Statistical Complexity in Protein Bioinformatics

Kanti V. Mardia

Department of Statistics, University of Leeds

1 Introduction

Problems in protein bioinformatics have led to various challenges in statistics. Here we focus basically on a few problems related to shape analysis and directional statistics.

Random rotations have been reappearing recently in various problems. In Section 2, we summarize some key points for the Fisher matrix distribution which is the most plausible distribution for random rotations. It requires a treatment of Haar measure which in practice needs a coordinate system. Eulerian angles are natural but the choice is not unique and neither is the Jacobian simple. We deal with the three dimensional case in detail motivated by applications in protein structure. Eulerian angles also appear in the registration of form (ie. when rigid tranformations are filtered out), and a Bookstein type registration is given in Section 3. This is important in matching two protein backbones. These sections involve Euler's angles of one type or another.

Another important problem is how to make inference for directional models, especially where the normalizing constant is intractable. Section 5 gives a recent approach to the saddlepoint approximation to directional distributions. Another problem is how to sample such distributions efficiently (especially when the number of parameters are large). One approach which seems to fit well in directional statistics is importance sampling, and we describe a simple case which would make a building block. Bayesian analysis for directional distributions is becoming popular and we comment on the current trend in Section 6. We end the paper with a discussion.

2 Random Rotations and the Matrix Fisher Distribution

2.1 Introduction

Distributions on rotations are becoming increasingly important and we give a brief overview (with some new results) of the most common distributions on rotations, namely the Fisher matrix distribution studied by Downs (1972) and Khatri and Mardia (1977). Since then Green and Mardia (2006) and Mardia (2009) have given some further properties.

Let X be a $p \times p$ rotation matrix so that

$$X^T X = X X^T = I, \;\; X \in SO(p) \tag{1}$$

with $|X| = 1$ and I is the identity matrix. We will write the uniform distribution on $SO(p)$ as

$$[dX], \;\; X \in SO(p) \tag{2}$$

which is the Haar measure scaled to have unit mass. The matrix Fisher distribution on a rotation X (Downs, 1974; Khatri and Mardia, 1977) has probability density function (pdf)

$$a(F) \exp\{tr F^T X\} [dX], \; X \in SO(p), \tag{3}$$

where F is a $p \times p$ parameter matrix. Note that (3) is usually defined for the general case of Stiefel manifold, and here we are dealing with a special case. Also the same distribution is applicable to $X \in O(p)$. For further details, see also Mardia and Jupp (2000). For another model, see León et al (2006).

2.2 Uniform Distributions

The Haar measure is well known in multivariate analysis (see, for example, Muirhead, 1982) but in practice we need to express this uniform distribution with respesct to a particular parameterization. Khatri and Mardia (1977) have provided an Euler type representation for any dimensions. We will consider first the general Euler parameterization. Let us define $p(p-1)/2$ Eulerian angles for X as

$$\theta_{ij}, i < j = 1, \ldots, p, \tag{4}$$

where

$$\left\{ -\pi \leq \theta_{i,i+1} \leq \pi,\ -\frac{1}{2}\pi \leq \theta_{ij} \leq \frac{1}{2}\pi,\ \ i = 1, 2, \ldots, p;\ \ j = i+2, \ldots, p \right\}.$$

Thus for $p = 3$, we have the three Eulerian angles θ_{12}, θ_{13} and θ_{23} where

$$-\pi \leq \theta_{12}, \theta_{23} \leq \pi \text{ and } -\frac{1}{2}\pi \leq \theta_{13} \leq \frac{1}{2}\pi. \tag{5}$$

Let $H_{ij}(\theta_{ij})$ be an orthogonal matrix such that in the diagonal places, there are unities except at (i,i)th and (j,j)th places in which there is $\cos\theta_{ij}$, and in the off-diagonal places, there are zeros except at (i,j)th and (j,i)th places in which there are $-\sin\theta_{ij}$ and $\sin\theta_{ij}$ respectively $(j > i)$. Define

$$H^{(j)} = H_{j-1,j}(\theta_{j-1,j}) \ldots H_{1j}(\theta_{1j}),\ \ H = H^{(p)} H^{(p-1)} \ldots H^{(2)}. \tag{6}$$

The matrix H then gives a general rotation matrix in terms of Euler angles. We can have different forms of orthogonal matrices which can be obtained by permuting the orders of multiplications mentioned in (6). There are such $\left(\frac{p(p-1)}{2}\right)!$ different orthogonal matrices which will give different (or same) Jacobian of transformations. Also, we can permute unities in $H_{ij}(.)$.

Suppose we select the independent elements of H as

$$\{h_{ij}(\theta_{ij}),\ \ i < j,\ \ j = 2, \ldots, p\}. \tag{7}$$

Further let H_{ii} be the sub-matrix obtained from H by taking the first i rows and i columns. Then Khatri and Mardia (1977) have shown that the Haar measure (with respect to $\Pi d\theta_{ij}$) becomes

$$[dH] = \pi^{-\frac{1}{2}p^2} \{\Gamma_p\left(\frac{1}{2}p\right)\} \prod_{i<j=1}^{p} \{dh_{ij}(\theta_{ij})/d\theta_{ij}\} / \prod_{i=1}^{p} |H_{ii}|_+ , \tag{8}$$

where $|H_{pp}| = |H| = 1$, and

$$\Gamma_p\left(\frac{1}{2}p\right) = \pi^{\frac{1}{4}p(p-1)} \prod_{j=1}^{p} \Gamma\left\{\frac{1}{2}(p-j+1)\right\}.$$

If the independent elements of H selected in (7) are different then the term $\prod_{i=1}^{p} |H_{ii}|_+$ needs to be replaced by the full determinant from the Jacobian of the linear transformation of the elements of the skew symmetric matrix $H(dH)^T$ to the independent elements in $(dH)^T$.

The most important case in practice is for $p = 3$ (for $p = 2$ we have the well known uniform distribution on circle). We now consider the case for $p = 3$ with the support given by (5) on the angles. Note that there are many variations in parameterization even in this case (see below). First note that we have from (6)

$$\begin{aligned} H^{(2)} &= H_{12}(\theta_{12}),\ H^{(3)} = H_{23}(\theta_{23})\ H_{13}(\theta_{13}), \\ H &= H_{23}(\theta_{23})\ H_{13}(\theta_{13})\ H_{12}(\theta_{12}) = H(\theta_{12}, \theta_{13}, \theta_{23}),\ \text{say}, \end{aligned} \tag{9}$$

$$H_{12}(\theta_{12}) = \begin{pmatrix} \cos\theta_{12} & -\sin\theta_{12} & 0 \\ \sin\theta_{12} & \cos\theta_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$H_{13}(\theta_{13}) = \begin{pmatrix} \cos\theta_{13} & 0 & -\sin\theta_{13} \\ 0 & 1 & 0 \\ \sin\theta_{13} & 0 & \cos\theta_{13} \end{pmatrix},\quad H_{23}(\theta_{23}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_{23} & -\sin\theta_{23} \\ 0 & \sin\theta_{23} & \cos\theta_{23} \end{pmatrix}.$$

It is found from (9) that here H is given by

$$\begin{pmatrix} \cos\theta_{13}\cos\theta_{12} & -\cos\theta_{13}\sin\theta_{12} & -\sin\theta_{13} \\ -\sin\theta_{23}\sin\theta_{13}\cos\theta_{12} + \cos\theta_{23}\sin\theta_{12} & \sin\theta_{23}\sin\theta_{13}\sin\theta_{12} + \cos\theta_{23}\cos\theta_{12} & -\sin\theta_{23}\cos\theta_{13} \\ \cos\theta_{23}\sin\theta_{13}\cos\theta_{12} + \sin\theta_{23}\cos\theta_{12} & \cos\theta_{23}\sin\theta_{13}\sin\theta_{12} + \sin\theta_{23}\cos\theta_{12} & \cos\theta_{23}\cos\theta_{13} \end{pmatrix}.$$

Thus in this representation, the independent elements from (7) are

$$h_{12}(\theta_{12}) = -\cos\theta_{13}\sin\theta_{12},\quad h_{13}(\theta_{13}) = -\sin\theta_{13},\quad h_{23}(\theta_{23}) = -\sin\theta_{23}\cos\theta_{13}. \tag{10}$$

Also we can compute

$$H_{11} = |h_{11}|,\quad H_{22} = \begin{vmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{vmatrix}. \tag{11}$$

Substituting these in (8), we find that the pdf of the uniform distribution on $(\theta_{12}, \theta_{13}, \theta_{23})$ is given by

$$(8\pi^2)^{-1}\cos\theta_{13}. \tag{12}$$

Indeed, this a particular case of the uniform distribution for any Stiefel manifold given in Khatri and Mardia (1977) with Euler parameterization.

Consider now the standard Eulerian transformation with three angles α, β, γ (see for example, Fisher et al, 1987, p.32)

$$A(\alpha, \beta, \gamma) = A_1(\gamma)A_2(\alpha)A_3(\beta) = (a_{ij})$$

where

$$A_1(\gamma) = \begin{pmatrix} \cos\gamma & \sin\gamma & 0 \\ -\sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix},\quad A_2(\alpha) = \begin{pmatrix} \cos\alpha & 0 & -\sin\alpha \\ 0 & 1 & 0 \\ \sin\alpha & 0 & \cos\alpha \end{pmatrix},$$

$$A_3(\beta) = \begin{pmatrix} \cos\beta & \sin\beta & 0 \\ -\sin\beta & \cos\beta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

where $0 \leq \beta, \gamma \leq 2\pi, \;\; 0 < \alpha < \pi$. It is seen that $A(\alpha, \beta, \gamma)$ is given by

$$\begin{pmatrix} \cos\alpha\cos\beta\cos\gamma - \sin\beta\sin\gamma & \cos\alpha\sin\beta\cos\gamma + \cos\beta\sin\gamma & -\sin\alpha\cos\gamma \\ -\cos\alpha\cos\beta\sin\gamma - \sin\beta\cos\gamma & -\cos\alpha\sin\beta\sin\gamma + \cos\beta\cos\gamma & \sin\alpha\sin\gamma \\ \sin\alpha\cos\beta & \sin\alpha\sin\beta & \cos\alpha \end{pmatrix}.$$

Here a set of independent variables is given by

$$a_{33} = \cos\alpha, \;\; a_{32} = \sin\alpha\sin\beta, \;\; a_{23} = \sin\alpha\sin\gamma. \tag{13}$$

If we identify (10) with (13), then

$$\alpha = \theta_{13} + \frac{\pi}{2}, \;\; \beta = \theta_{12}, \;\; \gamma = \theta_{23},$$

and the uniform pdf in (12) becomes

$$(8\pi^2)^{-1} \sin\alpha. \tag{14}$$

This result was derived using a geometrical argument by Miles (1965). Note that the results can be proved directly using the "conditional" Jacobians as in Mardia (2009a). Now if we substitute

$$H(\theta_{12}, \theta_{13}, \theta_{23}) = X$$

into (1), we get a distribution whose conditional distributions for θ_{12} given $(\theta_{13}, \theta_{23})$ and θ_{13} given $(\theta_{12}, \theta_{23})$ are univariate von Mises distributions but the distribution of θ_{23} given $(\theta_{12}, \theta_{13})$ has the Mardia-Gadsen distributional form. However if we use $X = A(\alpha, \beta, \gamma)$ then (β, γ) given α has a bivariate von Mises distribution (Rivest-Mardia type) with pdf proportional to

$$\exp\{\lambda_1 \cos\beta\cos\gamma + \lambda_2 \sin\beta\sin\gamma\}$$

but α given β and γ has the Watson distribution (marginal) with pdf proportional to

$$\exp\{\lambda_3 \cos^2\alpha\} \sin\alpha.$$

This link given by Habeck (2009) allows a faster sampling procedure since the Rivest-Mardia distribution can be reduced to a product of two independent von Mises distributions. We are investigating whether there exists another representation of $X = H(\theta, \phi, \psi)$ which may have, for example, the Fisher distribution for (θ, ϕ) conditional on a circular variable ψ whereas the circular variable ψ conditional on (θ, ϕ) may have a von Mises distribution!

2.3 The GM algorithm and Form

We now indicate how the matrix Fisher distribution has appeared in Bayesian alignments. Green and Mardia (2006) have aligned a pair of configurations using a full Bayesian aproach for unlabelled configurations in $\mathbb{R}^d$. Denote the j^{th} point in the x configuration by x_j where $j = 1, \ldots, m$. Similarly, y_k denotes the k^{th} point in the y configuration where $k = 1, \ldots, n$. Let A and τ denote the rotation matrix and translation vector to bring y into alignment with x. Furthermore denote prior distributions on these parameters by $p(A)$ and $p(\tau)$. We denote the prior for σ, parameterising noise in positions for x and y coordinates, by $p(\sigma)$. The joint posterior distribution for the model is

$$p(M, A, \tau, \sigma, x, y) \propto p(A)p(\tau)p(\sigma) \times \prod_{j,k:M_{jk}=1} \left(\kappa \frac{\phi(\{x_j - Ay_k - \tau\}/\sigma\sqrt{2})}{(\sigma\sqrt{2})^d} \right) \tag{15}$$

where $\phi(\cdot)$ is the standard normal probability density function and $\kappa > 0$ is a parameter representing the propensity of points to be matched. M is an unknown matrix for matching:

$$M_{jk} = \begin{cases} 1 & \text{if } x_j \text{ corresponds to } y_k, \\ 0 & \text{otherwise.} \end{cases}$$

Now for the labelled case, we have $m = n;\ j = k = 1, \ldots, m$ and $M = I$ so that we can rewrite (15), the joint posterior distribution for the model, as

$$p(A, \tau, \sigma, x, y) \propto p(A)p(\tau)p(\sigma)(\sigma)^{-md} \exp\left\{ -\frac{1}{\sigma^2} \sum_{j=1}^{m} ||x_j - Ay_j - \tau||^2 \right\}.$$

Hence we can now carry out the Bayesian inference on A, τ and σ following Green and Mardia (2006) for example. This gives a Bayesian perspective to matching labelled forms (cf. Habeck, 2009; Theobald and Wukkte, 2006).

3 Form Analysis and Bookstein-Type Coordinates

We indicate how Bookstein-type coordinates can be constructed for form analysis in three dimensions which are required in registering backbones (see, for example, Killian et al, 2007). For similarity shape, a set of Bookstein coordinates for three dimensions is given in Dryden and Mardia (1998, p.78). Let $X(k \times 3)$ be the configuration matrix with rows $x_i,\ i = 1, \ldots, k$. There are six degrees of freedom and from a naive point of view, it might be thought that the six degrees of freedom in x_1 and x_2 might be sufficient to determine the six degrees of freedom needed to specify Bookstein registration. However this is not the case as x_1 and x_2 are colinear, and planar information is required to specify the Bookstein registration. We first use the point x_1 as the origin so that

$$y_i = x_i - x_1, \quad i = 1, \ldots, k. \tag{16}$$

Then use y_2 (ie. x_2) to fix the colatitude and longitude of the points, ie. let (θ_2, ϕ_2) be the polar coordinates of y_2 (z-axis is the north pole). Then

$$u_i = R(\theta_2, \phi_2)y_i, \quad i = 1, \ldots, k \tag{17}$$

where

$$R(\theta, \phi) = \begin{pmatrix} \cos\theta\cos\phi & \cos\theta\sin\phi & -\sin\theta \\ -\sin\phi & \cos\phi & 0 \\ \sin\theta\cos\phi & \sin\theta\sin\phi & \cos\theta \end{pmatrix}.$$

Now rotate the new coordinates around the coordinates u_3 (or x_3), ie. if

$$u_3^T/||u_3|| = (\sin\theta_3 \sin\phi_3, \sin\theta_3 \cos\phi_3, \cos\theta_3).$$

Then the Bookstein-type coordinates for the form in the three dimensions (angle ϕ_3 with new x-axis) are

$$v_i = S(\phi_3)u_i, \quad i = 1, \ldots, k, \tag{18}$$

where

$$S(\phi) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & \sin\phi \\ 0 & -\sin\phi & \cos\phi \end{pmatrix},$$

and u_i are given by (15) and (16). In bioinformatics, these coordinates are termed bond-angle-torsion (BAT), see for example, Killian et al (2007) and we show elsewhere how these can be used in matching protein backbones.

For the two dimensional case, the first two new coordinates are $(0,0)$ and $(d,0)$ using x_1 and x_2, namely

$$u_i = S(\phi)y_i, \quad y_i = x_i - x_1, \quad i = 1, \ldots, n,$$

where now $S(\phi)$ is a 2×2 rotation matrix

$$S(\phi) = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix},$$

such that

$$u_2^T = (|x_2 - x_1|, 0) = (|y_2|, 0). \tag{19}$$

The construction extends to any dimenion.

4 Saddlepoint Approximations

We consider the simple example of calculating the normalizing constant for the von Mises distribution. The pdf of the von Mises distribution for zero mean is given by

$$\{c(\kappa)\}^{-1} \exp\{\kappa \cos\theta\}, \quad 0 < \theta < 2\pi, \tag{20}$$

where κ is the concentration parameter ($\kappa \geq 0$). The normalizing constant is complicated, namely, we have

$$c(\kappa) = 2\pi I_0(\kappa), \tag{21}$$

where $I_0(\cdot)$ is a Bessel function. This difficulty is more complex for the multivariate directional distributions on the torus, sphere, etc, see, for example, Mardia et. al. (2008). However, we can derive a saddlepoint approximation from Kume and Wood (2005) as a particular case of the Fisher-Bingham distribution. Here we give a direct derivation of their approximation for the von Mises case to get insight into their approach. We show that approximately

$$c(\kappa) \simeq (2\pi)^{\frac{1}{2}}(1+\kappa^2)^{-\frac{1}{4}} \exp\{(1+\kappa^2)^{\frac{1}{2}}\} = d(\kappa), \text{ say}. \tag{22}$$

Note that it is a well known fact for small κ and large κ that

$$c(0) = 2\pi, \quad c(\kappa) \simeq (2\pi)^{\frac{1}{2}}\kappa^{-\frac{1}{2}} \exp(\kappa).$$

Thus,

$$c(0)/d(0) \simeq 1.084, \quad c(\kappa)/d(\kappa) \simeq 1 \text{ for large } \kappa.$$

Hence the approximation covers at least the extreme range.

Consider $x \sim N(\kappa, 1), y \sim N(0,1)$ with $x = r\cos\theta, y = r\sin\theta$. Then, the conditional distribution of θ given $r = 1$ with pdf

$$f(\theta | r = 1) = \{c_0(\kappa)\}^{-1} \exp\{\kappa \cos\theta\},$$

where now

$$c_0(\kappa) = 2\pi g(1) \exp\{\frac{\kappa^2}{2} + \frac{1}{2}\} \tag{23}$$

and $g(1)$ is the value of the marginal pdf of r at $r = 1$. Now, r^2 has non-central χ^2_2 distribution with the non-centrality parameter $\lambda = \kappa^2/2$. We can now take the saddlepoint approximation for $g(r)$ to obtain $g(1)$ which then leads to (22). In fact, the saddlepoint approximation for the non-central χ^2_2 is given by (see, for example, Casella and Goutis, 1999)

$$g(1) \simeq (2\pi K''(t))^{-\frac{1}{2}} \exp\{K(t) - t\}, \tag{24}$$

where

$$t = -\frac{1}{2}(1+\kappa^2)^{\frac{1}{2}}, \;\; K(t) = \frac{2\lambda t}{1-2t} - \log(1-2t), \;\; K''(t) = \frac{4(1+2\lambda-2t)}{(1-2t)^3}.$$

On substituting $K(t)$ and $K''(t)$ into (28), we find that

$$g(1) \simeq \frac{(1-2t)^{\frac{1}{2}}}{2(2\pi)^{\frac{1}{2}}(1+\kappa^2-2t)^{\frac{1}{2}}} \exp\{-2t - \frac{\kappa^2}{2} - \frac{1}{2}\}. \tag{25}$$

Further, substituting $g(1)$ given by (25) into (23), we get $d(\kappa)$ given by (22). This has great potential for carrying out inference procedure for the multivariate directional distributions and is currently under investigation here and elsewhere. Another method for inference is using composite likelihoods (see, for example, Mardia et al, 2009).

5 Importance Sampling

Most of the common distributions in directional statistics belong to the canonical exponential family (when at least the “location parameter" is given). Thus the new methods to obtain MLE or normalizing constants for the exponential family play a significant role. Another feature of these models is that when the “association" parameters are zero then the marginals are easy to handle. The simplest examples are perhaps the sine and cosine bivariate von Mises models (Mardia et al, 2007). Insight into such a behaviour can be obtained by considering the well known bivariate normal distribution as shown below. The main idea is to use importance sampling as proposed by Geyer and Thompson (1992) and Geyer (1996). But here we follow the presentation of Green (1992).

Let θ be a true parameter vector in the exponential density

$$f_\theta(x) = \frac{1}{c(\theta)} \exp\{\theta^T t(x)\} \tag{26}$$

and let ψ be a suitable value of θ where $f_\psi(.)$ can be sampled. Then, we have

$$c(\theta) \simeq \frac{1}{n} \sum_{i=1}^{n} e^{(\theta-\psi)^T t(x_i)} \tag{27}$$

where $x_1, \ldots, x_n$ is a random sample from $f_\psi(.)$.

Let X be a bivariate normal with zero means, unit variances and correlation ρ. Here

$$c(\rho) = \int\int \exp\left\{-\frac{1}{2}(x_1^2 - 2\rho x_1 x_2 + x_2^2)/(1-\rho^2)\right\} dx_1 dx_2.$$

Here $\rho = 0$ is the case (two independent normal variables) where we can sample the distribution easily. Let

$$t_1(x) = -\frac{1}{2}(x_1^2 + x_2^2), \quad t_2(x) = x_1 x_2, \quad \theta_1 = 1/(1-\rho^2), \quad \theta_2 = \rho/(1-\rho^2). \tag{28}$$

Suppose $\psi_1 = 1, \ \ \psi_2 = 0$ (so that $\rho = 0$). Thus from (27), we have

$$c(\rho) \simeq \frac{1}{N}\sum_{i=1}^{N} \exp\left\{\left(\frac{1}{(1-\rho^2)} - 1\right) t_1(x_i) + \left(\frac{\rho}{1-\rho^2}\right) t_2(x_i)\right\} \tag{29}$$

where the first term has the coefficient $(\theta_1 - \psi_1)$ and the second $(\theta_2 - \psi_2)$. So now draw random samples from $x_1 \sim N(0,1)$ and $x_2 \sim N(0,1)$, both independent. Note that we know here that $c(\rho) = 2\pi(1-\rho^2)^{\frac{1}{2}}$. Hence the behaviour of $\hat{c}(\rho)$ can be studied. Also, we can obtain the MLE of ρ by this method. For some comments on general issues with this technique, see for example, Robert and Casella (1999, pp210–211). Here we have considered a "marginal approach" but when conditionals are easier then the Gibbs sampler is an alternative (see, for example, Mardia et al, 2007; Mardia et al, 2008). Incidentally, there are no problems in simulating wrapped normal distributions on the torus, for example (see, Kent and Mardia, 2009).

6 Conjugate Priors for Directional Distributions

6.1 Introduction

There has been renewed interest in directional Bayesian analysis since the paper of Mardia and El-Atoum (1976). One of the most recent papers on the topic is by Lennox et al (2009). Consider the von Mises distribution with p.d.f.

$$\{2\pi I_0(\kappa)\}^{-1} \exp\{\kappa \cos(\theta - \mu)\},$$

where $-\pi < \theta \leq \pi$, $-\pi < \mu \leq \pi$ and $\kappa \geq 0$, and $I_0(\kappa)$ is a Bessel function. Here μ is the mean direction and κ is the concentration (precision) parameter. It has been shown in Mardia and El-Atoum (1976) that for given κ, the conjugate prior for μ leads to the conditional distribution again as von Mises.

Guttorp and Lockhart (1988) have given the joint conjugate prior for μ and κ, and Mardia (2007) has considered a slight variant. However, the distribution for κ is not straightforward. Various suggestions have appeared; for example, take the prior for κ independently as a chi-square distribution, use the non-informative prior and so on.

6.2 Bivariate Distributions

The current interest in bivariate directional distributions has increased with their important applications in structural protein bioinformatics; Kent et al (2008) have recently given an overview. A general bivariate circular model, which we call the "full" bivariate von Mises (BVM) distribution, was introduced by Mardia (1975),

$$\begin{aligned} f(\theta,\phi) \propto \exp\big\{&\kappa_1 \cos(\theta-\mu) + \kappa_2 \cos(\phi-\nu) + \\ &\big[\cos(\theta-\mu), \ \ \sin(\theta-\mu)\big]\, A\, \big[\cos(\phi-\nu), \ \ \sin(\phi-\nu)\big]^T\big\}, \end{aligned} \tag{30}$$

where the angles $\theta, \phi \in (-\pi, \pi]$ lie on the torus, a square with opposite sides identified, and the matrix $A = (a_{ij})$ is 2×2. This model has eight parameters and allows for dependence between the two angles. It can be shown that the conjugate prior for (μ, ν) given κ_1, κ_2 and A leads to the posterior distribution for (μ, ν) of the same form as the full BVM. We have also obtained the normalizing constant for this general case to write down the full conjugate prior and the posterior (Mardia, 2009c) but the problem is more intricate as expected.

Recently the submodels of the full BVM with 5 parameters have been popular. We will mainly concentrate on the sine model where $a_{11} = a_{12} = a_{21} = 0, \quad a_{22} = \lambda$, that is the p.d.f. is given by

$$f(\theta, \phi) \propto \exp\{\kappa_1 \cos(\theta - \mu) + \kappa_2 \cos(\phi - \nu) + \lambda \sin(\theta - \mu) \sin(\phi - \nu)\} \tag{31}$$

where κ_1, κ_2 and λ are the precision parameters. Let $(\theta_i, \phi_i),\ i = 1, \ldots, n$, be a random sample from the sine model. Let us write

$$C_1 = \Sigma \cos\theta_i, \quad S_1 = \Sigma \sin\theta_i, \quad C_2 = \Sigma \cos\phi_i, \quad S_2 = \Sigma \sin\phi_i,$$
$$U_{11} = \Sigma \cos\theta_i \cos\phi_i, U_{21} = \Sigma \sin\theta_i \cos\phi_i, \quad U_{12} = \Sigma \cos\theta_i \sin\phi_i, U_{22} = \Sigma \sin\theta_i \sin\phi_i.$$

Suppose that the prior for (μ, ν) given the precision parameters is again the sine model with parameters $(\mu_0, \nu_0, \kappa_{01}, \kappa_{02}, \lambda_0)$. Then after some algebra it can be shown (Mardia, 2009c) that the posterior density for (μ, ν) is given by the full **BVM** distribution with the mean (μ_0^*, ν_0^*), concentration (κ_1^*, κ_2^*) and the matrix A as defined below.

$$\kappa_1^* \cos\mu_0^* = \kappa_{01} \cos\mu_0 + \kappa_1 C_1, \quad \kappa_1^* \sin\mu_0^* = \kappa_{01} \sin\mu_0 + \kappa_1 S_1,$$
$$\kappa_2^* \cos\nu_0^* = \kappa_{02} \cos\nu_0 + \kappa_2 C_2, \quad \kappa_2^* \sin\nu_0^* = \kappa_{02} \sin\nu_0 + \kappa_2 S_2$$

$$A = R(\mu_0^*) D R(\nu_0^*)^T, \quad R(\psi) = \begin{pmatrix} \cos\psi & \sin\psi \\ -\sin\psi & \cos\psi \end{pmatrix}, \quad D = (d_{ij}),$$

$$d_{11} = \lambda_0 \sin\mu_0 \sin\nu_0 + \lambda U_{22}, \quad d_{12} = -(\lambda_0 \sin\mu_0 \cos\nu_0 + \lambda U_{21}),$$
$$d_{21} = -(\lambda_0 \cos\mu_0 \sin\nu_0 + \lambda U_{12}), \quad d_{22} = \lambda_0 \cos\mu_0 \cos\nu_0 + \lambda U_{11}.$$

This result is in conflict with the result given by Lennox et al (2009) where the posterior density for (μ, ν) is claimed to be a sine distribution.

We can work in the same way with the cosine model (Mardia et al, 2007; Boomsma et al, 2008) but now the posterior distribution of the mean is another cosine submodel of the full BVM but with six parameters only. Various details will appear in Mardia (2009c).

6.3 The Multivariate von Mises Distribution

Mardia et al (2008) have given a multivariate sine model with its pdf of $\boldsymbol{\theta}^T = (\theta_1, \ldots, \theta_p)$ as

$$\{T(\boldsymbol{\kappa}, \boldsymbol{\Lambda})\}^{-1} \exp\{\boldsymbol{\kappa}^T c(\boldsymbol{\theta}, \boldsymbol{\mu}) + \frac{1}{2} s(\boldsymbol{\theta}, \boldsymbol{\mu})^T \Lambda \quad s(\boldsymbol{\theta}, \boldsymbol{\mu})\}, \tag{32}$$

where $-\pi < \theta_i \leq \pi, -\pi < \mu_i \leq \pi, \kappa_i \geq 0, -\infty < \lambda_{ij} < \infty, \boldsymbol{\kappa}^T = (\kappa_1, \ldots, \kappa_p)$,

$$c(\boldsymbol{\theta}, \boldsymbol{\mu})^T = (\cos(\theta_1 - \mu_1), \ldots, \cos(\theta_p - \mu_p)), \quad s(\boldsymbol{\theta}, \boldsymbol{\mu})^T = (\sin(\theta_1 - \mu_1), \ldots, \sin(\theta_p - \mu_p)),$$

and $(\boldsymbol{\Lambda})_{ij} = \lambda_{ij} = \lambda_{ji}, \quad i \neq j, \quad \lambda_{ii} = 0$, with $\{T(\boldsymbol{\kappa}, \boldsymbol{\Lambda})\}^{-1}$ a normalizing constant. We call this the multivariate von Mises density. Note that for $p = 1$, this is a univariate von Mises density and for $p = 2$, this density corresponds to the bivariate sine model. Further, for $p > 2$, the normalizing constant is not known in any closed form. For large concentrations we have ($\boldsymbol{\mu} = \mathbf{0}$ without any loss of generality)

$$\boldsymbol{\theta} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma}^{-1}), \quad \text{where } (\boldsymbol{\Sigma}^{-1})_{ii} = \kappa_i, \ (\boldsymbol{\Sigma}^{-1})_{ij} = -\lambda_{ij}, \ i \neq j.$$

Using the conjugate prior for the mean vector $\boldsymbol{\mu}$ for given $\boldsymbol{\kappa}$ and $\boldsymbol{\Lambda}$, then the posterior density of $\boldsymbol{\mu}$ is found to belong to an extension of (32) given in Mardia and Patrangenaru (2005). For $\boldsymbol{\kappa}, \boldsymbol{\Lambda}$ we can use the independent prior distribution as Wishart for Γ where $(\Gamma)_{ii} = \kappa_i$ and $(\Gamma)_{ij} = -\lambda_{ij}$, following the proposal for $p = 2$ by Lennox et al (2009). Again, full details of the results in this section will appear elsewhere (Mardia, 2009c).

7 Discussion

We have given here a few pointers to investigate some new “complex" problems in directional statistics and shape analysis. This style follows in the spirit set by LASR papers such as Mardia (2007), and Kent et al (2008). There has been a lot of activity in statistical protein bioinformatics and a full review will appear elsewhere (Mardia, 2009b). Boomsma et al (2008) and Frellsen et al (2009) are some of our examples of interdisciplinary work in the area but the field is moving fast with activities at various other centres. We hope that the field will continue to grow at the current speed!

Acknowledgements

I would like to thank Mike Gilson, Peter Green, Thomas Hamelryck, Peter Jupp, John Kent, Alfred Kume, Charles Taylor and Zhengzheng Zhang for helpful discussions.

References

Boomsma, W., Mardia, K.V., Taylor, C.C., Ferkinghoff-Borg, J., Krogh, A. and Hamelryck, T. (2008). A generative, probabilistic model of local protein structure. *PNAS*, **105**, pp8932–8937.

Casella, G. and Goutis, C. (1999). Explaining the saddlepoint approximations. *American Statistician*, **53**, pp216–224.

Downs, T.D. (1972). Orientation statistics. *Biometrika*, **59**, pp665–676.

Dryden, I.L. and Mardia, K.V. (1998). *Statistical Shape Analysis*. Wiley.

Fisher, N.T., Lewis, T.L, and Embleton, B.J.J. (1987). *Statistical Spherical Analysis of Spherical Data*. Cambridge University Press, Cambridge.

Frellsen, J., Moltke, I., Thiim, M., Mardia, K.V., Ferkinghoff-Borg, J. and Hamelryck, T. (2009). A probabilistic model of local RNA 3-D structure. *PLoS Computational Biology*, in press.

Geyer, C.J. and Thompson, E.A (1992). Constrained Monte Carlo maximum likelihood in dependent data. *J.Roy.Statist. Soc. B*, **54**, pp657–699.

Geyer, C.J. (1996). Estimation and optimization of functions. In Markov Chain Monte Carlo in Pratice, Editors Gilks, W.R., Richardson, S and Spiegelhalter, D.J. *Chapman and Hall, London*, pp241–258.

Green, P.J. (1992). Discussion to Geyer and Thompson (above). *J.Roy.Statist. Soc. B*, **54**, pp683-684.

Green, P.J. and Mardia, K.V. (2006). Bayesian alignment using hierarchical models, with applications in protein bioinformatics. *Biometrika*, **93**, pp235–254.

Guttorp, P. and Lockhart, R.A. (1988). Finding the location of a signal: a bayesian analysis. *J. Amer. Statist. Soc.*, **83**, pp322-330.

Habeck (2009). Generation of three dimensional random rotations in fitting and matching problems. *Computational Statistics*, online.

Kent, J.T. and Mardia, K.V. (2009). Principal component analysis for the wrapped normal torus model. In this proceedings.

Kent, J.T., Mardia, K.V. and Taylor, C.C. (2008). Modelling strategies for bivariate circular data. In: The Art and Science of Statistical Bioinformatics. Leeds, Leeds University Press.

Khatri, C.G. and Mardia, K.V. (1977). The von Mises-Fisher matrix distribution in orientation statistics. *J. Roy. Statist. Soc. Ser. B*, **39**, pp95–106.

Killian, B.J., Kravitz, J.Y., and Gilson, M.K. (2007). Extraction of configurational entropy from molecular simulations via an expansion approximation. *J. Chem. Physics*, **127**, 024107.

Kume, A. and Wood, A.T. (2005). Saddlepoint approximations for the Bingham and Fisher-Bingham normalising constants. *Biometrika*, **92**, pp465–476.

Lennox, K.P., Dahl, D.B., Vannucci, D.B. and Tsai, J.W. (2009). Density estimation for protein conformation angles using a bivariate von Mises distribution and Bayesian nonparametrics, *J. Amer. Statist. Soc.*, **104**, pp586–596.

León, C.A., Masse, J.-C., Rivest, L.-P. (2006). Statistical model for random rotations. *J. Mult. Analysis*, **97**, pp412–430.

Mardia, K.V. (1975). Statistics of directional data (with discussion). *J. Royal Statist. Soc. Series B*, **37**, pp349–393.

Mardia, K.V. (2007). On some recent advancements in applied shape analysis and directional statistics. In S. Barber, P.D. Baxter, & K.V.Mardia (eds), *Systems Biology & Statistical Bioinformatics*, pp.9–17 . Leeds, Leeds University Press.

Mardia, K.V. (2008). Holistic statistics and contemporary life sciences. In LASR Proceedings *The Art and Science of Statistical Bioinformatics*, S. Barber, P.D. Baxter, A. Gusnanto, and K.V. Mardia (Eds.), pp9–17, Leeds University Press.

Mardia, K.V. (2009a). Jacobians under constraints and statistical bioinformatics. *ISI Platinum Jubilee/S.N. Roy*, Volume 5, Chapter 4, World Scientific, Singapore.

Mardia, K.V. (2009b). Statistical challenges in protein bioinformatics, in preparation.

Mardia, K.V. (2009c). Bayesian directional statistics. Technical Report, Department of Statistics, University of Leeds.

Mardia, K.V. and El-Atoum, S.A.M. (1976). Bayesian inference for the von Mises-Fisher distribution. *Biometrika*, **63**, pp203–205.

Mardia, K.V., Hughes, G., Taylor, C.C. and Singh, H. (2008). A multivariate von Mises distribution with applications to bioinformatics. *Can. J. of Statist.*, **36**, pp99–109.

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

Mardia, K.V., Kent, J.T., Hughes, G., and Taylor, C.C. (2009). Maximum likelihood estimation using composite likelihoods for closed exponential families. *Biometrika*, in press.

Mardia, K.V. and Patrangenaru, V. (2005). Directions and projective shapes. *Ann. Statist.*, **33**, pp1666–1699.

Mardia, K.V., Taylor, C.C., and Subramaniam, G.K. (2007) Protein bioinformatics and mixtures of bivariate von Mises distributions for angular data. *Biometrics*, **63**, pp505–512.

Miles, R.E. (1965). On random rotations in R^3. *Biometrika*, **52**, pp636–639.

Muirhead, R.J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York.

Robert, C.P. and Casella, G. (1999). *Monte Carlo Statistical Methods*. Springer Verlag, New York.

Singh, H., Hnizdo, V., and Demchuk, E. (2002). Probabilistic model for two dependent circular variables. *Biometrika*, **89**, pp719–723.

Theobald, D.L. and Wuttke, D.S. (2006). Empirical Bayes hierarchical models for regularizing maximum likelihood estimation in the matrix Gaussian Procrustes problem. *Proceedings of the National Academy of Sciences*, **103**, pp18521–18527.