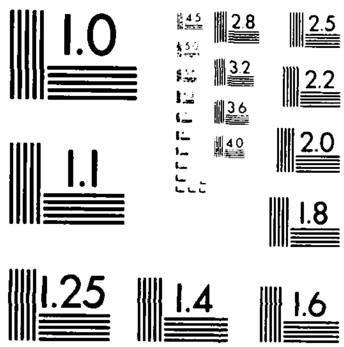
AD-A186 385 THE INFORMATION METRIC FOR UNIVARIATE LINEAR ELLIPTIC MODELS(U) PITTSBURGH UNIV PA CENTER FOR MULTIVARIATE ANALYSIS J BURBEA ET AL. JUN 87 TR-87-20 1/1

UNCLASSIFIED AFOSR-TR-87-0978 F49620-85-C-0008 F/G 12/3 NL

END
FILMED
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS
STANDARD REFERENCE MATERIAL 1010a
(ANSI and ISO TEST CHART No. 2)

Unclassified

DTIC FILE COPY (2)

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AD-A186 385

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFOSR-TR- 87-0978	2. GOVT ACCESSION NO.
	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): The information metric for univariate linear elliptic models	5. TYPE OF REPORT & PERIOD COVERED: ~~Technical~~ Interim - June 1987
	6. PERFORMING ORG. REPORT NUMBER: 87-20
7. AUTHOR(s): Jacob Burbea and Jose M. Oller	8. CONTRACT OR GRANT NUMBER(s): F49620-85-C-0008
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Center for Multivariate Analysis, Fifth Floor - Thackeray Hall, University of Pittsburgh, Pittsburgh, PA 15260	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61102F 2304 A5
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Office of Scientific Research NM, Department of the Air Force, Bolling Air Force Base, DC 20332 Bldg. 410	12. REPORT DATE: June 1987
	13. NUMBER OF PAGES: 18
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Same as 11	15. SECURITY CLASS. (of this report): Unclass
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

DTIC ELECTE OCT 0 2 1987 S D

18. SUPPLEMENTARY NOTES

19. *Key words and phrases:* elliptic distribution, Fisher information matrix, hyperbolic metric, information metric, Mahalanobis distance, Rao distance.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The information metric associated with a univariate linear elliptic family is shown to be, essentially, the Poincaré hyperbolic metric on a half-space whose geodesic Rao distance is an increasing hyperbolic function

20.

of a modified Mahalanobis distance. This result enables us to construct new statistical tests and to recover earlier results as special cases.

AD-A186 385

AFOSR-TR- 87-0978

Center for Multivariate Analysis

University of Pittsburgh

THE INFORMATION METRIC FOR UNIVARIATE LINEAR ELLIPTIC MODELS*

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

and

Jose M. Oller
Department of Statistics
University of Barcelona
08028 Barcelona, Spain

June 1987

Technical Report No. 87-20

Center for Multivariate Analysis
Fifth Floor Thackeray Hall
University of Pittsburgh
Pittsburgh, PA 15260

* This work is sponsored by the Air Force Office of Scientific Research under Contract F49620-85-C-0008. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

THE INFORMATION METRIC FOR UNIVARIATE LINEAR ELLIPTIC MODELS*

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

and

Jose M. Oller
Department of Statistics
University of Barcelona
08028 Barcelona, Spain

ABSTRACT

The information metric associated with a univariate linear elliptic family is shown to be, essentially, the Poincaré hyperbolic metric on a half-space whose geodesic Rao distance is an increasing hyperbolic function of a modified Mahalanobis distance. This result enables us to construct new statistical tests and to recover earlier results as special cases.

AMS 1980 subject classifications: Primary 62B10; Secondary 62H12.

Key words and phrases: elliptic distribution, Fisher information matrix, hyperbolic metric, information metric, Mahalanobis distance, Rao distance.

* This work is sponsored by the Air Force Office of Scientific Research under Contract F49620-85-C-0008. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

1. INTRODUCTION

The concepts of metrics and distances are fundamental in problems of statistical inference and in practical applications to study affinities among a given set of populations. A statistical model is specified by a family of probability distributions, described by a set of continuous parameters known as the parameter space. This model possesses some geometrical properties which are induced by the local information structures of the distributions. In particular, the Fisher information matrix of the given family of distributions gives rise to a Riemannian metric over the parameter space, whose geodesic distance, known as the Rao distance, plays a major role in multivariate statistical techniques. For the family of multivariate normal distributions with fixed shape but varying locations, this distance reduces to the well-known Mahalanobis distance. ~~We~~ refer to Burbea [1,2], and Burbea and Rao [3,4], and the references therein, for more details on these concepts and their derivations.

An interesting statistical model is provided by the family of elliptic distributions whose density functions have elliptical contours and which include the multivariate normal distributions as a subfamily. In this paper ~~we study~~ the information metric associated with an elliptic family whose shape varies linearly. It will be shown that this metric is essentially the Poincaré hyperbolic metric on a half-space, and that the resulting Rao distance is an increasing hyperbolic function of the generalized Mahalanobis distance. This will enable us to construct new statistical tests and to recover the recent results of Mitchell and Krzanowski [6] as a special case of our setting.

2. INFORMATIVE GEOMETRY OF ELLIPTIC DISTRIBUTIONS

We begin with a brief description of the general informative geometry that is induced by a parametric family $\mathcal{P}_\Theta = \{p(\cdot|\theta) : \theta \in \Theta\}$ of distributions for a random variable X, possibly vector-valued, with a sample space $\mathfrak{X}$. Here Θ is the parameter space, a manifold embedded in $\mathbf{R}^m$, with points $\theta \in \Theta$ coordinated by $\theta = [\theta_1, \ldots, \theta_m]^\dagger$, and satisfying the ordinary conditions of regular estimation. The elements $p(\cdot|\theta)$ of $\mathcal{P}_\Theta$ are probability distribution functions

$$p(x|\theta) = dP(x|\theta)/d\mu(x), \qquad (x \in \mathfrak{X},\ \theta \in \Theta)$$

where μ is a fixed positive σ-finite additive measure, defined on a σ-algebra of the subsets of $\mathfrak{X}$. In particular,

$$\int_{\mathfrak{X}} p(\cdot|\theta)d\mu = 1, \qquad (\theta \in \Theta).$$

It is also assumed that for a fixed $\theta \in \Theta$, the m functions

$$\ell_j(\cdot|\theta) = \partial \log p(\cdot|\theta)/\partial\theta_j, \qquad (j = 1,\ldots,m)$$

are linearly independent and are in $L^2\big(p(\cdot|\theta)d\mu\big)$. This, by the Cauchy-Schwarz inequality, implies that the elements

$$g_{jk}(\theta) = E_\theta\{\ell_j(\cdot|\theta)\ell_k(\cdot|\theta)\}, \qquad (j,k = 1,\ldots,m)$$

of the *information matrix* $G(\theta)$ are all finite, and that $G(\theta)$ is (strictly) positive-definite. It also implies that $\{\ell_j(\cdot|\theta)\}$, $j = 1,\ldots,m$, forms a basis for the tangent space T_θ at $\theta \in \Theta$, and, moreover that

$$ds^2(\theta) = d\theta^\dagger G(\theta)d\theta$$

is a Riemannian metric on Θ; called the *information metric* of the family $\mathcal{P}_\Theta$. This metric is invariant under the admissible transformations of the parameters as well as of the random variables, and the differential geometry associated with it is called *informative geometry*. The latter includes the evaluations of curvatures , geodesic curves and geodesic distances. The geodesic distance $S(\theta^{(1)}, \theta^{(2)})$ between the points $\theta^{(1)}$ and $\theta^{(2)}$ of Θ is known as the *Rao distance* between $p(\cdot|\theta^{(1)})$ and $p(\cdot|\theta^{(2)})$ of P_Θ. For a more detailed account, we refer to Burbea [1] (see also Burbea [2], Burbea and Rao [3,4], Oller [8], and Oller and Cuadras [9]). We also note, in passing, that in matrix-notation, $G(\theta)$ may also be expressed as

$$G(\theta) = E_\theta\left\{\frac{\partial}{\partial\theta} \log p(\cdot|\theta) \frac{\partial}{\partial\theta^\dagger} \log p(\cdot|\theta)\right\},$$

or as

$$G(\theta) = E_\theta\{\ell(\cdot|\theta)\ell(\cdot|\theta)^\dagger\}$$

where $\ell(\cdot|\theta) = [\ell_1(\cdot|\theta), \ldots, \ell_m(\cdot|\theta)]^\dagger$.

An n-dimensional random variable X is said to have an *elliptic distribution* with parameters $\mu = [\mu_1, \ldots, \mu_n]^\dagger$ and Σ, an $n\times n$ (strictly) positive-definite matrix, if its density is of the form

$$p(x|\mu, \Sigma) = \frac{\Gamma(n/2)}{\pi^{n/2}} \frac{1}{|\Sigma|^{1/2}} F\{(x-\mu)^\dagger \Sigma^{-1}(x-\mu)\} \tag{2.1}$$

where F is a nonnegative function on $\mathbb{R}_+ = (0,\infty)$ satisfying

$$\int_0^\infty r^{n/2-1} F(r)dr = 1. \tag{2.2}$$

In this case the sample space $\mathfrak{X}$ is $\mathbb{R}^n$ with $d\mu = dv$, the usual volume Lebesgue measure of $\mathbb{R}^n$. The parameter space Θ is now the $n(n+3)/2$-dimensional manifold $\mathbb{R}^n \times P(n, \mathbb{R})$, where $P(n,\mathbb{R})$ is the set of all $n\times n$ positive-definite matrices over $\mathbb{R}$.

The vector μ and the matrix Σ for the point (μ,Σ) in Θ may be expressed in terms of $E(X)$ and $Cov(X)$, provided the latter exist. In fact, the characteristic function $\phi_F(t) = E(e^{it\cdot X})$ of the above $p(\cdot|\mu, \Sigma)$ may be expressed as

$$\phi_F(t) = e^{it\cdot\mu}\Lambda_F(t^\dagger \Sigma t) \tag{2.3}$$

where

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

with

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

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

$$E(X) = i\,\frac{\partial}{\partial t}\,\phi_F(t)\Big|_{t=0}$$

and

$$E(XX^\dagger) = -\frac{\partial^2}{\partial t\partial t^\dagger}\,\phi_F(t)\Big|_{t=0}.$$

This gives $E(X) = \mu$ and $E(XX^\dagger) = \mu\mu^\dagger + c_F\Sigma$, where

$$c_F = -2\Lambda_F'(0) = \frac{1}{n}\int_0^\infty r^{n/2}F(r)dr \tag{2.4}$$

and hence $\mathrm{Cov}(X) = c_F\Sigma$. In particular, $E(X)$ exists if and only if $\int_0^\infty r^{n/2-1}F(r)dr < \infty$, and $\mathrm{Cov}(X)$ exists if and only if $\int_0^\infty r^{n/2}F(r)dr < \infty$, in which case $0 < c_F < \infty$. A normal distribution $N_n(\cdot|\mu, \Sigma)$ is an example of an elliptic distribution with

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

Other basic properties of elliptic distributions have been obtained by Kelker [5] and are summarized in Muirhead [7, pp.32-40].

We now turn to the information matrix $G_F(\mu,\Sigma)$ of the elliptic distribution (2.1). In this paper, however, we confine our attention to a submanifold of $\mathbf{P}(n, \mathbf{R})$ consisting of the cone $C(\Sigma_0) = \{\sigma^2\Sigma_0 : \sigma > 0\}$ where Σ_0 is a fixed element of $\mathbf{P}(n, \mathbf{R})$. The resulting parameter space is now $\mathbf{R}^n \times C(\Sigma_0)$ which is an $(n+1)$-dimensional submanifold of the full $n(n+3)/2$-dimensional manifold $\mathbf{R}^n \times \mathbf{P}(n, \mathbf{R})$. Note, however, that the former is not a geodesic submanifold, with respect to the information metric $d\theta^{\dagger} G_F(\theta) d\theta$, $\theta = (\mu,\Sigma)$, of the latter. A slight generalization is obtained by replacing μ in (2.1) with $\mu = A\beta$ where $\beta = [\beta_1, \ldots, \beta_m]^{\dagger}$ is a vector in $\mathbf{R}^m$ and A is a fixed $n\times m$ matrix of rank $m \leq n$. In particular, $A^{\dagger}A$ is a nonsingular $m\times m$ matrix, and the density in (2.1) is of the form

$$p(x|\beta, \sigma) = \frac{\Gamma(n/2)}{\pi^{n/2}} |\Sigma_0|^{-1/2} \sigma^{-n} F(\sigma^{-2}(x - A\beta)^{\dagger} \Sigma_0^{-1}(\cdot - A\beta)) \tag{2.5}$$

where F is a function from $\mathbf{R}_+$ into $\mathbf{R}_+$, satisfying (2.2). In this case, x is in the sample space $\mathbf{R}^n$ and (β,σ) is in the parameter space $\mathbf{R}_+^{m+1} = \mathbf{R}^m \times \mathbf{R}_+$ which is a half-space in $\mathbf{R}^{m+1}$.

In the setting of $m = n$, $A = I$ (the identity matrix of $\mathbf{R}^n$) and $\sigma \equiv 1$, the informative geometry of the distribution in (2.5), with β in the parameter space $\mathbf{R}^n$, was studied by Mitchell and Krzanowski [6]. The analysis in this paper will enable us to recover the results in the setting of [6] as a special case of our more general setting.

To find the information matrix $G_{(\beta,\sigma)} = G^{(n)}_{F:A,\Sigma_0}(\cdot,\cdot)$ of $p(\cdot|\beta, \sigma)$ in (2.5), we shall assume, in addition to (2.2), that F is also in $C^1(\mathbf{R}_+)$ with

$$\int_0^{\infty} r^{n/2} F(r) \{(\mathcal{L}F)(r)\}^2 dr < \infty \tag{2.6}$$

and

$$\int_0^\infty r^{n/2+1}F(r)\{(\mathcal{L}F)(r)\}^2 dr < \infty, \tag{2.7}$$

where $\mathcal{L}F = F'/F$ is the logarithmic derivative of F. Then, for $p = p(\cdot|\beta, \sigma)$, $G = G_{(\beta,\sigma)}$ and $E = E_{(\beta,\sigma)}$, we have

$$G = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} \tag{2.8}$$

where

$$G_{11} = E\{\frac{\partial}{\partial\beta} \log p \frac{\partial}{\partial\beta^\dagger} \log p\},$$

$$G_{12} = G_{21}^\dagger = E\{\frac{\partial}{\partial\beta} \log p \frac{\partial}{\partial\sigma} \log p\}$$

and

$$G_{22} = E\{\frac{\partial}{\partial\sigma} \log p \frac{\partial}{\partial\sigma} \log p\}.$$

Later, in Section 4, we shall show that conditions (2.2) and (2.6)-(2.7) guarantee the finiteness of the matrices G_{jk} ($1 \leq j, k \leq 2$) and the (strict) positive-definiteness of the information matrix G. Moreover, we shall also show that, in fact

$$G_{11} = a\sigma^{-2}A^\dagger\Sigma_0^{-1}A, \quad G_{12} = G_{21}^\dagger = 0, \quad G_{22} = b\sigma^{-2} \tag{2.9}$$

where

$$a = \frac{4}{n}\int_0^\infty r^{n/2}F(r)\{(\mathcal{L}F)(r)\}^2 dr$$

and

$$b = \int_0^\infty r^{n/2-1}\{n + 2r(\mathcal{L}F)(r)\}^2 F(r)dr.$$

In particular, $0 < a, b < \infty$.

To find the information metric ds^2 of $p(\cdot|\beta, \sigma)$, we consider the orthogonal diagonalization

$$VDV^\dagger = A^\dagger\Sigma_0^{-1}A$$

of the $m\times m$ positive-definite matrix $A^\dagger\Sigma_0^{-1}A = (\Sigma_0^{-1/2}A)^\dagger(\Sigma_0^{-1/2}A)$. Here V is a $m\times m$ orthogonal matrix and $D = \mathrm{diag}[\lambda_1, \ldots, \lambda_m]$ with $\lambda_j > 0$ $(j = 1,\ldots,m)$. Then the linear change of parameters

$$\tilde{\theta} = [\theta_1, \ldots, \theta_m]^\dagger = \sqrt{ab^{-1}}\, D^{1/2}V^\dagger\beta, \qquad \theta_{m+1} = \sigma$$

constitutes a diffeomorphism $(\beta,\sigma)\longmapsto\theta = (\tilde{\theta}, \theta_{m+1})$ of the parameter space $\mathbb{R}_+^{m+1}$ onto itself with the Jacobian $(ab^{-1})^{m/2}\,(\lambda_1 \cdots \lambda_m)^{1/2} > 0$. The Jacobian-matrix of the inverse of this transformation is

$$J = \begin{bmatrix} \sqrt{ba^{-1}}\, VD^{-1/2} & 0 \\ 0 & 1 \end{bmatrix},$$

and hence the information matrix $\hat{G}(\theta)$ in the new coordinates $\theta = [\theta_1, \ldots, \theta_m, \theta_{m+1}]^\dagger$ is

$$\hat{G}(\theta) = J^\dagger GJ = b\theta_{m+1}^{-2}I_{m+1},$$

where I_{m+1} is the identity matrix of $\mathbb{R}^{m+1}$. The information metric is

$$ds^2(\theta) = d\theta^\dagger\hat{G}(\theta)d\theta = \frac{b}{\theta_{m+1}^2}\sum_{j=1}^{m+1}(d\theta_j)^2, \tag{2.10}$$

which is, effectively, the *Poincaré hyperbolic metric* of the upper half-space $\mathbb{R}_+^{m+1} = \{[\theta_1, \ldots, \theta_{m+1}]^\dagger \in \mathbb{R}^{m+1}: \theta_{m+1} > 0\}$ (see, for example, Wolf [10]). It follows that the manifold of the family of distributions $p(\cdot|\theta)$ in

(2.5), $\theta \in \mathbb{R}_+^{m+1}$, is *isotropic* with a constant negative Riemannian curvature

$$\kappa = -1/b.$$

In particular, for any two points on this hyperbolic manifold, there exists one and only one geodesic line joining the two points.

The equations of the geodesics of the above information metric, in terms of its arc-length parameters, are found to be

$$\theta_k = \sqrt{b}\, C^{-2} B_k \tanh\left(\frac{s}{\sqrt{b}} + \epsilon\right) + D_k, \qquad (k = 1,\ldots,m),$$

$$\theta_{m+1} = C^{-1} \operatorname{sech}\left(\frac{s}{\sqrt{b}} + \epsilon\right),$$

where ϵ, B_k and D_k $(k = 1,\ldots,m)$ are real constants of integration, and

$$C = \left\{ b \sum_{k=1}^{m} B_k^2 \right\}^{1/2},$$

or $C = \infty$, in which case $\theta_k = D_k$ $(k = 1,\ldots,m)$ and $\theta_{m+1} = 0$. Note that since

$$\sum_{k=1}^{m} (\theta_k - D_k)^2 + \theta_{m+1}^2 = C^{-2},$$

the above geodesics are semi-circles of the upper half-space $\mathbb{R}_+^{m+1}$, with center $(D_1, \ldots, D_m, 0)$ and radius C^{-1}, and are orthogonal to the hypersurfaces $\theta_{m+1} = \xi$, $(\xi \geq 0)$.

The geodesic distance or the Rao distance ρ_{12} between two points $\theta^{(1)} = (\bar{\theta}^{(1)}, \theta_{m+1}^{(1)})$ and $\theta^{(2)} = (\bar{\theta}^{(2)}, \theta_{m+1}^{(2)})$ of $\mathbb{R}_+^{m+1}$ is then

$$\rho_{12} = \sqrt{b} \log \frac{1 + \Delta_{12}}{1 - \Delta_{12}} = 2\sqrt{b} \tanh^{-1}(\Delta_{12}) \qquad (2.11)$$

where

$$\Delta_{12} = \left\{ \frac{\|\bar{\theta}^{(1)} - \bar{\theta}^{(2)}\|^2 + b(\theta_{m+1}^{(1)} - \theta_{m+1}^{(2)})^2}{\|\bar{\theta}^{(1)} - \bar{\theta}^{(2)}\|^2 + b(\theta_{m+1}^{(1)} + \theta_{m+1}^{(2)})^2} \right\}^{1/2} .$$

Thus, using the old coordinates (β,σ) of $\mathbb{R}^{m+1}$, this Rao distance ρ_{12} between $p(\cdot|\beta_{(1)}, \sigma_1)$ and $p(\cdot|\beta_{(2)}, \sigma_2)$ admits the same form with

$$\Delta_{12} = \left\{ \frac{a(\beta_{(1)} - \beta_{(2)})^{\dagger} A^{\dagger} \Sigma_0^{-1} A(\beta_{(1)} - \beta_{(2)}) + b(\sigma_1 - \sigma_2)^2}{a(\beta_{(1)} - \beta_{(2)})^{\dagger} A^{\dagger} \Sigma_0^{-1} A(\beta_{(1)} - \beta_{(2)}) + b(\sigma_1 + \sigma_2)^2} \right\}^{1/2} . \qquad (2.12)$$

3. MAHALANOBIS DISTANCE

If σ in the distributions (2.5) is fixed, say $\sigma = \sigma_0 > 0$, then the parameter space of the distributions is restricted to $\mathbb{R}^m$. In this case the information metric in (2.10) reduces to the euclidean metric on $\mathbb{R}^m$

$$ds^2(\bar{\theta}) = b\sigma_0^{-2}\sum_{j=1}^{m}(d\theta_j)^2,$$

and thus the resulting Rao distance $\bar{\rho}_{12}$ between $p(\cdot|\beta_{(1)}, \sigma_0)$ and $p(\cdot|\beta_{(2)}, \sigma_0)$, in the manifold $\mathbb{R}^m\times\{\sigma_0\}$, is

$$\bar{\rho}_{12} = \sigma_0^{-1}\sqrt{a}\left\{(\beta_{(1)} - \beta_{(2)})^{\dagger}A^{\dagger}\Sigma_0^{-1}A(\beta_{(1)} - \beta_{(2)})\right\}^{1/2}. \qquad (3.1)$$

Since, however, $\mathbb{R}^m\times\{\sigma_0\}$ is clearly not a geodesic submanifold with respect to the nonreduced metric $ds^2(\theta)$, of $\mathbb{R}_+^{m+1} = \mathbb{R}^m\times\mathbb{R}_+$, $\bar{\rho}_{12}$ must exceed the nonreduced Rao distance ρ_{12} between $p(\cdot|\beta_{(1)}, \sigma_0)$ and $p(\cdot|\beta_{(2)}, \sigma_0)$.

In the general case that σ is not fixed, we introduce a modification of $\bar{\rho}_{12}$ in the form

$$d_{12} = (\sigma_1\sigma_2)^{-1/2}\sqrt{a}\left\{(\beta_{(1)} - \beta_{(2)})^{\dagger}A^{\dagger}\Sigma_0^{-1}A(\beta_{(1)} - \beta_{(2)})\right\}^{1/2}, \qquad (3.2)$$

which we call the *Mahalanobis generalized-distance* between $p(\cdot|\beta_{(1)}, \sigma_1)$ and $p(\cdot|\beta_{(2)}, \sigma_2)$. This quantity reduces to $\bar{\rho}_{12}$ when $\sigma_1 = \sigma_2 = \sigma_0$ and is directly related to the *classical Mahalanobis distance* M_{12} between $p(\cdot|\beta_{(1)}, \sigma_0)$ and $p(\cdot|\beta_{(2)}, \sigma_0)$, provided that Cov(X) of the distribution $p(\cdot|\beta, \sigma_0)$ in (2.5) exists. That is, besides (2.2) and (2.6)-(2.7), we must also assume that the quantity c_F, defined in (2.4), satisfies $0 > c_F < \infty$. In this case, $\mathrm{Cov}(X) = c_F\sigma_0^2\Sigma_0$ and $\bar{\rho}_{12} = \sqrt{a}\,c_F M_{12}$. The relationship between the Mahalanobis generalized-distance d_{12} and the Rao distance ρ_{12} can be read off from (2.11)-(2.12) and (3.2). This, after some algebraic manipulations, gives

$$\rho_{12} = 2\sqrt{b}\,\sinh^{-1}\frac{1}{2\sqrt{b}}\left\{d_{12}^2 + b(\sigma_1 - \sigma_2)^2/\sigma_1\sigma_2\right\}^{1/2}.$$

In particular, ρ_{12} is an increasing function of d_{12}, and

$$\rho_{12} \leq \{d_{12}^2 + b(\sigma_1 - \sigma_2)^2/\sigma_1\sigma_2\}^{1/2}.$$

We therefore conclude that the statistical tests based on either ρ_{12} or on d_{12} are completely equivalent when $\sqrt{\sigma_1/\sigma_2} - \sqrt{\sigma_2/\sigma_1}$ = const. In particular, this is so when σ_1 and σ_2 are fixed. Moreover, when $\sigma_1 = \sigma_2 = \sigma_0$, d_{12} reduces to $\bar{\rho}_{12}$ and we obtain the symmetric relationships

$$\rho_{12} = 2\sqrt{b}\,\sinh^{-1}(\bar{\rho}_{12}/2\sqrt{b})$$

and

$$\bar{\rho}_{12} = 2\sqrt{b}\,\sinh(\rho_{12}/2\sqrt{b}), \qquad (\sigma_1 = \sigma_2 = \sigma_0).$$

Especially, ρ_{12} is an increasing function of $\bar{\rho}_{12}$ and, of course, $\rho_{12} \leq \bar{\rho}_{12}$. Moreover, $\bar{\rho}_{12} \simeq \rho_{12}$ when $\rho_{12} << 2\sqrt{b}$.

When Cov(X) of $p(\cdot|\beta, \sigma_0)$ exists, the reduced Rao distance $\bar{\rho}_{12}$ in (3.1) was also discussed in Mitchell and Krzanowski [6] in the special setting of $m = n$, $A = I$ and $\sigma_0 = 1$. The discussion in [6], however, does not contain the above relationships between $\bar{\rho}_{12}$ and the fuller Rao distance ρ_{12}.

4. EVALUATIONS OF INTEGRALS

This section is devoted to the evaluations of the integrals appearing in this paper that are associated with the elliptic distribution (2.1) and its information matrix G in (2.8). It may therefore be regarded as an appendix to this paper.

To evaluate an integral of the form $\int_{\mathbb{R}^n} f dv$, we use polar coordinates

$$\int_{\mathbb{R}^n} f(x)dv(x) = \int_0^\infty r^{n-1}\left(\int_{S_n} f(rx)d\sigma(x)\right)dr \tag{4.1}$$

where dv is the volume Lebesgue measure of $\mathbb{R}^n$, $S_n = \{x \in \mathbb{R}^n: \|x\| = 1\}$ is the unit sphere of $\mathbb{R}^n$, and $d\sigma$ is its surface measure.

For $x = [x_1, \ldots, x_n]^\dagger \in \mathbb{R}^n$ and $\alpha = [\alpha_1, \ldots, \alpha_n]^\dagger \in \mathbb{Z}_+^n$, we use the multinomial notation of $x^\alpha = x_1^{\alpha_1} \ldots x_n^{\alpha_n}$ and $|\alpha| = \alpha_1 + \ldots + \alpha_n$. We also define a function $\delta: \mathbb{R}^n \longrightarrow \{0,1\}$ by letting $\delta(x) = 1$ if $x \in \mathbb{Z}_+^n$ and $\delta(x) = 0$ if $x \in \mathbb{R}^n \setminus \mathbb{Z}_+^n$.

LEMMA 4.1. *Let* $\alpha = [\alpha_1, \ldots, \alpha_n]^\dagger \in \mathbb{Z}_+^n$. *Then*

$$\int_{S_n} x^\alpha d\sigma(x) = \delta(\alpha/2)2\prod_{j=1}^n \Gamma\left((\alpha_j + 1)/2\right)/\Gamma\left((n + |\alpha|)/2\right).$$

In particular,

$$\sigma(S_n) = \int_{S_n} d\sigma = 2\pi^{n/2}/\Gamma(n/2).$$

Proof. Using (4.1), we find that

$$\int_{\mathbb{R}^n} x^\alpha e^{-\|x\|^2} dv(x) = \int_0^\infty r^{n+|\alpha|-1}e^{-r^2}\left(\int_{S_n} x^\alpha d\sigma(x)\right)dr$$

$$= \frac{1}{2}\Gamma\left(\frac{n+|\alpha|}{2}\right)\int_{S_n} x^\alpha d\sigma(x),$$

and thus

$$\int_{S_n} x^\alpha d\sigma(x) = \frac{2}{\Gamma(\frac{n+|\alpha|}{2})}\int_{\mathbb{R}^n} x^\alpha e^{-\|x\|^2} dv(x)$$

$$= \frac{2}{\Gamma(\frac{n+|\alpha|}{2})} \prod_{j=1}^{n} \int_{-\infty}^{\infty} t^{\alpha_j} e^{-t^2} dt.$$

If for some $1 \leq j \leq n$, α_j is odd, then the above product vanishes. Otherwise,

$$\int_{S_n} x^\alpha d\sigma(x) = \frac{2}{\Gamma(\frac{n+|\alpha|}{2})} \prod_{j=1}^{n} 2\int_0^{\infty} t^{\alpha_j} e^{-t^2} dt$$

$$= \frac{2}{\Gamma(\frac{n+|\alpha|}{2})} \prod_{j=1}^{n} \Gamma(\frac{\alpha_j+1}{2}),$$

and the lemma follows.

This lemma, together with (4.1), will enable us to prove that $p(\cdot|\mu, \Sigma)$ in (2.1) is a probability distribution, provided (2.2) is satisfied. Indeed, letting $y = \Sigma^{-1/2}(x - \mu)$, we have

$$\int_{\mathbb{R}^n} p(x|\mu, \Sigma)dv(x) = \frac{\Gamma(n/2)}{\pi^{n/2}}\int_{\mathbb{R}^n} F(\|y\|^2)dv(y)$$

$$= \frac{\Gamma(n/2)}{\pi^{n/2}}\int_0^{\infty} r^{n-1}F(r^2)(\int_{S_n} d\sigma)dr$$

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

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

Similarly, to prove (2.3), we observe that

$$\phi_F(t) = e^{it\cdot\mu}E(e^{it\cdot(X-\mu)}),$$

and so, using $y = \Sigma^{-1/2}(x-\mu)$ and $s = \Sigma^{1/2}t$,

$$E(e^{it\cdot(X-\mu)}) = \int_{\mathbb{R}^n} e^{it\cdot(x-\mu)}p(x|\mu,\Sigma)dv(x)$$

$$= \frac{\Gamma(n/2)}{\pi^{n/2}}\int_{\mathbb{R}^n} e^{is\cdot y}F(\|y\|^2)dv(y).$$

But, using (4.1) and Lemma 4.1 again, we obtain the well-known formula,

$$\int_{\mathbb{R}^n} e^{is\cdot y}F(\|y\|^2)dv(y) = (2\pi)^{n/2}\|s\|^{-(n-2)/2}\int_0^\infty r^{n/2}F(r^2)J_{n/2-1}(\|s\|r)dr,$$

and so

$$E(e^{it\cdot(X-\mu)}) = 2\Gamma(n/2)\int_0^\infty r^{n-1}F(r^2)K_{n/2-1}(r^2\|s\|^2)dr$$

$$= \Gamma(n/2)\int_0^\infty r^{n/2-1}F(r)K_{n/2-1}(r\|s\|^2)dr$$

$$= \Lambda_F(\|s\|^2) = \Lambda_F(t^\dagger\Sigma t),$$

and (2.3) follows.

We now consider the distribution $p(\cdot|\beta,\sigma)$ in (2.5), under the assumptions (2.2) and (2.6)-(2.7). To evaluate the information matrix G in (2.8), we calculate the matrices G_{jk} $(1 \leq j, k \leq 2)$ with the aim of proving (2.9). We let $Z = \sigma^{-1}\Sigma_0^{-1/2}(x - A\beta)$, to find

$$G_{11} = \frac{4}{\sigma^2} B^\dagger E((\mathcal{L}F)^2(\|Z\|^2)Z^\dagger Z)B,$$

$$G_{12} = G_{21}^\dagger = \frac{2}{\sigma^2} B^\dagger E((\mathcal{L}F)(\|Z\|^2)\left(n + 2\|Z\|^2(\mathcal{L}F)(\|Z\|^2)\right)Z)$$

and

$$G_{22} = \frac{1}{\sigma^2} E(\left(n + 2\|Z\|^2(\mathcal{L}F)(\|Z\|^2)\right)^2),$$

where $B = \Sigma_0^{-1/2}A$. We use (4.1) and Lemma 4.1 to compute the elements of the n×n matrix $E\{(\mathcal{L}F)^2(\|Z\|^2)Z^+Z\}$. The (i,j)-element is then

$$\frac{\Gamma(n/2)}{\pi^{n/2}}\int_{\mathbb{R}^n} (\mathcal{L}F)^2(\|z\|^2)z_iz_jF(\|z\|^2)dv(z)$$

$$= \frac{\Gamma(n/2)}{\pi^{n/2}}\int_0^\infty r^{n+1}(\mathcal{L}F)^2(r^2)F(r^2)\Big(\int_{S_n} z_j^2d\sigma(z)\Big)dr$$

$$= 2\delta_{ij}\frac{\pi^{(n-1)/2}\Gamma(3/2)}{\Gamma(n/2+1)}\frac{\Gamma(n/2)}{\pi^{n/2}}\int_0^\infty r^{n+1}(\mathcal{L}F)^2(r^2)F(r^2)dr$$

$$= \delta_{ij}\frac{1}{n}\int_0^\infty r^{n/2}F(r)\{(\mathcal{L}F)(r)\}^2dr = \delta_{ij}a/4,$$

and thus $G_{11} = a\sigma^{-2}A^+\Sigma_0^{-1}A$ as in (2.9).

Similarly,

$$E\{\Big(n+2\|Z\|^2(\mathcal{L}F)(\|Z\|^2)\Big)^2\}$$

$$= \frac{\Gamma(n/2)}{\pi^{n/2}}\int_0^\infty r^{n-1}\Big(n+2r^2(\mathcal{L}F)(r^2)\Big)^2F(r^2)\Big(\int_{S_n}(d\sigma)dr$$

$$= \int_0^\infty r^{n/2-1}\{n+2r(\mathcal{L}F)(r)\}^2dr = b,$$

anc .o $G_{22} = b\sigma^{-2}$ as in (2.9).

Finally, the n×1 expection-matrix appearing in the m×1 matrix G_{12} is finite by virtue of the Cauchy-Schwarz inequality and by the finiteness of G_{11} and G_{22}. It follows from (4.1) and Lemma 4.1 that $G_{12} = G_{21}^+ = 0$ as in (2.9).

REFERENCES

[1] BURBEA, J. (1986a). Informative geometry of probability spaces. *Expositiones Math.* **4**, 347-378.

[2] BURBEA, J. (1986b). Metrics and distances on probability spaces. *Encycy. Statist. Sci.* **7**, 241-248.

[3] 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.

[4] BURBEA, J. and RAO, C.R. (1984). Differential metrics in probability spaces. *Probability Math. Statist.* **3**, 241-266.

[5] KELKER, D. (1970). Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhyā A* **32**, 419-430.

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

[7] MUIRHEAD, R.J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York.

[8] OLLER, J.M. (1987). Information metric for extreme value and logistic probability distributions. *Sankhyā A* **49**, 17-23.

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

[10] WOLF, J.A. (1967). *Spaces of Constant Curvature*, McGraw-Hill, New York.

END

FILMED

DATE: 4-92

DTIC