A Multivariate von Mises Distribution with Applications to Bioinformatics

Kanti V. Mardia, Gareth Hughes & Charles C. Taylor
University of Leeds, Leeds, LS2 9JT

Harshinder Singh(*)

(*) formerly at West Virginia University, Morgantown, WV 26506; now deceased

Abstract

Motivated by problems of modeling torsional angles in molecules, Singh et al. (2002) proposed a bivariate circular model which is a natural torus analogue of the bivariate normal distribution and a natural extension of the univariate von Mises distribution to the bivariate case. In this paper, we propose a multivariate extension of the bivariate model of Singh et al. The results about conditional distributions are derived and the shapes of marginal distributions have been investigated for a special case. Procedures for the estimation of parameters of the proposed distribution include the method of moments, and pseudolikelihood; the efficiency of the latter is investigated in two and three dimensions. The methods are applied to real protein data of conformational angles.

KEYWORDS: Bias; Bessel functions; Circular mean; Circular variance; Gamma turns; Proteomics; Pseudolikelihood.

1 Introduction

Circular distributions are important in modeling angular variables in biology, astronomy, meteorology, earth sciences and several other areas (see Mardia & Jupp, 1999). Recently Demchuk and Singh (2001) emphasized the importance of circular distributions in molecular sciences for modeling torsional angles in molecules. The von Mises distribution is most prominent among the univariate circular distributions and its probability density function is given by $f(\theta) = \{2\pi I_0(\kappa)\}^{-1} \exp\{\kappa \cos(\theta - \mu)\}$, $-\pi \leq \theta < \pi$, where $\kappa \geq 0$ is the concentration parameter, μ is the mean angle and $I_p(\kappa)$ is the modified Bessel function of order p. The von Mises distribution is a natural circular analogue of the univariate normal distribution to which it reduces when the fluctuations in the circular variable are small. Other circular distributions studied in the literature include the wrapped normal, wrapped Cauchy, wrapped Poisson, cardoid, projected normal and a generalization of the von Mises distribution proposed by Cox (1975), which was investigated by

Yfantis & Borgmann (1982). There are a few bivariate circular distributions discussed in the literature, including a class of bivariate circular distributions proposed by Mardia (1975a, 1975b), a submodel of this class discussed by Jupp & Mardia (1980), another submodel considered by Rivest (1988), subsets of which were investiagted by Subramaniam (2005) and Mardia *et al.* (2007). A bivariate wrapped normal distribution was studied by Johnson & Wehrly (1977) and a wrapped multivariate normal distribution was discussed by Baba (1981). The multivariate wrapped normal has all marginals wrapped normal as well as the marginal bivariate distributions wrapped normal, and thus has a theoretical advantage. Another disadvantage of the bivariate wrapped normal is that if we wrap $c\theta_1, c\theta_2$ instead of θ_1, θ_2 then the two wrapped distributions are different though the original correlation is invariant. Furthermore, the maximum likelihood estimators of the parameters, even for the univariate case are not computationally feasible as is well known, and one has to resort to selecting some moment estimators which are not clear-cut for the bivariate parameters. Thus, it is not possible to carry out hypothesis tests in a meaningful way. The objective of this work is to introduce and study a natural multivariate generalization of the von Mises distribution.

Motivated by problems in molecular sciences, Singh *et al.* (2002) proposed a bivariate circular distribution which is a natural torus analogue of the bivariate normal distribution. The general form is proposed in Mardia & Patrangenaru (2005) but this particular case has some attractive properties among the "minimal" redundancy class. Let Θ_1 and Θ_2 be two circular random variables. The proposed probability density function is of the form

$$f(\theta_1, \theta_2) = \frac{\exp\{\kappa_1 \cos(\theta_1 - \mu_1) + \kappa_2 \cos(\theta_2 - \mu_2) + \lambda_{12} \sin(\theta_1 - \mu_1) \sin(\theta_2 - \mu_2)\}}{T(\kappa_1, \kappa_2, \lambda_{12})}, \tag{1}$$

for $-\pi \leq \theta_1, \theta_2 < \pi$, where κ_1, $\kappa_2 \geq 0$, $-\infty < \lambda_{12} < \infty$, $-\pi \leq \mu_1, \mu_2 < \pi$, and $T(\cdot)$ is a normalization constant. This probability density function is constructed from the bivariate normal distribution by replacing the quadratic and linear terms by their natural circular analogues. It can also be obtained from Rivest (1988) by taking one of the parameters as zero. This distribution is approximately bivariate normal when the fluctuations in the circular variables are small. The parameter λ_{12} is a measure of dependence between θ_1 and θ_2. In the bivariate case, there is only one such parameter, so we will simply use λ $(= \lambda_{12})$. If λ=0, then Θ_1 and Θ_2 are independent with each having univariate von Mises distribution. This distribution has a natural generalization for allowing multiple modes in the marginal distributions which is obtained by replacing $(\theta_i - \mu_i)$ by $l_i(\theta_i - \mu_i), i = 1, 2$, in (1) where l_1 and l_2 are positive integers.

Singh *et al.* (2002) obtained an expression for the normalization constant $T(\cdot)$ in (1) as

$$T(\kappa_1, \kappa_2, \lambda) = 4\pi^2 \sum_{m=0}^{\infty} \binom{2m}{m} \left(\frac{\lambda}{2}\right)^{2m} \kappa_1^{-m} I_m(\kappa_1) \kappa_2^{-m} I_m(\kappa_2). \tag{2}$$

They also showed that the conditional distributions of the above distribution are von Mises and marginal distributions are symmetric around their circular means. The marginal distributions are either unimodal or bimodal and they established conditions on the parameters which assure a specific shape. For the unimodal case, the marginals are approximately von Mises when the concentration parameters are large.

In this paper we propose a multivariate version of the model (1) which could be useful for jointly modeling several circular variables. The model provides a natural multivariate extension of the von Mises distribution and thus we call the model the multivariate von Mises distribution. In Section 2, we propose the multivariate circular model and discuss its properties. The one dimensional conditional distributions are von Mises and the bivariate conditional distributions belong to the above stated model (1). The shapes of the marginal distributions are investigated for the three variable case for many configurations of parameters and are seen to be unimodal or bimodal symmetrical distributions. If the fluctuations in the circular variables are very small, the model is approximately the multivariate normal distribution. In Section 3, we discuss three alternative procedures, namely, maximum pseudolikelihood method, the method of moments and a maximum likelihood method for estimating the parameters of the proposed multivariate distribution. Since the proposed distribution belongs to the natural exponential family, the maximum likelihood estimators and the moment estimators match for an important case. In Sections 4 and 5, we compare the performance of the maximum pseudolikelihood estimators with maximum likelihood estimators for the bivariate and trivariate von Mises distributions, repectively. In Section 6 we apply the methods to some real data from the field of proteomics, and we conclude with a discussion.

2 A Multivariate von Mises Distribution

We define a multivariate angular distribution which is an extension of Singh et al (2002) as follows. The probability density function of $\mathbf{\Theta}^T = (\Theta_1, \Theta_2, \ldots, \Theta_p)$ is given by

$$\{T(\boldsymbol{\kappa}, \mathbf{\Lambda})\}^{-1} \exp\{\boldsymbol{\kappa}^T c(\boldsymbol{\theta}, \boldsymbol{\mu}) + \frac{1}{2} s(\boldsymbol{\theta}, \boldsymbol{\mu})^T \Lambda \ s(\boldsymbol{\theta}, \boldsymbol{\mu})\}, \tag{3}$$

where $-\pi < \theta_i \leq \pi, -\pi < \mu_i \leq \pi, \kappa_i \geq 0, -\infty < \lambda_{ij} < \infty$,

$$\begin{aligned} c(\boldsymbol{\theta}, \boldsymbol{\mu})^T &= (\cos(\theta_1 - \mu_1), \cos(\theta_2 - \mu_2), \ldots, \cos(\theta_p - \mu_p)) \\ s(\boldsymbol{\theta}, \boldsymbol{\mu})^T &= (\sin(\theta_1 - \mu_1), \sin(\theta_2 - \mu_2), \ldots, \sin(\theta_p - \mu_p)) \end{aligned}$$

and

$$(\mathbf{\Lambda})_{ij} = \lambda_{ij} = \lambda_{ji}, \qquad i \neq j, \qquad \lambda_{ii} = 0,$$

with $\{T(\boldsymbol{\kappa}, \mathbf{\Lambda})\}^{-1}$ a normalizing constant (which is unknown in any explicit form for $p > 2$) so that (3) is a probability density function. We call this the multivariate von Mises density and denote it by $\mathbf{\Theta} \sim M_p(\boldsymbol{\mu}, \boldsymbol{\kappa}, \mathbf{\Lambda})$. We note that for $p = 1$, this is a univariate von Mises density and for $p = 2$, this density corresponds to the bivariate model (1) of Singh et al. (2002). Now we investigate the properties of the proposed multivariate von Mises distribution.

Without loss of generality, take $\boldsymbol{\mu} = \mathbf{0}$. We observe that for large concentrations in the circular variables, we have

$$\mathbf{\Theta} = (\Theta_1, \Theta_2, \ldots, \Theta_p)^T \sim N_p(\mathbf{0}, \mathbf{\Sigma}^{-1}), \quad \text{where } (\mathbf{\Sigma}^{-1})_{ii} = \kappa_i,\ (\mathbf{\Sigma}^{-1})_{ij} = -\lambda_{ij},\ i \neq j \tag{4}$$

and $N_p(\boldsymbol{\mu}, \boldsymbol{A})$ denotes a multivariate normal distribution with mean $\boldsymbol{\mu}$ and dispersion matrix $\boldsymbol{A}$.

The results about the conditional distributions of the proposed distribution are given in the following theorem which can be proved easily.

Theorem 1. *Let $\mathbf{\Theta}$ have the p-variate von Mises probability density function (3). Then, the conditional distribution of $\Theta_1, \Theta_2, \ldots, \Theta_r$ given $\Theta_j = \theta_j, j = r+1, r+2, \ldots, p$ is an extended r-variate von Mises distribution. In particular all univariate conditional distributions are von Mises.*

More generally, (but still taking $\boldsymbol{\mu} = \mathbf{0}$) we can obtain the conditional distribution of $\Theta_1, \Theta_2, \ldots, \Theta_r$ given $\theta_{r+1}, \theta_{r+2}, \ldots, \theta_p$ using the same approach as in the proof of Theorem 1. We obtain $f_{\text{cond}} = f(\Theta_1, \ldots, \Theta_r | \theta_{r+1}, \ldots, \theta_p)$ as

$$f_{\text{cond}} \propto \exp\left\{ \sum_{j=1}^{r} \kappa_j \cos\theta_j + \sum_{j=1}^{r-1} \sum_{l=j+1}^{r} \lambda_{jl} \sin\theta_j \sin\theta_l + \sum_{j=1}^{r} c_j \sin\theta_j \right\},$$

where c_j is constant with respect to $\theta_1, \ldots, \theta_r$, for $j = 1, \ldots, r$. Writing $\kappa_j = a_j \cos\nu_j$ and $c_j = a_j \sin\nu_j$ gives that

$$f_{\text{cond}} \propto \exp\left\{ \sum_{j=1}^{r} a_j \cos(\theta_j - \nu_j) + \sum_{j=1}^{r-1} \sum_{l=j+1}^{r} \lambda_{jl} \sin\theta_j \sin\theta_l \right\}.$$

Next writing $\phi_j = \theta_j - \nu_j$, so that $\theta_j = \phi_j + \nu_j$, gives

$$f_{\text{cond}} \propto \exp\left\{ \sum_{j=1}^{r} a_j \cos\phi_j + \sum_{j=1}^{r-1} \sum_{l=j+1}^{r} \lambda_{jl} \sin(\phi_j + \nu_j) \sin(\phi_l + \nu_l) \right\}.$$

Expanding each sine term and their product gives, for constants a_{jl}, b_{jl}, c_{jl} and d_{jl},

$$\begin{aligned} f_{\text{cond}} \propto \exp\Bigg\{ & \sum_{j=1}^{r} a_j \cos\phi_j + \sum_{j=1}^{r-1} \sum_{l=j+1}^{r} \big[a_{jl} \cos\phi_j \cos\phi_l \\ & + b_{jl} \sin\phi_j \sin\phi_l + c_{jl} \cos\phi_j \sin\phi_l + d_{jl} \sin\phi_j \cos\phi_l \big] \Bigg\}. \end{aligned} \tag{5}$$

Equation (5) is the density of Mardia & Patragenaru (2005) with $b_s = 0$ for all s.

For the bivariate von Mises distribution, Singh *et al.* (2002) proved that marginal distributions are symmetric around their circular means and are either unimodal or bimodal. When $\mathbf{\Theta}$ is distributed according to (3) it is clear that $\mathbf{\Theta}$ and $-\mathbf{\Theta}$ have the same distribution. This implies that all marginal distributions are symmetric. For the trivariate von Mises density, it does not appear possible to obtain an analytic expression for the univariate marginal of Θ_1 from this bivariate density, since it involves the integral of a Bessel function applied to a function of θ_1 and θ_2. However, to investigate the shape of the marginal density, we plotted the function $f_1(\theta_1)$ for many configurations of parameters. We observed in various numerical experiments that, in all cases, the density is either unimodal (quite similar to a von Mises distribution) or a bimodal symmetric distribution.

3 Approaches for Inference

Since we do not have an explicit expression for the normalizing constant, the maximum likelihood estimator has to work through a numerical optimization method where the normalizing constant is calculated at each stage for the iterative solution in the parameter values. We propose alternative strategies.

3.1 Method of Moments

Illustrating first for $p = 2$, we have $\hat{\mu}_1 = \overline{x}_{01}, \hat{\mu}_2 = \overline{x}_{02}$, where $\overline{x}_{01}$ and $\overline{x}_{02}$ are the circular means of the marginal distributions of Θ_1 and Θ_2 respectively. Next κ_1, κ_2 and λ_{12} are obtained from the following equations using an alternating algorithm. Namely

$$\nu_1 = \mathsf{E}[\cos(\Theta_1 - \mu_1)], \tag{6}$$

$$\nu_2 = \mathsf{E}[\cos(\Theta_2 - \mu_2)], \tag{7}$$

$$\nu_{12} = \mathsf{E}[\sin(\Theta_1 - \mu_1)(\sin(\Theta_2 - \mu_2)]. \tag{8}$$

The moment estimators of the above parameters are respectively given by

$$\overline{R}_1 = \frac{1}{n}\sum_{r=1}^{n}\cos(\theta_{1r} - \overline{x}_{01}), \qquad \overline{R}_2 = \frac{1}{n}\sum_{r=1}^{n}\cos(\theta_{2r} - \overline{x}_{02}),$$

and

$$\overline{S}_{12} = \frac{1}{n}\sum_{r=1}^{n}\sin(\theta_{1r} - \overline{x}_{01})\sin(\theta_{2r} - \overline{x}_{02}).$$

Expressions for the right hand side of (6) and (7) have been derived in Singh *et al.* (2002). Indeed, the RHS of equation (6) is given by $(1/T(\kappa_1, \kappa_2, \lambda_{12}))\ \partial T(\kappa_1, \kappa_2, \lambda_{12})/\partial\kappa_1$ where $T(\cdot)$ is given by (2). The RHS of equation (7) is similarly obtained. The expression for the RHS of Equation (8) can be shown to be $(1/T(\kappa_1, \kappa_2, \lambda_{12}))\partial T(\kappa_1, \kappa_2, \lambda_{12})/\partial\lambda_{12}$ which can be rewritten as

$$\sum_{m=1}^{\infty} m\binom{2m}{m}\left(\frac{\lambda_{12}^2}{\kappa_1\kappa_2}\right)^m I_m(\kappa_1)I_m(\kappa_2)\,/(\lambda_{12}T(\kappa_1, \kappa_2, \lambda_{12}))\,. \tag{9}$$

The algorithm may start with $\hat{\kappa}_1 = A_1^{-1}(\overline{R}_1), \hat{\kappa}_2 = A_1^{-1}(\overline{R}_2)$, where $A_\nu(y) = I_\nu(y)/I_0(y)$. The value of $\hat{\lambda}_{12}$ is obtained from equation (8) using these values of $\hat{\kappa}_1$ and $\hat{\kappa}_2$ through (9). The values of $\hat{\kappa}_2$ and $\hat{\lambda}_{12}$ are used in (6) to obtain new $\hat{\kappa}_1$. Using this new value of $\hat{\kappa}_1$ and the value of $\hat{\lambda}_{12}$ to obtain the new value of $\hat{\kappa}_2$ is obtained using (7). The procedure is continued till the convergence to the solution is achieved.

The extension to higher dimensions could be through all univariate and bivariate conditional distributions. In particular, for highly concentrated data, the covariance matrix in Equation (4) can be equated to the corresponding variables in the multivariate von Mises distribution. This gives moment estimators of $\boldsymbol{\Sigma}$ as $\hat{\boldsymbol{\Sigma}} = (\overline{S}_{ij})$ with

$$\overline{S}_{ii} = 2(1 - \overline{R}_i), \qquad \overline{S}_{ij} = \frac{1}{n}\sum_{r=1}^{n}\sin(\theta_{ir} - \overline{x}_{0i})\ \sin(\theta_{jr} - \overline{x}_{0j})$$

The interesting point is that $\overline{S}$ is the standard covariance matrix, and it should be interpreted as such. A difficulty may arise when $\boldsymbol{S}$ is not positive definite, but, in general, it will be positive definite if the diagonal terms of $\boldsymbol{\Sigma}$ are estimated by $\overline{S}_{ii} = \sum_r \sin^2\theta_{ir}/n$.

3.2 Maximum Pseudolikelihood Method

Define the pseudolikelihood (Besag, 1975), based on a random sample of n observations of $\boldsymbol{\theta} = (\Theta_1, \Theta_2, \ldots, \Theta_p)^T$, by

$$\mathrm{PL} = \prod_{j=1}^{p}\prod_{i=1}^{n} g_j(\Theta_{ji}|(\Theta_{1i}, \ldots, \Theta_{j-1,i}, \Theta_{j+1,i}, \ldots, \Theta_{pi}); \boldsymbol{q})$$

where $g_j(\cdot|\cdots;\boldsymbol{q})$ is the conditional distribution whose parameters will depend on j and $\boldsymbol{q}$ is an unknown parameter vector of length r. For the p-variate von Mises distribution we therefore have

$$\mathrm{PL} = (2\pi)^{-pn}\prod_{j=1}^{p}\prod_{i=1}^{n}\left[I_0(\kappa_{j\cdot\mathrm{rest}}^{(i)})\right]^{-1}\exp\left\{\kappa_{j\cdot\mathrm{rest}}^{(i)}\cos(\theta_{ji} - \mu_{j\cdot\mathrm{rest}}^{(i)})\right\}$$

where, for the ith observation, we have the coefficients of the conditional distributions which are given as functions of the parameters in the model:

$$\begin{aligned}
\mu_{j\cdot\mathrm{rest}}^{(i)} &= \mu_j + \tan^{-1}\left\{\left[\sum_{l\neq j}\lambda_{jl}\sin(\theta_{li} - \mu_l)\right]/\kappa_j\right\}, \\
\kappa_{j\cdot\mathrm{rest}}^{(i)} &= \left\{\kappa_j^2 + \left[\sum_{l\neq j}\lambda_{jl}\sin(\theta_{li} - \mu_l)\right]^2\right\}^{1/2}.
\end{aligned}$$

4 Pseudolikelihood and moments for the bivariate case

The method of moments estimators of the three parameters of (1) (with zero means) are found by solving the following three moments equations which come from Section 3.1.

$$\frac{1}{n}\sum_{i=1}^{n}\cos\theta_{1i} = \frac{\sum_{m=0}^{\infty}\binom{2m}{m}(\frac{\lambda_{12}^2}{\kappa_1\kappa_2})^m I_{m+1}(\kappa_1)I_m(\kappa_2)}{\sum_{m=0}^{\infty}\binom{2m}{m}(\frac{\lambda_{12}^2}{\kappa_1\kappa_2})^m I_m(\kappa_1)I_m(\kappa_2)}, \tag{10}$$

$$\frac{1}{n}\sum_{i=1}^{n}\cos\theta_{2i} = \frac{\sum_{m=0}^{\infty}\binom{2m}{m}(\frac{\lambda_{12}^2}{\kappa_1\kappa_2})^m I_{m+1}(\kappa_2)I_m(\kappa_1)}{\sum_{m=0}^{\infty}\binom{2m}{m}(\frac{\lambda_{12}^2}{\kappa_1\kappa_2})^m I_m(\kappa_1)I_m(\kappa_2)}, \tag{11}$$

$$\frac{1}{n}\sum_{i=1}^{n}\sin\theta_{1i}\sin\theta_{2i} = \frac{\sum_{m=1}^{\infty}m\binom{2m}{m}(\frac{\lambda_{12}^2}{\kappa_1\kappa_2})^m I_m(\kappa_1)I_m(\kappa_2)}{\lambda_{12}\sum_{m=0}^{\infty}\binom{2m}{m}(\frac{\lambda_{12}^2}{\kappa_1\kappa_2})^m I_m(\kappa_1)I_m(\kappa_2)}. \tag{12}$$

The usual maximum likelihood estimators are obtained by maximizing the likelihood function given by

$$(1/T(\kappa_1, \kappa_2, \lambda_{12}))^n\exp\left\{\kappa_1\sum_{i=1}^{n}\cos\theta_{1i} + \kappa_2\sum_{i=1}^{n}\cos\theta_{2i} + 2\lambda_{12}\sum_{i=1}^{n}\sin\theta_{1i}\sin\theta_{2i}\right\}.$$

It is easily seen that the likelihood equations match with the moment equations (10)–(12) given above, which is to be expected since this is an exponential family. Thus, the maximum likelihood estimators and the moment estimators coincide for the bivariate model (1).

We have used a numerical approach in order to assess the efficiency of the pseudolikelihood for the bivariate von Mises distribution. We considered the case in which $\mu_1 = \mu_2 = 0$ and examined how the efficiency depends on λ and $\kappa_1 = \kappa_2 = \kappa$. The approach is through comparison of information matrices — for the full likelihood, and for the partial likelihood — and is given in more detail in Mardia *et al.* (2007).

We computed the efficiency for various pairs of values $(\kappa,\ |\lambda|)$. For $\lambda = 0$ the efficiency is unity, since in this case $f(\theta_1, \theta_2) = f(\theta_1|\theta_2)f(\theta_2|\theta_1)$. We observed that efficiency is high for small $|\lambda|$. We know that the pseudolikelihood is fully efficient for the bivariate normal distribution with known $\mu_1 = \mu_2 = 0$ and unknown parameters σ and ρ. Since the bivariate von Mises distribution tends to a normal distribution for large κ values (Singh *at al.*, 2002), the efficiency should tend to unity as $\kappa \to \infty$. From Singh *at al.* (2002), for the bivariate case with $\mu_1 = \mu_2 = 0$ and $\kappa_1 = \kappa_2 = \kappa$, the corresponding parameters of the approximating normal distribution are $\sigma^2 = \kappa/(\kappa^2 - \lambda^2)$ and $\rho = \lambda/\kappa$. In order that $\sigma^2 > 0$ we require $\kappa > |\lambda|$. The aforementioned improvement in efficiency as κ increases should therefore only be expected to occur, for fixed λ, in the case that $\kappa > |\lambda|$. This was indeed observed, and the efficiency is greater than 0.9 for all all considered pairs such that $\kappa > \lambda$. For further details, see Mardia *et al.* (2007).

5 Parameter Estimation for the trivariate case

In this section we investigate data simulation and parameter estimation for the trivariate von Mises distribution. A Gibbs sampling technique (Mardia *et al.*, 2007) is used to simulate data, and the properties of pseudolikelihood estimates are investigated and compared with maximum likelihood estimates.

The pseudolikelihood for the trivariate von Mises distribution is given by

$$(2\pi)^{-3n} \prod_{i=1}^{n} \left[I_0(\kappa^{(i)}_{1\cdot 2,3}) I_0(\kappa^{(i)}_{2\cdot 1,3}) I_0(\kappa^{(i)}_{3\cdot 1,2}) \right]^{-1} \exp\left\{ \kappa^{(i)}_{1\cdot 2,3} \cos(\theta_{1i} - \mu^{(i)}_{1\cdot 2,3}) \right.$$
$$\left. + \kappa^{(i)}_{2\cdot 1,3} \cos(\theta_{2i} - \mu^{(i)}_{2\cdot 1,3}) \kappa^{(i)}_{3\cdot 1,2} \cos(\theta_{3i} - \mu^{(i)}_{3\cdot 1,2}) \right\} \tag{13}$$

where

$$\begin{aligned}
\mu^{(i)}_{1\cdot 2,3} &= \mu_1 + \tan^{-1}\{[\lambda_{12}\sin(\theta_{2i} - \mu_2) + \lambda_{13}\sin(\theta_{3i} - \mu_3)]/\kappa_1\} \\
\kappa^{(i)}_{1\cdot 2,3} &= \surd\{\kappa_1^2 + [\lambda_{12}\sin(\theta_{2i} - \mu_2) + \lambda_{13}\sin(\theta_{3i} - \mu_3)]^2\}
\end{aligned}$$

and analogously for $\mu^{(i)}_{2\cdot 1,3}$, $\kappa^{(i)}_{2\cdot 1,3}$, $\mu^{(i)}_{3\cdot 1,2}$ and $\kappa^{(i)}_{3\cdot 1,2}$.

For chosen parameters κ_1, κ_2, κ_3, λ_{12}, λ_{13} and λ_{23}, the following Gibbs sampling method will be used in order to simulate variates from the trivariate von Mises distribution. For the simulation method, the mean direction vector will be set to zero. For subsequent estimation, mean directions will be added to (zero-mean) simulated data.

- Firstly, two independent von Mises distributed vectors of length n_1 are simulated with concentration parameters κ_1 and κ_2 respectively. These are the initial vectors representing θ_1 and θ_2 respectively.
- Next, using the parameters κ_3, λ_{13} and λ_{23}, a vector of θ_3 values is simulated such that θ_{3i} is von Mises with mean direction $\mu_{3\cdot 1,2}^{(i)}$ and concentration parameter $\kappa_{3\cdot 1,2}^{(i)}$, as defined for Equation (13) and given the θ_1 and θ_2 vectors ($i = 1, \ldots, n_1$).
- The simulation proceeds by cycling through columns one at a time, each time replacing the values therein with new, von Mises distributed data generated conditionally on the current values in the other two columns. This loop is done n_2 times.
- n_3 such data sets are generated giving, in total, $n_1 n_3$ vectors $(\theta_1, \theta_2, \theta_3)$.

An analysis of our experiments indicates that a value of $n_2 = 50$ gives comparable accuracy to that achieved with $n_2 = 100$, and is enough to give reasonably accurate estimates. We also observed that a value of $n_1 = 200$ gives more accurate results than a value of $n_1 = 100$ for fixed n_2, n_3. Overall, the accuracy of the estimates and the effect of changing n_1 are encouraging signs that both the simulation method and the estimation method are effective. In general, the convergence rate of Gibbs sampling will need to be monitored. For example, Liu (2001, pp. 131–132) shows that, for the bivariate normal case with zero means, unit variances, and correlation ρ, the convergence rate is measured by ρ^2. Indeed, after t iterations, the variance of the variables is $1 - \rho^{4t-2}$ so that, if ρ is near 1, more iterations will be required. Alternatively, if the absolute value of the correlation is small, fewer iterations will be required. The same principle should also applying the case of von Mises distributions, we conjecture.

Table 1 displays maximum likelihood (ML) and pseudolikelihood (PL) estimates based on a simulated data set (with $n_1 = 100, n_2 = 50, n_3 = 1$) for each of four different parameter configurations (where $\mu_1 = \mu_2 = \mu_3 = 0$). The figures in brackets give approximate standard errors of estimates, and are calculated from the Hessian matrix. Maximum likelihood estimates are obtained by incorporating a numerical integration into each stage of the algorithm in order to evaluate the unknown normalizing constant.

Comparison of the MLEs and MPLEs themselves reveals very little difference in terms of accuracy. The approximate standard errors are also closely comparable, with the odd exception, for example for the estimates of λ_{23} for parameter configuration three. For the λ estimates in parameter configurations one, two and four, the given standard errors are in general slightly smaller for the pseudolikelihood estimates than for the full likelihood estimates. It should be noted however that these are approximate and the difference is small.

As an indication of the relative computational expense of the two methods, the estimation of parameters using the pseudolikelihood took less than 1 second for each of the configurations in Table 1, whilst the

	True	ML	(SD)	PL	(SD)
κ_1	2	2.23	(0.31)	2.23	(0.33)
κ_2	3	2.81	(0.38)	2.81	(0.36)
κ_3	1	1.09	(0.22)	1.17	(0.22)
λ_{12}	2	1.63	(0.51)	1.39	(0.48)
λ_{13}	2	2.09	(0.43)	2.37	(0.38)
λ_{23}	2	2.30	(0.46)	2.58	(0.46)
κ_1	0.5	0.46	(0.23)	0.46	(0.23)
κ_2	0.75	0.92	(0.26)	0.91	(0.26)
κ_3	0.25	0.01	(0.25)	0.01	(0.26)
λ_{12}	2.0	1.84	(0.71)	1.73	(0.64)
λ_{13}	3.0	3.14	(0.67)	3.25	(0.64)
λ_{23}	4.0	3.75	(0.67)	3.76	(0.64)
κ_1	2	2.72	(0.80)	2.73	(0.81)
κ_2	2	2.03	(0.76)	2.03	(0.77)
κ_3	2	2.34	(0.91)	2.33	(0.93)
λ_{12}	20	12.45	(6.65)	12.69	(7.45)
λ_{13}	30	38.60	(7.42)	38.51	(7.42)
λ_{23}	40	38.43	(3.90)	38.27	(7.42)
κ_1	2.0	2.52	(0.31)	2.52	(0.31)
κ_2	2.0	2.32	(0.28)	2.32	(0.28)
κ_3	2.0	2.12	(0.26)	2.12	(0.26)
λ_{12}	0.1	0.16	(0.32)	0.15	(0.26)
λ_{13}	0.1	-0.12	(0.31)	-0.12	(0.26)
λ_{23}	0.1	0.42	(0.31)	0.44	(0.27)

Table 1: Mean values and approximate standard errors of pseudolikelihood (PL) and full likelihood (ML) estimates based on a single sample of trivariate von Mises data for four sets of parameters.

corresponding figures for the full likelihood averaged about 25 minutes. These figures clearly indicate the need for an alternative to the full likelihood in the current situation, whilst the accuracy of the estimates show the pseudolikelihood to be a good candidate for this alternative. It should be expected that the computational expense of MLEs relative to PLEs is even greater for higher dimensional data.

We now use the trivariate von Mises distribution in order to model a protein data set, using both the full and the pseudolikelihood.

6 Application to Protein Data

The data to be analyzed in this section comprise the ϕ and ψ triplets of 497 (classic) Gamma turns (Whitford, 2005, pp. 46–48) in protein backbone chains.

Definition: *A Gamma turn is a three-residue sequence defined by the existence of a hydrogen bond between* CO *of residue i and* NH *of residue $i+2$. In addition, the ϕ and ψ angles of residue $i+1$ fall in the ranges $\phi_{i+1} \in [35°, 115°] = [0.61, 2.00]$* radians *and $\psi_{i+1} \in [-104°, -24°] = [-1.82, -0.42]$* radians, *respectively.*

Figure 1 displays correlation plots of the data, with circular plots on the main diagonal, pairwise plots on the upper panels and circular correlation values on the lower panels. Circular correlations are calculated by replacing $(x_i - \overline{x})$ and $(y_i - \overline{y})$ in Pearson's product moment correlation for X and Y by $\sin(x_i - \overline{x})$ and $\sin(y_i - \overline{y})$, where $\overline{x}$ and $\overline{y}$ in the latter two expressions are sample mean directions. p-values for testing the significance of the correlation coefficients are also given.

As an exploratory analysis, a univariate von Mises distribution with mean direction μ and concentration parameter κ is fitted separately to each of ϕ_j and ψ_j, $j = 1, 2, 3$. Maximum likelihood estimates of μ and κ are displayed in Table 2. Note that the estimate of κ is large for ϕ_2 and ψ_2 since these angles are (by definition of the Gamma turn) constrained.

	ϕ_1	ϕ_2	ϕ_3	ψ_1	ψ_2	ψ_3
$\hat{\mu}$	-1.64	1.20	-1.76	1.58	-1.02	1.03
$\hat{\kappa}$	1.58	31.46	1.67	0.44	8.19	0.31

Table 2: Marginal MLE's of μ (in radians) and κ for Gamma turn data

We now use both the pseudolikelihood approach and the full likelihood approach to fit a trivariate von Mises distribution to the ϕ and ψ angles, separately, of the Gamma turn data, the results of which are shown in Table 3.

Comparing Tables 2 and 3 we see that, when using a univariate von Mises, the maximum likelihood estimates for μ and κ are very similar to the estimates for μ and κ in the trivariate case, with a possible tendency for the univariate κ estimates to be slightly smaller. Moreover, the λ estimates are generally

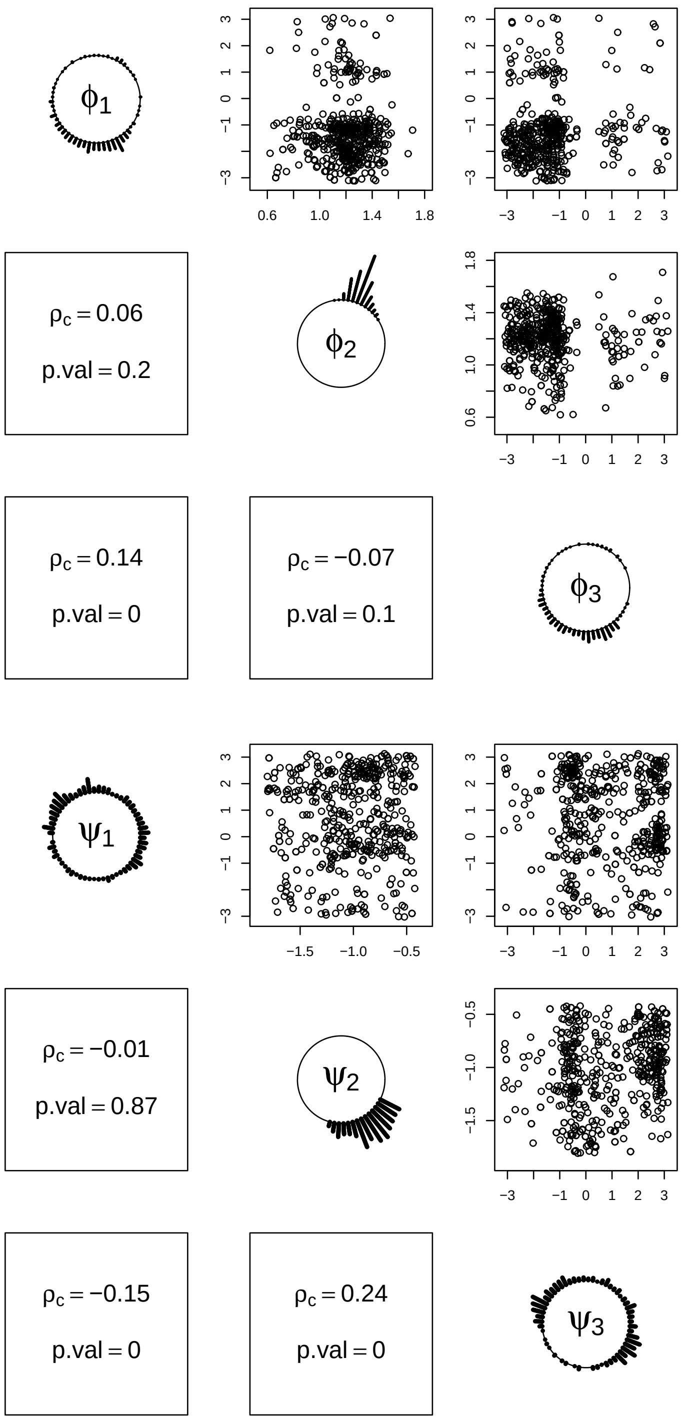

Figure 1: Matrix plot of ϕ (top panel) and ψ (bottom panel)angles for gamma turn data, with circular plots on main diagonal, pairwise plots on upper panels and correlations on lower panels

	MLE				PLE			
	ϕ		ψ		ϕ		ψ	
μ_1	–1.64	(0.04)	1.46	(0.13)	–1.63	(0.05)	1.46	(0.12)
μ_2	1.20	(0.03)	–1.02	(0.02)	1.20	(0.01)	–1.02	(0.02)
μ_3	–1.75	(0.05)	1.19	(0.11)	–1.73	(0.04)	1.27	(0.11)
κ_1	1.60	(0.09)	0.44	(0.07)	1.60	(0.09)	0.44	(0.07)
κ_2	31.72	(1.94)	8.92	(0.55)	31.73	(1.99)	8.87	(0.53)
κ_3	1.69	(0.12)	0.32	(0.07)	1.69	(0.10)	0.31	(0.07)
λ_{12}	0.55	(0.09)	0.23	(0.22)	0.65	(0.31)	0.15	(0.13)
λ_{13}	0.32	(0.12)	–0.44	(0.10)	0.39	(0.09)	–0.33	(0.06)
λ_{23}	–0.71	(0.18)	1.40	(0.21)	–0.79	(0.31)	1.13	(0.13)

Table 3: Maximum likelihood estimates (MLEs) and pseudolikelihood estimates (PLEs) (and their standard errors) for ϕ and ψ angles of Gamma turn data with $\boldsymbol{\phi}$ treated independently from $\boldsymbol{\psi}$. The units of the means are in radians.

less than 1 in magnitude, with the exception of λ_{23} for the ψ angles. However some also have small standard errors relative to their magnitude, suggesting that the “correlations” are statistically significant, though small.

We now extend the analysis of the Gamma turn data by developing likelihood ratio tests for the trivariate von Mises distribution.

6.1 Hypothesis Testing

An important part of fitting statistical models is the formulation and testing of hypotheses. In this section we compare likelihood ratio test statistics based on a von Mises maximum likelihood approach, a von Mises pseudolikelihood approach, and a normal approximation approach.

Since Gamma turns serve to reverse the direction of a polypeptide, it may be hypothesised that $\mu_1 = \mu_3$ for the ϕ and ψ angles of such a turn. A further test on the mean directions of the ϕ angles could be based on the hypothesis $\mu_1 = \mu_3 = \mu_2 - \pi$. Although this hypothesis may be criticized on the grounds that it has been generated by looking at the data, the focus of the present section is on formulating test procedures rather than the results of the tests *per se*. A test of independence of ϕ_1, ϕ_2 and ϕ_3 is equivalent to testing that all λ values are equal to zero. These three hypotheses will be tested in the three ways described above.

We outline the procedure for using a Normal approximation to the von Mises distribution in order to test the above hypotheses. Although κ values for angles 1 and 3 of both ϕ and ψ are reasonably small, the trivariate normal distribution given by Equation (4) will be used to test the hypothesis $\mu_1 = \mu_3$.

In (4), Θ is to be replaced by Φ or Ψ, depending on the variable of interest, whilst $\boldsymbol{\mu}_\Theta$ is the vector comprising the mean directions of Θ_1, Θ_2 and Θ_3. For $\hat{\boldsymbol{\mu}}_\phi$ and $\hat{\boldsymbol{\mu}}_\psi$ we take the parameters of the maximized pseudolikelihood. The approximate covariance matrices obtained for ϕ and ψ, using the normal approximation, are respectively

$$\hat{\boldsymbol{\Sigma}}_\psi = \begin{pmatrix} -6.28 & 1.43 & 12.06 \\ 1.43 & -0.11 & -1.96 \\ 12.06 & -1.96 & -16.98 \end{pmatrix} \qquad \hat{\boldsymbol{\Sigma}}_\phi = \begin{pmatrix} 0.67 & 0.01 & 0.15 \\ 0.01 & 0.03 & -0.01 \\ 0.15 & -0.01 & 0.63 \end{pmatrix} \tag{14}$$

The matrix $\hat{\boldsymbol{\Sigma}}_\psi$ is not positive definite. Therefore the hypothesis described will be carried out for the ϕ angles only.

<u>Test1:</u> $\mu_1 = \mu_3$

The test statistic, which is approximately χ^2_1 distributed under the null hypothesis and for large κ values, is $S = 2(l_{\text{full}} - l_{\text{red}})$, where l_{full} is the log–likelihood of the distribution (4) with mean vector and covariance matrix $\hat{\boldsymbol{\mu}}_\phi$ and $\hat{\boldsymbol{\Sigma}}_\phi$ respectively. l_{red} ("red" is as an abbreviation for *reduced*) is the corresponding value when the pseudolikelihood estimates are calculated with $\mu_1 = \mu_3$. We get $S = 2.91$, with a p–value of 0.093. We therefore accept the null hypothesis that the means are equal.

<u>Test2:</u> $\mu_1 = \mu_3 = \mu_2 - \pi$

Using the same procedure as for test 1, we obtain a test statistic of 191.8 on two degrees of freedom. This is clearly a very large value and we reject the null hypothesis. As seen in Table 3, the difference between the PLE of μ_2 and both μ_1 and μ_3 is slightly *less* than π, and μ_2 has a particularly small standard error.

<u>Test3:</u> $\lambda_{12} = \lambda_{13} = \lambda_{23} = 0$

Under this test of independence $\hat{\boldsymbol{\Sigma}}_\phi^{-1}$ for the reduced model is $\text{diag}(\hat{\boldsymbol{\kappa}})$, where $\hat{\boldsymbol{\kappa}}$ is the PLE of $(\kappa_1, \kappa_2, \kappa_3)$ when all λ values are equal to zero. $\hat{\boldsymbol{\kappa}}$ is calculated to be $(1.59, 31.46, 1.67)$ (values that correspond almost exactly with those in Table 2), and the resulting test statistic is 119.8 on three degrees of freedom. Again this value is highly significant, although perhaps not surprisingly so, given the small magnitude of the standard errors of the λ estimates relative to the magnitude of the estimates themselves.

Of importance in fitting a normal distribution to directional data is the point at which the circle is cut. In the current situation we have done this at π, although this should perhaps be done separately for each variable in such a way that the concentration parameter is maximized. Of course, the more concentrated the data, the easier it is to select a point at which to cut the circle, and the more appropriate the use of the normal approximation. Since in the current setting ϕ_1 and ϕ_3 are quite dispersed, the normal approximation is perhaps not entirely suitable, and the issue is not pursued.

The left half of Table 4 gives log–likelihood values for the full and reduced models and for each of the three approaches described: full (joint) von Mises likelihood, von Mises pseudolikelihood and normal approximation. The right half of the table displays the likelihood ratio test statistics based on the values in the left part of the table. As can be seen from the table, the same conclusions are reached regarding the acceptance or rejection of the null hypothesis in each case, namely that the only null hypothesis not rejected is for test 1, $\mu_1 = \mu_3$ (using a 5% significance level).

	Log-likelihood			Test statistic		
	VM joint	VM pseudo	Normal	VM joint	VM pseudo	Normal
Full	-1233.66	-1225.46	-1857.21			
Test 1	-1235.34	-1227.04	-1858.66	3.35	3.17	2.91
Test 2	-1259.57	-1250.30	-1953.10	51.81	49.70	191.79
Test 3	-1238.88	-1238.88	-1917.13	10.43	26.85	119.84

Table 4: Log-likelihood and log-pseudolikelihood values evaluated at MLE's and PLE's, and test statistics based on these values.

Of more interest, however, is comparison of log-likelihood values for the different methods, and the resulting differences in the test statistics. In particular, the log-pseudolikelihood appears to underestimate the full log-likelihood slightly (in absolute value) in the first three rows of Table 4. In each case, however, the size of the underestimation is similar, resulting in little difference in the test statistics for tests 1 and 2. In the case $\boldsymbol{\lambda} = 0$ (bottom row of the table), the full likelihood and pseudolikelihood values are the same. This is to be expected, since in this case each is the product of the same three independently distributed von Mises distributions. The effect is an inflated test statistic when the test is based on the pseudolikelihood.

7 Discussion

In this paper we have derived the efficiency of the pseudolikelihood for the bivariate von Mises distribution. This efficiency has been shown to depend on the value of parameters, and is reasonably high so long as λ is not very large. We have performed a thorough analysis of properties of pseudolikelihood estimates for the trivariate von Mises distribution, implementing a Gibbs sampling approach to data simulation and comparing PLEs with MLEs. For all parameter combinations considered, the two are shown to have similar properties in terms of accuracy and precision of estimates, whilst the former are calculated at a fraction of the computational cost. An analysis of protein fold data shows the trivariate model to be a reasonable fit for those data studied, and likelihood ratio tests have been developed for the trivariate distribution.

In the computation of the estimators, we observed that obtaining the maximum pseudolikelihood estimators is much faster than obtaining the maximum likelihood estimators. For large data sets in high dimensions where the computation of moments and the maximum likelihood estimators might take a very long time, the maximum pseudolikelihood estimators will have a distinct computational advantage.

We have selected only one example from Bioinformatics but the circular variables in protein structure are obiquious; conformational angles appear for example not only in Gamma turn but in different patterns (helices, β-sheets). Also each amino acid can have more than one conformational angle such as ϕ, ψ and χ-angles. Hughes (2007) contains another application to trivariate data in protemics concerning "χ" angles in Serine and Valine residues (the three angles, then being (ϕ, ψ, χ)). For this dataset, it was seen that a mixture distribution would be required to model the data, but a preliminary analysis was carried out by segmenting the data.

More appropriate for ϕ and ψ angles could be a six-dimensional distribution for Gamma turns. Alternatively, a time series on $\phi - \psi$ could be explored, or something using a neighbourhood structure. Some form of Hisenberg distribution (Ising-type model on circle) could be also appropriate to use the neighbourhood information with pdf proportional to $\exp\{\Sigma a_i^T x_i + \Sigma a_{ij} x_i^T x_j\}$ where $a_{ij} = 1$ if i and j are neighbours and $a_{ij} = 0$ otherwise. Here $x_i^T = (\cos\theta_i, \sin\theta_i)$ so it leads to the cosine type density. The estimation could be perhaps simplified by using a saddlepoint approximation to the normalizing constant (see, for example, Kent and Mardia, 2006, for the complex Bingham quartic distribution).

An alternative multivariate von Mises distribution is to extend the bivariate cosine model of Mardia *et al.* (2007). This leads to a multivariate cosine model, which can be defined for $\mathbf{\Theta} = (\Theta_1, \Theta_2, \ldots, \Theta_p)$ by

$$f(\mathbf{\Theta}) = C_p^{-1}(\boldsymbol{\kappa}, \mathbf{\Delta}) \exp\left\{\boldsymbol{\kappa}^T c(\boldsymbol{\theta}, \boldsymbol{\mu}) - s(\boldsymbol{\theta}, \boldsymbol{\mu})^T \mathbf{\Delta}\, s(\boldsymbol{\theta}, \boldsymbol{\mu}) - c(\boldsymbol{\theta}, \boldsymbol{\mu})^T \mathbf{\Delta}\, c(\boldsymbol{\theta}, \boldsymbol{\mu})\right\},$$

where $-\pi < \theta_j \leq \pi$, $-\pi < \mu_j \leq \pi$, $\kappa_j \geq 0$, $\delta_{jl} \geq 0$. The vectors $c(\boldsymbol{\theta}, \boldsymbol{\mu})$, $s(\boldsymbol{\theta}, \boldsymbol{\mu})$, $\boldsymbol{\mu}$ and $\boldsymbol{\kappa}$ are defined as for the multivariate Sine model (3), while $[\mathbf{\Delta}]_{jl} = \delta_{jl} = \delta_{lj}$ and $\delta_{jj} = 0$. The normalizing constant is $C_p^{-1}(\boldsymbol{\kappa}, \mathbf{\Delta})$. This model can be investigated in ways similar to those above for the multivariate Sine model. Comparisons analogous to those made by Mardia *et al.* (2007) for the bivariate Sine and Cosine models can then be made for the multivariate models. The two could also be compared, for example, in terms of the efficiency of the pseudolikelihood for parameter estimation. Although there is some debate about which model is preferred, we have made a compromise here which sacrifices conditional properties.

Acknowledgements

This work began during a visit by Kanti Mardia to Harishdar Singh (deceased) at NIOSH during which this new challenging area of Bioinformatics was introduced. The authors wish to express special thanks to Dan Sharpe and NIOSH for making the visit possible. The authors would also like to thank John Kent for many helpful comments, Thomas Hamelryck for useful discussions and for providing the Gamma turn data, and Wally Gilks for feedback on an earlier draft of this work. Gareth Hughes was supported by an EPSRC studentship.

References

Baba, Y. (1981). Statistics of angular data. Wrapped normal distribution model. *Proceedings of Institute of Statistical Mathematics*, **28**, 41–54. (in Japanese)

Besag, J. (1975). Statistical analysis of non-lattice data. *The Statistician*, **24**. 179–195.

Cox, D. R. (1975). Contribution to discussion of Mardia (1975a). *Journal of Royal Statistical Society* **B 37**, 380–381.

Davison, A.C. (2003). *Statistical Models.* Cambridge University Press.

Demchuk, E. & Singh, H. (2001). Statistical thermodynamics of hindered rotation from computer simulations. *Molecular Physics* **99**, 627–36.

Hughes, G. (2007). *Multivariate and Time Series Models for Circular Data with Applications to Protein Conformational Angles.* PhD thesis, University of Leeds.

Johnson, R. and Wehrly, T. (1977). Measures and models for angular correlation and angular-linear correlation. *Journal of Royal Statistical Society* **B 39**, 222–229.

Jupp, P. E. & Mardia, K. V. (1980). A general correlation coefficient for directional data and related regression problems. *Biometrika* **67**, 163–73.

Kent, J.T. and Mardia, K.V. (2006). The complex Bingham quartic distribution and shape analysis. *Journal of Royal Statistical Society Series B*, **68**, 747–765.

Liu, J.S. (2001). *Monte Carlo Strategies in Scientific Computing.* Springer-Verlag, New York.

Mardia, K. V. (1975a). Statistics of directional data (with Discussion). *Journal of Royal Statistical Society* **B 37**, 349–93.

Mardia, K. V. (1975b). Characterizations of directional distributions. In *Statistical Distributions in Scientific Work*, **3**, Ed. G. P. Patil, S. Kotz and J. K. Ord, pp. 365–85. Dordrecht: Reidel.

Mardia, K.V., Hughes, G. & Taylor, C.C. (2007). Efficiency of the pseudolikelihood for multivariate normal and von Mises distributions. *submitted for publication.*

Mardia, K. V. & Jupp, P. E. (1999). *Directional Statistics.* New York: Wiley.

Mardia, K.V. & Patrangenaru, V. (2005) Directions and projective shapes. *Annals of Statistics*, **33**, 1666–1699.

Mardia, K.V., Taylor, C.C. & Subramaniam, G. K. (2007) Protein Bioinformatics and Mixtures of Bivariate von Mises Distributions for Angular Data. *Biometrics* (in press).

Rivest, L. P. (1984). Symmetric distributions for dependent unit vectors. *Annals of Statistics*, **12**, 1050–1057.

Rivest, L. P. (1988). A distribution for dependent unit vectors. *Communications in Statistics* **A 17**, 461–83.

Singh, H., Hnizdo, V. and Demchuk, E. (2002). Probabilistic model for two dependent circular variables. *Biometrika*, **89**, 719–723.

Subramaniam, G. K. (2005). *von Mises Distributions with Applications in Speech Data*. PhD thesis, University of Leeds, Leeds, UK.

Whitford, D. (2005). *Proteins, Structure and Function*. Wiley, Chichester.

Yfantis, E. A. & Borgmann, L. E. (1982). An extension of the von Mises distribution. *Communications in Statistics* **A 11**, 1695–1706.