Biometrika (1997), **84**, 3, *pp.* 717–726
Printed in Great Britain

A decentred predictor for circular–circular regression

BY LOUIS-PAUL RIVEST
Département de mathématiques et de statistique, Université Laval, Ste-Foy, Québec G1K 7P4, Canada
e-mail: lpr@mat.ulaval.ca

SUMMARY

This paper considers the prediction of a y-angle, given an x-angle in a circular–circular regression. The predictor under study is a rotation of the decentred x-angle, defined as the direction of the sum of the unit vector corresponding to angle x plus a decentring vector. The parameters are defined as those which maximise the average residual cosine. This predictor is shown to perform well when the data come from a bivariate von Mises distribution or from a wrapped bivariate normal model. The large sample distributions of maximum cosine estimators are derived. Tests of fit for the rotational model, which predicts y by a rotation of x, and for the independence model where y is not related to x are presented. The proposed methods are illustrated by the analysis of an earthquake dataset where the direction of ground movement is regressed on the direction of steepest descent.

Some key words: Angular regression; Directional data; Earthquake ground movement; Test of independence; Von Mises distribution; Wrapped normal distribution.

1. INTRODUCTION

The literature on statistical methods for bivariate samples of angles (Fisher, 1993; Jupp & Mardia, 1989) has mostly been concerned with deriving correlation coefficients for measuring the dependency between angular samples. One model for circular–circular regression expresses the y-direction as a rotation of the x-direction. Further models are considered by Fisher & Lee (1992). Prediction of the y-direction using a rotation of the 'decentred' x-angle is considered in this paper. The decentred x-angle is the direction of the vector $(u_1 + \cos x, u_2 + \sin x)^{\mathrm{T}}$, where u_1 and u_2 are parameters to be estimated.

Decentred directions are useful tools for analysing directional data. They enter the construction of offset distributions (Mardia, 1972, Ch. 3). They were used by Jupp & Spur (1989) to estimate the source of a signal. Boulerice & Ducharme (1994) derive the distribution of a decentred direction $(u_1 + \cos x, u_2 + \sin x)^{\mathrm{T}}$ from that of x. The name 'decentred' comes from the following geometrical construction: if directions on the unit circle are recorded by an observer located at a point with coordinates $(-u_1, -u_2)$, then the direction $(\cos x, \sin x)^{\mathrm{T}}$ will be recorded by the observer as that of $(u_1 + \cos x, u_2 + \sin x)^{\mathrm{T}}$.

In § 2 we discuss the basic properties of the decentred predictor, while § 3 shows that decentred predictions closely approximate the conditional mean direction for two classes of bivariate symmetric distributions, the wrapped normals and a bivariate extension of the von Mises distribution. In § 4 we investigate the large sample distributions of the maximum cosine estimators of the parameters of the decentred model. Simple large sample tests for comparing the fit of the decentred model to those of the independence model and of the rotational model are proposed. An example concerned with the prediction of the direction of earthquake displacement in terms of the direction of steepest descent is presented in § 5.

2. BASIC PROPERTIES OF THE DECENTRED PREDICTOR

A convenient parameterisation for the decentring vector is $(u_1, u_2) = r(\cos\alpha, \sin\alpha)$, where r is a real number and α is an angle. In this notation, the direction of $(u_1 + \cos x, u_2 + \sin x)^{\mathrm{T}}$ can be written as the unit vector

$$\frac{1}{\{r^2+1+2r\cos(\alpha-x)\}^{\frac{1}{2}}}\begin{pmatrix} r\cos\alpha+\cos x \\ r\sin\alpha+\sin x \end{pmatrix}$$
$$=\frac{1}{\{r^2+1+2r\cos(x-\alpha)\}^{\frac{1}{2}}}\begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}\begin{pmatrix} r+\cos(x-\alpha) \\ \sin(x-\alpha) \end{pmatrix}.$$

A rotation of the above vector gives the general form of the decentred predicted direction as

$$\frac{1}{\{r^2+1+2r\cos(x-\alpha)\}^{\frac{1}{2}}}\begin{pmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{pmatrix}\begin{pmatrix} r+\cos(x-\alpha) \\ \sin(x-\alpha) \end{pmatrix}, \tag{1}$$

where β belongs to $[0, 2\pi)$. Rewriting (1) in terms of angles we obtain the following expression for the angle of the decentred predictor:

$$\begin{aligned}\theta(x;\, \beta, \alpha, r) &= \beta + \operatorname{atan}\{\sin(x-\alpha), r+\cos(x-\alpha)\} \pmod{2\pi} \\ &= \operatorname{atan}\{r\sin\beta+\sin(x+\beta-\alpha), r\cos\beta+\cos(x+\beta-\alpha)\} \pmod{2\pi},\end{aligned} \tag{2}$$

where $\operatorname{atan}(a, b)$ is the angle whose sine and cosine are proportional to a and b respectively. Thus $\operatorname{atan}(a, b) = \tan^{-1}(a/b) + h(b)\pi$, where h takes the value 1 if b is negative and 0 otherwise, and $\tan^{-1}$ is the inverse of the tangent function taking its values in $[-\pi/2, \pi/2]$. When r goes to ∞, (2) no longer depends on x. This is the independence model,

$$\theta(x;\, \beta, \alpha, \infty) = \beta. \tag{3}$$

Putting $r = 0$ in (2) yields the rotational model,

$$\theta(x;\, \beta, \alpha, 0) = \beta - \alpha + x. \tag{4}$$

When $r = \infty$, only β is identifiable while, when $r = 0$, only $\beta - \alpha$ is identifiable.

In (1) and (2), parameters α and β are related to the locations of x and y respectively. Parameter r determines the local slope of the decentred regression function. Differentiation of (2) shows that

$$\frac{d}{dx}\theta(x;\, \beta, \alpha, r) = \frac{r\cos(x-\alpha)+1}{r^2+2r\cos(x-\alpha)+1}.$$

At $x = \alpha$, the derivative is given by $1/(1+r)$. For x near α, (2) reduces to

$$\theta(x;\, \beta, \alpha, r) \simeq (x-\alpha)/(r+1).$$

Thus, when the x and the y directions are tightly clustered around α and β respectively, the decentred prediction model is approximately equivalent to a simple linear regression with a slope of $1/(1+r)$.

The range of possible angles for $\theta(x;\,\beta,\alpha,r)$ depends on r. When $r<1$, all directions are possible and $\theta(x;\,\beta,\alpha,r)$ goes around the circle once as x increases from 0 to 2π. The local slope varies between $1/(1+r)$ and $1/(1-r)$. When $r\geqslant 1$, the range of $\theta(x;\,\beta,\alpha,r)-\beta$ is contained in $[-\pi/2,\pi/2]$. The decentred predictions are clustered around β.

The following notation is used throughout the paper. Let σ_x and θ_x denote respectively the mean resultant length and the mean direction of the random angle x. In other words,

$$\sigma_x=\{E^2(\cos x)+E^2(\sin x)\}^{\frac{1}{2}}=E\{\cos(x-\theta_x)\},\quad \theta_x=\operatorname{atan}\{E(\sin x),E(\cos x)\}$$

and $E\{\sin(x-\theta_x)\}=0$. The mean direction θ_x is well defined only when σ_x is strictly positive. Also, if $\{x_i\colon i=1,\ldots,n\}$ is a random sample distributed as x, then $\bar{R}_x$ and $\bar{\theta}_x$ denote the empirical mean resultant length and mean direction respectively. This notation is easily extended: for instance, $\bar{\theta}_{y-x}$ is the mean direction of the sample $\{y_i-x_i\}$ while $\theta_{y|x}$ and $\sigma_{y|x}$ denote respectively the conditional mean direction and mean resultant length of y given x, which are defined as the direction and the length of the vector $E\{(\cos y,\sin y)^{\mathrm{T}}\,|\,x\}$.

3. Decentred predictions for symmetric distributions

Let $f(s,t)$ be the joint density of the random angles x and y. The dependence between x and y is assumed to be positive, that is $\sigma_{x-y}\geqslant\sigma_{x+y}$ (Rivest, 1982). If this condition is not met, it suffices to consider the distribution of y and $-x$. Maximising the average residual cosine is proposed in this section as a method for finding the decentred predictor for a particular density f. Comparisons between the decentred predictor and the conditional mean direction $\theta_{y|x}$ are also presented for two families of bivariate distributions.

The parameters of the decentred predictor for f are defined as the values maximising the average cosine residual angle defined as

$$L(\beta,\alpha,r)=E[\cos\{y-\theta(x;\,\beta,\alpha,r)\}],\tag{5}$$

where the expectation is taken with respect to the joint distribution of (x,y). In (5) the expectation with respect to the conditional distribution of y given x is easily evaluated. This yields

$$L(\beta,\alpha,r)=E[\sigma_{y|x}\cos\{\theta(x;\,\beta,\alpha,r)-\theta_{y|x}\}].\tag{6}$$

When x and y are independent, $\theta_{y|x}=\theta_y$ and $\sigma_{y|x}=\sigma_y$. Thus (6) implies that $L(\beta,\alpha,r)\leqslant\sigma_y$, with equality only if $r=\infty$ and $\beta=\theta_y$. The maximum cosine value of r is ∞ and the decentred predictor reduces to (3). When $y-x$ and x are independent one has $\theta_{y|x}=\theta_{y-x}+x$ and $r=0$ maximises (6); the decentred predictor then reduces to (4).

Suppose now that f is symmetric (Rivest, 1984) with respect to (θ_x,θ_y), that is,

$$f(s+\theta_x,t+\theta_y)=f(-s+\theta_x,-t+\theta_y)\tag{7}$$

for any (s,t) in $(0,2\pi)$. For r fixed, the partial derivatives of $L(\beta,\alpha,r)$ with respect to α and θ are null at $\alpha=\theta_x$ and $\beta=\theta_y$. For a large class of symmetric models these extrema are indeed maxima, and the optimal r is then the value maximising $L(\theta_y,\theta_x,r)$.

If $\theta_{y|x}$ is itself a decentred predictor, one has $\theta(x;\,\beta,\alpha,r)=\theta_{y|x}$ since, in (6), $L(\beta,\alpha,r)\leqslant E(\sigma_{y|x})$ with equality if $\theta(x;\,\beta,\alpha,r)=\theta_{y|x}$. In such instances, the conditional distribution of y can be expressed as the regression model

$$y = \theta(x; \beta, \alpha, r) + \varepsilon, \tag{8}$$

where ε has a distribution, possibly depending on x, with mean direction 0.

Consider, for instance, the symmetric exponential model of Rivest (1988), with $f(s, t)$ proportional to

$$\exp\{\kappa_1 \cos(s-\theta_x) + \kappa_2 \cos(t-\theta_y) + \gamma_1 \cos(s-\theta_x)\cos(t-\theta_y) + \gamma_2 \sin(s-\theta_x)\sin(t-\theta_y)\}, \tag{9}$$

where θ_x, θ_y are the mean directions, and κ_1, κ_2, γ_1, γ_2 are shape parameters. The conditional mean direction for this model is given by

$$\theta_{y|x} = \theta_y + \operatorname{atan}\{\gamma_2 \sin(x-\theta_x), \kappa_2 + \gamma_1 \cos(x-\theta_x)\}.$$

When $\gamma_1 = \gamma_2 = \gamma$, $\theta_{y|x}$ is a decentred predictor, and $(\beta, \alpha, r) = (\theta_y, \theta_x, \kappa_2/\gamma)$ maximises $L(\beta, \alpha, r)$. For this bivariate generalisation of the von Mises distribution, (8) holds with ε distributed as a von Mises distribution with a concentration parameter depending on x. Observe that, if in (9) one has $\gamma_2 = 0$ and $\kappa_2 > \gamma_1 > 0$, then $\theta_{y|x} = \theta_y$ and the maximum cosine value of r is ∞. This shows that the decentred predictor can be equal to (3) even when x and y are not independent.

Let $x = X \pmod{2\pi}$ and $y = Y \pmod{2\pi}$, where (X, Y) is distributed as a bivariate normal with mean vector (θ_x, θ_y), variances (γ_x^2, γ_y^2) and correlation coefficient ρ. The joint density of (x,y) is the wrapped bivariate normal; it is symmetric with respect to (θ_x, θ_y). The conditional mean direction of y given x is evaluated by first conditioning on X and then by taking the expectation with respect to the conditional distribution of X given x. This yields

$$\theta_{y|x} = \theta_y + \operatorname{atan}\left[\sum_{k=-\infty}^{\infty} \phi\left(\frac{x-\theta_x+2k\pi}{\gamma_x}\right)\cos\left\{\frac{\rho\gamma_y(x-\theta_x+2k\pi)}{\gamma_x}\right\},\right.$$
$$\left.\sum_{k=-\infty}^{\infty} \phi\left(\frac{x-\theta_x+2k\pi}{\gamma_x}\right)\sin\left\{\frac{\rho\gamma_y(x-\theta_x+2k\pi)}{\gamma_x}\right\}\right],$$

where $\phi(x) = \exp(-x^2/2)$. When the slope $\rho\gamma_y/\gamma_x$ is equal to 1 and $\frac{1}{2}$, the corresponding $\theta_{y|x}$ are decentred predictors given by $\theta_y + (x-\theta_x)$ and $\theta_y + (x-\theta_x)/2$ for x in $(\theta_x - \pi, \theta_x + \pi)$ respectively. Thus, from (6), the conditional mean direction and the decentred predictor are equal in these two instances. Numerical investigations revealed that, for the wrapped normal model, $\theta_{y|x}$ and $\theta(x; \theta_y, \theta_x, r)$ generally coincide almost perfectly except when γ_x is small, for x-values that are far from θ_x, provided that the slope $\rho\gamma_y/\gamma_x$ is less than 1·5. For larger slopes decentred predictors constructed with angle $[\frac{1}{2} + \rho\gamma_y/\gamma_x]x$, where $[z]$ denotes the integer part of z, approximate $\theta_{y|x}$ well. Thus, the wrapped bivariate normal is another instance where (8) holds, at least approximately, with the residual distribution depending on x.

4. Sampling properties of the estimated decentred model

4·1. *General*

Let $\{(x_i, y_i); i = 1, \ldots, n\}$ be a bivariate sample of angles drawn from some density $f(s, t)$. No restriction is imposed on $f(s, t)$ in this section. We propose nonparametric inference procedures that are valid, at least asymptotically, even when (8) is violated.

The maximum cosine estimators maximise an estimator of (5),

$$\hat{L}(\beta, \alpha, r) = \frac{1}{n}\sum_{i=1}^{n} \cos\{y_i - \theta(x_i; \beta, \alpha, r)\}.$$

They are maximum likelihood estimators under (8) when ε has the von Mises distribution with concentration parameters independent of x. From (2), $\cos\{y_i - \theta(x_i; \beta, \alpha. r)\}$ is equal to

$$\cos\beta \cos[y_i - \operatorname{atan}\{\sin(x_i - \alpha), r + \cos(x_i - \alpha)\}] + \sin\beta \sin[y_i - \operatorname{atan}\{\sin(x_i - \alpha), r + \cos(x_i - \alpha)\}].$$

This suggests that we calculate the estimates in two steps: first find $(\hat{\alpha}, \hat{r})$ by maximising the length of

$$\frac{1}{n}\sum_{i=1}^{n} \frac{1}{\{r^2 + 1 + 2r\cos(x_i - \alpha)\}^{\frac{1}{2}}} \begin{pmatrix} r\cos y_i + \cos(y_i - x_i + \alpha) \\ r\sin y_i + \sin(y_i - x_i + \alpha) \end{pmatrix}, \tag{10}$$

and then take $\hat{\beta}$ as the direction of the vector with maximum length.

4·2. *Large sample distributions of the maximum cosine estimators*

The density $f(s, t)$ is assumed to be such that the value r maximising $L(\beta, \alpha, r)$ is nonnull and finite; this ensures that β and α are identifiable. Methods for investigating the validity of this assumption are proposed in §§ 4·3 and 4·4. To investigate the asymptotic properties of $(\hat{\beta}, \hat{\alpha}, \hat{r})$ it is convenient to use the theory of estimating equations. Let $v(x; \beta, \alpha, r)$, be the vector of partial derivatives of $\theta(x; \beta, \alpha, r)$, with respect to β, α and r,

$$v(x; \beta, \alpha, r) = \frac{\partial}{\partial(\beta, \alpha, r)}\theta(x; \beta, \alpha, r) = \frac{1}{r^2 + 2r\cos(x - \alpha) + 1}\begin{pmatrix} r^2 + 2r\cos(x - \alpha) + 1 \\ -r\cos(x - \alpha) - 1 \\ -\sin(x - \alpha) \end{pmatrix}. \tag{11}$$

Differentiation of $\hat{L}(\beta, \alpha, r)$ yields the estimating equations for β, α and r,

$$\frac{1}{n}\sum_{i=1}^{n} v(x_i; \beta, \alpha, r)\sin\{y_i - \theta(x_i; \beta, \alpha, r)\} = 0. \tag{12}$$

The estimators $(\hat{\beta}, \hat{\alpha}, \hat{r})$ are solutions of (12). Thus, in large samples, one has

$$\begin{pmatrix} \hat{\beta} - \beta \\ \hat{\alpha} - \alpha \\ \hat{r} - r \end{pmatrix} \simeq A^{-1}\frac{1}{n}\sum_{i=1}^{n} v(x_i; \beta, \alpha, r)\sin\{y_i - \theta(x_i; \beta, \alpha, r)\},$$

where A, the matrix of the expected partial derivatives of the estimating equations, is equal to

$$E\left[\begin{pmatrix} 0 & 0 & 0 \\ 0 & (r - r^3)\sin(x - \alpha) & r^2\cos(x - \alpha) + 2r + \cos(x - \alpha) \\ 0 & r^2\cos(x - \alpha) + 2r + \cos(x - \alpha) & 2\sin(x - \alpha)\{r + \cos(x - \alpha)\} \end{pmatrix} \times \frac{\sin\{y - \theta(x; \beta, \alpha, r)\}}{\{r^2 + 2r\cos(x - \alpha) + 1\}^2}\right] - E[v(x; \beta, \alpha, r)v(x; \beta, \alpha, r)^{\mathrm{T}}\cos\{y - \theta(x; \beta, \alpha, r)\}].$$

Thus, $(\hat\beta, \hat\alpha, \hat r)$ is asymptotically normal with variance–covariance matrix given by

$$\operatorname{var}(\hat\beta, \hat\alpha, \hat r) = \frac{A^{-1}VA^{-1}}{n}, \tag{13}$$

where

$$V = E[v(x;\, \beta, \alpha, r)v(x;\, \beta, \alpha, r)^{\mathrm{T}} \sin^2\{y - \theta(x;\, \beta, \alpha, r)\}].$$

Under model (8), with homoscedastic errors, this variance–covariance matrix simplifies to

$$\operatorname{var}(\hat\beta, \hat\alpha, \hat r) = \frac{1}{n}\frac{E[\sin^2\{y - \theta(x;\, \beta, \alpha, r)\}]}{E^2[\cos\{y - \theta(x;\, \beta, \alpha, r)\}]} E\{v(x;\, \beta, \alpha, r)v(x;\, \beta, \alpha, r)^{\mathrm{T}}\}^{-1}.$$

When x is in a neighbourhood of α, the first two components of (11) are almost proportional. Thus, when the x-sample is clustered, the estimating equations are ill conditioned since the first two equations of (12) are almost identical. If the assumption of symmetry as given by (7) is tenable, one could set $\hat\beta = \bar\theta_y$, $\hat\alpha = \bar\theta_x$ and maximise $\hat L(\bar\theta_y, \bar\theta_x, r)$ to estimate r. Under symmetry, the matrix $\operatorname{var}(\hat\beta, \hat\alpha, \hat r)$ is block diagonal, with one block for $(\hat\beta, \hat\alpha)$ and one block for $\hat r$. Thus, when f is symmetric, the variance of $\hat r$ is given by the (3, 3) entry of $\operatorname{var}(\hat\beta, \hat\alpha, \hat r)$, namely,

$$\begin{aligned}\frac{1}{n}\Bigg[E\left[\frac{\sin^2(x-\theta_x)\sin^2\{y - \theta(x;\, \theta_y, \theta_x, r)\}}{\{r^2 + 2r\cos(x-\theta_x) + 1\}^2}\right] & \\ \times \{E([\sin^2(x-\theta_x)\cos\{y - \theta(x;\, \theta_y, \theta_x, r)\} - 2\sin(x-\theta_x)\{r + \cos(x-\theta_x)\}& \\ \times \sin\{y - \theta(x;\, \theta_y, \theta_x, r)\}]\{r^2 + 2r\cos(x-\theta_x) + 1\}^{-2})^2\}^{-1}\Bigg]&.\end{aligned}$$

The variance of $\hat\theta(x_p;\, \hat\beta, \hat\alpha, \hat r)$, the predicted y-angle at $x = x_p$, is easily calculated to be

$$\operatorname{var}\{\hat\theta(x_p;\, \beta, \alpha, r)\} = \frac{v(x_p;\, \hat\beta, \hat\alpha, \hat r)^{\mathrm{T}} \operatorname{var}(\hat\beta, \hat\alpha, \hat r) v(x_p;\, \hat\beta, \hat\alpha, \hat r)}{n}.$$

To derive prediction intervals, one could adapt some of the methods reviewed in Carroll & Ruppert (1988, Ch. 3) to study the variability of maximum cosine residuals.

Finally, note that all the variances reported in this section are easily estimated by replacing, in the variance formulae, the parameters by their estimates and the theoretical circular moments by their empirical counterparts. Since the underlying density f is usually unknown, estimating (13) is proposed to estimate the variance–covariance matrix of $(\hat\beta, \hat\alpha, \hat r)$.

4·3. *A test of fit for the rotational model*

The results of § 4·2 are valid only as long as the hypothesis H_0: $r = 0$ is false. In this section we construct a test for H_0. The proposed test statistic is the difference of the squared average residual cosine for the full decentred predictor minus that for the rotational model given by (4), $\hat L^2(\hat\beta, \hat\alpha, \hat r) - \hat L^2(\hat\beta_0, 0, 0)$, where $\hat\beta_0 = \bar\theta_{y-x}$ and $\hat L(\hat\beta_0, 0, 0) = \bar R_{y-x}$. The null distribution of this difference is given next; the proofs of Propositions 1 and 2 are given in Appendixes 1 and 2 respectively.

PROPOSITION 1. *If x and $y-x$ are independent, and if σ_{y-x} is nonnull, then asymptotically*

$$n\frac{\hat{L}^2(\hat{\beta},\hat{\alpha},\hat{r})-\hat{L}^2(\hat{\beta}_0,0,0)}{\sum \sin^2(y_i-x_i-\bar{\theta}_{y-x})/n}\sim\chi_2^2.$$

This test is consistent for any density f for which the maximum residual cosine satisfies $L(\beta,\alpha,r)>\sigma_{y-x}$. A score test, asymptotically equivalent to the statistic of Proposition 1, is easily constructed. Let S denote the 3×3 variance–covariance matrix of the z-sample, where

$$z_i=\{\sin(y_i-x_i-\hat{\theta}_{y-x}),\sin x_i,\cos x_i\}^{\mathrm{T}}\quad(i=1,\ldots,n),$$

and let $\hat{\rho}^2=\mathrm{tr}(S_{11}^{-1}S_{12}S_{22}^{-1}S_{21})$ be the squared canonical correlation between the first component of z, that is $\{\sin(y_i-x_i-\hat{\theta}_{y-x})\}$, and $\{(\sin x_i,\cos x_i)^{\mathrm{T}}\}$. Under H_0, it is shown in Appendix 1 that the difference between the statistic of Proposition 1 and $n\hat{\rho}^2$ is $o_p(1)$. Therefore the decentred model improves on the rotational model when the sines of the rotational residuals are linearly related to the unit vector of angle x.

4·4. *A test of independence*

To investigate if the decentred model provides a better fit than the independence model, see (3), one can test H_0: $1/r=0$. A suitable test statistic, similar to that of § 4·3, is given in the following proposition together with its large sample distribution.

PROPOSITION 2. *If x and y are independent and if σ_y is nonnull, then the asymptotic null distribution of the test statistic for independence,*

$$n\frac{\hat{L}^2(\hat{\beta},\hat{\alpha},\hat{r})-\hat{L}^2(\hat{\beta}_0,0,\infty)}{\sum \sin^2(y_i-\bar{\theta}_y)/n},$$

where $\hat{\beta}_0=\bar{\theta}_y$ and $\hat{L}(\hat{\beta}_0,0,\infty)=\bar{R}_y$, is χ_2^2.

In Proposition 2, we must have that $\sigma_y>0$. Boulerice & Ducharme (1994) showed that the mean resultant length of $\mathrm{atan}(\sin x, r+\cos x)$ is nonnull, even when x is uniformly distributed on the circle, except of course when r is null. This suggests that decentred predictions are appropriate for y only when σ_y is positive.

In Propositions 1 and 2 the assumption of independence is necessary to get a χ_2^2 asymptotic distribution. For instance, if f is given by (9) with $\gamma_2=0$ and $\kappa_2>\gamma_1>0$, then the maximum cosine parameter is $r=\infty$ even if x and y are not independent. For such a density, the test statistic of Proposition 2 does not have a limiting χ_2^2 distribution.

A score statistic asymptotically equivalent to the statistic of Proposition 2 can be obtained as in § 4·3. It is the squared canonical correlation between $\sin(y_i-\bar{\theta}_y)$ and $(\sin x_i,\cos x_i)$. This score statistic can be regarded as a component of the Jupp & Mardia (1980) independence statistic, which is the sum of the squared canonical correlations between $(\cos y_i,\sin y_i)$ and $(\cos x_i,\sin x_i)$.

5. EXAMPLE

The basic data, collected by Prof. Hamada from Tokai University in Japan, consist of two sets of aerial photographs taken before and after an earthquake in the town of Noshiro, Japan (Hamada & O'Rourke, 1992). On each set of photographs, coordinates (x, y, z), related to east–west and north–south distances and altitude, respectively, were calculated for 763 points of the site by photogrammetric analysis. The earthquake ground movement for the 763 points was obtained by subtracting the pre-earthquake from the post-earthquake coordinates. The directions of steepest descent were estimated as the directions of the gradients in a pre-earthquake local regression of z on x and y (Ruppert, 1997). The data under study are $\{x_i, y_i; i=1, \ldots, 763\}$, where x_i represents the pre-earthquake direction of steepest descent and y_i stands for the direction of lateral ground movement.

Since $\bar{R}_y = 0{\cdot}16$, σ_y is nonzero and independence can be tested with the statistic of § 4·4. The canonical correlation for independence is $\hat{\rho} = 0{\cdot}35$ yielding $n\hat{\rho}^2 = 93{\cdot}47$. The maximum average residual cosine for the decentred model is $\hat{L}(\hat{\beta}, \hat{\alpha}, \hat{r}) = 0{\cdot}474$; this leads to a value of 322·49 for the χ_2^2 test statistic of Proposition 2. Both tests are highly significant and the assumption of independence is rejected. For the rotational model $\bar{R}_{y-x} = 0{\cdot}454$. The statistic of Proposition 1 is 38·0 while the corresponding canonical correlation is 0·19, yielding a score test statistic of 27·5. Both tests are significant at the 0·01 level. Thus the rotational model does not fit well; decentring improves the predictions significantly. The maximum cosine estimates, the angles are expressed in radians, are $\hat{\beta} = -0{\cdot}89$, $\hat{\alpha} = -0{\cdot}98$, $\hat{r} = -0{\cdot}45$, and the estimated variance–covariance matrix, obtained from formula (13), is

$$v(\hat{\beta}, \hat{\alpha}, \hat{r}) = 10^{-3}\begin{pmatrix} 16{\cdot}3 & 15{\cdot}6 & 2{\cdot}2 \\ 15{\cdot}6 & 17{\cdot}4 & 0{\cdot}9 \\ 2{\cdot}2 & 0{\cdot}9 & 6{\cdot}4 \end{pmatrix}.$$

The hypothesis H_0: $\beta = \alpha$ is of interest. If it is true, then the predictor for y is the unrotated decentred x-angle. This is plausible for this dataset since ground movement should be closely related to the unrotated direction of steepest descent. A Wald test statistic is easily constructed for H_0, namely,

$$z_{\text{obs}} = \frac{\hat{\alpha} - \hat{\beta}}{\{v(\hat{\alpha}) + v(\hat{\beta}) - 2\,\text{cov}(\hat{\alpha}, \hat{\beta})\}^{\frac{1}{2}}},$$

where $v(.)$ and $\text{cov}(.,.)$ denote variances and covariances appearing in $v(\hat{\beta}, \hat{\alpha}, \hat{r})$. One obtains $z_{\text{obs}} = -1{\cdot}8$, which is not significant at the 0·05 level. Thus H_0 is not rejected; this suggests taking $\hat{\beta} = \hat{\alpha} = -0{\cdot}93$ for constructing the decentred predictor:

$$\hat{\theta}(x;\ -0{\cdot}93,\ -0{\cdot}93,\ -0{\cdot}45) = \text{atan}(0{\cdot}362 + \sin x;\ -0{\cdot}267 + \cos x).$$

Exploratory analysis of the residuals $\{y_i - \hat{\theta}(x_i;\ -0{\cdot}93,\ -0{\cdot}93,\ -0{\cdot}45)\}$ revealed the presence of heteroscedasticity.

The analysis shows that direction of steepest descent is not the only determinant of the direction of ground movement. Other factors, possibly related to the direction of earthquake propagation, influence ground movement. Further modelling, using the slope of steepest descent as an additional predictor variable, should improve on the decentred model.

6. DISCUSSION

The decentred predictor is a useful tool for circular regression. It has a simple geometrical interpretation, it provides good y-predictions for some important bivariate angular models, and the sampling properties of the maximum cosine estimators of the parameters are tractable. The application of the decentred predictor to angular time series (Breckling, 1989, p. 183; Fisher & Lee, 1994) is a promising area for future research.

Acknowledgement

This research was carried out while I was visiting the Biometrics Unit of Cornell University. I thank all the members of the Unit for their hospitality. I am grateful to T. O'Rourke and A. Schulman for making the dataset discussed in § 5 available to me. I am also grateful to a referee for his detailed comments. The financial support of the Natural Sciences and Engineering Research Council of Canada is gratefully acknowledged.

Appendix 1

Proof of Proposition 1

Without loss of generality assume that $E(\sin x) = E\{\sin(y-x)\} = 0$. If these expectations are nonzero, it suffices to subtract θ_x from x and $\theta_x + \theta_{y-x}$ from y to make them zero. Let $u_1 = r\cos\alpha$ and $u_2 = r\sin\alpha$. With this new parameterisation, a rotation of angle $-\alpha$ of (10) yields

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{(u_1^2+u_2^2+1+2u_1\cos x_i+2u_2\sin x_i)^{\frac{1}{2}}}\begin{pmatrix}u_1\cos y_i+u_2\sin y_i+\cos(y_i-x_i)\\ u_1\sin y_i-u_2\cos y_i+\sin(y_i-x_i)\end{pmatrix}.$$

Since rotating a vector does not change its length, $\hat{L}(\hat{\alpha},\hat{\beta},\hat{r})$ is the maximum length of the above vector. Its squared length can be written in terms of the angles $\tau(x)$, defined as $\operatorname{atan}(u_2+\sin x, u_1+\cos x)$, as follows:

$$\hat{L}^2(u_1,u_2) = \left[\frac{1}{n}\sum\cos\{y_i-\tau(x_i)\}\right]^2 + \left[\frac{1}{n}\sum\sin\{y_i-\tau(x_i)\}\right]^2.$$

To obtain an approximation to $\hat{L}^2(\hat{\beta},\hat{\alpha},\hat{r})$, one can expand $\hat{L}^2(u_1,u_2)$ in a Taylor series around $(0,0)$. This yields

$$\hat{L}^2(u_1,u_2) = \bar{R}^2_{y-x} + d(u_1,u_2)^{\mathrm{T}} + \tfrac{1}{2}(u_1,u_2)H(u_1,u_2)^{\mathrm{T}} + o_p(u^2), \tag{A1·1}$$

where d is the vector of partial derivatives of $\hat{L}^2(u_1,u_2)$ and H is the matrix of second derivatives, both evaluated at $u_1=u_2=0$. If H is negative definite, then an approximation to $\hat{L}^2(\hat{\beta},\hat{\alpha},\hat{r})$ is obtained by maximising the quadratic form in (A1·1). This yields $o_p(n^{-\frac{1}{2}})$ approximations to the values of (u_1,u_2) maximising $\hat{L}^2(u_1,u_2)$. Thus

$$\hat{L}^2(\hat{\beta},\hat{\alpha},\hat{r}) = \bar{R}^2_{y-x} - d^{\mathrm{T}}H^{-1}d/2 + o_p(n^{-1}).$$

To complete the proof, one needs to derive the asymptotic distribution of $d^{\mathrm{T}}H^{-1}d/2$. Let $w(x;u_1,u_2)$ be the vector of partial derivatives of $\tau(x)$,

$$w(x;u_1,u_2) = \frac{\partial\tau(x)}{\partial(u_1,u_2)} = \frac{1}{1+u_1^2+u_2^2+2u_1\cos x+2u_2\sin x}\begin{pmatrix}-u_2-\sin x\\ u_1+\cos x\end{pmatrix}.$$

The vector of partial derivatives of $L^2(u_1,u_2)$ with respect to u_1 and u_2 is

$$2\sum\frac{\cos\{y_i-\tau(x_i)\}}{n}\sum\frac{w(x_i;u_1,u_2)\sin\{y_i-\tau(x_i)\}}{n}$$
$$-2\sum\frac{\sin\{y_i-\tau(x_i)\}}{n}\sum\frac{w(x_i;u_1,u_2)\cos\{y_i-\tau(x_i)\}}{n}.$$

When $u_1=u_2=0$, $\tau(x)=x$. In evaluating d, only terms that are $O_p(n^{-\frac{1}{2}})$ when $u_1=u_2=0$ need to be kept, so that

$$d = 2\frac{\sigma(y-x)}{n}\sum\sin(y_i-x_i)\begin{pmatrix}-\sin x_i\\ \cos x_i-\sigma_x\end{pmatrix} + o_p(n^{-\frac{1}{2}}).$$

In a similar way, when evaluating H, one only needs to retain terms that are $O_p(1)$. This yields

$$H = 2\sigma^2(y-x)\begin{pmatrix} -E(\sin^2 x) & E(\cos x \sin x) \\ E(\cos x \sin x) & -E\{(\cos x - \sigma_x)^2\} \end{pmatrix} + o_p(1).$$

Since d is approximately equal to the average of independent identically distributed random vectors, by the Central Limit Theorem, its large sample distribution is bivariate normal with zero mean and variance–covariance matrix equal to $-2E\{\sin^2(y-x)\}H/n$. Therefore

$$n\frac{\hat{L}^2(\hat\beta, \hat\alpha, \hat r) - \hat{L}^2(\hat\beta_0, 0, 0)}{\sum \sin^2(y_i - x_i - \bar\theta_{y-x})/n} = nd^{\mathrm{T}}[-2E\{\sin^2(y-x)\}H]^{-1}d + o_p(1), \qquad (A1\cdot 2)$$

and the large sample distribution of the test statistic is χ^2_2.

If we use in (A1·2) the approximations

$$2S_{12} \simeq d/\sigma(y-x), \quad 2S_{22} \simeq H/\sigma^2(y-x), \quad S_{11} \simeq E\{\sin^2(y-x)\},$$

where S is the variance–covariance matrix of the z-sample defined in § 4·3, then the test statistic of Proposition 1 is seen to be asymptotically equivalent to the score statistic $n\hat\rho^2$ of § 4·3.

APPENDIX 2

Proof of Proposition 2

Reparameterising (13) in terms of $u_1 = \cos\alpha/r$ and $u_2 = \sin\alpha/r$, we obtain

$$\frac{1}{n}\sum_{i=1}^{n} \frac{1}{(u_1^2 + u_2^2 + 1 + 2u_1 \cos x_i + 2u_2 \sin x_i)^{\frac{1}{2}}} \begin{pmatrix} \cos y_i + u_1 \cos(y_i - x_i) + u_2 \sin y_i - x_i) \\ \sin y_i + u_1 \sin(y_i - x_i) - u_2 \cos y_i - x_i) \end{pmatrix}.$$

The argument follows that for Proposition 1 with the angles y_i and $y_i - x_i$ interchanged.

REFERENCES

BOULERICE, B. & DUCHARME, G. (1994). Decentred directional data. *Ann. Statist. Math.* **46**, 573–86.

BRECKLING, J. (1989). *The Analysis of Directional Time Series: Applications to Wind Speed & Direction*, Lecture Notes in Statistics #61. New York: Springer Verlag.

CARROLL, R. J. & RUPPERT, D. (1988). *Transformation and Weighting in Regression.* New York: Chapman & Hall.

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

FISHER, N. I. & LEE, A. J. (1992). Regression models for angular responses. *Biometrics* **48**, 665–77.

FISHER, N. I. & LEE, A. J. (1994). The analysis of directional time series. *J. R. Statist. Soc.* B **56**, 327–39.

HAMADA, M. & O'ROURKE, T. (1992). *Case Studies of Liquefaction & Lifeline Performance During Past Earthquake, 1: Japanese Case Studies.* SUNY at Buffalo, Red Jacket Quadrangle, Buffalo, NY 14261: National Center for Earthquake Engineering Research.

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

JUPP, P. E. & MARDIA, K. V. (1989). A unified view of the theory of directional statistics, 1975–1988. *Int. Statist. Rev.* **57**, 261–94.

JUPP, P. E. & SPURR, B. (1989). Statistical estimation of a shock center: Slate Islands astrobleme. *Math. Geol.* **21**, 191–7.

MARDIA, K. V. (1972). *Statistics of Directional Data.* New York: Academic Press.

RIVEST, L.-P. (1982). Some statistical methods for bivariate circular data. *J. R. Statist. Soc.* B **44**, 81–90.

RIVEST, L.-P. (1984). Symmetric distributions for dependent unit vectors. *Ann. Statist.* **12**, 1050–7.

RIVEST, L.-P. (1988). A distribution for dependent unit vectors. *Commun. Statist.* B **17**, 461–83.

RUPPERT, D. (1997). Local polynomial regression and its application in environmental statistics. In *Statistics for the Environment, 3*, Ed. V. Barnett and F. Turkman. To appear. New York: John Wiley.

[*Received February* 1996. *Revised December* 1996]