Fisher-Bingham and Kent distributions on Stiefel Manifolds

Andrew T.A. Wood

School of Mathematical Sciences, University of Nottingham

1 Introduction

In a paper by Kume et al. (2013) soon to be published, procedures for approximating the normalizing constant of two families distributions are presented. These families are the Fisher-Bingham distributions defined on Stiefel manifolds (see below), and Fisher-Bingham distributions defined on products of spheres. The term 'Fisher-Bingham' here means an exponential family distribution whose exponent is a sum of linear and quadratic functions of the data. The procedures used to calculate these normalizing constants are based on saddlepoint approximations; Matlab code for doing the calculations is available.

The principal goal of this presentation is to propose a subfamily, referred to as the Kent distribution, of the Fisher-Bingham family on a Stiefel manifold, cf. the Kent (1982) distribution on a sphere. The main charactistics of Kent distributions are (i) they have the same number of parameters as a general multivariate normal distribution defined on the tangent space; and (ii) under certain conditions, Kent distributions approach the multivariate normal as the concentration of the distribution increases. Kent distributions on Stiefel manifolds are of more than purely academic interest: in particular, they avoid certain difficulties (see Section 3) which sometimes arise when estimating the parameters of the Fisher matrix distribution (i.e. the linear exponential family) on a Stiefel manifold.

2 The Fisher-Bingham distribution on a Stiefel manifold

For $d \geq q$, the Stiefel manifold $\mathcal{V}_{d,q}$ is defined by $\mathcal{V}_{d,q} = \{Y\,(d \times q)\colon Y^\top Y = I_q\}$, where I_q is the $q \times q$ identity matrix. Note that $\mathcal{V}_{d,d}$ consists of the space of 3×3 orthogonal matrices. Write $X = (x_1, \ldots, x_q)$ for the $d \times q$ matrix with columns $x_1, \ldots, x_q$. The density of the Fisher–Bingham distribution on $\mathcal{V}_{d,q}$, with respect to Lebesgue (or geometric) measure on $\mathcal{V}_{d,q}$, is

$$\begin{aligned} f(X \mid A, B) &= \mathcal{C}(A,B)^{-1} \exp\left\{\mathrm{tr}(A^\top X) + \mathrm{vec}(X)^\top B \,\mathrm{vec}(X)\right\} \\ &= \mathcal{C}(A,B)^{-1} \exp\left(\textstyle\sum_{i=1}^{q} a_i^\top x_i + \sum_{i,j=1}^{q} x_i^\top B_{ij} x_j\right), \end{aligned} \tag{1}$$

where $A = (a_1, \ldots, a_q)$ is a $d \times q$ matrix, $B = (B_{ij})_{i,j=1}^{q}$ consists of q^2 blocks B_{ij} of dimension $d \times d$, and $\mathcal{C}(A,B)$ is chosen so that the integral of $f(X \mid A, B)$ over $\mathcal{V}_{d,q}$ is 1. Kume et al. (2013) provide saddlepoint approximations for the normalizing constant $\mathcal{C}(A,B)$ which achieve good numerical accuracy.

The family (1) contains a number of distributions of importance in directional statistics and the statistics of orientations. The Fisher matrix distribution, corresponding to the case where B in (1) is a matrix of zeros, has been considered by a number of authors including Downs (1972), Khatri & Mardia (1977), Jupp & Mardia (1979), Prentice (1986) and Wood (1993). It turns out that, in the Fisher matrix case, the saddlepoint approximations of Kume et al. (2013) are available in closed form. The Bingham version of (1), corresponding to the case where A

is the matrix of zeros, has been studied by Arnold & Jupp (2013); see also Hoff (2009), who describes a method for simulating from (1).

3 A limitation of the Fisher matrix distribution

A rigid body motion in $\mathbf{R}^d$ relative to a fixed coordinate system is defined by a pair (X, y) where X is a $d \times d$ rotation matrix which represents the change in orientation of the object, while $y \in \mathbf{R}^d$ is a vector which represents the translation of the object. In a recent paper, Oualkacha and Rivest (2012) develop statistical methods for the analysis of rigid body motion data, focusing on the question of how to define an average rigid body motion. They applied their approach to the analysis of an ankle-joint dataset; this dataset was first presented in Rivest et al. (2008). An issue not explored by Oualkacha and Rivest (2012) is the development of parametric models which incorporate dependence between X and y.

In Kume et al. (2013, Section 5), a parametric model involving (1) was fitted to the ankle-joint dataset. The data of Rivest et al. (2008) are highly concentrated, yet unexpectedly it turned out that the Fisher matrix model (obtained by taking B to be the matrix of zeros in (1)) did not provide a good fit, even though plots suggest the Gaussian tangent-space approximation should be reasonable. To understand what goes wrong, consider the singular value decomposition $A = Q\Omega R^\top$, write

$$Y = Q^\top X R = \exp(H) \approx I + H + H^2/2,$$

where $H = (h_{ij})$ is $d \times d$ is a skew-symmetric matrix, i.e. $h_{ii} = 0$ and $h_{ij} = -h_{ji}$.

Put $H = (Y - Y^\top)/2$. Then

$$\exp\{\text{tr}(A^\top X)\} \approx \exp\left[\text{tr}\left\{\Omega\left(I + H + \frac{1}{2}H^2\right)\right\}\right] \propto \exp\left\{-\frac{1}{2}\sum_{i<j} h_{ij}^2(\omega_i + \omega_j)\right\}.$$

It follows that, for large $\omega_1, \ldots, \omega_d$, the elements h_{ij}, for $i < j$, have approximately independent Gaussian distributions with zero mean and variance $\sigma_{ij}^2 = (\omega_i + \omega_j)^{-1}$. Conversely, however, the h_{ij} being asymptotically independent and Gaussian does not necessarily lead to a well-matched matrix Fisher distribution for X: for the case $d = 3$, $\omega_1 = (\sigma_{12}^{-2} + \sigma_{13}^{-2} - \sigma_{23}^{-2})/2$, $\omega_2 = (\sigma_{12}^{-2} + \sigma_{23}^{-2} - \sigma_{13}^{-2})/2$, $\omega_3 = (\sigma_{13}^{-2} + \sigma_{23}^{-2} - \sigma_{12}^{-2})/2$, and these expressions impose constraints on the σ_{ij} in order that $\omega_i > 0$. If some of the σ_{ij}^2 are sufficiently different from each other, then the positivity constraint $\omega_i > 0$ may not be achieved by all of the ω_i. For the data of Rivest et al. (2008), the sample analogues of the σ_{ij}^2 are given by $\hat{\sigma}_{12}^2 = 0.0008$, $\hat{\sigma}_{13}^2 = 0.0352$, $\hat{\sigma}_{13}^2 = 0.0708$, which corresponds to a negative ω_3. This explains why the Fisher matrix model does not produce a good fit here.

4 The Kent distribution on a Steifel manifold

Partly motivated by the difficulty identified in Section 3, we now propose a definition of the Kent distribution on a general Stiefel manifold $\mathcal{V}_{d,q}$. The advantage of using this model is that, in the common situation in which data are highly concentrated and unimodal, it behaves approximately like a multivariate normal distribution in the tangent space.

First of all we shall present the distribution in a special coordinate system chosen for convenience. Then the form of the distribution in general coordinates will be given.

Assume as before that $d \geq q$ and define the $d \times q$ matrix

$$M = [e_1, \ldots, e_q], \tag{2}$$

where e_r is a $d \times 1$ vector with all components zero except for component r, which is 1. Consider the $q \times d$ matrix $H = (h_{ij} : i = 1, \ldots, d; j = 1, \ldots q)$. Assume that H has the structure $H = [H_1^\top, H_2^\top]^\top$ where H_1 is a $q \times q$ skew-symmetric matrix and H_2 is a $(d-q) \times q$ matrix. A key point is that a general element of the tangent space to $\mathcal{V}_{d,q}$ at M defined in (2) can be represented in the form H.

For a matrix $X = (x_{ij} : i = 1, \ldots, d; j = 1, \ldots, q) \in \mathcal{V}_{d,q}$ define

$$h_{ij} = \begin{cases} (x_{ij} - x_{ji})/2 & \text{if} \quad 1 \leq i \leq q \\ x_{ij} & \text{if} \quad q < i \leq d. \end{cases} \tag{3}$$

Using (3), define the $q(q-1)/2 \times 1$ dimensional vector $u = \text{vec}_{LT}(H_1)$ and the $q(d-q) \times 1$ dimensional vector $v = \text{vec}(H_2)$, where vec is the standard operator which stacks the columns of a matrix to form a vector, and vec_{LT} is defined in a similar way except that it only stacks the lower triangular elements, i.e. the elements strictly below the diagonal. Now define w as $w = (u^\top, v^\top)^\top$.

Then the probability density function of the Kent distribution with respect to Lebesgue measure on $\mathcal{V}_{d,q}$ is given by

$$f_K(X; \lambda, M, \Xi) = \mathcal{C}_1(\lambda, \Xi)^{-1} \exp\left\{\lambda \, \text{tr}(M^\top X) + w^\top \Xi w\right\}, \tag{4}$$

where $\lambda > 0$, M is the special element of $\mathcal{V}_{d,q}$ defined in (2) and Ξ is a symmetric parameter matrix which satisfies $\text{tr}(\Xi) = 0$. Note that, due to (3), each component of w is a linear function of components of X, and therefore (4) is a sub-family of (1).

Let us now consider the general case. Here, M is a general element of $\mathcal{V}_{d,q}$. The singular value decomposition of M is given by $\sum_{i=1}^q \gamma_i s_i t_i^\top$ where the γ_i are non-negative singular values, and actually equal to 1 because $M \in \mathcal{V}_{d,q}$; $s_1, \ldots, s_q$ are orthonormal d-vectors; and $t_1, \ldots, t_q$ are orthonormal q-vectors. If $d > q$, let $s_{q+1}, \ldots, s_d$ be any set of d-vectors such that $s_1, \ldots, s_d$ form an orthonormal basis of $\mathbf{R}^d$. Then the general form of (4) is given by

$$f_K(Y; \lambda, M, \Xi) = \mathcal{C}(\lambda, \Xi)^{-1} \exp\left\{\lambda \, \text{tr}(M^\top Y) + w^\top \Xi w\right\}, \tag{5}$$

where w is defined in terms of X as before but now the x_{ij} are given by $x_{ij} = s_i^\top Y t_j$ for each $i = 1, \ldots, d$ and $j = 1, \ldots, q$.

We now check the dimension of the statistical model (5). Note that M has

$$q(d-q) + \frac{1}{2}q(q-1) = qd - \frac{1}{2}q(q+1)$$

degrees of freedom; Ξ, a symmetric matrix of dimension $qd - q(q+1)/2$ satisfying $\text{tr}(\Xi) = 0$, has

$$\frac{1}{2}\left\{qd - \frac{1}{2}q(q+1)\right\}\left\{qd - \frac{1}{2}q(q+1) + 1\right\} - 1$$

degrees of freedom; and λ, a real-valued quantity, has one degree of freedom. Adding these numbers together, it is seen that we end up with a statistical model which has the same number of parameters as the $\{qd - q(q+1)/2\}$-dimensional multivariate normal distribution.

Further discussion and explanation will be given during the presentation, along with details of an estimation procedure for the Kent distribution and some numerical results.

References

Arnold, R. and Jupp, P.E. (2013). Statistics of orthogonal frames. *Biometrika* (to appear).

Downs, T. D. (1972). Orientation statistics. *Biometrika* **59**, 66576.

Hoff, P. (2009). Simulation of the matrix Bingham–von Mises–Fisher distribution, with applications to multivariate and relational data. *J. Comp. Graphical Stats.*, **18**, 438–56.

Jupp, P.E. & Mardia, K.V. (1979). Maximum likelihood estimators for the matrix von Mises–Fisher and Bingham distributions. *Ann. Statist.* **7**, 599–606.

Kent, J.T. (1982). The Fisher–Bingham distribution on the sphere. *J.R. Statist. Soc.* B **44**, 285–99.

Kume, A., Preston, S.P. and Wood, A.T.A. (2013). Saddlepoint approximations for the normalizing constant of Fisher-Bingham distributions on products of spheres and Stiefel manifolds. *Biometrika* (to appear).

Khatri, C.G. & Mardia, K.V. (1977). The von Mises–Fisher matrix distribution in orientation statistics. *J.R. Statist. Soc.* B **39**, 95–106.

Oualkacha, K. & Rivest, L.-P. (2012). On the estimation of an average rigid body motion. *Biometrika* **99**, 585–598.

Prentice, M.J. (1986). Orientation statistics without parametric assumptions. *J.R. Statist. Soc.* B **48**, 214–22.

Rivest, L.-P., Baillargeon, S. & Pierrynowski, M.R. (2008). A directional model for the estimation of the rotation axes of the ankle joint. *J. Am. Statist. Assoc.* **103**, 1060–9.

Wood, A.T.A. (1993). Estimation of the concentration parameters of the Fisher matrix distribution on $SO(3)$ and the Bingham distribution on S_q, $q \geq 2$. *Aust. J. Statist.* **35**, 69–79.