Biometrika (2002), **89**, 3, *pp.* 683–697

Printed in Great Britain

Circular regression

By T. D. DOWNS

Research Statistics, Inc., PO Box 25231, Houston, Texas 77265-5231, U.S.A.

tkdowns@swbell.net

AND K. V. MARDIA

Department of Statistics, University of Leeds, Leeds LS2 9JT, U.K.

k.v.mardia@leeds.ac.uk

Summary

A new model for an angular regression link function is introduced. The model employs an angular scale parameter, incorporates proper and improper rotations as special cases, and is equivalent to the Möbius circle mapping for complex variables. Desirable properties of the circle mapping carry over to angular regression. Parameter estimation and inferential methods are developed and illustrated.

Some key words: Angular regression; Circle mapping; Möbius transformation; Primary symmetry point; Reflection; Rotation; Torus; Von Mises distribution.

1. Introduction

Circular regression methods have been used in many diverse applications including crystallography by MacKenzie (1957), vectorcardiography by Downs et al. (1970), predicting direction of ground movement during an earthquake by Rivest (1997) and studies of correlations among circadian biological rhythms, wherein a 24-hour clock is considered as a circle (Binkley, 1990; Downs, 1974; Moore-Ede et al., 1982). Medical applications include circadian timing of cancer chemotherapy to reduce the number and severity of toxic side effects (Hrushesky, 1985) and medical imaging (Jones & Silverman, 1989; Weir & Green, 1994). Recent work on the genetic and molecular aspects of mammalian circadian rhythms suggests the need for more powerful circular regression models (Lowrey et al., 2000; Shearman et al., 2000).

One of the earlier angular-linear regression models was proposed by Gould (1969). Mardia (1975) developed a nonparametric rank correlation coefficient for circular data. Johnson & Wehrly (1978) improved the Gould model by restricting the range of the independent variables to the half-open interval $(0, 2\pi]$. Stephens (1979) has applied a directional regression scheme to spatial rock magnetism data. Fisher & Lee (1992) generalised Johnson & Wehrly's model via a link function that maps the real line on to the unit circle. Recently Lund (1999) proposed a regression model where the independent variables consist of one circular variable and a set of linear variables, and Follman & Proschan (1999) provided a test of circadian circular uniformity of epileptic seizure times using correlated successive seizure times on the same individuals.

Most of the models cited above are rotational in nature, with the mean of the dependent direction assumed to be a rotation, proper or improper, of the independent direction. Non-rotational models have been proposed by Gould (1969), Johnson & Wehrly (1978) and Fisher & Lee (1992), all of whom used linear combinations of linear concomitant variables, and by Downs (1974) and Downs et al. (1970). All these non-rotational models are relatively intractable and difficult to interpret. A serious shortcoming of all existing models is the absence of any topologically appropriate method for angular scale changes. For rotational models, this shortcoming is the equivalent of forcing the slope of the regression curve to be ± 1 in simple linear regression. Our proposed regression model is relatively tractable and incorporates a topologically valid form of angular scale parameter, thus maintaining a one-to-one correspondence between the independent angle and the mean of the dependent angle. As the model parameters vary, the resulting models form a group, with closure and other desirable group properties. Furthermore, rotational models are special cases. Some spherical extensions are proposed in Downs & Mardia (2000).

To our knowledge, no bivariate angular probability distribution has been used to model this type of circular regression. A bivariate von Mises distribution was introduced by Mardia (1975), but we show in § 6·2 that the conditional probability distribution for that model is somewhat different from the conditional von Mises distribution used herein; we also discuss in § 6·2 a special case of this distribution which has been used by Rivest (1997) for a circular regression link function.

The circular regression model is defined in § 2. Properties of the model are described in § 3, parameter estimation and inferential methods in § 4 and practical examples in § 5. Additional topics are briefly discussed in § 6.

Unless otherwise specified all angles and their sums or differences are expressed as their principal values in the half-open interval $(-\pi, \pi]$ radians, and all half-angles of these principal values are in the half-open interval $(-\pi/2, \pi/2]$.

2. CIRCULAR REGRESSION MODEL

2·1. *Regression curve: Relationship of angular mean to independent angle*

Let α and β be angular location parameters, ω a slope parameter in the closed interval $[-1, 1]$, and u and v running angular variables. The mapping

$$\tan \tfrac{1}{2}(v-\beta) = \omega \tan \tfrac{1}{2}(u-\alpha), \tag{2·1}$$

which has the unique solution

$$v = \beta + 2 \operatorname{atan}\{\omega \tan \tfrac{1}{2}(u-\alpha)\}, \tag{2·2}$$

defines a one-to-one relationship between u and v provided ω is not zero. The locus of the points (u, v) satisfying (2·1) is a continuous closed curve winding once around a toroidal surface. Fisher & Lee (1992) suggested the link function $\mu = 2 \operatorname{atan} x$ for linear-circular regression, since it maps the linear variable x to $(-\pi, \pi]$. Our circular regression link has, in addition, the linear x as a function of angle u via $x = \omega \tan(u-\alpha)/2$.

Now assume that u is the fixed independent angle, v the dependent random angle and v in (2·1) replaced by μ, the mean direction for v given u. The resulting link function, or regression curve, is given by

$$\tan \tfrac{1}{2}(\mu-\beta) = \omega \tan \tfrac{1}{2}(u-\alpha), \tag{2·3}$$

which has the unique solution

$$\mu = \beta + 2 \operatorname{atan}\{\omega \tan \tfrac{1}{2}(u-\alpha)\}. \tag{2·4}$$

From (2·3) we get

$$\tan \mu'/2 = \omega \tan \mu'/2, \tag{2·5}$$

where $\mu' = \mu - \beta$, $u' = u - \alpha$. We call the version (2·5) the centred regression curve. We show how to construct μ' graphically from a given u'. Imagine a v'-circle with unit diameter and tangent to a horizontal t-axis at the point $(t = 0, v' = 0)$, and a u'-circle with diameter $\omega < 1$ and likewise tangent to the t-axis at the point $(t = 0, u' = 0)$. Given a point u' on the u'-circle, draw the line from the top of the u'-circle through the point u' thereon to intersect the t-axis at the point t, say. Next draw the line from the top of the v'-circle to the same point t on the t-axis, intersecting the v'-circle at the point v', say. Then $t = \tan v'/2$ and also $t = \omega \tan u'/2$, so $v' = \mu'$. Use $-u'$ if $\omega < 0$.

When we are numerically evaluating the tangent functions in (2·3), (2·4) and (2·5), principal values for u' and $u - \alpha$ must be used before division by 2, otherwise the half-angles $\frac{1}{2}(u - \alpha)$ and $u'/2$ will not lie in the interval $(-\pi/2, \pi/2]$ and numerical errors will result.

To obtain a simple complex form for the link (2·5), let $M' = \exp\{i(\mu - \beta)\}$ and $U' = \exp\{i(u - \alpha)\}$ be points on the unit circle. Substitute $(M' - 1)/i(M' + 1)$ and $\omega(U' - 1)/i(U' + 1)$ for the left-hand and right-hand sides of (2·5) respectively, and solve for M' to obtain

$$M' = (U' + \psi)/(\psi U' + 1), \tag{2·6}$$

where

$$\psi = (1 - \omega)/(1 + \omega), \quad \omega = (1 - \psi)/(1 + \psi), \quad i^2 = -1.$$

2·2. *Probability model for angular error: Properties of the von Mises distribution*

A random angle t has the von Mises distribution, with mean γ and nonnegative concentration parameter κ, when the density for t is

$$f(t) = \{2\pi I_0(\kappa)\}^{-1} \exp\{\kappa \cos(t - \gamma)\}, \tag{2·7}$$

in which case we write $t \sim M(\gamma, \kappa)$, where $I_0(\kappa) = \sum_{j=0}^{\infty} \{(\kappa/2)^j/j!)\}^2$ is the modified Bessel function of the first kind and order zero. The expected value of the unit vector $(\cos t, \sin t)$ is

$$E\{(\cos t, \sin t)\} = \rho(\cos \gamma, \sin \gamma), \tag{2·8}$$

where $\rho = I_1(\kappa)/I_0(\kappa)$, for $0 \leqslant \rho \leqslant 1$, and $I_1(\kappa)$ is the modified Bessel function of the first kind and order one. The parameter ρ is the precision, and increases from 0 to 1 as κ increases from 0 to ∞. The precision is the distance from the origin to the centre of gravity of the population of points on the unit circle, and the mean direction γ is the direction from the origin to the centre of gravity. Mardia & Jupp (2000, Ch. 3–5, 7) have given extensive details about properties, estimation and inference aspects of the von Mises distribution. The family of distributions (2·7) is closed under rotations of t, with the concentration parameter unchanged:

$$t \sim M(\gamma, \kappa) \Rightarrow t + \theta \sim M(\gamma + \theta, \kappa) \tag{2·9}$$

for any fixed angle θ in $(-\pi, \pi]$.

2·3. *Probability model for angular error: Angular error and the von Mises distribution*

We assume that v given u has the von Mises distribution with mean direction μ and nonnegative concentration parameter κ. To emphasise that μ is a function of the independent variable u and the parameters (α, β, ω), we write

$$v \mid u \sim M\{\mu(u;\, \alpha, \beta, \omega), \kappa\}, \tag{2·10}$$

where

$$\mu(u;\, \alpha, \beta, \omega) = \beta + \nu(u-\alpha;\, \omega), \quad \nu(u-\alpha;\, \omega) = 2\,\text{atan}\{\omega \tan \tfrac{1}{2}(u-\alpha)\}. \tag{2·11}$$

Applying (2·9) to (2·10) gives that

$$v \mid u \sim M\{\mu(u;\, \alpha, \beta, \omega), \kappa\} \Rightarrow v + b \mid u + a \sim M\{\mu(u+a;\, \alpha + a, \beta + b, \omega), \kappa\} \tag{2·12}$$

for any fixed angles a, b in $(-\pi, \pi]$, since, by (2·11),

$$(\beta + b) + \nu\{(u+a) - (\alpha + a), \omega\} = \mu(u + \alpha;\, \alpha + a, \beta + b, \omega).$$

Two important results derived from (2·12) are as follows:

$$e = \text{angular error} = v - \mu(u;\, \alpha, \beta, \omega) \sim M(0, \kappa), \quad E(\cos e, \sin e) = (\rho, 0); \tag{2·13}$$

$$t = \text{special transform} = e + \gamma \sim M(\gamma, \kappa), \quad E(\cos t, \sin t) = \rho(\cos\gamma, \sin\gamma). \tag{2·14}$$

3. Properties of the regression curve

3·1. *Special forms of the regression curve*

The special transform t is used whenever the slope parameter ω takes one of the special values, -1, 0 or 1. Then (2·12), (2·13) and (2·14) simplify to their special distributions, wherein only the mean direction $\gamma = (\beta - \omega\alpha)$ is estimable, and the individual parameters α and β are not. For $\omega = -1$, 0 or 1, we have that

$$v \mid u \sim M(\mu, \kappa),$$

where

$$\mu = \omega u + (\beta - \omega\alpha), \tag{3·1}$$

$$e = \text{angular error} = (v - \omega u) - (\beta - \omega\alpha) \sim M(0, \kappa), \tag{3·2}$$

$$t = \text{special transform} = (v - \omega u) \sim M(\gamma, \kappa), \tag{3·3}$$

where $\gamma = \beta - \omega\alpha$. The expression for μ in (3·1) has the slope-intercept form of a linear regression model. If $\omega = -1$ then μ is a rotation, through $\gamma = \beta + \alpha$, of a reflection of u in the horizontal axis. Also, $\omega = 0$ if and only if v is independent of u, in which case $\mu = \beta$, a constant, and α disappears. If $\omega = 1$ then μ is a rotation of u through $\gamma = \beta - \alpha$. Thus the three special models for the regression curve correspond to an improper rotation, no association and a proper rotation.

For nonspecial cases the mean μ of v and the independent angle u are negatively related, unrelated or positively related according to the algebraic sign of ω. If $u = \alpha$ then $\mu = \beta$ so the point (α, β) lies on the (u, v) regression curve, as does $(\alpha + \pi, \beta + \pi)$. The regression curve (2·3) is symmetric about both these points, and only these points. Thus α and β are location parameters for the points of symmetry of the regression curve and are not, in general, measures of central location for u or v. The point (α, β) will be called the primary symmetry point. The points $(0, 0)$ and (π, π) lie on the centred (u', μ') toroidal regression curve (2·5), and $(0, 0)$ is the primary symmetry point. Planar graphs of the centred model are shown in Fig. 1 for selected values of ω. No generality is lost by restricting the slope parameter ω to the interval $[-1, 1]$, for, if the magnitude of ω exceeded unity then, since $\tan\{(\alpha+\pi)/2\} = -\cot(\alpha/2)$, reparameterising (2·3) by adding π to each of α and β transforms the tangents therein to negative cotangents with ω unchanged. Taking reciprocals of both sides of this, and multiplying through by -1, transforms the negative cotangents back to positive tangents, but replaces ω with its reciprocal.

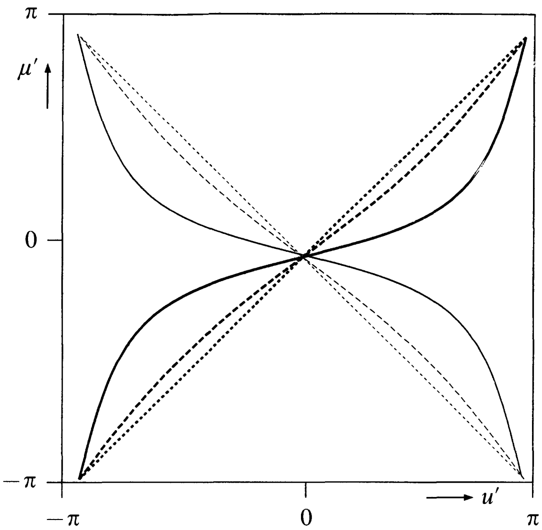

Fig. 1. Planar plots of μ' versus u' for the centred regression curve (2·5) for selected values of ω. Regression curves for negative ω are reflections in the horizontal line $\mu' = 0$ of the corresponding curves for positive ω. Short-dashed lines, $\omega = \pm 1$; long-dashed lines, $\omega = \pm\frac{3}{4}$; solid lines $\omega = \pm\frac{1}{4}$.

McCullagh (1996) has used mainly the real Möbius group $\mathrm{SL}(2, R)$ to investigate Möbius transformations of the Cauchy distribution. The group $\mathrm{SL}(2, R)$ is the real analogue of the complex Möbius group $\mathrm{SL}(2, C)$ of fractional linear transformations on the extended complex plane. These have the form

$$w = (az + b)/(cz + d),$$

where a, b, c, d, w and z are complex and the determinant $ad - bc$ is not zero. Some algebra shows, as in Needham (1997, pp. 176–8), that those elements of $\mathrm{SL}(2, C)$ which send the unit circle into itself are precisely those for which

$$\frac{w}{B} = \frac{z/A + \psi}{\psi z/A + 1}, \tag{3·4}$$

where $A = \exp(i\alpha)$, $B = \exp(i\beta)$. This is the noncentred version of (2·6) above, with $M' = w/B$, $U' = z/A$.

3·2. *Additional properties of the regression curve*

The three parameters (α, β, ω) or (α, β, ψ) are unique unless ω takes one of the three special values. To show this, consider (2·6) with $M' = M/B$, $U' = U/A$, so that

$$M = B(U + \psi A)/(\psi U + A). \tag{3·5}$$

Suppose that $M = B_1(U + \psi_1 A_1)/(\psi_1 U + A_1)$ and also that $M = B_2(U + \psi_2 A_2)/(\psi_2 U + A_2)$ for all values of U. Cross-multiplying and equating coefficients of U implies that the two sets of parameters are equal, so long as ψ is not 0, 1 or ∞, corresponding to the three special values 1, 0 and -1 of ω.

In addition, the complex regression curves (2·6) have an elegant linear representation: $M' + U'$ can be shown to be parallel to $U' + \psi$, which implies that M' and $-U'$ on the unit circle are collinear with ψ on the nonnegative real axis. The points M, $-U'B$ and ψB are also collinear.

Finally, using cross ratios we may show that the circle mapping (3·5) is completely determined, through (3·6) below, by any three distinct pairs (U_1, M_1), (U_2, M_2) and (U_3, M_3) of mapped points on the U and M circles:

$$\frac{(M - M_1)(M_2 - M_3)}{(M - M_3)(M_2 - M_1)} = \frac{(U - U_1)(U_2 - U_3)}{(U - U_3)(U_2 - U_1)}. \tag{3·6}$$

The values of the parameters (A, B, ψ) in (3·5) are readily deduced from (3·6).

4. ESTIMATION AND INFERENCE

4·1. *Graphical estimation*

Despite potential data distortion, a planar graph of v versus u can be useful for evaluating the regression model and for obtaining rough estimates of the regression parameters. Graphical consistency of the data with the regression curve model may be assessed in some cases. Non-monotonicity of the data discredits the model. Constant slopes of -1, 0 or 1 around the toroidal surface are supportive of one of the three special cases of the model, while constant slopes of other values discredit the model.

Data with a linear trend and shallow slope over a large range are consistent with a regression curve when ω is small and when the primary symmetry point is within the range of the data. Linear data with a sharp slope and a small range are again consistent with a regression curve when ω is small, but now the secondary symmetry point, $(\alpha + \pi, \beta + \pi)$, is within or near the range of the data. Indications of convexity or concavity in the data are also useful for assessing model validity and graphically estimating (α, β, ω). We will illustrate these points in § 5.

4·2. *Classification of regression curve models*

We classify our regression models as A, B or C according to the nature of the parameters (α, β, ω). Class A models have three functionally independent parameters, α, β and ω. Class B models have α and β functionally related by

$$\alpha \pm \beta = 0,$$

so the parameters (α, β, ω) have the form $(\alpha, \pm\alpha, \omega)$. The equations $\alpha \pm \beta = 0$ imply a kind of symmetry in the regression curves: the primary symmetry point (α, β) lies on the line $\mu = u$ when $\omega > 0$ or on the line $\mu = -u$ when $\omega < 0$. Class C models are the special models where ω takes one of the special values of -1, 0 or 1, and is presumed known, in which case we write $\omega=\omega_0$.

The three classes are hierarchical in that the parameter spaces of Class C models are contained in those of Class B models, which are in turn contained in those of Class A models.

4·3. *Maximum likelihood estimation for Class A and B models*

The loglikelihood function for a random sample of n pairs (u_j, v_j) from a Class A or B model is given by

$$l(\alpha, \beta, \omega, \kappa; v_1, \ldots, v_n) = -n \log I_0(\kappa) + \kappa \sum_j \cos\{v_j - \beta - \nu(u_j - \alpha; \omega)\} + \text{const.} \quad (4{\cdot}1)$$

Recall that, for Class B models, $\beta = \pm\alpha$. For both classes, the maximum likelihood estimator $\hat{\rho}$ of the precision parameter ρ is defined explicitly by

$$\hat{\rho}(\alpha, \beta, \omega) = (1/n) \sum_j \cos\{v_j - \beta - \nu(u_j - \alpha; \omega)\}. \quad (4{\cdot}2)$$

For a Class A model, we can obtain the estimator $\hat{\beta}$ of β which leads to

$$\hat{\rho}(\alpha, \hat{\beta}, \omega) = \frac{1}{n}\left(\left[\sum_j \cos\{v_j - \nu(u_j - \alpha; \omega)\}\right]^2 + \left[\sum_j \sin\{v_j - \nu(u_j - \alpha; \omega)\}\right]^2\right)^{\frac{1}{2}}. \quad (4{\cdot}3)$$

Thus, maximising the loglikelihood (4·1) is equivalent to maximising (4·3) with respect to (α, ω); that is we are using a profile likelihood. Furthermore, the maximum likelihood estimator $\hat{\kappa}$ of κ is given by

$$I_1(\hat{\kappa})/I_0(\hat{\kappa}) = \hat{\rho}(\hat{\alpha}, \hat{\beta}, \hat{\omega}). \quad (4{\cdot}4)$$

However, we will concentrate on $\hat{\rho}$ rather than $\hat{\kappa}$ in our examples. For Class B, the profile likelihood leads to maximising $\hat{\rho}(\alpha, \pm\alpha, \omega)$ with respect to α and ω.

4·4. *Maximum likelihood estimation for Class C models*

A sample of n pairs (u_j, v_j) from any Class C model, in which ω is fixed at ω_0, can be considered as a univariate sample from a special $M(\gamma, \kappa)$ distribution, by setting

$$t_j = (v_j - \omega_0 u_j), \quad \gamma = (\beta - \omega_0 \alpha), \quad (4{\cdot}5)$$

where

$$t_j \sim M(\gamma, \kappa) \quad (j = 1, 2, \ldots, n).$$

Maximum likelihood estimators $\hat{\gamma}$ and $\hat{\rho}$ of the special mean direction γ and precision ρ are known to be the sample centre of gravity analogues of their population counterparts in (2·8). Thus, for $\omega = \omega_0$ known, ρ and γ are estimated implicitly from

$$\hat{\rho}_0(\hat{\gamma}, \omega_0)(\cos\hat{\gamma}, \sin\hat{\gamma}) = (1/n) \sum_j (\cos t_j, \sin t_j). \quad (4{\cdot}6)$$

If (γ, ω) are known or assumed to be (γ_0, ω_0) then the precision ρ is estimated explicitly by

$$\hat{\rho}_0(\gamma_0, \omega_0) = (1/n) \sum_j \cos(t_j - \gamma_0). \quad (4{\cdot}7)$$

4·5. *Inference*

When $e \sim M(0, \kappa)$ and ρ is sufficiently large, it is known that

$$2\kappa(1-\cos e) \sim \chi^2(1), \tag{4·8}$$

approximately. The utility of (4·8) for regression parameter inference is illustrated with two scenarios. First, consider the important null hypothesis $H_0: \omega = 0$. The alternative hypothesis, $H_1: \omega \neq 0$, must have a parameter space that contains the parameter space under H_0. Then an approximate test of the null hypothesis can be obtained by noting that, for sufficiently large n and ρ, the terms of the analysis of dispersion identity,

$$2n\kappa\{1-\hat{\rho}_0(\hat{\gamma}, 0)\} = 2n\kappa[\{1-\hat{\rho}_0(\hat{\gamma}, 0)\} - \{1-\hat{\rho}(\hat{\alpha}, \hat{\beta}, \hat{\omega})\}] + 2n\kappa\{1-\hat{\rho}(\hat{\alpha}, \hat{\beta}, \hat{\omega})\}, \tag{4·9}$$

where $\hat{\rho}_0(\hat{\gamma}, 0)$ and $\hat{\rho}(\hat{\alpha}, \hat{\beta}, \hat{\omega})$ are defined in (4·6) and (4·3), will by Wilks' theorem be approximately distributed under the null hypothesis as

$$\chi^2(n-1) = \chi^2\{(n-1)-(n-3)\} + \chi^2(n-3), \tag{4·10}$$

so that

$$F = (n-3)\{\hat{\rho}(\hat{\alpha}, \hat{\beta}, \hat{\omega}) - \hat{\rho}_0(\hat{\gamma}, 0)\}/[2\{1-\hat{\rho}(\hat{\alpha}, \hat{\beta}, \hat{\omega})\}] \sim F(2, n-3),$$

approximately, since κ cancels out. Large F values discredit H_0. The chi-squared distributions in (4·8) and (4·10) are reminiscent of the Watson & Williams (1956) analysis-of-variance-like equalities for directional statistics, and it is expected that we may not require n to be large for these results to hold so long as ρ is large. In practice it will suffice if ρ is only moderately large.

The second scenario employs a generalisation to a regression framework of a procedure first proposed by Watson & Williams (1956) for testing equality of mean directions in two independent samples. We wish to test the hypothesis that the unknown regression parameters, d say in number, are the same for both samples. The relevant quantities for the first sample of size n_1, for the second of size n_2 and for the combined samples of size n are

$$Q_1^2 = 2n_1\kappa_1(1-\hat{\rho}_1), \quad Q_2^2 = 2n_2\kappa_2(1-\hat{\rho}_2), \quad Q^2 = 2n\kappa(1-\hat{\rho}).$$

If $\kappa_1 = \kappa_2$ then the precisions are all equal, and equal to ρ, say, and if ρ is moderately large then, as above, the terms of the identity

$$Q^2 = \{Q^2 - (Q_1^2 + Q_2^2)\} + (Q_1^2 + Q_2^2)$$

will be approximately distributed as

$$\chi^2(n-d) = \chi^2\{(n-d)-(n_1-d)-(n_2-d)\} + \chi^2\{(n_1-d)+(n_2-d)\}, \tag{4·11}$$

so that

$$F = (n-2d)(n_1\hat{\rho}_1 + n_2\hat{\rho}_2 - n\hat{\rho})/\{d(n-n_1\hat{\rho}_1 - n_2\hat{\rho}_2)\} \sim F(d, n-2d),$$

approximately, under the null hypothesis, with large F values discrediting the null hypothesis.

All the tests for the $M\{\mu(u; \alpha, \beta, \omega), \kappa\}$ regression model in this paper are simple variations of the above two scenarios.

We now calculate the information matrix. We write the loglikelihood with known parameters $(\alpha, \beta, \omega, \kappa)$ as

$$l = \text{const} - n\log I_0(\kappa) + \kappa\sum\cos(v_i - \mu_i), \tag{4·12}$$

where

$$\mu_i = \beta + 2\,\mathrm{atan}\{\omega \tan\tfrac{1}{2}(u_i - \alpha)\}. \tag{4·13}$$

Using the facts that

$$E\{\cos(v_i - \mu_i)\} = A(\kappa), \quad E\{\sin(v_i - \mu_i)\} = 0, \tag{4·14}$$

we find that various expressions for the information matrix simplify as we now summarise. Recall that $A(\kappa) = I_1(\kappa)/I_0(\kappa)$.

Let us now write $\theta^{\mathrm{T}} = (\beta, \alpha, \omega, \kappa) = (\theta_1, \theta_2, \theta_3, \theta_4)$ and suppose that $I = (I_{ij})$ is the Fisher information matrix for θ. We find that $I_{14} = I_{24} = I_{34} = 0$ so that $\hat{\theta}_1$, $\hat{\theta}_2$ and $\hat{\theta}_3$ are independent of $\hat{\theta}_4$ as expected, asymptotically. Hence, we can concentrate on I_{ij} $(i, j = 1, 2, 3)$. Write

$$I = \begin{bmatrix} C_{11} & 0 \\ 0^{\mathrm{T}} & C_{22} \end{bmatrix},$$

where $0^{\mathrm{T}} = (0, 0, 0)$, C_{11} is 3×3 and C_{22} is a scalar. Then it can be shown, using (4·12), (4·13) and (4·14), that

$$C_{22} = I_{44} = nA'(\kappa), \quad C_{11} = \kappa A(\kappa) B(\theta_1, \theta_2, \theta_3), \tag{4·15}$$

where the elements of the matrix $B(\beta, \alpha, \omega)$ are

$$b_{11} = n, \quad b_{12} = \sum (\mu_i)_\alpha, \quad b_{13} = \sum (\mu_i)_\omega,$$
$$b_{22} = \sum (\mu_i)_\alpha^2, \quad b_{23} = \sum (\mu_i)_\alpha (\mu_i)_\omega, \quad b_{33} = \sum (\mu_i)_\omega^2,$$

with

$$(\mu_i)_\omega = \frac{2 \tan\frac{1}{2}(u_i - \alpha)}{1 + \omega^2 \tan^2\frac{1}{2}(u_i - \alpha)}, \quad (\mu_i)_\alpha = -\frac{\omega \sec^2\frac{1}{2}(u_i - \alpha)}{1 + \omega^2 \tan^2\frac{1}{2}(u_i - \alpha)}.$$

Thus

$$\mathrm{cov}(\hat{\beta}, \hat{\alpha}, \hat{\omega}) \simeq \{I(\hat{\beta}, \hat{\alpha}, \hat{\omega})\}^{-1} = \{\hat{\kappa} A(\hat{\kappa})\}^{-1} \{B(\hat{\beta}, \hat{\alpha}, \hat{\omega})\}^{-1}, \tag{4·16}$$

$$\mathrm{var}(\hat{\kappa}) \simeq \{nA'(\hat{\kappa})\}^{-1}, \quad A'(\kappa) = 1 - A^2(\kappa) - \frac{A(\kappa)}{\kappa}, \tag{4·17}$$

$$\mathrm{cov}(\hat{\beta}, \hat{\kappa}) = \mathrm{cov}(\hat{\alpha}, \hat{\kappa}) = \mathrm{cov}(\hat{\omega}, \hat{\kappa}) = 0. \tag{4·18}$$

Other particular cases are much simpler. For example, for $\alpha = \beta$, all the terms are the same except that we replace $(\mu_i)_\alpha$ by $1 + (\mu_i)_\alpha$ everywhere, and there is no term for β.

5. Examples

Example 1: *Circadian biological rhythms.* We are indebted to Franz Halberg and Michael Smolensky for providing us with data from ten medical students in Austria. The students measured each of about 20 variables several times daily for a period of several weeks. The study period was split into two prime time periods as part of the study, and the peak time for systolic blood pressure was estimated separately for each student for each period, giving values S_1 and S_2. These data, shown in Table 1, are in degrees, with 15 degrees equal to one hour. The two blood pressure peak times should be nearly equivalent, if circumstances are the same for each of the two periods. This can be checked by regressing S_2 on S_1. First we test that there really is an association between the peak times for S_1 and S_2. The maximum likelihood estimates for the regression parameters are $\hat{\rho}(\hat{\alpha}, \hat{\beta}, \hat{\omega}) = \hat{\rho}(17^\circ, 6^\circ, 0{\cdot}67) = 0{\cdot}971$ under H_1, with $n - 3 = 7$ degrees of freedom. Asymptotic standard errors for $(\hat{\alpha}, \hat{\beta}, \hat{\omega})$ are $(19^\circ, 20^\circ, 0{\cdot}16)$. Consequently, the Wald test for $\omega = 0$ has $z = 0{\cdot}67/0{\cdot}16 = 4{\cdot}2$ with $P \ll 0{\cdot}001$. From Table 1, $\hat{\rho}_0(\hat{\gamma}, \omega_0) = \hat{\rho}_0(-45^\circ, 0) = 0{\cdot}718$ under H_0, with 9 degrees of freedom. Then under the null hypothesis the F statistic is approximately $F(2, 7)$.

Table 1. *Systolic blood pressure peak times (degrees) with sample angular means $\hat{\gamma}$ and precisions $\hat{\rho}$ for Example 1: 0° is midnight, 90° is 6 a.m. and −90° is 6 p.m.*

	Student											
	1	2	3	4	5	6	7	8	9	10	$\hat{\gamma}$	$\hat{\rho}$
S_1	−102	−17	−23	−39	−77	−47	−65	−166	2	−44	−52°	0·736
S_2	−98	−14	−39	1	−76	−48	−45	−175	−13	−32	−45°	0·718

S_1, systolic blood pressure peak time for period 1; S_2, systolic blood pressure peak time for period 2.

The observed value is $F = 7 \times (0{\cdot}971 - 0{\cdot}718)/\{2 \times (1 - 0{\cdot}971)\} = 30{\cdot}5$, giving $P \ll 0{\cdot}001$, so the data show a strong association and severely discredit the null hypothesis.

Next, we test the Class C null hypothesis that the blood pressure peak times for the two periods are identical, apart from random error. This will be true when $\gamma = 0$ and $\omega_0 = 1$. Then $\hat{\rho}_0(0, 1) = \sum_j \cos(v_j - u_j)/10 = 0{\cdot}962$ by (4·7), with $10 - 0 = 10$ degrees of freedom; and $\hat{\rho}(\hat{\alpha}, \hat{\beta}, \hat{\omega}) = 0{\cdot}971$ as above, with 7 degrees of freedom. We obtain $F = 7 \times (0{\cdot}971 - 0{\cdot}962)/\{(10 - 7)(1 - 0{\cdot}971)\} = 0{\cdot}724$, giving $P > 0{\cdot}5$ when referred to the $F(3, 7)$ distribution, indicating that the data do not conflict with the null hypothesis of identical peak times for the two periods.

Example 2: *Wind directions in Milwaukee.* Wind directions u and v were measured at a weather station in Milwaukee, Wisconsin, at 6 a.m. and at noon, respectively, on each day for 21 consecutive days. These data are from Fisher (1993, Table B.21). A noncentred planar graph of the data is shown in Fig. 2. Figure 2 suggests little or no association, but this is deceptive. The top boundary represents values of u for $v = \pi$, and the bottom boundary values of u for $v = -\pi$. Imagine curling the rectangular data plot so as to join these two equivalent lines and form a cylinder, and then curling and stretching the cylinder so as to form a torus. The four corners of the original rectangle are now all the same point on the topologically appropriate torus.

Data from separate days are considered independent. Take as null hypothesis the Class B symmetry hypothesis that $\alpha = \beta$ with, as alternative, the general Class A model. A graph of the sample precision versus the slope parameter for the likelihood profile grid search is shown in Fig. 3. The asymmetry and dramatic changes in precision therein suggest a strong positive correlation where the precision and slope estimates both seem to be about $\frac{1}{2}$. In contrast, the planar data plot of Fig. 3 wrongly suggests little or no association. Actual precision estimates are $\hat{\rho}(\hat{\alpha}, \hat{\beta}, \hat{\omega}) = \hat{\rho}(41^\circ, 41^\circ, 0{\cdot}367) = 0{\cdot}398$ with $n - 2 = 19$ degrees of freedom, and $\hat{\rho}(\hat{\alpha}, \hat{\beta}, \hat{\omega}) = \hat{\rho}(78^\circ, 119^\circ, 0{\cdot}581) = 0{\cdot}487$ with $n - 3 = 18$ degrees of freedom. Asymptotic standard errors from (4·16) are $(10^\circ, 0{\cdot}23)$ for $(\hat{\alpha}, \hat{\omega})$ and $(48^\circ, 55^\circ, 0{\cdot}23)$ for $(\hat{\alpha}, \hat{\beta}, \hat{\omega})$. The test statistic under the null hypothesis is $F = 18 \times (0{\cdot}487 - 0{\cdot}398)/\{1 \times (1 - 0{\cdot}487)\} = 3{\cdot}12$. Under H_0, $F \sim F(1, 18)$, approximately, giving $P = 0{\cdot}094$, so the data do not markedly discredit the null hypothesis of symmetry.

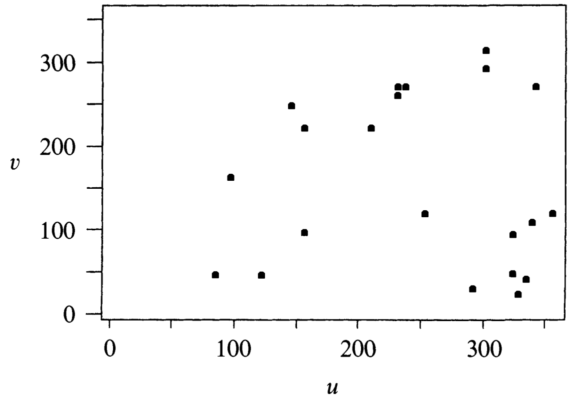

Fig. 2. Example 2. Deceptive planar data plot of wind directions, measured in degrees: v, measured at noon, versus u, measured at 6 a.m. Zero is north.

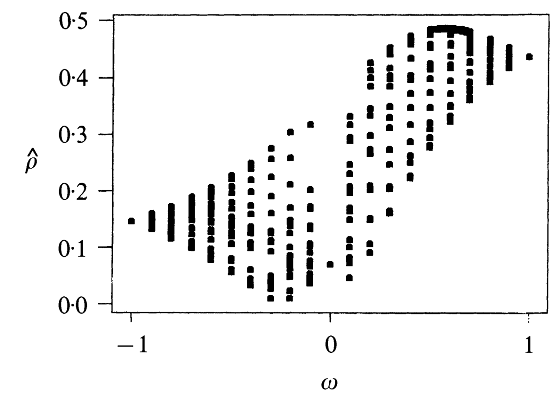

Fig. 3. Graph of the grid search for Example 2 for the maximum likelihood estimation where sample precisions $\hat{\rho}$ are plotted against the slope parameter ω, for 17 different values of α.

6. Additional topics

6·1. *Vectorial representation of this conditional density*

Define the product and quotient slope parameters ω_p and ω_q by

$$\omega_p^2 = \psi^2/(1+\psi^2), \quad \omega_q^2 = 1/(1+\psi^2), \tag{6·1}$$

where $\omega_p, \omega_q \geqslant 0$, so that $\psi = \omega_p/\omega_q$ and $\omega_p^2 + \omega_q^2 = 1$.

Knowledge of any one of the four parameters ω, ω_p, ω_q and ψ is sufficient to determine the other three. Next, define the three vectors x, y and η by

$$\begin{aligned} x^{\mathrm{T}} &= (x_1, x_2, x_3, x_4) = (\cos p/2, \sin p/2, \cos q/2, \sin q/2), \\ y^{\mathrm{T}} &= (y_1, y_2, y_3, y_4) = (\sin p/2, -\cos p/2, \sin q/2, -\cos q/2), \\ \eta^{\mathrm{T}} &= (\eta_1, \eta_2, \eta_3, \eta_4) \\ &= (\omega_p \cos\tfrac{1}{2}(\beta+\alpha), \omega_p \sin\tfrac{1}{2}(\beta+\alpha), \omega_q \cos\tfrac{1}{2}(\beta-\alpha), \omega_q \sin\tfrac{1}{2}(\beta-\alpha)), \end{aligned} \tag{6·2}$$

where $p = v + u$ and $q = v - u$. Note that x is perpendicular to y. The parameters (α, β, ω) can be obtained from the elements of η via

$$\begin{aligned}
&\omega_p^2 = \eta_1^2 + \eta_2^2, \quad \omega_q^2 = \eta_3^2 + \eta_4^2, \quad \omega = (\omega_q - \omega_p)/(\omega_q + \omega_p);\\
&\cos\alpha = (\eta_1\eta_3 + \eta_2\eta_4)/\omega_p\omega_q, \quad \cos\beta = (\eta_1\eta_3 - \eta_2\eta_4)/\omega_p\omega_q;\\
&\sin\alpha = (\eta_2\eta_3 - \eta_1\eta_4)/\omega_p\omega_q, \quad \sin\beta = (\eta_2\eta_3 + \eta_1\eta_4)/\omega_p\omega_q.
\end{aligned} \tag{6·3}$$

All the expressions determining (α, β, ω) in (6·3) are of second order in the elements of η, so the same values for (α, β, ω) will result whether one uses the elements of η or the elements of $-\eta$.

Now substitute ω_p/ω_q for ψ in (2·6) and divide the resulting expression for M' into V' to get an expression for the complex angular error E. Multiply and divide this expression for E by the complex conjugate of $(\omega_q U' + \omega_p)$ and rearrange to get

$$E = \exp(ie) = \frac{\{\omega_p(V'U')^{\frac{1}{2}} + \omega_q(V'/U')^{\frac{1}{2}}\}^2}{\delta(u')^2} = \frac{\{(x^{\mathrm{T}}\eta) + i(y^{\mathrm{T}}\eta)\}^2}{(x^{\mathrm{T}}\eta)^2 + (y^{\mathrm{T}}\eta)^2}, \tag{6·4}$$

where

$$\delta(u')^2 = |\omega_q U' + \omega_p|^2 = 1 + 2\omega_p\omega_q\cos u' = (x^{\mathrm{T}}\eta)^2 + (y^{\mathrm{T}}\eta)^2, \quad 2\omega_p\omega_q = \frac{1-\omega^2}{1+\omega^2}, \tag{6·5}$$

where $u' = u - \alpha$. Taking square roots in (6·4) yields $x^{\mathrm{T}}\eta/\delta(u') = \cos e/2$ as the real part of $E^{\frac{1}{2}}$ and $y^{\mathrm{T}}\eta/\delta(u') = \sin e/2$ as the imaginary part. Hence the conditional density of $v|u$ in vectorial form is

$$f(v|u) = \{2\pi I_0(\kappa)\exp(-\kappa)\}^{-1}\exp\{-2\kappa y^{\mathrm{T}}\eta\eta^{\mathrm{T}}y/\delta(u')^2\}. \tag{6·6}$$

Initially, we used this form of the density to obtain maximum likelihood estimates of the parameters, but found the likelihood profile grid search method to be better.

6·2. *Similarities with conditional bivariate von Mises distribution*

Mardia (1975) introduced a bivariate von Mises distribution wherein the conditional distribution of one angle v given the other angle u is also von Mises but had the form

$$\begin{aligned}
g(v|u) \propto \exp(&\kappa_1\cos v + \kappa_2\sin v + \kappa_{11}\cos u\cos v + \kappa_{12}\cos u\sin v\\
&+ \kappa_{21}\sin u\cos v + \kappa_{22}\sin u\sin v),
\end{aligned} \tag{6·7}$$

where the κ's are parameters. An explicit expression for our conditional von Mises distribution with $f(v|u) \propto \exp\{\kappa\cos(v-\mu)\}$ is obtained by getting $\cos(v-\mu)$ from the real part of (6·4), so that

$$f(v|u) \propto \exp\{(K_2\cos v' + K_{11}\cos u'\cos v' + K_{22}\sin u'\sin v')/\delta(u')\}, \tag{6·8}$$

where $\delta(u')$ is defined in (6·5), $u' = u - \alpha$, $v' = v - \beta$, and the K's are parameters depending on ρ and ω. The presence of $\delta(u')$ in (6·8), which is a function of u', detracts from the similarities between (6·7) and (6·8).

Rivest (1997) used a special case of Mardia's bivariate distribution (6·7) to obtain a conditional distribution for regression which has the same form as ours in (6·8), but in his case $\delta(u') = \text{constant}$. As a result his link function still contained three parameters, somewhat analogous to our (α, β, ω). Indeed, his regression function can be written as

$$\mu(u;\, \alpha, \beta, \psi) = \beta + \operatorname{atan}\left(\frac{a}{b}\right) + h(b)\pi, \tag{6·9}$$

where $a = \sin(u-\alpha)$, $b = \psi + \cos(u-\alpha)$, $\psi \geqslant 0$, $h(b) = 1$ if $b < 0$, and $h(b) = 0$ otherwise. There are two important distinctions between his model and ours. His α and β are functionally dependent on the mean directions of the dependent and independent directions, while ours are not. His ψ, like our ψ, ranges from 0 to infinity, and has $\psi = 0$ corresponding to a pure rotation and $\gamma = \infty$ to no association, while our ψ has 0 as a pure rotation, 1 as no association and infinity as a pure reflection. The most important difference is that our link function is a one-to-one mapping of the unit circle on to itself for all values of ψ except the special values 0, 1 and infinity, thus preserving group properties and topological validity. Rivest's link function is only one-to-one and onto when $\psi < 1$.

6·3. *Barber pole regression curves*

All our regression curves have treated the mean μ as a one-to-one invertible function of u. We extend this to a many-to-one 'barber pole' regression curve by supposing that μ runs through an integral number, h, of cycles, as u runs through one cycle. The equation of such a regression curve is achieved by modifying (2·3) to

$$\tan\tfrac{1}{2}(\mu - \beta) = \omega \tan\tfrac{1}{2}(hu - \alpha), \tag{6·10}$$

with unique solution

$$\mu = \beta + 2\operatorname{atan}\{\omega \tan\tfrac{1}{2}(hu-\alpha)\}. \tag{6·11}$$

The centred form of (6·10) will be

$$\tan\tfrac{1}{2}\mu' = \omega \tan\tfrac{1}{2}u'', \tag{6·12}$$

where $\mu' = \mu - \beta$ and $u'' = hu - \alpha$. A centred planar graph of (6·11) will have h branches squeezed into the interval $(-\pi, \pi)$ on the horizontal u-axis. If this planar graph is metamorphosed into a cylinder, the metamorphosed planar graph becomes a barber pole, lying on its side, with h visible stripes.

6·4. *Squared correlation coefficient*

Assume $v\,|\,u \sim M\{\mu(u;\, \alpha, \beta, \omega), \kappa\}$ conditionally as in (2·10), and $v \sim M(\gamma, \kappa)$ marginally as in (2·7). Define the conditional variance, $V(v\,|\,u)$, and marginal variance, $V(v)$, by

$$\begin{aligned} V(v\,|\,u) &= E[\{\cos v - \cos\mu(u;\, \alpha, \beta, \omega)\}^2 + \{\sin v - \sin\mu(u;\, \alpha, \beta, \omega)\}^2] \\ &= 2\{1 - \rho(\alpha, \beta, \omega)\}, \\ V(v) &= E[\{\cos v - \cos\mu(u;\, \alpha, \beta, 0)\}^2 + \{\sin v - \sin\mu(u;\, \alpha, \beta, 0)\}^2] \\ &= 2\{1 - \rho_0(\gamma, 0)\}. \end{aligned} \tag{6·13}$$

A squared circular correlation coefficient, R^2, analogous to the squared linear correlation coefficient, is defined as the proportion of $V(u)$ explained by u, namely

$$R^2 = 1 - \{V(v\,|\,u)/V(v)\} = \{\rho(\alpha, \beta, \omega) - \rho_0(\gamma, 0)\}/\{1 - \rho_0(\gamma, 0)\}. \tag{6·14}$$

Our circular regression techniques are not based on any bivariate angular distribution.

Nevertheless, circular regressions done both ways, $v|u$ and $u|v$, lead to different estimates $\hat{R}^2$ for R^2 in (6·14). For example, computing $\hat{R}^2$ for $v|u$ and again for $u|v$ in Example 1 yields

$$\hat{R}^2_{v|u} = \frac{0{\cdot}9712 - 0{\cdot}7184}{1 - 0{\cdot}7184} = 0{\cdot}900, \quad \hat{R}^2_{u|v} = \frac{0{\cdot}9651 - 0{\cdot}7359}{1 - 0{\cdot}7359} = 0{\cdot}868,$$

which are distinctly different values. The search for a circular correlation coefficient with properties analogous to those of the linear correlation coefficient requires further study. We are unaware of any bivariate angular distribution that would provide a suitable model for developing such a correlation coefficient.

6·5. *Multiple circular regression*

We present a recursive method for regressing the dependent angle on multiple independent angles using a sequence of circle transforms for μ as in (2·4).

Stage 1. Summarise the model for regressing v on a single independent angle u_1 as $v = \mu_1 + e$, where $\mu_1 = \mu_1(u_1; \alpha_1, \beta_1, \omega_1)$ is a circle transform of u_1 and e is an angular error. The fitted version of the model can be expressed as $\hat{v} = \hat{\mu}_1 + \hat{e}_1$, where $\hat{\mu}_1$ and $\hat{e}_1$ are obtained by maximising the sample precision.

Stage 2. A second independent angle, u_2, may explain some of the variability in the fitted errors e_1 above, via the model $\hat{e}_1 = \mu_2 + e_2$, where $\mu_2 = \mu_2(u_2; \alpha_2, \beta_2, \omega_2)$ is a circle transform of u_2 and e_2 is now a new angular error. The fitted version of this model can be expressed as $\hat{e}_1 = \hat{\mu}_2 + \hat{e}_2$, where $\hat{\mu}_2$ and $\hat{e}_2$ are obtained by maximising the sample precision for this second model. Combining these results we obtain the fitted model $\hat{v} = \hat{\mu}_1 + \hat{\mu}_2 + \hat{e}_2$ with sample precision at least as great as that for the initial model. If we continue in this manner, the circular regression can be extended recursively to an arbitrary number of independent angles. Note that the method depends on the order in which the independent variables are introduced.

Acknowledgement

The authors wish to thank John Kent and Kevin de Souza for helpful discussions, Research Statistics, Inc. for facilitating completion of this work, and the referees for their detailed comments which have improved the paper considerably.

References

Binkley, S. (1990). *The Clockwork Sparrow: Time, Clock and Calendars in Biological Organisms.* Upper Saddle River, NJ: Prentice Hall.

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

Downs, T. & Mardia, K. V. (2000). A family of spherical regression models. In *The Statistics of Directions, Shapes and Images*, Ed. J. T. Kent and R. G. Aykroyd, pp. 22–3. Leeds: Leeds University Press.

Downs, T., Liebman, J. & Mackay, W. (1970). Statistical methods for vectorcardiogram orientations. In *Proc. XI Int. Vector Cardiography Symposium*, Ed. I. Hoffman, pp. 216–22. Amsterdam: North Holland.

Fisher, N. I. (1993). *Statistical Analysis of Circular Data.* Cambridge: Cambridge University Press.

Fisher, N. I. & Lee, A. J. (1992). Regression models for an angular response. *Biometrics* **48**, 665–77.

Follmann, D. A. & Proschan, M. A. (1999). A simple permutation-type method for testing circular uniformity with correlated angular measurements. *Biometrics* **55**, 782–91.

Gould, A. L. (1969). A regression technique for angular variates. *Biometrics* **25**, 683–700.

HRUSHESKY, W. J. M. (1985). Circadian timing of cancer chemotherapy. *Science* **228**, 73–5.
JOHNSON, R. A. & WEHRLY, T. (1978). Some angular-linear distributions and related regression models. *J. Am. Statist. Assoc.* **73**, 602–6.
JONES, M. C. & SILVERMAN, B. W. (1989). An orthogonal series density estimation approach to reconstructing positron emission tomography images. *J. Appl. Statist.* **16**, 177–91.
LOWREY, P. L., SHIMOMURA, K., ANTOCH, M. P., YAMAZAKI, S., ZEMENIDES, P. D., RALPH, M. R., MENAKER, M. & TAKAHASHI, J. S. (2000). Positional syntenic cloning and functional characterization of the mammalian circadian mutation tau. *Science* **288**, 483–91.
LUND, U. (1999). Least circular distance regression for directional data. *J. Appl. Statist.* **26**, 723–33.
MACKENZIE, J. K. (1957). The estimation of an orientation relationship. *Acta Crystallog.* **10**, 61–2.
MARDIA, K. V. (1975). Statistics of directional data (with Discussion). *J. R. Statist. Soc.* B **37**, 349–93.
MARDIA, K. V. & JUPP, P. E. (2000). *Directional Statistics.* Chichester: Wiley.
MCCULLAGH, P. (1996). Möbius transformation and Cauchy parameter estimation. *Ann. Statist.* **24**, 787–808.
MOORE-EDE, M. C., SULZMAN, F. M. & FULLER, C. A. (1982). *The Clocks that Time Us: Physiology of the Circadian Timing System.* Cambridge, MA: Harvard University Press.
NEEDHAM, T. (1997). *Visual Complex Analysis.* Oxford: Clarendon Press.
RIVEST, L.-P. (1997). A decentred predictor for circular regression. *Biometrika* **84**, 717–26.
SHEARMAN, L. P., SRIRAM, S., WEAVER, D. R., MAYWOOD, E. S., CHAVES, I., ZHENG, B., KUME, K., LEE, C. C., VAN DER HORST, G. T. J., HASTINGS, M. H. & REPPERT, S. M. (2000). Interacting molecular loops in the mammalian circadian clock. *Science* **288**, 1013–8.
STEPHENS, M. A. (1979). Vector correlation. *Biometrika* **66**, 41–8.
WATSON, G. S. & WILLIAMS, E. J. (1956). On the construction of significance tests on the circle and the sphere. *Biometrika* **43**, 344–52.
WEIR, I. S. & GREEN, P. J. (1994). Modelling data from single-photon emission computerized tomography. In *Statistics and Images* **2**, Ed. K. V. Mardia, pp. 313–8. Abingdon: Carfax Publishing Co.

[*Received September* 2000. *Revised March* 2002]