The Annals of Statistics
1995, Vol. 23, No. 5, 1562–1581

INTRINSIC ANALYSIS OF STATISTICAL ESTIMATION

BY J. M. OLLER AND J. M. CORCUERA[1]

Universitat de Barcelona

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 measures of an estimator's performance, like the bias and the mean square error, are clearly dependent on the statistical model parametrizations.

In this paper the authors work with extended notions of mean value and moments of 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 approach an intrinsic version of the Cramér–Rao lower bound is obtained: a lower bound, which depends on the intrinsic bias and the curvature of the statistical model, for the mean square of the Rao distance, the invariant measure analogous to the mean square error. Further, the behavior of the mean square of the Rao distance of an estimator when conditioning with respect to a sufficient statistic is considered, obtaining intrinsic versions of the Rao–Blackwell and Lehmann–Scheffé theorems. Asymptotic properties complete the study.

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 by proposing probability measures that belong 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 interval estimates no matter what specific inferences may eventually be needed. In this approach estimators supply different methods of induction.

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 would arise if closely related properties, like unbiasedness or uniformly minimum variance estimation, were preserved under coordinate system transformations. Unfor-

Received June 1993; revised October 1994.

[1]Supported in part by DGYCIT Grant (Spain) PB 91-249.

AMS 1991 *subject classifications*. 62F10, 62B10, 62A99.

Key words and phrases. Bias, mean square Rao distance, mean value, moments, Riemannian centre of mass, information metric, Cramér–Rao lower bounds, Rao–Blackwell, Lehmann–Scheffé.

tunately, this is not the case, as the reader can easily check by considering different parametrizations in a given parametric family, essentially due to the nontensorial character of the bias and the mean square error.

In this situation a natural question arises: could notions analogous 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. First, we may try to grant a privilege to a 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 (1951). 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 what 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. 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 results and to establish relationships between the classical noninvariant measures and the invariant ones herein obtained.

A related and interesting study has been developed by Hendriks (1991) about estimators with values in a manifold, but his purpose and mathematical framework are quite different. It is also important to consider the works of Kendall (1990) and Emery and Mokobodzki (1991) in developing the stochastic calculus in manifolds.

2. Moments and mean values. Let $(\mathscr{X}, \mathscr{A}, P)$ be a probability space, where $\mathscr{X}$ is the sample space, $\mathscr{A}$ is a σ-algebra of subsets of $\mathscr{X}$ and P is a probability measure on $\mathscr{A}$. 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, $f: \mathscr{X} \to M$, also called a *random object* on M, that is, a map such that, for all open sets $W \subset M$, $f^{-1}(W) \in \mathscr{A}$. We will now introduce the notion of mean value and moments of f, making the fewest necessary assumptions and maintaining the intuitive notion of *centrality* measure, in a closely related idea of *center of mass* as we shall see later [see Karcher (1977), Kobayashi and Nomizu (1969), Kendall (1990, 1991) and Emery and Mokobodzki (1991)].

Let $\mathscr{T}_s^r$ be the set of all C^∞ tensor fields in M, of order $r + s$, r-times contravariant and s-times covariant. If we fix $p \in M$, any map X from $\mathscr{X}$ to $\mathscr{T}_s^r$ induces a map X_p such that X_p: $\mathscr{X} \to T_s^r(M_p)$ with $X_p(x) = (X(x))_p$, where $T_s^r(M_p)$ denotes the space of (r, s)-tensors on the tangent space at p (M_p) having a natural topological vector space structure. Considering the Borel σ-algebra on $\mathscr{T}_s^r$ induced by the Borel σ-algebras of the M_p, a simple definition follows.

Definition 2.1. A C^∞ *random* (r, s)-*tensor field* on M (X) is a measurable map from $\mathscr{X}$ to $\mathscr{T}_s^r$.

It follows from the definition that $\forall\ p \in M$ the induced map X_p is a measurable map on $(\mathscr{X}, \mathscr{A})$.

Let $\otimes$ stand for the tensor field product. In the present context it is natural to define the kth-*order moment* of the random tensor field X as the ordinary (kr, ks)-tensor field on M

$$\mathscr{M}^k(X) = E\Big(\overbrace{X \otimes \cdots \otimes X}^{k}\Big) = \int_{\mathscr{X}} \overbrace{X(x) \otimes \cdots \otimes X(x)}^{k} P(dx), \qquad k \in \mathbb{N},$$

provided the integral exists.

In order to consider the mean value of a random object, 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, for every $p \in M$, called the exponential map $\exp_p \colon M_p \to M$. 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.

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

DEFINITION 2.2. 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 shall consider star-shaped neighbourhoods $V(p)$ such that $V(p) = -V(p)$ (i.e., *balanced neighbourhoods*) when we have only an affine connection, and balls in the Riemannian case. Both shall be referred to as *balls* with center 0_p.

DEFINITION 2.3. The image $W(p)$, by the exponential map, of an 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.

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 2.4. An open set $W(p)$ is said to be a *regular normal* neighbourhood of p iff its intersection with any normal ball with center p remains normal.

We can ensure the existence of neighbourhoods of this kind in the Riemannian case. This is because for every point p there is a ball in M_p with center 0_p where the exponential map is a diffeomorphism. It is also easy to see that a neighbourhood $W(p)$ of p is regular normal iff the shortest geodesic that joins p with any other point in $W(p)$ is unique and lies in $W(p)$. Thus regular normal neighbourhoods are a generalization of neighbourhoods with this property to the affine case.

Then, given a random object f taking values on a (Hausdorff and connected) manifold, equipped with an affine connection, 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 $x \in \mathscr{X}$, but if it is defined almost surely, we can introduce the following intrinsic *mean value* concept.

DEFINITION 2.5. A point on the manifold $p \in M$ is a *mean value* of the random object f if and only if there is a regular normal neighbourhood of p (W) such that $P\{f \in W\} = 1$ and

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

REMARKS. We shall write $\mathfrak{M}(f)$ to denote any mean value or the set of mean values, depending on the context. These mean values are, essentially, what Emery and Mokobodzki (1991) call *exponential barycenters.*

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

EXAMPLE 2.6. Let M be $\mathbb{R}^n$. Identifying the points with their coordinates on the trivial chart, and considering the usual Euclidean affine connection, we find, for $q, p \in \mathbb{R}^n$, that $\exp_p^{-1}(q) = q - p$. Therefore we recover the classical definition $\mathfrak{M}(f) = E(f) = \int_{\mathscr{X}} f(x)P(dx)$.

EXAMPLE 2.7. 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

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

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 that there are two mean values, given by ξ and $-\xi$. Compare this result with the *mean direction* defined in Kent, Mardia and Bibby [(1979), pages 424–451]. See also Jupp and Mardia (1989) for a comprehensive exposition.

EXAMPLE 2.8. 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 can supply, in the Riemannian case, a scalar *dispersion* measure with respect to a mean value p, $E(\rho^2(f(x), p))$, the ordinary expected value of the squared Riemannian distance between $f(x)$ and p, $\rho^2(f(x), p)$, which may be regarded as a coordinate-free version of the variance of a real random variable.

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

DEFINITION 2.9. The kth-*order moment* of the random object f is an ordinary $(k, 0)$-tensor field on M defined by $\mathscr{M}^k(f)_p \equiv \mathscr{M}^k(\exp_p^{-1}(f))$, $\forall\, p \in M$, $k \in \mathbb{N}$, provided the integral exists.

We can establish a relationship between the defined *mean value* and the Riemannian *center of mass*

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

where $\mathscr{H}_f(p) = \int_{\mathscr{X}} \rho^2(f(x), p) P(dx)$. First of all we have the following propositions.

PROPOSITION 2.10. *If there is a $q \in M$ such that $\mathscr{H}_f(q)$ is defined, then the function $\mathscr{H}_f(p)$ is defined for all $p \in M$, it is differentiable and*

$$X_p \mathscr{H}_f = -2\left\langle X_p, \int_{\mathscr{X}} \exp_p^{-1}(f(x)) P(dx) \right\rangle,$$

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

PROOF. By the triangular inequality $\mathscr{H}_f(p) \leq 2\mathscr{H}_f(q) + 2\rho^2(p, q)$, the first part of the proposition follows.

Since, fixing $q \in M$, $\rho^2(\cdot, q)$ is a C^∞ function for all $X_p \in M_p$, we can write

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

where the last equality can easily be checked considering a geodesic spherical coordinate system with origin q. Then we have $|X_p \rho^2(\cdot, q)| \leq 2\|X_p\|\, \rho(p, q)$. The proposition follows by the mean value and dominated convergence theorem. □

Now the connection between mean values and center of mass mentioned above is given by the following proposition.

PROPOSITION 2.11. *Let* $(\mathscr{X}, \mathscr{A}, P)$ *be a probability space, let* $(M, \mathfrak{A})$ *be a complete Riemannian manifold and let* $f\colon \mathscr{X} \to M$ *be a measurable map, such that* P_f *is dominated by the Riemannian measure* V_R, $P_f \ll V_R$. *Let the function* $\mathscr{H}_f$ *be as before. Then* $\mathscr{H}_f$ *has a critical point at* $p \in M$ *if and only if* p *is a mean value of* f.

PROOF. It is straightforward from the previous proposition and the fact that the *cut locus* of any $p \in M$ is a Riemannian measure-zero set. [See Spivak (1979).] □

REMARKS. From Proposition 2.11 we show that the defined *mean value concept* is weaker than the *center of mass concept*. Notice also that to define the first we only need an affine connection, while the second requires a Riemannian structure.

Moreover, we could give sufficient conditions to ensure the existence of a mean value whenever $\mathscr{H}_f(\cdot)$ would be defined. If we have a *regular convex* set in a complete manifold M (i.e., a set $A \subset M$ such that for any p, $q \in A$ the shortest geodesic from p to q is unique in M and lies in A) and the probability is concentrated on it, then the random object will have at least a mean value in the interior of A. [See also Kendall (1990).]

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 $\{(\mathscr{X}, \mathscr{A}, P_\theta);\ \theta \in \Theta\}$ 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 statistical model M with the n-dimensional C^∞ real manifold structure induced by this map. Let us denote this manifold by $(M, \mathfrak{A})$, where $\mathfrak{A}$ is the atlas induced by the parametrizations, that is, the coordinates in the parameter space.

In the dominated case, which we shall assume 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, that is, 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. The manifold $(M, \mathfrak{A})$ 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, $\partial \log p(x;\theta)/\partial\theta^i$, $i = 1,\dots,n$, are linearly independent and belong to $L^\alpha(p(\cdot;\theta)\,d\mu)$ for a suitable $\alpha > 0$.
4. The partial derivatives of the required orders,
$$\frac{\partial}{\partial\theta^i},\quad \frac{\partial^2}{\partial\theta^i\,\partial\theta^j},\quad \frac{\partial^3}{\partial\theta^i\,\partial\theta^j\,\partial\theta^k},\dots,\qquad i,j,k = 1,\dots,n,$$
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 (1985), Atkinson and Mitchell (1981), Barndorff-Nielsen and Blaesild (1987), Burbea (1986), Burbea and Rao (1982) and Oller (1989), among many others.

In this context, given a sample size k, an *estimator* $\mathscr{U}$ *for the true density function* (or probability measure) $p = p(\cdot;\theta) \in M$ *of the statistical model* is a measurable map
$$\mathscr{U}: \mathscr{X}^k \mapsto M,$$
assuming that the probability measure on $\mathscr{X}^k$ is $(P)_k(dx) = p_{(k)}(x;\theta)\mu_k(dx) = \prod_{i=1}^k p(x_i;\theta)\mu(dx_i)$.

Definition 3.1. An estimator $\mathscr{U}$ is *intrinsically unbiased* if and only if p is a mean value of $\mathscr{U}$ computed with respect to the true probability measure $(P)_k$, whatever the true density function $p \in M$.

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

Given the random vector field $A_p(x) = \exp_p^{-1}(\mathscr{U}(x))$, which we shall call *the estimator vector field*, it is convenient, in order to measure the bias, to introduce the following.

Definition 3.2. The *bias vector field* is defined as $B_p = E_p(A_p(x))$, or in components notation,
$$B^\alpha(\theta) = \int_{\mathscr{X}^k} A^\alpha(x;\theta)\,p_{(k)}(x;\theta)\,\mu_k(dx),\qquad \alpha = 1,\dots,n,$$
provided that they exist.

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

Proposition 3.3. *An estimator* $\mathscr{U}$ *is intrinsically unbiased if and only if its bias tensor field* B *is null.*

REMARKS. It is interesting to point out that an estimator is intrinsically unbiased if and only if it is *stationary* in the sense of Hendriks (1991) if we take ρ^2 as the loss function.

When the manifold associated to a regular statistical model is Euclidean, it is obvious that an estimator $\mathscr{U}$ will be intrinsically unbiased if and only if it is unbiased in the *affine* coordinate system.

EXAMPLE 3.4 (The univariate exponential distribution). Let us consider the exponential density function parametrized as

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

The metric tensor component is given by $g_{11}(\lambda) = 1/\lambda^2$. Then, if we consider the maximum-likelihood estimator for the parameter λ given by $\overline{X}$, the ordinary sample mean computed from a sample of size k, we have

$$B^1(\lambda) = E_\lambda\left(\left[\exp_\lambda^{-1}(\overline{X})\right]^1\right) = E_\lambda\left(\log\left(\frac{\overline{X}}{\lambda}\right)\right) = \Psi(k) - \log k,$$

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 an intrinsically unbiased estimator given by

$$\hat{\lambda} = \frac{k\overline{X}}{\exp[\Psi(k)]}.$$

EXAMPLE 3.5. 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 to 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, $\mu = (\mu_1, \dots, \mu_n)'$ is a parameter vector and F is a nonnegative function on $\mathbb{R}_+ = [0, \infty)$ satisfying

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

We have to assume, in addition, that

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

where $\mathscr{L}F = F'/F$, in order to ensure the existence of the information metric for this parametric family of probability distributions, given by

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

See Mitchell and Krzanowski (1985) and Burbea and Oller (1988) for more details.

Since the metric tensor field is constant, the manifold is Euclidean and the geodesics are straight lines. Identifying the manifold points with their coordinates, and considering the estimator in μ-coordinates given by the sample mean $\bar{X}$, we may write

$$E_\mu\left(\exp_\mu^{-1}(\bar{X})\right) = E_\mu(\bar{X} - \mu) = 0.$$

Therefore $\bar{X}$ is intrinsically unbiased.

4. Lower bound of mean square Rao distance. In this section, the relationship between unbiasedness and the mean square of the Rao distance, the Riemannian distance between the probability measure estimates and the true one, is studied. The inequality obtained is an *intrinsic* version of Cramér–Rao lower bound, and it is based on the comparison theorems of Riemannian geometry [see Kobayashi and Nomizu (1969), Chavel (1984) and Karcher (1977), among others]. Some related results, as we shall comment later, can be seen in Hendriks (1991).

First it is necessary to remember some basic definitions and results in differential geometry.

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

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

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

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

and

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

we know that $\exp_p$ maps $\mathfrak{D}_p$ diffeomorphically onto D_p [see Hicks (1965)]. In fact, D_p is the maximal regular neighbourhood of p in the sense that any other regular neighbourhood of p is included in it.

DEFINITION 4.1. Given two vector fields X and Y in M such that for any $p \in M$ they span a two-dimensional subspace, the sectional Riemannian curvature K is defined as the scalar function

$$K(X,Y) = \langle R(X,Y)X, Y\rangle / (\langle X,X\rangle\langle Y,Y\rangle - \langle X,Y\rangle^2),$$

where $R(X,Y)$ is the (curvature) operator,

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

and ∇ is the Riemannian connection of M.

In the two-dimensional case it coincides with the *Gaussian curvature* of a surface. In the general case the sectional curvature at a point $p \in M$ is the Gaussian curvature of the surface generated by the geodesics that start at p

and are tangent to the two-dimensional subspace generated by $X(p)$ and $Y(p)$ in M_p. If $n = 1$, we adopt the convention $K = 0$.

THEOREM 4.2 (Intrinsic Cramér–Rao lower bound). *Let $\mathscr{U}$ be an estimator corresponding to an n-dimensional regular parametric family of density functions for a sample of size k. Assume that manifold M is simply connected and $(P)_k \circ \mathscr{U}^{-1}(M \setminus D_p) = 0 \ \forall\ p \in M$. Let A be the estimator vector field, and let B be the corresponding bias vector field $B = E(A)$. Let us assume that the mean square of the Rao distance between the true density and an estimate, $E(\rho^2(\mathscr{U}, p))$, exists for all $k \in \mathbb{N}$ and that the covariant derivative of B exists and can be obtained by differentiating under the integral sign. Then the following hold:*

(i) *In general we have*

$$E(\rho^2(\mathscr{U}, p)) \ge \frac{\{\operatorname{div}(B) - E(\operatorname{div}(A))\}^2}{kn} + \|B\|^2,$$

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

(ii) *If all sectional curvatures are bounded from above by a nonpositive constant $\mathscr{K}$ and* $\operatorname{div}(B) \ge -n$, *then*

$$\begin{aligned} &E(\rho^2(\mathscr{U}, p)) \\ (1) \qquad &\ge \frac{\{\operatorname{div}(B) + 1 + (n-1)\sqrt{-\mathscr{K}}\|B\|\coth(\sqrt{-\mathscr{K}}\|B\|)\}^2}{kn} + \|B\|^2. \end{aligned}$$

(iii) *If all sectional curvatures are bounded from above by a positive constant $\mathscr{K}$, $d(M) < \pi/2\sqrt{\mathscr{K}}$, where $d(M)$ is the diameter of the manifold, and* $\operatorname{div}(B) \ge -1$, *then*

$$\begin{aligned} &E(\rho^2(\mathscr{U}, p)) \\ (2) \qquad &\ge \frac{\{\operatorname{div}(B) + 1 + (n-1)\sqrt{\mathscr{K}}d(M)\cot(\sqrt{\mathscr{K}}d(M))\}^2}{kn} + \|B\|^2. \end{aligned}$$

In particular, for intrinsically unbiased estimators, we have the following:

(iv) *If all sectional curvatures are nonpositive, then*

$$E(\rho^2(\mathscr{U}, p)) \ge \frac{n}{k}.$$

(v) *If all sectional curvatures are bounded from above by a positive constant $\mathscr{K}$ and $d(M) < \pi/2\sqrt{\mathscr{K}}$, then*

$$E(\rho^2(\mathscr{U}, p)) \ge \frac{1}{kn}.$$

The expectations, at each point p, are computed with respect to the corresponding probability measure $(P)_k$.

PROOF. (i) Let C be any vector field. Then, applying the Cauchy–Schwarz inequality twice,

$$E(|\langle A-B,C\rangle|) \le E(\|A-B\|\,\|C\|) \le \sqrt{E(\|A-B\|^2)}\sqrt{E(\|C\|^2)},$$

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

Let $C(x;\theta) = \operatorname{grad}(\log p_{(k)}(x;\theta))$, where $\operatorname{grad}(\cdot)$ stands for the gradient operator. Taking expectations and using the repeated index convention,

$$E(\|C\|^2) = E\big(g_{\alpha\beta}g^{\beta\gamma}\,\partial_\gamma \log p_{(k)} g^{\alpha\lambda}\,\partial_\lambda \log p_{(k)}\big) = kg_{\alpha\beta}g^{\beta\gamma}g^{\alpha\lambda}g_{\gamma\lambda} = kn,$$

where $\partial_\alpha = \partial/\partial\theta^\alpha$. Furthermore, we also have

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

and

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

Thus

$$|E(\langle A,C\rangle)| \le \sqrt{E(\|A\|^2) - \|B\|^2}\,\sqrt{kn},$$

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

$$\operatorname{div}(B) = E(\operatorname{div}(A)) + E(\langle A,C\rangle),$$

therefore (i) follows.

Fixing x, we now choose a convenient coordinate system in order to estimate $\operatorname{div}(A)$. Given p and $\mathscr{U}(x)$, we choose a geodesic spherical coordinate system with origin $\mathscr{U}(x)$, defined almost surely since $(P)_k \circ \mathscr{U}^{-1}(M \setminus D_{\mathscr{U}(x)}) = 0$.

It is clear that the components of tensor A are $(-\rho, 0, 0, \dots, 0)$ when ρ, the Riemannian distance between p and $\mathscr{U}(x)$, is the first coordinate. Additionally, using the repeated index convention

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

where Γ^α_{ij} are the Christoffel symbols of the second kind and g is the determinant of the metric tensor. Then

$$\operatorname{div}(A) = -\left(1 + \rho\frac{\partial\log\sqrt{g}}{\partial\rho}\right),$$

(ii) *Sectional curvature nonpositive*. If the sectional curvatures are bounded from above by $\mathscr{K} \le 0$, as a corollary of Bishop's comparison theorem (I) [see Chavel (1984), pages 38–39, 66–69], we have

$$\frac{\partial\log\sqrt{g}}{\partial\rho} \ge (n-1)\sqrt{-\mathscr{K}}\coth\big(\sqrt{-\mathscr{K}}\rho\big), \tag{3}$$

yielding

$$E(\langle A, C\rangle) \geq \operatorname{div}(B) + 1 + (n-1)\sqrt{-\mathscr{K}}\|B\|\coth(\sqrt{-\mathscr{K}}\|B\|) \geq 0,$$

where the first inequality is due to the fact that the function $\|u\|\coth(\|u\|)$ is convex and the second is because $\|u\|\coth(\|u\|) > 1$ and $\operatorname{div}(B) \geq -n$.

(iii) *All sectional curvatures are bounded from above by a positive constant* $\mathscr{K}$, $d(M) < \pi/2\sqrt{\mathscr{K}}$ *and* $\operatorname{div}(B) \geq -1$. As a corollary of Bishop's comparison theorem (II) [Chavel (1984), pages 38–39, 71–73], we have

$$\frac{\partial \log \sqrt{g}}{\partial \rho} \geq (n-1)\sqrt{\mathscr{K}}\cot(\rho\sqrt{\mathscr{K}}) \geq 0. \tag{4}$$

Then, taking into account that $u \cot u$ is a monotone decreasing function, it turns out that

$$E(\langle A, C\rangle) \geq \operatorname{div}(B) + 1 + (n-1)\sqrt{\mathscr{K}}d(M)\cot(\sqrt{\mathscr{K}}d(M)) \geq 0,$$

where the last inequality follows from the conditions $d(M) < \pi/2\sqrt{\mathscr{K}}$ and $\operatorname{div}(B) \geq -1$ since $u \cot u > 0$, $0 < u < \pi/2$.

[(iv) and (v)] These follow trivially from (ii) and (iii), with $\operatorname{div}(B) = 0$ and $\|B\| = 0$. □

REMARKS. In the Euclidean case, $K = 0$, it is not necessary to impose the condition $\operatorname{div}(B) \geq -n$. It should also be pointed out that the case (i) of Theorem 4.2 can be deduced from Hendriks [(1991), Theorem 3.2] if $B = 0$, the squared Riemannian distance ρ^2 is the loss function and the Hendriks' mapping ϕ is the identity of M. However, his setup is restricted to unbiased estimators. In our context we can also obtain a Cramér–Rao tensorial inequality, independently of the bias structure, for the following tensor.

DEFINITION 4.3. The *dispersion tensor field* of an estimator $\mathscr{U}$ is defined as

$$S_p = E(A \otimes A)(p) \equiv E_p(A_p \otimes A_p) \quad \forall\, p \in M,$$

where $A_p = \exp_p^{-1}(\mathscr{U})$.

Observe that in the Euclidean case $S - B \otimes B = \operatorname{Cov}(\mathscr{U})$.

PROPOSITION 4.4. *With the same conditions as in Theorem 4.2 and where the inequality means that the difference between two sides is a nonnegative definite tensor, we have*

$$S \geq \frac{1}{k}\operatorname{Tr}^{2,4}\big[\mathrm{G}^{2,2}[(DB - E(DA)) \otimes (DB - E(DA))]\big] + B \otimes B,$$

where $\operatorname{Tr}^{i,j}$ *and* $\mathrm{G}^{i,j}$ *are, respectively, the contraction and raising operators on index* i, j; D *is the covariant derivative*; *and the expectations are evaluated at any* $p \in M$ *with respect to the corresponding probability measure* $(P)_k$.

PROOF. Let

$$T = A - B - \frac{1}{k}\mathrm{Tr}^{2,3}\big[\mathrm{G}_*^{2,2}\big[E(A \otimes C) \otimes C\big]\big];$$

then the proof is straightforward by taking the expectation, at each point $p \in M$, with respect to $(P)_k$, of the tensor $T \otimes T$, where $\mathrm{G}_*^{i,j}$ is the *lowering* operator on index i, j [see Hicks (1965) for notation]. □

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 distributions which satisfy the hypothesis of the last theorem, such as the multinomial [see Atkinson and Mitchell (1981)], the negative multinomial [see Oller and Cuadras (1985)] and the extreme value distributions [see Oller (1987)], 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,

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

Finally, since the mean square Rao distance is bounded from above by $d(M)^2$, $d(M)$ being the diameter of the manifold, it turns out from Theorem 4.2(v) that a necessary condition for an unbiased estimator is $d(M) \geq 1/\sqrt{kn}$.

5. Conditional mean values of manifold-valued maps and the Rao–Blackwell theorem. Classically, we can decrease the mean square error for a given estimator by taking the conditional mean value with 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 the conditional mean value in this context and then obtain intrinsic versions of the Rao–Blackwell and Lehmann–Scheffé theorems.

Let $(\mathscr{X}, \mathscr{A}, P)$ 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 taking values in M with respect to a σ-algebra $\mathscr{D}$ on the sample space $\mathscr{X}$.

Moreover, if the mean square Riemannian distance of f exists, we can write

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

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

If for each $x \in \mathscr{X}$ there were one and only one stationary point $p \in M$ of $E(\rho^2(\cdot, f)|\mathscr{D})(x)$, that is, a point $p \in M$ such that

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

we would have a map from $\mathscr{X}$ 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 5.1. Let f be a random object on M, and let $\mathscr{D}$ be a σ-algebra on $\mathscr{X}$; we shall define the conditional mean value of f with respect to $\mathscr{D}$ as a $\mathscr{D}$-measurable map Z such that

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

REMARKS. We shall write $\mathfrak{M}(f|\mathscr{D})$ to denote any conditional mean value or the set of conditional mean values, depending on the context. From the remark at the end of Section 2, a sufficient condition to ensure the mean value exists is that there should be 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 be regular versions of the conditional probability of f given $\mathscr{D}$.

The following propositions are immediate.

PROPOSITION 5.2. *If f is a $\mathscr{D}$-measurable map, then $\mathfrak{M}(f|\mathscr{D}) = f$ a.e.-P.*

PROPOSITION 5.3. *If f is independent of $\mathscr{D}$, then $\mathfrak{M}(f|\mathscr{D}) = \mathfrak{M}(f)$ a.e.-P.*

REMARK. It is necessary to point out that, in general, $\mathfrak{M}(\mathfrak{M}(f|\mathscr{D})) \neq \mathfrak{M}(f)$, as observed in Emery and Mokobodzky (1991) and Kendall (1990) and as is easy to see from simple counterexamples.

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

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

$$\Delta^2_{\mathscr{U}}(p) = E_p\big(\rho^2(\mathscr{U}, p)\big) \quad \text{and} \quad \Delta^2_{\mathfrak{M}(\mathscr{U}|\mathscr{D})}(p) = E_p\big(\rho^2(\mathfrak{M}(\mathscr{U}|\mathscr{D}), p)\big).$$

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 theorem.

THEOREM 5.4 (Intrinsic Rao–Blackwell). *If we fix $p \in N$ and the square of the Rao distance $\rho^2(\cdot, p)$ is a convex function, then $\Delta^2_{\mathfrak{M}(\mathscr{U}|\mathscr{D})}(p) \le \Delta^2_{\mathscr{U}}(p)$.*

PROOF. This proof is adapted from Kendall (1990). By convexity, for all positive t,

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

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

$$\frac{d\gamma}{ds}(0)t = \exp_m^{-1}(q) \quad \text{and} \quad \operatorname{grad}(\rho^2)(0) = -2\exp_m^{-1}(p),$$

the above inequality can be written

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

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

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

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

$$\begin{aligned}\Delta^2_{\mathscr{U}}(p) &= E_p(\rho^2(\mathscr{U}, p)) = E_p\big(E_p(\rho^2(\mathscr{U}, p)|\mathscr{D})\big) \\ &\ge E_p(\rho^2(\mathfrak{M}(\mathscr{U}|\mathscr{D}), p)) = \Delta^2_{\mathfrak{M}(\mathscr{U}|\mathscr{D})}(p).\end{aligned}$$ □

REMARKS. It is interesting to note that if N is simply connected and the sectional curvatures in N are at most 0, or $\mathscr{K} > 0$ with $d(N) < \pi/2\sqrt{\mathscr{K}}$, then

$$\Delta^2_{\mathscr{U}}(p) \ge \Delta^2_{\mathfrak{M}(\mathscr{U}|\mathscr{D})}(p),$$

since these are sufficient conditions to ensure N is a regular convex set and the squared Riemannian distance is convex. 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 squared Riemannian distance, and then it is not necessarily true that the mean of the squared Riemannian distance between the true density and the estimated one should decrease when conditioning on $\mathscr{D}$.

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

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

Even though the bias tensor is not preserved in general when we condition with respect to a sufficient statistic, a theorem (which is analogous to the Lehmann–Scheffé theorem) can be formulated in the intrinsic context. We need first to redefine the notion of completeness.

DEFINITION 5.5. A sufficient statistic T is said to be *complete*, for M, if and only if, for any M-valued measurable maps f and g

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

implies that $f(T) = g(T)$ (a.e. $\forall\, p \in M$).

Then, with the same conditions as in Theorem 5.4, we directly have the following proposition.

PROPOSITION 5.6 (Intrinsic Lehmann–Scheffé). *Let $\mathscr{U}$ be an estimator that is a function of a complete sufficient statistic for M. Then it is the uniformly minimum Rao distance estimator for a fixed bias vector after conditioning.*

6. Asymptotic properties. First notice that, given a sequence of random objects taking values on an 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 rth-mean convergence, as 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, we can reduce the study of these convergences, with the exception of the rth mean, to the convergence of random sequences taking values on $\mathbb{R}^n$.

Since we prefer to use the term *estimator* instead the more cumbersome term *estimator sequence*, we shall redefine the estimator as a family of measurable maps

$$\mathscr{U} = \{\mathscr{U}_k : \mathscr{X}^k \mapsto M,\, k \in \mathbb{N}\}.$$

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

DEFINITION 6.1. An estimator $\mathscr{U}$ is *asymptotically intrinsically unbiased in a large sense* if and only if it is intrinsically unbiased asymptotically, that is, we can construct a sequence of mean values of $\mathscr{U}_k$ which converges to p. When the sequence of mean values is uniquely defined, we may write

$$\lim_{k\to\infty} \mathfrak{M}_p(\mathscr{U}_k) = p = p(\cdot;\theta) \quad \text{whatever } p \in M,$$

where $\mathfrak{M}_p$ stands for the mean value of $\mathscr{U}$ computed with respect to the true probability measure $(P)_k$, and we shall say that $\mathscr{U}$ is *asymptotically intrinsically unbiased.*

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

$$\sup_{k \in \mathbb{N}} E_p\big(\rho^2(\mathscr{U}_k, p)\big) < \infty$$

and the covariant derivative of the vector field $\mathscr{E}(q) = E_p(\exp_q^{-1}(\mathscr{U}_k))$ exists and can be obtained by differentiating under the integral sign. We shall also assume that the associated manifold of the regular parametric family of densities is complete, simply connected and has sectional curvatures K bounded from above and below [i.e., $\kappa < K < \mathscr{K}$ and if $\mathscr{K} > 0$, $d(M) < \pi/2\sqrt{\mathscr{K}}$, $d(M)$ being the diameter of the manifold]. Notice that in Theorem 4.2 we had analogous conditions, and it can be shown that these conditions are sufficient to ensure the existence and uniqueness of a mean value and that in fact the mean value is a centre of mass.

PROPOSITION 6.2. *A regular estimator $\mathscr{U}$ is asymptotically intrinsically unbiased if and only if, for the corresponding bias vector field,*

$$B_k(p) = E_p\big(\exp_p^{-1}(\mathscr{U}_k)\big),$$

which depends on the sample size k, we have

$$\lim_{k \to \infty} B_k(p) = 0 \quad \forall\, p \in M.$$

PROOF. Since $\operatorname{grad}(\mathscr{H}_k)(p) = -2B_k(p)$, then, following Karcher [(1977), Theorem (1.5)],

$$C(\mathscr{K})\,\rho\big(\mathfrak{M}_p(\mathscr{U}_k), p\big) \le \|B_k(p)\| \le c(\kappa)\,\rho\big(\mathfrak{M}_p(\mathscr{U}_k), p\big),$$

where $C(\mathscr{K})$ and $c(\kappa)$ are constants that depend on the curvature bound and where we have used the inequality $u \tanh u \le 1 + u$, $u \in \mathbb{R}^+$, and the fact that the second-order moments are uniformly bounded.

REMARK. If we consider the maximum-likelihood estimator for the univariate exponential distribution, we obtain (see Example 3.4) that $B_k(p) = \Psi(k) - \log k$, where $\Psi(k) = \Gamma'(k)/\Gamma(k)$. Then, since $\lim_{k \to \infty} k/\exp(\Psi(k)) = 1$, it turns out that it is asymptotically unbiased.

Now we introduce a definition of normal distribution on a manifold. There are several ways to build distributions on a manifold [see Jupp and Mardia (1989)].

As usual in this paper we are going to consider only random objects Z that take values, almost surely, on regular normal neighbourhoods (see Definition 2.4) of any point in a complete manifold M. For random objects of this kind 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 6.3. Let Z be a random object valued on a complete manifold M. We shall say that Z is normally distributed with respect to p and with parameters (η, Σ), where η is a vector and Σ is a positive definite 2-contravariant tensor in M_p, if there is a random vector $Y \sim N(\eta, \Sigma)$ on M_p, such that $Z = \exp_p(Y)$. We shall write $Z \sim N(\eta, \Sigma)_p$.

Notice that if $\eta = 0$ and the cut locus of p is empty, then p is a mean value of Z. This definition is in fact an extension of wrapped normal distribution as used in directional statistics [Jupp and Mardia (1989)]. It seems natural to introduce the concept of asymptotically normal distribution as follows. Let $\{Z_k\}_{k \in \mathbb{N}}$ be a sequence of M-valued random objects. Then we have the following definition.

DEFINITION 6.4. Let M be a complete manifold. A random sequence $\{Z_k\}_{k \in \mathbb{N}}$ 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 Σ in M_p such that

$$\left\{s_k \exp_p^{-1}(Z_k)\right\}_{k \in \mathbb{N}} \to_{\mathscr{L}} Y \quad \text{with } Y \sim N(0, \Sigma),$$

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

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

$$\left\{\exp_p\left(s_k \exp_p^{-1}(Z_k)\right)\right\}_{k \in \mathbb{N}} \to_{\mathscr{L}} Z \quad \text{with } Z \sim N(0, \Sigma)_p,$$

but if $\{X_k\}_{k \in \mathbb{N}} \to_{\mathscr{L}} Z$, it is not necessarily true that $\exp_p^{-1}(X_k)$ converges in law to a normal distribution.

We also say that *the estimator $\mathscr{U}$ is s_k-asymptotically normally distributed* if its corresponding random M-valued sequence is asymptotically normally distributed.

PROPOSITION 6.5. *Let $\mathscr{U}$ be an s_k-asymptotically normally distributed estimator of a regular parametric family of probability distributions, with mean $p \in M$. Also, assume that there is an $\varepsilon \in \mathbb{R}^+$ such that*

$$\sup_{k \in \mathbb{N}} E_p\left(\rho^{1+\varepsilon}(\mathscr{U}_k, p)\right) < \infty.$$

Then $\mathscr{U}$ is asymptotically intrinsically unbiased.

PROOF. It is straightforward since $\rho(\mathscr{U}_k, p) = \|\exp_p^{-1}(\mathscr{U}_k)\|$ are uniformly integrable. □

THEOREM 6.6. *With the assumptions of Proposition* 6.5, *maximum-likelihood estimators are asymptotically intrinsically unbiased.*

PROOF. This is an immediate consequence of Proposition 6.5, assuming sufficient conditions to ensure $\sup_{k \in \mathbb{N}} E_p(\rho^{1+\varepsilon}(\mathscr{U}_k, p)) < \infty$, for an $\varepsilon \in \mathbb{R}^+$, $\mathscr{U}$ being the maximum-likelihood estimator, by observing that maximum-likelihood estimators are $\sqrt{k}$-asymptotically normally distributed. In fact,

$$\sqrt{k}\, \exp_p^{-1}(\mathscr{U}_k) \to_{\mathscr{L}} N(0, (g^{\alpha\beta})),$$

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 exponential map at a point p of the manifold M. Let $(U, \theta(\cdot))$ be a local chart, where $\theta(p) = \theta_0$. Then, if we write $\hat{\theta}_{(k)} = \theta(\mathscr{U}_k)$,

$$\left(\exp_p^{-1}(\mathscr{U}_k)\right)^{\alpha} = \hat{\theta}^{\alpha}_{(k)} - \theta^{\alpha}_0 + \tfrac{1}{2}\Gamma^{\alpha}_{ij}\left(\hat{\theta}^{i}_{(k)} - \theta^{i}_0\right)\left(\hat{\theta}^{j}_{(k)} - \theta^{j}_0\right) + O(\rho^3).$$

Then with certain obvious conditions we can relate the intrinsic and the ordinary bias as follows.

PROPOSITION 6.7. *Let* $(U, \theta(\cdot))$ *be a local chart, where* $\theta(p) = \theta_0$; *let* $\mathscr{U}$ *be an estimator such that* $\sqrt{k}\, \exp_p^{-1}(\mathscr{U}_k)$ *converges in distribution to a random vector with mean zero and second-order moments*; $\sup_{k \in \mathbb{N}} E(k^2\rho^{3+\varepsilon}(\mathscr{U}_k, p)) < \infty$ *and the Christoffel symbols and their derivatives are uniformly bounded on the support of* $\{\mathscr{U}_k\}$. *Then*

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

with $B^{\alpha}(p) = E_p([\exp_p^{-1}(\mathscr{U}_k)]^{\alpha})$ *and* $\operatorname{Bias}(\hat{\theta}_{(k)}) = E_p(\hat{\theta}_{(k)} - \theta_0)$.

7. Concluding remarks. The parametrization invariance of an inference procedure has been valued as an important and desirable property by several authors [see Barndorff-Nielsen (1988) and Amari (1985), among others]. Notice, for instance, that we need this property if we want to use the *parametric bootstrap* in a consistent way. 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 the same conclusions if the tools used to reach a conclusion like the bias, the mean square error and so on depend on the parametrization. The Rao distance distance has been used as a tool in different approaches, but now we emphasize its use as the convenient distance between estimates, namely, the appropriate scale at which to observe and compare the estimates and consequently the estimators. The Rao distance has permitted us to build the intrinsic bias and mean square error, the invariant quantities whose classical analogues are, as we have shown, some kind of approximations.

Acknowledgments. We would like to thank Professor W. S. Kendall at Warwick University and the referees' reports for their interesting comments and suggestions.

REFERENCES

AMARI, S. (1985). *Differential-Geometrical Methods in Statistics. Lecture Notes in Statist.* **28**. Springer, New York.

ATKINSON, C. and MITCHELL, A. F. S. (1981). Rao's distance measure. *Sankhyā Ser. A* **43** 345–365.

BARNDORFF-NIELSEN, O. E. (1988). *Parametric Statistical Models and Likelihood. Lecture Notes in Statist.* **50**. Springer, New York.

BARNDORFF-NIELSEN, O. E. and BLAESILD, P. (1987). Strings: mathematical theory and statistical examples. *Proc. Roy. Soc. London Ser. A* **411** 155–176.

BURBEA, J. (1986). Informative geometry of probability spaces. *Exposition. Math.* **4** 347–378.

BURBEA, J. and OLLER, J. M. (1988). The information metric for univariate linear elliptic models. *Statist. Decisions* **6** 209–221.

BURBEA, J. and RAO, C. R. (1982). Entropy differential metric, distance and divergence measures in probability spaces: a unified approach. *J. Multivariate Anal.* **12** 575–596.

CHAVEL, I. (1984). *Eigenvalues in Riemannian Geometry.* Academic Press, New York.

EMERY, M. and MOKOBODZKI, G. (1991). Sur le barycentre d'une probabilité dans une variété. In *Séminaire de Probabilités XXV. Lecture Notes in Math.* **1485** 220–233. Springer, Berlin.

HENDRIKS, H. (1991). A Cramér–Rao type lower bound for estimators with values in a manifold. *J. Multivariate Anal.* **38** 245–261.

HICKS, N. J. (1965). *Notes on Differential Geometry.* Van Nostrand Reinhold, New York.

JUPP, P. E. and MARDIA, K. V. (1989). A unified view of the theory of directional statistics, 1975–1988. *Internat. Statist. Rev.* **57** 261–294.

KARCHER, H. (1977). Riemannian center of mass and mollifier smoothing. *Comm. Pure Appl. Math.* **30** 509–541.

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

KENDALL, W. S. (1991). Convexity and hemisphere. *J. London Math. Soc.* (2) **43** 567–576.

KENT, J. T., MARDIA, K. V. and BIBBY, J. M. (1979). *Multivariate Analysis.* Academic Press, New York.

KOBAYASHI, S. and NOMIZU, K. (1969). *Foundations of Differential Geometry.* Wiley, New York.

LEHMANN, E. L. (1951). A general concept of unbiasedness. *Ann. Math. Statist.* **22** 587–592.

MITCHELL, A. F. S. and KRZANOWSKI, W. J. (1985). The Mahalanobis distance and elliptic distributions. *Biometrika* **72** 464–467.

OLLER, J. M. (1987). Information metric for extreme value and logistic probability distributions. *Sankhyā Ser. A* **49** 17–23.

OLLER, J. M. (1989). Some geometrical aspects of data analysis and statistics. In *Statistical Data Analysis and Inference* (Y. Dodge, ed.) 41–58. North-Holland, Amsterdam.

OLLER, J. M. and CUADRAS, C. M. (1985). Rao's distance for multinomial negative distributions. *Sankhyā Ser. A* **47** 75–83.

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

DEPARTAMENT D'ESTADÍSTICA
UNIVERSITAT DE BARCELONA
08028 BARCELONA
SPAIN