$$

Similarly for a curve $\gamma:(-\varepsilon,\varepsilon)\longrightarrow\mathscr{M}_+(\Omega,\Theta)$, $\gamma(0)=\mu$ and $\dot\gamma(0)=A$, we define **affine m-connection** $\widetilde\nabla^{(m)}_A B|_\mu$ as:

$$\widetilde\nabla^{(m)}_A B|_\mu = \lim_{t\to 0}\frac{1}{t}(\Pi^{(m)}_{\gamma(t),\mu}(B_{\gamma(t)})-B)\in T_\mu\mathscr{M}_+(\Omega,\Theta),$$

also, by (30) in the previous relation, we get:

$$\widetilde\nabla^{(m)}_A B|_\mu = (\mu, \lim_{t\to 0}\frac{1}{t}(\sum_i b_{\gamma(t),i}\delta^i - \sum_i b_{\mu,i}\delta^i)) = (\mu, \frac{\partial b}{\partial a_\mu}(\mu)), \tag{33}$$

as we have that the main model for information geometry is the space of probability measures $\mathscr{P}_+(\Omega,\Theta)$ with the induced Fisher metric, we must consider connection on $\mathscr{P}_+(\Omega,\Theta)$. For this, we need to project the connections on the tangent space of probability measure as follows:

$$\nabla^{(m)}_A B = \widetilde\nabla^{(m)}_A B, \tag{34}$$

$$\nabla^{(e)}_A B = (\mu, \sum_i \frac{\partial b_i}{\partial a_\mu}(\mu)\delta^i - \sum_i \frac{1}{\mu_i}b_{\mu,i}a_{\mu,i}\delta^i + \sum_i g_\mu(A_\mu,B_\mu)\mu_i\delta_i), \tag{35}$$

Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales
49(193):844-871, octubre-diciembre de 2025. doi: https://doi.org/10.18257/raccefyn.3260

where the first projection is the same because $\widetilde{\nabla}^{(m)}$ belong to $T\mathscr{P}_+(\Omega,\Theta)$ whenever it is evaluated in vector fields of $\mathscr{P}_+(\Omega,\Theta)$, while the second connection must be projected with the extra term, $\sum_i g_\mu(A_\mu,B_\mu)\mu_i\delta_i$. A direct consequence of this definition is that:

$$Ag_\mu(B,C) = g_\mu(\nabla_A^{(m)}B,C) + g_\mu(B,\nabla_A^{(e)}C),$$

for $A=\dfrac{\partial}{\partial\mu_k}, B=\dfrac{\partial}{\partial\mu_l}, C=\dfrac{\partial}{\partial\mu_s}$ in $\mathscr{P}_+(\Omega,\Theta)$. This relation allows us to prove the following property along vector fields

Proposition 0.15. *The connections $\nabla^{(m)}$ and $\nabla^{(e)}$ satisfy the condition:*

$$Ag_\mu(B,C) = g_\mu(\nabla_A^{(m)}B,C) + g_\mu(B,\nabla_A^{(e)}C),$$

for any A,B,C vector field in $\mathscr{P}_+(\Omega,\Theta)$. In addition, $\nabla^{(m)}$ and $\nabla^{(e)}$ are torsion-free.

A proof can be found in (**Calin & Udrişte**, 2014, Proposition 1.10.4), but for convenience of the reader we give a coordinate free proof of the claim,

Proof. For the first claim, we just note that we have proved for a basis of local vector fields. Now, we will use the fact on $A=\dfrac{\partial}{\partial\mu_k}, B=\dfrac{\partial}{\partial\mu_l}, C=\dfrac{\partial}{\partial\mu_s}$ and extend (via Leibniz rule for the connections and the vector fields) as $C^\infty(\mathscr{P}_+(\Omega,\Theta))$-module. For this, consider b a smooth function in $\mathscr{P}_+(\Omega,\Theta)$, and we will prove the relation for bA as follows:

$$b(Ag_\mu(B,C)) = g_\mu(b\nabla_A^{(m)}B,C) + g_\mu(B,b\nabla_A^{(e)}C) = g_\mu(\nabla_{bA}^{(m)}B,C) + g_\mu(B,\nabla_{bA}^{(e)}C);$$

and in a similar way we will prove for bB:

$$\begin{aligned} Ag_\mu(bB,C) &= A(bg_\mu(B,C)) = (Ab)g_\mu(B,C) + bAg_\mu(B,C) \\ &= (Ab)g_\mu(B,C) + b(g_\mu(\nabla_A^{(m)}B,C) + g_\mu(B,\nabla_A^{(e)}C)) \\ &= g_\mu((Ab)B + b\nabla_A^{(m)}B,C) + g_\mu(bB,\nabla_A^{(e)}C) \\ &= g_\mu(\nabla_A^{(m)}bB,C) + g_\mu(bB,\nabla_A^{(e)}C), \end{aligned}$$

yielding the desired result. For bC, the verification is the same as for bB. Finally, extend the relation linearly and the relation holds.

For the second claim, recall that the torsion of a connection is the tensor determined by $\nabla_AB - \nabla_BA - [A,B]$, so we must verify that $\nabla_AB - \nabla_BA = [A,B]$, for m-connection and e-connection. From the defining relation (34), we just note that for any vector fields $A = \sum_i a^i\frac{\partial}{\partial e^i}$ y $B = \sum_j b^j\frac{\partial}{\partial e^j}$, we get:

$$\nabla_A^{(m)}B - \nabla_B^{(m)}A = \sum_{i,j} a^i\frac{\partial b_j}{\partial e^i} - b^j\frac{\partial a_i}{\partial e^j},$$

which is the same as $[A,B]$. Similarly for the relation in (35), then we get torsion free in both cases. □

Remark 0.16. *The previous fact is a well known fact and a key starting point on the theory of information geometry these connections. A proof can be found in (**Calin & Udrişte**, 2014, Proposition 1.10.4). All the proofs known in the literature use Christhoffel symbols, but our proof avoids this geometrical notion and shows that depends only on the local direction because the identity $Ag_\mu(B,C) = g_\mu(\nabla_A^{(m)}B,C) + g_\mu(B,\nabla_A^{(e)}C)$ is tensorial.*

Geometric concepts for information geometry

Motivated by the geometry on statistical models, the Fisher metric and dual connections, we will define *statistical manifolds* as an abstraction of this structures for any Riemannian manifold. The main goal of this section is to present the difference between statistical model and statistical manifold. The technique we will use is to endow those manifolds with a particular geometric structure and verify that such structures coincide with the statistical model.

0.6 Statistical models and dual structure

This section takes as reference (**Amari & Nagaoka**, 2000, section 3.1) and (**Ay et al.**, 2017, section 4.2) to study the dual structure and the notion of torsion-free connections for any Riemannian manifold extending the notion of statistical models previously defined.

Definition 0.17. *Two affine connections $\nabla^{(1)}$ and $\nabla^{(-1)}$ on a Riemannian manifold (M,g) are called dual connection if for any three vector fields they satisfy:*

$$Zg(X,Y) := g(\nabla^{(1)}_Z X, Y) + g(X, \nabla^{(-1)}_Z Y). \tag{36}$$

In this case, the triple $(g,\nabla^{(1)},\nabla^{(-1)})$ is called dual structure on M.

Recall that the torsion of an affine connection ∇ is $\mathrm{Tor}(X,Y) = \nabla_X Y - \nabla_Y X - [X,y]$, and is called torsion-free if $\mathrm{Tor}(X,Y) = 0$. A Riemannian manifold (M,g) is called **statistical manifold** if it is endowed with a pair of torsion-free dual connections $(\nabla^{(1)},\nabla^{(-1)})$.

Lemma 0.18. *Let (M,g) be a Riemannian manifold and $\nabla^{(1)}$ an affine connection, then there exists a unique dual connection $\nabla^{(1)}$ with respect to g.*

Proof. The existence of $\nabla^{(-1)}$ is a direct consequence of the identity (36) defining the dual connection. Now, consider $\bar{\nabla}$ and $\nabla^{(-1)}$ dual connections to $\nabla^{(1)}$ with respect to (M,g), that is, for all tangent vectors X,Y,Z in M, we have:

$$\begin{aligned} g(X,Y) &= g(\nabla^{(1)}ZX, Y) + g(X, \bar{\nabla}ZY) \\ Zg(X,Y) &= g(\nabla^{(1)}ZX, Y) + g(X, \nabla^{(-1)}ZY); \end{aligned}$$

in particular we get $0 = g(X, (\bar{\nabla} - \nabla^{(-1)})_Z Y)$, which finally yields $\bar{\nabla} = \nabla^{(-1)}$. Both conclusions depend on the non-degeneracy of the metric tensor g. □

We will present a summary of well known facts of geometry of dual connections, but we state them in a general version and give coordinate-free proofs that are not available in the literature.

Theorem 0.19. *Let (M,g) be a Riemannian manifold, and $(\nabla^{(1)},\nabla^{(-1)})$ any two connections in M.*

1. *$(\nabla^{(1)},\nabla^{(-1)})$ is dual structure if and only if $(\alpha\nabla^{(1)} + \beta\nabla^{(-1)}, \beta\nabla^{(1)} + \alpha\nabla^{(-1)})$ is dual structure for any combination such that $\alpha + \beta = 1$.*
2. *$(\nabla^{(1)},\nabla^{(-1)})$ are torsion-free if and only if $(\alpha\nabla^{(1)} + \beta\nabla^{(-1)}, \beta\nabla^{(1)} + \alpha\nabla^{(-1)})$ are torsion-free for any combination such that $\alpha + \beta = 1$,*
3. *If $(\nabla^{(1)},\nabla^{(-1)})$ are dual and torsion free, then $2\nabla^{(0)} = \nabla^{(1)} + \nabla^{(-1)}$,*
4. *whenever $2\nabla^{(0)} = \nabla^{(1)} + \nabla^{(-1)}$, then $\nabla^{(1)}$ is torsion free if and only if $\nabla^{(-1)}$ is also torsion free.*

Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales
49(193):844-871, octubre-diciembre de 2025. doi: https://doi.org/10.18257/raccefyn.3260

Proof. Before we give the proofs of each item, we must remark that any $\mathbb{R}$-linear combination of affine connection is again affine connection, just because all the defining conditions (as in Remark ?) are preserved by linearity.
*Proof of (1):*First assume that $(\nabla^{(1)}, \nabla^{(-1)})$ are dual structures, then by linearity of the metric we get

$$\begin{aligned} g &= ((\alpha\nabla^{(1)} + \beta\nabla^{(-1)})_X Y, Z) + g(Y, (\beta\nabla^{(1)} + \alpha\nabla^{(-1)})_X Z) \\ &= \alpha(g(\nabla^{(1)}_X Y, Z) + g(Y, \nabla^{(-1)}_X Z)) + \beta(g(\nabla^{(-1)}_X Y, Z) + g(Y, \nabla^{(1)}_X Z)) = Xg(Y,Z). \end{aligned}$$

For the other direction, just note that

$$\begin{aligned} \nabla^{(1)} &= \tilde{\alpha}(\alpha\nabla^{(1)} + \beta\nabla^{(-1)}) + \tilde{\beta}(\beta\nabla^{(1)} + \alpha\nabla^{(-1)}) \\ \nabla^{(-1)} &= \tilde{\beta}(\alpha\nabla^{(1)} + \beta\nabla^{(-1)}) + \tilde{\alpha}(\beta\nabla^{(1)} + \alpha\nabla^{(-1)}), \end{aligned}$$

with $\tilde{\alpha} = \dfrac{\alpha}{\alpha-\beta}$ and $\tilde{\beta} = \dfrac{-\beta}{\alpha-\beta}$. Note that we also have that $\tilde{\alpha} + \tilde{\beta} = 1$, and the results follow from the previous claim.

Proof of (2): This verification follows the same argument as previous item.

*Proof of (3):*Just note that for $\alpha = \dfrac{1}{2} = \beta$, we get $(\alpha\nabla^{(1)} + \beta\nabla^{(-1)}, \beta\nabla^{(1)} + \alpha\nabla^{(-1)})$ is dual structure and torsion-free (previous items). Indeed, in this case, we have:

$$\alpha\nabla^{(1)} + \beta\nabla^{(-1)} = \beta\nabla^{(1)} + \alpha\nabla^{(-1)},$$

which means that it is *self-dual*, or in other words it is a metric connection. By fundamental theorem in Riemannian geometry, we get that $\alpha\nabla^{(1)} + \beta\nabla^{(-1)}$ is the Levi-Civita connection, and the claim holds.

Proof of (4): An easy verification from the identity $2\nabla^{(0)} = \nabla^{(1)} + \nabla^{(-1)}$, leads us to note that:

$$\nabla^{(1)}_X Y - \nabla^{(1)}_Y X = 2\nabla^{(0)}_X Y - \nabla^{(-1)}_X Y - (2\nabla^{(0)}_Y X - \nabla^{(-1)}_X Y) = 2[X,Y] - (\nabla^{(-1)}_X Y - \nabla^{(-1)}_Y X),$$

which yields that $\nabla^{(1)}_X Y - \nabla^{(1)}_Y X - [X,Y] = [X,Y] - (\nabla^{(-1)}_X Y - \nabla^{(-1)}_Y X)$, and the claim holds. □

In what follows, we work only with this type of combination, in particular, we will consider the family of **α-connection**, with $\alpha \in [-1,1]$, as:

$$(\nabla^{(\alpha)} = \frac{1+\alpha}{2}\nabla^{(1)} + \frac{1-\alpha}{2}\nabla^{(-1)}, \nabla^{(-\alpha)} = \frac{1-\alpha}{2}\nabla^{(1)} + \frac{1+\alpha}{2}\nabla^{(-1)}) \tag{37}$$

or in its equivalent version:

$$\nabla^{(\alpha)} = \nabla^{(-1)} + (\frac{1+\alpha}{2})(\nabla^{(1)} - \nabla^{(-1)}), \quad \nabla^{(-\alpha)} = \nabla^{(1)} - (\frac{1+\alpha}{2})(\nabla^{(1)} - \nabla^{(-1)}). \tag{38}$$

As a direct consequence of the previous theorem, we get:

Corollary 0.20. *If $(g, \nabla^{(1)}, \nabla^{(-1)})$ is a torsion-free dual structure in M, then $(g, \nabla^{(\alpha)}, \nabla^{(-\alpha)})$ is also torsion-free dual structure, for any $-1 \leq \alpha \leq 1$. and we get that,*

$$\nabla^{(0)} = \frac{1}{2}(\nabla^{(\alpha)} + \nabla^{(-\alpha)}). \tag{39}$$

Example 0.21. *Using the notation:*

$$\nabla^{(1)} := \nabla^{(e)} \qquad \nabla^{(-1)} := \nabla^{(m)}$$

we recall that the statistical model $(\mathscr{P}_+(\Omega,\Theta), g, \nabla^{(-1)}, \nabla^{(1)})$ *is also a statistical manifold. Additionally, the previous definition yields that* α*-connections are torsion-free and dual structure. Furthermore, using* (32) *and* (33)*, applied to B in the direction of A (in* $\mathscr{P}_+(\Omega,\Theta)$*) gives local representation as:*

$$\nabla^{(\alpha)}_A B|_\mu = \Big(\mu, \sum_i \Big(\frac{\partial b_i}{\partial a_\mu}(\mu) - \frac{1+\alpha}{2} g_\mu(A,B)\Big)\delta^i\Big), \quad \nabla^{(-\alpha)}_A B|_\mu = \Big(\mu, \sum_i \Big(\frac{\partial b_i}{\partial a_\mu}(\mu) - \frac{1-\alpha}{2} g_\mu(A,B)\Big)\delta \tag{40}$$

As final remark in this section is that the usual prove of the previous results lie on Christoffel symbols for the two dual connections on statistical manifolds, however, we give a coordinate-free proof of these facts and even in the more general setting of statistical structure (**Ay et al.**, 2017, cf. Section 4.2).

0.7 Tensor Amari-Chentsov

We now want to measure the difference between two dual structures $(\nabla^{(1)}, \nabla^{(-1)})$, in explicit way we want to compute $\mathscr{T} = \nabla^{(-1)} - \nabla^{(1)} : \mathfrak{X}^2(M) \to \mathfrak{X}(M)$. We can study this difference by using the metric tensor g, i.e we define the following tensor: or $T : \mathfrak{X}^3(M) \to C^\infty(M)$:

$$T(X,Y,Z) = g(\nabla^{(-1)}_X Y - \nabla^{(1)}_X Y, Z) = g(\mathscr{T}(X,Y), Z) \tag{41}$$

Note the special case of $\nabla^{(0)}$ of Levi-Civita connection, this tensor vanishes identically. In the Literature, T (simialry $\mathscr{T}$) is known as **Amari-Chentsov** tensor.

Example 0.22. *Observe,* $\nabla^{(-1)}_A B|_\mu = \frac{\partial b}{\partial a_\mu}(\mu)$ *and* $\nabla^{(1)}_A B|_\mu = \frac{\partial b}{\partial a_\mu}(\mu) - g_\mu(A,B)$*, the difference between them on* $A \sim (\mu, a_\mu), B \sim (\mu, b_\mu) \in T_\mu \mathscr{M}_+(\Omega,\Theta)$ *is:*

$$\mathscr{T}(A,B) := \nabla^{(-1)}_A B|_\mu - \nabla^{(1)}_A B|_\mu = g_\mu(A,B) = \sum_{i\in I} \frac{1}{\mu_i} a_i b_i \tag{42}$$

Using the Fisher metric with respect to other tangent vector $C = (\mu, c_\mu)$ *on* $\mu \in \mathscr{M}_+(\Omega,\Theta)$*, the relation* (41) *yields (**Ay et al.**, 2017, section 2.5.1):*

$$\mathbf{T}_\mu(A_\mu, B_\mu, C_\mu) = \sum_{i\in I} \mu_i \frac{a_{\mu,i}}{\mu_i} \frac{b_{\mu,i}}{\mu_i} \frac{c_{\mu,i}}{\mu_i}. \tag{43}$$

In this way the Amari-Chentsov in $\mathscr{M}_+(\Omega,\Theta)$ *is:*

$$T_\xi = E_p\Big[\frac{\partial}{\partial \xi_i} \log p \frac{\partial}{\partial \xi_j} \log p \frac{\partial}{\partial \xi_k} \log p\Big] \tag{44}$$

Proposition 0.23. *(**Ay et al.**, 2017, Theorem 4.1) The Amari-Chentsov tensor T from a torsion-free dual structure* $(g, \nabla^{(1)}, \nabla^{(-1)})$ *is a symmetric 3-tensor.*

We remark that all proof of this claim lies on the Christoffel symbols, but here we will do it globally.

Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales
49(193):844-871, octubre-diciembre de 2025. doi: https://doi.org/10.18257/raccefyn.3260

Proof. First, by using the dual structure, we get:

$$\begin{aligned} T(X,Y,Z) &= g(\nabla^{(-1)}_X Y - \nabla^{(1)}_X Y, Z) - g(\nabla^{(1)}_X Y, Z) = Xg(Y,Z) - g(Y,\nabla^{(1)}_X Z) - (Xg(Y,Z) - g(Y,\nabla^{(-1)}_X Z)) \\ &= g(Y,\nabla^{(-1)}_X Z) - g(Y,\nabla^{(1)}_X Z) = T(X,Z,Y). \end{aligned}$$

Also note that the torsion-free condition implies that:

$$T(X,Y,Z) = g(\nabla^{(-1)}_X Y, Z) = g(([X,Y] + \nabla^{(-1)}_Y X) - ([X,Y] + \nabla^{(1)}_Y X), Z) = T(Y,X,Z).$$

These two relations show that T is symmetric on the three components. It remains to verify that it is tensorial, but this follows directly from the fact that:

$$T(X,Y,fZ) = g(\nabla^{(-1)}_X Y - \nabla^{(1)}_X Y, fZ) = fg(\nabla^{(-1)}_X Y - \nabla^{(1)}_X Y, Z) = fT(X,Y,Z).$$

From this identity, we conclude:

$$fT(X,Y,Z) = fT(X,Z,Y) = T(X,Z,fY) = T(X,fY,Z)$$

and similar for $fT(X,Y,Z) = T(fX,Z,Y)$. □

Definition 0.24. *(**Ay et al.**, 2017, Definition 4.2) A **statistical structure** on a manifold M consists of a metric g and a 3-tensor T that is symmetric in all arguments.*

Indeed, both structures are the same, as it is showed in the following results. We give complete and explicit construction to show the dependence on the Levi-Civita connection.

Proposition 0.25. *(**Ay et al.**, 2017, Theorem 4.2) Each statistical structure (M,g,T) induces a statistical manifold $(M,g,\nabla^{(1)},\nabla^{(-1)})$, i.e dual and torsion-free.*

Proof. The idea is to use an auxiliary connection to construct two connections, and conditions from the auxiliary translate to the new ones. Let us denote by $\bar{\nabla}$ an auxiliary connection and define two new connections:

$$\nabla_Z X = \bar{\nabla}_Z X - \frac{1}{2}\mathscr{T}(Z,X), \quad \nabla^*_Z X = \bar{\nabla}_Z X + \frac{1}{2}\mathscr{T}(Z,X). \tag{45}$$

It is a straightforward computation to verify that are $\mathbb{R}$-linear. So it remains to study the behavior with respect to product with $f \in C^\infty(M)$. We give the proof for ∇ and for ∇^* the computation is the same.

$$\begin{aligned} \nabla_Z(fX) &= \bar{\nabla}_Z(fX) - \frac{1}{2}\mathscr{T}(Z,fX) = f(\bar{\nabla}_Z X - \frac{1}{2}\mathscr{T}(Z,X)) + Z(f)X = f\nabla_Z X + Z(f)X \\ \nabla_{fZ}(X) &= \bar{\nabla}_{fZ}(X) - \frac{1}{2}\mathscr{T}(fZ,X) = f(\bar{\nabla}_Z X - \frac{1}{2}\mathscr{T}(Z,X)) = f\nabla_Z X. \end{aligned}$$

Then, ∇ and ∇^* are affine connections. Now, we study the torsion tensor of ∇:

$$\begin{aligned} \nabla_Z X - \nabla_X Z - [Z,X] &= \bar{\nabla}_Z X - \frac{1}{2}\mathscr{T}(Z,X) - (\bar{\nabla}_X Z - \frac{1}{2}\mathscr{T}(X,Z)) - [Z,X] \\ &= \bar{\nabla}_Z X + \bar{\nabla}_X Z - \frac{1}{2}(\mathscr{T}(X,Z) - \mathscr{T}(Z,X)) - [Z,X] \\ &= \bar{\nabla}_Z X - \bar{\nabla}_X Z - [Z,X]. \end{aligned}$$

And a similar computation for the torsion of the tensor ∇^* gives us $\nabla^*_Z X - \nabla^*_X Z - [Z,X] = \bar{\nabla}_Z X - \bar{\nabla}_X Z - [Z,X]$. Thus, if we impose that $\bar{\nabla}$ is torsion-free, we also have same property for both ∇ and ∇^*.

The duality condition follows a similar argument, just by noticing that:

$$\begin{aligned} g(\nabla_Z X,Y)+g(X,\nabla^*_Z Y) &= [g(\bar{\nabla}_Z X,Y)-\frac{1}{2}T(Z,X,Y)]+[g(X,\bar{\nabla}_Z Y)+\frac{1}{2}T(Z,Y,X)] \\ &= g(\bar{\nabla}_Z X,Y)+g(X,\bar{\nabla}_Z Y)-\frac{1}{2}T(Z,X,Y)+\frac{1}{2}T(Z,Y,X) \\ &= g(\bar{\nabla}_Z X,Y)+g(X,\bar{\nabla}_Z Y). \end{aligned}$$

Therefore, if it is also assumed that $\bar{\nabla}$ is a metric connection, we have:

$$g(\nabla_Z X,Y)+g(X,\nabla^*_Z Y)=Zg(X,Y).$$

From previous discussion, if we choose $\bar{\nabla}$ as the Levi-Civita connection for (M,g), then the symmetric tensor $\mathscr{T}$ yields (from (45)) $(M,g,\nabla^{(1)}=\nabla,\nabla^{(-1)}=\nabla^*)$ a statistical manifold. □

Furthermore, these two procedures are inverse of each other, since given the dual structure (∇,∇^*) regarding the metric g we have $\mathscr{T}=\nabla^*-\nabla$, the statistical structure $(g,\mathscr{T})$ and this produces the dual structure, as follows:

$$\nabla'=\nabla^{(0)}-\frac{1}{2}\mathscr{T}=\frac{1}{2}\nabla+\frac{1}{2}\nabla^*-\frac{1}{2}(\nabla^*-\nabla)=\nabla$$

$$\nabla'^*=\nabla^{(0)}+\frac{1}{2}\mathscr{T}=\frac{1}{2}\nabla+\frac{1}{2}\nabla^*+\frac{1}{2}(\nabla^*-\nabla)=\nabla^*$$

On the other hand: $(g,\mathscr{T})$ produces $(\nabla,\nabla^*)=(\nabla^{(0)}-\frac{1}{2}\mathscr{T},\nabla^{(0)}+\frac{1}{2}\mathscr{T})$ in turn gives,

$$\mathscr{T}'=(\nabla^{(0)}+\frac{1}{2}\mathscr{T})-(\nabla^{(0)}-\frac{1}{2}\mathscr{T})=\mathscr{T}$$

In conclusion,

Theorem 0.26. *For any Riemannian metric (M,g), there is a one-to-one relation between torsion-free dual structures (statistical manifold) and statistical structure.*

0.8 Canonical divergence

A divergence function is a non negative function $D:M\times M\to\mathbb{R}$ so that $D|_\Delta=0$ and for any two vector fields X,Y in M we get $X^1Y^2D|_\Delta>0$ where the superscript X^1 and Y^2 represent the lifting to $M\times\{q\}$ and $\{p\}\times M$ respectively. We will say that a divergence function *generates* the structure $(g,\nabla^{(1)},\nabla^{(-1)},T)$ if $g=g^{(D)}$, $(\nabla^{(1)}=\nabla^{(D)},\nabla^{(-1)}=\nabla^{(D^*)})$ and $T=T^{(D)}$ for the following relations:

$$g^{(D)}(V,W)|_p := -D(V\|W)(p), \tag{46}$$

$$g^{(D)}(\nabla^{(D)}_V W,Z) := -D(VW\|Z), \tag{47}$$

$$g^{(D)}(\nabla^{(D^*)}_V W,Z) := -D^*(VW\|Z), \text{ with } D^*(p,q)=D(q,p), \tag{48}$$

$$T^D(X,Y,Z) := -D(XY\|Z)+D(Z\|XY). \tag{49}$$

where we are using the usual notation:

Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales
49(193):844-871, octubre-diciembre de 2025. doi: https://doi.org/10.18257/raccefyn.3260

$$D(V_1\cdots V_n\|W_1\cdots W_m)(p) := (V_1)^1\cdots(V_n)^1(W_1)^2\cdots(W_m)^2 D|_\Delta.$$

The problem of the existence of a divergence function that generates a statistical manifold (and also the statistical structure) is already solved by (**Amari & Nagaoka**, 2000, Amari et al)., indeed there exist many of such functions. For more details and references we refer to (**Ay et al.**, 2017, section 4.4).

As a consequence of the isostatistical immersion theorem, we have the following corollary:

Corollary 0.27. *(**Ay et al.**, 2017, Corollary 4.5) For any statistical manifold (M,g,T) we can find a divergence D of M which defines g and T by the formulas* (46)-(49)*.*

A related question is if there exists a natural choice among all of these divergence functions. The answer comes as **canonical divergence**, that, in addition to the definition of divergence must satisfy:

1. D generates the dualistic structure $(g,\nabla^{(1)},\nabla^{(-1)})$,
2. D is one half of the squared Riemannian distance, i.e. $2D(p,q) = d(p,q)^2$, when the statistical manifold is self-dual, namely when $\nabla^{(1)},\nabla^{(-1)}$ are equal and coincide with the Levi–Civita connection,
3. D is the canonical divergence, when $(M,g,\nabla^{(1)},\nabla^{(-1)})$ is dually flat; this is *Bregman divergence.*

It was computed that canonical divergence is induced by the geodesics in the following way: On a manifold M which has the associated connection ∇ concerning the metric g, given a pair of (closed enough) points $q,p\in M$, there exists a unique curve $\nabla-$geodesic.

$$\gamma_{q,p} : [0,1] \longrightarrow M,$$

which satisfies $\gamma_{q,p}(0) = p$ y $\gamma_{q,p}(1) = q$.

Remark 0.28. *This is equivalent to saying that for each pair of points q and p there exists a unique vector $X(q,p)\in T_qM$ that satisfies $\exp_q(X(q,p)) = p$ where* exp *denotes the exponential map associated with ∇.*

Theorem 0.29. *(**Ay et al.**, 2017, section 4.4.2) Giving an affine connection ∇ and a metric g on M, it is possible to define the **canonical divergence** $D : M\times M\to\mathbb{R}$ associated to (g,∇) as:*

$$D(p\|q) = \int_0^1 t\|\dot\gamma_{p,q}(t)\|^2 dt, \tag{50}$$

for γ the geodesic with initial point p and ends in q. In similar way, it also defines the dual canonical divergence as $D^*(p\|q) := D(q\|p)$.

As a direct consequence, for the statistical model, we get:

Corollary 0.30. *(**Ay et al.**, 2017, Theorem 4.8) The statistical model $(\mathscr{P}_+(\Omega,\Theta),g,\nabla^{(e)},\nabla^{(m)},T)$ coincides with*

$(\mathscr{P}_+(\Omega,\Theta),g^{(D)},\nabla^{(D)},\nabla^{*(D^*)},T^{(D)})$.

A particular class of canonical divergence in $\mathscr{P}_+(\Omega,\Theta)$ is the *KL-divergence* which is defined by

$$D_{KL}(\mu\|\nu) = \sum_{i\in I}\mu_i\log\frac{\mu_i}{\nu_i}. \tag{51}$$

The idea behind the proof is that for a torsion-free dual structure $(\nabla^{(1)}, \nabla^{(-1)})$, the formulae in (50) can be expressed as:

$$D(p\|q) = \Psi(p) + \varphi(q) - \vartheta^i(p)\eta_i(q),$$

for suitable function $\Psi, \varphi, \vartheta^i$ and η_i, and verify that this description of D coincides with equation (51) defining the KL-divergence. For more details we reefeer to (**Ay et al.**, 2017, Section 4.4.3).

A key property at this point is that the Kullback–Leibler (KL) divergence is closely associated with many of the geometric concepts arising in information theory (**Amari & Nagaoka**, 2000, section 3.2). In particular, the KL divergence between a mixture distribution and its components plays a central role in variational inference and in expectation-maximization (EM) algorithms for learning latent structures. Moreover, the KL divergence provides a bridge between global information, as quantified by Shannon entropy, and local information, as characterized by the Fisher information metric (as divergence in $\mathscr{P}_+$) and

$$D_{KL}(p\|q) = h(p,q) - h(p)$$

where $h(p,q)$ is the *differential cross-entropy* of p and q, and $h(P)$ is the *marginal differential entropy* of p. See proof in (**Ay et al.**, 2017, section 4.3).

0.9 Hessian geometry

On a usual Riemann manifold, we mean by Hessian of a function the 2-order differential operator given by the differential and the metric tensor. This notion can be promoted to a more general setting, where we again have second order derivative, i.e the notion of derivation. As we have seen, this can be realized by a connection ∇ on M. We will say that a Riemannian metric g on a flat manifold (M,∇) is a **Hessian metric** if for any point $x \in M$, there exists a local function φ (on an open set around x) such that $g = \nabla d\varphi$. A Hessian structure for which there exists a global 1-form α so that $g = \nabla\alpha$, then g is called **Koszul type** with respect to ∇. A nice description of such structure is the follwing equivalent statement in (**Shima**, 2007)

Proposition 0.31. *For a flat torsion free manifold (M,∇) equipped with Riemannian metric g, the following conditions are equivalent:*

1. *g is Hessian metric for ∇*
2. *$(\nabla_X g)(Y,Z) = (\nabla_Y g)(X,Z)$*
3. *$g(\gamma_X Y, Z) = g(Y, \gamma_X z)$ where $\gamma_X Y = \nabla^{(0)}_X Y - \nabla_X Y$*

An additional equivalent way, that is far from this notes but it worth to mention, is the existence of Kähler metric on the tangent bundle TM.

The notion of information geometric Hessian structure, also is generalized as follows:

Definition 0.32. *A* **nearly statistical** *structure on a manifold M is a pair (h,∇) of a 2-tensor field h and a connection on M, so that*

$$(\nabla_X h)(Y,Z) = (\nabla_Y h)(X,Z) \tag{52}$$

for all X,Y,Z vector fields on M. The pair (h,∇) with torsion T_∇, is a **quasi statistical** *structure if*

$$d^\nabla h(X,Y,Z) := (\nabla_X h)(Y,Z) - (\nabla_Y h)(X,Z) - h(T_\nabla(X,Y),Z) \tag{53}$$

vanishes for all vector fields on M.

Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales
49(193):844-871, octubre-diciembre de 2025. doi: https://doi.org/10.18257/raccefyn.3260

A direct consequence of the definition is the direct relation among the new notions and the torsion of the connection:

Lemma 0.33. *Consider a (0,2) tensor h and a connection ∇, then any two of the following hypothesis*

$$(a)(h,\nabla) \text{ is nearly statistical} \quad (b)(h,\nabla) \text{ is quasi statistical} \quad (c)Im(T_\nabla) \subset Ker(h)$$

imply the third one. In particular, if h is non-degenerated, then condition (c) can be change by (c') torsion-frenees of ∇.

A straighforward computation on a Riemannian manifold (M,g) with dual structure $(\nabla^{(1)},\nabla^{(-1)})$ yields

$$(\nabla^{(1)}_X g)(Y,Z) = (\nabla^{(1)}_Y g)(X,Z) + T_{AC}(X,Z,Y) - T_{AC}(Y,Z,X)$$
$$(d^{\nabla^{(1)}} g)(X,Y,Z) = T_{AC}(X,Z,Y) - T_{AC}(Y,Z,X) - g(T_{\nabla^{(1)}}(X,Y),Z).$$

Therefore, for any statistical structure (i.e torsion-free dual structure or T_{AC} symmetric), the pair $(g,\nabla^{(1)})$ is nearly statistical and quasi statistical structure. On the other direction, we can verify the following equivalence:

Proposition 0.34. *Let (M,g) be a Riemannian manifold with dual structure $(\nabla^{(1)},\nabla^{(-1)})$, then (g,T_{AC}) is statistical structure if and only if $(g,\nabla^{(1)})$ is nearly statistical and $T_{AC}(X,Y,\cdot) = T_{AC}(Y,X,\cdot)$.*

Proof. The proof just follows that

$$T_{AC}(X,Y,\cdot) = T_{AC}(Y,X,\cdot), \qquad \text{and} \qquad T_{AC}(X,\cdot,Y) = T_{AC}(Y,\cdot,X)$$

generate the whole symmetries of T_{AC}. □

We conclude this section by showing a closed relation beteween nearly-statistical structures and other type of geometric structure called *almost-tensor solitons* in Equation 54:

Proposition 0.35. *Let (M,g) a Riemannian manifold with a metric connection ∇ (i.e $\nabla g = 0$) so that*

$$\nabla.\xi + J = \lambda Id \tag{54}$$

ξ is tensor field, $J : TM \to TM$ and λ is a smooth function. We get that, (h_ξ,∇) is nearly statistical if and only if

$$R(X,Y)\xi + J(T(X,Y)) = \lambda T(X,Y)$$

for any X,Y vector fields in M.

Proof. From the fact that $\nabla g = 0$, and the definition of torsion and curvature, we have

$$\begin{aligned}(\nabla_X h_\xi)(Y,Z) - (\nabla_Y h_\xi)(X,Z) &= g((\nabla_{[X,Y]} + R(X,Y))\xi + \nabla_{[Y,X]+T(Y,X)}\xi, Z)\\ &= g(R(X,Y)\xi + \lambda T(Y,X) - J(T(Y,X)), Z)\end{aligned}$$

where the last equality comes from the almost-tensor-soliton. As the relation holds for all vector field Z, the result is proved. □

0.10 Immersion

In the previous section, we conclude that there is a one-to-one relation between statistical structures and statistical manifolds, but both of them are motivated by the statistical model $(P_+(I), g)$. A natural question arises when we want to see any statistical structure as a statistical model. For this question, we should consider the following notion:

Definition 0.36. *(**Ay et al.**, 2017, Definition 4.9) Let h be a smooth application of a statistical manifold (M_1, g_1, T_1) to a statistical manifold (M_2, g_2, T_2). The map h will be called isostatistical immersion if it is an immersion of M_1 into M_2 such that $g_1 = h^*(g_2)$ and $T_1 = h^*(T_2)$.*

A first result related to this notion is an original extension of the statement in (**Ay et al.**, 2017, Lemma 4.6) for Fisher metric. It is important to highlight that (**Ay et al.**, 2017, Lemma 4.6) just prove item a., and in this note we give a proof for the rest of the claims:

Proposition 0.37. *Let Ω be a measurable space and $p_i : \Omega \times M_i \longrightarrow [0,1]$ measurable functions with $\int_{\Omega \times M_i} p_i(x;\xi_i)dx = 1$ (for $i = 1,2$), such that $h : (M_1, g_1, T_1) \longrightarrow (M_2, g_2, T_2)$ is an isostatistical immersion between this two statistical structures, then we have:*

a. If g_1 and g_2 are Fischer metrics, then $h^* p_2(x;\xi_2) = p_1(x;\xi_1)$.

b. If $h^* p_2(x;\xi_2) = p_1(x;\xi_1)$, and g_2 is Fisher metric, then g_1 it also a Fisher metric.

c. If $h^* p_2(x;\xi_2) = p_1(x;\xi_1)$ and T_2 is Amari-Chentsov tensor, then T_1 is Amari-Chentsov tensor.

Proof. As we commented before the statement of the proposition, we just prove the second and third claim:

b. Given $h^* p_2 = g_1$, and Fisher metric g_2, we have,

$$g_1(V_1,V_2) = E_{h^* p_2}[\frac{\partial}{\partial V_1}\log h^* p_2(x;\xi_2)\frac{\partial}{\partial V_2}\log h^* p_2(x;\xi_2)] = E_{p_1}[\frac{\partial}{\partial V_1}\log p_1(x;\xi_1)\frac{\partial}{\partial V_2}\log p_1(x_1;\xi$$

then g_1 is Fisher metric

c. Given $h^* T_2 = T_1$ and T_1 is Amari-Chentsov tensor.

$$\begin{aligned} T_1(V_1,V_2,V_3) &= E_{h^* p_2}\Big[\frac{\partial}{\partial V_1}\log h^* p_2(x;\xi_2)\frac{\partial}{\partial V_2}\log h^* p_2(x;\xi_2)\frac{\partial}{\partial V_3}\log h p_2(x;\xi_2)\Big] \\ &= E_{h^* p_1}\Big[\frac{\partial}{\partial V_1}\log p_1(x;\xi_1)\frac{\partial}{\partial V_2}\log p_1(x;\xi_1)\frac{\partial}{\partial V_3}\log h^* p_1(x;\xi_1)\Big] \end{aligned}$$

then T_1 is Amari-Chentsov tensor.

□

Applications

Here we will give a brief exposition of some applications of the Fisher metric in statistical theory. The aim of this last section is to open new horizons in the research and use of this new technique and to join different branches of mathematics and statistics. We give explicits reference for convenience of the reader.

Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales
49(193):844-871, octubre-diciembre de 2025. doi: https://doi.org/10.18257/raccefyn.3260

The 1-dimensional statistical model

(**Ly, Marsman, Verhagen, Grasman, & Wagenmakers**, 2017) In the case of Bernoulli model, we have $\Omega = \{0,1\}$ and $p(x,\mu) = \mu^x(1-\mu)^{1-x}$, then from (20) we get

$$g_\mu = \frac{1}{\mu(1-\mu)} : \mathbb{R} \to \mathbb{R}$$

as metric tensor in one dimension. Despite this seems a simply computation, in Bayesian probability theory this will have a very nice interpretation. One problem of interest in this theory is the *principle of insufficient reason*, this is a rule for assigning prior probabilities when there is no reason to favor one possibility over another. In particular this suggest to extend data by a Bernoulli distribution with *prior* of parameter of interest. After that, the Bayes' theorem can be stated as

$$F(\mu|X) = \frac{p(x,\mu)F(\mu)}{\int_X p(x,\mu)F(\mu)d\mu},$$

where $F(\mu|X)$ is knwon as *poteriori density* of μ and the integral is known as *marginal likelihood*. In some data models, we might not be able to derive such a distribution analytically, as the integration involved in the marginal likelihood is typically hard to perform. However, Jeffrey´s rule (**Jeffreys**, 1946) interpretation in information geometry depends on the fisher metric along the derivative of $\ln p$, and in the 1-dimensional particular case this can be seen as

$$F(\mu) = \frac{\sqrt{g_\mu}}{\int_\Omega \sqrt{g_\mu} d\mu}$$

for the Fisher metric. For further references on the interpretation of Jeffrey´s rule in information geometry, we refer to (**Li, Sun, & Peng**, 2022) and (**Snoussi & Mohammad-Djafari**, 2003).

Riemannian Hamiltonian Monte Carlo

(**Betancourt**, 2013) Hamiltonian Monte Carlo is a powerful Markov Chain Monte Carlo (MCMC) method for sampling probability distributions. It originates in physics but gains deep insight in information geometry. HMC augments the original probability space of interest (say, the posterior distribution $\pi(\theta)$) by introducing auxiliary momentum variables p, and simulates dynamics on this phase space using Hamiltonian mechanics where the main tool are the position and momentum coordinates. The Hamiltonian is typically

$$H(\theta,p) = -\ln(\pi(\theta)) + \frac{1}{2}p^T M^{-1} p,$$

where M is a mass matrix. The mass matrix M interprets a global decorrelation and the potential (similar to the potential function in classical mechanics) comes as $log|M|$. The common Hamiltonian Monte Carlo algorithm considers the measure-preserving underdamped Langevin process [Stoltz, Rousset] and hamiltonian trajectories are related to geodesic for the metric on M.

Riemannian Manifold HMC (RMHMC) uses a position-dependent metric, typically related to the Fisher information matrix or the negative Hessian of the log-posterior. But this introduces a technical challenge: the Hessian may not be positive definite, which is essential for defining a valid Riemannian metric. A solution comes with the SoftAbs metric (introduced by Betancourt in 2013) which provides a smooth, positive-definite approximation of the Hessian using a *soft absolute value* function applied to the eigenvalues. If the initial metric has eigenvalues λ_i, we choose suitable regularizing values α so that the new metric

$$X = Q\text{diag}_{abs}(\frac{\lambda_i}{\tanh(\alpha\lambda_i)})Q^T$$

is positive-defined.

For more detail we refeer to (**Stoltz, Rousset, et al.**, 2010). and (?, ?)

q-exponential and bivariated Student's t-distributions

(**Sakamoto & Matsuzoe**, 2015) From a usual computation, we know that $\lim_{q\to 1^+}\exp_q(t) = e^t$ for the function $\exp_q(t) = (1+(1-q)t)^{\frac{1}{1-q}}$. This allows us to define $\exp_q$ as q-**exponential**. This same idea can be extended to exponential family, where we define q-exponential family as:

$$p(x;\vartheta) = \exp_q[\gamma(x) + \sum_{i=1}^{n} f_i(x)\vartheta^i - \psi(\vartheta)],$$

In the same spirit of exponential family, we can define q-Fisher metric as $g_{ij} = \frac{\partial}{\partial\vartheta_i}\frac{\partial}{\partial\vartheta_j}\psi$. In addition, we can adapt the construction of Amari-Chentsov tensor (44) to induce $q(\alpha)$-connection (for $\alpha\in[-1,1]$) in the same way as we did with the Amari-Chentsov tensor:

$$g(\nabla_X^{q(\alpha)}Y,Z) = g(\nabla_X^{q(0)}Y,Z) - \frac{\alpha}{2}T_q(X,Y,Z)$$

for the Levi-Civita connection $\nabla^{q(0)}$ of the Fisher metric form q-exponential family. As usual, we define the q-exponential connection and q-mixture connection as:

$$\nabla_X^{q(e)} = \nabla_X^{q(1)}, \qquad \nabla_X^{q(m)} = \nabla_X^{q(-1)}$$

respectively.

Example 0.38. *Recall that n-dimensional Student's t-distribution with degree of freedom ν or a q-Gaussian distribution is given by*

$$p_q(x,\mu,\bar{\mu}) = \frac{\Gamma(\frac{1}{q-1})}{(\pi\nu)^{d/2}\Gamma(\frac{\nu}{2})\sqrt{|\bar{\mu}|}}\Big(1+\frac{1}{\nu}(x-\mu)^T\bar{\mu}^{-1}(x-\mu)\Big)^{\frac{1}{q-1}}.$$

With the change of variables

$$z_q = \frac{(\pi\nu)^{d/2}\Gamma(\frac{\nu}{2})\sqrt{|\bar{\mu}|}}{\Gamma(\frac{1}{q-1})}, \qquad R = \frac{z_q^{q-1}}{(1-q)d+2}\bar{\mu}^{-1}, \text{ and} \qquad \theta = 2R\mu$$

this is also q-exponential:

$$p_q(x,\mu,\bar{\mu}) = \exp_q\Big(\sum_{i=1}^{d}\theta^i x_i - \sum_{i=1}^{d}R_{ii}x_i^2 - 2\sum_{ij}R_{ij}x_ix_j - \frac{1}{4}\theta^TR^{-1}\theta + \ln_q(\frac{1}{z_q})\Big).$$

The associated Fisher metric comes as second derivatives of $\psi(\theta) = \frac{1}{4}\theta^TR^{-1}\theta - \ln_q(\frac{1}{z_q})$.
◇

A direct consequence of the example and from previous construction is the following theorem, which gives a new result on the product of random variables:

Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales
49(193):844-871, octubre-diciembre de 2025. doi: https://doi.org/10.18257/raccefyn.3260

Theorem 0.39. *Let X_1 and X_2 two random variables with t-Student distribution $p_i(x_i)$ respectively with same parameter q. Then there exists a bivariate t-Student distribution $p(x_1,x_2)$ such that X_1 and X_2 are independent with*

$$p = p_1 \otimes p_2 \otimes (-c)$$

where $c = (\ln_q(\frac{1}{z_q(\sigma_1)}) + \ln_q(\frac{1}{z_q(\sigma_2)})) - \ln_q(\frac{1}{z_q})$ and z_q is the m-normalization function for the bivariate Student's t-distribution.

0.11 Some results on MLE

Maximum likelihood estimator (MLE) is a well known estimator in statistics. The asymptotic properties of MLEs on Euclidean spaces is of great interest, but their studies on manifolds are still insufficient. If $X_1,\dots,X_n$ are independent random samples from a family of distributions $p(x,\theta)$, a maximum likelihood estimator is any $\hat{\theta}_n$ which solves

$$\max_\theta \frac{1}{n}\sum_{i=1}^{n} \log p(X_i,\theta) = \frac{1}{n}\sum_{i=1}^{n} \log p(X_i,\hat{\theta}_n).$$

Example 0.40. *When we work in the space of symmetric positive define matrices (SPD) with the Fisher metric (in this particular example called Rao's distance), it is possible to compute the MLE as*

$$\hat{\theta}_n = \arg min_\theta \sum_{i=1}^{n} d^2(X_i,\theta)$$

which is unique (as MLE) and converge in probability to some element on SPD. $\diamond$

To study asymptotic efficiency of MLE, in (**Heyde**, 1997)there was defined a generalization of estimating functions on a measurable space Ω with family of probabilities $(p_m)_{m\in M}$ parametrized by M. An **estimating form** on Ω along M is a function $\omega : \Omega \times M \to T^*M$ such that

$$E_\theta[\omega(x,\theta)X_\theta] = 0$$

for all $X_\theta \in T_\theta M$. Note that given a function l on M (parametrized by Ω) it is possible to get its derivative $dl : \Omega \times M \to T^*M$ and we can define the matrix $\Gamma = \begin{pmatrix} E & F \\ G & H \end{pmatrix}$ comparing ω and dl via a fixed $\theta_0 \in M$,

$$\begin{aligned} E_{a,b} &= E_{\theta_0}[dl(x,\theta)e_a dl(x,\theta)e_b], \\ F_{a,b} &= E_{\theta_0}[dl(x,\theta)e_a \omega(x,\theta)e_b] = G_{b,a}, \\ H_{a,b} &= E_{\theta_0}[\omega(x,\theta)e_a \omega(x,\theta)e_b], \end{aligned}$$

where $\{e_a\}$ is a frame of TM. This data gives the following estimates:

Proposition 0.41. *If Γ is positive defined, then $E^{-1} < -(E[\nabla dl(e_a,e_b)]^{-1})HF^{-1}$ and E^{-1} is the limit distribution when $\omega = dl$. In addition, if $(a_1,\dots,a_n)$ are the coordinates of $Log_{\theta_0}(\bar{\theta})$ with $E[Log_{\theta_0}(\bar{\theta})] = 0$, then the Cramer-Rao inequality* (23) *holds for $\mathscr{F}^{-1}$ and curvature terms where*

$$\mathscr{F}_{ij} = E[d(\sum_{i=1}^{n} \log p(x_i,\theta_0)e_i d(\sum_{i=1}^{n} \log p(x_i,\theta_0)e_j].$$

Clustering Patterns

(**Amari & Cichocki**, 2010) Nulla in ipsum. Praesent eros nulla, congue vitae, euismod ut, commodo a, wisi. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Aenean nonummy magna non leo. Sed felis erat, ullamcorper in, dictum non, ultricies ut, lectus. Proin vel arcu a odio lobortis euismod. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Proin ut est. Aliquam odio. Pellentesque massa turpis, cursus eu, euismod nec, tempor congue, nulla. Duis viverra gravida mauris. Cras tincidunt. Curabitur eros ligula, varius ut, pulvinar in, cursus faucibus, augue. Clustering is an important technique in data analysis. It is used to group patterns or data points into subsets, called clusters, that share similar characteristics. One way to implement clustering using information geometry is via *divergence*, which is an asymmetric measure of the difference between probability distributions.

We represent patterns as vectors x in a pattern manifold $X \subset \mathbb{R}^n$, and we study the case where a divergence is defined between two patterns x and x'.

The dual ϕ-divergence between two patterns x and x'

$$D_\phi[x : x'] = \phi(x) - \phi(x') - \nabla\phi(x') \cdot (x - x'),$$

is constructed from a dual convex function ϕ.

A cluster C consists on k patterns $x_1, \cdots, x_k$ in X. The goal is to determine a representative for C that is as close as possible to all its members. To achieve this, we compute the average of ϕ-divergences among C a vector η, given by:

$$D_\phi[C : \eta] = \frac{1}{k} \sum_{x_i \in C} D_\phi[x_i : \eta].$$

and the procedure finish by minimizing the term $D_\phi[C : \eta]$. Indeed, it is possible to obtain:

Theorem 0.42. *(**Amari & Cichocki**, 2010, Theorem 11.1) The ϕ-center of cluster C is given by*

$$\eta_C = \frac{1}{k} \sum x_i$$

for any ϕ.

Remark 0.43. *It is possible to generalize the situation by considering that, instead of a cluster C, a probability distribution $p(x)$ over x is given. In this case, the center of the distribution is defined as the minimizer of the expression:*

$$D_\phi[p : \eta] = \int D_\phi[x : \eta] p(x)\, dx.$$

Thus, the center is simply the expectation of x for any ϕ, given by:

$$\eta_p = \int x p(x)\, dx.$$

Next step is to propose an algorithm that classifies patterns into clusters based on their centers. This approach, known as the k-**means clustering algorithm**, iteratively refines the cluster centers to minimize variance within each cluster and better represent the data structure. It can be summarized in the following steps:

1. **Initial Step:** Choose m cluster centers $\eta_1, \ldots, \eta_m$ arbitrarily such that they are all different.

Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales
49(193):844-871, octubre-diciembre de 2025. doi: https://doi.org/10.18257/raccefyn.3260

2. **Classification Step:** For each x_i, calculate the ϕ-divergences to the m cluster centers. Assign x_i to cluster C_h that minimizes the ϕ-divergence:

$$x_i \in C_h : D_\phi[x_i : \eta_h] = \min_j \{D_\phi[x_i : \eta_j]\}.$$

   Thus, new clusters $C_1, \ldots, C_m$ are formed.

3. **Renewal Step:** Calculate the ϕ-centers of the renewed clusters, obtaining new cluster centers $\eta_1, \ldots, \eta_m$.

4. **Termination Step:** Repeat the above procedures until convergence.

It is known that the procedures terminate within a finite number of steps, giving a good clustering result, although there is no guarantee that it is optimal.

Conclusion and Remarks

In this manuscript, we focus on particular concepts of geometry and probability, namely information geometry. However, there is a long list of geometric structures that can be used or adapted to deeper notions in probability and statistics structures. A non exhaustive list of these connections includes: support vector machine, Hessian structures, time series, classification of a Stochastic process, clustering, geometric deep learning, deformed entropy, cross-entropy, relative entropy among others. More applications and recent research on the subject can be found in the series of book on *geometric science information* and in the link `https://franknielsen.github.io/GSI/`

Conflict of interest

We, the authors, declare that we do not have any type of *conflict or interest* in the research for this work.

Author contributions

This project's conception and development were a joint effort by both authors, who also worked together to write and revise the manuscript.

References

Amari, S.-i., & Cichocki, A. (2010). Information geometry of divergence functions. *Bulletin of the polish academy of sciences. Technical sciences*, ***58***(1), 183–195.

Amari, S.-i., & Nagaoka, H. (2000). *Methods of information geometry* (Vol. 191). American Mathematical Soc.

Ay, N., Jost, J., Vân Lê, H., & Schwachhöfer, L. (2017). *Information geometry* (Vol. 64). Springer.

Betancourt, M. (2013). A general metric for riemannian manifold hamiltonian monte carlo. In *International conference on geometric science of information* (pp. 327–334).

Calin, O., & Udrişte, C. (2014). *Geometric modeling in probability and statistics* (Vol. 121). Springer.

Garatejo Escobar, O. C. (2024). *Estructura dual en modelos estadísticos sobre productos warped* (Unpublished doctoral dissertation). Universidad Nacional de Colombia.

Heyde, C. C. (1997). *Quasi-likelihood and its application: a general approach to optimal parameter estimation*. Springer.

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, ***186***(1007), 453–461.

Jost, J., & Jost, J. (2008). *Riemannian geometry and geometric analysis* (Vol. 42005). Springer.

Kolp, M., Snoeck, M., Vanderdonckt, J., & Wautelet, Y. (2019). An overview of scientific areas in research challenges in information science. *International Conference on Research Challenges in Information Science*, ***13***(), 1–5.

Li, M., Sun, H., & Peng, L. (2022). Fisher–rao geometry and jeffreys prior for pareto distribution. *Communications in Statistics-Theory and Methods*, ***51***(6), 1895–1910.

Ly, A., Marsman, M., Verhagen, J., Grasman, R. P., & Wagenmakers, E.-J. (2017). A tutorial on fisher information. *Journal of Mathematical Psychology*, ***80***(), 40–55.

Nielsen, F. (2020). *An elementary intrduction to information geometry, sony computer science laboratories inc.* Japan, MDPI Journal.

Nielsen, F. (2022). The many faces of information geometry. *Not. Am. Math. Soc*, ***69***(1), 36–45.

Rao, C. R. (1992). Information and the accuracy attainable in the estimation of statistical parameters. In *Breakthroughs in statistics: Foundations and basic theory* (pp. 235–247). Springer.

Sakamoto, M., & Matsuzoe, H. (2015). A generalization of independence and multivariate student'st-distributions. In *International conference on geometric science of information* (pp. 740–749).

Shima, H. (2007). *The geometry of hessian structures*. World Scientific.

Snoussi, H., & Mohammad-Djafari, A. (2003). Information geometry and prior selection. In *Aip conference proceedings* (Vol. 659, pp. 307–327).

Stoltz, G., Rousset, M., et al. (2010). *Free energy computations: A mathematical perspective*. World Scientific.