Discrimination for spherical data

John T. Kent and Kanti V. Mardia

Department of Statistics, University of Leeds

1 Introduction

Consider two populations of measurements in p-dimensional Euclidean space. A standard multivariate normal model allows different mean vectors but a common covariance matrix for the two groups, $N_p(\mu_1, \Sigma)$ and $N_p(\mu_2, \Sigma)$. This model is the foundation for much multivariate analysis, including the multivariate analysis of variance (MANOVA), Hotelling's T^2 test and Fisher's linear discriminant rule (e.g. Mardia et al., 1979). There is some mild simplification to the theory under an assumption of isotropy, $\Sigma \propto I_p$, but in most applications Σ is not restricted in this way. The purpose of this paper is to develop similar models for spherical data.

2 The Fisher Bingham distribution on the sphere

The unit sphere S_2 in $\mathbb{R}^3$ is defined by $\{\mathbf{x} = (x_1, x_2, x_3)^T : x_1^2 + x_2^2 + x_3^2 = 1\}$. An analogue on S_2 of the bivariate normal distribution on $\mathbb{R}^2$ is the FB$_5$ distribution of Kent (1982). In standardized form the probability density function takes the form

$$f(\mathbf{x}; \kappa, \beta, I) = \frac{1}{c(\beta, \kappa)} \exp\{\kappa x_3 + \beta(x_1^2 - x_2^2)\}, \tag{1}$$

where $0 \le 2\beta \le \kappa$. For the standardized version, the modal direction is the north pole $(0, 0, 1)^T$, and the major and minor axis of the covariance structure lie on the x_1 and x_2 axes, respectively. The general form of the distribution, denoted FB$_5(\kappa, \beta, \Gamma)$, where $\Gamma =$ is a 3×3 rotation matrix, is obtained from the standardized form by transforming $\mathbf{x}$ to $\mathbf{y} = \Gamma\mathbf{x}$. The third column of Γ gives the modal direction of $\mathbf{y}$. If $P = I - \boldsymbol{\gamma}_{(3)}\boldsymbol{\gamma}_{(3)}^T$ denote projection onto the plane perpendicular to $\boldsymbol{\gamma}_{(3)}$, then the major and minor axes of ellipse determined by the rank 2 covariance matrix $E\{P\mathbf{y}\mathbf{y}^T P\}$ are given by the first two columns of Γ.

The special case $\beta = 0$ reduces to the Fisher distribution on the sphere and corresponds to the isotropic bivariate normal distribution in $\mathbb{R}^2$. Discrimination between two Fisher distributions with equal concentration is much simpler than for the FB$_5$ distribution (see Section 4) and does not require the complications of parallel transport that we describe next.

3 Parallel transport

Since a covariance structure lies in the tangent plane at a mean direction, it is not possible for covariance structures at different mean directions to be directly compared with one another. Instead the concept of parallel transport is needed to bring the tangent spaces together. This concept is most easily explained by letting one mean direction be the north pole, $\boldsymbol{\mu}_0 = (0, 0, 1)^T$ and the other direction be an arbitrary point on the sphere

$$\boldsymbol{\mu}_1 = (\sin\theta_1 \cos\phi_1, \sin\theta_1 \sin\phi_1, \cos\theta_1)^T, \tag{2}$$

not lying at the north or south pole, where $0 < \theta_1 < \pi$ denotes the colatitude and $\phi_1 \in [0, 2\pi)$ denotes the longitude.

For $i, j, = 1, 2, 3,\ i \neq j$, and an angle ψ, let $R_{ij}(\psi)$ denote the 3×3 rotation matrix with nonzero elements $r_{ii} = r_{jj} = \cos\psi$, $r_{ij} = -r_{ji} = \sin\psi$, and with the remaining diagonal element equal to one. Then construct the rotation matrix

$$R = R_{21}(\phi_1)R_{13}(\theta_1)R_{12}(\phi_1), \tag{3}$$

and let $\boldsymbol{\tau} = (-\sin\phi_1, \cos\phi_1, 0)^T$. It is easy to check that $R\boldsymbol{\mu}_0 = \boldsymbol{\mu}_1$, and $R\boldsymbol{\tau} = \boldsymbol{\tau}$, these two conditions determining R. Since $\boldsymbol{\tau}^T\boldsymbol{\mu}_0 = 0, \boldsymbol{\tau}^T\boldsymbol{\mu}_1 = 0$, R represents rotation about the axis $\boldsymbol{\tau}$ perpendicular to the plane determined by $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_1$, and R rotates $\boldsymbol{\mu}_0$ to $\boldsymbol{\mu}_1$. It can be checked that $R^T = R_{21}(\phi_1 + \pi)R_{13}(\theta_1)R_{12}(\phi_1 + \pi)$ represents the reverse rotation, taking $\boldsymbol{\mu}_1$ to $\boldsymbol{\mu}_0$.

We are now ready to set out the parameterize the two-group model in three steps.

(a) Start with a standardized $\text{FB}_5(\kappa, \beta, I)$ distribution as in equation (1), with modal direction at the north pole and principal axes equal to the first two coordinate axes.

(b) Consider a vector $\boldsymbol{\mu}_1$ in equation (2) with polar coordinates (θ_1, ϕ_1), restricted to lie in the northern hemisphere, so that the colatitude satisfies $0 < \theta_1 < \pi/2$. Construct two new distributions by parallel transport with modal directions equi-spaced about the north pole with polar coordinates (θ_1, ϕ_1) and $(\theta_1, \phi_1 + \pi)$: the two distributions are $\text{FB}_5(\kappa, \beta, R)$ and $\text{FB}_5(\kappa, \beta, R^T)$, with mean directions

$$\boldsymbol{\mu}_1 = R\boldsymbol{\mu}_0 \quad \text{and} \quad \boldsymbol{\mu}_2 = R^T\boldsymbol{\mu}_0. \tag{4}$$

The north pole $\boldsymbol{\mu}_0$ will be called the "central mean direction" because it points towards the vector average of the mean directions of the two groups.

(c) Lastly, using a general rotation matrix Ω, shift the central mean direction to an arbitrary point on the sphere, and allow the principal axes at the central mean direction to have an arbitrary orientation, yielding the two distributions $\text{FB}_5(\kappa, \beta, \Omega R)$ and $\text{FB}_5(\kappa, \beta, \Omega R^T)$.

It is instructive to compare this model with the Euclidean case. For the bivariate normal model there are two parameters for each of two mean vectors and 3 parameters needed to specify the common covariance matrix, making a total of 7 parameters. For the spherical model there are two concentration parameters (κ and β), two angular parameters (θ_1 and ϕ_1) for the parallel transport matrix R in (3) used to construct the two means in standardized cooordinates in (4), and 3 parameters to specify Ω, again making 7 parameters. Indeed it can be shown that when these two distributions are concentrated on a small part of the sphere, the two-group FB_5 model projected onto the tangent space of the sphere closely matches the bivariate normal model.

4 Estimation, testing and discrimination

Here we give a simple method of moment estimation, which can be used both as an estimation method in its own right, or as a starting point for ML estimation. The data take the form of unit vectors $\mathbf{y}_{11}, \ldots, \mathbf{y}_{1n_1}$ from group 1 and $\mathbf{y}_{21}, \ldots, \mathbf{y}_{2n_2}$ from group 2.

The fitting procedure can be sketched as follows. Rotate the whole dataset so that the sample central mean direction is at the north pole. Then use parallel transport on each group separately to make it centered at the north pole. Using moment estimation on the pooled dataset (Kent, 1982), fit an FB_5 distribution. A final rotation about the north pole is needed to ensure that the major and minor axes of fitted FB_5 distribution lie along the first two coordinate axes. After

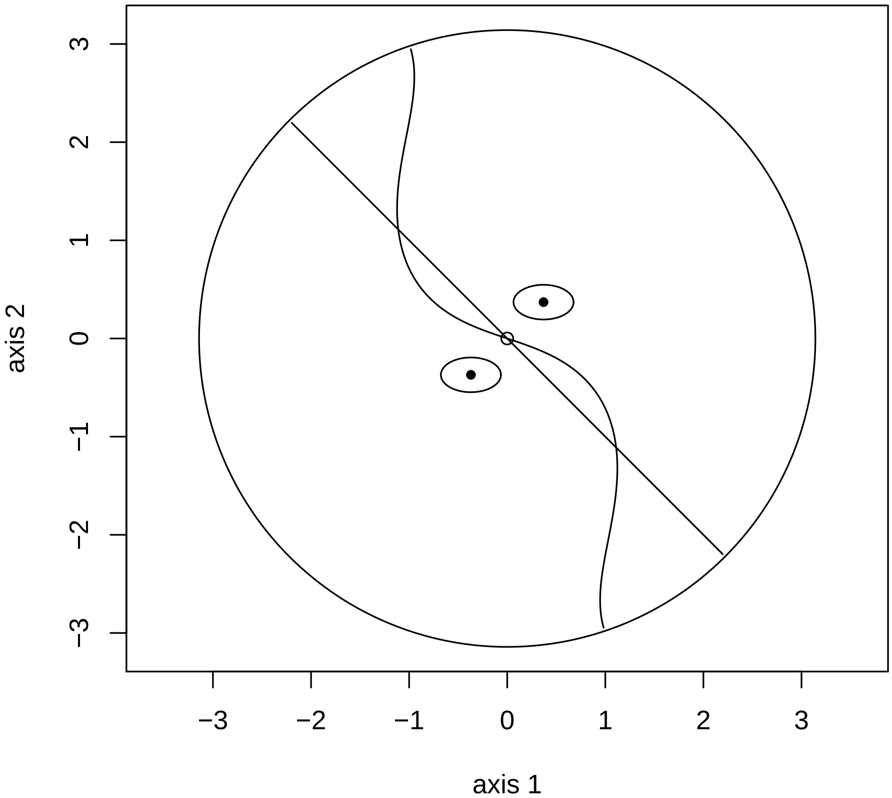

Figure 1: Azimuthal equidistant projection of the sphere in coordinates $(\theta \cos \phi, \theta \sin \phi)$, with the north pole given by the open circle at the center of the figure and the south pole by the circle at radius π. The elongated "S"-shaped curve through the north pole gives the discriminant boundary between two FB_5 distributions with parameters $\theta_1 = \pi/6$, $\phi_1 = \pi/4$, $2\beta/\kappa = 0.5$ and represented by the two ellipses centered at the black dots. The straight line through the north pole is the great circle discriminant boundary for two Fisher distributions, relevant when $2\beta/\kappa = 0$.

this fitting procedure, the data have been transformed into the standardized form of part (b) in Section 3.

The first task when presented with two groups of data is to test $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$, i.e. that the two mean directions are equal. Ignoring the β parameter (i.e. treating it as 0), yields the standard Watson-Williams test from directional data analysis to compare the mean directions of two Fisher distributions (e.g. Mardia and Jupp, 2000, p. 219). By incorporating the estimate of β, it should be possible to get more faithful size and higher power, at least for large samples.

Once the difference between the two groups has been established, it makes sense to consider discrimination. Here we limit attention to the maximum likelihood discriminant rule. The boundary between the two regions on the sphere is given by a curve on the sphere where the ratio of two probability densities equals 1.

In the standardized version of the problem, the boundary is simple if $\phi_1 = 0, \pi/2, \pi$ or $3\pi/2$, so that the difference in group means is aligned with one of the principal axes. In any of these cases the boundary simplifies to the great circle passing through the central mode at the north pole and perpendicular to the great circle between $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$. This boundary is also applicable in the simplified case $\beta = 0$, under which FB_5 reduces to the Fisher distribution.

However, for nonspecial values of ϕ_1 when $\beta \neq 0$, the boundary is messy to describe analytically. An illustration is given in Figure 1 where the actual boundary (the elongated "S"-shaped curve passing through the north pole), calculated numerically, is compared to the corresponding great circle boundary (the straight line passing through the north pole) under the Fisher model (i.e. taking $\beta = 0$). This example has been chosen to give a situation where the difference between the two boundaries is substantial.

5 Discussion

As far as we are aware, the first use of parallel transport on the sphere for a statistical problem was given in Jupp and Kent (1987) for a nonparametric smoothing problem, where it was described in terms of "unrolling" a path on the sphere. Subsequent developments have included applications to regression problems and extensions to more complicated manifolds (e.g., Le, 2003; Kume et al., 2007; Pauley, 2012; and Su et al., 2012). Mardia et al. (2004) have used parallel transport investigate a problem involving edgels in landmark-based planar shape analysis. The current paper is somewhat different in emphasis because it involves the parallel transport of distributions.

References

Jupp, P.E. and Kent, J.T. (1987). Fitting smooth paths to spherical data. *Appl. Statist.*, bf 36, 34–46.

Kent, J.T. (1982). The Fisher-Bingham distribution on the sphere. *J. Roy. Statist. Soc. B*, **44**, 71–80.

Kume, A., Dryden, I.L., and Le, H. (2007). Shape-space smoothing splines for planar landmarks. *Biometrika*, **94**, 513–528.

Le, H. (2003). Unrolling shape curves. *J. Lond. Math. Soc*, **68**, 511–526.

Mardia, K.V., Kent, J.T. and Bibby, J.M. (1979). *Multivariate Analysis*. Academic Press, London.

Mardia, K.V. and Jupp, P.E. (2000). *Directional Statistics*. Wiley, Chichester.

Mardia, K.V., Kirkbride, J. and Bookstein, F.L. (2004). Statistics of shape, direction and cylindrical variables. *J. Appl. Statist.*, **31**, 465–479.

Pauley, M. (2011). *Cubics, Curvature and Asymptotics*. PhD thesis, University of Western Australia.

Su, J., Dryden, I.L., Klassen, E., Le, H. and Srivastava, A. (2012). Fitting optimal curves to time indexed, noisy observations of stochastic processes on nonlinear manifolds. *J. Image and Vision Computing*, **30**, 428–442