Submitted to the Annals of Statistics

RIEMANNIAN STATISTICS GEOMETRY: A COUNTERPART APPROACH OF INFERENCE GEOMETRY

By Luisberis Velazquez

Universidad Católica del Norte

Riemannian statistics geometry is proposed in this work as a counterpart approach of inference geometry. This geometry framework is inspired on the existence of a notable analogy between the general theorems of *inference theory* and the the *general fluctuation theorems* associated with a parametric family of distribution functions $dp(I|\theta)$, which describes the stochastic behavior of a set of *continuous stochastic variables* driven by a set of control parameters θ. In this approach, statistical properties are rephrased as purely geometric notions derived from the *Riemannian structure* on the manifold $\mathcal{M}_\theta$ of stochastic variables I. Consequently, this theory arises as an alternative framework for applying the powerful methods of differential geometry for the statistical analysis.

1. Introduction. Inference theory supports the introduction of a Riemannian metric [1]:

$$ds_F^2 = g_{\alpha\beta}(\theta)d\theta^\alpha d\theta^\beta \tag{1.1}$$

to characterize the *statistical distance* between two close members of a generic parametric family of distribution functions:

$$dp(I|\theta) = \rho(I|\theta)dI \tag{1.2}$$

in terms of the so-called *Fisher's information matrix* [2]. The existence of this type of Riemannian formulation was pioneering suggested by Rao [3], which is referred to as *inference geometry* in the literature.

The goal of this paper is to demonstrate that the same parametric family (1.2) supports the introduction of a Riemannian metric:

$$ds^2 = g_{ij}(I|\theta)dI^i dI^j \tag{1.3}$$

to characterize the statistical distance between two close sets of stochastic continuum variables I and $I + dI$ for fixed values of control parameters θ.

AMS 2000 subject classifications: Primary 97K70, 70G45; secondary 62Fxx
Keywords and phrases: Geometrical Methods in Statistics

This novel approach, hereafter referred to as *Riemannian statistics geometry*, is inspired on the existence of a notable analogy between inference theory and the so-called *fluctuation theory*[1] [4].

2. Motivations. Let us start from the parametric family of distribution functions (1.2), which describes the stochastic behavior of a set of continuous variables I driven by a set θ of control parameters. Let us denote by $\mathcal{M}_\theta$ the manifold constituted by all admissible values of the stochastic variables I that are accessible for a given value θ of control parameters, which is hereafter assumed as a simply connected domain. Moreover, let us denote by $\mathcal{P}$ the manifold constituted by all admissible values of control parameters θ. The parametric family of distribution functions (1.2) can be analyzed from two different perspectives:

- The study of fluctuating behavior of stochastic variables $I \in \mathcal{M}_\theta$, which is the main interest of ***fluctuation theory*** [4];
- The analysis of the relationship between this fluctuating behavior and the external control described in terms of parameters $\theta \in \mathcal{P}$, which is the interest of ***inference theory*** [2].

2.1. *Fluctuation theory.* Let us admit that the probability density $\rho(I|\theta)$ is everywhere finite and differentiable, and obeys the following conditions for every point I_b located on the boundary $\partial\mathcal{M}_\theta$ of the manifold $\mathcal{M}_\theta$, $I_b \in \partial\mathcal{M}_\theta$:

$$\lim_{I\rightarrow I_b} \rho(I|\theta) = \lim_{I\rightarrow I_b} \frac{\partial}{\partial I^i}\rho(I|\theta) = 0. \tag{2.1}$$

The probability density $\rho(I|\theta)$ can be considered to introduce the *differential forces* $\eta_i(I|\theta)$ as follows:

$$\eta_i(I|\theta) = -\frac{\partial}{\partial I^i}\log\rho(I|\theta). \tag{2.2}$$

By definition, the differential forces $\eta_i(I|\theta)$ vanish in those points where the probability density $\rho(I|\theta)$ exhibits its local maxima or its local minima. The global (local) maximum of the probability density can be regarded as a *stable (metastable) equilibrium points* $\bar{I}$, which can be obtained from the following *stationary and stability conditions*:

$$-\frac{\partial}{\partial I^i}\log\rho(\bar{I}|\theta) = 0,\ -\frac{\partial^2}{\partial I^i \partial I^j}\log\rho(\bar{I}|\theta) \succ 0, \tag{2.3}$$

[1]Denomination taken from statistical physics

where $A_{ij} \succ 0$ denotes that the matrix A_{ij} is positive definite. In general, the differential forces $\eta_i(I|\theta)$ characterize the deviation of a given point $I \in \mathcal{M}_\theta$ from these local equilibrium points. Analogously, it is worth to introduce the *response matrix* $\chi_{ij}(I|\theta)$:

$$\chi_{ij}(I|\theta) = \partial_j \eta_i(I|\theta), \tag{2.4}$$

where $\partial_i A = \partial A/\partial I^i$, which describes the response of differential forces $\eta_i(I|\theta)$ under an infinitesimal change of the variable I^j.

As stochastic variables, the expectation values of the differential forces $\eta_i = \eta_i(I|\theta)$ identically vanish:

$$\langle \eta_i \rangle = 0, \tag{2.5}$$

and these quantities also obey the *fundamental and the associated fluctuation theorems*:

$$\left\langle \eta_i \delta I^j \right\rangle = \delta_i^j, \tag{2.6}$$

$$\langle \chi_{ij} \rangle = \langle \eta_i \eta_j \rangle, \tag{2.7}$$

where δ_i^j is the Kronecker delta. These last fluctuation relations are directly derived from the following identity:

$$\langle \partial_i A(I|\theta) \rangle = \langle \eta_i(I|\theta) A(I|\theta) \rangle \tag{2.8}$$

substituting the cases $A(I|\theta) = 1$, I^i and η_i, respectively. Here, $A(I)$ is a differentiable function defined on the continuous variables I with definite expectation values $\left\langle \partial A(I|\theta)/\partial I^i \right\rangle$ that obeys the following the boundary condition:

$$\lim_{I \to I_b} A(I)\rho(I|\theta) = 0. \tag{2.9}$$

Moreover, equation (2.8) follows from the integral expression:

$$\int_{\mathcal{M}_\theta} \frac{\partial \upsilon^j(I|\theta)}{\partial I^j} \rho(I|\theta) dI = \oint_{\partial \mathcal{M}_\theta} \rho(I|\theta) \upsilon^j(I|\theta) \cdot d\Sigma_j - \int_{\mathcal{M}_\theta} \upsilon^j(I|\theta) \frac{\partial \rho(I|\theta)}{\partial I^j} dI \tag{2.10}$$

derived from the **intrinsic exterior calculus** of the manifold $\mathcal{M}_\theta$ and the imposition of the constraint $\upsilon^j(I|\theta) \equiv \delta_i^j A(I|\theta)$. It is easy to realize that the identity (2.5) and the associated fluctuation theorem (2.7) are just the *stationary and stability equilibrium conditions* (2.3) *written in term of statistical expectation values*, respectively. In particular, the positive definite character of the self-correlation matrix $M_{ij}(\theta) = \langle \eta_i(I|\theta)\eta_j(I|\theta)\rangle$ implies the positive definition of the matrix:

$$\langle \partial_i \eta_j(I|\theta)\rangle = \left\langle -\frac{\partial^2}{\partial I^i \partial I^j} \log \rho(I|\theta) \right\rangle \succ 0, \tag{2.11}$$

Remarkably, the fundamental fluctuation theorem (2.6) suggests the *statistical complementarity* of the variable I^i and its conjugated differential force $\eta_i = \eta_i(I|\theta)$. Using the Schwartz inequality $\langle \delta A \delta B\rangle^2 \leq \langle \delta A^2\rangle \langle \delta B^2\rangle$, one obtains the following inequality:

$$\Delta I^i \Delta \eta_i \geq 1, \tag{2.12}$$

where $\Delta x = \sqrt{\langle \delta x^2\rangle}$ is the *statistical uncertainty* of the quantity x. This last result exhibits the same mathematical appearance of *Heisenberg's uncertainty relation* $\Delta q \Delta p \geq \hbar$ *in quantum physical theories* [5, 6, 7]. Equation (2.12) can be generalized considering the inverse $M^{ij}(\theta)$ of the self-correlation matrix of the differential forces $M_{ij}(\theta) = \langle \eta_i(I|\theta)\eta_j(I|\theta)\rangle$. Denoting by $C^{ij}(\theta) = \langle \delta I^i \delta I^j\rangle$ the self-correlation matrix of the stochastic variables I, it is possible to obtain the following matrical inequalities:

$$C^{ij}(\theta) - M^{ij}(\theta) \succeq 0. \tag{2.13}$$

Accordingly, the self-correlation matrix $C^{ij}(\theta)$ of stochastic variables I is inferior bound by the inverse $M^{ij}(\theta)$ of the self-correlation matrix of the differential forces η_i.

2.2. *Inference theory.* Inference theory can be described as the problem of deciding how well a set of outcomes $\mathcal{I} = \left\{ I^{(1)}, I^{(2)}, \ldots I^{(m)} \right\}$ obtained from independent measurements fits a proposed distribution function $dp(I|\theta)$ [2]. This question is fully equivalent to infer the values of control parameters θ from this last experimental information. To make inferences about control parameters, one employs *estimators* $\hat{\theta}^\alpha = \theta^\alpha(\mathcal{I})$, that is, functions on the outcomes $\mathcal{I} \in \mathcal{M}_\theta^m$, where $\mathcal{M}_\theta^m = \mathcal{M}_\theta \otimes \mathcal{M}_\theta \ldots \otimes \mathcal{M}_\theta$ (m-times the external product of the manifold $\mathcal{M}_\theta$). The values of these functions pretend to be the best guess for θ^α.

Let us admit that the probability density $\rho(I|\theta)$ is everywhere differentiable and finite on the manifold $\mathcal{P}$ of control parameters θ. Let us start introducing the statistical expectation values $\langle A(\mathcal{I}|\theta)\rangle$ as follows:

$$\langle A(\mathcal{I}|\theta)\rangle = \int_{\mathcal{M}_\theta^m} A(\mathcal{I}|\theta)\, \varrho(\mathcal{I}|\theta)\, d\mathcal{I}, \tag{2.14}$$

where $d\mathcal{I} = dI^{(1)}dI^{(2)}...dI^{(m)}$ and $\varrho\left(\mathcal{I}|\theta\right)$ is the so-called *likelihood function*:

$$\varrho\left(\mathcal{I}|\theta\right) = \prod_{i=1}^{m} \rho(I^{(i)}|\theta). \tag{2.15}$$

Taking the partial derivative $\partial_\alpha = \partial/\partial\theta^\alpha$ of Eq.(2.14), one obtains the following mathematical identity:

$$\left\langle \partial_\alpha A\left(\mathcal{I}|\theta\right)\right\rangle - \partial_\alpha \left\langle A\left(\mathcal{I}|\,\theta\right)\right\rangle = \left\langle A\left(\mathcal{I}|\,\theta\right)\upsilon_\alpha\left(\mathcal{I}|\,\theta\right)\right\rangle, \tag{2.16}$$

where:

$$\hat{\upsilon}_\alpha = \upsilon_\alpha\left(\mathcal{I}|\,\theta\right) = -\frac{\partial}{\partial\theta^\alpha}\log\rho\left(\mathcal{I}\,|\theta\,\right) \tag{2.17}$$

are the components of the *score vector* $\hat{\upsilon} = \{\hat{\upsilon}_\alpha\}$. Substituting $A\left(\mathcal{I}|\,\theta\right) = 1$ into Eq.(2.16), one arrives at the vanishing of the expectation values of the score vector components:

$$\left\langle \upsilon_\alpha\left(\mathcal{I}|\,\theta\right)\right\rangle = 0. \tag{2.18}$$

Let us consider now any *unbiased estimator* $A(\mathcal{I}|\theta) = \theta^\alpha(\mathcal{I})$ of the parameter θ^α, $\langle\theta^\alpha(\mathcal{I})\rangle = \theta^\alpha$, as well as the score vector component $A(\mathcal{I}|\theta) = \upsilon_\beta(\mathcal{I}|\theta)$. Substituting these quantities into identity (2.16), it is possible to obtain the following results:

$$\left\langle \delta\theta^\alpha\left(\mathcal{I}\right)\upsilon_\beta\left(\mathcal{I}\,|\theta\,\right)\right\rangle = -\delta^\alpha_\beta, \tag{2.19}$$

$$\left\langle \partial_\beta\upsilon_\alpha\left(\mathcal{I}|\theta\right)\right\rangle = \left\langle \upsilon_\alpha\left(\mathcal{I}|\,\theta\right)\upsilon_\beta\left(\mathcal{I}|\,\theta\right)\right\rangle. \tag{2.20}$$

It is easy to realize that the identities (2.18) and (2.20) can be regarded as the stationary and stability conditions of the known method of *maximum likelihood estimators* [2]. According to this method, the best values of the parameters θ should maximize the logarithm of the *likelihood function* $\varrho\left(\mathcal{I}|\theta\right)$ for a given set of outcomes $\mathcal{I}$. Such an exigence leads to the following *stationary and stability conditions*:

$$-\frac{\partial}{\partial\theta^\alpha}\log\varrho\left(\mathcal{I}|\bar{\theta}\right) = 0, -\frac{\partial^2}{\partial\theta^\alpha\partial\theta^\beta}\log\varrho\left(\mathcal{I}|\bar{\theta}\right) \succ 0, \tag{2.21}$$

which should be solved to obtain the *maximum likelihood estimators* $\hat{\theta}^\alpha_{mle} = \bar{\theta}^\alpha(\mathcal{I})$. On the other hand, the identity (2.19) also suggests the statistical complementarity between the estimator $\hat{\theta}^\alpha = \theta^\alpha(\mathcal{I})$ and its conjugated score vector component υ_α. Using the Schwartz inequality, one obtains the following uncertainty-like inequality:

$$\Delta\hat{\theta}^\alpha \Delta\upsilon_\alpha \geq 1. \tag{2.22}$$

This result can be easily improved introducing the inverse matrix $g^{\alpha\beta}(\theta)$ of the self-correlation matrix $g_{\alpha\beta}(\theta) = \langle \upsilon_\alpha(\mathcal{I}|\theta)\upsilon_\beta(\mathcal{I}|\theta)\rangle$ and the auxiliary quantity $X^\alpha = \delta\hat{\theta}^\alpha - g^{\alpha\beta}\upsilon_\beta$. Thus, one can compose the positive definite form:

$$\left\langle (\lambda_\alpha X^\alpha)^2 \right\rangle = \left\langle X^\alpha X^\beta \right\rangle \lambda_\alpha \lambda_\beta \geq 0, \tag{2.23}$$

which leads to the positive definition of the matrix:

$$\left\langle \delta\hat{\theta}^\alpha \delta\hat{\theta}^\beta \right\rangle - g^{\alpha\beta}(\theta) \succeq 0. \tag{2.24}$$

This last inequality is the famous *Cramer-Rao theorem* of inference theory [2, 3] that imposes an inferior bound to the efficiency of unbiased estimators $\hat{\theta}^\alpha$, where the self-correlation matrix $g_{\alpha\beta}(\theta)$:

$$g_{\alpha\beta}(\theta) = \langle \upsilon_\alpha(\mathcal{I}|\theta)\upsilon_\beta(\mathcal{I}|\theta)\rangle \tag{2.25}$$

is the Fisher's information matrix referred to in the introductory section.

2.3. *Analogy between inference theory and fluctuation theory.* Fluctuation theory and inference theory provides two different but complementary characterizations for a given parametric family of distribution functions $dp(I|\theta)$. Apparently, these two statistical frameworks can be regarded as *dual counterpart approaches* because of the great analogy between their main definitions and theorems, which are summarized in Table 1. Here, it was introduced the gradient operators $\partial_i \to \nabla_I$ and $\partial_\alpha \to \nabla_\theta$, the diadic products $\mathbf{A}\cdot\mathbf{B} = A_i B_j \mathbf{e}^i \cdot \mathbf{e}^j$ and $\xi \cdot \psi = \xi_\alpha \psi_\beta \epsilon^\alpha \cdot \epsilon^\beta$ and the Kroneker delta $\delta^i_j \to \mathbf{1}_I$ and $\delta^\alpha_\beta \to \mathbf{1}_\theta$. Remarkably, the analogy between fluctuation theory and inference theory is uncomplete, specifically, in regard to their respective *geometric features*.

The parametric family $dp(I|\theta)$ is expressed in the representations $\mathcal{R}_I$ and $\mathcal{R}_\theta$ of the manifolds $\mathcal{M}_\theta$ and $\mathcal{P}$, respectively. Equivalently, the same parametric family can be also rewritten using the representations $\mathcal{R}_\Theta$ and $\mathcal{R}_\nu$ of the manifolds $\mathcal{M}_\theta$ and $\mathcal{P}$, which implies the consideration of the coordinate changes $\Theta(I) : \mathcal{R}_I \to \mathcal{R}_\Theta$ and $\nu(\theta) : \mathcal{R}_\theta \to \mathcal{R}_\nu$. Under these parametric changes, the Fisher's inference matrix (2.25) behaves as the components of a second rank covariant tensor:

$$g_{\gamma\delta}(\nu) = \frac{\partial\theta^\alpha}{\partial\nu^\gamma}\frac{\partial\theta^\beta}{\partial\nu^\delta} g_{\alpha\beta}(\theta). \tag{2.26}$$

TABLE 1
Analogy between inference theory and fluctuation theory.

Inference theory	**Fluctuation theory**
$\upsilon\left(\mathcal{I}\vert\theta\right)=-\nabla_{\theta}\log\rho\left(\mathcal{I}\vert\theta\right)$	$\eta\left(I\vert\theta\right)=-\nabla_{I}\log\rho\left(I\vert\theta\right)$
$\left\langle\upsilon\left(\mathcal{I}\vert\theta\right)\right\rangle=0$	$\left\langle\eta\left(I\vert\theta\right)\right\rangle=0$
$\left\langle\upsilon\left(\mathcal{I}\vert\theta\right)\cdot\delta\hat{\theta}\right\rangle=-\mathbf{1}_{\theta}$	$\left\langle\eta\left(I\vert\theta\right)\cdot\delta\mathbf{I}\right\rangle=\mathbf{1}_{I}$
$\left\langle\nabla_{\theta}\cdot\upsilon\left(\mathcal{I}\vert\theta\right)\right\rangle=\left\langle\upsilon\left(\mathcal{I}\vert\theta\right)\cdot\upsilon\left(\mathcal{I}\vert\theta\right)\right\rangle$	$\left\langle\nabla_{I}\cdot\eta\left(I\vert\theta\right)\right\rangle=\left\langle\eta\left(I\vert\theta\right)\cdot\eta\left(I\vert\theta\right)\right\rangle$

The existence of these last transformation rules guarantees the invariance of the inference metric (1.1). The relevance of the distance notion (1.1) can be understood considering the asymptotic expression of the distribution function of the most efficient unbiased estimators $\hat{\theta}_{best}(\mathcal{I})$:

$$dQ^{m}(\vartheta|\theta)=\int_{\mathcal{M}^{m}_{\theta}}\delta\left[\vartheta-\hat{\theta}_{best}(\mathcal{I})\right]\varrho(\mathcal{I}|\theta)d\mathcal{I} \tag{2.27}$$

when $m\gg 1$ and $g_{\alpha\beta}(\theta)\propto m$:

$$dQ^{m}(\vartheta|\theta)\simeq\exp\left[-\frac{1}{2}g_{\alpha\beta}(\theta)\Delta\vartheta^{\alpha}\Delta\vartheta^{\beta}\right]\sqrt{\left|\frac{g_{\alpha\beta}(\theta)}{2\pi}\right|}d\vartheta, \tag{2.28}$$

where $\Delta\vartheta^{\alpha}=\vartheta^{\alpha}-\theta^{\alpha}$. Clearly, the distance notion (1.1) provides the *distinguishing probability* between two close distribution functions of the parametric family $dp(I|\theta)$ through the inferential procedure.

At first glance, the analogy between fluctuation theory and inference theory *strongly suggests the existence of a counterpart approach of inference geometry in the framework of fluctuation theory*, that is, the existence of a Riemannian metric (1.3) to characterize the statistical separation between close points I and $I+dI\in\mathcal{M}_{\theta}$.

Unfortunately, the underlying analogy is insufficient to introduce the particular expression of the metric tensor $g_{ij}(I|\theta)$. For example, it is easy to check that the fluctuation theorems (2.5)-(2.7) can be extended to the new coordinate representation $\mathcal{R}_{\Theta}$, e.g.: the associated fluctuation theorem:

$$\left\langle\partial_{i}\eta_{j}(\Theta|\theta)\right\rangle=\left\langle\eta_{i}(\Theta|\theta)\eta_{j}(\Theta|\theta)\right\rangle, \tag{2.29}$$

where $\partial_{i}=\partial/\partial\Theta^{i}$ and $\eta_{i}(\Theta|\theta)$:

$$\eta_{i}(\Theta|\theta)=-\frac{\partial}{\partial\Theta^{i}}\log\rho(\Theta|\theta). \tag{2.30}$$

However, the self-correlation matrix $\tilde{M}_{ij}(\theta)=\left\langle\eta_{i}(\Theta|\theta)\eta_{j}(\Theta|\theta)\right\rangle$ associated with the new representation $\mathcal{R}_{\Theta}$ is not related by *local transformation rules* to its counterpart expression $M_{ij}(\theta) = \langle \eta_i(I|\theta)\eta_j(I|\theta)\rangle$ in the old representation $\mathcal{R}_I$. The self-correlation matrix of the differential forces $M_{ij}(\theta) = \langle \eta_i(I|\theta)\eta_j(I|\theta)\rangle$ is a matrix function defined on the manifold $\mathcal{P}$ of control parameters θ, while the metric tensor $g_{ij}(I|\theta)$ is a *tensorial entity* defined on the manifolds $\mathcal{M}_\theta$ and $\mathcal{P}$.

The introduction of the metric tensor $g_{ij}(I|\theta)$ is not a trivial question. This fact explains why the existence of this counterpart approach of inference geometry is previously unknown in the literature. The above reasonings evidence that the definition of the metric tensor $g_{ij}(I|\theta)$ cannot involve integral expressions over the manifold $\mathcal{M}_\theta$ as the case of Fisher's inference matrix (2.25), but some kind of *covariant differential equations* defined from the probability density $\rho(I|\theta)$ of the parametric family (1.2).

3. Riemannian statistics geometry. Riemannian statistics geometry can be formulated starting from a set of axioms that combine the stochastic nature of the manifold $\mathcal{M}_\theta$ and the notions of differential geometry. This section is devoted to discuss these set of axioms as well as their most direct consequences.

3.1. *Postulates of Riemannian statistics geometry.*

Axiom 1. *The manifold of the stochastic variables $\mathcal{M}_\theta$ possesses a **Riemannian structure**, that is, it is provided of a **metric tensor** $g_{ij}(I|\theta)$ and a **torsionless covariant differentiation** D_i that obeys the following constraints:*

$$D_k g_{ij}(I|\theta) = 0. \tag{3.1}$$

Definition 1. *The Riemannian structure on the manifold $\mathcal{M}_\theta$ allows to introduce the **invariant volume element** as follows:*

$$d\mu(I|\theta) = \sqrt{\left|\frac{g_{ij}(I|\theta)}{2\pi}\right|}dI, \tag{3.2}$$

where $|g_{ij}(I|\theta)|$ denotes the absolute value of the metric tensor determinant.

Axiom 2. *There exist a differentiable scalar function $\mathcal{S}(I|\theta)$ defined on the manifold $\mathcal{M}_\theta$, hereafter referred to as the **information potential**, whose knowledge determines the distribution function $dp(I|\theta)$ of the stochastic variables $I \in \mathcal{M}_\theta$ as follows:*

$$dp(I|\theta) = \exp\left[\mathcal{S}(I|\theta)\right] d\mu(I|\theta). \tag{3.3}$$

Definition 2. *Let us consider an arbitrary curve given in parametric form $I(t) \in \mathcal{M}_\theta$ with fixed extreme points $I(t_1) = P$ and $I(t_2) = Q$. Adopting the following notation:*

$$\dot{I}^i = \frac{dI^i(t)}{dt}, \tag{3.4}$$

*the **length** Δs of this curve can be expressed as:*

$$\Delta s = \int_{t_1}^{t_2} \sqrt{g_{ij}\left[I(t)|\theta\right] \dot{I}^i\left(t\right) \dot{I}^j\left(t\right)} dt. \tag{3.5}$$

Definition 3. *It is say that the curve $I(t) \in \mathcal{M}_\theta$ exhibits an **unitary affine parametrization** when its parameter t satisfies the following constraint:*

$$g_{ij}[I(t)|\theta]\dot{I}^i(t)\dot{I}^j(t) = 1. \tag{3.6}$$

Definition 4. *A **geodesic** is the curve $I_g\left(t\right)$ with minimal length (3.7) between two fixed arbitrary points $(P, Q) \in \mathcal{M}_\theta$. Moreover, the **distance** $\mathfrak{D}_\theta(P, Q)$ between these two points (P, Q) is given by the length of its associated geodesic $I_g(t)$:*

$$\mathfrak{D}_\theta(P, Q) = \int_{t_1}^{t_2} \sqrt{g_{ij}\left[I_g(t)|\theta\right] \dot{I}_g^i\left(t\right) \dot{I}_g^j\left(t\right)} dt. \tag{3.7}$$

Definition 5. *Let us consider a differentiable curve $I(t) \in \mathcal{M}_\theta$ with an unitary affine parametrization. The **information dissipation** $\Phi(t)$ along the curve $I(t)$ is defined as follows:*

$$\Phi(t) = \frac{d\mathcal{S}\left[I(t)|\theta\right]}{dt}. \tag{3.8}$$

Axiom 3. *The length Δs of any interval (t_1, t_2) of an arbitrary geodesic $I_g(t) \in \mathcal{M}_\theta$ with an unitary affine parametrization is given by the negative of the variation of its information dissipation $\Delta\Phi(t)$:*

$$\Delta s = -\Delta\Phi(t) = \Phi(t_1) - \Phi(t_2). \tag{3.9}$$

Axiom 4. *The probability density $\rho(I|\theta)$ associated with distribution function (3.3) vanishes with its first partial derivatives for any point on the boundary $\partial\mathcal{M}_\theta$ of the manifold $\mathcal{M}_\theta$.*

3.2. *Analysis of axioms and their direct consequences.* **Axiom 1** postulates the *Riemannian nature* of the manifold $\mathcal{M}_\theta$, that is, the existence of the metric tensor $g_{ij}(I|\theta)$. Even, this axiom specifies the *Riemannian structure* of the manifold $\mathcal{M}_\theta$ starting from the knowledge of the metric tensor $g_{ij}(I|\theta)$, e.g.: the *covariant differentiation* D_i and the *curvature tensor* $R_{ijkl}(I|\theta)$. As discussed elsewhere [8], equation (3.1) is an strong constraint of Riemannian geometry that determines the *affine connections* Γ^k_{ij} employed to introduce the covariant differentiation D_i, specifically, the so-called *Levi-Civita connection*:

$$\Gamma^k_{ij}\left(I|\theta\right) = g^{km}\frac{1}{2}\left(\frac{\partial g_{im}}{\partial I^j} + \frac{\partial g_{jm}}{\partial I^i} - \frac{\partial g_{ij}}{\partial I^m}\right). \tag{3.10}$$

The knowledge of the affine connections Γ^k_{ij} allows the introduction of the *curvature tensor* $R^l_{ijk} = R^l_{ijk}(I|\theta)$ of the manifold $\mathcal{M}_\theta$:

$$R^l_{ijk} = \frac{\partial}{\partial I^i}\Gamma^l_{jk} - \frac{\partial}{\partial I^j}\Gamma^l_{ik} + \Gamma^l_{im}\Gamma^m_{jk} - \Gamma^l_{jm}\Gamma^m_{ik}, \tag{3.11}$$

which is also derived from the knowledge of the metric tensor $g_{ij}(I|\theta)$ and its first and second partial derivatives. Using the curvature tensor $R^l_{ijk}(I|\theta)$, it is possible to obtain its fourth-rank covariant form $R_{ijkl} = g_{lm}R^m_{ijk}$:

$$\begin{aligned} R_{ijkl} = \frac{1}{2}\left(\frac{\partial^2 g_{il}}{\partial I^j\partial I^k} + \frac{\partial^2 g_{jk}}{\partial I^i\partial I^l} - \frac{\partial^2 g_{jl}}{\partial I^i\partial I^k} - \frac{\partial^2 g_{ik}}{\partial I^j\partial I^l}\right) + \\ + g_{mn}\left(\Gamma^m_{il}\Gamma^n_{jk} - \Gamma^m_{jl}\Gamma^n_{ik}\right), \end{aligned} \tag{3.12}$$

the *Ricci curvature tensor* $R_{ij}(I|\theta)$:

$$R_{ij}(I|\theta) = R^k_{kij}(I|\theta) \tag{3.13}$$

as well as the *curvature scalar* $R(I|\theta)$:

$$R(I|\theta) = g^{ij}(I|\theta)R^k_{kij}(I|\theta) = g^{ij}(I|\theta)g^{kl}(I|\theta)R_{kijl}(I|\theta). \tag{3.14}$$

The curvature scalar $R(I|\theta)$ has a paramount relevance in Riemannian geometry [8] because of it is the only invariant derived from the first and second partial derivatives of the metric tensor $g_{ij}(I|\theta)$.

Axiom 2 postulates the probabilistic nature of the manifold $\mathcal{M}_\theta$, in particular, the existence of the distribution function $dp(I|\theta)$. Simultaneous, this axiom also provides the formal definition of the information potential $\mathcal{S}(I|\theta)$.

The probability density $\rho(I|\theta)$ of the parametric family (1.2) obeys the transformation rule of a tensorial density:

$$\rho(\Theta|\theta) = \rho(I|\theta)\left|\frac{\partial\Theta}{\partial I}\right|^{-1} \tag{3.15}$$

under coordinate change $\Theta(I):\mathcal{R}_I \to \mathcal{R}_\Theta$ of the manifold $\mathcal{M}_\theta$. The covariance of the metric tensor $g_{ij}\left(I|\theta\right)$:

$$g_{ij}\left(\Theta|\theta\right) = \frac{\partial I^m}{\partial\Theta^i}\frac{\partial I^n}{\partial\Theta^j}g_{mn}\left(I|\theta\right) \tag{3.16}$$

implies that the pre-factor of the invariant volume element (3.2) also behaves as a tensorial density:

$$\sqrt{\left|\frac{g_{ij}\left(\Theta|\theta\right)}{2\pi}\right|} = \sqrt{\left|\frac{g_{ij}\left(I|\theta\right)}{2\pi}\right|}\left|\frac{\partial\Theta}{\partial I}\right|^{-1}. \tag{3.17}$$

Admitting that the metric tensor determinant $|g_{ij}(I|\theta)|$ is non-vanishing everywhere, it is possible to introduce *probability weight* $\omega(I|\theta)$:

$$\omega(I|\theta) = \rho(I|\theta)/\sqrt{\left|\frac{g_{ij}\left(I|\theta\right)}{2\pi}\right|}, \tag{3.18}$$

which represents a *scalar function* defined on the manifold $\mathcal{M}_\theta$. Since the manifold $\mathcal{M}_\theta$ possesses a Riemannian structure, it is natural to replace the integration over the usual volume element dI by the invariant volume element $d\mu(I|\theta)$. This consideration allows to rephrase the parametric family (1.2) in the following equivalent representation:

$$dp(I|\theta) = \omega(I|\theta)d\mu(I|\theta), \tag{3.19}$$

which explicitly exhibits the covariance of the distribution function under the coordinate reparametrizations of the manifold $\mathcal{M}_\theta$. The information potential $\mathcal{S}(I|\theta)$ referred to in **Axiom 2** is simply defined by the logarithm of the probability weight $\omega(I|\theta)$:

$$\mathcal{S}(I|\theta) = \log\omega(I|\theta), \tag{3.20}$$

which also represents a scalar function defined on the manifold $\mathcal{M}_\theta$.

By itself, the Riemannian character of the manifold $\mathcal{M}_\theta$ allows to overcome a geometry inconsistence of the notion of *information entropy* employed in *information theory* [9]. The information entropy $S\left[p|\theta\right]$ for a distribution function $p(X_k|\theta)$ for discrete variables $X = \{X_k\}$ is defined as follows:

$$S\left[p|\theta\right]=-\sum_{k}p(X_k|\theta)\log p(X_k|\theta). \tag{3.21}$$

However, its counterpart extension for a distribution function defined on continuous variables like (1.2), the so-called *differential entropy*:

$$S_I\left[p|\theta\right]=-\int_{\mathcal{M}_\theta}\rho(I|\theta)\log\rho(I|\theta)dI, \tag{3.22}$$

undergoes an important inconsistence. In fact, such a definition crucially depends on the coordinate representation of the manifold $\mathcal{M}_\theta$:

$$S_I\left[p|\theta\right]\neq S_\Theta\left[p|\theta\right]=-\int_{\mathcal{M}_\theta}\rho(\Theta|\theta)\log\rho(\Theta|\theta)d\Theta. \tag{3.23}$$

The above inconsistence is simply avoided redefining the information entropy $S\left[p|\theta\right]$ in terms of the probability weight $\omega(I|\theta)$ instead of the probability density $\rho(I|\theta)$:

$$S\left[p|\theta\right]=-\int_{\mathcal{M}_\theta}\omega(I|\theta)\log\omega(I|\theta)d\mu(I|\theta). \tag{3.24}$$

Clearly, this last expression does not depend on the coordinate representation of the Riemannian manifold $\mathcal{M}_\theta$.

For a given probability density $\rho(I|\theta)$, the concrete definition of the information potential $\mathcal{S}(I|\theta)$ depends on the underlying metric tensor $g_{ij}(I|\theta)$ defined on the manifold $\mathcal{M}_\theta$. So far, it was postulate the existence of the metric tensor $g_{ij}(I|\theta)$, but its specific definition is still arbitrary. Such an ambiguity is eliminated after considering the **Axiom 3**, which establishes a direct connection between the distance notion (1.3) and the information dissipation (3.8). In other words, this axiom imposes a strong constraint between the metric tensor $g_{ij}(I|\theta)$ and the information potential $\mathcal{S}(I|\theta)$.

THEOREM 3.1. *The metric tensor $g_{ij}(I|\theta)$ can be identified with the negative of the* ***covariant Hessian*** *$\mathcal{H}_{ij}(I|\theta)$ of the information potential $\mathcal{S}\left(I|\theta\right)$:*

$$g_{ij}\left(I|\theta\right)=-\mathcal{H}_{ij}(I|\theta)=-D_iD_j\mathcal{S}\left(I|\theta\right). \tag{3.25}$$

COROLLARY 1. *The information potential $\mathcal{S}(I|\theta)$ is locally concave everywhere and the metric tensor $g_{ij}(I|\theta)$ is positive definite on the manifold $\mathcal{M}_\theta$.*

PROOF. The searching of the curve with minimal length (3.7) between two arbitrary points (P,Q) is a variational problem that leads to the following ordinary differential equations [8]:

$$\dot{I}^k_g(t)D_k\dot{I}^i_g(t)=\ddot{I}^i_g(t)+\Gamma^i_{mn}\left[I_g(t)|\theta\right]\dot{I}^m_g(t)\dot{I}^n_g(t)=0, \tag{3.26}$$

which describes the geodesic $I_g(t)$ with an unitary affine parametrization. Equations (3.8) and (3.9) can be rephrased as follows:

$$\Delta s=-\Delta\Phi(s)\rightarrow\frac{d\Phi(s)}{ds}=\frac{d^2\mathcal{S}}{ds^2}=-1. \tag{3.27}$$

Taking into account the geodesic differential equations (3.26), the constraint (3.27) can be rewritten as:

$$\frac{d^2\mathcal{S}}{ds^2}=\ddot{I}^k_g\frac{\partial\mathcal{S}}{\partial I^k}+\dot{I}^i_g\dot{I}^j_g\frac{\partial^2\mathcal{S}}{\partial I^i\partial I^j}=\dot{I}^i_g\dot{I}^j_g\left\{\frac{\partial^2\mathcal{S}}{\partial I^i\partial I^j}-\Gamma^k_{ij}\frac{\partial\mathcal{S}}{\partial I^k}\right\}, \tag{3.28}$$

which involves the *covariant Hessian* $\mathcal{H}_{ij}$:

$$\mathcal{H}_{ij}=D_iD_j\mathcal{S}=\frac{\partial^2\mathcal{S}}{\partial I^i\partial I^j}-\Gamma^k_{ij}\frac{\partial\mathcal{S}}{\partial I^k}. \tag{3.29}$$

Equations (3.28)-(3.29) can be combined with constraint (3.6) to obtain the following expression:

$$(g_{ij}+\mathcal{H}_{ij})\dot{I}^i_g\dot{I}^j_g=0. \tag{3.30}$$

Its covariant character leads to Eq.(3.25). Corollary 1, that is, the concave character of the information potential $\mathcal{S}(I|\theta)$ and the positive definition of the metric tensor $g_{ij}(I|\theta)$ are direct consequences of equation (3.27). □

COROLLARY 2. *The metric tensor $g_{ij}\left(I|\theta\right)$ can be obtained from a given information potential $\mathcal{S}\left(I|\theta\right)$ through the following set of covariant differential equations:*

$$g_{ij}=-\frac{\partial^2\mathcal{S}}{\partial I^i\partial I^j}+\frac{1}{2}g^{km}\left(\frac{\partial g_{im}}{\partial I^j}+\frac{\partial g_{jm}}{\partial I^i}-\frac{\partial g_{ij}}{\partial I^m}\right)\frac{\partial\mathcal{S}}{\partial I^k}. \tag{3.31}$$

The admissible solutions derived from the nonlinear problem (3.31) should be everywhere finite and differentiable, including also on boundary of the manifold $\mathcal{M}_\theta$.

Axiom 4 talks about the asymptotic behavior of the distribution function (3.3) for any point I_b on the boundary $\partial\mathcal{M}_\theta$:

$$\lim_{I\rightarrow I_b}\rho(I|\theta)=\lim_{I\rightarrow I_b}\frac{\partial}{\partial I^i}\rho(I|\theta)=0. \tag{3.32}$$

Such conditions play a fundamental role in the character of stationary points (maxima and minima) of the information potential $\mathcal{S}(I|\theta)$.

Remark 1. *The boundary conditions (3.32) are independent from the* admissible coordinate representation $\mathcal{R}_I$ *of the manifold* $\mathcal{M}_\theta$*, and they implies the vanishing of the probability weight* $\omega(I|\theta)$ *on the boundary* $\partial\mathcal{M}_\theta$ *of the manifold* $\mathcal{M}_\theta$*.*

Proof. This remark is a direct consequence of the transformation rule of the probability density (3.15) as well as the ones associated with its partial derivatives:

$$\frac{\partial\rho\left(\Theta|\theta\right)}{\partial\Theta^i}=\frac{\partial I^j}{\partial\Theta^i}\left\{\frac{\partial\rho\left(I|\theta\right)}{\partial I^j}-\rho\left(I|\theta\right)\frac{\partial}{\partial I^j}\log\left|\frac{\partial\Theta}{\partial I}\right|\right\}\left|\frac{\partial\Theta}{\partial I}\right|^{-1} \tag{3.33}$$

under a coordinate change $\Theta(I):\mathcal{R}_I\rightarrow\mathcal{R}_\Theta$ with Jacobian $|\partial\Theta/\partial I|$ finite and differentiable everywhere. Since the metric tensor determinant $|g_{ij}(I|\theta)|$ is non-vanishing everywhere, **Axiom 4** directly implies the vanishing of the probability weight $\omega(I|\theta)$ on the boundary $\partial\mathcal{M}_\theta$ of the manifold $\mathcal{M}_\theta$. □

3.3. *Section summary.* Before to end this section, it is worth to summarize the main partial results so far obtained. The interest of this paragraph was to describe the mathematical recipe to introduce a metric tensor $g_{ij}(I|\theta)$ for Riemannian statistics geometry. Clearly, the way to accomplish this task for Riemannian statistics geometry is much complex than rule to obtain the metric tensor $g_{\alpha\beta}(\theta)$ for inference geometry:

- Starting from the knowledge of the probability density $\rho(I|\theta)$ of parametric family (1.2), one can employ definition (3.18) to obtain the information potential $\mathcal{S}(I|\theta)=\log\omega(I|\theta)$.
- Then, one can consider the covariant differential equations (3.31) to derive the metric tensor $g_{ij}(I|\theta)$.

According to some preliminary analysis, the boundary conditions associated with **Axiom 4** are also important to satisfy in order to guarantee the existence and uniqueness of the solution of the problem (3.31). Due its complexity, this last question cannot be fully analyzed in this work. Instead, it will be discussed some consequences derived from the existence of a given solution $g_{ij}(I|\theta)$.

4. Geometric analysis of distribution functions.

4.1. *Gaussian representation.*

DEFINITION 6. *The covariant form of the* ***gradiental forces*** $\psi_i\left(I|\theta\right)$ *are defined from the information potential* $\mathcal{S}\left(I|\theta\right)$ *as follows:*

$$\psi_i\left(I|\theta\right)=-D_i\mathcal{S}\left(I|\theta\right)\equiv-\partial\mathcal{S}\left(I|\theta\right)/\partial I^i. \tag{4.1}$$

Using the metric tensor $g^{ij}(I|\theta)$, it is possible to obtain its contravariant counterpart $\psi^i(I|\theta)$:

$$\psi^i\left(I|\theta\right)=g^{ij}\left(I|\theta\right)\psi_j\left(I|\theta\right), \tag{4.2}$$

as well as its the square norm $\psi^2=\psi^2(I|\theta)$:

$$\psi^2(I|\theta)=\psi^i\left(I|\theta\right)\psi_i\left(I|\theta\right). \tag{4.3}$$

THEOREM 4.1. *The information potential* $\mathcal{S}(I|\theta)$ *can be expressed in terms of the square norm of the gradiental forces as follows:*

$$\mathcal{S}(I|\theta)=\mathcal{P}(\theta)-\frac{1}{2}\psi^2(I|\theta), \tag{4.4}$$

where $\mathcal{P}(\theta)$ *is a certain function on control parameters* θ *hereafter referred to as the* ***gaussian potential****.*

PROOF. Let us introduce the scalar function $\mathcal{P}(I|\theta)$:

$$\mathcal{P}(I|\theta)=\mathcal{S}(I|\theta)+\frac{1}{2}g^{ij}(I|\theta)\psi_i(I|\theta)\psi_j(I|\theta). \tag{4.5}$$

It is easy to verify that its covariant derivatives:

$$\begin{aligned}D_k\mathcal{P}(I|\theta)=D_kS(I|\theta)+\frac{1}{2}\left\{\psi_i(I|\theta)\psi_j(I|\theta)D_kg^{ij}(I|\theta)+\right.\\ \left.+g^{ij}(I|\theta)\left[\psi_i(I|\theta)D_k\psi_j(I|\theta)+\psi_j(I|\theta)D_k\psi_i(I|\theta)\right]\right\}\end{aligned} \tag{4.6}$$

vanish as direct consequences of the metric tensor properties (3.1) and (3.25), as well as the definition (4.1) of the gradiental forces $\psi_i(I|\theta)$. Since the covariant derivatives of any scalar function are given by the usual partial derivatives:

$$D_k\mathcal{P}(I|\theta)=\frac{\partial}{\partial I^k}\mathcal{P}(I|\theta)=0, \tag{4.7}$$

the scalar function $\mathcal{P}(I|\theta)$ only depends on the control parameters:

$$\mathcal{P}(I|\theta) \equiv \mathcal{P}(\theta). \tag{4.8}$$

This last result leads to equation (4.4). Mathematically speaking, the scalar function (4.5) can be regarded as a first integral of the set of covariant differential equations (3.31). □

Corollary 3. *The value of information potential $\mathcal{S}(I|\theta)$ at all its extreme points derived from the stationary condition:*

$$\psi^2(\bar{I}|\theta) = 0 \tag{4.9}$$

is exactly given by the gaussian potential $\mathcal{P}(\theta)$.

Corollary 4. *The equilibrium distribution function (3.3) admits the following* ***gaussian representation****:*

$$dp(I|\theta) = \frac{1}{\mathcal{Z}(\theta)} \exp\left[-\frac{1}{2}\psi^2(I|\theta)\right] d\mu(I|\theta). \tag{4.10}$$

Here, the factor $\mathcal{Z}(\theta)$ is related to the gaussian potential as follows:

$$\mathcal{P}(\theta) = -\log \mathcal{Z}(\theta), \tag{4.11}$$

which shall be hereafter referred to as the ***gaussian partition function****.*

4.2. *Maximum and completeness theorems.*

Theorem 4.2. *The information potential $\mathcal{S}(I|\theta)$ exhibits a unique stationary point $\bar{I}$, which corresponds to its global maximum.*

Proof. Since the vanishing of the scalar weight of distribution function:

$$\omega(I|\theta) = \exp\left[\mathcal{S}(I|\theta)\right] \tag{4.12}$$

on the boundary $\partial\mathcal{M}_\theta$, as well as its character nonnegative, finite and differentiable on the simply connected compact manifold $\mathcal{M}_\theta$, it is easy to realize that the entropy $\mathcal{S}(I\theta)$ should exhibit at least a stationary point $\bar{I}$ where takes place the stationary condition (4.9). Since the information potential $\mathcal{S}(I\theta)$ is a concave function, its stationary points can only correspond to local maxima. Let us suppose the existence of a least two stationary points $\bar{I}_1$ and $\bar{I}_2$, which can always be connected with a certain geodesic $I_g(t)$. According to constraint (3.27), the information dissipation $\Phi(t)$ is a monotonous function along the curve $I_g(t)$. Therefore, $\Phi(t)$ should exhibits different values at the stationary points $\bar{I}_1$ and $\bar{I}_2$, which is absurdum since the information dissipation $\Phi(t)$ identically vanishes for any stationary point of the information potential $\mathcal{S}(I|\theta)$:

$$\Phi(t) = -\dot{I}^i(t)\psi_i[I(t)|\theta]. \tag{4.13}$$

Consequently, there exist only one stationary point that corresponds with the global maximum of the information potential $\mathcal{S}(I|\theta)$. □

THEOREM 4.3. *Any hyper-surface of constant information potential $\mathcal{S}(I|\theta)$ is just the boundary of a n-dimensional sphere $S^n(\bar{I}, \ell) \subset \mathcal{M}_\theta$ centered at the point $\bar{I}$ with global maximum information potential, where n is the dimension of the manifold $\mathcal{M}_\theta$. Moreover, the information potential $\mathcal{S}$ depends on the radio ℓ of this n-dimensional sphere as follows:*

$$\mathcal{S} = \mathcal{P}(\theta) - \frac{1}{2}\ell^2, \tag{4.14}$$

PROOF. By definition, the vector field $\upsilon^i(I|\theta)$:

$$\upsilon^i(I|\theta) = \frac{\psi^i(I|\theta)}{\psi(I|\theta)} \tag{4.15}$$

is the unitary normal vector of the hyper-surface with constant information potential $\mathcal{S}(I|\theta)$. It is easy to verify that the vector field $\upsilon^i(I|\theta)$ obeys the geodesic equations (3.26):

$$\upsilon^k(I|\theta)D_k\upsilon^i(I|\theta) = \frac{\upsilon^k(I|\theta)}{\psi(I|\theta)}\left[\delta^i_k - \upsilon^i(I|\theta)\upsilon_k(I|\theta)\right] = 0. \tag{4.16}$$

Hence, $\upsilon^i(I|\theta)$ can be regarded as the tangent vector:

$$\frac{dI^i_g(s|\mathbf{e})}{ds} = \upsilon^i[I_g(s|\mathbf{e})|\theta] \tag{4.17}$$

of geodesic family $I_g(s|\mathbf{e})$ with unitary affine parametrization centered at the point $\bar{I}$ with maximum scalar entropy $\mathcal{S}(I|\theta)$, $I_g(s = 0|\mathbf{e}) = \bar{I}$, where the parameters $\mathbf{e}$ distinguish geodesics with different directions at the origin. The information dissipation $\Phi(s|\mathbf{e})$ along any of these geodesics is given by the negative of the norm of the gradiental forces:

$$\Phi(s|\mathbf{e}) = -\frac{dI^i(s|\mathbf{e})}{ds}\psi_i\left[I_g(s|\mathbf{e})|\theta\right] = -\psi[I_g(s|\mathbf{e})|\theta]. \tag{4.18}$$

Considering equation (3.9), the norm $\psi(I|\theta)$ can be related to the length Δs of the geodesic connecting the point I with point $\bar{I}$ with maximum information potential, that is, the *distance* $\mathfrak{D}_\theta(I,\bar{I})$ between the points I and $\bar{I}$:

$$\psi(I|\theta)=\mathfrak{D}_\theta(I,\bar{I}). \tag{4.19}$$

According to the gaussian decomposition (4.4), the hyper-surface with constant information potential $\mathcal{S}(I|\theta)$ is also the hyper-surface where the norm of gradiental generalized forces $\psi(I|\theta)$ is kept constant, that is, the boundary of a n-dimensional sphere $S^n(\bar{I},\ell)$ centered at the point $\bar{I}$ with maximum information potential. □

COROLLARY 5. *The distribution function (3.3) can be expressed in the following* ***Riemannian gaussian representation****:*

$$dp(I|\theta)=\frac{1}{\mathcal{Z}(\theta)}\exp\left[-\frac{1}{2}\ell^2(I)\right]d\mu(I|\theta), \tag{4.20}$$

where $\ell(I)=\mathfrak{D}_\theta(I,\bar{I})$ is the separation distance between the points $\bar{I}$ and I. Consequently, the knowledge of the metric tensor $g_{ij}(I|\theta)$ and the point $\bar{I}$ with maximum information potential $\mathcal{S}(I|\theta)$ fully determines the distribution function $dp(I|\theta)$.

COROLLARY 6. *For points I asymptotically close to the point $\bar{I}$ with maximum information potential $\mathcal{S}(I|\theta)$, the distribution function (3.3) admits the following gaussian approximation:*

$$dp(I|\theta)\simeq\exp\left[-\frac{1}{2}g_{ij}(\bar{I}|\theta)\Delta I^i\Delta I^i\right]\sqrt{\left|\frac{g_{ij}(\bar{I}|\theta)}{2\pi}\right|}dI. \tag{4.21}$$

PROOF. For points I asymptotically close to the point $\bar{I}$ with maximum information potential $\mathcal{S}(I|\theta)$, the separation distance $\ell(I)=\mathfrak{D}_\theta(I,\bar{I})$ between the points $\bar{I}$ and I admits the following approximation:

$$\ell^2(I)\simeq g_{ij}(\bar{I}|\theta)\Delta I^i\Delta I^i, \tag{4.22}$$

where $\Delta I^i=I^i-\bar{I}^i$, which can be directly obtained from definition (3.7). In this approximation level, the normalization condition implies the following estimation for gaussian partition function $\mathcal{Z}(\theta)\simeq 1$. □

4.3. *Section summary.* The radio ℓ of the n-dimensional sphere $S^n(\bar{I},\ell)$ of the **theorem 4.3** and the invariant volume element $d\mu(I|\theta)$ are purely *geometric notions* derived from the knowledge of the metric tensor $g_{ij}(I|\theta)$ and the point $\bar{I}$ with maximum information potential $\mathcal{S}(I|\theta)$. Therefore, *the information potential $\mathcal{S}(I|\theta)$ and all its associated quantities represents geometric notions derived from the Riemannian structure of the manifold $\mathcal{M}_\theta$.* At first glance, the approximation formula (4.21) can be regarded as a *counterpart expression* of the asymptotic distribution (2.28) of inference theory. Expressions (4.21) clarifies the statistical relevance of the distance notion (1.3): this geometric measure is associated with the *occurrence probability* of a fluctuation $\Delta I = I - \bar{I}$ around the point $\bar{I}$ with maximum information potential $\mathcal{S}(I|\theta)$. Interestingly, equation (4.21) admits the exact improvement (4.20). This is a remarkable fact taking into consideration that the metric tensor $g_{ij}(I|\theta)$ of Riemannian statistics geometry is more difficult to calculate than the metric tensor $g_{\alpha\beta}(\theta)$ of inference geometry.

5. Final remarks and open problems. As already discussed, Riemannian statistics geometry constitutes a counterpart approach of inference geometry. The introduction of this geometry approach allows to rephrase a generic parametric family of distribution functions $dp(I|\theta)$ in terms of purely geometric notions derived from the Riemannian structure of the manifold $\mathcal{M}_\theta$ of stochastic continuous variables I. Consequently, this connection allows an alternative application of the powerful tools of Riemannian geometry for the statistical analysis.

There exist important problems that deserves a careful attention in future works. For example, one of the most important notions of Riemannian geometry in the *curvature notion* of the manifold $\mathcal{M}_\theta$. It is natural to expect that this geometry concept also plays a fundamental role from the viewpoint statistical nature of the manifold $\mathcal{M}_\theta$. At first glance, the curvature notion should be associated in some way with the existence of *irreducible* (absolute) *correlations* among the variables I entering in the statistical description. In fact, both curvature and correlation notions can be only defined for a manifold $\mathcal{M}_\theta$ with dimension $n \geq 2$. Although this conjecture is very suggestive, the same one should be rigorously demonstrate in the framework of Riemannian statistics geometry. On the other hand, it is worth remarking that the underlying analogy between inference theory and fluctuation theory allows us to conjecture the following improvement:

$$dQ^m(\vartheta|\theta) = \frac{1}{\mathcal{Z}(\theta)} \exp\left[-\frac{1}{2}\ell^2(\vartheta)\right] \sqrt{\left|\frac{g_{\alpha\beta}(\vartheta)}{2\pi}\right|} d\vartheta \tag{5.1}$$

for the asymptotic distribution (2.28) of inference theory. Here, $\ell(\vartheta)$ should represent the distance between the points ϑ and θ calculated with the metric tensor $g_{\alpha\beta}(\theta)$ defined on the manifold $\mathcal{P}$ of control parameters θ. Of course, this conjecture should be carefully analyzed in the framework of inference geometry.

Acknowledgements. Velazquez thanks the financial support of CONICYT/Programa Bicentenario de Ciencia y Tecnología PSD **65** (Chilean agency).

References.

[1] Amari Sh. (1990) *Geometrical Methods in Statistics: 22 Lecture notes in statistics*, Springer Verlag.
[2] Fisher R. A. (1922) *Philosophical Transactions, Royal Society of London*, (A), **222**, 309-368.
[3] Rao C. R. (1945) *Bull. Calcutta Math. Soc.* **37**, 81-91.
[4] Velazquez L and Curilef S (2010) *J. Stat. Mech.* **P12031**.
[5] Hilgevoord J. and Uffink J. (2001) *The uncertainty principle*, in *The Stanford encyclopedia of philosophy*, (ed.) Zalta E. N.
[6] Heisenberg W. (1927) *Zeitschrift für Physik* **43** 172.
[7] Kennar E. H. (1927) *Zeitschrift für Physik* **44** 326.
[8] Berger A. (2002) *A panoramic view of Riemannian geometry* , Springer.
[9] Thomas M. Cover, Joy A. Thomas. (1991) *Elements of Information Theory* Wiley, New York

Departamento de Física, Av. Angamos 0610, Antofagasta, Chile
E-mail: lvelazquez@ucn.cl