UNIVERSITAT DE BARCELONA

ON RAO DISTANCE ASYMPTOTIC DISTRIBUTION

by

Jacob Burbea and *Josep M. Oller*

AMS Subject Classification: 62B10, 62H12

Mathematics Preprint Series No. 67
June 1989

ON RAO DISTANCE ASYMPTOTIC DISTRIBUTION

Jacob Burbea

Department of Mathematics and Statistics
University of Pittsburgh
Pittsburgh, PA 15260, USA

and

Josep M. Oller

Department of Statistics
University of Barcelona
08028 Barcelona, Spain

ABSTRACT

This paper is concerned with the study of the asymptotic properties of Rao distance maximum-likelihood estimation obtained from two independent samples of two statistical populations, under certain regularity conditions. These results allow the construction, by means of the Rao distance, of a statistical test to compare different populations, which may be regarded as an alternative to the standard likelihood ratio test. Furthermore, a geometrical model is supplied which can be used to obtain graphical outputs through numerical taxonomy methods or multidimensional scaling techniques. Finally, some of the previously-mentioned methods are illustrated by means of a specific example based on the univariate normal distribution.

1. INTRODUCTION

Frequently in statistics and data analysis it is not enough to use hypothesis test theory in order to solve the wide class of problems in applied research, since it is often interesting not only to decide between several alternative hypotheses but also to compare different experimental conditions, studying how different they are. For this purpose it is convenient to use geometrical models which help us to visualize the analogies and differences between them.

On the other hand, the use of geometrical models is implicit in many statistical techniques, like analysis of variance and multivariate analysis. All these considerations have been pointed out earlier, from the works of Pearson (1902, 1926), Mahalanobis (1930, 1936) and Fisher (1925) among others.

Later, in many areas of applied research and also in statistics, many works appeared which make use of geometrical models. Consider for instance Rao (1945) and Matusita (1964) in statistics, Rao (1948) in anthropology, Prevosti et al. (1975) in genetics, Legendre and Legendre (1979) in ecology, among many others in different fields of study. In all these cases the distance used is considered as a measure of the information we have about the differences between the objects compared.

In the above-mentioned paper of Rao, (1945), he introduced a Riemannian metric, through the Fisher information matrix, over a manifold of parametric probability distributions, and proposed the Riemannian distance induced by the metric as a measure of dissimilarity between two probability distributions. Since then, many statisticians have attempted to construct a geometrical theory of probability spaces. We may mention the works of Efron (1975) who introduced the concept of statistical curvature, Amari (1968, 1980, 1982, 1985), Čenčov (1965, 1982), who introduced new affine connections, Atkinson and Mitchell (1981), Burbea and Rao (1982, 1984), Burbea (1985) and also the works of Campbell (1985, 1986), Oller and Cuadras (1985), Oller (1987), Burbea and Oller (1988), Mitchell (1988) and Skovgaard (1984) who studied different aspects of the information metric, among others.

In the present paper we study the asymptotic distribution of certain natural transformations of maximum likelihood estimation of the Rao distance, obtained from two independent samples of two statistical populations, under certain regularity conditions. From

these results we propose some statistical hypothesis tests in order to compare two or more parametric statistical populations, which may be viewed as an alternative to the standard likelihood ratio tests. Lastly, we apply these methods to the problem of comparing two univariate normal statistical populations.

2. NOTATION AND REMARKS

We first introduce some notation. Let χ be a sample space and $\mathcal{A}$ a σ-algebra of subsets of χ. Let μ be a σ finite additive positive measure on the measurable space $(\chi, \mathcal{A})$ which we shall call a **reference measure**. We denote by $\mathcal{L}^r \equiv \mathcal{L}^r(\chi : \mu)$ the usual Lebesgue space of μ-measurable real valued functions p on χ so that:

$$\|p\|_{\mu,r} \equiv \{\int_\chi |p(x)|^r \, d\mu(x)\}^{\frac{1}{r}} \qquad r > 0$$

and let $\mathcal{D}$ be the convex subset of $\mathcal{L}^1$ consisting in all μ-almost everywhere positive functions $p \in \mathcal{L}^1$ such that $\|p\|_{\mu,1} = 1$. In the probability context, $\mathcal{D}$ is the set of density functions corresponding to the measure μ.

Now we introduce the following definition:

Definition 1.

Given a reference measure μ, an **n-parametric family of probability density functions** is a C^∞ n-dimensional manifold (D, a_D) where $D \subset \mathcal{D}$ and a_D is a maximal C^∞ atlas for D. Since this concept is μ-dependent, we shall use the following notation: $((D, a_D), \mu)$.

Clearly, from the Lebesgue-Radon-Nykodim theorem, a parametric family of probability density functions is a representation of a parametric family of probability measures on $(\chi, \mathcal{A})$.

Frequently the manifold (D, a_D) is defined directly, by an auxiliary function f such that $f : \chi \times \Theta \longrightarrow \mathbf{R}$, $f(x, \theta) \geq 0$ for μ-almost all $x \in \chi$, $\forall \theta \in \Theta$, where Θ is an open set of $\mathbf{R}^n$, and additionally $\int_\chi f(x, \theta) \, d\mu(x) = 1 \quad \forall \theta \in \Theta$. In this case the manifold (D, a_D) is defined as:

$$D = \{p \in \mathcal{D} \; : \; p = f(\cdot, \theta) \quad \theta \in \Theta\}$$

and the atlas a_D is the maximal atlas which contains the chart (θ, D) where $\theta : D \longrightarrow \mathbf{R}^n$ and $\theta(p) = \tilde{\theta}$ with $p = f(\cdot, \tilde{\theta})$. We shall call Θ the **parametric space,** which is obviously an n-dimensional manifold on $\mathbf{R}^n$.

Given an n-parametric family of probability density functions, $((D,\, a_D),\, \mu)$ and given a local chart, $(\theta,\mathcal{U})$, $\mathcal{U} \subset D$, it is customary to write $p(x|\tilde{\theta})$ or $p(\cdot|\tilde{\theta})$ for $p \in \mathcal{U}$, where $\tilde{\theta} = \theta(p)$: that is, the components of $\tilde{\theta}$ are the coordinates of p under the coordinate system induced by θ. Also, given $x \in \chi$ and given a certain real function h, if we define a function on D through $\Phi : \mathcal{U} \longrightarrow \mathbf{R}$, $\Phi(p) = h\big(p(x)\big)$, we usually write:

$$\partial_{\theta_i} h\big(p(x|\theta)\big) \equiv D_i\big(\Phi \circ \theta^{-1}\big)\big(\theta(p)\big) \qquad i = 1, ..., n$$

moreover,

$$\partial^2_{\theta_i \theta_j}\, h\big(p(x|\theta)\big) \equiv D_{ij}\left(\Phi \circ \theta^{-1}\right)\big(\theta(p)\big), \qquad \text{and so on.}$$

Also, $E_\theta\big(s(X)\big) \equiv \int_\chi s(x)\; p(x|\theta)\; d\mu(x)$ stands for the mean value of $s(X)$ with respect to the density $p(\cdot|\theta)$, provided the above integral exists. In particular let g_{ij} be the coefficients of the **Fisher information matrix,** $G = (g_{ij})$, defined as $g_{ij}(\theta) \equiv E_\theta\big(\partial_{\theta_i} log\, p(X|\theta)\; \partial_{\theta_j}\; log\, p(X|\theta)\big) \quad i,j = 1, ..., n$ provided they exist. Notice that g_{ij} are both dependent on a local chart $(\theta,\mathcal{U})$ and the point $p \in \mathcal{U}$. It will be more rigorous to write $g_{ij}\big(\theta(p)\big)$, but it is customary to write $g_{ij}(\theta)$, taking into account that $\theta = \theta(p)$.

Let us now consider an independent sample of size k from the sample space χ, $X \in \chi^k$, $X = (x_1, ..., x_k)$. The joint density function is given by:

$$p(x_1, x_2, ..., x_k \,|\, \theta) = p(x_1 \,|\, \theta) : ... \cdot p(x_k \,|\, \theta)$$

and given a sample $X \in \chi^k$, the **likelihood function** L_X, on D, is defined as: $L_X : D \longrightarrow \mathbf{R}$, $L_X(p) = \prod_{i=1}^k p(x_i|\theta)$, where $\theta = \theta(p)$. It is customary to write $L_X(\theta)$ instead of $L_X(p)$, in other words a coordinate-dependent approach is usually followed, but it is also possible to consider the previous coordinate-free approach.

The **likelihood equations** are given by

$$\partial_{\theta_i}\; log\, L_X(\theta) = \sum_{j=1}^{k} \partial_{\theta_i}\; log\, p(x_j \,|\, \theta) = 0 \qquad i = 1, ..., n$$

here "X" is fixed and "θ" are the unknowns. From these equations we usually obtain the **maximum-likelihood estimation** of the parameters θ. It is also possible to consider a coordinate-free approach in order to define the maximum-likelihood estimation of a density. Consider L_X or a monotonous transformation of the likelihood function, as the log-likelihood, $\ell(p) = log\,\big(L_X(p)\big)$, which under the usual regularity condition is a C^1 function of D. Finding the solutions of likelihood equations is equivalent findig the **critical points** of ℓ, that is finding the points $p \in M$ such that the differential of ℓ, ℓ_*, at p is the

zero linear map. In other words, p^* is a critical point of ℓ, if and only if $\theta^* = \theta(p^*)$ is a solution of the likelihood equations, which means that every critical point of ℓ defines a solution of likelihood equations and the converse.

It is useful to introduce some additional notation. We denote by $\{X_n\}_{n\in\mathbf{N}} \xrightarrow{\mathcal{L}} X$ and $\{X_n\}_n \xrightarrow{\mathcal{P}} X$, that the random sequence $\{X_n\}_{n\in\mathbf{N}}$ **converges in law** or **in probability,** respectively, to the random variable or vector X.

We understand by $X \sim N_k(\mu, \Sigma)$ that X is a k-dimensional random vector distributed as a k-variate normal distribution with μ mean vector and Σ covariance matrix.

For the purpose of our paper, we shall restrict our study to certain n-parametric families of probability density functions which satisfy certain regularity conditions.

Let $((D, a_D), \mu)$ be an n-parametric family of probability density functions. We shall consider the following assumptions:

A1.- (D, a_D) is a connected Hausdorff manifold.

There exists a sub-atlas $\mathcal{Y}$ contained in a_D such that for every local chart (θ, V) of the atlas $\mathcal{Y}$, then:

A2.- $\partial_{\theta_i} \log p(x|\theta)$, $\partial^2_{\theta_i\theta_j} \log p(x|\theta)$, $\partial^3_{\theta_i\theta_j\theta_k} \log p(x|\theta)$ $\quad i,j,k = 1,...,n$ almost certainly exists for all $p \in V$, $\theta = \theta(p)$.

A3.- The moments of $\partial_{\theta_i} log\ p(x|\theta)$ $\quad i = 1,\ldots,n$ exist up to second order satisfying: $E_\theta\big(\partial_{\theta_i} \log p(X|\theta)\big) = 0$ and $E_\theta\Big(\big(\partial_{\theta_i} \log p(X|\theta)\big)^2\Big) < \infty \qquad i = 1,...,n$ for all $p \in V$, $\quad \theta = \theta(p)$.

A4.- The functions $\partial_{\theta_1} log\, p(x|\theta), ..., \partial_{\theta_n} log\, p(x|\theta)$ are linearly independent, that is if $\lambda_1\, \partial_1\ log\, p(x|\theta) + ... + \lambda_n\, \partial_n\ log\, p(x|\theta) = 0$ almost everywhere, then $\lambda_1 = ... = \lambda_n = 0$.

A5.- The components of the Fisher information matrix, $g_{ij}(\theta)$, are C^∞ functions, whatever coordinate system we consider, that is for every local chart (θ, V).

A6.- The expected value of $\frac{1}{p(x/\theta)}\partial^2_{\theta_i\theta_j}\ p(x|\theta)$ $\quad i,j = 1,...,n$ exists and is equal to zero: $E_\theta\big(\frac{1}{p(x/\theta)}\partial^2_{\theta_i\theta_j} p(X|\theta)\big) = 0$.

A7.- There exists a function $M(x)$ such that:

$$|\partial^3_{\theta_i\theta_j\theta_k} log\, p(x|\theta)| < M(x) \qquad i,j,k = 1,...,n$$

almost surely, for all $p \in V$, $\quad \theta = \theta(p)$ and $E_\theta\big(M(X)\big) < \infty$ for all θ values.

We shall call an n-parametric family of probability density functions that satisfies A1-A7 a **regular n-parametric family of probability density functions.**

Under the above assumptions, the coefficients of the Fisher information matrix are

finite, at each point $p \in D$ and for every local chart (θ, V), and the matrix G, which depends on the point $p \in D$ and a local chart (θ, V), is a strictly positive symmetric matrix. Furthermore, the coefficients of G, g_{ij}, are the components of a second order covariant symmetric tensor field on the manifold (D, a_D). Since at each point G is a strictly positive definite matrix, this tensor field may be considered as the metric tensor field of the manifold, and therefore (D, a_D) is a Riemannian manifold. We denote by $\langle v, w \rangle_p$ the inner product of two tangent vectors and $ds^2 = \sum_{ij=1}^{n} g_{ij}\, d\theta^i \cdot d\theta^j$ the quadratic differential form.

The Riemannian distance is called the Rao distance of the parametric family of probability density functions, which is a representation model for a parametric family of probability measures.

We now introduce some additional differential geometry notation. Let p be a point of $D, p \in D$, and let D_p be the tangent space at p. Given $v \in D_p$, there exists a unique geodesic γ_v through p, whose tangent vector at p is v. This geodesic is defined in a neighbourhood of p. In this context the **exponential map** at p, exp_p is defined as: $exp_p(v) = \gamma_v(1)$ for all $v \in D_p$ such that 1 is in the domain of γ_v. Consider an open neighbourhood $\mathcal{U}$ of $0 \in D_p$ such that the exponential map will be a diffeomorphism, and let $e_1, ..., e_n$ be an orthonormal basis for D_p. Let $x^1, ..., x^n$ be defined as $x^i(v) = \langle v, e_i \rangle_p \quad i = 1, ..., n, \quad v \in D_p$. We define a coordinate system on $exp_p(\mathcal{U})$, through $x^1 \circ exp_p^{-1}, ..., x^n \circ exp_p^{-1}$. In other words, the coordinates of the point $exp_p(\lambda_1 e_1 + ... + \lambda_n e_n)$ are $(\lambda_1, ..., \lambda_n)$ for all points such that $\lambda_1 e_1 + ... + \lambda_n e_n \in \mathcal{U}$. This coordinate system is called a **normal coordinate system.**

Let us now define $S_p \subset D_p$ as $S_p = \{v \in M_p : \|v\|_p = 1\}$, where $p \in D$, for a complete manifold D. For each $v \in S_p$ we define

$$C(v) = \sup \{t > 0 : d(p, \gamma_v(t)) = t\}$$

It is well known, see for instance Spivak (1979) or Chavel (1984), that $C(v) > 0$ for all $v \in S_p$. Notice that if $t \in (0, C(v))$ then $d(p, \gamma_v(t)) = t$. Also, for $t \in (0, C(v)), \gamma_v$ must be the only minimizing geodesic connecting p to $\gamma(t)$. When $C(v) < +\infty$, $\gamma_v(C(v))$ is called the **cut point** of p along γ_v. Now define a set $\Delta_p \subset D_p$ as $\Delta_p = \{t\, v \in D_p : 0 \leq t < C(v),\ v \in S_p\}$ and also $\mathcal{U}_p \subset D$, as: $\mathcal{U}_p = exp_p(\Delta_p)$. The boundary of $\mathcal{U}_p, \partial\mathcal{U}_p$, is called the **cut locus** of p in D. Notice that Δ_p is a star-shaped open region in D_p which is mapped smoothly, by exp_p onto $\mathcal{U}_p$, an open subset of D.

Now we shall introduce **geodesic spherical coordinates.** Consider a point $q \in \mathcal{U}_p, q \neq p$, such that $q = \gamma_v(t_o), v \in S_p, t_o \in (0, C(v))$. Assume we are given a

coordinate system in an open neighbourhood of v on S_p defined through a local map on S_p, (ξ, V), where V satisfies $V \subset S_p$, $v \in V$, $\xi : V \longrightarrow \mathbb{R}^{n-1}$, ξ a 1-1 map onto an open set in $\mathbb{R}^{n-1}$, $\xi(V)$. We may also assume $\xi(v) = 0 \in \mathbb{R}^{n-1}$. Now we can define a coordinate system on an open neighbourhood of q on $\mathcal{U}_p$, assigning to the point $exp_p\big(t \cdot \xi^{-1}(\theta)\big)$ the coordinates $(t, \theta) \in \mathbb{R}^n$, $\theta \in \mathbb{R}^{n-1}$. Observe that the point q has coordinates $(t_o, 0)$. More formally, if we define the map $\Psi : B_q \longrightarrow \mathbb{R}^n$ such that:

$$z \longrightarrow \Psi(z) = \big(\|z\|_p\,,\ \xi(z/\|z\|_p)\big)$$

where $B_q = \{z \in \Delta_p : z = \lambda \cdot y \quad \lambda > 0,\ y \in V\} \subset D_p$ and if we let $\pi^1, ..., \pi^n$ be the usual coordinate system on $\mathbb{R}^n$, $\pi^i : \mathbb{R}^n \longrightarrow \mathbb{R}$, $\pi^i(a) = a_i \quad i = 1, ..., n$, a geodesic spherical coordinate is given by: $\pi^1 \circ \Psi \circ exp_p^{-1}, ...,$ $\pi^n \circ \Psi \circ exp_p^{-1}$ which domain is included in $\mathcal{U}_p - \{p\}$, say $W_q = exp_p(B_q)$. Let r be a point on W_q, $r \in W_q$, the natural basis induced by the above coordinate system in D_r is given by:

$$(\partial_i)_r f = D_i\big(f \circ exp_p \circ \Psi^{-1}\big)\big(\Psi \circ exp_p^{-1}(r)\big) \qquad i = 1, ..., n$$

where f is smooth function defined in a neighbourhood of r. It is well known, see for instance Cheeger and Ebin (1975), that in a geodesic spherical system of coordinates we have:

$$\langle (\partial_1)_r\,, (\partial_1)_r \rangle = 1 \qquad \langle (\partial_1)_r\,, (\partial_i)_r \rangle = 0 \qquad i = 2, ..., n$$

and therefore the quadratic differential form is given by:

$$ds^2 = dt^2 + \sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij}\ d\theta^i\ d\theta^j$$

as a consequence of Gauss' lemma. Notice that without loss of generality we may assume that in the point q, the metric tensor coefficients $g_{ij}(q) = \delta_{ij} \quad i, j = 2, ..., n$ where δ_{ij} are the usual Kronecker deltas.

3. PREVIOUS RESULTS

In the present section we shall develop some lemmas that we shall need latter. Some references are Hicks (1965), Helgason (1978), Ash (1972), Laha and Rohatgi (1979), Billingsley (1968) and Giné and Araujo (1980).

Lemma 1

Let $\{U_n\}_{n\in\mathbb{N}}$ be a sequence of k-dimensional random vectors, which satisfy $\sqrt{n}(U_n-\theta)\xrightarrow{\mathcal{L}} X\sim N_k(0,\Sigma)$, where $\mathcal{L}$ stands for the convergence in law, $\theta\in\mathbb{R}^k$, X is a random vector distributed as a k-variate normal distribution and Σ is a $k\times k$ strictly positive definite symmetric matrix.

Let g be a real function defined on A, $g:A\longrightarrow\mathbb{R}$, where $A\subset\mathbb{R}^k$ is an open set, and $\theta\in A$. Assume that g is differentiable at $x=\theta$ and let $g'(\theta)\equiv c$ be the Jacobian matrix $(1\times k)$ at $x=\theta$.

i) If $c\neq 0$, then $\sqrt{n}\big(g(U_n)-g(\theta)\big)\xrightarrow{\mathcal{L}} Z\sim N_1(0,\,c\,\Sigma\,c')$ that is, the above sequence, defined with probability tending to one when $n\longrightarrow\infty$, converges in law to a random variable Z distributed as univariate normal distribution with zero mean and $c\,\Sigma\,c'>0$ variance.

ii) If $c=0$, then $\sqrt{n}\big(g(U_n)-g(\theta)\big)\xrightarrow{\mathcal{P}} 0$, that is, converges in probability and therefore in law to zero.

iii) Assume additionally that g is C^2 in a neighbourhood of θ, $c=0$, and let $H(\theta)\equiv A$ be the Hessian matrix of g at $x=\theta$, where A is a $k\times k$ matrix of rank r, $0<r\leq k$. Then the sequence $n\big(g(U_n)-g(\theta)\big)$ defined with probability tending to one when $n\longrightarrow\infty$, converges in law to a random variable J, which is a linear combination of the squares of independent random variables $Z_1,...,Z_r$, each one having a univariate normal distribution with zero mean and unit variance. The coefficients of the linear combination are the non null eigenvalues of the matrix $\frac{1}{2}\,\Sigma^{1/2}\,A\,\Sigma^{1/2}$, that is:

$$n\big(g(U_n)-g(\theta)\big)\xrightarrow{\mathcal{L}} J=\lambda_1\,Z_1^2+...+\lambda_r\,Z_r^2$$

where $Z_i\sim N_1(0,1)$ is stochastically independent and $\lambda_1,...,\lambda_r$ are the r non-null eigenvalues of $\frac{1}{2}\,\Sigma^{1/2}\,A\,\Sigma^{1/2}$.

The proofs of sections i) and ii) are, esentially in Rao (1973) p. 387 section. iii) is similar to section iii) of lemma 3 and is omitted here.

Lemma 2

Let $\{U_n\}_{n\in\mathbb{N}}$ and $\{V_n\}_{n\in\mathbb{N}}$ be two stochastically independent k-dimensional random vector sequences, satisfying:

$$\{U_n\}_{n\in\mathbb{N}}\xrightarrow{\mathcal{L}} X\sim N_k(0,\Sigma)\quad,\quad\{V_n\}_{n\in\mathbb{N}}\xrightarrow{\mathcal{L}} Y\sim N_k(0,\,\Sigma)$$

that is U_n and V_n converge in law to X and Y, respectively, where X and Y are k-dimensional random vectors identically distributed as a k-variate normal distribution with zero mean and Σ covariance matrix, a $k\times k$ symmetric and strictly positive definite matrix.

Let a_{nm} and b_{nm} be two double sequences of real numbers, satisfying: $a_{nm}^2 + b_{nm}^2 = 1$. Then,

$$\{a_{nm}\, U_n + b_{nm}\, V_m\}_{(n,m)\in\mathbb{N}^2} \xrightarrow{\mathcal{L}} Z \sim N_k(0,\, \Sigma)$$

that is, the double sequence $a_{nm}\, U_n + b_{nm}\, V_m$ converges in law to a random vector Z distributed as $N_k(0,\, \Sigma)$.

Proof:

It follows from the fact that the characteristic functions sequences of U_n and V_m, converge uniformly in every $\mathbb{R}^k$ compact set to the zero mean k-variate normal characteristic function:

$$\varphi(t) = e^{-\frac{1}{2}\, t'\, \Sigma\, t} \qquad t \in \mathbb{R}^k$$

since $|a_{nm}| \leq 1$, $|b_{nm}| \leq 1$ and taking into account that $a_{nm}^2 + b_{nm}^2 = 1$ we have:

$$\varphi(a_{nm}t) \cdot \varphi(b_{nm}t) = e^{-\frac{1}{2}\, a_{nm}^2\, t'\, \Sigma\, t}\; e^{-\frac{1}{2}\, b_{nm}^2\, t'\, \Sigma\, t} = e^{-\frac{1}{2}\, t'\, \Sigma\, t}$$

and therefore $a_{nm}\, U_n + b_{nm}\, V_m$ converges in law to a k-variate normal distribution $N_k(0,\, \Sigma)$.

Lemma 3

Let $\{U_n\}_{n\in\mathbb{N}}$ and $\{V_n\}_{n\in\mathbb{N}}$ be two stochastically independent k-dimensional random sequences, satisfying:

$$\sqrt{n}(U_n - \theta_1) \xrightarrow{\mathcal{L}} X \sim N_k(0,\, \Sigma) \quad \text{and} \quad \sqrt{m}(V_m - \theta_2) \xrightarrow{\mathcal{L}} Y \sim N_k(0,\, \Sigma)$$

where $\theta_1,\, \theta_2 \in \mathbb{R}^k$, X and Y are k-dimensional random vectors, distributed as a k-variate normal distribution, with zero mean and Σ covariance matrix, a $k \times k$ symmetric and strictly positive definite matrix.

Let g be a real function defined on $A \times B$, $g : A \times B \longrightarrow \mathbb{R}$, where $A,\, B \subset \mathbb{R}^k$ are open sets, $\theta_1 \in A$, $\theta_2 \in B$. Assume that g is differentiable at $(\theta_1,\, \theta_2)$ and the Jacobian is $g'(\theta_1,\, \theta_2) \equiv (c_1,\, c_2)$, where c_1 and c_2 are $1 \times k$ matrix such that $c_1\, \Sigma\, c_1' = c_2\, \Sigma\, c_2' \equiv c\, \Sigma c'$, for a certain c, $1 \times k$, matrix.

i) If $c_i \neq 0$ then $\frac{\sqrt{nm}}{\sqrt{n+m}}\, \left(g(U_n,\, V_m) - g(\theta_1,\, \theta_2)\right) \xrightarrow{\mathcal{L}} Z \sim N_1(0,\, c\, \Sigma\, c')$ that is, the above double sequence, defined with probability tending to one as $n \longrightarrow \infty$, converges in law to a univariate normal distribution, with zero mean and $c\, \Sigma\, c' > 0$ variance.

ii) If $c_1 = c_2 = 0$, then $\frac{\sqrt{nm}}{\sqrt{n+m}}\, \left(g(U_n,\, V_m) - g(\theta_1,\, \theta_2)\right) \xrightarrow{\mathcal{P}} 0$ that is, it converges in probability and also in law to zero.

iii) We further assume that g is C^2 in a neighbourhood of (θ_1, θ_2), $c = 0$ and the Hessian matrix of g at (θ_1, θ_2), $H(\theta_1, \theta_2) \equiv \begin{pmatrix} A & -A \\ -A & A \end{pmatrix}$ where A is a symmetric $k \times k$ matrix of rank r, $0 < r \leq k$. Then, the double sequence $\frac{nm}{n+m}\left(g(U_n, V_m) - g(\theta_1, \theta_2)\right)$ converges in law to a random variable J which is a linear combination of the squares of the independent random variables $Z_1, ..., Z_r$, each one having a univariate distribution with zero mean and unit variance. The coefficients of this linear combination are the non-null eigenvalues of the matrix $\frac{1}{2}\,\Sigma^{1/2}\,A\,\Sigma^{1/2}$, that is:

$$\frac{nm}{n+m}\left(g(U_n, V_m) - g(\theta_1, \theta_2)\right) \xrightarrow{\mathcal{L}} J = \lambda_1 Z_1^2 + ... + \lambda_r Z_r^2$$

where $Z_i \sim N_1(0,1)$ is stochastically independent, and $\lambda_1, ..., \lambda_r$ are the r non-null eigenvalues of $\frac{1}{2}\,\Sigma^{1/2}\,A\,\Sigma^{1/2}$.

Proof:

i) Let us define $h(u,v) = \frac{g(u,v) - g(\theta_1,\theta_2) - Dg(\theta_1,\theta_2)(u-\theta_1, v-\theta_2)}{\|(u,v)-(\theta_1,\theta_2)\|}$

where $\|\ \|$ stands for the usual Euclidean norm. This function h is defined in an open neighbourhood of (θ_1, θ_2) excluding the point (θ_1, θ_2). Now define the double sequence of real random variables,

$$\varepsilon_{nm} = \begin{cases} h(U_n, V_m) & \text{if} \quad \| (U_n, V_m) - (\theta_1, \theta_2) \| \neq 0 \\ 0 & \text{if} \quad \| (U_n, V_m) - (\theta_1, \theta_2) \| = 0 \end{cases}$$

this double sequence is defined with probability tending to one when $n \longrightarrow \infty$, since (U_n, V_m) converges in probability to (θ_1, θ_2). Furthermore, since g is differentiable at (θ_1, θ_2), clearly ε_{nm} converges in probability to zero. On the other hand, since

$$\frac{\sqrt{nm}}{\sqrt{n+m}} \| (U_n, V_m) - (\theta_1, \theta_2) \| \leq \left(\|\sqrt{n}\,(U_n - \theta_1)\|^2 + \|\sqrt{m}\,(V_m - \theta_2)\|^2 \right)^{1/2}$$

and the last expression is a continuous function of $\sqrt{n}\,(U_n - \theta_1)$ and $\sqrt{m}\,(V_m - \theta_2)$ and, by lemma 1, has a limiting distribution, it follows that $\frac{\sqrt{nm}}{\sqrt{n+m}} \| (U_n, V_m) - (\theta_1, \theta_2) \|$ is bounded in probability, and thus $\frac{\sqrt{nm}}{\sqrt{n+m}} \|(U_n, V_m) - (\theta_1, \theta_2)\| \cdot \varepsilon_{nm}$, and its absolute value, converges in probability to zero. Let us now define

$$R_{nm} = \frac{\sqrt{nm}}{\sqrt{n+m}}\, Dg(\theta_1, \theta_2)(U_n - \theta_1, V_m - \theta_2) =$$

$$= \frac{\sqrt{m}}{\sqrt{n+m}} \left\{\sqrt{n}\, c_1(U_n - \theta_1)\right\} + \frac{\sqrt{n}}{\sqrt{n+m}} \left\{\sqrt{m}\, c_2(V_m - \theta_2)\right\}.$$

Since $\sqrt{n}\, c_1(U_n - \theta_1)$ and $\sqrt{m}\, c_2(V_m - \theta_2)$ converge in law to a univariate normal distribution with zero mean and $c\,\Sigma\,c' > 0$ variance, by lemma 2 it follows that the double sequence R_{nm} converges in law to a univariate normal distribution with zero mean and $c\,\Sigma\,c'$ variance. Therefore taking into account that the double sequence:

$$\left| \frac{\sqrt{nm}}{\sqrt{n+m}} \left(g(U_n, V_m) - g(\theta_1, \theta_2)\right) - R_{nm} \right| = \frac{\sqrt{nm}}{\sqrt{n+m}} \left\| (U_n, V_m) - (\theta_1, \theta_2) \right\| \, |\varepsilon_{nm}|$$

converges in probability to zero, it follows that $\frac{\sqrt{nm}}{\sqrt{n+m}} \left(g(U_n, V_n) - g(\theta_1, \theta_2)\right)$ has the same limiting distribution as R_{nm}, and the proof of section i) is concluded.

Section ii) follows inmediately since if $c_i = 0$, then $R_{nm} = 0$ and thus

$$\frac{\sqrt{nm}}{\sqrt{n+m}} \left(g(U_n, V_m) - g(\theta_1, \theta_2)\right) \quad \text{converges in probability to zero,}$$

and therefore also converges in law to zero.

Let us now consider section iii).

First we introduce the function $h(u, v) = \frac{g(u,v) - g(\theta_1,\theta_2) - \frac{1}{2} H(\theta_1,\theta_2)(u-\theta_1, v-\theta_2)}{\|(u,v)-(\theta_1,\theta_2)\|^2}$, where $H(\theta_1, \theta_2)(u - \theta_1, v - \theta_2)$ is the Hessian quadratic form. This function is defined in an open neighbourhood of (θ_1, θ_2) not containing the point (θ_1, θ_2). Now define the double sequence of random variables,

$$\varepsilon_{nm} = \begin{cases} h(U_n, V_m) & \text{if} \quad \|(U_n, V_m) - (\theta_1, \theta_2)\| \neq 0 \\ 0 & \text{if} \quad \|(U_n, V_m) - (\theta_1, \theta_2)\| = 0 \,. \end{cases}$$

This double sequence is defined with probability tending to one when $n \longrightarrow \infty$, since (U_n, V_m) converges in probability to (θ_1, θ_2). Furthermore, since $c = 0$ clearly ε_{nm} converges in probability to zero. Also, using similar considerations as in section i) $\frac{nm}{n+m} \|(U_n, V_m) - (\theta_1, \theta_2)\|^2$ is bounded in probability and thus, $\frac{nm}{n+m} \|(U_n, V_m) - (\theta_1, \theta_2)\|^2 \, \varepsilon_{nm}$, and their absolute value, converges in probability to zero. Let us now define

$$R_{nm} = \frac{1}{2} \frac{nm}{n+m} H(\theta_1, \theta_2)(U_n - \theta_1, V_m - \theta_2)$$

and,

$$T_{nm} = \frac{\sqrt{m}}{\sqrt{n+m}} \left\{\sqrt{n}\,(U_n - \theta_1)\right\} - \frac{\sqrt{n}}{\sqrt{n+m}} \left\{\sqrt{m}\,(V_m - \theta_2)\right\}$$

under our assumptions we have $R_{nm} = \frac{1}{2} T'_{nm}\, A\, T_{nm}$, where here, U_n, V_m, θ_1, and θ_2 are column vectors.

By lemma 2, it follows that T_{nm} converges in law to a k-variate normal distribution, with zero mean vector and Σ covariance matrix.

On the other hand, since A is a symmetric $k \times k$ matrix of rank r, $0 < r \leq k$, and Σ is a regular matrix, $\frac{1}{2}\Sigma^{1/2} A \Sigma^{1/2}$ has rank r and can be factorized into: $\frac{1}{2}\Sigma^{1/2} A \Sigma^{1/2} = P D P'$, $P P' = P' P = I$ and $D = diag(\lambda_1, ..., \lambda_r, 0, ..., 0)$ and therefore $Q_{nm} = P' \Sigma^{-1/2} T_{nm}$ converges in law to a random vector Z normally distributed with zero mean vector and identity covariance matrix, and $\frac{1}{2} T'_{nm} A T_{nm} = Q'_{nm} D Q_{nm}$ converges in law to the real random variable $Z' D Z$, and thus, since

$$\left| \frac{nm}{n+m} \left(g(U_n, V_m) - g(\theta_1, \theta_2)\right) - R_{nm} \right| = \frac{nm}{n+m} \| (U_n, V_m) - (\theta_1, \theta_2) \|^2 \, |\varepsilon_{nm}|$$

converges in probability to zero, then, the limiting distribution of $\frac{nm}{n+m} \left(g(U_n, V_m) - g(\theta_1, \theta_2)\right)$ is $Z' D Z = \lambda_1 Z_1^2 + ... + \lambda_r Z_r^2$, where each Z_i are normally distributed, with zero mean and unit variance, and $Z_1, ..., Z_r$ are stochastically independent. This concludes the proof of section iii).

We shall use the notation of pages 5 and 6 for lemmas 4, 5 and 6.

Lemma 4

Let (D, a_D) be a Hausdorff, connected and complete n-dimensional Riemannian manifold. Let p, q be two points of D, $p, q \in D$, such that, $p \neq q$, $q \in \mathcal{U}_p$. Then there exists a local chart (ϕ, V_q), $V_q \subset \mathcal{U}_p$, $p \notin V_q$ such that:

i) $\langle \partial_{\phi_i}, \partial_{\phi_j} \rangle_q = \delta_{ij}$ $i, j = 1, ..., n$, where δ_{ij} are the Kronecker deltas, and ∂_{ϕ_i}, $i = 1, ..., n$ is the basis vector field corresponding to the coordinate system defined through the local chart (ϕ, V_q), that is:

$$(\partial_{\phi_i})_m f = D_i (f \circ \phi^{-1})(\phi(m)) \qquad i = 1, ..., n$$

for any f, C^∞ function in a neighbourhood of m, $m \in V_q$.

ii) If we define the real function $g : W_q \longrightarrow \mathbf{R}$, such that:

$$g(x) = d(p, \phi^{-1}(x))$$

where $W_q = \phi(V_q) \subset \mathbf{R}^n$, is an open neighbourhood of $\phi(q)$, and d is the Riemannian distance on D, then g is differentiable at $\phi(q)$, and the Jacobian matrix is $g'(\phi(q)) = (1, 0, ..., 0)$.

Proof:

Section i) is trivial. Section ii) follows from the fact that it is possible to define a geodesic spherical coordinate system, as in pages 5 and 6 of section 2, in a neighbourhood

of q not containing p. It is enough to let ϕ be defined as $\phi = \psi \circ exp_p^{-1}$, with the same notation of pages 5 and 6, then, in a neighbourhood of $\phi(q)$, clearly we have $D_1 g(x) = 1$ and $D_\alpha g(x) = 0 \quad \alpha = 2, ..., n$. Then section ii) follows since g is differentiable with continuity in a neighbourhood of $\phi(q)$.

Lemma 5

Let (D, a_D) be a Hausdorff, connected and complete n-dimensional Riemannian manifold. Let p, q be two points of D, $p, q \in D$, such that $p \neq q$, $q \in \mathcal{U}_p$. Then, there exist two local charts (ϕ, V_q), (τ, V_p), where $V_q \subset \mathcal{U}_p \quad p \notin V_q$ and $V_p \subset \mathcal{U}_q \quad q \notin V_p$, such that:

i) $\langle \partial_{\phi_i}, \partial_{\phi_j} \rangle_q = \delta_{ij} \qquad \langle \partial_{\tau_i}, \partial_{\tau_j} \rangle_p = \delta_{ij} \qquad i, j = 1, ..., n$

where $(\partial_{\phi_i})_m f = D_i(f \circ \phi^{-1})(\phi(m)) \quad i = 1, ..., n$ for any f, C^∞ function in a neighbourhood of $m \in V_q$, and $(\partial_{\tau_i})_r g = D_i(g \circ \tau^{-1})(\tau(r)) \quad i = 1, ..., n$ for any g, C^∞ function in a neighbourhood of $r \in V_p$.

ii) If we define the real function $g : W_p \times W_q \longrightarrow \mathbf{R}$ such that: $g(x, y) = d(\tau^{-1}(x), \phi^{-1}(y))$, where $W_p = \tau(V_p)$, $W_q = \phi(V_q)$, are open sets of $\mathbf{R}^n$, such that $(\tau(p), \phi(q)) \in W_p \times W_q$, and d is the Riemannian distance on D, then g is differentiable at $(\tau(p), \phi(q))$, and the Jacobian matrix is $g'(\tau(p), \phi(q)) = (A, A)$ where A is a $1 \times n$ matrix, $A = (1, 0, 0, ..., 0)$.

Proof:

By observing that if $q \in \mathcal{U}_p$, then $p \in \mathcal{U}_q$, the proposition is a consequence of the previous lemma.

Lemma 6

Let (D, a_D) be a Hausdorff connected and complete Riemannian manifold. Let p be a point of D, $p \in D$. Consider $v, w \in D_p$ and let d be the Riemannian distance on D. Then:

$$\lim_{v,w \to 0} \frac{d(exp_p(v), exp_p(w))}{\|v - w\|_p} = 1 .$$

Proof:

Let (ϕ, V) be a local chart, satisfying $p \in V$ and

$$\phi = (x^1 \circ exp_p^{-1}, ..., x^n \circ exp_p^{-1})$$

where $x^1, ..., x^n$ are defined as: $x^i : D_p \longrightarrow \mathbf{R}$, $x^i(v) = \langle v, e_i \rangle_p \quad i = 1, ..., n$, where $e_1, ..., e_n$ is an orthonormal basis for D_p. The coordinate system defined is called a normal coordinate system.

Let $B(p, \delta)$ be defined as $B(p, \delta) = \{q \in D : d(p,q) < \delta\}$, $\delta \in \mathbb{R}^+$, and let us define $\bar{B}(p, \delta) = \{q \in D : d(p,q) \leq \delta\}$.

It is well known that there exists $\varepsilon \in \mathbb{R}^+$ such that $B(p, \delta)$ is strictly geodesically convex; that is given two points $q, r \in B(p,\delta)$, there exists one and only one geodesic between them in $B(p, \delta)$, for all δ such that $0 < \delta < \varepsilon$, and this is the minimal geodesic between q and r on D.

Let us now define $G(q) = \left(g_{ij}(q)\right)_{n\times n}$, the $n \times n$ matrix whose coefficients are the components of the metric tensor field in the normal coordinate system previously defined, that is $g_{ij}(q) = \langle \partial_{\phi_i}, \partial_{\phi_j} \rangle_q \quad i,j = 1,...,n$, where ∂_{ϕ_i} are defined as in the previous lemmas. Notice that $G(p) = I$. Now define

$$A(\delta) = \{\eta \in \mathbb{R} : \eta \quad \text{is an eigenvalue of} \quad G(q) \quad q \in \bar{B}(p, \delta)\}$$

and $\lambda(\delta) = inf A(\delta)$, $\mu(\delta) = sup\, A(\delta)$. Clearly $\lambda(\delta)$ and $\mu(\delta)$ are continuous functions on $\mathbb{R}^+$ such that $\lim_{\delta\to 0+} \mu(\delta) = \lim_{\delta\to 0+} \lambda(\delta) = 1$.

Therefore, for all $v, w \in D_p$ such that $exp_p(v)$, $exp_p(w) \in B(p, \delta)$, the minimal geodesic which joins both points, $exp_p(v)$ and $exp_p(w)$, say γ_{vw}, is in $B(p,\delta)$, and its length, $|\gamma_{vw}|$ is equal to the Riemannian distance between $exp_p(v)$ and $exp_p(w)$, and satisfies:

$$\sqrt{\lambda(\delta)}\, \|v - w\|_p \leq |\gamma_{vw}| \leq \sqrt{\mu(\delta)}\, \|v - w\|_p$$

and thus, since:

$$\sqrt{\lambda(\delta)} \leq \frac{d\big(exp_p(v),\ exp_p(w)\big)}{\|v - w\|_p} \leq \sqrt{\mu(\delta)}$$

and taking into account that $\lim_{\delta\to 0+} \mu(\delta) = \lim_{\delta\to 0+} \lambda(\delta) = 1$, we have:

$$\lim_{v,w\to 0} \frac{d\big(exp_p(v),\ exp_p(w)\big)}{\|v - w\|_p} = 1.$$

4. MAIN RESULTS

Theorem 1

Let $((D, a_D), \mu)$ be a regular n-parametric family of probability density functions. Let X be an independent sample of size k obtained from a population whose density is $p_o \in D$. Let $\tilde{p}_k$ be a consistent sequence of critical points of the log-likelihood

function. Then $k\,\rho^2(p_o, \tilde{p}_k) \xrightarrow{\mathcal{L}} Y$ where ρ is the Rao distance between the estimated point $\tilde{p}_k$ and the true point p_o, and Y is a random variable distributed as a chi-squared distribution with n degrees of freedom, χ^2_n.

Remark:

It is possible to use a coordinate-dependent approach:

Let (θ, V) be a local chart including the true density, $p_o \in V$, and let $\tilde{\theta}_k$ be a consistent sequence of roots of likelihood equations. Then, $k\rho^2\big(\theta^{-1}(\theta_o), \theta^{-1}(\tilde{\theta}_k)\big) \xrightarrow{\mathcal{L}} Y$, where $\theta_o = \theta(p_o)$, $\tilde{\theta}_k = \theta(\tilde{p}_k)$ and ρ and Y are as before.

Proof:

Consider a normal coordinate system in a suitable neighbourhood of p_o, $\theta^1, ..., \theta^n$. Let θ_o be the coordinates of p_o, $\theta_o = \theta(p_o)$. In a neighbourhood of p, the Rao distance between p_o and any point of the neighbourhood will be:

$$\rho(p, p_o) = \rho\big(\theta^{-1}(\theta), \theta^{-1}(\theta_o)\big) = \sqrt{\sum_{i=1}^{n} (\theta^i - \theta^i_o)^2}$$

and the Fisher information matrix at p_0 is the identity matrix. It is well known that if $\tilde{\theta}_k$ is a consistent sequence of roots of likelihood equations, then

$$\sqrt{k}\,(\tilde{\theta}_k - \theta_o) \xrightarrow{\mathcal{L}} Y \sim N_n\big(0, G^{-1}(\theta_o)\big) = N_n(0, I)$$

therefore

$$k\,\rho^2\big(\theta^{-1}(\tilde{\theta}_k), \theta^{-1}(\theta_o)\big) = k(\tilde{\theta}_k - \theta_o)'(\tilde{\theta}_k - \theta_o) \xrightarrow{\mathcal{L}} Y'Y$$

where $Y'Y$ is distributed as a chi-squared distribution with n degrees of freedom.

Finally, we may notice that $\rho\big(\theta^{-1}(\tilde{\theta}_k), \theta^{-1}(\theta_o)\big) = \rho(\tilde{p}_k, p_o)$ and the proof is concluded.

Theorem 2

Let $\big((D, a_D), \mu\big)$ be a regular n-parametric family of probability density functions. Let X_1 and X_2 be two independent samples of size k_1 and k_2 respectively obtained from the same statistical population whose density is $p_o \in D$. Let $\tilde{p}^{(1)}_{k_1}$ and $\tilde{p}^{(2)}_{k_2}$ be two consistent sequences of critical points of the log-likelihood functions of samples 1 and 2 respectively. Then $\frac{k_1 k_2}{k_1 + k_2}\,\rho^2(\tilde{p}^{(1)}_{k_1}, \tilde{p}^{(2)}_{k_2}) \xrightarrow{\mathcal{L}} Y$, where ρ is the Rao distance between $\tilde{p}^{(1)}_{k_1}$ and $\tilde{p}^{(2)}_{k_2}$, and Y is a random variable distributed as a chi-squared distribution with n degrees of freedom, χ^2_n.

Remark:

A coordinate-dependent approach would be:

Let (θ, V) be a local chart including the true density, $p_o \in V$, and let $\tilde{\theta}^{(1)}_{k_1}$ and $\tilde{\theta}^{(2)}_{k_2}$ be two consistent sequences of roots of likelihood equations computed from samples 1 and 2, respectively. Then

$$\frac{k_1\, k_2}{k_1 + k_2} \rho^2 \left(\theta^{-1}(\tilde{\theta}^{(1)}_{k_1}),\ \theta^{-1}(\tilde{\theta}^{(2)}_{k_2})\right) \xrightarrow{\mathcal{L}} Y\,,$$

where $\tilde{\theta}^{(1)}_{k_1} = \theta(\tilde{p}^{(1)}_{k_1})$, $\tilde{\theta}^{(2)}_{k_2} = \theta(\tilde{p}^{(2)}_{k_2})$, and ρ and Y as before.

Proof:

Consider a normal coordinate system of coordinates $\theta^1, ..., \theta^n$ in a suitable neighbourhood of p_o. Let ρ be the Rao distance, and let $\theta_o = \theta(p_o)$. If $\tilde{\theta}^{(1)}_{k_1}$ and $\tilde{\theta}^{(2)}_{k_2}$ are two consistent sequences of roots of likelihood equations computed from samples 1 and 2 respectively, then, taking into account that the Fisher information matrix at p_o is the identity matrix, we have:

$$\sqrt{k_1}\ (\tilde{\theta}^{(1)}_{k_1} - \theta_o) \xrightarrow{\mathcal{L}} X \sim N_n(0, I)\ ,\quad \sqrt{k_2}\ (\tilde{\theta}^{(2)}_{k_2} - \theta_o) \xrightarrow{\mathcal{L}} Y \sim N_n(0, I)$$

therefore, if we consider $g(\tilde{\theta}^{(1)}_{k_1}\,,\ \tilde{\theta}^{(2)}_{k_2}) = \|\tilde{\theta}^{(1)}_{k_1} - \tilde{\theta}^{(2)}_{k_2}\|^2_{p_o} = (\tilde{\theta}^{(1)}_{k_1} - \tilde{\theta}^{(2)}_{k_2})'(\tilde{\theta}^{(1)}_{k_1} - \tilde{\theta}^{(2)}_{k_2})$, since the vectors θ may be identified with vectors in D_{p_o}, the tangent space at p_o, through the exponential map, by applying lemma 3, we conclude that

$$\frac{k_1 k_2}{k_1 + k_2}\ g(\theta^{(1)}_{k_1}\,,\ \theta^{(2)}_{k_2}) \xrightarrow{\mathcal{L}} J$$

where J is a random variable distributed as a chi-squared distribution with n degrees of freedom, since the Jacobian of g at (θ_o, θ_o) is zero and the Hessian matrix is $H(\theta_o, \theta_o) = \begin{pmatrix} A & -A \\ -A & A \end{pmatrix}$ with $A = 2I$.

Now consider the Riemannian distance between $\theta^{-1}(\tilde{\theta}^{(1)}_{k_1})$ and $\theta^{-1}(\tilde{\theta}^{(2)}_{k_2})$. Taking into account the definition of a normal coordinate system, we have, by identifying $\tilde{\theta}^{(1)}_{k_1}$ and $\tilde{\theta}^{(1)}_{k_2}$, through the exponential map, by their corresponding elements of D_{p_o},

$$\rho\left(\theta^{-1}(\tilde{\theta}^{(1)}_{k_1}),\ \theta^{-1}(\tilde{\theta}^{(2)}_{k_2})\right) = \rho\big(exp_{p_o}(\tilde{\theta}^{(1)}_{k_1}),\ exp_{p_o}(\tilde{\theta}^{(2)}_{k_2})\big)\,.$$

Let us call $\tilde{D}_{12} = \|\tilde{\theta}^{(1)}_{k_1} - \tilde{\theta}^{(2)}_{k_2}\|_p$ and $\tilde{\rho}_{12} = \rho(\theta^{-1}(\tilde{\theta}^{(1)}_{k_1}),\ \theta^{-1}(\tilde{\theta}^{(2)}_{k_2}))$ therefore:

$$\frac{k_1\, k_2}{k_1 + k_2}\ |\tilde{\rho}^2_{12} - \tilde{D}^2_{12}| \ = \frac{k_1\, k_2}{k_1 + k_2}\ \tilde{D}^2_{12}\ \left|\frac{\tilde{\rho}^2_{12}}{\tilde{D}^2_{12}} - 1\right| \xrightarrow{\mathcal{P}} 0$$

since $\frac{k_1 k_2}{k_1+k_2}\,\tilde{D}_{12}^2$ has a limiting distribution and $\left|\frac{\tilde{\rho}_{12}^2}{\tilde{D}_{12}^2}-1\right| \xrightarrow{\mathcal{P}} 0$ by lemma 6 section 2. Then the statistic $\frac{k_1 k_2}{k_1+k_2}\,\tilde{\rho}_{12}^2$, has the same limiting distribution as $\frac{k_1 k_2}{k_1+k_2}\,\tilde{D}_{12}$, a chi-squared distribution with n degrees of freedom.

A coordinate-free approach is obtained by considering the sequences $\tilde{p}_{k_i}^{(i)} = \theta^{-1}(\tilde{\theta}_{k_i}^{(i)})$ $i = 1,2$. These are sequences of critical points of log-likelihood functions of samples 1 and 2, respectively, and $\rho(\tilde{p}_{k_1}^{(1)}, \tilde{p}_{k_2}^{(2)}) = \tilde{\rho}_{12}$.

Theorem 3

Let $((D, a_D), \mu)$ be a regular n parametric family of probability density functions. Assume that there exists a local chart (θ, V) such that, under the coordinate system defined by θ, the Rao distance is given by:

$$\rho^{-1}(\theta^{-1}(\beta), \theta^{-1}(\gamma)) = \sqrt{\sum_{i=1}^{n} (\beta^i - \gamma^i)^2} \quad \beta, \gamma \in \theta(V) \subset \mathbf{R}^n .$$

Let X_1 and X_2 be two independent samples of size k_1 and k_2 respectively, $p_1, p_2 \in V$, $\theta_1 = \theta(p_1)$, $\theta_2 = \theta(p_2)$, $p_1 \neq p_2$.
Let $\tilde{\theta}_{k_1}^{(1)}$ and $\tilde{\theta}_{k_2}^{(2)}$ be two consistent sequences of roots of likelihood equations computed from samples 1 and 2 respectively. Then,

$$\frac{\sqrt{k_1 k_2}}{\sqrt{k_1 + k_2}} \left(\rho(\theta^{-1}(\tilde{\theta}_{k_1}^{(1)}), \theta^{-1}(\tilde{\theta}_{k_2}^{(2)})) - \rho(p_1, p_2)\right) \xrightarrow{\mathcal{L}} Z \sim N_1(0,1).$$

Remark:

It is customary to say that $\frac{\sqrt{k_1 k_2}}{\sqrt{k_1+k_2}}\,\rho(\theta^{-1}(\tilde{\theta}_{k_1}^{(1)}), \theta^{-1}(\tilde{\theta}_{k_2}^{(2)}))$ converges asymptotically to a univariate normal distribution with $\frac{\sqrt{k_1 k_2}}{\sqrt{k_1+k_2}}\,\rho(p_1, p_2)$ mean and unit variance or it is also usual to say that the limiting distribution of $\frac{k_1 k_2}{k_1+k_2}\,\rho^2(\theta^{-1}(\tilde{\theta}_{k_1}^{(1)}), \theta^{-1}(\tilde{\theta}_{k_2}^{(2)}))$ is a non central chi-squared distribution with n degrees of freedom and a non-centrality parameter $\lambda = \frac{k_1 k_2}{k_1+k_2}\,\rho^2(p_1, p_2)$.

Proof:

Under the above assumptions, and defining $g(\tilde{\theta}_{k_1}^{(1)}, \tilde{\theta}_{k_2}^{(2)}) = \|\tilde{\theta}_{k_1}^{(1)} - \tilde{\theta}_{k_2}^{(2)}\| = \rho(\theta^{-1}(\tilde{\theta}_{k_1}^{(1)}), \theta^{-1}(\tilde{\theta}_{k_2}^{(2)}))$, and since:

$$\sqrt{k_1}\,(\tilde{\theta}_{k_1}^{(1)} - \theta_1) \xrightarrow{\mathcal{L}} X \sim N_n(0, I) \quad , \quad \sqrt{k_2}\,(\tilde{\theta}_{k_2}^{(1)} - \theta_2) \xrightarrow{\mathcal{L}} Y \sim N_n(0, I)$$

because the Fisher information matrix is equal to the identity matrix in all V, in particular in $\theta(p_1)$, $\theta(p_2)$. Then, since $g(\theta_1, \theta_2) \neq 0$, and the Jacobian of g at (θ_1, θ_2) is

$g'(\theta_1, \theta_2) = (A, -A)$, with $A \neq 0$, and $A\,I\,A' = (-A)\,I\,(-A)' = 1$, where A is a $1 \times n$ matrix, we may apply lemma 3, and we conclude that

$$\sqrt{\frac{k_1 k_2}{k_1 + k_2}}\left(\rho\big(\theta^{-1}(\tilde{\theta}^{(1)}_{k_1}),\ \theta^{-1}(\tilde{\theta}^{(2)}_{k_2})\big) - \rho(p_1, p_2)\right) \xrightarrow{\mathcal{L}} Z \sim N_1(0,1)$$

and the proof is completed.

Corollary 1

Assume that the parametric family has the following form:

$$p(x_1, ..., x_n \mid \theta^1, ..., \theta^n) = p_1(x_1 \mid \theta^1) \ldots p_n(x_n \mid \theta^n)$$

which means it is the joint density function of n independent random variables whose density is a member of a one-parameter family of probability densities, $p_i \in \big((D_i, a_{D_i}), \mu_i\big)$ $i = 1, ..., n$.

Assume also that each $\big((D_i, a_{D_i}), \mu_i\big)$ is a regular uniparametric family of probability density functions. Then, the hypothesis of Theorem 3 is satisfied and then, if $\tilde{\theta}^{(1)}_{k_1}$ and $\tilde{\theta}^{(2)}_{k_2}$ are consistent sequences of roots of likelihood equations from two independent samples X_1, X_2 of size k_1, k_2 of two populations whose densities are p_1, p_2, with $p_1 \neq p_2$, the sequence:

$$\sqrt{\frac{k_1 k_2}{k_1 + k_2}}\left(\rho\big(\theta^{-1}(\tilde{\theta}^{(1)}_{k_1})\big),\ \theta^{-1}(\tilde{\theta}^{(2)}_{k_2}) - \rho(p_1, p_2)\right) \xrightarrow{\mathcal{L}} Z \sim N_1(0,1)$$

where ρ is the Rao distance.

Proof:

Under the above hypothesis, the metric tensor field components are given by: $g_{\mu\nu} = \delta_{\mu\nu}\, h_\mu(\theta_\mu)$, where $h_\mu(\theta_\mu) = E\big((\partial_{\theta_\mu}\ log\, p_\mu)^2\big)$ $\mu = 1, ..., n$ and $\delta_{\mu\nu}$ are the Kronecker deltas.

The Christoffel symbols of the second kind, connection coefficients, are zero, except the coefficients of the form:

$$\Gamma^{\alpha}_{\alpha\,\alpha} = \partial_{\theta_\alpha}\ log\sqrt{h_\alpha} \quad \alpha = 1, ..., n, \quad \text{which only depends on } \theta_\alpha,$$

and therefore, the Riemann-Christoffel curvature tensor vanishes.

Then, since (D, a_D) is simply-connected, as the (D_i, a_{D_i}) are connected, the manifold is Euclidean and there exists a coordinate system θ, such that the distance between two points of coordinates α and β is given by: $\rho\left(\theta^{-1}(\alpha), \theta^{-1}(\beta)\right) = \sqrt{\sum_{i=1}^{n}(\alpha_i - \beta_i)^2}$ and then theorem 3 applies.

Corollary 2

Let $((D, a_D), \mu)$ be a regular n-parametric family of probability density functions, such that:

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

where F is a convenient function, Σ_o are fixed $n \times n$ strictly positive definite symmetric matrices, $\theta \in \mathbb{R}^n$ is an n-dimensional parameter vector, the sample space χ is $\mathbb{R}^n$ and the reference measure μ is the usual Lebesgue measure of $\mathbb{R}^n$. Then, the hypothesis of Theorem 3 is satisfied and therefore, if $\tilde{\theta}^{(1)}_{k_1}$ and $\tilde{\theta}^{(2)}_{k_2}$ are consistent sequences of roots of likelihood equations from two independent samples X_1, X_2 of size k_1, k_2 from two populations whose densities are p_1, p_2, with $p_1 \neq p_2$, the sequence $\sqrt{\frac{k_1 k_2}{k_1+k_2}} \left(\rho\left(\theta^{-1}(\tilde{\theta}^{(1)}_{k_1})\right), \, \theta^{-1}(\tilde{\theta}^{(2)}_{k_2}) - \rho(p_1, p_2)\right) \xrightarrow{\mathcal{L}} Z \sim N_1(0,1)$ where ρ is the Rao distance.

Proof:

It is easy to see that under the coordinate system given by the parameter θ, the metric tensor field is constant on all the manifold, and therefore, since this manifold is simply-connected, there exists a coordinate system defined by a global chart (β, D) such that the Riemannian distance is given by:

$$\rho\left(\beta^{-1}(\gamma), \, \beta^{-1}(\mu)\right) = \sqrt{\sum_{i=1}^{n} (\gamma_i - \mu_i)^2}$$

and thus, theorem 3 applies.

Theorem 4

Let $((D, a_D), \mu)$ be a regular n-parametric family of probability density functions. Assume that (D, a_D) is a complete manifold, with respect to the Rao distance. Let X_1 and X_2 be two independent samples of size k_1 and k_2 respectively, obtained from two statistical populations whose densities are p_1 and p_2 respectively, $p_1, p_2 \in D$, $p_1 \neq p_2$ and $p_2 \in \mathcal{U}_{p_1}$ (and thus $p_1 \in \mathcal{U}_{p_2}$) where the $\mathcal{U}_{p_i}$ are defined as in page 5.

Let $\tilde{p}^{(1)}_{k_1}$ and $\tilde{p}^{(2)}_{k_2}$ be two consistent sequences of critical points of the log-likelihood functions of samples 1 and 2 respectively. Then:

$$\sqrt{\frac{k_1 \, k_2}{k_1 + k_2}} \left(\rho(\tilde{p}^{(1)}_{k_1}, \, \tilde{p}^{(2)}_{k_2}) - \rho(p_1, p_2)\right) \xrightarrow{\mathcal{L}} Z \sim N_1(0,1)$$

where ρ is the Rao distance.

Remark 1

A coordinate-dependent approach would be:

Let (ϕ, V) and (τ, W) be two local charts such that $p_1 \in V$, $p_2 \in W$ and let $\tilde{\phi}_{k_1}^{(1)}$ and $\tilde{\tau}_{k_2}^{(2)}$ be two consistent sequences of roots of likelihood equations computed from samples 1 and 2 respectively. Then

$$\sqrt{\frac{k_1\, k_2}{k_1 + k_2}}\, \left(\rho\left(\phi^{-1}(\tilde{\phi}_{k_1}^{(1)}),\ \tau^{-1}(\tilde{\tau}_{k_2}^{(2)})\right) - \rho(p_1, p_2)\right) \xrightarrow{\mathcal{L}} Z \sim N_1(0,1).$$

Remark 2

It is customary to say that $\sqrt{\frac{k_1\, k_2}{k_1+k_2}}\, \rho(\tilde{p}_{k_1}^{(1)},\ \tilde{p}_{k_2}^{(2)})$ converges asymptotically to a univariate normal distribution with $\frac{\sqrt{k_1\, k_2}}{\sqrt{k_1+k_2}}\, \rho(p_1, p_2)$ mean and unit variance, or also, we usually say that the limiting distribution of $\frac{k_1\, k_2}{k_1+k_2}\, \rho^2(\tilde{p}_{k_1}^{(1)},\ \tilde{p}_{k_2}^{(2)})$ is a non-central chi-squared distribution with n degrees of freedom and a non-centrality parameter $\lambda = \frac{k_1\, k_2}{k_1+k_2}\, \rho^2(p_1, p_2)$.

Proof:

By lemma 5, there exist two local charts (ϕ, V_{p_2}), (τ, V_{p_1}), where $V_{p_2} \subset \mathcal{U}_{p_1}$, $p_1 \notin V_{p_2}$ and $V_{p_1} \subset \mathcal{U}_{p_2}$, $p_2 \notin V_{p_1}$, with the same notation as in page 5 and 12, such that:

$$\langle \partial_{\phi_i},\ \partial_{\phi_j} \rangle = \delta_{ij} \qquad \langle \partial_{\tau_i},\ \partial_{\tau_j} \rangle = \delta_{ij} \qquad i,j = 1, \ldots, n$$

and if $g : W_{p_1} \times W_{p_2} \longrightarrow \mathbb{R}$ is defined as $g(x, y) = \rho\left(\tau^{-1}(x),\ \phi^{-1}(y)\right)$ where $W_{p_i} = \tau(V_{p_i})$ $i = 1, 2$ are open sets of $\mathbb{R}^n$ and ρ is the Rao distance, then $g'\left(\tau(p_1),\ \phi(p_2)\right) = (A, A)$ where $A = (1, 0, \ldots 0)$, a $1 \times n$ matrix.

We also know that if $\tilde{\phi}_{k_2}^{(2)}$ and $\tilde{\tau}_{k_1}^{(1)}$ are two consistent roots of likelihood equations computed from samples X_1 and X_2 respectively, then, $\sqrt{k_1}\ (\tilde{\tau}_{k_1}^{(1)} - \tau(p_1)) \xrightarrow{\mathcal{L}} X \sim N_n(0, I)$ and $\sqrt{k_2}\ (\tilde{\phi}_{k_2}^{(2)} - \tau(p_2)) \xrightarrow{\mathcal{L}} Y \sim N_n(0, I)$, then, by lemma 3, since $g'\left(\tau(p_1),\ \phi(p_2)\right) = (A,\ A) \neq 0$, $A\, I\, A' = 1$, and therefore:

$$\sqrt{\frac{k_1\, k_2}{k_1 + k_2}}\, \left\{g(\tilde{\tau}_{k_1}^{(1)},\ \tilde{\phi}_{k_2}^{(2)}) - g\left(\tau(p_1),\ \phi(p_2)\right)\right\} \xrightarrow{\mathcal{L}} Z \sim N_1(0,1)$$

or equivalently,

$$\sqrt{\frac{k_1\, k_2}{k_1 + k_2}}\, \left\{\rho\left(\tau^{-1}(\tilde{\tau}_{k_1}^{(1)}),\ \phi^{-1}(\tilde{\phi}_{k_2}^{(2)})\right) - \rho(p_1, p_2)\right\} \xrightarrow{\mathcal{L}} Z \sim N_1(0,1).$$

A coordinate-free approach is obtained by considering the sequences of critical points of the log-likelihood function, $\tilde{p}_{k_1}^{(1)} = \tau^{-1}(\tilde{\tau}_{k_1}^{(1)})$ and $\tilde{p}_{k_2}^{(2)} = \phi^{-1}(\tilde{\phi}_{k_2}^{(2)})$ which obviously satisfy $\rho(\tilde{p}_{k_1}^{(1)}, \tilde{p}_{k_2}^{(2)}) = \rho(\tau^{-1}(\tilde{\tau}_{k_1}^{(1)}), \phi^{-1}(\tilde{\phi}_{k_2}^{(2)}))$ and thus $\sqrt{\frac{k_1 k_2}{k_1+k_2}}\, \rho(\tilde{p}_{k_1}^{(1)}, \tilde{p}_{k_2}^{(2)})$ has the same limiting distribution as before.

Corollary 3

Let $((D, a_D), \mu)$ be a regular n-parametric family of probability density functions. Assume that (D, a_D) is a simply connected complete Riemannian manifold, with the Riemannian manifold structure induced by Fisher information matrix. Additionally assume that all sectional curvatures are-non positive. Then if X_1 and X_2 are two independent samples of size k_1 and k_2 respectively, obtained from two statistical populations whose densities are p_1 and p_2, $p_1, p_2 \in D$, $p_1 \neq p_2$, and let $\tilde{p}_{k_1}^{(1)}$, $\tilde{p}_{k_2}^{(2)}$ be two consistent sequences of critical points of the log likelihood functions of samples X_1 and X_2 respectively. Then,

$$\sqrt{\frac{k_1 k_2}{k_1 + k_2}}\, (\rho(\tilde{p}_{k_1}^{(1)}, \tilde{p}_{k_2}^{(2)}) - \rho(p_1, p_2)) \xrightarrow{\mathcal{L}} Z \sim N_1(0,1)$$

where ρ is the Rao distance.

Proof:

Under these assumptions, given any point, p_1 or p_2, for instance, its cut locus is empty and then $p_1 \in \mathcal{U}_{p_2}$ or $p_2 \in \mathcal{U}_{p_1}$ is always true. Therefore Theorem 4 applies.

Corollary 4

Let $((D, a_D), \mu)$ be a regular n-parametric family of probability density functions. Assume that (D, a_D) is simply-connected complete Riemannian manifold, with the Riemannian manifold structure induced by the Fisher information matrix. Let X_1 and X_2 be two independent samples of size k_1 and k_2 respectively, obtained from two statistical populations, whose densities are p_1 and p_2, $p_1, p_2 \in D$, $p_1 \neq p_2$. Assume that p_1 or p_2 have no conjugate points along any geodesic. Let $\tilde{p}_{k_1}^{(1)}$ and $\tilde{p}_{k_2}^{(2)}$ be two consistent sequences of critical points of the log-likelihood functions of samples X_1 and X_2 respectively. Then,

$$\sqrt{\frac{k_1 k_2}{k_1 + k_2}}\, (\rho(\tilde{p}_{k_1}^{(1)}, \tilde{p}_{k_2}^{(2)}) - \rho(p_1, p_2)) \xrightarrow{\mathcal{L}} Z \sim N_1(0,1)$$

where ρ is the Rao distance.

Remark:

Notice that the result is obtained by replacing the conjugate points conditions on p_1 and p_2 with a stricter condition on D, e.g. (D, a_D) is a manifold without conjugate points (any point along any geodesic).

Proof:

Under the above assumptions for any point, p_1 or p_2, for instance, its cut locus is empty and then $p_1 \in \mathcal{U}_{p_2}$ or $p_2 \in \mathcal{U}_{p_1}$ is always true, and therefore, theorem 4 applies.

Remark

Several examples of n-parametric families of probability density functions may be considered and the problem of the asymptotic distribution of the Rao distance solved by applying Theorems 2, 4 or their corollaries. For instance, the univariate normal case or the multinomial negative case.

Remark:

Theorem 2 may be considered as a limiting case of Theorem 4. It is usual to say, the asymptotic distribution of the statistic: $\frac{k_1 k_2}{k_1+k_2} \rho^2 (\tilde{p}^{(1)}_{k_1}, \tilde{p}^{(2)}_{k_2})$ is a non-central chi-squared distribution with n-degrees of freedom and a non-centrality parameter $\lambda = \frac{k_1 k_2}{k_1+k_2} \rho^2(p_1, p_2)$. If $p_1 = p_2$ then $\lambda = 0$ and we obtain a central chi-squared distribution as in theorem 2, but theorem 2 is not redundant.

5. EXAMPLE

As an example we may consider the biparametric family of univariate normal distributions. This is a regular parametric family and, additionally, is simply-connected and together with the Riemannian metric induced by the Fisher information matrix is a complete Riemannian manifold with negative Gaussian curvature, see Burbea and Rao (1982). Consider the problem of comparing two univariate normal populations, given two samples from each one. Let $(\bar{x}_1, s_1)$ and $(\bar{x}_2, s_2)$ be the likelihood estimation of the parameters (μ_i, σ_i), $i = 1,2$, the usual coordinate system of the univariate normal case. In order to decide between the null hypothesis $H_o : \mu_1 = \mu_2 \quad \sigma_1 = \sigma_2$ against the alternative, $H_1 : \mu_1 \neq \mu_2$ or $\sigma_1 \neq \sigma_2$, we may define the critical region of this test

through the inequality $U = \frac{n_1 n_2}{n_1+n_2} \; D^2 > A_\varepsilon$ where D is defined as:

$$D = \sqrt{2}\; log\; \frac{1+\Delta}{1-\Delta} = 2\sqrt{2}\; tanh^{-1}(\Delta)$$

$$\Delta = \left\{ \frac{(\bar{x}_2 - \bar{x}_1)^2 + 2(s_1 - s_2)^2}{(\bar{x}_2 - \bar{x}_1)^2 + 2(s_1 + s_2)^2} \right\}^{1/2}$$

and n_1, n_2 are the sample sizes, and A_ε is determined in such a way to satisfy:

$$P(U > A_\varepsilon \mid H_o) = \varepsilon$$

that is ε is the significance level of the test. In order to find A_ε we may consider Theorem 2, then we know that U converges in law to a chi-squared distribution with 2 degrees of freedom, and if the sample size is not small we may assume that U is approximately distributed as a chi-square with 2 degrees of freedom.

In order to study the minimum sample size that is needed to use the asymptotic distribution as a good approximation, computer simulation was used to obtain samples, with different parameters and sample size, from univariate normal statistical populations. The maximum-likelihood estimation of the Rao distance between two samples was calculated together with the empirical distribution function (EDF) from U statistics, which according to the Glivenko-Cantelli theorem almost certainly uniformly converges to the true distribution function. Therefore, it is possible to compare the EDF with the asymptotic distribution function (ADF), the distribution function of a chi-squared distribution with 2 degrees of freedom.

Random numbers distributed normally were obtained using a modified Box-Müller method, Yakowitz (1977). The following tables of results were obtained by means of two sequences of 10^4 independent samples of size $n = 15$ or 30 from a univariate normal statistical population, $N(\mu, \sigma^2)$. For each of these 10^4 sample pairs, the maximum likelihood estimation of the parameters μ and σ^2 was calculated and for each pair of likelihood estimations, the U statistics with $n_1 = n_2 = n$ i.e. $U = \frac{n}{2} D^2$ were calculated.

TABLE I

$n = 15, \quad \mu = 0, \quad \sigma = 1$

U	*EDF*	*ADF*	*Differences*
0.0100	0.0059	0.0050	0.0009
0.0201	0.0113	0.0100	0.0013
0.0506	0.0259	0.0250	0.0009
0.1030	0.0468	0.0500	-0.0032
0.2110	0.0969	0.1000	-0.0031
0.5750	0.2380	0.2500	-0.0120
1.3900	0.4611	0.5000	-0.0389
2.7700	0.7081	0.7500	-0.0419
4.6100	0.8672	0.9000	-0.0328
5.9900	0.9239	0.9500	-0.0261
7.3800	0.9548	0.9750	-0.0202
9.2100	0.9770	0.9900	-0.0130
10.600	0.9861	0.9950	-0.0089

Maximum absolute value differences: -0.0419 (approx. 4%)

TABLE II

$n = 30, \quad \mu = 0, \quad \sigma = 1$

U	*EDF*	*ADF*	*Differences*
0.0100	0.0045	0.0050	-0.0005
0.0201	0.0105	0.0100	0.0005
0.0506	0.0261	0.0250	0.0011
0.1030	0.0486	0.0500	-0.0014
0.2110	0.0995	0.1000	-0.0005
0.5750	0.2482	0.2500	-0.0018
1.3900	0.4789	0.5000	-0.0211
2.7700	0.7235	0.7500	-0.0265
4.6100	0.8817	0.9000	-0.0183
5.9900	0.9373	0.9500	-0.0127
7.3800	0.9650	0.9750	-0.0100
9.2100	0.9851	0.9900	-0.0049
10.600	0.9926	0.9950	-0.0024

Maximum absolute value differences: -0.0265 (approx. 2.5%)

TABLE III

$n = 15, \quad \mu = 0, \quad \sigma = 0.1$

U	*EDF*	*ADF*	*Differences*
0.0100	0.0043	0.0050	-0.0007
0.0201	0.0097	0.0100	-0.0003
0.0506	0.0234	0.0250	-0.0016
0.1030	0.0454	0.0500	-0.0046
0.2110	0.0928	0.1000	-0.0072
0.5750	0.2268	0.2500	-0.0232
1.3900	0.4622	0.5000	-0.0378
2.7700	0.7009	0.7500	-0.0491
4.6100	0.8623	0.9000	-0.0377
5.9900	0.9229	0.9500	-0.0271
7.3800	0.9542	0.9750	-0.0208
9.2100	0.9780	0.9900	-0.0120
10.600	0.9867	0.9950	-0.0083

Maximum absolute value differences: -0.0491 (approx. 5%)

TABLE IV

$n = 30, \quad \mu = 0, \quad \sigma = 0.1$

U	*EDF*	*ADF*	*Differences*
0.0100	0.0045	0.0050	-0.0005
0.0201	0.0090	0.0100	-0.0010
0.0506	0.0235	0.0250	-0.0015
0.1030	0.0470	0.0500	-0.0030
0.2110	0.0954	0.1000	-0.0046
0.5750	0.2339	0.2500	-0.0161
1.3900	0.4815	0.5000	-0.0185
2.7700	0.7268	0.7500	-0.0232
4.6100	0.8808	0.9000	-0.0192
5.9900	0.9360	0.9500	-0.0140
7.3800	0.9630	0.9750	-0.0120
9.2100	0.9826	0.9900	-0.0074
10.600	0.9903	0.9950	-0.0047

Maximum absolute value differences: -0.0232 (approx. 2.5%)

TABLE V

$n = 15, \quad \mu = 0, \quad \sigma = 10$

U	*EDF*	*ADF*	*Differences*
0.0100	0.0044	0.0050	-0.0006
0.0201	0.0099	0.0100	-0.0001
0.0506	0.0250	0.0250	0.0000
0.1030	0.0491	0.0500	-0.0009
0.2110	0.0955	0.1000	-0.0005
0.5750	0.2317	0.2500	-0.0183
1.3900	0.4655	0.5000	-0.0345
2.7700	0.7090	0.7500	-0.0410
4.6100	0.8630	0.9000	-0.0370
5.9900	0.9226	0.9500	-0.0274
7.3800	0.9545	0.9750	-0.0205
9.2100	0.9776	0.9900	-0.0124
10.600	0.9861	0.9950	-0.0089

Maximum absolute value differences: -0.041 (approx. 4%)

TABLE VI

$n = 30, \quad \mu = 0, \quad \sigma = 10$

U	*EDF*	*ADF*	*Differences*
0.0100	0.0043	0.0050	-0.0007
0.0201	0.0086	0.0100	-0.0014
0.0506	0.0244	0.0250	-0.0006
0.1030	0.0492	0.0500	-0.0008
0.2110	0.0952	0.1000	-0.0048
0.5750	0.2423	0.2500	-0.0077
1.3900	0.4860	0.5000	-0.0140
2.7700	0.7347	0.7500	-0.0153
4.6100	0.8919	0.9000	-0.0081
5.9900	0.9418	0.9500	-0.0082
7.3800	0.9691	0.9750	-0.0059
9.2100	0.9846	0.9900	-0.0054
10.600	0.9916	0.9950	-0.0034

Maximum absolute value differences: -0.0153 (approx. 1.5%)

TABLE VII

$n = 15\,, \quad \mu = 1\,, \quad \sigma = 1$

U	*EDF*	*ADF*	*Differences*
0.0100	0.0046	0.0050	-0.0004
0.0201	0.0094	0.0100	-0.0006
0.0506	0.0215	0.0250	-0.0035
0.1030	0.0439	0.0500	-0.0061
0.2110	0.0887	0.1000	-0.0113
0.5750	0.2297	0.2500	-0.0203
1.3900	0.4722	0.5000	-0.0278
2.7700	0.7107	0.7500	-0.0393
4.6100	0.8685	0.9000	-0.0315
5.9900	0.9247	0.9500	-0.0253
7.3800	0.9560	0.9750	-0.0190
9.2100	0.9767	0.9900	-0.0133
10.600	0.9853	0.9950	-0.0097

Maximum absolute value differences: -0.0393 (approx. 4%)

TABLE VIII

$n = 30\,, \quad \mu = 1\,, \quad \sigma = 1$

U	*EDF*	*ADF*	*Differences*
0.0100	0.0053	0.0050	0.0003
0.0201	0.0095	0.0100	-0.0005
0.0506	0.0239	0.0250	-0.0011
0.1030	0.0474	0.0500	-0.0026
0.2110	0.0966	0.1000	-0.0034
0.5750	0.2407	0.2500	-0.0093
1.3900	0.4793	0.5000	-0.0207
2.7700	0.7254	0.7500	-0.0246
4.6100	0.8810	0.9000	-0.0190
5.9900	0.9358	0.9500	-0.0142
7.3800	0.9655	0.9750	-0.0095
9.2100	0.9850	0.9900	-0.0050
10.600	0.9909	0.9950	-0.0041

Maximum absolute value differences: -0.0246 (approx. 2.5%)

If the results shown in the previous tables are taken into account, it is reasonable to take the value 7.12 as a critical value for a sample size near 15 and a significance level of 5% and for a sample size near 30 with the same significance level, a value of 6.48. For an intermediate sample size, interpolation can be used and for a larger sample size, the asymptotic approximation. If a significance level of 1% and a sample size of 15 is considered, the critical value is approximately 11.54 and for a sample size of 30 with the same significance level, the critical value is 10.28. A more detailed study will appear on this subject in the near future.

The test power can be evaluated in a specific example. Consider two independent univariate normal statistical populations, $N(0,1)$ and $N(1,1)$ respectively. The Rao distance between them is $\rho_{12} = 0.9803$. 10^4 independent samples of size n were generated from each population and, for each pair of samples, the estimation (D_{12}) of the Rao distance (ρ_{12}) was calculated, thus obtaining the empirical distribution function (EDF) of $W = \sqrt{\frac{n_1 n_2}{n_1+n_2}}\,(D_{12} - \rho_{12})$ which was compared with the asymptotic distribution (ADF), a univariate normal $N(0,1)$. The results are given in tables IX and X for sample sizes $n_1 = n_2 = n = 15$ and $n_1 = n_2 = n = 30$ respectively.

TABLE IX

sample size $n = 15$ $\rho = 0.9803$

W	EDF	ADF	W	EDF	ADF
-2.685	0.000	0.004	0.352	0.500	0.637
-1.272	0.050	0.102	0.487	0.550	0.686
-0.929	0.100	0.176	0.618	0.600	0.731
-0.685	0.150	0.247	0.755	0.650	0.774
-0.497	0.200	0.309	0.906	0.700	0.817
-0.338	0.250	0.367	1.007	0.750	0.841
-0.187	0.300	0.425	1.262	0.800	0.896
-0.042	0.350	0.484	1.470	0.850	0.929
0.087	0.400	0.534	1.733	0.900	0.958
0.224	0.450	0.589	2.171	0.950	0.985

Maximum absolute value differences: 0.139

TABLE X

sample size $n = 30$ $\rho = 0.9803$

W	EDF	ADF	W	EDF	ADF
-3.797	0.000	$> 10^{-3}$	0.223	0.500	0.588
-1.380	0.050	0.084	0.355	0.550	0.638
-1.062	0.100	0.144	0.487	0.600	0.686
-0.826	0.150	0.205	0.626	0.650	0.734
-0.648	0.200	0.259	0.781	0.700	0.782
-0.470	0.250	0.320	0.944	0.750	0.827
-0.303	0.300	0.381	1.122	0.800	0.869
-0.156	0.350	0.438	1.323	0.850	0.907
-0.028	0.400	0.489	1.575	0.900	0.942
0.103	0.450	0.541	1.974	0.950	0.976

Maximum absolute value differences: 0.091

Let us consider now the case $n = 15$ and assume a significance level $\varepsilon = 0.05$. The critical value for U statistics is approximately 7.12 as we have said before, therefore, taking into account that:

$$U > 7.12 \iff W > -0.016$$

From table IX we conclude that the test power is approximately, by interpolation, 0.64.

If we consider the case $n = 30$, the critical value for U statistics is approximately 6.48, and since

$$U > 6.48 \iff W > -1.251$$

From table X we now conclude that the test power is approximately about 0.93.

REFERENCES

AMARI, S. (1968): Theory of information spaces. A geometrical foundation of the analysis of comunication systems. *RAAG Memoirs, 4,* 373-418.

AMARI, S. (1980): Theory of Information Space: a Differential-Geometrical Foundation of Statistics. *RAAG Reports, 106,* 1-53.

AMARI, S. (1982): Geometrical theory of asymptotic ancillarity and conditional inference. *Biometrika, 69,* 1-17.

AMARI, S. (1985): *Differential-Geometrical Methods in Statistics.* Springer-Verlag, New York.

ASH, R. (1972): *Real Analysis and Probability.* Academic Press, New York.

ATKINSON, C. and MITCHELL, A.F.S. (1981): Rao's distance measure. *Sankhyà,* 43 A, 345-365.

BHATTACHARYYA, A. (1946): On a measure of divergence between two multinomial populations. *Sankhyà, 7,* 401-406.

BILLINGSLEY, P. (1968): *Convergence of Probability Measures.* John Wiley, New York.

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.

BURBEA, J. (1986): Informative Geometry of Probability Spaces. *Expositiones Mathematicae, 4,* 347-378.

BURBEA, J. and OLLER, J.M. (1988): The information metric for univariate linear elliptic models. *Statistics and Decision, vol. 6, no. 3,* 209-222.

CAMPBELL, L.L. (1985): The relation between information theory and the differential geometry approach to statistics, *Inform. Sci. 35,* 199-210.

CAMPBELL, L.L. (1986): An extended Cencov characterization of the information metric, *Proc. Amer. Math. Soc. 98,* 135-141.

ČENČOV, N.N. (1965): Categories of mathematical statistics (in Russian), *Doklady Acad. Nauk, SSSR 164,* 3.

ČENČOV, N.N. (1972): Statistical Decision Rules and Optimal Conclusions (in Russian), *Nauka, Moskva,* translated in English (1982), AMS, Providence, Rhode Island.

CHAVEL, I. (1984): *Eigenvalues in Riemannian Geometry.* Academic Press, Inc. London.

CHEEGER, J. and EBIN, D.G. (1975): *Comparison Theorems in Riemannian Geometry.* North Holland, Oxford.

EFRON, B. (1975): Defining the Curvature of a Statistical Problem (With Application to Second Order Efficiency) (with discussion). *Annals of Statistics, 3,* 1189-1242.

FISHER, R.A. (1925): Theory of statistical estimation, *Proc. Camb. Phil. Soc. 22* (1925), 700-725.

GINÉ, E. and ARAUJO, A. (1980): *Central limit theorem for Real and Banach valued Random Variables.* John Wiley, New York.

HELGASON, S. (1978): *Differential Geometry, Lie Groups and Symmetric Spaces.* Academic Press, New York.

HICKS, N.J. (1965): *Notes on Differential Geometry.* Van Nostran Reinhold Comp., London.

LAHA, R.G. and ROHATGI, V.K. (1979): *Probability Theory.* John Wiley, New York.

LEGENDRE and LEGENDRE (1979): *Ecologie numérique. V. 2. La structure des données écologiques.* Masson.

MAHALANOBIS, P.C. (1930): On tests and measures of group divergence. *Jour. of the Assiatic Soc. of Bengal [N.S. XXVI],* 541-588.

MAHALANOBIS, P.C. (1936): On the Generalized Distance in Statistics. *Proc. Natl. Inst. Sci. India, 2 (1),* 49-55.

MATUSITA, K. (1957): Decision rule based on the distance for the classification problem. *Ann. Inst. Statist. Math. 8,* 67-77.

MATUSITA (1964): Distance and Decision Rules. *Ann. Inst. Statist. Math. 16,* 305-315.

MITCHELL, A.F.S. (1988): Statistical Manifolds of Univariate Elliptic Distributions. *International Statistical Review, 56, 1,* 1-16.

OLLER, J.M. and CUADRAS, C.M. (1985 b): Rao's distance for negative multinomial distributions. *Sankhyà, 47, A,* 75-83.

OLLER, J.M. (1987): Information metric for extreme value and logistic probability distributions. *Sankhyà, 49, A,* 17-23.

PEARSON, K. (1901): On lines and Planes of Closest Fit to Systems of Points in Space. *Phil. Mag., Ser. 6, 2(11),* 559-572.

PEARSON, K. (1926): On the coefficient of racial likeness. *Biometrika 18,* 105-117.

PREVOSTI, A., OCAÑA, J., and ALONSO, G. (1975): Distances between populations of *Drosophila subobscura.* Based on Chromosome Arrangement Frequencies. *Theor. and Appl. Genetics 45,* 231-241.

RAO, C.R. (1945): Information and accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc. 37,* 81-91.

RAO, C.R. (1948): The utilization of multiple measurements in problems of biological classification. *J. Roy. Statist. Soc. B, 10,* 159-193.

RAO, C.R. (1973): *Linear Statistical Inference and its Applications,* Wiley, New York.

SKOVGAARD, L.T. (1984): A Riemannian Geometry of the Multivariate Normal Model. *Scand. J. Statist, 11,* 211-223.

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

YAKOWITZ, S.J. (1977): *Computational Probability and Simulation.* Addison-Wesley Publ. Co. Inc.