UNIVERSITAT DE BARCELONA

INTRINSIC ANALYSIS OF THE STATISTICAL ESTIMATION

by

J.M. Oller and *J.M. Corcuera*

AMS 1980 Subjects Classifications: 62F10, 62B10, 62A99

Mathematics Preprint Series No. 212
July 1996

Intrinsic Analysis of the Statistical Estimation

J.M. Oller and J.M. Corcuera
Departament d'Estadística
Universitat de Barcelona
08028 Barcelona, Spain

January 1994

Abstract

The parametric statistical models with suitable regularity conditions have a natural Riemannian manifold structure, given by the information metric. Since the parameters are merely labels for the probability measures, an inferential statement should be formulated through intrinsic objects, invariant under reparametrizations. In this context the estimators will be random objects valued on the manifold corresponding to the statistical model. In spite of these considerations, classical important measures of an estimator's performance, like the bias and the mean square error, are clearly dependent of the statistical model parametrizations.

In this paper the concept of moment of a random variable is extended to random fields on an n-dimensional C^∞ real manifold, and the notion of mean value is extended to random objects which take values on a Hausdorff and connected manifold, equipped with an affine connection. In particular, the Riemannian manifold case is considered. This extension is applied to the bias and the mean square error study in statistical point estimation theory.

Under this aproach several basic results are obtained: local and global lower bounds for the mean square of the Rao distance, the invariant measure analogous to the mean square error, which depend on the intrinsic bias and the curvature of the statistical model. Also the behaviour of the mean square of the Rao distance of an estimator when conditioning respect to a sufficient statistic is considered, obtaining intrinsic versions of the Rao-Blackwell and Lehmann-Scheffée theorems. Asymptotic properties complete the study.

AMS 1980 subjects classifications: 62F10, 62B10, 62A99.

Keywords and phrases: Bias, mean square Rao distance, mean value, Riemannian centre of mass, moments, information metric, Cramér–Rao lower bounds, Cramér–Rao type integral inequalities, Rao–Blackwell, Lehmann–Scheffée.

1 Introduction

Estimation can be defined as the theory that concerns making inductions from the data and inferences about inductions. In parametric statistical estimation theory we make inductions, from the data set, by proposing probability measures that belongs to a parametric family, the parameters being only a name and playing no role in the induction process. The inferences are usually in the form of point and intervales estimates and no matter what specific inferences may eventually be needed. In this approach estimators supply different methods of induction.

On the other hand, as it is well known, the bias and the mean square error are the most commonly used measures of performance of an estimator. These concepts are clearly dependent on the coordinate system or model parametrization. No difficulty arises from this as long as closely related properties, like unbiasedness or uniformly minimum variance estimation are preserved under coordinate system transformations. Unfortunately, this is not the case, essentially due to the non tensorial character of the bias and the mean square error.

Example 1.1 Let a statistical model be defined through the parametric family of densities

$$p(x;\alpha,\beta) = \frac{\beta^\alpha x^{\alpha-1}}{\Gamma(\alpha)}\exp\{-\beta x\} \quad x,\beta \in \mathbb{R}^+,$$

where $\alpha > 0$ is a known constant. Taking into account that, for a sample size k,

$$S = \sum_{i=1}^{k} X_i$$

is a sufficient statistic for the model, the estimator, say estimator $\mathcal{W}$, which is unbiased and UMV for β, with $k\alpha > 1$, is given by

$$\beta(\mathcal{W}) = \frac{k\alpha - 1}{S}$$

But parametrizing the same statistical model as

$$p(x;\alpha,\lambda) = \frac{x^{\alpha-1}}{\lambda^\alpha\Gamma(\alpha)}\exp\{-\frac{x}{\lambda}\} \quad x,\lambda \in \mathbb{R}^+,$$

where $\alpha > 0$ is again a known constant, the above estimator $\mathcal{W}$ would give the estimation for λ

$$\lambda(\mathcal{W}) = \frac{S}{k\alpha - 1}$$

which is biased. The corresponding estimator, say $\mathcal{U}$ which is now unbiased and UMV for λ is given by

$$\lambda(\mathcal{U}) = \frac{S}{k\alpha}$$

However, if we used this estimator to give an estimation of β we would obtain

$$\beta(\mathcal{U}) = \frac{k\alpha}{S}$$

which is biased.

Furthermore, if we compute the mean square error of these estimators, $\mathcal{W}$ and $\mathcal{U}$, under both parametrizations, the following table, assuming $k\alpha > 2$, summarizes the above discussion:

Parameters	Estimators $\mathcal{W}$	Estimators $\mathcal{U}$
β	$\text{Bias}(\beta(\mathcal{W})) = 0$	$\text{Bias}(\beta(\mathcal{U})) = \frac{\beta}{\alpha - 1}$
	$\text{MSE}(\beta(\mathcal{W})) = \frac{\beta^2}{k\alpha - 2} <$	$\text{MSE}(\beta(\mathcal{U})) = \frac{(k\alpha + 2)\beta^2}{(k\alpha - 1)(k\alpha - 2)}$
	(Not attaining Cramér-Rao lower bound: $\frac{\beta^2}{k\alpha}$)	
	Estimator $\mathcal{W}$ is preferable to Estimator $\mathcal{U}$	
λ	$\text{Bias}(\lambda(\mathcal{W})) = \frac{\lambda}{k\alpha - 1}$	$\text{Bias}(\lambda(\mathcal{U})) = 0$
	$\text{MSE}(\lambda(\mathcal{W})) = \frac{(k\alpha + 1)\lambda^2}{(k\alpha - 1)^2} >$	$\text{MSE}(\lambda(\mathcal{U})) = \frac{\lambda^2}{k\alpha}$
	(Attaining Cramér-Rao lower bound: $\frac{\lambda^2}{k\alpha}$)	
	Estimator $\mathcal{U}$ is preferable to Estimator $\mathcal{W}$	

This example show some problems, paradoxes or inconsistencies of classical unbiased minimun variance estimation, essentially due to the dependence on the coordinate system or model parametrization of this statistical criterion. Therefore seems desirable to modify these classical notions in an intrinsic or invariant way.

In this situation a natural question arises: could analogous notions to the bias and the mean square error be formulated depending only on the estimation procedure employed? There are several ways to attempt to achieve this purpose. First, we may try to choose a privileged coordinate system, but it would be difficult to justify the choice. Second, we may define a loss function, extrinsic to the statistical model, invariant under reparametrizations, and proceed as in Lehmann, see [24]. This may be a reasonable procedure from a decision theoretical point of view, but for statistical inference purposes it might be better to work exclusively with concepts intrinsic to statistical model.

The aim of that we shall refer to as *Intrinsic Analysis* of the statistical estimation, is to develop a statistical estimation theory analogous to the classical one, based on geometrical structures of the statistical models. Then one goal of the Intrinsic Analysis is to supply invariant tools in order to analyse the performance of an estimator, and another is to obtain results that are analogous to classical ones and to establish relationships between the classical non invariant measures and the invariant herein obtained.

In this paper, taking into account the Riemannian structure of the regular parametric statistical models, an *intrinsic* bias measure is obtained by considering the mean value of random manifold–valued maps. The mean square of the Riemannian, or Rao, distance is the invariant analogous to the mean square error.

The first part of the paper is concerned with the moments of a random field on an n–dimensional C^∞ real manifold, and also the mean value concept of a random object which takes values on a (Hausdorff and connected) manifold equipped with an affine connection, through the exponential map. We emphasize the analogies and differences between moments and mean values, and we consider, in particular, the Riemannian case.

The second part is the application of these results to the study of some invariant measures analogous to the bias and mean square error corresponding to a statistical estimator. The third and fourth parts are devoted to the development of intrinsic versions of the local and global Cramér-Rao lower bounds. In the fifth part we study the behaviour of the mean square Rao distance of an estimator when it is conditioned by a sufficient statistic, in order to obtain intrinsic versions of the Rao-Blackwell and Lehmann–Scheffée theorems. Finally some asymptotic properties, specially related with the maximum–likelihood estimator, are studied.

2 Moments and mean values

Let $(\mathcal{X}, \mathcal{A}, P)$ be a probability space, where $\mathcal{X}$ is the sample space, $\mathcal{A}$ is a σ-algebra of subsets of $\mathcal{X}$ and P is a probability measure on $\mathcal{A}$. Let $(M, \mathfrak{A})$ be an n-dimensional C^∞ real manifold, where $\mathfrak{A}$ is the atlas for M.

Let f be a measurable map, $f : \mathcal{X} \to M$, also called a *random object on* M, that is, a map such that for all open sets $W \subset M$, $f^{-1}(W) \in \mathcal{A}$. We will now introduce the notion of mean value and moments of f, assuming the fewest necessary asumptions and maintaining the intuitive notion of *centrality* measure, in a closely related idea of *center of mass* as we shall see later (see Karcher [19], Kobayashi and Nomizu [22], Kendall [21] and Hendriks [15]).

If there exists a global chart (M, ϕ) we may try to define the mean value of f as:

$$E(f) = \phi^{-1}\left(\int_{\mathcal{X}} (\phi \circ f)(x)\, P(dx)\right),$$

but this naive approach it is not satisfactory since $E(f)$ would be dependent, in general, on the coordinate system. Only linear transformations would preserve $E(f)$.

In order to solve this problem, let us first introduce some concepts. Let A be a set of M, and $\mathcal{F}_A^{(p,q)}$ the set of all C^∞ tensor fields in any open subset of A, of order $p+q$, p times contravariant and q times covariant. If we fix $m \in A$, any map X from $\mathcal{X}$ to $\mathcal{F}_A^{(p,q)}$ induces a map X_m, such that $X_m : \mathcal{X} \to T_q^p(M_m)$ with $X_m(x) = (X(x))_m$, where $T_q^p(M_m)$ denotes the space of (p,q)-tensors on the tangent space at m, M_m, having a natural topological vector space structure. Considering the Borel σ-algebra on $\mathcal{F}_A^{(p,q)}$ induced by the Borel σ-algebras of the M_m, a simple definition follows,

Definition 2.1 *A* C^∞ random (p,q)-tensor field *on* A, X, *is a measurable map from* $\mathcal{X}$ *to* $\mathcal{F}_A^{(p,q)}$.

It follows from the definition that $\forall m \in A$, the induced map X_m is a measurable map on $(\mathcal{X}, \mathcal{A})$.

Moreover, any random tensor field may be characterized by its $n^{(p+q)}$ components with respect to any coordinate system, $\theta^1, \ldots, \theta^n$,

$$X^{\alpha_1,\ldots,\alpha_p}_{\beta_1,\ldots,\beta_q}(x; \theta^1, \ldots, \theta^n) \qquad \alpha_1, \ldots, \alpha_p,\ \beta_1, \ldots, \beta_q = 1, \ldots, n,$$

which are clearly fixed x, C^∞ functions of $\theta^1, \ldots, \theta^n$, and, fixed θ, real valued measurable functions on $(\mathcal{X}, \mathcal{A})$.

Let $\otimes$ stand for the tensor field product. In the present context it is natural to define:

Definition 2.2 *The* k-order moment *of the random tensor field* X *is an ordinary* (kp, kq)*-tensor field on* A *defined by*

$$\mathcal{M}^k(X) = \int_{\mathcal{X}} X(x) \overbrace{\otimes \ldots \otimes}^{k} X(x)\, P(dx)\,, \quad k \in \mathbb{N},$$

provided the existence of the above integral.

Notice that $\mathcal{M}^k(X)$ may be computed explicitly through its components in any coordinate system. The components of $\mathcal{M}^k(X)$, with respect to a coordinate system $\theta^1, \ldots, \theta^n$, will be given by

$$M^{\alpha_1^1,\ldots,\alpha_p^k}_{\beta_1^1,\ldots,\beta_q^k}(\theta) = \int_{\mathcal{X}} \left(X^{\alpha_1^1,\ldots,\alpha_p^1}_{\beta_1^1,\ldots,\beta_q^1}(x;\theta) \cdot \ldots \cdot X^{\alpha_1^k,\ldots,\alpha_p^k}_{\beta_1^k,\ldots,\beta_q^k}(x;\theta) \right) P(dx).$$

This is in fact the simplest and also the most natural extension of the k-order moment to a random tensor field. In particular, the 1-order moment is the expectation of X.

We can also write

$$\mathcal{M}^k(X) = E(X \overbrace{\otimes \ldots \otimes}^{k} X),$$

where the tensor product of random tensor fields is naturally defined from the tensor product of ordinary tensor fields.

In a similar way we could define the *central moments*, which exhibit classical properties, for instance:

$$\mathcal{M}_c^2(X) = E((X - EX) \otimes (X - EX)) = E(X \otimes X) - E(X) \otimes E(X).$$

In the case that X is a vector field, the components of this tensor, with respect to a coordinate system, may be written in matrix notation, obtaining the covariance matrix, Σ_X,

$$\Sigma_X = E(XX') - E(X)E(X)',$$

identifying, in the previous equation, the vectors with their components and X being a column vector and X' the corresponding row vector.

In order to consider the mean value of a random object, measurable map, which takes values on a C^∞ real manifold, we have to introduce an additional structure on the manifold: we shall assume that it is equipped with an affine connection. Typical examples of manifolds with an affine connection are Riemannian manifolds.

Associated with an affine connection there is a map, called the exponential map $\exp_p : M_p \to M$ It is defined for all v in an open star-shaped neighbourhood of $0_p \in M_p$.

Additionally it is also well known that this map, in general, has no inverse, although there are important particular cases where one exists. Nevertheless, we can always restrict the map in an open neighbourhood of $0_p \in M_p$, such that the inverse is well defined, the exponential map being a local diffeomorphism. Further information can be found in 9.1 of the Appendix.

Let us precise the kind of neighbourhoods that we consider suitable to define the mean value of a random object.

Definition 2.3 *A neighbourhood $W(p)$ of $p \in M$ is said to be* normal *if $W(p)$ is the diffeomorphic image, by the exponential map, of an open star-shaped neighbourhood of $0_p \in M_p$.*

Notice that a normal neighbourhood $W(p)$ of p has the property that every $q \in W(p)$ can be joined to p by a *unique* geodesic in $W(p)$.

In the vector space M_p we will consider star-shaped neighbourhoods, $V(p)$, such that $V(p) = -V(p)$, in the case we have only an affine connection, and balls in the Riemannian case. They shall be referred to as *balls* with center 0_p, even in the affin case.

Definition 2.4 *The image, $W(p)$, by the exponential map, of a open ball $V(p)$ with center 0_p, is said to be a* normal ball *with center p if $W(p)$ is a normal neighbourhood of p.*

Notice that, in the Riemannian case, the *shortest* geodesic that joins p with any $q \in W(p)$, $W(p)$ being a normal ball with center p, is unique in M and lies in $W(p)$. However we can consider more general neighbourhoods with this property.

Definition 2.5 *An open set $W(p)$ is said to be a* regular normal *neighbourhood of p if and only if its intersection with any normal ball with center p remains normal.*

In the Riemannian case we can assure the existence of these kind of neighbourhoods. Since every point p has a neighbourhood where the exponential map is a diffeomorphism we can obtain a normal neighbourhood of $0_p \in M_p$. Let $V(p)$ be the corresponding star-shaped neigbourhood in M_p. Then, consider some ball with center $0_p \in M_p$ where the exponential map is injective. If we restrict the map to the intersection with $V(p)$ we obtain a regular normal neighbourhood of p. It is easy to see that in the Riemannian case a neighbourhood $W(p)$ of p is regular normal if and only if the shortest geodesic that joins p with any other point in $W(p)$ is unique and lies in $W(p)$, then the regular normal neighbourhoods are a generalization of neighbourhoods with these property to the affin case.

Denote by $\mathfrak{S}_p = \{\xi \in M_p : ||\xi||_p = 1\}$, and for each $\xi \in \mathfrak{S}_p$ we define

$$C_p(\xi) = \sup\{s > 0 \ : \ \rho\,(p, \gamma_\xi(s)) = s\},$$

where ρ is the Riemannian distance and γ_ξ is a geodesic defined in an open interval containing zero, such that $\gamma_\xi(0) = p$ and with tangent vector equal to ξ at the origin. Then if we set

$$\mathfrak{D}_p = \{s\xi \in M_p : 0 \leq s < C_p(\xi) \ ; \ \xi \in \mathfrak{S}_p\}$$

and

$$D_p = \exp_p(\mathfrak{D}_p),$$

it is known that $\exp_p$ maps $\mathfrak{D}_p$ diffeomorphically onto D_p, see 9.1. of the Appendix. In fact D_m is the maximal regular neighbourhood of m in the sense that any other regular neighbourhood of m is included in it.

Then, given a random variable f taking values on a (Hausdorff and connected) manifold, equipped with an affine connection (which may be the Levi-Civita connection corresponding to a Riemannian manifold), there is a natural way to define a random vector, fixed $p \in M$, given by $\exp_p^{-1}(f(x))$. This vector is not necessarily defined for all $x \in \mathcal{X}$, but if it is defined almost surely, we can introduce the following *mean value* concept,

Definition 2.6 *A point on the manifold $p \in M$ is a* mean value *of the random variable f and we shall write $p = \mathfrak{M}\,(f)$, if and only if there is a regular normal neighbourhood of p where f takes values almost surely $[P]$, and we have*

$$\int_{\mathcal{X}} \exp_p^{-1}(f(x)) \ P(dx) = 0_p.$$

Let us remark that this is an *intrinsic mean value* definition, independent of the coordinate system.

Moreover, in the case where M is a complete Riemannian manifold, if P_f is the probability measure induced by the measurable map in M, and P_f is dominated by the Riemmannian measure, for any $p \in M$ we will have a regular normal neighbourhood of p, with probability $[P]$ equal to one, where $\exp_p^{-1}(f(x))$ will be defined. This is an immediate consequence of the image measure theorem and that the cut locus of p in M is a Riemannian measure zero set.

In the following we will use the notation $\exp_p^{-1}(\cdot)$ to indicate the inverse of the exponential in some regular normal neighbourhood of p.

We shall consider now several examples.

Example 2.7 Let M be $\mathbb{R}^n$. Identifying the points with their coordinates corresponding to the trivial chart, and considering the usual Euclidean affine connection, we have, for $z, m \in \mathbb{R}^n$, that $\exp_m^{-1}(z) = (z - m)_m$. In order to find the mean value of a random variable f we have to solve the following equation

$$\int_{\mathcal{X}} (f(x) - m)_m \, P(dx) = 0_m,$$

but this equation has the unique, trivial solution

$$m = \int_{\mathcal{X}} f(x) \, P(dx),$$

provided the existence of the latter integral. Therefore we recover the classical definition

$$\mathfrak{M}(f) = E(f) = \int_{\mathcal{X}} f(x) \, P(dx).$$

Moreover, the second order central moment of $\exp_m^{-1}(f(x))$ can be written, in matrix notation and omitting the subindex m, as

$$\begin{aligned} \Sigma_f &= \mathcal{M}_c^2(\exp_m^{-1}(f(x))) = E\left((f(x) - m)(f(x) - m)'\right) \\ &= E(ff') - E(f)E(f)', \end{aligned}$$

which is the usual covariance matrix.

Example 2.8 Another interesting example is given by considering the mean values of the Von Mises distribution. In this case the manifold is the unit n–dimensional Euclidean sphere, with the connection induced by the natural embedding into the Euclidean space $\mathbb{R}^n$. The probability measure induced in the manifold is absolutely continuous with respect to the surface measure on the sphere and the corresponding density function (Radon-Nikodym derivative) is given by

$$p(x; \xi, \lambda) = \alpha_n(\lambda) \exp\{\lambda \xi' x\} \quad x, \xi \in S_n = \{z \in \mathbb{R}^n : z'z = 1\}, \ \lambda \in \mathbb{R}^+,$$

where $\alpha_n(\lambda) = \lambda^{k/2-1}/(2\pi)^{k/2} I_{k/2-1}(\lambda)$ is a normalization constant, $I_{k/2-1}$ being the modified Bessel function of the first kind and order $k/2 - 1$. In this case it is clear the existence of two mean values, given by ξ and $-\xi$. Compare this result with the *mean direction* defined in Mardia *et al.* [23, 424-451]. See also Jupp and Mardia [18], for a comprehensive exposition.

Example 2.9 Consider a random variable uniformly distributed on a circle, with the connection induced by the natural embedding into the Euclidean manifold $\mathbb{R}^2$. Then, *all* points on the circle are mean values.

We would supply, in the Riemannian case, a scalar *dispersion* measure with respect to a mean value m: the ordinary expected value of the Riemannian distance square between $f(x)$ and m, which may be regarded as an invariant version, independent of the coordinate system, of the variance of a real random variable. It is also possible to define a dispersion measure with respect to an arbitrary reference point of a Riemannian manifold, as the mean value of the square of the Riemannian distance between $f(x)$ and the selected reference point.

We may also observe that, with this extension of the concept of mean value, we maintain the intuitive and appealing meaning of *centrality measure*, even though we do not have the *linear* properties of the expectation. However, this is natural since we cannot identify, in general, M with its tangent spaces. Similarly we will have a dissociation between the mean value and the concept of first order moment. The moments of a random map f, which takes values on M, should be defined as

Definition 2.10 *The k-*order moment *of the random map f is an ordinary $(k,0)$-tensor field on A defined by*

$$\mathcal{M}^k(f)_m = \int_X \exp_m^{-1}(f(x)) \overbrace{\otimes \ldots \otimes}^{k} \exp_m^{-1}(f(x))\; P(dx)\,, \quad \forall m \in A,\ \ k \in \mathbb{N}$$

provided the existence of the above integral.

There is a relationship between the defined *mean value* and the classical *center of mass*, $\mathfrak{C}$,

$$\mathfrak{C} = \underset{m \in M}{\arg\min}\ \ \mathcal{H}_f(m),$$

where $\mathcal{H}_f(m) = \int_X \rho^2(m, f(x))\, P(dx)$. First of all we have the following propositions:

Proposition 2.11 *If exists some $m_0 \in M$ such that $\mathcal{H}_f(m_0)$ is defined, then the function $\mathcal{H}_f(m)$ is defined for all $m \in M$.*

Proof:

By the triangular inequality

$$\mathcal{H}_f(m) \leq 2\mathcal{H}_f(m_0) + 2\rho^2(m, m_0),$$

and proposition follows. ∎

Suppose it exists a point m_0 such that $\mathcal{H}_f(m_0) < \infty$, then:

Proposition 2.12 *The fuction $\mathcal{H}_f(m)$ is differentiable and*

$$X_m\mathcal{H}_f = -2\langle X_m, \int_{\mathcal{X}} \exp_m^{-1}(f(x))\, P(dx)\rangle.$$

whenever $\exp_p^{-1}(\cdot)$ *is well defined for all* $p \in M$ *almost surely-P.*

Proof:
For all $X_m \in M_m$, since, fixing $q \in M$, $\rho^2(\cdot, q)$ is a C^∞ function, we can write

$$\begin{aligned} X_m\,\rho^2(\cdot, q) &= X_m\|\exp_{(\cdot)}^{-1}(q)\|^2 = 2\langle\nabla_{X_m} \exp_{(\cdot)}^{-1} q, \exp_m^{-1}(q)\rangle \\ &= -2\langle X_m, \exp_m^{-1}(q)\rangle, \end{aligned}$$

where the last equality can be easily checked considering a geodesic spherical coordinate system with origin q. Then, we have

$$|X_m\,\rho^2(\cdot, q)| \le 2\|X_m\|\rho(m, q),$$

thus, if U_m is a neighbourhood of m with diameter D, by the triangular inequality

$$|X_{m'}\,\rho^2(\cdot, q)| \le 2\|X_{m'}\|(\rho(m, q) + D) \quad \forall m' \in U_m \quad X_{m'} \in M_{m'}. \tag{1}$$

Let X be a C^∞ vector field such that $X(m) = X_m$ and consider, in a neighbourhood of m included in U_m, the integral curve of X, $\gamma(t)$, such that $\gamma(0) = m$ and the tangent vector in m is X_m, by the mean value theorem

$$X_m\mathcal{H}_f = \lim_{m'\to m} \int_{\mathcal{X}} X_{m'}\rho^2(\cdot, f(x))P(dx),$$

where $X_{m'} = X(m')$ and m' is on γ, then by the dominated converge theorem

$$X_m\mathcal{H}_f = -2\int_{\mathcal{X}} \langle X_m, \exp_m^{-1}(f(x))\rangle P(dx).$$

Finally, the continuity of $X_m\mathcal{H}_f$ follows from the inequality 1 and, again, by the the dominated converge theorem. ∎

Now we can established the conection between mean values and center of mass above–mentioned.

Proposition 2.13 *Let $(\mathcal{X}, \mathfrak{a}, P)$ be a probability space, $(M, \mathfrak{A})$ be a complete Riemannian manifold and $f : \mathcal{X} \to M$ a measurable map, such that P_f is dominated by the Riemannian measure V_R, $P_f \ll V_R$. Let the function $\mathcal{H}_f$ be defined as*

$$\mathcal{H}_f(m) = \int_{\mathcal{X}} \rho^2(m, f(x))\, P(dx)$$

defined for all $m \in M$. Then $\mathcal{H}_f$ has a critical point at $m \in M$ if and only if $m = \mathfrak{M}(f)$.

Proof:
$\mathcal{H}_f$ has a critical point at m if and only if $X_m \mathcal{H}_f = 0 \quad \forall X_m \in M_m$, then, since the *cut locus* of any $p \in M$ is a Riemmanian measure zero set, see Spivak [31], by the previous proposition,

$$0 = X_m \mathcal{H}_f = -2\langle X_m, \int_{\mathcal{X}} \exp_m^{-1}(f(x)) \rangle P(dx) \qquad \forall X_m \in M_m ,$$

which is satisfied if and only if

$$\int_{\mathcal{X}} \exp_m^{-1}(f(x))\; P(dx) = 0,$$

and then the proposition follows. ∎

From last proposition we show that the defined *mean value concept* it is weaker than the *center of mass concept.* Notice also that for defining the first we only need an affine connection, while the second requires a Riemannian structure.

At this point it is natural to ask in which conditions we will have a mean value. We can give sufficient conditions to assure we have a mean value.

Definition 2.14 *Let M be a complete manifold, a set A is said to be a* regular convex *set if and only if for any p, $q \in A$ the shortest geodesic from p to q is unique in M and lies in A.*

Notice that an open regular convex set is a regular normal neighbourhood of all its points.

Proposition 2.15 *Let A be regular convex set in a complete manifold M. Then, any minimal geodesic that joins a point $p \in \partial A$ and $q \in \acute{A}$ cannot be tangent to ∂A. Where ∂A is the boundary of A.*

Proof:

Let $p \in \partial A$ be the point of tangency of one geodesic tangent to ∂A, suppose there exists q a point in $\mathring{A}$, the interior of A, close to p, joined by this geodesic. We can always suppose q so close to p as we need, since we can take p as the point where the geodesic line "leaves" the boundary of A. Let $B_\epsilon(q)$ be an open ball, with center q and radius ϵ, since $\exp_p^{-1}$ is a diffeomorphism in some neighbourhood of p that contains $B_\epsilon(q)$, for $s = 1$ there will exist a $\delta > 0$ such that for all $v \in B_\delta(0_p)$

$$\tilde{q} = \exp_p(s(\exp_p^{-1}(q) + v))$$

will be in $B_\epsilon(q) \subset \mathring{A}$. However, for s small, there will be points of the shortest geodesic line joining p and $\tilde{q}$ outside A, contradicting that A is a regular convex set. This is due to the fact that if the geodesic line is tangential to ∂A, we could find geodesic lines with origin p and points outside A with tangent vector as close to $\exp_p^{-1}(q)$ as we want, so the difference between the tangent vectors would be in $B_\delta(0_p)$. ∎

Proposition 2.16 *Let $(\mathcal{X}, \mathcal{A}, P)$ be a probability space, $(M, \mathfrak{A})$ be a complete manifold and $f : \mathcal{X} \to M$ a measurable map, let A a regular convex set such that $P\{f \in \mathring{A}\} = 1$. Suppose*

$$\mathcal{H}_f(m) = \int_{\mathcal{X}} \rho^2(m, f(x))\, P(dx) < \infty.$$

Then f has a mean value $\mathfrak{M}(f) \in \mathring{A}$.

Proof:

Note first that there will be a compact set $C \subset \bar{A}$, where $\bar{A} = A \bigcup \partial A$, such that

$$\inf_{m \in \bar{A}} \mathcal{H}_f(m) = \min_{m \in C} \mathcal{H}_f(m).$$

Otherwise, let $K \subset \bar{A}$ be a compact set such that $P(K) > 0$. There will be a sequence $\{p_n\}_{n \in \mathbb{N}}$, $p_n \in \bar{A}$ such that $\lim_{n \to \infty} \rho(p_n, K) = \infty$ and

$$\inf_{m \in \bar{A}} \mathcal{H}_f(m) = \lim_{n \to \infty} \mathcal{H}_f(p_n) \geq \lim_{n \to \infty} \rho^2(p_n, K)\, P(K) = \infty,$$

contradicting that $\mathcal{H}_f(m)$ exists. Now, by the above proposition, if p belongs to the boundary of A

$$\text{grad } \mathcal{H}_f(p) = -\int_A \exp_p^{-1}(f(x))\, P(dx)$$

is an average over outward pointing vectors, therefore p will not be a minimum. Then the minimum will be in the interior and the proposition follows. ∎

3 The intrinsic bias and the mean square Rao distance

We now apply the concepts mentioned previously to develop intrinsic measures analogous to the bias and the mean square error of an estimator.

Let $\{\chi, \mathcal{a}, P_\theta\,;\,\theta \in \Theta\}$ be a parametric statistical model, where Θ, the parameter space, is an n-dimensional C^∞ real manifold. Usually Θ is an open set of $\mathbb{R}^n$ and in this case it is customary to use the same symbol, θ, to denote points and coordinates.

We shall suppose a one-to-one map $\theta \mapsto p(\cdot\,;\theta)$ and we shall consider the set of all probability measures in the statiscal model, M, with the n-dimensional C^∞ real manifold structure induced by this map. Let us denote this manifold by $(M, \mathfrak{A})$, being $\mathfrak{A}$ the atlas induced by the parametrizations, that is the coordinates in the parameter space.

In the dominated case, which we shall assumed hereafter, the probability measures can be represented by density functions. Then let us assume, for a fixed σ-finite reference measure μ, that $P_\theta \ll \mu$, $\forall\theta \in \Theta$ and denote by $p(\cdot\,;\theta)$ a density function with respect to μ, i.e., a certain version of the Radon-Nikodym derivative $dP_\theta/d\mu$. Now, through the identification $P_\theta \mapsto p(\cdot\,;\theta)$, the points in M can be considered either densities or probability measures. Additionally, we shall assume certain regularity conditions:

1. $(M, \mathfrak{A})$, is a connected Hausdorff manifold.
2. When x is fixed, the real function on M $\xi \mapsto p(x;\xi)$ is a C^∞ function.
3. For every local chart (W, θ), the functions in x, $\partial \log p(x;\theta)/\partial\theta^i$ $i = 1, \ldots, n$, are linearly independent, and belong to $L^\alpha(p(\cdot\,;\theta)\,d\mu)$ for an suitable $\alpha > 0$.
4. The partial derivatives of the required orders
$$\partial/\partial\theta^i, \quad \partial^2/\partial\theta^i\partial\theta^j, \quad \partial^3/\partial\theta^i\partial\theta^j\partial\theta^k, \quad \ldots, i, j, k = 1, \ldots, n,$$
and the integration with respect to $d\mu$ of $p(x;\theta)$ can always be interchanged.

When all these conditions are satisfied, for a version of the density function , we shall say that the parametric statistical model is *regular*, and in this case the manifold $(M, \mathfrak{A})$ has a natural Riemannian structure, given by its *information metric*. Then, there is an affine connection defined on the manifold, the Levi-Civita connection, naturally associated with the statistical model. For further details, see Amari [2], Atkinson and Mitchell [3], Barndorff-Nielsen [4], Barndorff-Nielsen and Blaesild [6], Burbea [8], Burbea and Rao [10], Castillo [11] and Oller [28], among many others.

Therefore, a regular parametric statistical model can be viewed as a Riemmanian manifold. In this context, an *estimator $\mathcal{U}$ for the true density function (or probability measure) $p_0 = p(\cdot;\theta_0) \in M$ of the statistical model* is a family of measurable maps

$$\mathcal{U} = \{\mathcal{U}_k : \chi^k \mapsto M,\ k \in \mathbb{N}\}$$

such that the true probability measure on χ^k is $(P_0)_k(dx) = p(x_1;\theta_0)\cdot\ldots\cdot p(x_k;\theta_0)\,\mu_k(dx)$.

Observe that, corresponding to an estimator, there is a sequence of random objects taking values on a convenient representation manifold of the statistical model and the converse.

Moreover, if k is fixed, corresponding to an estimator $\mathcal{U}$ of the true density function $p_0 = p(\cdot\,;\theta_0)$, we can associate a natural C^∞ vector (first–order contravariant tensor) field induced on the manifold through the inverse, provided its existence, of the Riemannian connection exponential map $A_p(x) = \exp_p^{-1}(\mathcal{U}_k(x))$, see the previous section.

Assuming that $p_0 = p(\cdot\,;\theta_0)$, is the true density function, we are now able to introduce the following definition

Definition 3.1 *An estimator $\mathcal{U}$ is* intrinsically unbiased, *if and only if, p_0 is a mean value of $\mathcal{U}_k$, $\forall k \in \mathbb{N}$ and whatever $p_0 \in M$ is the true density function, i.e., $\mathfrak{M}_{p_0}(\mathcal{U}_k) = p_0$, where $\mathfrak{M}_{p_0}$ stands for the mean value of $\mathcal{U}_k$ computed with respect to the true probability measure $(P_0)_k$.*

Notice that the definition of unbiased estimator, unlike the classical one, is invariant with respect to any coordinate change or reparametrization.

We may try to compute the moment tensor fields corresponding to an estimator vector field, provided their existence, and to obtain, for the first order moment, the ***expectation tensor field*** of the estimator. Let $p_0 = p(x;\theta_0)$ be the true, but unknown, density function corresponding to the true probability measure P_{θ_0}, then we have $\mathcal{E}_p = E_{p_0}(A_p) = E_{p_0}\left(\exp_p^{-1}(\mathcal{U}_k)\right)$. In components notation, with respect to the parametrization given by $\theta^1,\ldots,\theta^n$, if we let $p_{(k)}(x;\theta_0)$ be the corresponding μ_k-density function for a k-size independent random sample, we have

$$\mathcal{E}^\alpha(\theta) = \int_{\chi^k} A^\alpha(x;\theta)\,p_{(k)}(x;\theta_0)\,\mu_k(dx) \qquad \alpha = 1,\ldots,n\,,$$

where $A^1(x;\theta),\ldots,A^n(x;\theta)$ are the components of $A_p(x) = \exp_p^{-1}(\mathcal{U}_k(x))$, and the dependence on k is omitted in the notation. Notice that, for all θ, the integral is always computed with respect the same reference measure μ_k.

It is convenient, in order to measure the bias, to introduce the following

Definition 3.2 *The* bias tensor field *is defined as* $B_p = E_p\left(\exp_p^{-1}(\mathcal{U}_k)\right)$, *or in components notation,*

$$B^\alpha(\theta) = \int_{\chi^k} A^\alpha(x;\theta)\, p_{(k)}(x;\theta)\, d\mu_k(x) \qquad \alpha = 1,\ldots,n\,.$$

provided their existence.

Notice that $\|B\|^2$ would supply a scalar measure of the intrinsic unbiasedness. Also observe that

$$B^\alpha(\theta_0) = \mathcal{E}^\alpha(\theta_0) \qquad \alpha = 1,\ldots,n\,,$$

Clearly we have the following:

Proposition 3.3 *An estimator $\mathcal{U}$ is intrinsically unbiased if and only if its bias tensor field is null, that is*

$$B^\alpha(\theta) = 0 \qquad \alpha = 1,\ldots,n \quad \forall\theta \in \Theta.$$

We are going to establish some relationships between the classical definition of unbiasedness and the new one.

Theorem 3.4 *Let us consider a regular statistical model such that the density function manifold is simply-connected and complete, and assume additionally that all the sectional curvatures are zero. Then there exists a global coordinate system $\theta^1,\ldots,\theta^n$ such that the corresponding metric tensor field components are constant and,* under this coordinate system, *an estimator $\mathcal{U}$ is unbiased if and only if it is intrinsically unbiased.*

Proof:
The existence of a global coordinate system $\theta^1,\ldots,\theta^n$ such that the corresponding metric tensor field components are constant is a well known result, see for instance Kobayashi and Nomizu [22, pag 105, vol. II]. Then the conclusion follows, since the geodesics are straight lines, and the manifold is essentially like $\mathbb{R}^n$.

∎

The Riemannian distance that we obtain from the information metric is known as the Rao distance. The mean of the squared Rao distance, which we shall call the *mean square Rao distance*, is the natural intrinsic version of the mean square error. If we consider loss functions that depend on the statistical model and not on external considerations, the Rao distance appears in a natural way and with desirable properties, as can be apreciated in Oller [28]. This is the reason by which it plays a fundamental role in our approach.

3.1 Some examples

We present here some examples, in which we calculate the bias of several estimators.

Example 3.5 *The univariate Exponential distribution.*
Let us consider the exponential density function parametrized as

$$p(x;\lambda) = \frac{1}{\lambda}\exp\{-\frac{x}{\lambda}\} \qquad x, \lambda \in \mathbb{R}^+.$$

The metric tensor component is given by $g_{11}(\lambda) = 1/\lambda^2$. Clearly, if we let $\theta = \log\lambda$, the new tensor components will become $\overline{g}_{11}(\theta) = 1$. Let us now consider the maximum-likelihood estimator for the parameter λ computed from a sample of size k given by $\overline{X}_k$, the ordinary sample mean. The corresponding maximum-likelihood estimator for θ is given by $\log \overline{X}_k$. Since the metric tensor is constant under the coordinate system given by θ, the bias tensor, if we let $S = k\overline{X}_k$, is given by

$$B^1(\theta) = E\left(\log\left(\frac{S}{k}\right) - \theta\right) = \int_{\mathbb{R}^+}\left(\log\left(\frac{s}{k}\right) - \theta\right)\frac{s^{k-1}}{\Gamma(k)\, e^{k\theta}}\exp\{-\frac{s}{e^\theta}\}\, ds,$$

and with the change $u = s/e^\theta$, this yields

$$B^1(\theta) = \frac{1}{\Gamma(k)}\int_{\mathbb{R}^+}\log(u)\; u^{k-1}e^{-u}\, du - \log k = \frac{\Gamma'(k)}{\Gamma(k)} - \log k = \Psi(k) - \log k,$$

where $\Psi(k) = \Gamma'(k)/\Gamma(k)$, Γ being the usual gamma function. Therefore it is a biased estimator. However, we can easily correct the bias, obtaining in this case a strictly intrinsically unbiased estimator. With respect to the parametrization given by θ, the corrected estimator will be

$$\hat{\theta} = \log \overline{X}_k - \Psi(k) + \log k,$$

and with respect to the original parametrization it is

$$\hat{\lambda} = \frac{k\overline{X}_k}{e^{\Psi(k)}}.$$

Example 3.6 *The univariate Poisson distribution.*
Let us consider the Poisson density function parametrized as

$$p(x;\lambda) = e^{-\lambda}\,\frac{\lambda^x}{x!} \qquad \lambda \in \mathbb{R}^+,\ x \in \mathbb{N}.$$

The metric tensor component is given by $g_{11}(\lambda) = 1/\lambda$. It is clear that if we let $\theta = 2\sqrt{\lambda}$, the new tensor components will become $\bar{g}_{11}(\theta) = 1$. Let us now consider the maximum-likelihood estimator for the parameter λ obtained from a sample of size k given by $\overline{X}_k$, the ordinary sample mean. The corresponding maximum-likelihood estimator for θ is given by $2\sqrt{\overline{X}_k}$. Since the metric tensor is constant under the coordinate system given by θ, the bias tensor, if we let $S = k\overline{X}_k$, is given by

$$B^1(\theta) = E\left(2\sqrt{\frac{S}{k}} - \theta\right) = 2\left(\sum_{j=1}^{\infty}\sqrt{\frac{j}{k}}\, e^{-k\frac{\theta^2}{4}}\, \frac{(k\frac{\theta^2}{4})^j}{j!}\right) - \theta,$$

which is clearly biased. Moreover, since the equation

$$E(f(S)) = \sum_{j=0}^{\infty} f(j)\, e^{-k\lambda}\, \frac{(k\lambda)^j}{j!} = 2\sqrt{\lambda}$$

equivalent to

$$\sum_{j=0}^{\infty} f(j)\, \frac{(k\lambda)^j}{j!} = \frac{2}{\sqrt{k}}\sqrt{k\lambda}e^{k\lambda},$$

where f is an arbitrary function, has no solution because $\sqrt{z}e^z$ it is not an analytic function, we conclude that for univariate Poisson distribution there does not exist an intrinsically unbiased estimator based on the sufficient statistic S.

Example 3.7 Consider the multivariate elliptic probability distributions, with fixed dispersion matrix $\Sigma = \Sigma_0$, that is the parametric family with density functions, in $\mathbb{R}^n$ with respect the Lebesgue measure, given by

$$p(x;\mu) = \frac{\Gamma(n/2)}{\pi^{n/2}}\, |\Sigma_0|^{-1/2}\, F\left((x-\mu)'\Sigma_0^{-1}(x-\mu)\right),$$

where Σ_0 is a fixed $n \times n$ strictly positive-definite matrix, $\mu = (\mu_1, \ldots, \mu_n)'$ is a parameter vector, $\Gamma(n/2)$ is the usual gamma function, and F is a non-negative function on $\mathbb{R}_+ = [0, \infty)$ satisfying:

$$\int_0^{\infty} r^{n/2-1}\, F(r)\, dr \;=\; 1.$$

The vector μ and the matrix Σ_0 may be expressed in terms of $E(X)$ and $\operatorname{cov}(X)$, provided the latter exists. In fact, let be $t = (t_1, \ldots, t_n)'$; the characteristic function $\phi_F(t) = E(\exp\{it'X\})$ of the above introduced parametric family of probability distributions, which may be expressed as

$$\phi_F(t) = \exp\{it'\mu\}\; \Lambda_F(t'\Sigma_0 t),$$

where

$$\Lambda_F(s) = \Gamma(n/2) \int_0^\infty r^{n/2-1} F(r)\, K_{n/2-1}(rs)\, dr \qquad s \in \mathbb{R},$$

with

$$K_\nu(s) = 2^\nu \frac{J_\nu(\sqrt{s})}{(\sqrt{s})^\nu} = \sum_{m=0}^{\infty} \frac{(-s)^m}{4^m\, m!\, \Gamma(m+\nu+1)},$$

and where J_ν is the ordinary Bessel function of order ν.

Formally, therefore

$$E(X) = -i \left. \frac{\partial \phi_F(t)}{\partial t} \right|_{t=0} \quad \text{and} \quad E(XX') = - \left. \frac{\partial^2 \phi_F(t)}{\partial t\, \partial t'} \right|_{t=0}.$$

This gives $E(X) = \mu$ and $E(XX') = \mu\mu' + c_F\, \Sigma_0$, where

$$c_F = -2\Lambda_F'(0) = \frac{1}{n} \int_0^\infty r^{n/2}\, F(r)\, dr,$$

and hence $Cov(X) = c_F\, \Sigma_0$. In particular, $E(X)$ exists if and only if $\int_0^\infty r^{n/2-1/2}\, F(r)\, dr < \infty$; additionally $Cov(X)$ exists if and only if we have $\int_0^\infty r^{n/2}\, F(r)\, dr < \infty$ in which case $0 < c_F < \infty$.

A non-degenerate multivariate normal distribution $N_n(\mu, \Sigma_0)$ is an example of a multivariate elliptic distribution with

$$F(s) = \frac{1}{2^{n/2}\Gamma(n/2)} \exp\{-s/2\}, \qquad \Lambda_F(s) = \exp\{-s/2\}, \quad c_F = 1.$$

Other basic properties of elliptic probability distributions have been obtained by Kelker [20] and are sumarized in Muirhead [26, pp. 32-40]. We have to assume, in addition, that

$$a = \frac{4}{n} \int_0^\infty t^{n/2}\, (\mathcal{L}F)^2(t)\, F(t)\, dt < \infty,$$

where $\mathcal{L}F = F'/F$, in order to ensure the existence of the Fisher information matrix, which is given by

$$E\left(\frac{\partial \log p}{\partial \mu} \frac{\partial \log p}{\partial \mu'} \right) = a\, \Sigma_0^{-1},$$

see Mitchell and Krzanowski [25] and Burbea and Oller [9] for more details.

Therefore, the information metric for this parametric family of probability distributions is given by

$$ds^2 = a\, d\mu'\, \Sigma_0^{-1} d\mu.$$

Since the metric tensor field given by the information matrix is constant, the manifold is Euclidean, and the geodesics are straight lines. Identifying the manifold points with their coordinates, the geodesic which starts at μ_0 when $t = 0$ and reaches μ_1 when $t = 1$ is given by

$$\mu(t) = (\mu_1 - \mu_0)\, t + \mu_0 \qquad t \in \mathbb{R},$$

and if we let $p_0 = p(\cdot; \mu_0)$, and $p_1 = p(\cdot; \mu_1)$, we have

$$\exp_{p_0}^{-1}(p_1) = (\mu_1 - \mu_0)_{p_0},$$

where, in the last equation, we have identified the tangent vectors at p_0 with their components corresponding to the canonical basis induced by the coordinate system. Considering the estimator for μ given by

$$\overline{X}_k = \frac{1}{k}\sum_{i=1}^{k} x_i,$$

and omitting the subindex p_0 for the tangent vectors at p_0, we may write

$$E_{p_0}\left(\overline{X}_k - \mu_0\right) = \frac{1}{k}\sum_{i=1}^{k} E_{p_0}(x_i) - \mu_0 = 0.$$

Therefore $\overline{X}_k$ is intrinsically unbiased.

The following simple example shows how different the squared-error loss and the square of the Rao distance can work for a fixed parametrization.

Example 3.8 Let a statistical model be defined by the Pascal family of densities

$$p(x; \theta) = (1-\theta)^x \theta, \quad x \in \mathbb{N} \cup \{0\},\ \theta \in (0,1).$$

Let $\tilde{\theta}$ be an unbiased estimator in this parametrization, for a sample of size $k = 1$. Then

$$E(\tilde{\theta}) = \sum_{x=0}^{\infty} \tilde{\theta}(x)(1-\theta)^x \theta = \theta$$

implies that

$$\tilde{\theta}(0) = 1, \quad \tilde{\theta}(x) = 0, \quad \forall x \geq 0.$$

This is the UMV unbiased estimator for θ, but, up to this important property, it not seems a reasonable estimator. On the other hand, the MLE estimator is

$$\hat{\theta} = \frac{1}{x+1}, \quad x \in \mathbb{N} \cup \{0\},$$

that seems better than the first one. However if we compute the mean square errors, (MSE), in order to compare the precision of these estimators we obtain the following:

$$\mathrm{MSE}(\tilde{\theta}) = \theta(1-\theta),$$

and

$$\begin{aligned}\mathrm{MSE}(\hat{\theta}) &= \sum_{x=0}^{\infty}\left(\frac{1}{x+1}-\theta\right)(1-\theta)^x\theta = \frac{\theta}{1-\theta}\sum_{x=0}^{\infty}\frac{(1-\theta)^{x+1}}{(x+1)^2} \\ &\quad + \frac{2\theta^2}{1-\theta}\sum_{x=0}^{\infty}\frac{(1-\theta)^{x+1}}{x+1}+\theta^2 \\ &= \frac{\theta}{1-\theta}\int_{1-\theta}^{0}\frac{\log(1-t)}{t}\,dt + \frac{2\theta^2}{1-\theta}\log\theta+\theta^2.\end{aligned}$$

Using *Mathematica*, version 1.2, we obtain that $\mathrm{MSE}(\tilde{\theta})-\mathrm{MSE}(\hat{\theta})$ is a positive function in $(\theta_0, 1)$ and negative in $(0, \theta_0)$, where approximately $\theta_0 = 0.1606$.

In this sense the squared-error loss function does not distinguish clearly between this two estimators. In fact, it can be shown that $\tilde{\theta}$ is an admissible estimator with respect to the squared-loss function. On the contrary if we use the square of the Rao distance as loss function we obtain, since

$$\rho(\theta,\eta) = 2\log\frac{1+|\sqrt{1-\theta}-\sqrt{1-\eta}|-\sqrt{(1-\theta)(1-\eta)}}{\sqrt{\theta\eta}},$$

that

$$E\left(\rho^2(\tilde{\theta},\theta)\right) = \infty.$$

Thus $\tilde{\theta}$ is, with respect to the mean square Rao distance loss, and inadmissible estimator. Moreover, as the reader can checked easily,

$$\begin{aligned}&E\left(\rho^2(\hat{\theta},\theta)\right) = \\ &\sum_{x=0}^{\infty}\left\{2\log\frac{1+|\sqrt{1-\frac{1}{x+1}}-\sqrt{1-\theta}|-\sqrt{(1-\frac{1}{x+1})(1-\theta)}}{\sqrt{\frac{\theta}{x+1}}}\right\}^2(1-\theta)^x\theta \\ &< \infty \quad \forall\theta>0,\end{aligned}$$

which shows the superiority of the MLE estimator.

4 Lower bound of mean square Rao distance

In this section, the relationship between the unbiasedness and the mean square of the Rao distance between the density, or probability measures, estimates and the true one are studied, obtaining an analogous intrinsic version of Cramér-Rao lower bound, based on the comparison theorems of Riemannian geometry, see 9.2 of the Appendix. Some analogous results but in a different approach can be seen in Hendriks [15].

With the same notacion as in the previous sections, we have the following main result.

Theorem 4.1 (Intrinsic Cramér-Rao lower bound) *Let $\mathcal{U}$ be an estimator corresponding to a n-dimensional regular parametric family of density functions for a sample size k. Assume that $(P)_k\left(\mathcal{U}_k^{-1}(M\setminus D_p)\right) = 0 \quad \forall p \in M$. Let A be the estimator tensor field and let B be the corresponding bias tensor field, $B = E(A)$. Let us assume that the mean square of the Rao distance between the true density and an estimate, $E\left(\rho^2(\mathcal{U}_k, p)\right)$, exists, and the covariant derivative of $E_p(A_p)$ exists and can be obtained by differentiating under the integral sign. Then,*

1. *In general we have*

$$\frac{\{\mathrm{div}(B) - E(\mathrm{div}(A))\}^2}{kn} + \|B\|^2 \leq E(\|A\|^2) = E\left(\rho^2(\mathcal{U}_k, p)\right),$$

   where $\mathrm{div}(\cdot)$ *stands for the divergence operator.*

2. *If all the sectional curvatures are zero, $K = 0$, then*

$$\frac{(\mathrm{div}(B) + n)^2}{kn} + \|B\|^2 \leq E\left(\rho^2(\mathcal{U}_k, p)\right).$$

3. *If all the sectional curvatures are non-positive, $K \leq \mathcal{K} < 0$ and $-n < \mathrm{div}(B)$, then*

$$\frac{(\mathrm{div}(B) + n)^2}{kn}\, \mathcal{S}_n^2 + \|B\|^2 \leq E\left(\rho^2(\mathcal{U}_k, p)\right),$$

   where

$$\mathcal{S}_n = 1 + \frac{(n-1)\left(\sqrt{-\mathcal{K}}\,\|B\| \coth\left(\sqrt{-\mathcal{K}}\,\|B\|\right) - 1\right)}{\mathrm{div}(B) + n}.$$

4. *If all sectional curvatures are less than or equal to a positive constant* $\mathcal{K}$, $d(M) < \pi/2\sqrt{\mathcal{K}}$, $d(M)$ *being the diameter of the manifold, and* $-1 \leq \operatorname{div}(B)$, *then*

$$\frac{(\operatorname{div}(B)+n)^2}{kn}\mathcal{T}_n^2 + \|B^2\| \leq E\left(\rho^2(\mathcal{U}_k, p)\right),$$

where

$$\mathcal{T}_n = \frac{2\sqrt{1 - \dfrac{4(n-1)\|B\|^2\mathcal{K}}{(n+\operatorname{div}(B))\pi^2}}}{1+\sqrt{1+16(n-1)\mathcal{K}\dfrac{\left(n+\operatorname{div}(B)-4(n-1)\frac{\|B\|^2\mathcal{K}}{\pi^2}\right)}{kn\pi^2}}}.$$

In particular, for intrinsically unbiased *estimators, we have*

4. *If all sectional curvatures are non-positive, then*

$$\frac{n}{k} \leq E\left(\rho^2(\mathcal{U}_k, p)\right).$$

4. *If all sectional curvatures are less or equal than a positive constant* $\mathcal{K}$ *and* $d(M) < \pi/2\sqrt{\mathcal{K}}$, *then*

$$\frac{4}{\left(1+\sqrt{1+16(n-1)\mathcal{K}/(k\pi^2)}\right)^2}\frac{n}{k} \leq E\left(\rho^2(\mathcal{U}_k, p)\right).$$

Proof:
Let $A^\alpha(x;\theta)$, $\alpha = 1,\ldots,n$, be the components of $\exp_p^{-1}(\mathcal{U}_k)$, $B^\alpha = E(A^\alpha)$ and $C^\alpha(x;\theta)$, $\alpha = 1,\ldots,n$, the components of any first order contravariant random tensor field. Then, by the Cauchy–Schwartz inequality,

$$|\langle A-B, C\rangle| \leq \|A-B\|\;\|C\|,$$

where $\langle\,,\rangle$ and $\|\;\|$ stand for the inner product and the norm defined on each tangent space. Additionally,

$$E\left(|\langle A-B, C\rangle|\right) \leq E\left(\|A-B\|\;\|C\|\right) \leq \sqrt{E\left(\|A-B\|^2\right)}\,\sqrt{E\left(\|C\|^2\right)},$$

again by the Cauchy–Schwartz inequality, and where the expectations, at each point p, are computed with respect to the corresponding probability measure $p_{(k)}\,d\mu_k$.

Let be $C(x;\theta) = \mathrm{grad}(\log p_{(k)}(x;\theta))$, where $\mathrm{grad}(\cdot)$ stands for the gradient operator. In components notation, and freely using the repeated index summation convention, we may write

$$C^{\alpha}(x;\theta) = g^{\alpha\beta}(\theta)\,\frac{\partial \log\, p_{(k)}(x;\theta)}{\partial\,\theta^{\beta}},$$

$g^{\alpha\beta}(\theta)$ being the components of the contravariant fundamental tensor field and where p is the joint sample density function. Therefore we have

$$\|C\|^2 = g^{\alpha\beta}\,\frac{\partial \log\, p_{(k)}}{\partial\,\theta^{\beta}}\,\frac{\partial \log\, p_{(k)}}{\partial\,\theta^{\alpha}},$$

taking expectations, and using matrix notation,

$$\begin{aligned} & E\left(\|C\|^2\right) = E(C'G^{-1}C) = E\left(\mathrm{tr}(C'G^{-1}C)\right) = E\left(\mathrm{tr}(G^{-1}CC')\right) = \\ = \; & \mathrm{tr}\left(G^{-1}E(CC')\right) = k\;\mathrm{tr}(G^{-1}G) = k\;\mathrm{tr}I = kn. \end{aligned}$$

On the other hand we also have

$$|E(\langle A,C\rangle)| = |E(\langle A-B,C\rangle)| \le E\left(|\langle A-B,C\rangle|\right)$$

and

$$E\left(\|A-B\|^2\right) = E(\|A\|^2) - \|B\|^2.$$

Therefore

$$|E\left(\langle A,C\rangle\right)| \le \sqrt{E\left(\|A\|^2\right) - \|B\|^2}\quad \sqrt{k\,n},$$

but $\|A\|^2 = \rho^2(p,\mathcal{U}_k)$, where ρ is the Riemannian distance, also called in this case the Rao distance. Then

$$|E(\langle A,C\rangle)| \le \sqrt{E\left(\rho^2(p,\mathcal{U}_k)\right) - \|B\|^2}\quad \sqrt{k\,n}.$$

On the other hand

$$\langle A,C\rangle = g_{\alpha\beta}A^{\alpha}C^{\beta} = g_{\alpha\beta}A^{\alpha}g^{\beta\gamma}\,\frac{\partial \log\, p_{(k)}}{\partial\,\theta^{\gamma}} = A^{\alpha}\,\frac{\partial \log\, p_{(k)}}{\partial\,\theta^{\alpha}},$$

thus,

$$E(\langle A,C\rangle) = \int_{\chi^k} A^{\alpha}\,\frac{\partial \log\, p_{(k)}}{\partial\,\theta^{\alpha}}\,p_{(k)}\,d\mu_k = \int_{\chi^k} A^{\alpha}\,\frac{\partial\, p_{(k)}}{\partial\,\theta^{\alpha}}\,d\mu_k.$$

Notice that $A^{\alpha}\partial p_{(k)}/\partial\theta^{\alpha}$ is a function of x which is independent of the coordinate system: when x is fixed it is a scalar function on the manifold.

Additionally, since

$$\int_{\chi^k} A^\alpha p_{(k)}\, d\mu_k \;=\; B^\alpha \quad \alpha = 1, \ldots, n,$$

taking the *covariant derivative* we obtain,

$$\int_{\chi^k} \left\{ \frac{\partial A^\alpha}{\partial \theta^i} + \Gamma^\alpha_{ij}\, A^j \right\} p_{(k)}\, d\mu_k + \int_{\chi^k} A^\alpha \frac{\partial p_{(k)}}{\partial \theta^i}\, d\mu_k = \frac{\partial B^\alpha}{\partial \theta^i} + \Gamma^\alpha_{ij}\, B^j$$

where Γ^α_{ij} are the Christoffel symbols of the second kind.

If we carry out an index contraction we shall obtain a scalar equation:

$$\int_{\chi^k} \left\{ \frac{\partial A^\alpha}{\partial \theta^\alpha} + \Gamma^\alpha_{\alpha j}\, A^j \right\} p_{(k)}\, d\mu_k + \int_{\chi^k} A^\alpha \frac{\partial p_{(k)}}{\partial \theta^\alpha}\, d\mu_k = B^\alpha{}_{,\alpha}$$

or equivalently, since $A^\alpha{}_{,\alpha} = \operatorname{div}(A)$, and $B^\alpha{}_{,\alpha} = \operatorname{div}(B)$,

$$E(\operatorname{div}(A)) + \int_{\chi^k} A^\alpha \frac{\partial p_{(k)}}{\partial \theta^\alpha}\, d\mu_k = \operatorname{div}(B)\,,$$

which is invariant with respect to coordinate changes. That is, both integrands depend on x, but are independent of the coordinate system. Therefore *1* follows.

Fixing x, we are going to choose a convenient coordinate system. Given p and $\mathcal{U}_k(x)$, we choose a geodesic spherical coordinates system with origin $\mathcal{U}_k(x)$, i.e. a system (ρ, u) as discussed in 9.4 of the Appendix, and defined almost surely, since $(P)_k \left(\mathcal{U}_k^{-1}(M \setminus D_{\mathcal{U}_k(x)})\right) = 0$.

It is clear that the components of tensor A are $(-\rho, 0, 0, \ldots, 0)$ when ρ, the Riemannian distance between p and $\mathcal{U}_k(x)$, is the first coordinate. Additionally,

$$\frac{\partial A^\alpha}{\partial \theta^\alpha} = -1 \quad \text{and} \quad \Gamma^\alpha_{\alpha j}\, A^j \;=\; -\rho\, \Gamma^\alpha_{\alpha 1} = -\frac{\partial \log \sqrt{g}}{\partial \rho}\, \rho\,,$$

where g is the determinant of the metric tensor. Then

$$\int_{\chi^k} A^\alpha \frac{\partial p_{(k)}}{\partial \theta^\alpha}\, d\mu_k = \operatorname{div}(B) + \int_{\chi^k} \left\{ 1 + \rho\, \frac{\partial \log \sqrt{g}}{\partial \rho} \right\} p_{(k)}\, d\mu_k\,,$$

Now we consider several cases.

Case 4.1.1 *Sectional curvature equal to zero.*

As a corollary of Bishop's comparison theorem, see Theorem 9.8 of the Appendix, or by direct calculation, we have

$$\frac{\partial \log \sqrt{g}}{\partial \rho} = \frac{n-1}{\rho},$$

yielding

$$\int_{\chi^k} A^\alpha \frac{\partial p_{(k)}}{\partial \theta^\alpha}\, d\mu_k = \operatorname{div}(B) + n.$$

Then, we have

$$|\operatorname{div}(B) + n| \leq \sqrt{E\left(\rho^2(\mathcal{U}_k, p)\right) - \|B\|^2}\, \sqrt{k\, n},$$

which turns out that

$$\frac{(\operatorname{div}(B) + n)^2}{kn} + \|B\|^2 \leq E\left(\rho^2(\mathcal{U}_k, p)\right).$$

Case 4.1.2 *All sectional curvatures less than zero, $K < \mathcal{K} < 0$ and $-n \leq \operatorname{div}(B)$.*

By 10 in Subsection 9.4 of the Appendix, we have

$$\frac{\partial \log \sqrt{g}}{\partial \rho} \geq (n-1)\sqrt{-\mathcal{K}} \coth(\sqrt{-\mathcal{K}}\rho),$$

yielding

$$\begin{aligned} &\int_{\chi^k} A^\alpha \frac{\partial p_{(k)}}{\partial \theta^\alpha} d\mu_k \geq \\ &\operatorname{div}(B) + \int_{\chi^k} \left\{1 + (n-1)\sqrt{-\mathcal{K}}\|A\| \coth\left(\sqrt{-\mathcal{K}}\|A\|\right)\right\} p_{(k)} d\mu_k \\ &\geq \operatorname{div}(B) + 1 + (n-1)\sqrt{-\mathcal{K}}\|B\| \coth\left(\sqrt{-\mathcal{K}}\|B\|\right), \end{aligned}$$

where the second inequality is due to the fact that the function $u \coth u$ is a convex function, and we can apply the Jensen inequality, and that $\|B\| \leq E\|A\|$, by the Cauchy–Schwartz inequality. Therefore if $-n \leq \operatorname{div}(B)$, since $1 < u \coth u$, $\forall u > 0$,

$$\frac{\left\{\operatorname{div}(B) + 1 + (n-1)\sqrt{-\mathcal{K}}\|B\| \coth\left(\sqrt{-\mathcal{K}}\|B\|\right)\right\}^2}{kn} + \|B\|^2 \leq E\left(\rho^2(\mathcal{U}_k, p)\right).$$

Case 4.1.3 *All sectional curvatures are positive and less than a fixed positive constant $K \leq \mathcal{K}$, $d(M) < \pi/2\sqrt{\mathcal{K}}$, and $-1 \leq \operatorname{div}(B)$.*

From 11 in Subsection 9.4 of the Appendix we have

$$\frac{\partial \log \sqrt{g}}{\partial \rho} \geq (n-1)\sqrt{\mathcal{K}} \cot(\rho\sqrt{\mathcal{K}}),$$

yielding

$$\int_{\chi^k} A^\alpha \frac{\partial p_{(k)}}{\partial \theta^\alpha}\, d\mu_k \geq \operatorname{div}(B) + \int_{\chi^k} \left(1 + (n-1)\sqrt{\mathcal{K}}\rho \cot(\rho\sqrt{\mathcal{K}})\right)\, p_{(k)}\, d\mu_k,$$

but $|u \cot u| \geq 1 - 4u^2/\pi^2$, $0 < u \leq \pi/2$, and therefore, taking into account that $0 \leq \rho \leq \pi/2\sqrt{\mathcal{K}}$, we have

$$\int_{\chi^k} A^\alpha \frac{\partial p_{(k)}}{\partial \theta^\alpha}\, d\mu_k \geq \operatorname{div}(B) + n - (n-1)\frac{4\mathcal{K}E(\rho^2)}{\pi^2},$$

since $(n-1)\dfrac{4\mathcal{K}E(\rho^2)}{\pi^2} - n \leq -1 \leq \operatorname{div}(B)$, we have

$$\left(\operatorname{div}(B) + n - (n-1)\frac{4\mathcal{K}E(\rho^2)}{\pi^2}\right)^2 \leq \left(E\left(\rho^2(\mathcal{U}_k, p)\right) - \|B\|^2\right) kn,$$

and finally, solving the inequation, it turns out that

$$\frac{4\left(n + \operatorname{div}(B) - 4(n-1)\|B\|^2\mathcal{K}/\pi^2\right)^2}{\left(\sqrt{kn} + \sqrt{kn + 16(n-1)\mathcal{K}\dfrac{(n + \operatorname{div}(B) - 4(n-1)\|B\|^2\mathcal{K}/\pi^2)}{\pi^2}}\right)^2} + \|B^2\|$$
$$\leq E\left(\rho^2(\mathcal{U}_k, p)\right).$$

The cases *5.* and *6.* follow trivially from cases *2.*, *3.* and *4.*, with $\operatorname{div}(B) = 0$ and $\|B\| = 0$. ∎

Remarks. Notice that all the one–dimensional manifolds corresponding to one–parameter families of probability distributions are always Euclidean. Moreover, there are some well-known families of probability distribution which satisfy the hypothesis of last theorem, like multinomial, see Atkinson and Mitchell [3], negative multinomial distribution, see Oller and Cuadras [29], or extreme value distributions, see Oller [27], among many others.

Additionally, it is easy to check that in the multivariate normal case, with known covariance matrix, the sample mean is an estimator which attains the intrinsic Cramér–Rao lower bound,

$$\begin{aligned} E\left(\rho^2(\overline{X}_k-\mu_0)\right) &= E\left((\overline{X}_k-\mu_0)'\Sigma^{-1}(\overline{X}_k-\mu_0)\right) = \\ &= E\left(\mathrm{tr}(\Sigma^{-1}(\overline{X}_k-\mu_0)(\overline{X}_k-\mu_0)')\right) = \\ &= \mathrm{tr}\left(\Sigma^{-1}E\left((\overline{X}_k-\mu_0)(\overline{X}_k-\mu_0)'\right)\right) = \mathrm{tr}(\frac{1}{k}I) = \frac{n}{k}. \end{aligned}$$

Furthermore, observe the effect of the sectional curvature on the precision of the statistical estimates. Finally, since the mean square Rao distance is bounded from above by D^2, D being the diameter of the manifold, it turns out, from the fact that $|\mathrm{div}(()A)| \geq 1$, that a necessary condition to have an unbiased estimator is $D \geq \frac{1}{\sqrt{kn}}$.

5 Global estimator efficiency

Whichever loss function is considered, it is well known that, in general, there is no estimator which a risk function that is uniformly smaller than any other. Therefore, given an estimator, it seems reasonable in order to measure its performance over a certain region of the statistical model, to compute the integral of the mean square Rao distance, and then divide this quantity by the Riemannian volume of the region considered. More precisely, let $\mathcal{B} \subset M$ be a measurable subset, with $V(\mathcal{B}) \neq 0$, where V is the Riemannian measure, then we shall denote the*Riemannian average of the mean square Rao distance* by

$$\mathcal{R}^2_{\mathcal{U}_k}(\mathcal{B}) = \frac{\int_{\mathcal{B}} E\left(\rho^2(\mathcal{U}_k, p)\right) V(dp)}{\int_{\mathcal{B}} V(dp)}$$

the performance index obtained is a weighted average of the mean squared distance. This approach is compatible with a Bayesian point of view, assuming a prior uniform with respect to the Riemannian volume, see Jeffreys [17]; a similar approach can be found also in Prakasa Rao [30] and Cencov [12].

In this section we are going to find some lower bounds for the above–mentioned performance estimator measure. First, we shall start with some general results.

Proposition 5.1 *Let X be a C^∞ vector field and f be a C^∞ almost everywhere positive real function, with respect to the Riemannian measure, V, defined at least over a Riemannian ball with center p and radius $R > 0$, S_R. Assume that* $\mathrm{div}(X) \leq -a$, *where a is a strictly positive real number. Then*

$$0 < \frac{a}{\mathrm{vol}\,(S_R)} \int_0^R \int_{S_r} f\,dV\,dr \leq \frac{1}{\mathrm{vol}\,(S_R)} \int_{S_R} f\,\|X\|\,dV \quad + \\ + \quad \frac{1}{\mathrm{vol}\,(S_R)} \int_0^R \int_{S_r} \|X\|\,\|\mathrm{grad}(f)\|\,dVdr.$$

Proof:

Since

$$\mathrm{div}(fX) = f\mathrm{div}(X) \; + \; \langle X, \mathrm{grad}(f)\rangle,$$

and for $0 < r < R$,

$$\int_{S_r} f\mathrm{div}(X)\,dV \leq -a\int_{S_r} f\,dV,$$

we have

$$\int_{S_r} \mathrm{div}(fX)\,dV - \int_{S_r} \langle X, \mathrm{grad}(f)\rangle\,dV \leq -a\int_{S_r} f\,dV.$$

Moreover, as a consequence of *Gauss's divergence theorem*,

$$\int_{S_r} \mathrm{div}(fX)\,dV = \int_{\partial S_r} \langle fX, \nu\rangle\,dA,$$

where ν denote the outward unit normal vector field, and dA the Riemannian measure induced on ∂S_r, and taking into account, by the Cauchy–Schwartz inequality, that

$$|\langle X, \mathrm{grad}(f)\rangle| \leq \|X\|\,\|\mathrm{grad}(f)\|,$$

$$|\langle fX, \nu\rangle| \leq f\,\|X\|,$$

it turns out that

$$0 < a\int_{S_r} f\,dV \leq \int_{\partial S_r} f\,\|X\|\,dA + \int_{S_r} \|X\|\,\|\mathrm{grad}(f)\|\,dV.$$

Now, following standard integration rules in spherical coordinates, we have

$$\int_0^R \left(\int_{\partial S_r} f\,\|X\|\,dA\right)\,dr = \int_{S_R} f\,\|X\|\,dV,$$

and thus, varying r from 0 to R, integrating and dividing by $\mathrm{vol}\,(S_R) = \int_{S_R} dV$,

$$0 < \frac{a}{\mathrm{vol}\,(S_R)} \int_0^R \int_{S_r} f\, dV\, dr \leq \frac{1}{\mathrm{vol}\,(S_R)} \int_{S_R} f \,\|X\|\, dV \quad + \\ \quad + \quad \frac{1}{\mathrm{vol}\,(S_R)} \int_0^R \int_{S_r} \|X\| \,\|\mathrm{grad}(f)\|\, dV dr.$$

∎

Theorem 5.2 *With the same conditions as in Theorem (4.1) , if we let*

$$\mathcal{R}^2_{\mathcal{U}_k}(S_R) = \frac{1}{\mathrm{vol}\,(S_R)} \int_{S_R} E\left(\rho^2(\mathcal{U}_k, p)\right)\, dV,$$

we obtain the following lower bound for the Riemannian average of the mean square Rao distance

$$0 < \left\{ \frac{a \displaystyle\int_0^R \mathrm{vol}\,(S_r)\, dr}{\mathrm{vol}\,(S_R) + \sqrt{kn}\,\sqrt{\mathrm{vol}\,(S_R)} \displaystyle\int_0^R \sqrt{\mathrm{vol}\,(S_r)}\, dr} \right\}^2 \leq \mathcal{R}^2_{\mathcal{U}_k}(S_R), \tag{2}$$

where $a = n$ *if the sectional curvatures are non-positive and* $a = 1$ *if the supreme of the sectional curvatures,* $\mathcal{K}$*, is positive and the diameter of the manifold satisfies* $d(M) < \pi/2\sqrt{\mathcal{K}}$.

Proof:
First of all, observe that $A_p(x) = \exp_p^{-1}(\mathcal{U}_k(x))$ is a C^∞ random vector field and $p = p_{(k)}(x;\theta)$, the likelihood function which defines the statistical model, is a random C^∞ function. Then, following the same steps as in theorem (4.1), we can choose a geodesic spherical coordinate system with origin $\mathcal{U}_k(x)$; under this coordinate system, using the repeated index summation convention, we have

$$\frac{\partial A^\alpha}{\partial \theta^\alpha} = -1 \quad \text{and} \quad \Gamma^\alpha_{\alpha j}\, A^j = -\rho\, \Gamma^\alpha_{\alpha 1} = -\frac{\partial \log\sqrt{g}}{\partial \rho}\, \rho,$$

where g is the determinant of the metric tensor. Then

$$\mathrm{div}(A) = -1 - \rho\, \frac{\partial \log\sqrt{g}}{\partial \rho}.$$

In the Euclidean case

$$\frac{\partial \log \sqrt{g}}{\partial \rho} = \frac{n-1}{\rho},$$

and thus $\operatorname{div}(A) = -n$.

When the sectional curvatures are non positive, we obtain

$$\frac{\partial \log \sqrt{g}}{\partial \rho} \geq \frac{n-1}{\rho},$$

and therefore $\operatorname{div}(A) \leq -n$.

Finally, when the supreme of the sectional curvatures, $\mathcal{K}$, is positive and the diameter of the manifold satisfies $d(M) < \pi/2\sqrt{\mathcal{K}}$, we have

$$\frac{\partial \log \sqrt{g}}{\partial \rho} \geq 0,$$

and then we obtain $\operatorname{div}(A) \leq -1$.

In any case, $\operatorname{div}(A) < -a$ with $a = n$ or $a = 1$, depending on the sectional curvature sign. Therefore, we can apply the previously obtained formulas, and integrating with respect to $d\mu$, the reference measure, we obtain

$$0 < \frac{a}{\operatorname{vol}(S_R)} \int_{\chi^k} \left(\int_0^R \int_{S_r} p_{(k)}\, dV\, dr \right) d\mu_k \leq \frac{1}{\operatorname{vol}(S_R)} \int_{\chi^k} \left(\int_{S_R} p_{(k)}\, \|A\|\, dV \right) d\mu_k + \\ + \frac{1}{\operatorname{vol}(S_R)} \int_{\chi^k} \left(\int_0^R \int_{S_r} \|A\|\, \|\operatorname{grad}(p_{(k)})\|\, dV\, dr \right) d\mu_k$$

By observing that

$$\int_{\chi^k} p_{(k)}\, d\mu_k = 1, \qquad \text{and} \qquad \|\operatorname{grad}(p_{(k)})\| = \|\operatorname{grad}(\log p_{(k)})\|\, p_{(k)}$$

interchanging integrals, it follows that

$$0 < \frac{a}{\operatorname{vol}(S_R)} \int_0^R \operatorname{vol}(S_r)\, dr \leq \frac{1}{\operatorname{vol}(S_R)} \int_{S_R} E(\|A\|)\, dV \quad + \\ + \quad \frac{1}{\operatorname{vol}(S_R)} \int_0^R \int_{S_r} E\left(\|A\|\, \|\operatorname{grad}(\log p)\|\right) dV\, dr,$$

where E is the ordinary expectation operator with respect to the probability measure $p_{(k)}\, d\mu_k$. By Cauchy–Schwartz inequality,

$$E(\|A\|\ \|\mathrm{grad}(\log p_{(k)})\|\ ) \leq \sqrt{E(\|A\|^2)}\ \sqrt{E(\|\mathrm{grad}(p_{(k)})\|^2)},$$

and additionally

$$E(\|\mathrm{grad}(\log p_{(k)})\|^2) = k\, n,$$

we have

$$\begin{aligned} 0 < \frac{a}{\mathrm{vol}\,(S_R)} \int_0^R \mathrm{vol}\,(S_r)\, dr \quad & \leq \frac{1}{\mathrm{vol}\,(S_R)} \int_{S_R} E(\|A\|)\, dV \quad + \\ & + \quad \frac{\sqrt{kn}}{\mathrm{vol}\,(S_R)} \int_0^R \left(\int_{S_r} \sqrt{E(\|A\|^2)}\, dV \right)\, dr. \end{aligned}$$

Moreover, by Jensen inequality

$$E(\|A\|) \leq \sqrt{E(\|A\|^2)},$$

and

$$\frac{1}{\mathrm{vol}\,(S_r)} \int_{S_r} \sqrt{E(\|A\|^2)}\, dV \quad \leq \quad \sqrt{\frac{1}{\mathrm{vol}\,(S_r)} \int_{S_r} E(\|A\|^2) dV},$$

for any $0 \leq r \leq R$, and then

$$\begin{aligned} 0 < \frac{a}{\mathrm{vol}\,(S_R)} \int_0^R \mathrm{vol}\,(S_r)\, dr \quad & \leq \quad \sqrt{\frac{1}{\mathrm{vol}\,(S_R)} \int_{S_R} E(\|A\|^2)\, dV} \quad + \\ & + \quad \frac{\sqrt{kn}}{\mathrm{vol}\,(S_R)} \int_0^R \sqrt{\mathrm{vol}\,(S_r)}\ \sqrt{\int_{S_r} E(\|A\|^2)\, dV}\ dr. \end{aligned}$$

Taking into account that

$$r \mapsto \sqrt{\int_{S_r} E(\|A\|^2)\, dV},$$

is a positive monotonous increasing function of r, since $E\left(\|A\|^2\right) = E\left(\rho^2(\mathcal{U}_k, p)\right)$, if we let

$$\mathcal{R}^2_{\mathcal{U}_k}\,(S_R) = \frac{1}{\mathrm{vol}\,(S_R)} \int_{S_R} E\left(\rho^2(\mathcal{U}_k, p)\right)\, dV,$$

then

$$0 < \frac{a}{\mathrm{vol}\,(S_R)} \int_0^R \mathrm{vol}\,(S_r)\, dr \quad \leq$$

$$\leq \quad \left(1 + \frac{\sqrt{kn}}{\sqrt{\mathrm{vol}\,(S_R)}} \int_0^R \sqrt{\mathrm{vol}\,(S_r)}\, dr\right) \sqrt{\mathcal{R}^2_{\mathcal{U}_k}\,(S_R)}.$$

■

Remarks. It is interesting to notice that while the local bounds obtained in Theorem (4.1) could vanish, the global bound, for balls with radius greater than zero, is always positive. On the other hand the curvature effects are present here through the volume of a Riemannian ball. Proposition (9.11) in the Appendix implies, in manifolds with constant sectional curvature, that for small balls, the bound will decrease with the curvature since the order of the numerator in the inequality (2) will be the order of the bound. Moreover further investigations using the expressions obtained in subsection (9.10) for the volume of a Riemannian ball could reveal more precise information on the curvature effects.

Corollary 5.3 *When the parametric statistical model is an Euclidean manifold we have the following lower bound for the Riemannian average of the mean square Rao distance*

$$0 < \left\{ \frac{n\,(n+2)\,R}{(n+1)\,\left(n+2+2\sqrt{kn}R\right)} \right\}^2 \leq \mathcal{R}^2_{\mathcal{U}_k}\,(S_R)\,.$$

If the Euclidean manifold, M, is complete and simply connected, we obtain the following lower bound over all the manifold

$$\frac{n\,(n+2)^2}{4\,k\,(n+1)^2} \leq \lim_{R\to\infty} \mathcal{R}^2_{\mathcal{U}_k}\,(S_R) \equiv \mathcal{R}^2_{\mathcal{U}_k}(M).$$

Proof:

Since

$$\mathrm{vol}\,(S_r) = \frac{2\pi^{n/2} r^n}{n\,\Gamma(n/2)},$$

we have

$$\int_0^R \sqrt{\mathrm{vol}\,(S_r)}\,dr \;=\; \left(\frac{8\pi^{n/2}R^{n+2}}{n(n+2)^2\,\Gamma(n/2)}\right)^{1/2},$$

and

$$\int_0^R \mathrm{vol}\,(S_r)\,dr \;=\; \frac{2\pi^{n/2}R^{n+1}}{n(n+1)\,\Gamma(n/2)},$$

then

$$0 < \left\{\frac{n\,(n+2)\,R}{(n+1)\,\left(n+2+2\sqrt{kn}R\right)}\right\}^2 \leq \mathcal{R}^2_{\mathcal{U}_k}\,(S_R)\,.$$

We derive the second statement taking limit when $R \to \infty$. ■

Example 5.4 As an example, consider the n–variate normal distribution with known covariance matrix Σ. Given a sample of size k, the Riemannian density of the mean square Rao distance corresponding to the sample mean $\overline{X}_k$ is $\mathcal{R}^2_{\mathcal{U}_k}\,(S_R) = n/k$, which is clearly greater than $n(n+2)^2/4k(n+1)^2$.

6 Conditional mean values of manifold valued maps and the Rao–Blackwell theorem

We have already obtained a lower bound for the mean square Rao distance, now we are going to study how we can decrease the mean square Rao distance for a given estimator.

Classically, this is achieved by taking the conditional mean value respect to a sufficient statistic. We shall follow a similar procedure here, but now our random objects are valued on a manifold and thus we will have to explain the meaning of a conditional mean value in this context and then obtain intrinsic versions of the Rao–Blackwell and Lehmann-Scheffée theorems.

Let $(\mathcal{X}, \mathcal{A}, P)$ be a probability space. Let $(M, \mathfrak{A})$ be a complete, (Hausdorff and connected) C^∞, n–dimensional Riemannian manifold. Then M will be a complete separable metric space (a Polish space) and we will have a regular version of the conditional probability of any random object, f, valued on M with respect to a σ-algebra $\mathcal{D}$ on the sample space, $\mathcal{X}$.

Moreover if the mean square Rao distance of f exists, we can define

$$E(\rho^2(m,f)|\mathcal{D})(x) = \int_M \rho^2(m,t)P_{f|\mathcal{D}}(x,dt),$$

where $P_{f|\mathcal{D}}(x,B)$ is the regular conditional probability of f given $\mathcal{D}$, $x \in \mathcal{X}$, B a Borelian set in M.

If for each $x \in \mathcal{X}$ there were one and only one extreme $p \in M$ of $E(\rho^2(m,f)|\mathcal{D})(x)$, or equivalently a point $p \in M$ such that

$$\int_M \exp_p^{-1}(t)P_{f|\mathcal{D}}(x,dt) = 0_p,$$

we would have a map from $\mathcal{X}$ to M that would assign a mean value for each x. It is clear that if the image of this map were countable, the map would be measurable, but since we have a dense countable set on M it turns out that this map is always measurable. This justifies the following definition.

Definition 6.1 *Let f be a random object on M and $\mathcal{D}$ a σ-algebra on $\mathcal{X}$; we shall define the conditional mean value of f with respect to $\mathcal{D}$ as a $\mathcal{D}$-measurable map, Z, such that*

$$E(\exp_Z^{-1}(f(\cdot))|\mathcal{D}) = 0_Z.$$

We shall write $\mathfrak{M}(f|\mathcal{D}) = Z$.

Remarks. From 2.16 a sufficient condition to ensure that the mean value exists is to have an open regular convex subset $N \subset M$ such that $P\{f \in N\} = 1$. Also we can extend the previous results to the case where M is not complete, since N is diffeomorphic to an open set in $\mathbb{R}^n$ and then there will exist regular versions of the conditional probability of f given $\mathcal{D}$.

The following propositions are immediate.

Proposition 6.2 *If f is a $\mathcal{D}$-measurable map then $\mathfrak{M}(f|\mathcal{D}) = f$ a.e.-P*

Proposition 6.3 *If f is independent of $\mathcal{D}$ then $\mathfrak{M}(f|\mathcal{D}) = \mathfrak{M}(f)$. a.e.-$P$*

Remark. It is necessary to point out that, in general, $\mathfrak{M}(\mathfrak{M}(f|\mathcal{D})) \neq \mathfrak{M}(f)$, as Kendall [21] already noticed and as it is easy to see with simple counterexamples.

Let us apply these notions to statistical point estimation. Given a parametric statistical model $\{\mathcal{X}, \mathcal{A}, P_\theta\ \theta \in \Theta\}$, let $M = \{p : p = p(\cdot;\theta), \theta \in \Theta\}$ be the associated manifold with the Riemannian metric given by Fisher's information matrix. We shall assume that the model is regular and that there exists an open regular convex subset $N \subset M$ such that $\mu(M \setminus N) = 0$ (μ being the dominating reference measure in the model).

Let $\mathcal{D}$ be a *sufficient* σ-algebra for the statistical model. Given a sample of size k and an estimator $\mathcal{U}_k$ we can now consider the estimator $\mathfrak{M}(\mathcal{U}_k|\mathcal{D})$. Let

$$\begin{aligned}\Delta^2_{\mathcal{U}_k}(p) &= E_p(\rho^2(\mathcal{U}_k, p)),\\ \Delta^2_{\mathcal{U}_k|\mathcal{D}}(p) &= E_p\left(\rho^2(\mathfrak{M}(\mathcal{U}_k|\mathcal{D}), p)\right).\end{aligned}$$

Taking into account that a function $h(q)$, $q \in M$ on the manifold is said to be convex if $h(\gamma(t)), t \in \mathbb{R}$ is an ordinary convex function for any geodesic line $\gamma(t)$, we have the following theorems.

Theorem 6.4 (Intrinsic Rao–Blackwell) *If fixed $p \in N$ the square of the Rao distance $\rho^2(p, \cdot)$ is a convex function then $\Delta^2_{\mathcal{U}|\mathcal{D}}(p) \le \Delta^2_{\mathcal{U}}(p)$.*

Proof:

This proof is adapted from Kendall [21]. By convexity, for all positive t

$$\begin{aligned}\rho^2(\gamma(t), p) &\ge \rho^2(\gamma(0), p) + \left.\frac{d\rho^2(\gamma(s), p)}{ds}\right|_{s=0} \cdot t\\ &= \rho^2(\gamma(0), p) + \left\langle \mathrm{grad}(\rho^2)(0), \frac{d\gamma}{ds}(0)\right\rangle \cdot t,\end{aligned}$$

then writing $m = \gamma(0)$ and $q = \gamma(t)$, since

$$\frac{d\gamma}{ds}(0)\, t = \exp_m^{-1}(q),$$

and

$$\mathrm{grad}(\rho^2)(0) = -2\exp_m^{-1}(p),$$

the above inequality can be written

$$\rho^2(q, p) \ge \rho^2(m, p) - 2\left\langle \exp_m^{-1}(p), \exp_m^{-1}(q)\right\rangle,$$

then taking $m = \mathfrak{M}(\mathcal{U}_k|\mathcal{D})$ and integrating with respect to $P_{\mathcal{U}_k|\mathcal{D}}(x, dq)$ we obtain

$$\int_M \rho^2(q, p) P_{\mathcal{U}_k|\mathcal{D}}(x, dq) \ge \rho^2(\mathfrak{M}(\mathcal{U}_k|\mathcal{D}), p),$$

since
$$\int_M \exp_m^{-1}(q) P_{\mathcal{U}_k|\mathcal{D}}(x, dq) = 0_m.$$
Finally taking expectations we obtain

$$\begin{aligned}\Delta^2_{\mathcal{U}_k}(p) &= E_p(\rho^2(\mathcal{U}_k, p)) = E_p\left(E_p(\rho^2(\mathcal{U}_k, p)|\mathcal{D})\right) \\ &\geq E_p\left(\rho^2(\mathfrak{M}(\mathcal{U}_k|\mathcal{D}), p)\right) = \Delta^2_{\mathcal{U}_k|\mathcal{D}}(p).\end{aligned}$$

■

Theorem 6.5 *If the sectional curvatures in N are at most 0, or $\mathcal{K} > 0$ with $d(N) < \pi/2\sqrt{\mathcal{K}}$, then*
$$\Delta^2_{\mathcal{U}_k}(p) \geq \Delta^2_{\mathcal{U}_k|\mathcal{D}}(p)$$

Proof:

From 9.3 in the Appendix, we are in the conditions in which the square of the Riemannian distance is convex. Thus from the previous theorem the result follows.

■

Remarks. If some curvatures are positive and we do not impose conditions about the diameter of the regular convex set, $N \subset M$, we cannot be sure about the convexity of the Riemannian distance and then it is not necessarily true that the mean of Riemannian distance between the true density and the estimated one should decrease when conditioning to $\mathcal{D}$.

On the other hand we can improve the efficiency of the estimators by conditioning with respect to a sufficient σ-algebra $\mathcal{D}$, obtaining $\mathfrak{M}(\mathcal{U}_k|\mathcal{D})$, but the bias is not preserved in general, in contrast to the classical Rao–Blackwell theorem. In other words, if $\mathcal{U}_k$ were intrinsically unbiased, $\mathfrak{M}(\mathcal{U}_k|\mathcal{D})$ would not, in general, be intrinsically unbiased since
$$\mathfrak{M}(\mathfrak{M}(\mathcal{U}_k|\mathcal{D})) \neq \mathfrak{M}(\mathcal{U}_k).$$
However, the norm of the bias tensor of $\mathfrak{M}(\mathcal{U}_k|\mathcal{D})$ would be bounded. If we let $B_{\mathfrak{M}(\mathcal{U}_k|\mathcal{D})}$ be this bias tensor, by the Cauchy-Schwartz inequality,

$$\|B_{\mathfrak{M}(\mathcal{U}_k|\mathcal{D})}(p)\|^2 \leq \Delta^2_{\mathcal{U}_k|\mathcal{D}}(p) \leq \Delta^2_{\mathcal{U}_k}(p).$$

Eventhough the bias tensor is not preserved in general when we condition with respect to a sufficient statistic, a theorem, wich is analogous to the Lehmann–Scheffée one, can be formulated in the intrinsic context. We need first to redefine the completeness notion.

Definition 6.6 *A sufficient statistic T is said to be* complet, *for M, iff*

$$\mathfrak{M}_p(f(T)) = \mathfrak{M}_p(g(T)) \quad \forall p \in M$$

implies that $f(T) = g(T)$ (a.e. μ).

Then, with the same conditions as in the previous theorem, we have the following proposition

Proposition 6.7 (Intrinsic Lehmann–Scheffée) *Let $\mathcal{U}$ be an estimator that is function of a complete sufficient statistic for M, then, it is the uniformily minimum Rao distance estimator for a fixed bias tensor.*

Proof:
The proof is trivial from the previous definition and Theorem. ∎

7 Asymptotic properties

First of all notice that, given a sequence of random variables taking values on a n–dimensional C^∞ (Hausdorff and connected) manifold with Riemannian structure, the definition of the different types of stochastic convergences is straightforward: weak, in probability, almost sure, or in r–th mean convergence, like in any metric space. Moreover, since the topology induced by the Riemannian metric is the same as the topology induced by the atlas, if a global chart exists, taking coordinates, we can reduce the study of these convergences, with the exception of the r-th mean, to the convergence of random sequences taking values on $\mathbb{R}^n$.

We have seen that the estimators often are intrinsically biased but we are going to show that the intrinsic bias tends to zero for large samples in important cases such as the maximum-likelihood estimators.

Definition 7.1 *An estimator $\mathcal{U}$ is* asymptotically intrinsically unbiased *if and only if it is intrinsically unbiased asymptotically, that is, we can construct a sequence of mean* *values of $\mathcal{U}_k$ which converges to p_0. When the sequence of mean values is uniquely defined, we may write*

$$\lim_{k\to\infty} \mathfrak{M}_{p_0}(\mathcal{U}_k) = p_0 = p(\cdot\,;\theta_0) \qquad \text{whatever} \quad p_0 \in M.$$

and we shall say that $\mathcal{U}$ is asymptotically strictly intrinsically unbiased.

In the two following propositions we shall suppose that the estimator $\mathcal{U}$ is *regular* in the sense that

$$\sup_{k\in\mathbb{N}} E_{p_0}\left(\rho^2(\mathcal{U}_k, p_o)\right) < \infty$$

and the covariant derivative of the vector field $\mathcal{E}(p) = E_{p_0}(\exp_p^{-1}(\mathcal{U}_k))$ exists and can be obtained differentiating under the integral sign. We shall also assume that the associated manifold of the regular parametric family of densities has sectional curvatures K bounded from above and below, i.e.: $\kappa < K < \mathcal{K}$ and the diameter of the manifold $d(M) < \pi/2\sqrt{\mathcal{K}}$ if $\mathcal{K} > 0$ and ∞ otherwise. Notice that in theorem 4.1 we had analogous conditions and also that these conditions are sufficient to assure a convex geometry, see 9.3 in the Appendix, and thus that the mean value is in fact a centre of mass.

Proposition 7.2 *An estimator $\mathcal{U}$ is asymptotically intrinsically unbiased if and only if for the corresponding bias tensor field, $B_k(p_0) = E_{p_0}(\exp_{p_0}^{-1}(\mathcal{U}_k))$ which depends on the sample size k, we have*

$$\lim_{k\to\infty} B_k^\alpha(p_0) = 0 \qquad \alpha = 1,\ldots,n \quad \forall p_0 \in M.$$

Proof:

Suppose first that $\lim\limits_{k\to\infty} \|B_k(p_0)\| = 0$. Let $\mathcal{H}_k(p) = E_{p_0}\left(\rho^2(\mathcal{U}_k, p)\right)$ and let $\gamma(t)$ be a geodesic line such that $\gamma(0) = \mathfrak{M}_{p_0}(\mathcal{U}_k)$ and $\gamma(1) = p_0$. Then

$$\begin{aligned}\|\mathrm{grad}(\mathcal{H}_k)(\gamma(1))\|\|\dot\gamma(1)\| \;&\geq\; \langle\, \mathrm{grad}(\mathcal{H}_k)(\gamma(1)), \dot\gamma(1)\,\rangle = \frac{d}{dt}\mathcal{H}_k(\gamma(t))\Big|_{t=1} \\ &= \int_0^1 \frac{d^2}{dt^2}\mathcal{H}_k(\gamma(t))\,dt,\end{aligned}$$

since

$$\frac{d}{dt}\mathcal{H}_k(\gamma(t))\Big|_{t=0} = 0,$$

because $\gamma(0) = \mathfrak{M}_{p_0}(\mathcal{U}_k)$ is a local minimum of $\mathcal{H}_k(p)$. Additionally, by the regularity conditions of $\mathcal{U}_k$ we can write

$$\int_0^1 \frac{d^2}{dt^2}\mathcal{H}_k(\gamma(t))\,dt == \int_0^1 \left(E_{p_0}\left(\frac{d^2}{dt^2}\rho^2(\mathcal{U}_k,\gamma(t))\right)\right)dt.$$

By 8 in Subsection 9.3 of the Appendix,

$$\frac{d^2}{dt^2}\rho^2(\mathcal{U}_k,\gamma(t)) \geq C(\mathcal{K})\cdot\rho^2(\gamma(0),\gamma(1)),$$

with $C(\mathcal{K}) > 0$. Thus we obtain

$$\|\mathrm{grad}(\mathcal{H}_k)(\gamma(1))\|\|\dot\gamma(1)\| \geq C(\mathcal{K})\cdot\rho^2(\gamma(0),\gamma(1)).$$

Finally, since $\|\dot\gamma(1)\| = \rho(\gamma(0),\gamma(1))$ and

$$\mathrm{grad}(\mathcal{H}_k)(\gamma(1)) = -2\int_{\mathcal{X}}\exp_{p_0}^{-1}(\mathcal{U}_k(x))p_{(k)}(x\,;\theta_0)\,\mu_k(dx) = -2B_k(p_0),$$

we have

$$\|B_k(p_0)\| \geq \frac{1}{2}C(\mathcal{K})\cdot\rho(\mathfrak{M}_{p_0}(\mathcal{U}_k),p_0),$$

then, taking limits, we obtain

$$\lim_{k\to\infty}\rho(\mathfrak{M}_{p_0}(\mathcal{U}_k),p_0) = 0.$$

Suppose now that $\lim_{k\to\infty}\mathfrak{M}_{p_0}(\mathcal{U}_k) = p_0$

$$\begin{aligned}\|B_k(p_0)\| &= \frac{1}{2}\|\mathrm{grad}(\mathcal{H}_k)(\gamma(1))\| = \frac{1}{2}\int_0^1\frac{d}{dt}\|\mathrm{grad}(\mathcal{H}_k)(\gamma(t))\|\,dt\\ &\leq \frac{1}{2}\int_0^1\|\frac{\nabla}{dt}\mathrm{grad}(\mathcal{H}_k)(\gamma(t))\|\,dt\end{aligned}$$

since $\mathrm{grad}(\mathcal{H}_k)(\gamma(0)) = 0$, and where the last inequality is due to the fact that for any C^1 vector field X, $\|X\|' \leq \|X'\|$. Herein we denote the covariant derivative of X, along a curve determined from the context, by X'. Then, since

$$\frac{1}{2}\mathrm{grad}(\mathcal{H}_k)(\gamma(t)) = -E_{p_0}(\exp_{\gamma(t)}^{-1}(\mathcal{U}_k)) = E_{p_0}(\frac{\partial}{\partial s}c(s,t)\Big|_{s=1})$$

with $c(s,t) = \exp_{\mathcal{U}_k}(s \cdot \exp^{-1}_{\mathcal{U}_k}(\gamma(t)))$, we obtain, by the regularity conditions on $\mathcal{U}_k$ and with the same notation as in 9.3 of the Appendix,

$$\|B_k(p_0)\| \leq \int_0^1 \left(E_{p_0}\left(\|\frac{\nabla}{\partial s}\frac{\partial}{\partial t}c(1,t)\|\right)\right) dt = \int_0^1 E_{p_0}(\|J'(1)\|)\, dt.$$

From 9.3 in the Appendix, we know that

$$J^{tan}(1) = (\nabla/\partial s)J^{tan}(1),$$

and

$$\begin{aligned} \langle \frac{\nabla}{\partial s}J^{tan}(1), J^{tan}(1) \rangle &= \langle J^{tan}(1), J^{tan}(1) \rangle \geq 0 \\ \langle \frac{\nabla}{\partial s}J^{tan}(1), J^{nor}(1) \rangle &= -\langle \frac{\nabla}{\partial s}J^{nor}(1), J^{tan}(1) \rangle = 0. \end{aligned}$$

Also, from Proposition 9.5 of the Appendix, we can bound the covariant derivative of the normal component of J, and taking into account the geodesic speed, we have

$$\begin{aligned} \|J'(1)\| &\leq \|J(1)\| \cdot \left(\max\left(1, \|c'(1,t)\|\left(\frac{2S'_\kappa}{S_\kappa} - \frac{S'_K}{S_K}\right)(\|c'(1,t)\|)\right)\right) \\ &\leq \rho(\mathfrak{M}_{p_0}(\mathcal{U}_k), p_0)2 \cdot (1 + \sqrt{|||}\rho(\mathcal{U}_k, p_0)), \end{aligned}$$

where the second inequality follows by $tS'_\kappa(t)/S_\kappa(t) \leq 1 + \sqrt{||||}t, \quad || \in \mathbb{R}$, as the reader can check easily from its definitions in 4 in Subsection 9.2 of the Appendix.

Finally, since the second order moments of $\mathcal{U}_k$ are uniformly bounded

$$\|B_k(p_0)\| \leq C \cdot \rho(\mathfrak{M}_{p_0}(\mathcal{U}_k), p_0)$$

with $C = 1 + \sup_{k \in \mathbb{N}} E_{p_0}(\rho(\mathcal{U}_k, po))$, and taking limits the proposition follows. ∎

Remark. Notice that in fact we need only that the first moment be uniformly bounded. However we maintain this condition in order to be coherent with the conditions in theorem 4.1. Nevertheless, it seems quite sensible to demand this condition for any good estimator.

Example 7.3 For the univariate exponential distribution we obtained, see example 3.5, that $B_k(p) = \Psi(k) - \log k$, where $\Psi(k) = \Gamma'(k)/\Gamma(k)$, then, since $\lim_{k\to\infty} k/e^{\Psi(k)} = 1$, it turns out that the maximum-likelihood estimator is asymptotically unbiased.

Definition 7.4 *An estimator U is an α-*consistent estimator *if and only if*

$$\lim_{k\to\infty} E_{p_0}\left(\rho^{\alpha}(\mathcal{U}_k, p_0)\right) = 0 \qquad \text{whatever} \quad p_0 \in M.$$

Theorem 7.5 *Let $\mathcal{U}$ be an α-consistent estimator for a regular parametric family, with $\alpha \geq 1$. Then $\mathcal{U}$ is asymptotically intrinsically unbiased.*

Proof:
By hypothesis,

$$\lim_{k\to\infty} E_{p_0}\left(\rho^{\alpha}(\mathcal{U}_k, p_0)\right) = 0 \quad \text{with} \quad \alpha \geq 1,$$

where ρ is the Riemannian distance. Therefore, by Jensen inequality,

$$\lim_{k\to\infty} E_{p_0}\left(\rho\left(\mathcal{U}_k, p_0\right)\right) = 0,$$

and taking into account that $\rho\left(\mathcal{U}_k, p_0\right) = \|A_{p_0}\|_{p_0}$, where A_{p_0} is the estimator vector field, it follows that

$$\lim_{k\to\infty} E_{p_0}\left(A_{p_0}\right) = 0,$$

obtaining the required result. ∎

Now we introduce a definition of normal distribution on a manifold. There are several ways to built distributions on a manifold, for a comprehensive exposition see Jupp [18]. First we define a normally distributed random vector on the tangent space .

Definition 7.6 *Let Y be a random vector on the tangent space M_p, where p is a fixed point in M. We shall say that Y is normally distributed if there exists a vector η and a positive definite 2-contravariant tensor Σ such that for any coordinate system, $(Y^{\alpha}) \sim \mathrm{N}\left((\eta^{\beta}), (\Sigma^{\gamma\delta})\right)$. We shall write $Y \sim \mathrm{N}(\eta, \Sigma)$.*

Remark Notice that this definition is independent of the coordinate system. This is possible due to the fact that the parameters, η and Σ, in a normal distribution change as a vector and a 2-contravariant tensor, respectively, when we change the coordinates of the random vector Y. Now we can define the meanning of a normal distribution on a complete manifold M.

As usually in the paper we are going to consider only random objects Z that take values, almost surely, on regular neigbourhoods, see definition 2.5, of any point in a complete manifold M. For this kind of random objects the random vector field $\exp_p^{-1}(Z)$, $p \in M$ will be almost surely well defined and we shall suppose that $\exp_p^{-1}(Z)$, $p \in M$ is defined in this sense.

Definition 7.7 *Let Z be a random object valued on a complete manifold M. We shall say that Z is normally distributed with respect p and with parameters (η, Σ), if there is a random vector $Y \sim \mathrm{N}(\eta, \Sigma)$, on M_p, such that $Z = \exp_p(Y)$. We shall write $Z \sim \mathrm{N}(\eta; \Sigma)_p$.*

Notice that if $\eta = 0$ then $\mathfrak{M}(Z) = p$. We now introduce the concept of asymptotically normal distribution in this context. Let $\{Z_k\}_{k\in\mathrm{N}}$ be a sequence of M-valued random variables, then

Definition 7.8 *Let M a complete manifold, a random sequence $\{Z_k\}_{k\in\mathrm{N}}$ is said to be s_k-asymptotically normally distributed with mean $p \in M$ if and only if there is a positive definite 2-contravariant tensor in M_p, Σ such that*

$$\left\{s_k \exp_p^{-1}(Z_k)\right\}_{k\in\mathrm{N}} \xrightarrow{\mathcal{L}} Y \qquad \text{with} \qquad Y \sim \mathrm{N}(0, \Sigma),$$

where $\mathcal{L}$ stands for the weak convergence or convergence in law, and $\{s_k\}_{k\in\mathrm{N}}$ is a sequence of positive real numbers with $\lim_{k\to\infty} s_k = \infty$

Remark. Notice that if $\{Z_k\}_{k\in\mathrm{N}}$ is s_k-asymptotically normal with mean p then

$$\left\{\exp_p(s_k \exp_p^{-1}(Z_k))\right\}_{k\in\mathrm{N}} \xrightarrow{\mathcal{L}} Z \qquad \text{with} \qquad Z \sim \mathrm{N}(0, \Sigma)_p,$$

but if we have that $\{V_k\}_{k\in\mathrm{N}} \xrightarrow{\mathcal{L}} Z$ is not necessarily true that $\exp_p^{-1}(V_k)$ converges in law to a normal distribution. We also say that *the estimator $\mathcal{U}$* is *s_k-asymptotically normally distributed* if its corresponding random M-valued sequence is asymptotically normally distributed.

Proposition 7.9 *Let $\mathcal{U}$ be an s_k-asymptotically normally distributed estimator, of a regular parametric family of probability distributions, with mean $p_0 \in M$. Also, assume that*

$$\sup_{k\in\mathrm{N}} E_{p_0}\left(\rho^{1+\epsilon}(\mathcal{U}_k, p_0)\right) < \infty \;\; \textit{for an } \epsilon \in \mathbb{R}^+.$$

Then, $\mathcal{U}$ is asymptotically intrinsically unbiased.

Proof:

Since $\mathcal{U}$ is s_k-asymptotically normal then $\rho(p, \mathcal{U}_k) \xrightarrow{P} 0$. The sequence of random variables $\rho(p_0, \mathcal{U}_k) = \|A_k(p_0)\|$ is uniformly integrable because $\rho(p_0, \mathcal{U}_k) = \|A_k(p_0)\|$; it follows that $E_{p_0}(\|A_k(p_0)\|) \to 0$ and since $0 < \|E_{p_0}(A_k(p_0))\| < E_{p_0}(\|A_k(p_0)\|)$, the proposition is derived. ∎

Theorem 7.10 *Assuming the previous theorem assumptions, maximum-likelihood estimators are asymptotically intrinsically unbiased.*

Proof:

This is an immediate consequence of the previous theorem, assuming sufficient conditions to ensure the $\sup_{k\in\mathbb{N}} E_{p_0}\left(\rho^{1+\epsilon}(\mathcal{U}_k, p_0)\right) < \infty$, for an $\epsilon \in \mathbb{R}^+$, by observing that the maximum-likelihood estimators are $\sqrt{k}$-asymptotically normally distributed. In fact,

$$\sqrt{k}\exp_p^{-1}(\mathcal{U}_k) \xrightarrow{\mathcal{L}} N\left(0, (g^{\alpha\beta})\right),$$

where $(g^{\alpha\beta})$ is the contravariant version of the metric tensor. ∎

From the equations of the geodesics it is easy to obtain a power expansion of the inverse of exponencial map in a point p of the manifold M. The equations of the geodesics in a coordinate neigbourhood of a point p and with unit tangent vector u are:

$$\frac{d^2 x^\alpha}{dt^2} + \Gamma^\alpha_{ij}\frac{dx^i}{dt}\frac{dx^j}{dt} = 0,$$

with $(x^\alpha(0)) = p$, $(\dot{x}^\alpha(0)) = u$. Thus

$$\ddot{x}^\alpha(0) + \Gamma^\alpha_{ij}(0)u^i u^j = 0.$$

Then we can obtain all derivates at the point p recursively:

$$\begin{aligned}
\ddot{x}^\alpha(0) &= -\Gamma^\alpha_{ij}(0)u^i u^j \\
\dddot{x}^\alpha(0) &= -\dot{\Gamma}^\alpha_{ij}(0)u^i u^j - 2\Gamma^\alpha_{ij}(0)\ddot{x}^i x^j \\
&= -\dot{\Gamma}^\alpha_{ij}(0)u^i u^j + 2\Gamma^\alpha_{ij}(0)\Gamma^i_{lr}(0)u^l u^r u^j \\
&= \left(-\partial_j\Gamma^\alpha_{lr} + 2\Gamma^\alpha_{ij}\Gamma^i_{lr}\right)(0)u^l u^r u^j,
\end{aligned}$$

and so on.

On the other hand

$$x^\alpha(t) - x^\alpha(0) = \dot{x}^\alpha(0)t + \frac{1}{2}\ddot{x}^\alpha(0)t^2 + \frac{1}{6}\dddot{x}^\alpha(0)t^3 + O(t^4),$$

where $f(t) = O(t^4)$ if $\lim_{t\to 0}(f(t)/t^4) = k > 0$ and we use the convention that, when the expression, say, $O(t^4)$, is used several times in an argument, different quantities may be involved on each occasion. Moreover,

$$A^\alpha = (\exp_p^{-1}(x(t)))^\alpha = t\dot{x}^\alpha(0),$$

so, this yields

$$\Delta x^\alpha = A^\alpha - \frac{1}{2}\Gamma^\alpha_{ij}A^iA^j + \left(\frac{-1}{6}\partial_j\Gamma^\alpha_{lr} + \frac{1}{3}\Gamma^\alpha_{ij}\Gamma^i_{lr}\right)A^lA^rA^j + O(t^4)$$

But, in fact, we are interested in expressing A as a power expansion. Then we should invert the above expression. This can be done iteratively. At first order

$$A^\alpha = \Delta x^\alpha + O(t^2),$$

at second order

$$A^\alpha = \Delta x^\alpha + \frac{1}{2}\Gamma^\alpha_{ij}\Delta x^i\Delta x^j + O(t^3),$$

at third order

$$A^\alpha = \Delta x^\alpha + \frac{1}{2}\Gamma^\alpha_{ij}\Delta x^i\Delta^j + \frac{1}{6}\left(\partial_j\Gamma^\alpha_{lr} + \Gamma^\alpha_{ij}\Gamma^i_{lr}\right)\Delta x^l\Delta x^r\Delta x^j + O(t^4),$$

and so on.

If we generalize the O notation to random variables, writing $Y_k = O_P(X_k)$ if the sequence of random variables $\{Y_k/X_k\}$ is bounded in probability, we can say the following:

Proposition 7.11 *Let $(U,\theta(\cdot))$ a local chart, where $\theta(p)=\theta_0$, $\mathcal{U}$ such that $\sqrt{k}\exp_p^{-1}(\mathcal{U}_k)$ converges in distribution to a random vector with mean zero and second order moments. Then, if we write $\hat\theta_{(k)} = \theta(\mathcal{U}_k)$*

$$\begin{aligned}A^\alpha_k(p)^\alpha &= \hat\theta^\alpha_{(k)} - \theta^\alpha_0 + \frac{1}{2}\Gamma^\alpha_{ij}(\hat\theta^i_{(k)} - \theta^i_0)(\hat\theta^j - \theta^j_0)\\ &\quad + \frac{1}{6}\left(\partial_j\Gamma^\alpha_{lr} + \Gamma^\alpha_{ij}\Gamma^i_{lr}\right)(\hat\theta^l_{(k)} - \theta^l_0)(\hat\theta^r - \theta^r_0)(\hat\theta^j_{(k)} - \theta^j_0) + O_P(k^{-2}),\end{aligned}$$

where $A^\alpha_k(p) = \left(exp_p^{-1}(\mathcal{U}_k)\right)^\alpha$, and the Christoffel symbols and its derivatives are calculated in p.

Proof:

First of all, since $\sqrt{k}\exp_p^{-1}(\mathcal{U}_k)$ converges in distribution to a random vector with mean zero and second order moments, $\mathcal{U}_k \xrightarrow{P} p$. Then the norm of the remainder term in the Taylor expansion at third order is $R_k \cdot t^4$, where $R_k \xrightarrow{P} f(p)$, t can be chosen as the arclength, i.e. $t = \|\exp_p^{-1}(\mathcal{U}_k)\|$ and $f(p)$ is a function that depends only on p but not on k.

From the hypothesis and by the Slustky theorem $k^2t^4 \xrightarrow{\mathcal{L}} \|Y\|^4$, such that Y have a distribution function with mean zero and variance Σ_0. Finally, since k^2t^4 converges in distribution $k^2t^4 = O_P(1)$ and equivalently $t^4 = O_P(k^{-2})$. Then the proposition follows. ■

With certain obvious conditions we can say something similar for the moments.

Proposition 7.12 *In the above conditions if* $\sup_{k\in\mathbb{N}} E(k^2\rho^{3+\epsilon}(\hat{\theta}_{(k)}, \theta_0)) < \infty$ *and the Christoffel symbols and its derivates are uniformily bounded on the support of* $\{\mathcal{U}_k\}$

$$B^\alpha(p) = \mathrm{Bias}^\alpha(\hat{\theta}_{(k)}) + \tfrac{1}{2}\Gamma^\alpha_{ij}\left\{\mathrm{Bias}^i(\hat{\theta}_{(k)})\mathrm{Bias}^j(\hat{\theta}_{(k)}) + \mathrm{Cov}\left(\hat{\theta}^i_{(k)}, \hat{\theta}^j_{(k)}\right)\right\} + O(k^{-3/2}).$$

with $B^\alpha(p) = E_p(A^\alpha)$ *and* $\mathrm{Bias}(\hat{\theta}_{(k)}) = E_p(\hat{\theta}_{(k)} - \theta_0)$.

8 Concluding remarks

The parametrization invariance of an inference procedure has been valued as an important and desirable property by several authors, see Barndorff-Nielsen [5], S. Amari [2] among others. Notice, for instance, that we need this property if we want to use, in a consistent way, the *parametric bootstrap*. Basically the parametrization invariance means that the inference procedure yields the same conclusion in any coordinate or parameter system. But what does "same conclusion" mean?. We cannot talk about same conclusions if the tools used to reach a conclusion like the bias, the mean square error, etc.. depend on the parametrization. It is pointed out that the classical bias and mean square error measures are not intrinsic quantities and therefore, in this sense, inconvenient. Therefore, the defined bias measure and the mean square of the Rao distance allow us to investigate the estimator properties in a more objective way. Unfortunately in many common and simple cases intrinsically unbiased estimators do not exist, although it is possible to correct the bias locally, obtaining a new estimator with its corresponding bias tensor null at one fixed point, p_0, provided that the bias tensor field is defined at p_0. Observe that for a fixed sample of size k, in order to correct the bias of an estimator $\mathcal{U}_k(x)$, at a fixed point p_0, it is sufficient to define the modified estimator

$$\tilde{\mathcal{U}}_k(x) = \exp_{p_0}\left(\exp_{p_0}^{-1}(\mathcal{U}_k(x)) - B(p_0)\right)$$

where $B(p_0)$ is the bias tensor corresponding to $\mathcal{U}_k(x)$ at p_0. This could be used in testing hypothesis theory, when the null hypothesis is simple, correcting the estimator bias under the null hypothesis, and allowing the construction of tests which would be invariant under reparametrizations.

It is also possible to give an average measure of the bias, like the integral of the square of the norm of the bias tensor, over the manifold, as a scalar bias measure:

$$\mathcal{B}^2_{\mathcal{U}_k} = \int_M \|B(p)\|^2 \, V_R(dp)$$

where V_R is the Riemannian measure over the manifold. Notice that this definition is independent of the coordinate system, and with possible Bayesian interpretations.

Moreover, it will be interesting to characterize the parametric families which allow an estimator to attain the intrinsic lower bound for the mean square of the Rao distance.

Rao distance has been used as a tool in different approaches, but now we emphasize its use as the right distance between estimates, namely the appropriate scale to observe and compare the estimates and consequently the estimators, even if the samples belong to the same population. The estimates are in the same manifold where the populations are. Note that the distance between estimates depends on the statistical model from wich the sample have been drawn and that if we considered the estimates located in the tangent space of the true density we would obtain, as distance between estimates, the *Mahalanobis distance*. This being a first approximation in our context.

In this work we have established a way to compare different estimators, then the following step will be to find the best estimators according to these principles.

9 Appendix

In this Appendix we present a set of notions and results that belong to the differential geometry and which are necessary to prove the statements in the paper. The reader can find further information in Spivak [31], Kobayashi and Nomizu [22], Hicks [16], Chavel [13], Cheeger [14], Bishop [7] and Karcher [19] among others.

Let $(M, \mathfrak{A})$ be a n-dimensional connected C^∞ real manifold, $\mathfrak{A}$ is the atlas, TM denotes its tangent bundle with projection map $\pi : TM \to M$, where $\pi(\xi) = p$ if $\xi \in M_p$, the tangent space at p. Assume that there is an affine connection ∇ on the manifold. Let $c\colon(\alpha,\beta) \to M$ be a smooth path in M. A vector field X along c is a map $X\colon(\alpha,\beta) \to TM$ such that $\pi \circ X = c$, i.e.: $X(s) \in M_{c(s)}$ for all $s \in (\alpha,\beta)$. The tangent vector field corresponding to c is given by the map $t \mapsto c' = c_*(d/ds|_{s=t})$, where c_* is the differential of c and d/ds is the standard derivation operator on the real line. For the sake of simplicity we shall often identify the fields along curves or surfaces with their images. For instance, we shall write c' instead of the map $t \mapsto c'(t)$.

To define the derivative of X along c, $(\nabla/ds)X$, also called *covariant derivative*, let $(U, x) \in \mathfrak{A}$ be a local chart on M such that $c((\alpha, \beta)) \cap U \neq \emptyset$, and let Γ_{ij}^k be the Christoffel symbols corresponding to the affine connection ∇, with respect the local chart (U, x), defined through

$$\nabla_{\partial_j} \partial_i = \sum_{k=1}^{n} \Gamma_{ij}^k \partial_k \qquad i, j = 1, \ldots, n,$$

where $\partial_1, \ldots, \partial_n$ are the basis vector fields corresponding to the local chart. Let $X = \sum_{j=1}^n \eta^j \ (\partial_j \circ c)$, and $c^j = x^j \circ c$, x^j being the j-coordinate function.

The derivative of X is another vector field along c given by

$$\frac{\nabla}{ds} X = \sum_{k=1}^{n} \left\{ (\eta^k)' + \sum_{i,j=1}^{n} (\Gamma_{ij}^k \circ c) \, \eta^i (c^j)' \right\} (\partial_k \circ c),$$

$(\eta^k)'$ and $(c^j)'$ being the usual derivatives on $\mathbb{R}$. This definition is coordinate independent and therefore the vector field $(\nabla/ds)X$ is well–defined, provided the existence of $\eta^k, k = 1, \ldots, n$ derivatives.

Some well known propertys are

$$\begin{aligned} \frac{\nabla}{ds}(X + Y) &= \frac{\nabla}{ds} X + \frac{\nabla}{ds} Y \\ \frac{\nabla}{ds}(fX) &= f'X + f \frac{\nabla}{ds} X, \end{aligned}$$

where X and Y are smooth vector fields along c and f is a real valued C^1 function, $f \colon (\alpha, \beta) \to \mathbb{R}$. If the manifold is Riemannian, and ∇ is the Levi–Civita connection, we also have

$$\frac{d}{ds} \langle X, Y \rangle = \langle \frac{\nabla}{ds} X, Y \rangle + \langle X, \frac{\nabla}{ds} Y \rangle.$$

Given an arbitrary connection ∇ on the manifold, the curves whose tangent vector field remains constant along them, like the straight lines in a Euclidean space, are the *geodesics*, defined by $(\nabla/ds)c' = 0$.

9.1 The exponential map

The exponential map, $\exp_p : M_p \to M$, corresponding to ∇, is defined through the corresponding geodesics as follows. Let p be a point of the manifold, $p \in M$, M_p be the tangent space at p, $\xi \in M_p$, and let $\gamma : [0,1] \to M$ be a geodesic such that

$$\gamma(0) = p \quad \text{and} \quad \gamma'(0) = \gamma_* \left(\frac{d}{ds}\Big|_{s=0} \right) = \xi,$$

where $d/ds|_{s=0}$ is the standard derivation operator on the real line, at $s = 0$. Then, the exponential map is given by $\exp_p(\xi) = \gamma(1)$, defined for all ξ in an open star-shaped neighbourhood of $0_p \in M_p$.

Notice that $\exp_p$ maps the straight lines which start at 0_p to geodesics starting at p, and since M_p is also a manifold and any tangent vector $\eta \in M_p$ may be identified as a parallel vector field on M_p and therefore as an element of $(M_p)_\zeta$ for arbitrary $\zeta \in M_p$, we have $(\exp_p)_{*|0}(\eta) = \eta$, where $(\exp_p)_{*|0}$ is the restriction of $(\exp_p)_*$ to the tangent space $(M_p)_0$. This shows, from the inverse function theorem, that $\exp_p$ is a local diffeomorphism.

Now we are going to focus on Riemannian manifolds with their natural Levi–Civita connection. Denote by $\mathfrak{S}_p(r) \subset M_p$

$$\mathfrak{S}_p(r) = \{\xi \in M_p : \|\xi\|_p = r\},$$

where $r > 0$, and for each $\xi \in \mathfrak{S}_p \equiv \mathfrak{S}_p(1)$ we define

$$\mathcal{C}_p(\xi) = \sup\{s > 0 \, : \, \rho\,(p, \gamma_\xi(s)) = s\},$$

where ρ is the Riemannian distance and γ_ξ is a geodesic defined in an open interval containing zero, such that $\gamma_\xi(0) = p$ and with tangent vector equal to ξ at the origin. Then if we set

$$\mathfrak{D}_p = \{s\xi \in M_p : 0 \le s < \mathcal{C}_p(\xi) \, ; \, \xi \in \mathfrak{S}_p\}$$

and

$$D_p = \exp_p(\mathfrak{D}_p),$$

we have the following proposition:

Proposition 9.1 $\exp_p$ *maps* $\mathfrak{D}_p$ *diffeomorphically onto* D_p.

Proof:

It will be sufficient to show that $\exp_p$ is injective since is obviously surjective and smoothness follows by the smooth dependency of geodesics with respect the inicial conditions.

Let γ_ξ a geodesic segment connecting p to m, i.e. $\exp_p(s\xi) = m$, with $s = \rho(p, m)$. Consider a normal ball of some radius ϵ at $m \in D_p$. Suppose there is another curve γ (parametrized by the arclenght) from p to m with lenght $s = |\gamma|$. If $\gamma(s - \epsilon) \neq$

$\gamma_\xi(s-\epsilon)$ the corresponding curves would eventually form a corner with the extension of γ_ξ beyond m. By cutting across this corner we would get shorter curves from p to $\gamma_\xi(s+\epsilon)$ than $s+\epsilon$, contradicting the minimality of γ_ξ beyond m. ∎

Moreover, if the manifold is also complete, the boundary of $\mathfrak{D}_p$, $\partial\mathfrak{D}_p$, is mapped by the exponential map onto ∂D_p, called the *cut locus of p in M*. It is also interesting to note that the cut locus of m has zero n-dimensional Riemannian measure in M (essentially due to the Sard theorem), and M is the disjoint union of D_m and ∂D_m. For more details can be found in Hicks [16] or Spivak [31].

Additionally, let us consider the 1-parameter family of geodesics $c(s,t)$ defined as

$$c(s,t) = \exp_p(s\,\zeta(t)),$$

where $\zeta(t)$ is any curve in $\mathfrak{S}_p$ with $\zeta(0) = \xi$, defined for s sufficiently small. Denote by

$$c' = c_*\left(\frac{\partial}{\partial s}\right), \qquad \dot{c} = c_*\left(\frac{\partial}{\partial t}\right),$$

where $\partial/\partial s$ and $\partial/\partial t$ are the ordinary partial derivation operators on $\mathbb{R}^2$. Observe that

$$\frac{\partial}{\partial s}\langle c', \dot{c}\rangle = \langle \frac{\nabla}{\partial s}c', \dot{c}\rangle + \langle c', \frac{\nabla}{\partial s}\dot{c}\rangle,$$

where $\nabla/\partial s$ and $\nabla/\partial t$ are the covariant derivatives along the curves $c(\cdot,t)$ and $c(s,\cdot)$ respectively. Then, since, fixing t, $c(\cdot,t)$ is a geodesic, $(\nabla/\partial s)c' = 0$, the Levi-Civita connection is torsion free and the Lie bracket $[c', \dot{c}] = 0$, (where $[c', \dot{c}]f = (\frac{\partial}{\partial s}\frac{\partial}{\partial t} - \frac{\partial}{\partial t}\frac{\partial}{\partial s})(f \circ c)$ for any C^2 real function f on M), we have $(\nabla/\partial s)\dot{c} = (\nabla/\partial t)c'$, and therefore

$$\frac{\partial}{\partial s}\langle c', \dot{c}\rangle = \langle c', \frac{\nabla}{\partial t}c'\rangle = \frac{1}{2}\frac{\partial}{\partial t}\langle c', c'\rangle = 0,$$

since $\|c'\| = \|\zeta(t)\| = 1$. Therefore the scalar product $\langle c', \dot{c}\rangle$ is independent of s, and for $s = 0$, we have $c(0,t) = p$, and thus $\dot{c}(0,t) = 0$, obtaining $\langle c', \dot{c}\rangle = 0$. This result is known as *the Gauss lemma*, and if we let $\gamma'_\xi(s) = (\exp_p)_{*|s\xi}(\xi)$ it could be expressed as

$$\langle (\exp_p)_{*|s\xi}(\eta), \gamma'_\xi(s)\rangle = 0,$$

where $\eta \in (\mathfrak{S}_p(s))_{s\xi}$. Therefore the curves obtained fixing $s = a$, $w(t) = c(a,t) = \exp_p(a\,\zeta(t))$ are orthogonal to the radial geodesics obtained fixing $t = b$, $\gamma(s) = c(s,b) = \exp_p(s\,\zeta(b))$ and the map $(\exp_p)_* : (M_p)_v \to M_{\exp_p(v)}$, although it does not preserve angles, maps orthogonal vectors to orthogonal vectors.

9.2 Jacobi fields

With the same basic notation as in the previous subsection, let us consider a smooth 1-parameter family of geodesics, $c(s,t)$ on M, such that $c(s,t)$ is a geodesic for every t. Let us denote by $c' = c_*(\partial/\partial s)$ and $\dot{c} = c_*(\partial/\partial t)$. Then we have, for an arbitrary vector field X along $s \mapsto c(s,t)$,

$$R(c',\dot{c})X = \frac{\nabla}{\partial t}\frac{\nabla}{\partial s}X - \frac{\nabla}{\partial s}\frac{\nabla}{\partial t}X,$$

where R is the *curvature tensor*, since the Lie bracket, $[\dot{c}, c'] = 0$.

Therefore, since fixing t, $c(s,t)$ is a geodesic, we have

$$0 = \frac{\nabla}{\partial s}c' = \frac{\nabla}{\partial t}\frac{\nabla}{\partial s}c' = \frac{\nabla}{\partial s}\frac{\nabla}{\partial t}c' + R(c',\dot{c})c',$$

and finally, since $(\nabla/\partial t)c' = (\nabla/\partial s)\dot{c}$, the vector field $\dot{c}$ along the geodesic $s \to c(s,t)$ satisfies the second order differential equation

$$\frac{\nabla^2}{\partial s^2}\ \dot{c} + R(c',\dot{c})c' = 0,$$

where $\dfrac{\nabla^2}{\partial s^2} \equiv (\dfrac{\nabla}{\partial s})\dfrac{\nabla}{\partial s}$.

In general, if $c = c(s)$ is a geodesic on M, a *Jacobi field along c*, Y is a C^∞ vector field along c satisfying *Jacobi's equation*:

$$\frac{\nabla^2}{ds^2}Y + R(c',Y)c' = 0,$$

where $c' = c_*(d/ds)$ and ∇/ds is the covariant derivative along c.

Since this equation is linear in Y, the set of all Jacobi fields along c is a vector space $\mathcal{F}_J$ over $\mathbb{R}$ of dimension equal to $2n$, n being the dimension of M. A Jacobi field Y along a geodesic is determined by its value, Y, and $(\nabla/ds)Y$ in an arbitrary geodesic point. Moreover, if X and Y are Jacobi fields,

$$\frac{d}{ds}\left\{\langle\frac{\nabla}{ds}X,Y\rangle - \langle\frac{\nabla}{ds}Y,X\rangle\right\} = \langle\frac{\nabla^2}{ds^2}X,Y\rangle - \langle\frac{\nabla^2}{ds^2}Y,X\rangle =$$

$$= -\langle R(c',X)c',Y\rangle + \langle R(c',Y)c',X\rangle = 0,$$

by a well known property of the curvature tensor, therefore the Wronskian

$$\langle\frac{\nabla}{ds}X,Y\rangle - \langle\frac{\nabla}{ds}Y,X\rangle = const.$$

and, in particular, if $X(s_0) = Y(s_0) = 0$ we will have

$$\langle \frac{\nabla}{ds} X, Y \rangle - \langle \frac{\nabla}{ds} Y, X \rangle = 0. \tag{3}$$

Additionally, for any Jacobi field Y there exist two real constants a and b such that

$$\langle Y, c' \rangle = a + bs,$$

since c' is also a Jacobi field. If $a = b = 0$ we obtain all the normal Jacobi fields (orthogonal to c'), which form a subspace of $\mathcal{F}_J$ with dimension $2n - 2$.

Therefore we can decompose any $Y \in \mathcal{F}_J$ along the geodesic $c(s)$ into its normal component and its tangential component: $Y = Y^{nor} + Y^{tan}$, both components also being Jacobi fields.

For a tangential Jacobi field, as a consequence of the Jacobi equation, we have,

$$Y(s) = (a + bs)\, c'(s).$$

All the results on Jacobi fields can be formulated in terms of unit speed geodesics, i.e.: with $\|c'\| = 1$, since if $Y(s)$ is a Jacobi field along $c(s)$ then $J(s) = Y(rs)$ is a Jacobi field along the geodesic $c(rs)$, with $J(0) = Y(0)$ and $(\nabla/\partial s)J(0) = r\, (\nabla/\partial s)Y(0)$.

In order to study the behaviour of Jacobi fields we can introduce the following differential equation, as we shall see later. Let $\kappa : \mathbb{R} \to \mathbb{R}$ be a continuous function and consider the differential equation $f'' + \kappa f = 0$. Let us denote by $\boldsymbol{S}_\kappa$ the solution of this equation with $\boldsymbol{S}_\kappa(0) = 0$ and $\boldsymbol{S}'_\kappa(0) = 1$, and by $\boldsymbol{C}_\kappa$ the solution satisfying $\boldsymbol{C}_\kappa(0) = 1$ and $\boldsymbol{C}'_\kappa(0) = 0$.

It is easy to verify that if $\kappa = K$, where K is a real constant, then

$$\boldsymbol{S}_K(t) = \begin{cases} \dfrac{\sin(\sqrt{K}t)}{\sqrt{K}} & \text{if } K > 0, \\ t & \text{if } K = 0, \\ \dfrac{\sinh(\sqrt{-K}t)}{\sqrt{-K}} & \text{if } K < 0, \end{cases} \tag{4}$$

and $\boldsymbol{C}_K = \boldsymbol{S}'_K$.

Given a normal (orthogonal to c') Jacobi field Y, let us define

$$f_\kappa = \|Y\|(0)\, \boldsymbol{C}_\kappa + \|Y\|'(0)\, \boldsymbol{S}_\kappa. \tag{5}$$

Let us also introduce the sectional Riemannian curvature bounds, along the geodesic c, $\delta(s) \le K \le \Delta(s)$ (for arbitrary linearly independent 2–planes), where K is the sectional Riemannian curvature,

$$K(X, Y) = \frac{\langle R(X,Y)X, Y \rangle}{\langle X, X \rangle \langle Y, Y \rangle - \langle X, Y \rangle^2}.$$

Then, we have the following comparison's theorems

Theorem 9.2 (Rauch comparison theorem) *Let Y be a normal Jacobi field along a unit speed geodesic $c(s)$, then it satisfies the following inequalities, as long as it does not vanish and $f_\Delta(t) > 0$ $0 < t < s$:*

$$\left(\frac{\|Y\|(s)}{f_\Delta(s)}\right)' \geq 0, \qquad \frac{\|Y\|'(s)}{\|Y\|(s)} \geq \frac{f'_\Delta(s)}{f_\Delta(s)}, \qquad \|Y\|(s) \geq f_\Delta(s).$$

Where the second inequality can be written as

$$f_\Delta(s)\ \langle\frac{\nabla}{\partial s}Y(s), Y(s)\rangle \geq f'_\Delta(s)\ \langle Y(s), Y(s)\rangle.$$

Proof: For all $t \in (0, s)$

$$\begin{aligned} \|Y\|''(t) &= \left\{\frac{1}{\|Y\|}\langle\frac{\nabla}{\partial t}Y, Y\rangle\right\}'(t) = \frac{1}{\|Y\|}\langle -R(c', Y)c', Y\rangle(t) \\ &\quad + \frac{1}{\|Y\|^3}\left\{\langle\frac{\nabla}{\partial t}Y, \frac{\nabla}{\partial t}Y\rangle\langle Y, Y\rangle - \langle\frac{\nabla}{\partial t}Y, Y\rangle^2\right\}(t) \\ &\geq \qquad -\Delta(s)\|Y\|(t) \end{aligned}$$

by Cauchy–Schwartz inequality. Therefore, we have

$$\|Y\|''(t) + \Delta(t)\|Y\|(t) \geq 0,$$

and since $f_\Delta(t) > 0$

$$(\ f_\Delta(t)\|Y\|'(t) - f'_\Delta(t)\|Y\|(t)\ )' \geq 0,$$

since $f''_\Delta + \Delta f_\Delta = 0$. Then, since $\|Y\|(0) = f_\Delta(0)$ and $\|Y\|'(0) = f'_\Delta(0)$, integrating from 0 to s the inequalities follow.

Proposition 9.3 *Let Y be a normal Jacobi field along the geodesic $c(t)$ and let X a normal field along c with $X(0) = Y(0) = 0$ and $X(s) = Y(s)$ then $\psi_Y(s) \leq \psi_X(s)$, where*

$$\psi_X(s) \equiv \int_0^s \left\{\langle\frac{\nabla}{dt}X, \frac{\nabla}{dt}X\rangle - \langle R(c', X)c', X\rangle\right\} dt.$$

and equality holds if and only if X=Y.

Proof:

Let $Y_1, Y_2, \ldots, Y_{n-1}$ be Jacobi fields linearly independent vanishing at $t = 0$. Therefore

$$X = \sum_{i=1}^{n-1} f_i Y_i, \quad Y = \sum_{i=1}^{n-1} a_i Y_i,$$

where a_i are constants. We have

$$\begin{aligned}
\langle\frac{\nabla}{dt}X,\frac{\nabla}{dt}X\rangle &= \langle\sum_{i=1}^{n-1} f_i'Y_i+\sum_{i=1}^{n-1} f_i\frac{\nabla}{dt}Y_i,\sum_{j=1}^{n-1} f_j'Y_j+\sum_{j=1}^{n-1} f_j\frac{\nabla}{dt}Y_j\rangle \\
&= \langle\sum_{i=1}^{n-1} f_i'Y_i,\sum_{j=1}^{n-1} f_j'Y_j\rangle+2\langle\sum_{i=1}^{n-1} f_i'Y_i,\sum_{j=1}^{n-1} f_j\frac{\nabla}{dt}Y_j\rangle \\
&\quad +\langle\sum_{i=1}^{n-1} f_i\frac{\nabla}{dt}Y_i,\sum_{j=1}^{n-1} f_j\frac{\nabla}{dt}Y_j\rangle,
\end{aligned}$$

$$\langle R(c',X)c',X\rangle=\sum_{i=1}^{n-1} f_i\langle R(c',Y_i)c',X\rangle,$$

and

$$\begin{aligned}
\frac{d}{dt}\langle X,\sum_{j=1}^{n-1} f_j\frac{\nabla}{dt}Y_j\rangle &= \langle\sum_{i=1}^{n-1} f_i'Y_i+\sum_{i=1}^{n-1} f_i\frac{\nabla}{dt}Y_i,\sum_{j=1}^{n-1} f_j\frac{\nabla}{dt}Y_j\rangle \\
&\quad +\langle X,\sum_{j=1}^{n-1} f_j'\frac{\nabla}{dt}Y_j+\sum_{j=1}^{n-1} f_j\frac{\nabla^2}{dt^2}Y_j\rangle \\
&= \langle\sum_{i=1}^{n-1} f_i'Y_i,\sum_{j=1}^{n-1} f_j\frac{\nabla}{dt}Y_j\rangle+\langle\sum_{i=1}^{n-1} f_i\frac{\nabla}{dt}Y_i,\sum_{j=1}^{n-1} f_j\frac{\nabla}{dt}Y_j\rangle \\
&\quad +\langle X,\sum_{j=1}^{n-1} f_j'\frac{\nabla}{dt}Y_j\rangle-\langle\sum_{i=1}^{n-1} f_iY_i,\sum_{j=1}^{n-1} f_jR(c',Y_j)c'\rangle.
\end{aligned}$$

Therefore combining the above equations we obtain

$$\begin{aligned}
&\langle\frac{\nabla}{dt}X,\frac{\nabla}{dt}X\rangle-\langle R(c',X)c',X\rangle=\langle\sum_{i=1}^{n-1} f_i'Y_i,\sum_{j=1}^{n-1} f_j'Y_j\rangle \\
&+\langle\sum_{i=1}^{n-1} f_i'Y_i,\sum_{j=1}^{n-1} f_j\frac{\nabla}{dt}Y_j\rangle+\frac{d}{dt}\langle X,\sum_{j=1}^{n-1} f_j\frac{\nabla}{dt}Y_j,\rangle-\langle X,\sum_{j=1}^{n-1} f_j'\frac{\nabla}{dt}Y_j\rangle \\
= &\ \langle\sum_{i=1}^{n-1} f_i'Y_i,\sum_{j=1}^{n-1} f_j'Y_j\rangle+\frac{d}{dt}\langle X,\sum_{j=1}^{n-1} f_j\frac{\nabla}{dt}Y_j\rangle \\
&+\sum_{i=1}^{n-1}\sum_{j=1}^{n-1} f_i'f_j\left\{\langle Y_i,\frac{\nabla}{dt}Y_j,\rangle-\langle Y_j\frac{\nabla}{dt}Y_i\rangle\right\}.
\end{aligned}$$

where the last term is zero by 3. Thus

$$\psi_X(s) = \int_0^s \langle \sum_{i=1}^{n-1} f_i' Y_i, \sum_{j=1}^{n-1} f_j' Y_j \rangle \, dt \; + \; \langle X, \sum_{j=1}^{n-1} f_j \frac{\nabla}{dt} Y_j \rangle (s).$$

similarly

$$\psi_Y(s) = \int_0^s \langle \sum_{i=1}^{n-1} a_i' Y_i, \sum_{j=1}^{n-1} a_j' Y_j \rangle \, dt + \langle Y, \sum_{j=1}^{n-1} a_j \frac{\nabla}{dt} Y_j \rangle (s) = \langle Y, \sum_{j=1}^{n-1} a_j \frac{\nabla}{dt} Y_j \rangle (s),$$

since a_i are constants. Finally, since $f_i(s) = a_i$, we have

$$\psi_X(s) = \int_0^s \langle \sum_{i=1}^{n-1} f_i' Y_i, \sum_{j=1}^{n-1} f_j' Y_j \rangle \, dt + \psi_Y(s),$$

consequently

$$\psi_Y(s) \le \psi_X(s),$$

and the equality holds if and only if $f_i' = 0$ for $i = 1, \ldots, n-1$ and hence $X = Y$. ∎

Theorem 9.4 *Let Y be a normal Jacobi field along a unit speed geodesic $c(s)$ that vanishes at $s = 0$. Assume also that for any normal vector field Z along c*

$$\delta(s) \le \frac{\langle R(c', Z)c', Z \rangle}{\|Z\|^2},$$

then we have the following inequalities

$$\frac{\|Y\|'(s)}{\|Y\|(s)} \le \frac{f_\delta'(s)}{f_\delta(s)} \quad \left(\frac{\|Y\|(s)}{f_\delta(s)} \right)' \le 0, \quad \|Y\|(s) \le f_\delta(s),$$

where the second inequality may be written as

$$f_\delta(s) \, \langle \frac{\nabla}{\partial s} Y(s), Y(s) \rangle \le f_\delta'(s) \, \langle Y(s), Y(s) \rangle.$$

Proof:

Define

$$u(s) = \langle Y, Y \rangle (s), \quad v(s) = f_\delta^2(s),$$

and

$$\mu(s) = \frac{\psi_Y(s)}{u(s)}, \quad \nu(s) = \frac{\int_0^s \left((f_\delta')^2 - \delta f_\delta^2\right) dt}{v(s)}.$$

Notice that

$$\int_0^s \left((f_\delta')^2 - \delta f_\delta^2\right) dt = (f_\delta' f_\delta)(s) - (f_\delta' f_\delta)(0) = (f_\delta' f_\delta)(s),$$

since $f_\delta f_\delta'' + \delta f_\delta^2 = 0$ and $f_\delta(0) = 0$. Then

$$\frac{du}{ds} = 2\langle \frac{\nabla}{\partial s} Y, Y\rangle = 2\psi_Y = 2\mu u, \quad \frac{dv}{ds} = 2 f_\delta' f_\delta = 2\nu v.$$

And solving the diferential equations we obtain

$$u(s) = u(\epsilon)\exp\{2\int_\epsilon^s \mu(t)\, dt\}, \quad v(s) = v(\epsilon)\exp\{2\int_\epsilon^s \nu(t)\, dt\}.$$

Using twice the l'Hôpital rule

$$\lim_{\epsilon\to 0} \frac{u(\epsilon)}{v(\epsilon)} = \frac{\|Y'\|^2(0)}{(f_\delta'(0))^2} = 1,$$

and therefore

$$\frac{u(s)}{v(s)} = \exp\{2\int_0^s (\mu(t) - \nu(t))\, dt\}$$

Now we are going to see that $\mu(t) \leq \nu(t)$. Let W be a parallel vector field along $c(t)$ such that $W(s) = Y(s)$. Additionally, we introduce the vector field $Z = f_\delta \frac{W}{f_\delta(s)}$. Observe that $Z(0) = 0$ and $Z(s) = W(s) = Y(s)$, then by the above proposition

$$\psi_Y(s) \leq \psi_Z(s)$$

and taking into account that

$$\begin{aligned} \psi_Z(s) &= \int_0^s \left\{\langle f_\delta' \frac{W}{f_\delta(s)}, f_\delta' \frac{W}{f_\delta(s)}\rangle - \langle R(c', f_\delta \frac{W}{f_\delta(s)})c', f_\delta \frac{W}{f_\delta(s)}\rangle\right\} dt \\ &\leq \frac{\langle W, W\rangle}{(f_\delta(s))^2}\int_0^s \left\{(f_\delta')^2 - \delta f_\delta^2\right\} dt = u(s)\nu(s) \end{aligned}$$

we have

$$\mu(s) = \frac{\psi_Y(s)}{u(s)} \leq \nu(s)$$

and thus $u(s)/v(s)$ is monotonous decreasing, with $u(s)/v(s) \leq 1$, and the inequalities follow. ■

Proposition 9.5 *Let Y be, as above, a normal Jacobi field along $c(t)$ such that $Y(0) = 0$. Assume that δ and Δ are lower and upper curvatures bounds along $c(t)$. Let k be a continuous function. Then, as long as $S_\Delta > 0$ and $S'_\kappa > 0$*

$$||Y'||(t) \leq \begin{cases} ||Y||(t) \cdot \left(\dfrac{2 \cdot S'_\kappa}{S_\kappa} - \dfrac{S'_\Delta}{S_\Delta} \right)(t) & \text{if } \kappa \leq \frac{1}{2}(\delta + \Delta) \\ ||Y'(0)|| \cdot S'_\kappa(t) & \text{if } \kappa \geq \frac{1}{2}(\delta + \Delta) \end{cases} \tag{6}$$

Proof:

Let f_κ be as in 5, and Z a parallel vector field along $c(t)$ such that $||Z|| = 1$, then

$$\langle f_\kappa Y' - f'_\kappa Y, Z \rangle (0) = 0.$$

On the other hand

$$\langle f_\kappa Y' - f'_\kappa Y, Z \rangle' = f_\kappa \langle \kappa Y - R(c', Y)c', Z \rangle,$$

and since $\langle (\kappa - R(c', \cdot)c')\, Y, Z \rangle$ is a symmetric bilineal form such that

$$\langle (\kappa - R(c', \cdot)c')\, Y, Y \rangle \leq \max(\Delta - \kappa, \kappa - \delta) ||Y||^2,$$

we have

$$\langle f_\kappa Y' - f'_\kappa Y, Z \rangle' \leq f_\kappa \max(\Delta - \kappa, \kappa - \delta) ||Y||,$$

Then by the two above theorems

$$\langle f_\kappa Y' - f'_\kappa Y, Z \rangle'(t) \leq \begin{cases} \dfrac{||Y||}{f_\Delta}(s) \left\{ (\Delta - \kappa) f_\kappa f_\Delta \right\}(t) & \text{if } \kappa \leq \frac{1}{2}(\delta + \Delta) \text{ and} \\ & f_\Delta(t) > 0,\ 0 < t \leq s \\ \left\{ (\kappa - \delta) f_\kappa f_\delta \right\}(t) & \kappa \geq \frac{1}{2}(\delta + \Delta) \end{cases}$$

$$= \begin{cases} \dfrac{||Y||}{f_\Delta}(s) \left(f'_\kappa f_\Delta - f_\kappa f'_\Delta \right)'(t) & \text{if } \kappa \leq \frac{1}{2}(\delta + \Delta) \text{ and} \\ & f_\Delta(t) > 0,\ 0 < t \leq s \\ \left(f_\kappa f'_\delta - f'_\kappa f_\delta \right)'(t) & \kappa \geq \frac{1}{2}(\delta + \Delta) \end{cases}$$

Integrating from 0 to s, choosing Z such that $Z(s) = Y'(s)/\|Y'\|(s)$ and taking into account that, by hypothesis, $f'_\kappa > 0$, we have

$$\|Y'\|(s) \leq \begin{cases} \|Y\| \left(2\frac{f'_\kappa}{f_\kappa} - \frac{f'_\Delta}{f_\Delta}\right)(s) & \kappa \leq \frac{1}{2}(\delta+\Delta) \\ f'_\delta(s) & \kappa \geq \frac{1}{2}(\delta+\Delta) \end{cases}$$

where we have applied the Cauchy-Schwartz inequality. The last inequality is due to the fact that $\|Y\|(s) \leq f_\delta(s)$. Finally putting the expressions for f_κ and f_Δ the proposition follows. ■

9.3 Convex geometry conditions

Now we are ready to analyze the convexity of the square Riemannian distance, ρ^2, equivalent to the convexity of the real function $\rho^2(p, \gamma(t))$ for any geodesic γ and arbitrary $p \in M$.

We have the following important proposition:

Proposition 9.6 *Let N be a regular convex set in a Riemannian manifold M. If the sectional curvatures in N are at most 0, or $\mathcal{K} > 0$ with $d(N) < \pi/2\sqrt{\mathcal{K}}$, then the square of the Riemannian distance is convex.*

Proof:

Let γ be a geodesic on N and consider the family of geodesics from $p \in N$ to $\gamma(t)$ defined by $c_p(s,t) = \exp_p(s \exp_p^{-1}(\gamma(t)))$. Let us denote

$$c'_p = (c_p)_*\left(\frac{\partial}{\partial s}\right) \qquad \dot{c}_p = (c_p)_*\left(\frac{\partial}{\partial t}\right).$$

It is well known that $\rho(p,\gamma(t)) = \|c'_p(s,t)\|$ is independent of s and the mapping $s \mapsto \dot{c}_p(s,t)$ is a family of Jacobi fields, since $c_p(s,t)$ is a smooth 1-parameter family of geodesics. Then,

$$\frac{\partial}{\partial t}\rho^2(p,\gamma(t)) = \frac{\partial}{\partial t}\langle c'_p(s,t), c'_p(s,t)\rangle = \frac{\partial}{\partial t}\left\{\int_0^1 \langle c'_p(s,t), c'_p(s,t)\rangle\, ds\right\} =$$

$$= 2\int_0^1 \langle \frac{\nabla}{\partial t} c'_p, c'_p\rangle\, ds = 2\int_0^1 \langle \frac{\nabla}{\partial s}\dot{c}_p, c'_p\rangle\, ds,$$

since $\|c_p'\|$ is independent of s and $(\nabla/\partial t)c_p' = (\nabla/\partial s)\dot{c}_p$, where $\nabla/\partial s$ and $\nabla/\partial t$ are the covariant derivatives along the curves $c_p(\cdot,t)$ and $c_p(s,\cdot)$ respectively. Taking into account that $(\nabla/\partial s)c_p' = 0$, we have

$$\frac{\partial}{\partial s}\langle \dot{c}_p, c_p'\rangle = \langle \frac{\nabla}{\partial s}\dot{c}_p, c_p'\rangle,$$

therefore, by observing $\dot{c}_p(0,t) = 0$, we obtain

$$\frac{\partial}{\partial t}\rho^2(p,\gamma(t)) = 2\langle \dot{c}_p(1,t), c_p'(1,t)\rangle.$$

Differentiating again, since $(\nabla/\partial t)\dot{c}_p(1,t) = (\nabla/\partial t)\dot{\gamma} = 0$ it results in

$$\frac{\partial^2}{\partial t^2}\rho^2(p,\gamma(t)) = 2\langle \dot{c}_p(1,t), \frac{\nabla}{\partial s}\dot{c}_p(1,t)\rangle,$$

denoting the Jacobi field $\dot{c}_p$ along $s \to c_p(s,t)$ by $J(s)$, then $J(0) = 0$, $J(1) = \dot{\gamma}(t)$ is independent of p, and $(\nabla/\partial s)\dot{c}_p(1,t) = (\nabla/\partial s)J(1)$. We can decompose J into its normal and tangential component: $J = J^{nor} + J^{tan}$. For the tangential component, since $J^{tan}(0) = 0$, we obtain

$$J^{tan}(s) = b\, s\, c_p'(s,t),$$

and

$$\frac{\nabla}{\partial s}J^{tan}(s) = b\, c_p'(s,t) + b\, s\, \frac{\nabla}{\partial s}c'(s,t) = b\, c_p'(s,t),$$

resulting in $J^{tan}(1) = (\nabla/\partial s)J^{tan}(1)$.

For the normal component, from the Rauch comparison theorem, and taking into account the geodesic speed, we have

$$\begin{aligned} f_\Delta(\|c_p'(1,t)\|)\,\langle \frac{1}{\|c_p'(1,t)\|}\frac{\nabla}{\partial s}J^{nor}(1), J^{nor}(1)\rangle &\geq \\ \geq f_\Delta'(\|c_p'(1,t)\|)\,\langle J^{nor}(1), J^{nor}(1)\rangle&. \end{aligned}$$

Therefore, combining these results we get

$$\begin{aligned} \langle \frac{\nabla}{\partial s}J^{tan}(1), J^{tan}(1)\rangle &= \langle J^{tan}(1), J^{tan}(1)\rangle \geq 0 \\ \langle \frac{\nabla}{\partial s}J^{tan}(1), J^{nor}(1)\rangle &= -\langle \frac{\nabla}{\partial s}J^{nor}(1), J^{tan}(1)\rangle = 0, \end{aligned}$$

and

$$\langle \frac{\nabla}{\partial s}J^{nor}(1), J^{nor}(1)\rangle \geq \|c_p'(1,t)\|\,\frac{f_\Delta'(\|c_p'(1,t)\|)}{f_\Delta\|c_p'(1,t)\|)}\,\|J^{nor}(1)\|^2,$$

resulting in

$$\begin{aligned}\frac{1}{2}\frac{\partial^2}{\partial t^2}\rho^2(p,\gamma(t)) &= \langle\, J(1), \frac{\nabla}{\partial s}J(1)\,\rangle \\ &\geq \|c_p'(1,t)\|\,\frac{f_\Delta'(\|c_p'(1,t)\|)}{f_\Delta(\|c_p'(1,t)\|)}\,\|J^{nor}(1)\|^2 + \|J^{tan}(1)\|^2.\end{aligned}$$

We are going to consider two cases. First, let us assume $\Delta(s) = 0$. In this case, since $f_\Delta(0) = 0$, $f_\Delta(s) = b\, s$, obtaining

$$\frac{1}{2}\frac{\partial^2}{\partial t^2}\rho^2(p,\gamma(t)) \geq \|J(1)\|^2 = \rho^2(\gamma(0),\gamma(1)) > 0, \tag{7}$$

and $\rho^2(p,\gamma(t))$ is a convex function.

Second, let us assume $\Delta(s) = \mathcal{K}$ where $\mathcal{K} > 0$, and additionally the manifold has a diameter $d(M) < \pi/2\sqrt{\mathcal{K}}$. In this case, $f_\Delta(0) = 0$, $f_\Delta(s) = \sin(\sqrt{\mathcal{K}}\, s)$, and therefore

$$1 > \|c_p'(1,t)\|\frac{f_\Delta'(\|c_p'(1,t)\|)}{f_\Delta(\|c_p'(1,t)\|)} = \frac{\sqrt{\mathcal{K}}\,\|c_p'(1,t)\|\,\cos(\sqrt{\mathcal{K}}\,\|c_p'(1,t)\|)}{\sin(\sqrt{\mathcal{K}}\,\|c_p'(1,t)\|)} > 0,$$

since $\|c_p'(1,t)\| = \rho(p,\gamma(t)) < d(M)$ and $0 < \sqrt{\mathcal{K}}\,\|c_p'(1,t)\| < \pi/2$, obtaining

$$\begin{aligned}\frac{1}{2}\frac{\partial^2}{\partial t^2}\rho^2(p,\gamma(t)) &\geq \frac{\sqrt{\mathcal{K}}\,\|c_p'(1,t)\|\,\cos(\sqrt{\mathcal{K}}\,\|c_p'(1,t)\|)}{\sin(\sqrt{\mathcal{K}}\,\|c_p'(1,t)\|)}\|J(1)\|^2 \\ &= \frac{\sqrt{\mathcal{K}}\,\|c_p'(1,t)\|\,\cos(\sqrt{\mathcal{K}}\,\|c_p'(1,t)\|)}{\sin(\sqrt{\mathcal{K}}\,\|c_p'(1,t)\|)}\rho^2(\gamma(0),\gamma(1)) > 0, \end{aligned}\tag{8}$$

and again, $\rho^2(p,\gamma(t))$ is a convex function. ∎

9.4 Geodesic spherical coordinates

In order to describe the notion of *spherical coordinates*, first we have to introduce the following property.

Proposition 9.7 *Let Y be a Jacobi field along $\gamma_\xi(s) = \exp_p(s\,\xi)$, with $p \in M$, $\xi \in M_p$ determined by the initial conditions $Y(0) = 0$, $(\nabla/ds)Y(0) = \eta$. Then*

$$Y(s) = (\exp_p)_{*|s\xi}(s\eta).$$

Proof:
Let us consider the 1-parameter family of geodesics $c(s,t)$ defined as

$$c(s,t) = \exp_p(s\,\zeta(t)),$$

where $\zeta(t)$ is a path in M_p with $\zeta(0) = \xi$ and $\zeta'(0) = \eta$, identifying, as usual, the elements of M_p as elements of any $(M_p)_{s\xi}$, where $(M_p)_{s\xi}$ is the tangent space at $s\xi$. In this case we know that the vector field

$$Z(s) = \dot{c}(s,0) = c_*\left(\frac{\partial}{\partial t}\Big|_{t=0}\right),$$

is a Jacobi field, and

$$Z(s) = (\exp_p)_{*|s\xi}(s\,\zeta'(0)) = s\,(\exp_p)_{*|s\xi}(\eta),$$

therefore $Z(0) = 0$, and with the same basic notation as in the previous subsection,

$$\left(\frac{\nabla}{\partial s}Z\right)(0) = \left(\frac{\nabla}{\partial s}\dot{c}\right)(0,0) = \left(\frac{\nabla}{\partial t}c'\right)(0,0) = \zeta'(0) = \eta,$$

concluding that

$$(\exp_p)_{*|s\xi}(s\eta) = Y(s).$$

■

Given a point p in a Riemannian manifold, we are now able to introduce now *geodesic spherical coordinates* on D_p, through $\exp_p|\mathfrak{D}_p$, the restriction of $\exp_p$ on $\mathfrak{D}_p$. Let us assume that there is a coordinate system on $\mathfrak{S}_p$, $\xi = \xi(u)$ where u varies over a domain in $\mathbb{R}^{n-1}$. A coordinate system on D_p is defined through

$$v(\rho, u) = \exp_p(\rho\,\xi(u)).$$

Denote by

$$\gamma_\xi'(\rho) = v_*\left(\frac{\partial}{\partial \rho}\right) \quad \text{and} \quad Y_\alpha(\rho,\xi) = v_*\left(\frac{\partial}{\partial u^\alpha}\right),$$

then, for every $\xi \in \mathfrak{S}_p$, Y_α, $\alpha = 1,\ldots,n-1$ are Jacobi fields along γ_ξ determined by the initial conditions

$$Y_\alpha(0,\xi) = 0 \quad \text{and} \quad \left(\frac{\nabla}{\partial \rho}\right)Y_\alpha(0,\xi) = \xi_*\left(\frac{\partial}{\partial u^\alpha}\right), \quad \alpha = 1,\ldots,n-1,$$

$v_*(\partial/\partial u^\alpha) \in M_{\exp_p(\rho\xi)}$ being orthogonal to $\gamma'_\xi(\rho)$, as a consequence of Gauss's lemma. Therefore in a neighbourhood of $\exp_p(\rho\,\xi(u))$ there exist a coordinate system (ρ, u), such that the corresponding basis vector field

$$\partial_1 = \gamma'_\xi(\rho), \partial_2 = Y_1(\rho,\xi), \ldots, \partial_n = Y_{n-1}(\rho,\xi)$$

satisfies

$$g_{11}(\rho,u) = \langle\partial_1,\partial_1\rangle = 1,\ g_{1\alpha}(\rho,u) = g_{\alpha 1}(\rho,u) = \langle\partial_1,\partial_\alpha\rangle = 0,\ \alpha = 2,\ldots,n,$$

where $g_{\alpha\beta}$ are the metric tensor components, and ∂_α, $\alpha = 2,\ldots,n$ are Jacobi fields along the geodesic $\gamma(s) = \exp_p(\rho\,\xi(u))$. Therefore the Riemannian metric may be expressed as

$$ds^2 = d\rho^2 + \sum_{\alpha,\beta=2}^{n} g_{\alpha\beta}(\rho,u)du^\alpha\, du^\beta, \tag{9}$$

where

$$g_{\alpha\beta}(\rho,u) = \langle Y_\alpha(\rho,\xi(u)), Y_\beta(\rho,\xi(u))\rangle, \qquad \alpha,\beta = 2,\ldots,n.$$

We have the following important theorem:

Theorem 9.8 (Bishop's comparison theorem) *Let κ and $\mathcal{K}$ be lower and upper curvature bounds in a region where* $\det G > 0$ *then in this region we have*

$$(n-1)\frac{f'_\kappa(\rho)}{f_\kappa(\rho)} \geq \frac{\partial}{\partial\rho}\log\sqrt{\det G} \geq (n-1)\frac{f'_{\mathcal{K}}(\rho)}{f_{\mathcal{K}}(\rho)}.$$

Proof:

If we let $G = (g_{\alpha\beta})_{n\times n}$ be the fundamental tensor components in matrix form, along the geodesic $\gamma_\xi(\rho)$ we have

$$\frac{\partial}{\partial\rho}\log\sqrt{\det G} = \frac{1}{2\,\det G}\sum_{\alpha\leq\beta=2}^{n}\frac{\partial\,\det G}{\partial g_{\alpha\beta}}\,\frac{\partial g_{\alpha\beta}}{\partial\rho} =$$

$$= \frac{1}{2}\sum_{\alpha,\beta=2}^{n} g^{\alpha\beta}\,\frac{\partial}{\partial\rho}\langle Y_\alpha, Y_\beta\rangle \;=\; \sum_{\alpha,\beta=2}^{n} g^{\alpha\beta}\langle Y_\alpha, \frac{\nabla}{\partial\rho}Y_\beta\rangle,$$

where $g^{\alpha\beta}$ are the coefficients of G^{-1}. Throught a lineal transformation of the fields Y_α, thus without changing the value of the previous expression, we can make Y_α to be orthogonal at a point (ρ,u). Then, we have

$$\frac{\partial}{\partial\rho}\log\sqrt{\det G} = \sum_{\alpha=2}^{n}\frac{1}{\|Y_\alpha\|^2}\,\langle Y_\alpha, \frac{\nabla}{\partial\rho}Y_\alpha\rangle,$$

then, since we have the conditions to apply the previous theorems,

$$(n-1)\frac{f'_\kappa(\rho)}{f_\kappa(\rho)} \geq \frac{\partial}{\partial\rho}\log\sqrt{\det G} \geq (n-1)\frac{f'_\mathcal{K}(\rho)}{f_\mathcal{K}(\rho)},$$

now, since the Jacobi fields vanish at the origin, $f_\kappa(\rho) = \boldsymbol{S}_\kappa$ and $f_\mathcal{K} = \boldsymbol{S}_\mathcal{K}$ and the theorem follows. ∎

We are now going to consider two cases. First, let us assume $\mathcal{K} \leq 0$. In this case $f_\mathcal{K}(\rho) = \boldsymbol{S}_\mathcal{K}(\rho) = \sinh(\sqrt{-\mathcal{K}}\rho)$, and $f'_\mathcal{K}(\rho) = \boldsymbol{C}_\mathcal{K}(\rho) = \sqrt{-\mathcal{K}}\cosh(\sqrt{-\mathcal{K}}\rho)$ if $\mathcal{K} < 0$, and thus

$$\frac{\partial}{\partial\rho}\log\sqrt{\det G} \geq (n-1)\sqrt{-\mathcal{K}}\coth(\sqrt{-\mathcal{K}}\rho), \tag{10}$$

the case $\mathcal{K} = 0$ can be obtained directly or by continuity, resulting in

$$\frac{\partial}{\partial\rho}\log\sqrt{\det G} \geq \frac{n-1}{\rho}.$$

Second case, let us assume $\mathcal{K}(\rho) = \mathcal{K}$ where $\mathcal{K} > 0$, and additionally the manifold has a diameter $d(M) < \pi/2\sqrt{\mathcal{K}}$. In this case $f_\mathcal{K}(\rho) = \boldsymbol{S}_\mathcal{K}(\rho) = \sin(\sqrt{\mathcal{K}}\rho)$, and $f'_\mathcal{K}(\rho) = \boldsymbol{C}_\mathcal{K}(\rho) = \sqrt{\mathcal{K}}\cos(\sqrt{\mathcal{K}}\rho)$, and therefore

$$\frac{\partial}{\partial\rho}\log\sqrt{\det G} \geq (n-1)\sqrt{\mathcal{K}}\cot(\sqrt{\mathcal{K}}\rho) > 0. \tag{11}$$

9.5 Comparison theorems and volumes

We can use Bishop's theorem in order to obtain the volume of a ball of radius r in a Riemannian manifold whose sectional curvatures are constant and to give bounds of this volume when the sectional curvatures are bounded. We have the following propositions:

Proposition 9.9 *If the sectional curvatures are constant and equal to $\mathcal{K}$, the volume of a Riemannian ball of radius r and center $p \in M$ is given by*

$$\mathrm{vol}(B_p(r)) = \frac{2\pi^{n/2}}{\Gamma(n/2)}\int_0^r \boldsymbol{S}_\mathcal{K}^{n-1}(t)\,dt.$$

Proof:

From the expression 9 in subsection 9.4, and writing $G_\mathcal{K}$ for the metric tensor, we obtain

$$\mathrm{vol}(B_p(r)) = \int_0^r\int_{\xi^{-1}(S_n)}\sqrt{\det G_\mathcal{K}}\,du\,d\rho.$$

Where S_n is the unit sphere in M_p. On the other hand, by Bishop's comparison theorem, when the sectional curvatures are constant

$$\frac{\partial}{\partial \rho} \log \sqrt{\det G_{\mathcal{K}}(\rho, u)} = (n-1) \frac{S'_{\mathcal{K}}}{S_{\mathcal{K}}}(\rho),$$

then, integrating this expression, we have

$$\sqrt{\det G_{\mathcal{K}}(\rho, u)} = S_{\mathcal{K}}^{n-1}\, \Omega_{\mathcal{K}}(u).$$

But, in fact, $\Omega_{\mathcal{K}}$ does not depend on $\mathcal{K}$. Obviously

$$\lim_{\rho \to 0} \frac{\sqrt{\det G_{\mathcal{K}}(\rho, u)}}{\rho^{n-1}\Omega(u)} = 1,$$

where $\Omega(u)\, du$ is the area element of the unit esphere in a Euclidean manifold, and, since

$$\lim_{\rho \to 0} \frac{S_{\mathcal{K}}^{n-1}}{\rho^{n-1}} = 1,$$

we conclude that $\Omega_{\mathcal{K}} = \Omega$. Thus we may write

$$\mathrm{vol}(B_p(r)) = \int_{\xi^{-1}(S_n)} \Omega(u)\, du \int_0^r S_{\mathcal{K}}^{n-1}(\rho)\, d\rho.$$

Finally, it is easy to check that

$$\int_{\xi^{-1}(S_n)} \Omega(u)\, du = \frac{2\pi^{n/2}}{\Gamma(n/2)},$$

and the proposition follows. ∎

Proposition 9.10 *When the sectional curvatures are constant and equal to $\mathcal{K}$ and $\mathcal{K}S_{\mathcal{K}}^2(r) < 1$ we have the following expression for the volume of a Riemannian ball of radius r.*

$$\mathrm{vol}(B(r)) = \frac{2\pi^{n/2}}{n\,\Gamma(n/2)} S_{\mathcal{K}}^n(r) \left\{ 1 + \sum_{j=1}^{\infty} \frac{n\Gamma(j+\frac{1}{2})}{\sqrt{\pi}(n+2j)} \frac{\mathcal{K}^j S_{\mathcal{K}}^{2j}(r)}{j!} \right\}. \tag{12}$$

Proof:

From the previous proposition

$$\mathrm{vol}(B(r)) = \frac{2\pi^{n/2}}{\Gamma(n/2)} \int_0^r S_{\mathcal{K}}^{n-1}(t)\,dt$$

Then, taking into account, from the definition of $S_{\mathcal{K}}$, that

$$\{S'_{\mathcal{K}}(t)\}^2 + \mathcal{K}\{S_{\mathcal{K}}(t)\}^2 = 1$$

and doing $y = S_{\mathcal{K}}^2(t)/S_{\mathcal{K}}^2(u)$, we have

$$\int_0^r S_{\mathcal{K}}^{n-1}(t)dt = \frac{1}{2} S_{\mathcal{K}}^n(r) \int_0^1 y^{\frac{n-2}{2}} \left(1 - \mathcal{K} S_{\mathcal{K}}^2(r) y\right)^{-\frac{1}{2}} dy$$

On the other hand we have a relationship between this kind of integrals and the generalized hypergeometric functions, see Abramowitz [1],

$$F(a,b;c;z) \equiv \sum_{j=0}^{\infty} \frac{(a)_j (b)_j}{(c)_j} \frac{z^j}{j!} \quad |z| < 1$$

where $(a)_j = a(a+1)\cdots(a+j-1)$, given by

$$F(a,b;c;z) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 t^{b-1}(1-t)^{c-b-1}(1-tz)^{-a}\,dt, \; Re(c) > Re(b) > 0.$$

Then this leads to

$$\begin{aligned} \int_0^r S_{\mathcal{K}}^{n-1}(t)dt &= \frac{1}{2} S_{\mathcal{K}}^n(r) \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n+2}{2})} F\left(\frac{1}{2}, \frac{n}{2}; \frac{n+2}{2}; \mathcal{K} S_{\mathcal{K}}^2(r)\right) = \\ &= \frac{1}{2} S_{\mathcal{K}}^n(r) \frac{1}{\Gamma(\frac{1}{2})} \sum_{j=0}^{\infty} \frac{\Gamma(\frac{1}{2}+j)}{\frac{n}{2}+j} \frac{\mathcal{K}^j S_{\mathcal{K}}^{2j}(r)}{j!}, \end{aligned}$$

and the proposition follows. ■

Proposition 9.11 *Let* $\mathrm{vol}(B_p(r))$ *be the volume of a ball* $B_p(r)$ *with center* p *and radius* r*. Then*

$$\mathrm{vol}_\kappa(B_{\underline{p}}(r)) \geq \mathrm{vol}(B_p(r)) \geq \mathrm{vol}_{\mathcal{K}}(B_{\bar{p}}(r)),$$

where $\mathrm{vol}_\kappa(B_{\underline{p}}(r))$ *and* $\mathrm{vol}_{\mathcal{K}}(B_{\bar{p}}(r))$ *are the volumes of balls of radius* r *and arbitrary centers* $\underline{p}$ *and* $\bar{p}$ *respectively, in manifolds with constant sectional curvatures* κ *and* $\mathcal{K}$*.*

Proof:

If we integrate, from ρ_0 to ρ, the inequalities in Bishop's comparison theorem we obtain

$$\frac{S_\kappa^{n-1}(\rho)}{S_\kappa^{n-1}(\rho_0)} \geq \frac{\sqrt{\det G(\rho, u)}}{\sqrt{\det G(\rho_0, u)}} \geq \frac{S_\mathcal{K}^{n-1}(\rho)}{S_\mathcal{K}^{n-1}(\rho_0)}.$$

Moreover

$$\lim_{\rho_0 \to 0} \sqrt{\det G(\rho_0, u)} \frac{S_\kappa^{n-1}(\rho)}{S_\kappa^{n-1}(\rho_0)} \geq \sqrt{\det G(\rho, u)} \geq \lim_{\rho_0 \to 0} \sqrt{\det G(\rho_0, u)} \frac{S_\mathcal{K}^{n-1}(\rho)}{S_\mathcal{K}^{n-1}(\rho_0)},$$

and, since

$$\lim_{\rho_0 \to 0} \frac{\sqrt{\det G(\rho_0, u)}}{S_\kappa^{n-1}(\rho_0)} = \lim_{\rho_0 \to 0} \frac{\sqrt{\det G(\rho_0, u)}}{S_\mathcal{K}^{n-1}(\rho_0)} = \Omega(u),$$

with $\Omega(u)$ as in the proof of 9.9, we conclude

$$S_\kappa^{n-1}(\rho)\Omega(u) \geq \sqrt{\det G(\rho, u)} \geq S_\mathcal{K}^{n-1}(\rho)\Omega(u),$$

and the desired result follows. ∎

References

[1] M. Abramowitz. *Handbook of Mathematical Functions.* Dover Publications, Inc., New York, 1970.

[2] S. Amari. *Differential-Geometrical Methods in Statistics*, volume 28 of *Lecture notes in statistics.* Springer Verlag, New York, 1985.

[3] C. Atkinson and A.F.S. Mitchell. Rao's distance measure. *Sankhyà*, 43 A:345–365, 1981.

[4] O. E. Barndorff-Nielsen. *Differential geometry in statistical inference*, volume 10 of *Lecture Notes - Monographs Series*, chapter Differential and integral geometry in statistical inference, pages 95–161. Institute of Mathematical Statistics, Hayward, California, 1987.

[5] O. E. Barndorff-Nielsen. *Parametric Statistical Models and Likelihood*, volume 50 of *Lecture notes in statistics.* Springer Verlag, New York, 1988.

[6] O. E. Barndorff-Nielsen and P. Blaesild. Strings: mathematical theory and statistical examples. *Proc. Roy. Soc. London*, 411 A:155–176, 1987.

[7] R. Bishop and R. Crittenden. *Geometry of Manifolds*. Academic Press, New York, 1964.

[8] J. Burbea. Informative geometry of probability spaces. *Expositiones Mathematicae*, 4:347–378, 1986.

[9] J. Burbea and J.M. Oller. The information metric for univariate linear elliptic models. *Statistics & Decisions*, 6:209–221, 1988.

[10] J. Burbea and C.R. Rao. Entropy differential metric, distance and divergence measures in probability spaces: a unified approach. *J. Multivariate Anal.*, 12:575–596, 1982.

[11] J. Castillo. Sufficient geometrical conditions for cramér-rao inequality. *submitted*, 1992.

[12] N.N. Cencov. *Statistical Decision Rules and Optimal Infererence*. Nauka, Moscow (en ruso, versi n inglesa, 1982, Math. Monographs, 53, Amer. Math. Soc., Providence), 1972.

[13] I. Chavel. *Eigenvalues in Riemannian Geometry*. Pure and applied mathematics. Academic Press, Inc., Orlando, Florida, 1984.

[14] J. Cheeger and D. G. Ebin. *Comparison Theorems in Riemannian Geometry*. North-Holland, New York, 1975.

[15] H. Hendriks. A cramér-rao type lower bound for estimators with values in a manifold. *J. Multivariate Anal.*, 38:245–261, 1991.

[16] N.J. Hicks. *Notes on Differential Geometry*. Van Nostrand Reinhold Comp., London, 1965.

[17] H. Jeffreys. An invariant form for the prior probability in estimation problems. *Proc. Royal Soc.*, 186 A:453–461, 1946.

[18] P.E. Jupp and K.V. Mardia. A unified view of the theory of directional statistics, 1975-1988. *International Statistical Review*, 57(3):261–294, 1989.

[19] H. Karcher. Riemannian center of mass and mollifier smoothing. *Comm. Pure Appl. Math.*, 30:509–541, 1977.

[20] D. Kelker. Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhyà*, 32 A:419–430, 1970.

[21] W.S. Kendall. Probability, convexity and harmonic maps with small image i: uniqueness and fine existence. *Proc. London Math. Soc.*, 61(3):371–406, 1990.

[22] S. Kobayashi and K. Nomizu. *Foundations of differential geometry.* Interscience Publishers, John Wiley, New York, 1969.

[23] J.T. Kent K.V. Mardia and J.M. Bibby. *Multivariate Analysis.* Academic Press, London, 1979.

[24] E.L. Lehmann. A general concept of unbiasedness. *Ann. Math. Statist.*, 22:587–592, 1951.

[25] A.F.S. Mitchell and W.J. Krzanowski. The mahalanobis distance and elliptic distributions. *Biometrika*, 72:464–467, 1985.

[26] R.J. Muirhead. *Aspects of Multivariate Statistical Theory.* John Wiley, New York, 1982.

[27] J.M. Oller. Information metric for extreme value and logistic probability distributions. *Sankhyà*, 49 A:17–23, 1987.

[28] J.M. Oller. *Statistical data analysis and inference*, chapter Some geometrical aspects of data analysis and statistics, pages 41–58. Elsevier science publishers B. V., North Holland, Amsterdam, 1989.

[29] J.M. Oller and C.M. Cuadras. Rao's distance for multinomial negative distributions. *Sankhyà*, 47 A:75–83, 1985.

[30] B.L.S. Prakasa Rao. Cram r-rao type integral inequalities for estimators of functions of multidimensional parameter. *Sankhyà*, 54 A:53–73, 1992.

[31] M. Spivak. *A Comprehensive Introduction to Differential Geometry, Vol I–V.* Publish or Perish, Berkeley, 1979.

Relació dels últims Preprints publicats:

- 195 *Separatrix splitting for a high-frequency perturbation of the pendulum.* V. G. Gelfreich. 1991 Mathematics Subject Classification: 34C37, 34C29, 58F30, 70H05. January 1996.

- 196 *A second order Stratonovich differential equation with boundary conditions.* Aureli Alabert and David Nualart. AMS Subject Classification: 60H10, 60H07. January 1996.

- 197 *A Gentzen system equivalent to the BCK-logic.* Romà J. Adillon and Ventura Verdú. Mathematics Subject Classification: 03B20. January 1996.

- 198 *On Gentzen systems associated with the finite linear MV-algebras.* Àngel J. Gil, Antoni Torrens Torrell and Ventura Verdú. Mathematics Subject Classification: 03B22, 03F03, 03G20, 03G99. February 1996.

- 199 *Exponentially small splitting of separatrices under fast quasiperiodic forcing.* Amadeu Delshams, Vassili Gelfreich, Àngel Jorba and Tere M. Seara. AMS classification scheme numbers: 34C37, 58F27, 58F36, 11J25. February 1996.

- 200 *Existence and regularity of the density for solutions to stochastic differential equations with boundary conditions.* Arturo Kohatsu-Higa and Marta Sanz-Solé. AMS 1990 SCI: 60H07, 60H10, 60H99. March 1996.

- 201 *A forcing construction of thin-tall Boolean algebras.* Juan Carlos Martínez. 1991 Mathematics Subject Classification: 03E35, 06E99, 54612. March 1996.

- 202 *Weighted continuous metric scaling.* C. M. Cuadras and J. Fortiana. AMS Subject Classification: 62H25, 62G99. April 1996.

- 203 *Homoclinic orbits in the complex domain.* V.F. Lazutkin and C. Simó. AMS Subject Classification: 58F05. May 1996.

- 204 *Quasivarieties generated by simple MV-algebras.* Joan Gispert and Antoni Torrens. Mathematics Subject Classification: 03B50, 03G99, 06F99, 08C15. May 1996.

- 205 *Regularity of the law for a class of anticipating stochastic differential equations.* Carles Rovira and Marta Sanz-Solé. AMS Subject Classification: 60H07, 60H10. May 1996.

- 206 *Effective computations in Hamiltonian dynamics.* Carles Simó. AMS Subject Classification: 58F05, 70F07. May 1996.

- 207 *Small perturbations in a hyperbolic stochastic partial differential equation.* David Márquez–Carreras and Marta Sanz–Solé. AMS Subject Classification: 60H15, 60H07. May 1996.

- 208 *Transformed empirical processes and modified Kolmogorov-Smirnov tests for multivariate distributions.* A. Cabaña and E. M. Cabaña. AMS Subject Classification: Primary: 62G30, 62G20, 62G10. Secondary: 60G15. June 1996.

- 209 *Anticipating stochastic Volterra equations.* Elisa Alòs and David Nualart. AMS Subject Classification: 60H07, 60H20. June 1996.

- 210 *On the relationship between α-connections and the asymptotic properties of predictive distributions.* J.M. Corcuera and F. Giummolè. AMS 1980 Subjects Classifications: 62F10, 62B10, 62A99. July 1996.

- 211 *Global efficiency.* J.M. Corcuera and J.M. Oller. AMS 1980 Subjects Classifications: 62F10, 62B10, 62A99. July 1996.