The Canadian Journal of Statistics
La revue canadienne de statistique

A Multivariate Von Mises Distribution with Applications to Bioinformatics

Author(s): Kanti V. Mardia, Gareth Hughes, Charles C. Taylor and Harshinder Singh

Source: *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, Mar., 2008, Vol. 36, No. 1 (Mar., 2008), pp. 99-109

Published by: Statistical Society of Canada

Stable URL: https://www.jstor.org/stable/20445295

REFERENCES
Linked references are available on JSTOR for this article:
https://www.jstor.org/stable/20445295?seq=1&cid=pdf-reference#references_tab_contents
You may need to log in to JSTOR to access the linked references.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms

Statistical Society of Canada is collaborating with JSTOR to digitize, preserve and extend access to *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*

JSTOR

This content downloaded from
193.205.142.15 on Mon, 05 May 2025 08:33:37 UTC
All use subject to https://about.jstor.org/terms

The Canadian Journal of Statistics
Vol. 36, No. 1, 2008, Pages 99–109
La revue canadienne de statistique

A multivariate von Mises distribution with applications to bioinformatics

Kanti V. MARDIA, Gareth HUGHES, Charles C. TAYLOR and Harshinder SINGH

Key words and phrases: Bessel function; circular mean; directional data; gamma turn; proteomics; pseudo-likelihood; von Mises distribution; Wald test.

MSC 2000: Primary 62H11, 62H15; secondary 62F40.

Abstract: Motivated by problems of modelling torsional angles in molecules, Singh, Hnizdo & Demchuk (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. The authors present here a multivariate extension of the bivariate model of Singh, Hnizdo & Demchuk (2002). They study the conditional distributions and investigate the shapes of marginal distributions for a special case. The methods of moments and pseudo-likelihood are considered for the estimation of parameters of the new distribution. The authors investigate the efficiency of the pseudo-likelihood approach in three dimensions. They illustrate their methods with protein data of conformational angles.

Une loi de von Mises multivariée et ses applications en bioinformatique

Résumé : Motivés par des questions concernant la modélisation d'angles de torsion de molécules, Singh, Hnizdo & Demchuk (2002) ont proposé un modèle circulaire bivarié qui est à la fois l'analogue naturel pour le tore de la loi normale bivariée et une généralisation bivariée naturelle de la loi de von Mises univariée. Les auteurs présentent ici une version multidimensionnelle du modèle bivarié de Singh, Hnizdo & Demchuk (2002). Ils en étudient les lois conditionnelles et précisent la forme des marges dans un cas spécial. La méthode des moments et celle de la pseudo-vraisemblance sont envisagées pour l'estimation des paramètres de la nouvelle loi. Les auteurs étudient l'efficacité de l'approche par pseudo-vraisemblance en trois dimensions. Ils illustrent leur propos au moyen de mesures d'angle reflétant la forme de protéines.

1. INTRODUCTION

Circular distributions are important in modelling angular variables in biology, astronomy, meteorology, earth sciences, and several other areas (Mardia & Jupp 1999). 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. 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 is considered by Rivest (1988), subsets of which were investigated by Subramaniam (2005) and Mardia, Taylor & Subramaniam (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. A 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.

This content downloaded from
193.205.142.15 on Mon, 05 May 2025 08:33:37 UTC
All use subject to https://about.jstor.org/terms

Recently Demchuk & Singh (2001) emphasized the importance of bivariate circular distributions in molecular sciences for modelling torsional angles in molecules. Motivated by problems in molecular sciences, Singh, Hnizdo & Demchuk (2002) proposed a bivariate circular distribution which is a natural torus analogue of the bivariate normal distribution. Let Θ_1 and Θ_2 be two circular random variables. The probability density function is of the form

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

for $-\pi \leq \theta_1, \theta_2 < \pi$, where $\kappa_1, \kappa_2 \geq 0$, $-\infty < \lambda_{12} < \infty$, $-\pi \leq \mu_1, \mu_2 < \pi$, and $T(\cdot)$ is a normalization constant, which is given by (Singh, Hnizdo & Demchuk 2002),

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

In this paper we propose a multivariate extension of the von Mises distribution which could be useful for jointly modelling several circular variables. 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 "sin" model (Mardia, Taylor & Subramaniam 2007). 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 for estimating the parameters of the proposed multivariate distribution, namely, the maximum pseudo-likelihood method, the method of moments and a maximum likelihood method. Since the proposed distribution belongs to the natural exponential family, the maximum likelihood estimators and the moment estimators match for an important case. In Section 4, we use simulations to compare the performance of the maximum pseudo-likelihood estimators with maximum likelihood estimators for the trivariate von Mises distribution. Our objective here is to ascertain whether the conclusions of Mardia, Hughes & Taylor (2007) would extend to three dimensions: in that work, for the bivariate von Mises we found analytically that the estimators were reasonably similar so long as the product precision parameter λ_{12} is not large relative to the concentration parameters. In Section 5 we apply the methods to some real data from the field of proteomics, and in Section 6 we conclude with a discussion.

2. A MULTIVARIATE VON MISES DISTRIBUTION

We define a multivariate angular distribution which is an extension of Singh, Hnizdo & Demchuk (2002) as follows. The probability density function of $\boldsymbol{\Theta}^\top = (\Theta_1, \Theta_2, \dots, \Theta_p)$ is given by

$$\{T(\boldsymbol{\kappa}, \boldsymbol{\Lambda})\}^{-1} \exp\{\boldsymbol{\kappa}^\top c(\boldsymbol{\theta}, \boldsymbol{\mu}) + s(\boldsymbol{\theta}, \boldsymbol{\mu})^\top \boldsymbol{\Lambda} s(\boldsymbol{\theta}, \boldsymbol{\mu})/2\}, \quad (3)$$

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

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

and

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

with $\{T(\boldsymbol{\kappa}, \boldsymbol{\Lambda})\}^{-1}$ 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 $\boldsymbol{\Theta} \sim M_p(\boldsymbol{\mu}, \boldsymbol{\kappa}, \boldsymbol{\Lambda})$. We note that for $p = 1$, this is a univariate von Mises density and for $p = 2$, this density corresponds to the bivariate model of Singh, Hnizdo & Demchuk (2002). Now we investigate the properties of the proposed multivariate von Mises distribution.

This content downloaded from
193.205.142.15 on Mon, 05 May 2025 08:33:37 UTC
All use subject to https://about.jstor.org/terms

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

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

and $\mathsf{N}_p(\boldsymbol{\mu}, \mathbf{A})$ denotes a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{A}$.

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

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

Specifically, (taking $\boldsymbol{\mu} = \mathbf{0}$) we can obtain the conditional distribution of $\Theta_1, \Theta_2, \ldots, \Theta_r$ given $\theta_{r+1}, \theta_{r+2}, \ldots, \theta_p$ as

$$\begin{aligned} & f(\Theta_1, \ldots, \Theta_r \,|\, \theta_{r+1}, \ldots, \theta_p) \\ = \;& f_{\text{cond}} \propto \exp\left\{ \sum_{j=1}^{r} \kappa_j \cos\theta_j + \sum_{j=1}^{r-1} \sum_{\ell=j+1}^{r} \lambda_{j\ell} \sin\theta_j \sin\theta_\ell + \sum_{j=1}^{r} c_j \sin\theta_j \right\}, \end{aligned}$$

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

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

Next, writing $\phi_j = \theta_j - \nu_j$, and then expanding each sine term and their product gives, for constants $a_{j\ell}$, $b_{j\ell}$, $c_{j\ell}$ and $d_{j\ell}$,

$$\begin{aligned} f_{\text{cond}} \propto \exp\Bigg\{ & \sum_{j=1}^{r} a_j \cos\phi_j + \\ & \sum_{j=1}^{r-1} \sum_{\ell=j+1}^{r} \left[a_{j\ell} \cos\phi_j \cos\phi_\ell + b_{j\ell} \sin\phi_j \sin\phi_\ell + c_{j\ell} \cos\phi_j \sin\phi_\ell + d_{j\ell} \sin\phi_j \cos\phi_\ell \right] \Bigg\}, \end{aligned}$$

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

For the bivariate von Mises distribution, Singh, Hnizdo & Demchuk (2002) proved that marginal distributions are symmetric around their circular means and are either unimodal or bimodal. When $\boldsymbol{\Theta}$ is distributed according to (3), it is clear that $\boldsymbol{\Theta}$ and $-\boldsymbol{\Theta}$ 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.

This content downloaded from 193.205.142.15 on Mon, 05 May 2025 08:33:37 UTC
All use subject to https://about.jstor.org/terms

3. APPROACHES FOR INFERENCE

Given a sample of data $\boldsymbol{\theta}_i = (\theta_{1i}, \ldots, \theta_{pi})^\top$, $i = 1, \ldots, n$ from distribution (3), we need to estimate the parameters of the model. 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.

We consider all univariate and bivariate conditional distributions. We have $\hat{\mu}_i = \bar{x}_{0i}$, $i = 1, \ldots, p$, where $\bar{x}_{0i}$ are the circular means of the marginal distributions of Θ_i, $i = 1, \ldots, p$. 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 $\widehat{\boldsymbol{\Sigma}} = (\bar{S}_{ij})$ with

$$\bar{S}_{ij} = \frac{1}{n} \sum_{r=1}^{n} \sin(\theta_{ir} - \bar{x}_{0i}) \sin(\theta_{jr} - \bar{x}_{0j}).$$

The interesting point is that $\bar{S}$ is the standard covariance matrix, and it should be interpreted as such. By taking the inverse of $\widehat{\boldsymbol{\Sigma}}$, we then obtain estimates of κ_i, $i = 1, \ldots, p$ and λ_{ij} $(i \neq j)$ using (4).

3.2. Maximum pseudo-likelihood method.

Define the pseudo-likelihood (Besag 1975), based on a random sample of n observations of $\boldsymbol{\theta} = (\Theta_1, \ldots, \Theta_p)^\top$, by

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

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

$$\mathrm{PL} = (2\pi)^{-pn} \prod_{j=1}^{p} \prod_{i=1}^{n} \big[I_0(\kappa_{j\cdot\mathrm{rest}}^{(i)})\big]^{-1} \exp\big\{\kappa_{j\cdot\mathrm{rest}}^{(i)} \cos(\theta_{ji} - \mu_{j\cdot\mathrm{rest}}^{(i)})\big\}, \tag{5}$$

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

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

Parameter estimation based on the pseudo-likelihood approach proceeds by maximizing PL with respect to the $p + p(p+1)/2$ unknown parameters.

4. COMPARISON OF ESTIMATION METHODS FOR THE TRIVARIATE CASE

Mardia, Hughes & Taylor (2007) have analytically investigated the efficiency of the maximum pseudo-likelihood estimators in the bivariate case. The estimators were seen to be reasonably close so long as the product precision parameter λ_{12} is not large relative to the concentration parameters. Here, we compare these estimation methods for the trivariate case using simulated

This content downloaded from
193.205.142.15 on Mon, 05 May 2025 08:33:37 UTC
All use subject to https://about.jstor.org/terms

data. The simulations reported below are not extensive since our aim is to examine whether the broad conclusions for $p = 2$ extend to the case $p = 3$. We first outline a method of generating data using a Gibbs sampling technique, and then we examine the pseudo-likelihood estimates and compare these to maximum likelihood estimates and approximate moment estimates.

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. We simulate repeatedly values of θ_1, θ_2 and θ_3 based on the conditional distribution (univariate von Mises) of one variable given the other two. For the first cycle, θ_1 and θ_2 are simulated from univariate von Mises distributions with concentration parameters κ_1 and κ_2, respectively. When the number n_1 of cycles is large, these angles have the desired trivariate von Mises distribution. The process is repeated n_2 times giving a data set comprising n_2 triplets $(\theta_1, \theta_2, \theta_3)$.

The pseudo-likelihood for the trivariate von Mises distribution is given by Equation (5) with $p = 3$. Estimates based on the maximization of this can be compared with the standard maximum likelihood approach. In our experiments we first investigated suitable choices of n_1 and n_2 and then the main objective, which is to compare the pseudo-likelihood estimates with maximum likelihood estimates.

TABLE 1: Estimated values and approximate standard errors (SE) of pseudo-likelihood (PL) and full likelihood (ML) estimates based on a single sample of trivariate von Mises data for four sets of parameters.

	True	ML	SE	PL	SE	Moment
κ_1	2	2.66	(0.38)	2.81	(0.53)	1.66
κ_2	3	2.84	(0.39)	2.81	(0.44)	1.93
κ_3	1	0.98	(0.21)	0.93	(0.19)	0.94
λ_{12}	2	2.33	(0.55)	2.64	(0.80)	0.72
λ_{13}	2	2.58	(0.45)	2.57	(0.50)	0.37
λ_{23}	2	1.49	(0.48)	1.17	(0.51)	0.34
κ_1	0.5	0.82	(0.26)	0.82	(0.30)	0.75
κ_2	0.75	0.71	(0.26)	0.71	(0.26)	0.73
κ_3	0.25	0.39	(0.26)	0.40	(0.28)	0.64
λ_{12}	2.0	2.36	(0.73)	2.24	(0.68)	0.24
λ_{13}	3.0	3.27	(0.71)	3.36	(0.64)	0.16
λ_{23}	4.0	3.49	(0.70)	3.53	(0.69)	0.16
κ_1	2	2.65	(0.97)	2.65	(0.98)	0.66
κ_2	2	1.66	(0.81)	1.65	(0.85)	0.63
κ_3	2	2.01	(0.85)	2.02	(0.92)	0.65
λ_{12}	20	36.85	(8.63)	36.76	(6.99)	0.17
λ_{13}	30	40.01	(8.55)	40.15	(8.49)	0.20
λ_{23}	40	23.66	(8.61)	23.64	(7.87)	0.16
κ_1	2.0	1.84	(0.23)	1.84	(0.23)	1.52
κ_2	2.0	1.83	(0.23)	1.83	(0.23)	1.52
κ_3	2.0	1.94	(0.24)	1.94	(0.23)	1.63
λ_{12}	0.1	0.15	(0.28)	0.14	(0.28)	0.03
λ_{13}	0.1	0.17	(0.28)	0.16	(0.28)	0.06
λ_{23}	0.1	0.12	(0.28)	0.12	(0.30)	0.06

Table 1 displays maximum likelihood (ML) and pseudo-likelihood (PL) estimates based on a

This content downloaded from
193.205.142.15 on Mon, 05 May 2025 08:33:37 UTC
All use subject to https://about.jstor.org/terms

simulated data set (with $n_1 = 50$, $n_2 = 100$) for each of four different parameter configurations. In each case we used $\mu_1 = \mu_2 = \mu_3 = 0$ which are assumed known, since we do not believe that the estimation of $\boldsymbol{\mu}$ will affect our results. Maximum likelihood estimates are obtained by incorporating a numerical integration into each stage of the algorithm in order to evaluate the unknown normalizing constant. The figures in brackets give approximate standard errors of estimates. For the maximum likehood these are calculated from the Hessian matrix. For the pseudo-likelihood, the standard errors are calculated from a jackknife estimate of the covariance matrix.

An analysis of our experiments indicates that a value of $n_1 = 50$ (loops) leads to estimates of comparable accuracy to those obtained with $n_1 = 100$, for various parameter configurations. This indicates that a value of $n_1 = 50$ is sufficient to simulate data successfully for those parameter values used. 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 apply in the case of von Mises distributions, we conjecture.

Comparison of the maximum likelihood and the pseudo-likelihood estimators themselves reveals very little difference in the two estimates. The approximate standard errors are also closely comparable. It can be seen that the standard errors of the pseudo-likelihood are closer to the standard errors of the maximum likelihood when the λ are small relative to the κ. Thus, the results for the trivariate case are consistent with those in Mardia, Hughes & Taylor (2007).

As an indication of the relative computational expense of the two likelihood methods, the estimation of parameters using the pseudo-likelihood took less than 1 second for each of the configurations in Table 1, whilst the 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 pseudo-likelihood to be a good candidate for this alternative. It should be expected that the computational expense of ML estimators relative to PL estimators is even greater for higher dimensional data.

Finally, note that the moment estimates are, in general, very poor here—except for the last parameter configuration. This is to be expected, since the concentration parameters are not large here. We now use the trivariate von Mises distribution to model a protein data set, comparing the full likelihood and pseudo-likelihood.

5. APPLICATION TO PROTEIN DATA

5.1. Gamma turn 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 $(x_i - \bar{x})$ and $(y_i - \bar{y})$ in Pearson's product moment correlation for X and Y by $\sin(x_i - \bar{x})$ and $\sin(y_i - \bar{y})$, where $\bar{x}$ and $\bar{y}$ in the latter two expressions are sample mean directions. p-values for testing the significance of the correlation coefficients are also given.

This content downloaded from
193.205.142.15 on Mon, 05 May 2025 08:33:37 UTC
All use subject to https://about.jstor.org/terms

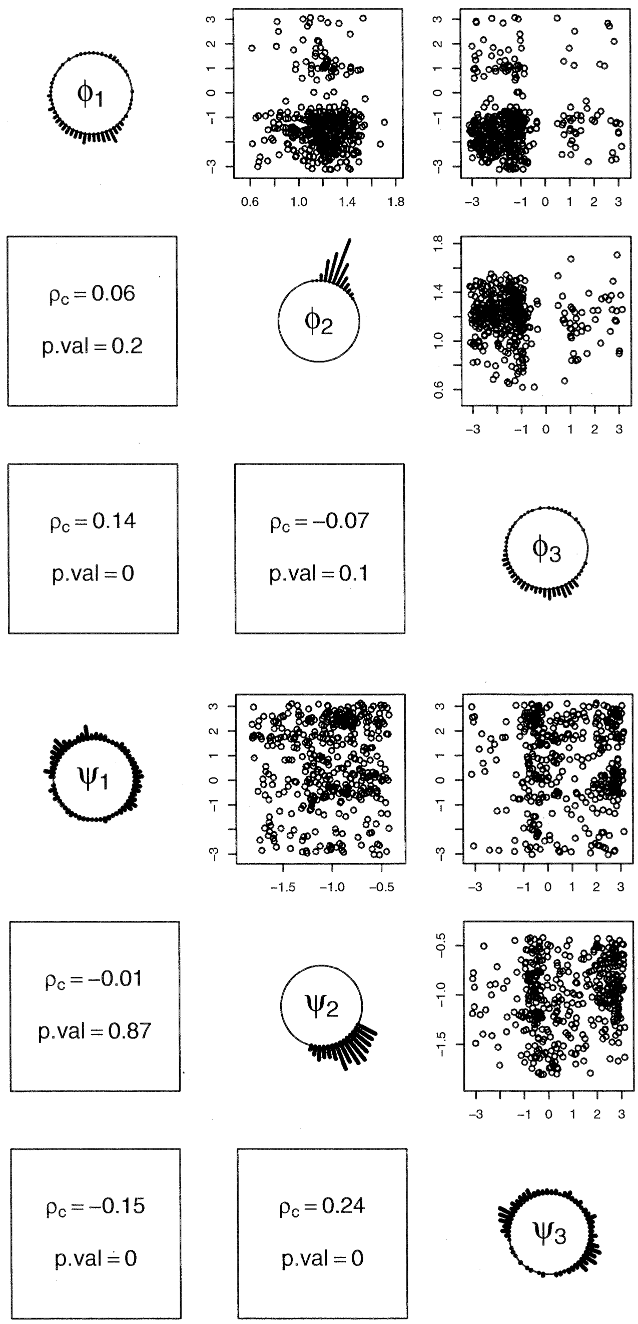

FIGURE 1: Matrix plot of ϕ (top panel) and ψ (bottom panel) angles for gamma turn data, with circular plots on main diagonal, pairwise plots on upper panels, and correlations on lower panels.

This content downloaded from
193.205.142.15 on Mon, 05 May 2025 08:33:37 UTC
All use subject to https://about.jstor.org/terms

5.2. Estimation.

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.

TABLE 2: Marginal maximum likelihood estimators of μ (in radians) and κ for gamma turn data. Standard errors are given in parentheses.

	ϕ_1	ϕ_2	ϕ_3	ψ_1	ψ_2	ψ_3
$\hat{\mu}$	−1.64 (0.05)	1.20 (0.01)	−1.76 (0.04)	1.58 (0.15)	−1.02 (0.02)	1.03 (0.21)
$\hat{\kappa}$	1.58 (0.09)	31.46 (1.98)	1.67 (0.10)	0.44 (0.07)	8.19 (0.50)	0.31 (0.06)

We now use the pseudo-likelihood approach, the full likelihood approach and the approximate moment estimators 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. Both the ML and the PL estimates were calculated by maximizing the respective likelihoods for all nine parameters simultaneously.

TABLE 3: Maximum likelihood estimates (MLE) and pseudo-likelihood estimates (PLE) (and their standard errors) for ϕ and ψ angles of gamma turn data with ϕ treated independently from ψ. The units of the means are in radians.

	MLE				PLE				Moment	
	ϕ		ψ		ϕ		ψ		ϕ	ψ
μ_1	−1.64	(0.04)	1.46	(0.13)	−1.63	(0.04)	1.46	(0.16)	−1.64	1.58
μ_2	1.20	(0.03)	−1.02	(0.02)	1.20	(0.01)	−1.02	(0.02)	1.20	−1.02
μ_3	−1.75	(0.05)	1.19	(0.11)	−1.73	(0.04)	1.27	(0.12)	−1.75	1.03
κ_1	1.60	(0.09)	0.44	(0.07)	1.60	(0.11)	0.44	(0.06)	3.45	1.82
κ_2	31.72	(1.94)	8.92	(0.55)	31.73	(2.32)	8.87	(0.45)	32.50	9.05
κ_3	1.69	(0.12)	0.32	(0.07)	1.69	(0.11)	0.31	(0.05)	3.04	1.49
λ_{12}	0.55	(0.09)	0.23	(0.22)	0.65	(0.43)	0.15	(0.20)	0.75	0.12
λ_{13}	0.32	(0.12)	−0.44	(0.10)	0.39	(0.12)	−0.33	(0.10)	0.47	−0.25
λ_{23}	−0.71	(0.18)	1.40	(0.21)	−0.79	(0.43)	1.13	(0.21)	−0.83	0.89

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 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, which suggests that the "correlations" are statistically significant, though small. The standard errors for the full maximum likelihood estimates are based on the respective Hessian matrices. The standard errors for the pseudo-likelihood estimates are based on the jackknife estimate of the covariance matrix. As expected, the approximate moment estimates are poor, since two of the concentration parameters (κ_1 and κ_3) are low. For this reason, we no longer consider the normal approximation as suitable for these data.

We now extend the analysis of the gamma turn data by considering a likelihood ratio test and an approximate test for the trivariate von Mises distribution.

This content downloaded from
193.205.142.15 on Mon, 05 May 2025 08:33:37 UTC
All use subject to https://about.jstor.org/terms

5.3. Hypothesis testing.

An important part of fitting statistical models is the formulation and testing of hypotheses. In this section we compute likelihood ratio test statistics based on a von Mises maximum likelihood approach and use an approximate Wald-type test. To illustrate the methods, various hypotheses will be tested for the ϕ angles only.

Since gamma turns serve to reverse the direction of a polypeptide, it may be hypothesized 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 whether all λ values are equal to zero.

The left half of Table 4 gives log-likelihood values and corresponding test statistics for the full and restricted models. Each test is carried out by comparing the maximized log-likelihood under the full model, with that corresponding to the test of interest. To test $\mathcal{H}_0 : \mu_1 = \mu_3$ the likelihood ratio test (LRT) statistic is 3.30, which gives a p-value of 0.07. The LRT for $\mathcal{H}_0 : \mu_1 = \mu_3 = \mu_2 - \pi$ is 52.07, so this hypothesis is emphatically rejected. As seen in Table 3, the difference between the MLE of μ_2 and both μ_1 and μ_3 is slightly *less* than π, and μ_2 has a particularly small standard error. The test statistic (LRT $= 11.04$) for independence is also highly significant (test 3).

5.4. Approximations.

Given the heavy computational burden of obtaining the ML estimates and the relatively cheap cost of obtaining the PL estimates, we now consider using the latter as approximations. The third column of Table 4 gives the (full) log-likelihood values corresponding to the estimate found to maximize the pseudo-likelihood. Of some interest is the comparison of log-likelihood values for the two estimates. In particular, the log-likelihood for the pseudo-likelihood estimates will inevitably be less than or equal to the maximum value. In the case $\boldsymbol{\lambda} = \mathbf{0}$ (last row of Table 4), the full likelihood and pseudo-likelihood 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.

TABLE 4: Log-likelihood values evaluated at MLEs (joint) and PLEs (pseudo), both for the full model, and three restricted models. Doubled differences are based on these values, and "Wald" is a Wald-type test statistic using a jackknife estimate of the covariance matrix. The restricted models are: 1. $\mu_1 = \mu_3$; 2. $\mu_1 = \mu_3 = \mu_2 - \pi$; 3. $\lambda_{12} = \lambda_{13} = \lambda_{23} = 0$

Log-likelihoods and Test Statistics

	Joint		Pseudo	
	ML	2× Diff	PL	Wald
Full	−1233.36		−1233.66	
Model 1	−1235.01	3.30	−1235.34	4.87
Model 2	−1259.40	52.07	−1259.57	69.34
Model 3	−1238.88	11.04	−1238.88	9.79

In the absence of the maximum likelihood estimator, it might be tempting to use the third column of Table 4 to derive test statistics, but the theory underpinning such an approach has not been worked out. So, instead, we obtain a Wald-type statistic (shown in column 4 of Table 4) using $(\mathbf{q}-\mathbf{q}_0)^\top \mathbf{V}(\mathbf{q})^{-1}(\mathbf{q}-\mathbf{q}_0)$ in which $\mathbf{q}$ is the pseudo-likelihood estimate of the parameters $(\mu_1, \mu_2, \mu_3, \lambda_{12}, \lambda_{13}, \lambda_{23}, \kappa_1, \kappa_2, \kappa_3)$ under the full model, $\mathbf{q}_0$ is the estimate under the reduced model, and $\mathbf{V}(\mathbf{q})$ is the jackknife estimate of the covariance matrix.

This content downloaded from
193.205.142.15 on Mon, 05 May 2025 08:33:37 UTC
All use subject to https://about.jstor.org/terms

As can be seen from Table 4, if we use the Wald-type test statistic, then broadly similar conclusions, i.e., the two statistics give p-values which are similar, are reached regarding the acceptance or rejection of each null hypothesis as for the more accurate (but much more expensive) likelihood ratio test.

6. DISCUSSION

We have examined some properties of pseudo-likelihood estimates for the trivariate von Mises distribution, implementing a Gibbs sampling approach to data simulation and comparing pseudo-likelihood with maximum likelihood estimators. 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 a likelihood ratio test has been illustrated for the trivariate distribution.

In the computation of the estimators, we observed that obtaining the maximum pseudo-likelihood 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 pseudo-likelihood estimators will have a distinct computational advantage.

We have selected only one example from bioinformatics, but circular variables in protein structure are ubiquitous; conformational angles appear, e.g., not only in gamma turns but in different patterns (helices, β-sheets). Also each amino acid is associated with more than one conformational angle such as χ-angles. Hughes (2007) contains another application to trivariate data in proteomics concerning χ-angles in Serine and Valine residues (the three angles, then being (ϕ, ψ, χ)). For this data set, 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 Heisenberg distribution (Ising-type model on circle), see, for example, Ellis (1985, p. 131), could be also appropriate to use the neighbourhood information with probability density function proportional to $\exp\{\Sigma a_i^\top x_i + \Sigma a_{ij} x_i^\top x_j\}$, where $a_{ij} = 1$ if i and j are neighbours, and $a_{ij} = 0$ otherwise. Here $x_i^\top = (\cos\theta_i, \sin\theta_i)$ so it leads to the cosine type density. The estimation could be perhaps simplified by using a saddlepoint approximation to the normalizing constant (see, e.g., Kent & Mardia (2006) for the complex Bingham quartic distribution).

An alternative multivariate von Mises distribution is to extend the bivariate cosine model of Mardia, Taylor & Subramaniam (2007). This leads to a multivariate cosine model, which can be defined for $\boldsymbol{\Theta} = (\Theta_1, \dots, \Theta_p)$ by

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

where $-\pi < \theta_j \le \pi$, $-\pi < \mu_j \le \pi$, $\kappa_j \ge 0$, $\delta_{j\ell} \ge 0$. 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 $[\boldsymbol{\Delta}]_{j\ell} = \delta_{j\ell} = \delta_{\ell j}$ and $\delta_{jj} = 0$. The normalizing constant is $C_p^{-1}(\boldsymbol{\kappa}, \boldsymbol{\Delta})$. This model can be investigated in ways similar to those above for the multivariate Sine model. Comparisons analogous to those made by Mardia, Taylor & Subramaniam (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 pseudo-likelihood 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 with Harshinder Singh at the National Institute of Occupational Safety and Health (NIOSH). The authors wish to express special thanks to Dan Sharpe and NIOSH

This content downloaded from
193.205.142.15 on Mon, 05 May 2025 08:33:37 UTC
All use subject to https://about.jstor.org/terms

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 a studentship from the Engineering and Physical Science Research Council.

REFERENCES

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

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

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

R. Ellis (1985). *Entropy, Large Deviations and Statistical Mechanics.* Springer, New York.

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

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

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

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

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

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

K. V. Mardia, G. Hughes & C. C. Taylor (2007). Efficiency of the pseudolikelihood for multivariate normal and von Mises distributions. http://www.maths.leeds.ac.uk/statistics/research/reports/2007/STAT07-02.pdf.

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

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

K. V. Mardia, C. C. Taylor & G. K. Subramaniam (2007). Protein Bioinformatics and Mixtures of Bivariate von Mises Distributions for Angular Data. *Biometrics*, 63, 505–512.

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

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

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

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

Received 1 March 2007
Accepted 29 September 2007

Kanti V. MARDIA: sta6kvm@leeds.ac.uk
Charles C. TAYLOR: c.c.taylor@leeds.ac.uk
Department of Statistics, University of Leeds
Leeds LS2 9JT, United Kingdom

Gareth HUGHES: g.hughes@bristol.ac.uk
Department of Clinical Science at South Bristol
Level 6, UBHT Education Centre, Upper Maudlin Street
Bristol BS2 8AE, United Kingdom

Harshinder SINGH: Deceased February 4, 2005.

This content downloaded from
193.205.142.15 on Mon, 05 May 2025 08:33:37 UTC
All use subject to https://about.jstor.org/terms