Ann. Inst. Statist. Math.
Vol. 57, No. 4, 789–802 (2005)

INFERENCES BASED ON A BIVARIATE DISTRIBUTION WITH VON MISES MARGINALS

GRACE S. SHIEH[1] AND RICHARD A. JOHNSON[2]

[1] *Institute of Statistical Science, Academia Sinica, Taipei 115, Taiwan, R.O.C.,*
e-mail: gshieh@stat.sinica.edu.tw
[2] *Department of Statistics, University of Wisconsin-Madison, WI 53706, U.S.A.,*
e-mail: rich@stat.wisc.edu

(Received April 21, 2004; revised August 24, 2004)

Abstract. There is very little literature concerning modeling the correlation between paired angular observations. We propose a bivariate model with von Mises marginal distributions. An algorithm for generating bivariate angles from this von Mises distribution is given. Maximum likelihood estimation is then addressed. We also develop a likelihood ratio test for independence in paired circular data. Application of the procedures to paired wind directions is illustrated. Employing simulation, using the proposed model, we compare the power of the likelihood ratio test with six existing tests of independence.

Key words and phrases: Angular observations, maximum likelihood estimation, models of dependence, power, testing.

1. Introduction

Testing for independence in paired circular (bivariate angular) data has been explored by many statisticians; see Fisher (1995) and Mardia and Jupp (2000) for thorough accounts. Paired circular data are either both angular or both cyclic in nature. For instance, wind directions at 6 a.m. and at noon recorded at a weather station for 21 consecutive days (Johnson and Wehrly (1977)), estimated peak times for two successive measurements of diastolic blood pressure (Downs (1974)), among others. After independence in circular variables been rejected, modeling circular correlation (*association*) and making inferences become of primary interest.

In this article, we investigate inference procedures under a special case of the bivariate families proposed by Wehrly and Johnson (1980) (see also Johnson and Wehrly (1977)). This case is the first bivariate correlated model with von Mises marginal distributions. The probability density function (pdf) of von Mises is in closed form, and von Mises is parallel to normal distribution for univariate data.

Some literatures related to this article are the following. Mardia (1975*a*, 1975*b*) proposed a bivariate von-Mises model for paired angles (θ_1, θ_2) with probability density functions proportional to

$$\exp\{\kappa_1 \cos(\theta_1 - \mu_1) + \kappa_2 \cos(\theta_2 - \mu_2) + (\cos\theta_1, \sin\theta_1)^T \boldsymbol{A}(\cos\theta_2, \sin\theta_2)\},$$

where $0 \le \kappa_i < \infty$, $0 \le \mu_i < 2\pi$ with $i = 1, 2$ and $\boldsymbol{A}$ is a 2×2 matrix. Although the conditional distributions of θ_1 (θ_2) given θ_2 (θ_1) are von Mises-Fisher, the marginal distributions are not von Mises. Saw (1983) constructed some distributions for dependent unit vectors; Rivest (1988) provided thorough inference procedures for a bivariate generalization of the Fisher-von Mises distribution. Singh *et al.* (2002) replaced the quadratic term and the linear term in the bivariate normal density by their angular analogues and proposed an interesting probabilistic model for bivariate circular variables.

In particular, we investigate inference procedures under the bivariate distribution with von Mises marginals (BVM) having probability density function:

$$f_{12}(\theta_1,\theta_2) = 2\pi f_1(\theta_1)f_2(\theta_2)\frac{1}{2\pi I_0(\kappa_{12})}e^{\kappa_{12}\cos(2\pi[F_1(\theta_1)-F_2(\theta_2)]-\mu_{12})}, \tag{1.1}$$

where $0 \le \theta_1,\ \theta_2 < 2\pi$, $\kappa_{12} \ge 0$, $0 \le \mu_{12} < 2\pi$ and $I_0(\cdot)$ denotes the modified Bessel function of order zero. The marginal densities

$$f_j(\theta_j) = \frac{1}{2\pi I_0(\kappa_j)}e^{\kappa_j\cos(\theta_j-\mu_j)} \quad \text{for} \quad j = 1,2$$

are von Mises with mean μ_j and dispersion parameter κ_j, respectively, and will be denoted by $\mathrm{VM}(\mu_j,\kappa_j)$ henceforth. This joint density involves the cumulative distributions $F_j(\theta_0) = \int_0^{\theta_0} f_j(\theta)d\theta$, but with todays computing power this is not a serious drawback. We note that this class of bivariate circular distribution can include more general marginals, for example *t*-distribution (Shimizu and Iida (2002)). However, the pdf of von Mises distribution is in closed form, so we focus on von Mises marginals here.

In Section 2, we introduce the BVM model and an algorithm to generate the model is given. 2D contours plots of the joint probability of some BVM models are demonstrated. In Section 3, we investigate the maximum likelihood estimation and derive a likelihood ratio test for independence under the BVM model. Next, we apply the model and inference procedures to paired wind directions data in Section 4. Using the BVM model, the power of a likelihood ratio test is compared to six existing tests (Johnson and Shieh (2002)) in Section 5.

2. Some properties of the BVM model

2.1 *Roles of* μ_{12} *and* κ_{12}

We first discuss the copula presentation of the BVM model and then we utilize the copula to interpret the roles of two parameters μ_{12} and κ_{12}. Let (u,v) be random variables on the unit square with density function:

$$f(u,v) = e^{\kappa_{12}\cos(2\pi[u-v]-\mu_{12})}/I_0(\kappa_{12}), \tag{2.1}$$

where $\kappa_{12} \ge 0$ and $0 \le \mu_{12} < 2\pi$. It is easily checked that the marginal distribution of U and V are uniform so that $f(u,v)$ is the density function of a copula. Let F_1 and F_2 denote von Mises marginal distributions, then $(F^{-1}(u), F^{-1}(v))$ has the proposed BVM joint density function. From (2.1), it is clear that $2\pi(u-v)$ follows a $\mathrm{VM}(\mu_{12},\kappa_{12})$. Thus $\Phi_1 = 2\pi F_1(\Theta_1)$ given θ_2 is $\mathrm{VM}(\mu_{12}+2\pi F_2(\theta_2),\kappa_{12})$. When $\mu_{12} = 0$, Φ_1 centers on $2\pi F_2(\theta_2)$, and the dependence of Θ_1 on Θ_2 is through the magnitude of κ_{12}. The conditional density of $\Theta_1 \mid \Theta_2 = \theta_2$ is given in Subsection 2.2.

The proposed model does reduce to the bivariate normal distribution as fluctuations in θ_1 and θ_2 are sufficiently small provided that $\mu_{12} = 2\pi[F_1(\mu_2) - F_2(\mu_2)]$ and f_i is continuous at μ_i for $i = 1,2$. With approximations $\cos(\theta_i - \mu_i) \approx 1 - (\theta_i - \mu_i)^2/2$ and

$\sin(\theta_i - \mu_i) \approx (\theta_i - \mu_i)$ for $i = 1, 2$, the proposed joint pdf reduces to $C\exp(\kappa_1 + \kappa_2 + \kappa_{12} - \frac{1}{2}\{[\kappa_1 + \kappa_{12}/I_0^2(\kappa_1)](\theta_1 - \mu_1)^2 + [\kappa_2 + \kappa_{12}/I_0^2(\kappa_2)](\theta_2 - \mu_2)^2 + 2\kappa_{12}/[I_0(\kappa_1)I_0(\kappa_2)](\theta_1 - \mu_1)(\theta_2 - \mu_2)\})$, where C is a normalizing constant. The parameters of the approximating bivariate normal correspond to those of f_{12} as follows:

$$\rho_{12} = -\kappa_{12}\{[\kappa_{12} + \kappa_1 I_0^2(\kappa_1)][\kappa_{12} + \kappa_2 I_0^2(\kappa_2)]\}^{-1/2},$$
$$\sigma_1^2 = [\kappa_{12} + \kappa_2 I_0^2(\kappa_2)]\{\kappa_1\kappa_2 I_0^2(\kappa_2) + \kappa_{12}[\kappa_1 + \kappa_2 I_0^2(\kappa_2)/I_0^2(\kappa_1)]\}^{-1}$$

and

$$\sigma_2^2 = [\kappa_{12} + \kappa_1 I_0^2(\kappa_1)]\{\kappa_1\kappa_2 I_0^2(\kappa_1) + \kappa_{12}[\kappa_2 + \kappa_1 I_0^2(\kappa_1)/I_0^2(\kappa_2)]\}^{-1}.$$

The bivariate normal is unimodal thus a necessary condition for unimodality of the proposed jpdf is $\mu_{12} = 2\pi[F_1(\mu_1) - F_2(\mu_2)]$. After computing the first and second derivatives of f_{12}, we applied the unimodality condition and test for extrema to show that (μ_1, μ_2) is the mode. Setting the first derivatives $\partial f_{12}/\partial\theta_1 = -f_{12}(\kappa_1 \sin(\theta_1 - \mu_1) + \kappa_{12}\sin\{2\pi[F_1(\theta_1) - F_2(\theta_2)] - \mu_{12}\} \times 2\pi f_1)$ and $\partial f_{12}/\partial\theta_2 = -f_{12}(\kappa_2 \sin(\theta_2 - \mu_2) - \kappa_{12}\sin\{2\pi[F_1(\theta_1) - F_2(\theta_2)] - \mu_{12}\} \times 2\pi f_2)$ to zero, we had the solutions (μ_1, μ_2) and $(\mu_1 + \pi, \mu_2 + \pi)$ (mod 2π) under the unimodality condition. Since $\partial^2 f_{12}/\partial\theta_1^2|_{(\mu_1,\mu_2)} = -f_{12}[\kappa_1 + \kappa_{12}/I_0^2(\kappa_1)] < 0$ and $\{\partial^2 f_{12}/\partial\theta_1^2 \partial^2 f_{12}/\partial\theta_2^2 - [\partial^2 f_{12}/\partial\theta_1\partial\theta_2]^2\}\,|_{(\mu_1,\mu_2)} = f_{12}^2[\kappa_1\kappa_2 + \kappa_{12}(2\pi)^2(\kappa_2 f_1^2 + \kappa_1 f_2^2)] > 0$, (μ_1, μ_2) is the mode. Similarly, we obtained that $(\mu_1 + \pi, \mu_2 + \pi)$ is the minimum.

Our numerical calculations suggest that the joint pdf of Θ_1 and Θ_2 is uni-modal provided that $\mu_{12} = \mu_1 - \mu_2$ which is consistent with the aforementioned condition since F_1 and F_2 are uniform on $[0, 1]$. If $\mu_{12} = 0$, then $\mu_1 = \mu_2$ is required for the unimodality of f_{12}. Thus μ_{12} is also a shape parameter beside κ_1, κ_2 and κ_{12}. 2D contours of a BVM with fixed $\mu_1 = \pi$, $\mu_2 = 0$, $\kappa_1 = \kappa_2 = 3.0$ and $\kappa_{12} = 4.0$, and with μ_{12} varying from 0 (bi-modal) to π (uni-modal) are illustrated in Fig. 1, respectively.

The parameter κ_{12} models the circular correlation of Θ_1 and Θ_2. This is illustrated by the 2D contours of the jpdf of BVM$(\pi, 3.0, \pi, 3.0, 0, \kappa_{12})$ in Fig. 2. As κ_{12} increases from 0, 1, 4 to 7, the circular association of Θ_1 and Θ_2 increases since the 2D contours of the jpdf get narrower and center around the 45° line. There is no unique theoretical definition of association for bivariate circular random variables. Instead, a few variety of circular-circular association estimates based on samples were proposed, for instance r_1 to r_6 in Section 5. Thus to illustrate the relationship between κ_{12} and the association

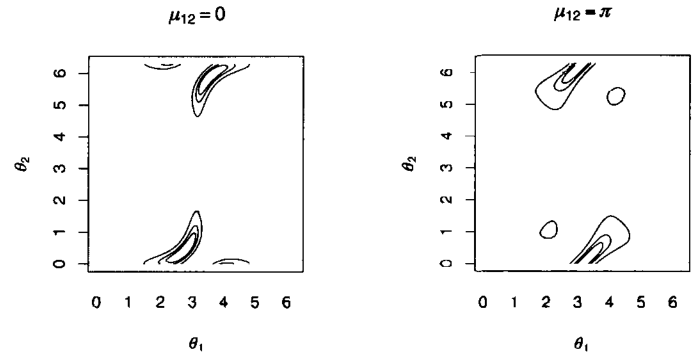

Fig. 1. Joint pdf of θ_1 and θ_2 from BVM$(\pi, 3.0, 0, 3.0, \mu_{12}, 4)$.

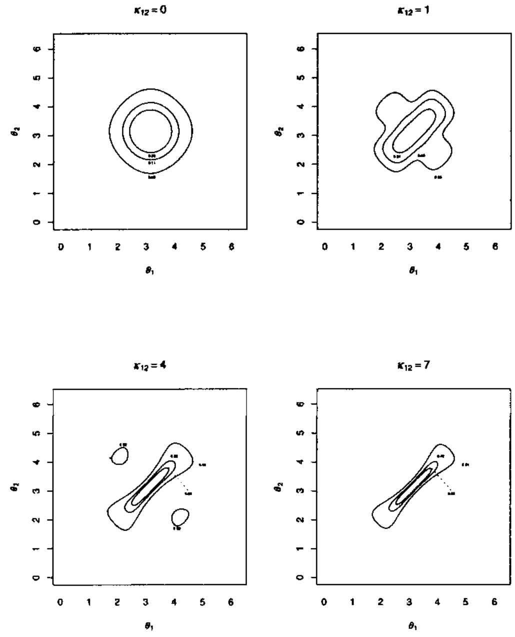

Fig. 2. Joint pdf of θ_1 and θ_2 from BVM$(\pi, 3.0, \pi, 3.0, 0, \kappa_{12})$.

between Θ_1 and Θ_2, we have simulated 1,000 copies of r_1 to r_6 under the BVM model with $\mu_1 = \mu_2 = \pi$, $\mu_{12} = 0$ and $\kappa_1 = \kappa_2 = 1.5$. Noticeably, the average of 1,000 r_5's increases from 0.00 (rounded to 2 decimal place), 0.12, 0.24, 0.57 to 0.69 when κ_{12} varies from 0, 0.5, 1, 2 to 3; the other five association measures also show the same increasing trend.

2.2 *BVM random vector generator*

We suppress the parameters and write $f_i(\theta_i)$ and $F_i(\theta_i)$, $i = 1, 2$ for the marginal density functions and cumulative distribution functions, respectively. A simple integration verifies that $f_{12}(\theta_1, \theta_2)$ does indeed have these marginal distributions. Consequently, the conditional density of $\Theta_1 \mid \Theta_2 = \theta_2$ is

$$f_{1|2}(\theta_1 \mid \theta_2) = 2\pi f_1(\theta_1) \frac{1}{2\pi I_0(\kappa_{12})} e^{\kappa_{12}\cos(2\pi[F_1(\theta_1) - F_2(\theta_2)] - \mu_{12})}.$$

Under the BVM distribution, $\Phi_1 = 2\pi F_1(\Theta_1)$, given θ_2 has a von Mises distribution with κ_{12} and location parameter $\mu_{12} + 2\pi F_2(\theta_2)$. This fact helps us generate pairs of angles (Θ_1, Θ_2) having the joint density function $f_{12}(\cdot, \cdot)$. In particular,

1. Generate V_2 as von Mises$(0, \kappa_2)$ and set $\Theta_2 = V_2 + \mu_2 \pmod{2\pi}$.
2. Generate V_1 as von Mises$(0, \kappa_{12})$ and, given $\Theta_2 = \theta_2$, set

$$W_1 = V_1 + \mu_{12} + 2\pi F_2(\theta_2) \pmod{2\pi}$$

so W_1 has the conditional distribution of $2\pi F_1(\Theta_1)$.

3. Finally, set

$$\Theta_1 = F_1^{-1}\left(\frac{W_1}{2\pi}\right).$$

Then, (Θ_1, Θ_2) are distributed as f_{12}.

More generally, we can make the transformation $\boldsymbol{\theta} \rightarrow \boldsymbol{\beta}$.

$$\beta_1 = \theta_1$$
$$\beta_2 = 2\pi(F_1(\theta_1) - F_2(\theta_2))$$

which has Jacobian $(-2\pi f_2(\theta_2))$. Consequently, the joint density is

$$\frac{1}{2\pi I_0(\kappa_1)} e^{\kappa_1 \cos(\beta_1 - \mu_1)} \frac{1}{2\pi I_0(\kappa_{12})} e^{\kappa_{12} \cos(\beta_2 - \mu_{12})}.$$

Since the joint density of β_1 and β_2 factor into two parts, we conclude that Θ_1 is independent of $2\pi(F_1(\Theta_1) - F_2(\Theta_2))$ which has a von Mises (κ_{12}, μ_{12}) distribution. Similarly, we can show that Θ_2 is independent of $2\pi(F_1(\Theta_1) - F_2(\Theta_2))$. A BVM random vector generator written in Fortran 77 can be downloaded from `http://www.stat.sinica.edu.tw/~gshieh/`.

The following is an algorithm to generator a BVM$(\mu_1, \kappa_1, \mu_2, \kappa_2, \mu_{12}, \kappa_{12})$ random vectors.

Step 1. We approximate the area under the von Mises density function (a curve) by a polynomial of certain degree, using the `gauleg` function in Press (1999), in each small interval $(\frac{(i-1)}{N}2\pi, \frac{i}{N}2\pi)$, where $N = 500$ and $i = 1, \ldots, N$. We note that N can be other large constant.

Step 2. We tabulate all approximate cumulative areas under a VM$(0, \kappa_2)$ density and denote them $F_{2N}(\frac{i}{500}2\pi)$, $i = 1, \ldots, 500$.

Next, we generate a random uniform variable from $(0,1)$, and denote it U_2. When the generated u_2 in the interval $[F_{2N}(\frac{i-1}{500}2\pi), F_{2N}(\frac{i}{500}2\pi))$, we linearly interpolate between $\frac{i-1}{500}2\pi$ and $\frac{i}{500}2\pi$ to get a corresponding angle v_2. Then we can shift the mean of V_2 to μ_2 by $V_2 + \mu_2 \pmod{2\pi}$ and denote it by θ_2.

Step 3. Similarly, we can generate another von Mises random variable V_1 which follows VM(μ_{12}, κ_{12}). Then, convoluting V_1 and $2\pi F_{2N}(\theta_2) \pmod{2\pi}$,

$$W_1 = V_1 + 2\pi F_{2N}(\theta_2) \pmod{2\pi}.$$

Recall that F_{2N} is the tabulated cumulative area under a VM$(0, \kappa_2)$ from Step 2.

Step 4. Similar to Step 2, for $i = 1, \ldots, 500$ we tabulate $F_{1N}(\frac{i}{500}2\pi)$ which denotes the approximate cumulative area under a VM$(0, \kappa_1)$. Finally, we invert $W_1/(2\pi)$ according to the tabled values of $F_{1N}(\cdot)$ to get θ_1.

2.3 *2D contours and other properties*

Contour plots of the joint density provide insight into the manner in which the parameter κ_{12} indexes the amount of probability that concentrates along the curve

$$2\pi(F_1(\theta_1 \mid \kappa_1, \mu_1) - F_2(\theta_2 \mid \kappa_2, \mu_2)) = \mu_{12} \pmod{2\pi}.$$

When the marginal distributions are identical, probability concentrates along the line $\theta_1 = \theta_2$ when $\mu_{12} = 0$.

For the case of identical marginal distributions, we have drawn the contours on $[0, 2\pi) \times [0, 2\pi)$ for the values 0, 1, 4 and 7 of κ_{12}. We have selected $\mu_1 = \mu_2 = \pi$ so the peak is in the center of the region. We have selected $\mu_{12} = 0$ since f_{12} takes its maximum at $\mu_{12} = 0$ when $F_1 = F_2$. The contours in Fig. 2 pertain to marginal distributions having the common $\kappa_1 = \kappa_2 = 3.0$. The approximate 50%, 70% and 90% contours nearby which their corresponding probability labeled are shown. Going from independence $\kappa_{12} = 0$ to $\kappa_{12} = 1$, we see how the circular contours distort somewhat. The densities for the intermediate values between 1 and 6 have some small secondary modes. At $\kappa_{12} = 7$ and beyond we see the concentration along a line. The contours are clearly symmetric about the line $\theta_1 = \theta_2$ and the line $2\pi - \theta_1 = \theta_2$. The contours for the increasing values $\kappa_{12} = 0$, 1, 4 and 7 are also symmetric about these two lines. Increasing κ_{12} concentrates the probability about the equal angular line. 2D contour plots instead of 3D perspective plots are illustrated since contour plots indicate confidence regions more clearly.

3. Estimation and testing

Although some of the parameters enter the joint density in a rather complicated manner, modern maximization programs enable us to easily obtain the maximum likelihood estimates (MLEs). We applied the FORTRAN code FFSQP, Version 3.7 by Zhou *et al.* (1997) to obtain the MLEs of unknown parameters in model (1.1) and call the program the MLE algorithm henceforth. Then, estimates of their variances and covariances can be obtained from derivatives of the likelihood or by numerical integration of the squares and products of the partial derivatives given below.

The code FFSQP does not prevent the estimates from being trapped in local maxima, thus we have applied a technique which is popular in the artificial intelligence community to get an approximate global maximum. To obtain the MLEs of a k-dimensional ($k > 1$) parameter, we first partition the range of each parameter coordinate 2π into m equal intervals. Next, we repeat the algorithm N (a big number) times; at each time one starts the algorithm with a different initial vector value. The initial value is sampled from N cubes out of the total m^k cubes. A global maximum is equal to the maximum of the N local ones. We further repeat the above process with N_1 ($> N$) different initial values. If the two global maxima obtained from the two experiments (with N and N_1 different initial values) converge, then we stop and declare that the approximate global maximum obtained. An alternative way to estimate these six parameters is to estimate μ_i and κ_i for $i = 1, 2$ from the marginals since the likelihood function given μ_i and κ_i $i = 1, 2$ is concave in μ_{12} and κ_{12}. Then one plugs the estimates in the joint likelihood function to estimate μ_{12} and κ_{12} as proposed in Subsection 3.3 of Rivest (1988). The latter algorithm reduces this optimization problem in R^6 to three optimization problems in R^2 and hence is more efficient.

We write $\boldsymbol{\eta} = (\kappa_1, \mu_1, \kappa_2, \mu_2, \kappa_{12}, \mu_{12})'$ for the vector of six unknown parameters. The parameter space is then

$$\boldsymbol{\Omega} = \{0 \le \kappa_1 < \infty, 0 \le \mu_1 < 2\pi, 0 \le \kappa_2 < \infty, 0 \le \mu_2 < 2\pi,\\ 0 \le \kappa_{12} < \infty, 0 \le \mu_{12} < 2\pi\}.$$

Let $l_n(\boldsymbol{\eta})$ denote the log-likelihood function, $\boldsymbol{U}_n(\boldsymbol{\eta})$ and $-\boldsymbol{I}_n(\boldsymbol{\eta})$ the first and the second derivatives of $l_n(\boldsymbol{\eta})$, respectively. Furthermore, we denote the neighborhood of the true parameter $\boldsymbol{\eta}_0$ by $B(\boldsymbol{\eta}_0)$. The regularity conditions in Self and Liang (1987) imply the root-n consistency of the MLEs. We restate the regularity conditions as follows:

(i) almost sure existence of the first three derivative of $l_n(\boldsymbol{\eta})$ with respect to $\boldsymbol{\eta}$ on the intersection of $B(\boldsymbol{\eta}_0)$ and $\boldsymbol{\Omega}$,

(ii) on the intersection of $B(\boldsymbol{\eta}_0)$ and $\boldsymbol{\Omega}$, n^{-1} times the absolute value of the third derivatives of $l_n(\boldsymbol{\eta})$ is bounded by a function $f(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$, where $Ef(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) < \infty$, $\boldsymbol{X}_i = (\theta_{1i}, \theta_{2i})^T$ and

(iii) $E(\boldsymbol{I}_n(\boldsymbol{\eta})/n) = \boldsymbol{I}(\boldsymbol{\eta})$ is positive definite on $B(\boldsymbol{\eta}_0)$ and $\boldsymbol{I}(\boldsymbol{\eta}_0)$ is equal to the variance-covariance matrix of $n^{-1/2}\boldsymbol{U}_n(\boldsymbol{\eta}_0)$.

When the vector of parameters $\boldsymbol{\eta}$ belongs to the interior of the parameter set, the regularity conditions hold and the consistency of the MLEs follows. Then asymptotic multivariate normality of the maximum likelihood estimators follows from Lemma 1 and Theorem 2 of Self and Liang (1987).

THEOREM 3.1. *Let $\boldsymbol{\eta}$ belong to the interior of the parameter set. The maximum likelihood estimator, $\hat{\boldsymbol{\eta}}$, is asymptotically multivariate normal, and*

$$\sqrt{n}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}) \text{ converges in distribution } N_6(\mathbf{0}, \boldsymbol{I}^{-1}),$$

where $\boldsymbol{I}$ is the Fisher information matrix with entries

$$I_{ik}(\boldsymbol{\eta}) = E\left[\frac{\partial}{\partial \eta_i} \ln f_{12}(\Theta_1, \Theta_2 \mid \boldsymbol{\eta}) \frac{\partial}{\partial \eta_k} \ln f_{12}(\Theta_1, \Theta_2 \mid \boldsymbol{\eta})\right].$$

Let $A(\kappa) = I_0'(\kappa)/I_0(\kappa)$. The six partial derivatives are straightforward and are omitted.

Remark. If some of the parameters are known, then asymptotic normality holds for the reduced set with the corresponding entries in the information matrix.

Numerical integration can be used to obtain each of the entries $\boldsymbol{I}$. However, this could be difficult and it is better to use the estimated, or empirical, information with terms

$$\frac{1}{n}\sum_{j=1}^{n} \frac{\partial}{\partial \eta_i} \ln f_{12}(\theta_{1j}, \theta_{2j} \mid \boldsymbol{\eta}) \frac{\partial}{\partial \eta_k} \ln f_{12}(\theta_{1j}, \theta_{2j} \mid \boldsymbol{\eta}),$$

where each partial derivative is evaluated at the maximum likelihood estimator.

Different asymptotics apply if we include $\kappa_{12} = 0$ in the parameter space. If κ_1 and κ_2 are still bounded away from 0, or known, then according to the results in Self and Liang (1987), Theorem 2 and Case 2 of p. 606.

$\sqrt{n}\hat{\kappa}_{12}$ has limiting distribution $Z_1 I[Z_1 > 0]$ where Z_1 has a normal distribution with variance determined from $\boldsymbol{I}^{-1}$. The cases where $\kappa_1 = 0$ and/or $\kappa_2 = 0$ are treated similarly.

We next consider likelihood ratio tests. There are really three cases, each progressively more complicated. In all cases, the test for independence is equivalent to testing

$$H_0 : \kappa_{12} = 0 \quad \text{versus} \quad H_1 : \kappa_{12} > 0.$$

Case 1. The marginal distributions are known. Let L be the likelihood, the product of the joint pdf's, assuming that μ_1 and μ_2 are known, then

$$L_{H_1}/L_{H_0} = [I_0(\kappa_{12})]^{-n} e^{\kappa_{12} \sum_i \cos(2\pi[F_1(\theta_{1i}) - F_2(\theta_{2i})] - \mu_{12})}.$$

The likelihood ratio test, denoted by $-2\ln\lambda_n$, is based on

$$\sum_i \cos(2\pi[F_1(\theta_{1i}) - F_2(\theta_{2i})] - \hat{\mu}_{12}),$$

where the maximum likelihood estimator $\hat{\mu}_{12} = 2\pi[\overline{F_1} - \overline{F_2}]$. If μ_{12} is also known, this is a uniformly most powerful test since we are then in the exponential family case. Otherwise, since the marginals are known, we treat $2\pi[F_1(\theta_{1i}) - F_2(\theta_{2i})] = \phi_i$ as a single angle. From the form of $-2\ln\lambda_n$, this test can be viewed as the maximum likelihood test for the concentration parameter in a one-sample case (p. 126 of Mardia and Jupp (2000)) and it reduces to

$$2n[\hat{\kappa}_{12}\overline{R}_\phi - \ln I_0(\hat{\kappa}_{12})],$$

where $\overline{R}_\phi = R_\phi/n$, the resultant $R_\phi = \sqrt{C_\phi^2 + S_\phi^2}$, $C_\phi = \sum_{i=1}^{\infty} \cos\phi_i$ and $S_\phi = \sum_{i=1}^{\infty} \sin\phi_i$.

We note that the null hypothesis raises the boundary problem, and the limit distribution of $-2\ln\lambda_n$ is not that of χ_2^2. Instead, according to Self and Liang (1987), the correct limiting null distribution of $-2\ln\lambda_n$ is that of the random variable $Z^2 I[Z > 0]$, where Z has a standard normal distribution and $I(A)$ is an indicator function.

Case 2. Von Mises marginals with location parameters μ_1, μ_2 and μ_{12} known. The likelihood ratio test is denoted by $-2\ln\lambda_n$, where

$$\lambda_n = \frac{\sup_{\kappa_1,\kappa_2} \prod_i f_1(\theta_{1i} \mid \kappa_1) f_2(\theta_{2i} \mid \kappa_2)}{\sup_{\kappa_1,\kappa_2,\kappa_{12}} \prod_i f_{12}(\theta_{1i}, \theta_{2i} \mid \kappa_1, \kappa_2, \kappa_{12})},$$

provided that κ_1 and κ_2 are not 0. The limiting null distribution of $-2\ln\lambda_n$ is again that of $Z^2 I[Z > 0]$.

Case 3. All six parameters of the distribution are unknown. We do the full likelihood ratio test with numerator of λ_n

$$\sup_{\kappa_1,\kappa_2,\mu_1,\mu_2} \prod_i f_1(\theta_{1i} \mid \kappa_1, \mu_1) f_2(\theta_{2i} \mid \kappa_2, \mu_2)$$

and denominator of λ_n

$$\sup_{\kappa_1,\kappa_2,\kappa_{12},\mu_1,\mu_2,\mu_{12}} \prod_i f_{12}(\theta_{1i}, \theta_{2i} \mid \kappa_1, \kappa_2, \kappa_{12}, \mu_1, \mu_2, \mu_{12}).$$

There are three sub-cases: (a) when $\kappa_1 \neq 0$ and $\kappa_2 \neq 0$, by Case 5 in Theorem 3 of Self and Liang (1987), the limiting null distribution of $-2\ln\lambda_n$ is that of $Z^2 I[Z > 0]$, (b) when $\kappa_1 = 0$ or $\kappa_2 = 0$ but $\kappa_1 \neq \kappa_2$, then the distribution is $(1/2)F_{\chi_1^2} + (1/2)F_{\chi_2^2}$ (Case 6 of Theorem 3 in Self and Liang (1987)), and (c) when both $\kappa_1 = 0$ and $\kappa_2 = 0$, then the distribution is

$$(1/8)\sum_{k=0}^{3}\binom{3}{k}F_{\chi^2_{k+3}},$$

where $F_{\chi^2_k}$ denotes the cdf of random variable χ^2_k (Case 9 of Theorem 3 in Self and Liang (1987)).

4. An application

In this section, we apply our model and algorithm to paired wind directions in Johnson and Wehrly (1977). The wind directions at 6 a.m. and 12 noon were measured each day at a weather station in Milwaukee for 21 consecutive days. The wind directions in degrees are listed in Table 1 and plotted as dots in Fig. 3. We denote the wind directions at 6 a.m. by Θ_1 and those at noon Θ_2.

Testing for independence in Θ_1 and Θ_2 is equivalent to testing $H_0 : \kappa_{12} = 0$ under model (1.1). We apply the likelihood ratio test in Section 3 (Case 3), with 4 parameters under the null hypothesis and 6 parameters under the alternative. The likelihood ratio test statistic $(-2\ln\lambda_n)$ equals 11.3. The test is significant against both 99% critical values of Case 3(a) $(2.33)^2$ (assuming $\kappa_i \neq 0$, $i = 1, 2$). Thus the circular association between Θ_1 and Θ_2 is significant at 99% significance level. This agrees with the results

Table 1. Wind directions at 6 a.m. (θ_1) and at noon (θ_2).

θ_1	356	97	211	232	343	292	157	302	335	302	324
θ_2	119	162	221	259	270	29	97	292	40	313	94
θ_1	85	324	340	157	238	254	146	232	122	329	
θ_2	45	47	108	221	270	119	248	270	45	23	

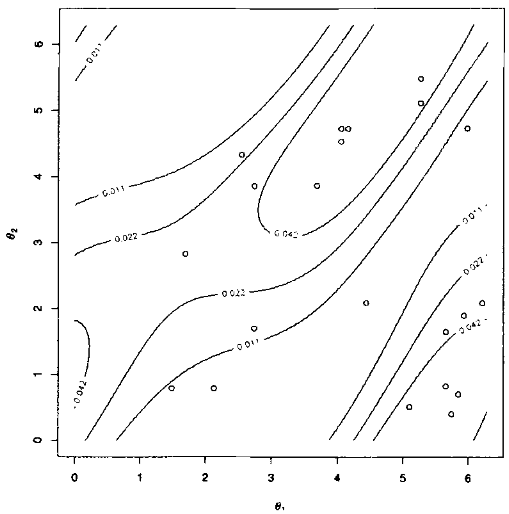

Fig. 3. 2D Contour plots of BVM(4.8, 0.6, 4.6, 0.2, 5.1, 1.2).

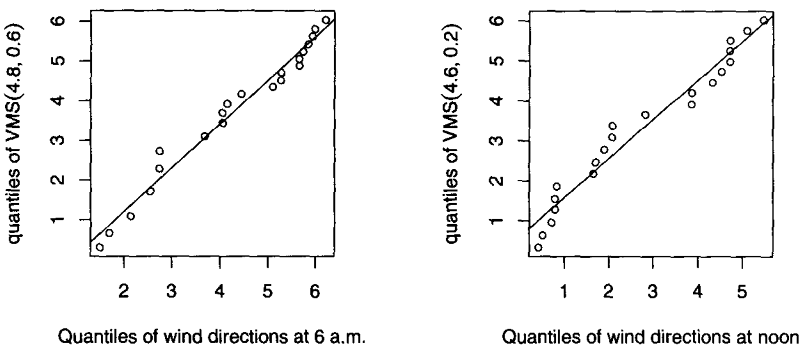

Fig. 4. QQ-plots of θ_1 and θ_2 versus von Mises quantiles.

of applying *circular Kendall's tau* and *circular Spearman's rho* in p. 149 of Fisher (1995). Next question of interest is: which BVM model do these wind directions fit?

We have fitted the data by the MLE algorithm with all six parameters unknown in (1.1). The algorithm was run twice where the first run used 1,000 and the second 4,096 different initial vector values. The MLEs obtained by the two experiments converge at the first decimal place. Thus we do not need to execute the algorithm at more initial values. The *approximate* MLEs obtained are $\hat{\mu}_1 = 4.8$, $\hat{\kappa}_1 = 0.6$, $\hat{\mu}_2 = 4.6$, $\hat{\kappa}_2 = 0.2$, $\hat{\mu}_{12} = 5.1$ and $\hat{\kappa}_{12} = 1.2$. We plot the wind directions data on the approximate 50%, 70% and 90% 2D contours of the jpdf of BVM(4.8, 0.6, 4.6, 0.2, 5.1, 1.2) in Fig. 3. The data seem to fit the BVM model well. QQ-plots of quantiles of Θ_1 versus VM(4.8, 0.6) and quantiles of Θ_2 versus VM(4.6, 0.2) in Fig. 4 also show that it is reasonable to assume that the marginals are von Mises.

5. Empirical power study

We consider tests of independence under alternatives $\kappa_{12} > 0$. Powers of the likelihood ratio test in Case 2 of Section 4, that use both the empirical critical value (from 5,000 simulations, denoted by r_{7E}) and the theoretical critical value of $Z^2 I[Z > 0]$ at the 95% level (r_7), are compared with powers of the other six tests considered in Johnson and Shieh (2002). This likelihood ratio test is expected to do best since some parameters are assumed to be known. The general case, Case 3 of the likelihood ratio test, took too long to simulate adequately.

Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ and $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ be two random samples of observations on two p-dimensional spheres ($S^{p-1} \times S^{p-1}$), respectively. The following tests have been proposed for testing independence of bivariate p-dimensional data. Here, we apply them to bivariate circular data (the case that $p = 2$). We note that $\boldsymbol{x}_i = (\cos(\theta_{1i}), \sin(\theta_{1i}))^T$, for $i = 1, \ldots, n$, $\overline{\boldsymbol{x}} = \sum_{i=1}^n \boldsymbol{x}_i/n$, $\boldsymbol{y}_i = (\cos(\theta_{2i}), \sin(\theta_{2i}))^T$ and $\overline{\boldsymbol{y}} = \sum_{i=1}^n \boldsymbol{y}_i/n$.

Stephens (1979), following Mackenzie (1957), proposed

$$r_1 = \max_{\boldsymbol{Q}} \frac{\boldsymbol{x}_i' \boldsymbol{Q} \boldsymbol{y}_i}{n} = \sum_{i=1}^{p} \sqrt{\gamma_i}, \tag{5.1}$$

where $\boldsymbol{Q}$ is an orthogonal matrix and $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_p > 0$ are the eigenvalues of $n^{-1}\boldsymbol{X}\boldsymbol{Y}'\boldsymbol{Y}\boldsymbol{X}'$, where $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n]$ and $\boldsymbol{Y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n]$. If $\boldsymbol{Q}$ is restricted to be a rotation so $\det\{\boldsymbol{Q}\} = 1$, the statistics is denoted by r_1^+.

$$r_1^+ = \begin{cases} r_1, & \text{if } \det\{\boldsymbol{X}\boldsymbol{Y}'\} > 0, \\ \sqrt{\gamma_1} + \sqrt{\gamma_2} + \cdots + \sqrt{\gamma_{p-1}} - \sqrt{\gamma_p}, & \text{otherwise.} \end{cases} \tag{5.2}$$

Fisher and Lee (1983, 1986) proposed the statistic

$$r_2 = \frac{\sum_{(p)} \det[\boldsymbol{x}_{i_1}, \ldots, \boldsymbol{x}_{i_p}] \det[\boldsymbol{y}_{i_1}, \ldots, \boldsymbol{y}_{i_p}]}{\sqrt{\sum_{(p)} \det^2[\boldsymbol{x}_{i_1}, \ldots, \boldsymbol{x}_{i_p}] \sum_{(p)} \det^2[\boldsymbol{y}_{i_1}, \ldots, \boldsymbol{y}_{i_p}]}}, \tag{5.3}$$

where $\sum_{(p)}$ denotes the sum over $1 \le i_1 \le \cdots < i_p \le n$. Other well known statistics depend on the centered observations $\boldsymbol{x}_i - \overline{\boldsymbol{x}}$ and $\boldsymbol{y}_i - \overline{\boldsymbol{y}}$ through the usual sample covariance matrix $\hat{\boldsymbol{\Sigma}}$.

Johnson and Wehrly (1977) introduced the statistic

$$r_3 = \lambda_1(\hat{\boldsymbol{\Sigma}}_{11}^{-1}\hat{\boldsymbol{\Sigma}}_{12}\hat{\boldsymbol{\Sigma}}_{22}^{-1}\hat{\boldsymbol{\Sigma}}_{21}), \tag{5.4}$$

which is the largest eigenvalue of the sample covariance matrix of the centered vectors. Jupp and Mardia (1980, 1989) considered

$$r_4 = \mathrm{tr}(\hat{\boldsymbol{\Sigma}}_{11}^{-1}\hat{\boldsymbol{\Sigma}}_{12}\hat{\boldsymbol{\Sigma}}_{22}^{-1}\hat{\boldsymbol{\Sigma}}_{21}), \tag{5.5}$$

where $\mathrm{tr}(A)$ denotes the trace of matrix A. Johnson and Shieh (2002) study the centered version of r_2,

$$r_5 = \frac{\sum_{(p)} \det[\boldsymbol{x}_{i_1} - \overline{\boldsymbol{x}}, \ldots, \boldsymbol{x}_{i_p} - \overline{\boldsymbol{x}}] \det[\boldsymbol{y}_{i_1} - \overline{\boldsymbol{y}}, \ldots, \boldsymbol{y}_{i_p} - \overline{\boldsymbol{y}}]}{\sqrt{\sum_{(p)} \det^2[\boldsymbol{x}_{i_1} - \overline{\boldsymbol{x}}, \ldots, \boldsymbol{x}_{i_p} - \overline{\boldsymbol{x}}] \sum_{(p)} \det^2[\boldsymbol{y}_{i_1} - \overline{\boldsymbol{y}}, \ldots, \boldsymbol{y}_{i_p} - \overline{\boldsymbol{y}}]}}. \tag{5.6}$$

Let $\mathrm{sgn}(x) = 1, 0$ or -1 if $x >, =$ or < 0. Finally, we include the rank statistic

$$r_6 = \binom{n}{3}^{-1} \sum_{1 \le i < j < k \le n} \mathrm{sgn}(\theta_{1i} - \theta_{1j})\,\mathrm{sgn}(\theta_{1j} - \theta_{1k})\,\mathrm{sgn}(\theta_{1k} - \theta_{1i}) \\ \times \mathrm{sgn}(\theta_{2i} - \theta_{2j})\,\mathrm{sgn}(\theta_{2j} - \theta_{2k})\,\mathrm{sgn}(\theta_{2k} - \theta_{2i})$$

developed by two different arguments (see Fisher and Lee (1982), and Shieh *et al.* (1994)).

Critical values of r_1 to r_6 have been estimated by generating 5,000 simulations of 50 and 100 pairs of angles under model (1.1) with $\mu_1 = \mu_2 = \pi$, $\mu_{12} = 0$, and both κ_1 and κ_2 varying from 0.5, 1.5 to 3.0.

Large values of κ_{12} correspond to strong dependence between θ_1 and θ_2 and hence large power of each test. We increased κ_{12} from 0.25 to 5.00 to obtain power in each case. All cases simulated show that, as expected, the likelihood ratio test is the most powerful among all tests in both sample sizes studied. Due to space limit, we only present result of sample size 50 (Tables 2–4); results of sample size 100 can be obtained from the corresponding author.

When both θ_1 and θ_2 are concentrated around their means ($\kappa_1 = \kappa_2 \ge 1.5$) and $\kappa_{12} \ge 0.75$, r_5 is more powerful than its uncentered version r_2 except for one case in Table 3 with $\kappa_{12} = 0.75$. Whereas in all other cases, r_2 is more powerful than r_5. With

Table 2. Empirical powers of some tests for independence with data simulated from BVM($\pi, 0.5, \pi, 0.5, 0, \kappa_{12}$) and sample size 50.

κ_{12}	r_1	r_1^+	r_2	r_3	r_4	r_5	r_6	r_{7E}	r_7
0.00	0.050	0.050	0.050	0.051	0.050	0.050	0.051	0.050	0.054
0.25	0.173	0.233	0.224	0.114	0.126	0.151	0.121	0.334	0.343
0.50	0.530	0.643	0.599	0.347	0.420	0.500	0.367	0.788	0.797
0.75	0.859	0.913	0.883	0.682	0.777	0.828	0.710	0.973	0.975
1.00	0.979	0.990	0.981	0.912	0.965	0.970	0.923	0.998	0.998
2.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
5.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 3. Empirical powers of some tests for independence with data simulated from BVM($\pi, 1.5, \pi, 1.5, 0, \kappa_{12}$) and sample size 50.

κ_{12}	r_1	r_1^+	r_2	r_3	r_4	r_5	r_6	r_{7E}	r_7
0.00	0.051	0.050	0.050	0.050	0.050	0.050	0.051	0.050	0.053
0.25	0.115	0.165	0.166	0.108	0.106	0.116	0.121	0.331	0.339
0.50	0.256	0.360	0.373	0.271	0.287	0.314	0.367	0.790	0.797
0.75	0.483	0.597	0.617	0.514	0.558	0.591	0.710	0.975	0.976
1.00	0.699	0.789	0.798	0.731	0.795	0.803	0.923	0.999	0.999
2.00	0.987	0.994	0.993	0.996	1.000	0.998	1.000	1.000	1.000
3.00	0.999	0.999	0.999	1.000	1.000	1.000	1.000	1.000	1.000
5.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 4. Empirical powers of some tests for independence with data simulated from BVM($\pi, 3.0, \pi, 3.0, 0, \kappa_{12}$) and sample size 50.

κ_{12}	r_1	r_1^+	r_2	r_3	r_4	r_5	r_6	r_{7E}	r_7
0.00	0.050	0.050	0.050	0.051	0.050	0.050	0.051	0.050	0.053
0.25	0.091	0.115	0.127	0.101	0.099	0.104	0.121	0.329	0.337
0.50	0.154	0.214	0.246	0.207	0.220	0.231	0.367	0.790	0.798
0.75	0.272	0.348	0.417	0.390	0.417	0.427	0.710	0.976	0.977
1.00	0.404	0.488	0.570	0.569	0.615	0.622	0.923	0.999	0.999
2.00	0.745	0.806	0.873	0.944	0.971	0.959	1.000	1.000	1.000
3.00	0.879	0.912	0.949	0.995	0.999	0.992	1.000	1.000	1.000
5.00	0.955	0.969	0.988	1.000	1.000	0.999	1.000	1.000	1.000

κ_{12} fixed, increasing both κ_1 and κ_2, does not change the power of the rank test and the likelihood ratio test (both r_7 and r_{7E}). The rank test is well known to be scale free; the likelihood ratio test based on the proposed model is also scale free since it is based on the joint distribution derived from the copula in (2.1). However, the power of each of the other tests (r_1 to r_5) decreases, and we leave what causes this phenomenon as an open question. When $\kappa_1 \neq \kappa_2$, for instance $0.5 = \kappa_1 < \kappa_2 = 1.5$ or vice versa, for any fixed κ_{12} the power of tests r_1 to r_5 appears to be between those of r_1 to r_5 with $\kappa_1 = \kappa_2 = 0.5$ in Table 2 and with $\kappa_1 = \kappa_2 = 1.5$ in Table 3. However, the power of r_6 and the likelihood ratio test should remain invariant. We note that the power of all seven tests will be invariant with respect to the location parameters μ_i for $i = 1, 2$ as shown in the dependence of Θ_1 on Θ_2 in Section 2. Thus our simulation study applies to a wide range of situations.

6. Concluding remarks

We investigated inferences procedures under a bivariate circular model with von Mises marginals. An algorithm to generate paired circular data and some 2D contour plots of the model were provided. MLEs for parameters involved and tests for independence were studied. A simulation study showed that the proposed likelihood ratio test is more powerful than six existing tests for independence between paired circular data. The copula in (2.1) entitles the likelihood ratio test scale free for circular-circular dependence. Furthermore, the neat closed form of the von Mises marginals enables the proposed model being extended to other bivariate models easily. Especially, marginals of gene locations in circular genomes were found to be skewed (Horimoto *et al.* (1998) and Horimoto *et al.* (2001)). Thus to compare similarity between two circular genomes, we shall extend this bivariate family to one that has skewed marginals. We leave this as an open question.

Acknowledgements

We thank an associate editor and two referees' constructive comments that helped improve the article much. Thanks also goes to Drs Zhou, Tits and Lawrence for letting us use the FFSQP code in the simulation, and Y. C. Jiang for computation assistance when the manuscript was re-written during 2001-2002. Grace Shieh's research was supported in part by NSC grant 89-2118-M001-006.

References

Downs, T. D. (1974). Rotational angular correlation, *Biorhythms and Human Reproduction* (eds. M. Ferin, F. Halberg and L. van der Wiele), 97–104, Wiley, New York.

Fisher, N. I. (1995). *Statistical Analysis of Circular Data*, Cambridge University Press, Cambridge.

Fisher, N. I. and Lee, A. J. (1982). Nonparametric measures of angular-angular correlation, *Biometrika*, **69**, 315–321.

Fisher, N. I. and Lee, A. J. (1983). A correlation coefficient for circular data, *Biometrika*, **70**, 327–332.

Fisher, N. I. and Lee, A. J. (1986). Correlation coefficients for random variables on a unit sphere or hypersphere, *Biometrika*, **73**, 159–164.

Horimoto, K., Suyama, M., Toh, H., Mori, K. and Otsuka, J. (1998). A method for comparing circular genomes from gene locations: Application to mitochondrial genomes, *Bioinformatics*, **14**, 789–802.

Horimoto, K., Fukuchi, S. and Mori, K. (2001). Comprehensive comparison between locations of orthologous genes on archaeal and bacterial genomes, *Bioinformatics*, **17**, 791–802.

Johnson, R. A. and Shieh, G. S. (2002). On tests of independence for spherical data-invariance and centering, *Statistics and Probability Letters*, **57**, 327–335.

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

Jupp, P. J. and Mardia, K. V. (1980). A general correlation coefficient for directional data and related regression problems, *Biometrika*, **67**, 163–173.

Jupp, P. J. and Mardia, K. V. (1989). A unified view of the theory of directional statistics, *International Statistical Review*, **57**, 1975–1988.

MacKenzie, J. K. (1957). The estimation of an orientation relationship, *Acta Crystallographica*, **10**, 61–62.

Mardia, K. V. (1975*a*). Statistics of directional data, *Journal of the Royal Statistical Society, Series B, Statistical Methodology*, **37**, 349–393.

Mardia, K. V. (1975*b*). Characterizations of directional distributions, *Statistical Distributions in Scientific Work*, *3* (eds. G. P. Patil, S. Kotz and J. K. Ord), 365–385, Reidel, Dordrecht.

Mardia, K. V. and Jupp, P. J. (2000). *Directional Statistics*, John Wiley and Sons, New York.

Press, W. H. (1999). *Numerical Recipes in Fortran*, 2nd ed., Cambridge University Press, New York.

Rivest, L.-P. (1988). A distribution for dependent unit vectors, *Communications in Statistics. Theory and Methods*, **17**, 461–483.

Saw, J. C. (1983). Dependent unit vectors, *Biometrika*, **70**, 665–671.

Self, S. G. and Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions, *Journal American Statistical Association*, **82**, 605–610.

Shieh, G. S., Johnson, R. A. and Frees, E. W. (1994). Testing independence of bivariate circular data and weighted degenerate U-statistics, *Statistica Sinica*, **4**, 729–747.

Shimizu, K. and Iida, K. (2002). Pearson type VII distribution on spheres, *Communications in Statistics*, **31**, 513–526.

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

Stephens, M. A. (1979). Vector correlation, *Biometrika*, **66**, 41–48.

Wehrly, T. and Johnson, R. A. (1980). Bivariate models for dependence of angular observations and a related Markov process, *Biometrika*, **67**, 255–256.

Zhou, J. L., Tits, A. L. and Lawrence, C. T. (1997). A fortran code for solving constrained nonlinear (minimax) optimization problems, generating iterates satisfying all inequality and linear constrains, Technical Report, No. TR-92-107r2, Electrical Engineering Department and Institute for Systems Research, University of Maryland.