Pergamon

Appl. Math. Lett. Vol. 8, No. 1, pp. 65–69, 1995

0893-9659/95 $9.50 + 0.00

0893-9659(94)00112-X

Statistical Tests Based on Geodesic Distances

M. L. MENÉNDEZ
Departamento de Matemática Aplicada, E.T.S. de Arquitectura
Universidad Politécnica de Madrid, 28040 Madrid, Spain

D. MORALES AND L. PARDO
Departamento de Estadística e I.O., Facultad de Matemáticas
Universidad Complutense de Madrid, 28040 Madrid, Spain

M. SALICRÚ
Departamento de Estadística, Avd. Diagonal 645
Universidad de Barcelona, 08028 Barcelona, Spain

(Received June 1994; accepted July 1994)

Abstract—Burbea and Rao [1] gave some general methods for constructing quadratic differential metrics on probability spaces. Using these methods, they obtain the Fisher information metric as a particular case. In this paper, a procedure to test statistical hypotheses is proposed on the basis of geodesic distances. An example is given and the asymptotic distribution of the test statistics are obtained; so, this method can be used in those cases where it is not possible to get the exact distribution of the test statistics.

Keywords—Information metric, Geodesic distance between probability distributions, Maximum likelihood estimators, Asymptotic distributions.

1. INTRODUCTION

Let $(\mathfrak{X}, \beta_{\mathfrak{X}}, P_\theta; \theta \in \Theta)$ be a statistical space, where Θ is an open subset of $\mathbb{R}^M$. We shall assume that there exists a generalized probability density function $f(x,\theta)$ for the probability P_θ with respect to a σ-finite measure μ, and we shall suppose that $f(x,\theta)$ is a smooth function verifying conditions of Cramer-Rao (C.R.) and the conditions for the consistency of a maximum likelihood estimator (M.L.E.) [2, pp. 194,223]. Taking into account that each population can be characterized by a particular point θ of Θ, we may interpret P_θ as a manifold and consider $\theta = (\theta_1, \dots, \theta_M)$ as a coordinate system. We consider a general positive quadratic differential form

$$ds^2(\theta) = \sum_{i,j=1}^{M} g_{ij}(\theta)\, d\theta_i\, d\theta_j, \tag{1}$$

and we suppose that it is invariant under transformation of θ. For example, if we consider

$$g_{ij}(\theta) = \int_{\mathfrak{X}} \frac{1}{f(x,\theta)} \frac{\partial f(x,\theta)}{\partial \theta_i} \frac{\partial f(x,\theta)}{\partial \theta_j}\, d\mu(x),$$

we have the Riemannian metric defined via Fisher information matrix by Rao [3]. Burbea and Rao [1] and Burbea [4] gave some general methods for constructing Riemannian metrics on probability spaces.

The research in this paper was supported in part by DGICYT Grants No. PB91–0387 and PB91–0155. Their financial support is gratefully acknowledged.

In the conditions given for the expression (1), we have a Riemannian space, with line element $ds = (ds^2)^{1/2}$ and fundamental tensor $g_{ij}(\theta)$. If we consider a curve in Θ joining θ_a and θ_b, i.e.,

$$\theta(t) = (\theta_1(t), \dots, \theta_M(t)), \qquad t_a \le t \le t_b,$$

with $\theta(t_a) = \theta_a$ and $\theta(t_b) = \theta_b$, then the distance between the probability density functions $f(x, \theta_a)$ and $f(x, \theta_b)$ along the curve $\theta(t)$ is given by

$$S(\theta_a, \theta_b) = \left| \int_{t_a}^{t_b} \left[\sum_{i=1}^{M} \sum_{j=1}^{M} g_{ij}(\theta) \frac{d\theta_i}{dt} \frac{d\theta_j}{dt} \right]^{1/2} dt \right|,$$

where, for ease of exposition, we have written θ, θ_i and θ_j instead of $\theta(t), \theta_i(t)$ and $\theta_j(t)$, respectively. In our case, this is usually a geodesic pseudo-distance (a pseudo-distance satisfies all the postulates of distance except that it may vanish for elements which are distant). In particular, the curve joining θ_a and θ_b, for which $S(\theta_a, \theta_b)$ is shortest, is of interest. Such a curve is called a geodesic and is given as the solution of the differential equations (the Euler-Lagrange equations)

$$\sum_{i=1}^{M} g_{ij}(\theta) \frac{d^2\theta_i}{dt^2} + \sum_{i,j=1}^{M} [i, j; k] \frac{d\theta_i}{dt} \frac{d\theta_j}{dt} = 0, \qquad j = 1, \dots, M,$$

where $[i, j; k]$ is the Christoffel symbol of the first kind, and it is defined by

$$[i, j; k] = \frac{1}{2} \left(\frac{\partial g_{ik}(\theta)}{\partial \theta_j} + \frac{\partial g_{ik}(\theta)}{\partial \theta_i} - \frac{\partial g_{ij}(\theta)}{\partial \theta_k} \right); \qquad i, j, k = 1, \dots, M.$$

The geodesic distance between θ_a and θ_b was proposed by Rao to measure the distance between distributions with parameters θ_a and θ_b. We can also observe that the geodesic distance between probability distributions are typically as follows:

$$S(\theta_1, \theta_2) = |h(\theta_1) - h(\theta_2)|. \tag{2}$$

In this paper, we propose a test based on the distance (2), and we obtain the asymptotic distribution of the distance $S(\widehat{\theta}_1, \widehat{\theta}_2)$, where $\widehat{\theta}$ is the M.L.E. of θ, for those cases where it is not possible to get the exact distribution of the test statistic.

2. ASYMPTOTIC DISTRIBUTION OF GEODESIC DISTANCES

When dealing with parametric distributions, statistical tests based on the geodesic distance can be constructed by substituting one or the two parameters by convenient estimators. To clarify this idea, let us consider the Pareto distribution (x_0 fixed) and the Rao distance $S(\theta_1, \theta_2)$. To test the hypothesis $H_0 : \theta = \theta_0$, is equivalent to test the hypothesis $S(\theta, \theta_0) = 0$, so we can use the statistic $T = S(\widehat{\theta}, \theta_0) = |\log \widehat{\theta} - \log \theta_0|$, where

$$\widehat{\theta} = n \left(\sum_{i=1}^{n} \log \frac{x_i}{x_0} \right)^{-1}$$

is the maximum likelihood estimator of θ. We reject the null hypothesis, at a level α, if $T_1 > c_\alpha$, where $P_{\theta_0}(T_1 > c_\alpha) = \alpha$. A straightforward calculus yields to the following decision rule

$$\phi(x_1, .., x_n) = \begin{cases} 1, & \text{if } c_1 < T_2(x_1, \dots x_n) < c_2, \\ 0, & \text{if } T_2(x_1, \dots x_n) < c_1 \text{ or } T_2(x_1, \dots x_n) > c_2, \end{cases}$$

where

(1) $F_{\chi^2_{2n}}(c_2) - F_{\chi^2_{2n}}(c_1) = 1 - \alpha$,

(2) $c_1 c_2 = 4n^2$,

and

$$T_2(x_1, \ldots, x_n) = 2\theta_0 \sum_{i=1}^{n} \log \frac{x_i}{x_0}.$$

Finally, note that Rao distance test coincides with the unbiased uniformly most powerful test if we change condition (2) by

(2′) $F_{\chi^2_{2(n+1)}}(c_2) - F_{\chi^2_{2(n+1)}}(c_1) = 1 - \alpha$.

In general, it will be not possible to get the exact distribution of the statistic $S(\widehat{\theta}_1, \widehat{\theta}_2)$; so, we will have to use its asymptotic distribution. To do this, we consider the function

$$S^*(\theta_1, \theta_2) = \phi\left(|h(\theta_1) - h(\theta_2)|^2\right),$$

ϕ being an increasing function with $\phi(0) = 0$. The null hypothesis $H_0 : \theta = \theta_0$ or $H_0 : \theta_1 = \theta_2$ will be rejected if $S^*(\widehat{\theta}, \theta_0)$ or $S^*(\widehat{\theta}_1, \widehat{\theta}_2)$ are greater than a critical value. If we write $f \in C^i(B)$ to denote that the real function f has continuous partial derivatives of i^{th} order on the set B, then we obtain the following result.

THEOREM 2.1. *Let $\widehat{\theta}_1$ and $\widehat{\theta}_2$ be M.L.E. of θ_1 and θ_2 based on independent random samples of sizes n and m, respectively, where $m/(n+m) \xrightarrow[n,m\to\infty]{} \lambda \in (0,1)$. Assume the regularity conditions of C.R. and for the consistency of an M.L.E. hold.*

(a) *If $\sigma^2 = \lambda T^t I_F(\theta_1)^{-1} T + (1-\lambda) S^t I_F(\theta_2)^{-1} S > 0$, $\phi \in C^1[0,\infty)$ and $h \in C^1(\Theta)$, then*

$$\left(\frac{nm}{n+m}\right)^{1/2} \left[S^*(\widehat{\theta}_1, \widehat{\theta}_2) - S(\theta_1, \theta_2)\right] \xrightarrow[n,m\to\infty]{L} N(0,\sigma)^2,$$

where $I_F(\theta)$ is the Fisher information matrix evaluated at θ, $T = (t_1, \ldots, t_M)^t$, $S = (s_1, \ldots, s_M)^t$,

$$\begin{aligned} a(\theta_1,\theta_2) &= 2(h(\theta_1) - h(\theta_2))\phi'((h(\theta_1) - h(\theta_2)^2), \\ t_i &= \frac{\partial S^*(\theta_1,\theta_2)}{\partial \theta_{1i}} = a(\theta_1,\theta_2)\frac{\partial h(\theta_1)}{\partial \theta_{1i}}, \\ s_i &= \frac{\partial S^*(\theta_1,\theta_2)}{\partial \theta_{2i}} = -a(\theta_1,\theta_2)\frac{\partial h(\theta_2)}{\partial \theta_{2i}}, \\ i &= 1, \ldots, M. \end{aligned}$$

(b) *If $\theta_1 = \theta_2$, $\phi'(0) > 0$, $\phi \in C^2[0,\infty)$ and $h \in C^2(\Theta)$, then*

$$\frac{nm}{n+m} \frac{S^*\left(\widehat{\theta}_1, \widehat{\theta}_2\right)}{\phi'(0)} \xrightarrow[n,m\to\infty]{L} \sum_{i=1}^{M} \beta_i \chi_1^2,$$

where the χ_1^2's are independent and the β_i's are the non-null eigenvalues of the matrix

$$A I_F(\theta_1)^{-1}, \quad A = (a_{ij})_{i,j=1,\ldots M} \quad \text{and} \quad a_{ij} = \frac{\partial h(\theta_1)}{\partial \theta_{1i}} \frac{\partial h(\theta_1)}{\partial \theta_{1j}}, \qquad i,j = 1, \ldots, M.$$

PROOF.

(a) By the mean value theorem

$$S^*\left(\widehat{\theta}_1,\widehat{\theta}_2\right) = S^*(\theta_1,\theta_2) + \sum_{i=1}^{M} \frac{\partial S^*(\gamma^*)}{\partial \theta_{1i}} \left(\widehat{\theta}_{1i} - \theta_{1i}\right) + \sum_{i=1}^{M} \frac{\partial S^*(\gamma^*)}{\partial \theta_{2i}} \left(\widehat{\theta}_{2i} - \theta_{2i}\right),$$

where $\gamma = (\theta_1^t, \theta_2^t)^t$, $\gamma^* = (\theta_1^{*t}, \theta_2^{*t})^t$, $\widehat{\gamma} = (\widehat{\theta}_1^t, \widehat{\theta}_2^t)^t$ and $\|\gamma^* - \gamma\|_2 < \|\widehat{\gamma} - \gamma\|_2$.

We conclude (c.f. [5, p. 385]) that

$$\left(\frac{nm}{n+m}\right)^{1/2} \left(S^*\left(\widehat{\theta}_1,\widehat{\theta}_2\right) - S^*(\theta_1,\theta_2)\right) =_a \left(\frac{nm}{n+m}\right)^{1/2} \left(T^t\left(\widehat{\theta}_1 - \theta_1\right) + S^t\left(\widehat{\theta}_2 - \theta_2\right)\right),$$

where $=_a$ means "asymptotically distributed as." Finally, applying the Central Limit Theorem, the result follows.

(b) If $\theta_1 = \theta_2$, then $\sigma^2 = S^*(\theta_1,\theta_2) = 0$. By the mean value theorem

$$S^*\left(\widehat{\theta}_1,\widehat{\theta}_2\right) = (\widehat{\gamma} - \gamma)^t \begin{pmatrix} C & -C \\ -C & C \end{pmatrix} (\widehat{\gamma} - \gamma) = \left(\widehat{\theta}_1 - \widehat{\theta}_2\right)^t C \left(\widehat{\theta}_1 - \widehat{\theta}_2\right),$$

where $C = (c_{ij})_{i,j=1,\dots,M}$, $c_{ij} = \frac{\partial S^*(\gamma^*)}{\partial \theta_{1i} \partial \theta_{1j}}$ and γ, γ^* and $\widehat{\gamma}$ are defined and verify the condition given above. We conclude (c.f. [5, p. 385]) that

$$\frac{nm}{n+m} \frac{S^*\left(\widehat{\theta}_1,\widehat{\theta}_2\right)}{\phi'(0)} =_a \left(\widehat{\theta}_1 - \widehat{\theta}_2\right)^t A \left(\widehat{\theta}_1 - \widehat{\theta}_2\right),$$

so the result follows.

COROLLARY 2.1. *If the assumptions of Theorem 2.1 hold and θ_2 is known, then*

(a) $n^{1/2}\left(S^*\left(\widehat{\theta}_1,\theta_2\right) - S^*(\theta_1,\theta_2)\right) \xrightarrow[n,m\to\infty]{L} N\left(0, T^t I_F(\theta_1)^{-1} T\right)$

(b) $n \dfrac{S^*\left(\widehat{\theta}_1,\widehat{\theta}_2\right)}{\phi'(0)} \xrightarrow[n,m\to\infty]{L} \displaystyle\sum_{i=1}^{M} \beta_i \chi_1^2.$

REMARK 2.1.

(a) The asymptotic distribution in Theorem 2.1(b) and Corollary 2.1(b) is proportional to a chi-square distribution when $\theta \in \mathbb{R}$.

(b) For testing the null hypothesis $H_0 : \theta_1 = \theta_2 (H_0 : \theta = \theta_0)$, the asymptotic power function can be obtained from Theorem 2.1(a) (Corollary 2.1(a)). The proposed test procedures are consistent in the sense of Fraser because the asymptotic power function tends to one as $n \to \infty$ in the alternative hypothesis.

(c) Probabilities $P(\sum_{i=1}^{M} \beta_i \chi_1^2 > t)$ can be calculated by computer simulation. Rao and Scott [6] suggest to consider the approximate distribution of $\sum_{i=1}^{M} \beta_i \chi_1^2$; which is given by $\overline{\beta} = \chi_M^2$, where $\overline{\beta} = \sum_{i=1}^{M} \beta_i / M$. In this case, we can easily calculate the value $\overline{\beta}$, since $\sum_{i=1}^{M} \beta_i = \operatorname{tr}(A I_F(\theta)^{-1})$.

REFERENCES

1. J. Burbea and C.R. Rao, Entropy differential metric, distance and divergence measures in probability spaces: A unified approach, *J. Multivariate Analysis* **12**, 575–596 (1982).
2. S. Zacks, *The Theory of Statistical Inference*, John Wiley, (1971).
3. C.R. Rao, Information and accuracy attainable in the estimation of statistical parameters, *Bull. Calcuta Math. Soc.* **37**, 81–91 (1945).
4. J. Burbea, Informative geometry of probability spaces, Technical Report No. 84-52, University of Pittsburgh, (1984).
5. C.R. Rao, *Linear Statistical Inference and its Applications*, John Wiley, (1973).
6. J.N.K. Rao and A.J. Scott, The analysis of categorical data from complex samplex surveys: Chi-squared tests for goodness of fit and independence in two-ways tables, *J. Amer. Stat. Assoc.* **76**, 221–230 (1981).