Modelling strategies for bivariate circular data

John T. Kent*, Kanti V. Mardia, & Charles C. Taylor

Department of Statistics, University of Leeds

1 Introduction

On the torus there are two common approaches to constructing a distribution which is the analogue of the bivariate normal distribution in the plane. These approaches are often termed the sine and cosine models, respectively, and in addition the cosine model comes in two versions. Each approach has its strengths and weaknesses. In this paper we develop a hybrid version which combines the strengths of each approach. The development of bivariate circular models has recently become important in applications to protein structure in bioinformatics (e.g. Mardia et al., 2007, and Boomsma et al., 2008).

A general bivariate circular model, which we call the "full" bivariate von Mises distribution, was introduced by Mardia (1975a),

$$f(\theta_1, \theta_2) \propto \exp\left\{\kappa_1 \cos(\theta_1 - \mu_1) + \kappa_2 \cos(\theta_2 - \mu_2) + \begin{bmatrix}\cos(\theta_1 - \mu_1) & \sin(\theta_1 - \mu_1)\end{bmatrix} A \begin{bmatrix}\cos(\theta_2 - \mu_2) \\ \sin(\theta_2 - \mu_2)\end{bmatrix}\right\}, \tag{1}$$

where the angles $\theta_1, \theta_2 \in (-\pi, \pi)$ lie on the torus, that is, a square with opposite sides identified, and where A is a 2×2 matrix. This model has eight parameters and allows for dependence between the two angles. However, since the analogous bivariate normal distribution in the plane contains only 5 parameters, this bivariate circular model seems overparameterized. Hence various submodels have been proposed.

The starting point is the 6-parameter model of Rivest (1988) and Mardia (1975b) obtained by setting the off-diagonal elements of A equal to 0 in (1), $a_{12} = a_{21} = 0$. In the each of the subsequent models, one degree of freedom in removed from the Rivest-Mardia model, leaving 5 parameters to mimic the bivariate normal distribution. To simplify the presentation, put the mean angle parameters equal to 0, $\mu_1 = \mu_2 = 0$, so that there are three remaining parameters to describe the concentration of each angle and their interaction. These models are
(i) the cosine model with positive interaction

$$f(\theta_1, \theta_2) \propto \exp\{\kappa_1 \cos\theta_1 + \kappa_2 \cos\theta_2 + \gamma_1 \cos(\theta_1 - \theta_2)\}, \tag{2}$$

(ii) the cosine model with negative interaction

$$f(\theta_1, \theta_2) \propto \exp\{\kappa_1 \cos\theta_1 + \kappa_2 \cos\theta_2 + \gamma_2 \cos(\theta_1 + \theta_2)\}, \tag{3}$$

(iii) the sine model (Singh et al., 2002)

$$f(\theta_1, \theta_2) \propto \exp\{\kappa_1 \cos\theta_1 + \kappa_2 \cos\theta_2 + \delta \sin\theta_1 \sin\theta_2\}. \tag{4}$$

Mardia et al. (2007) gives a systematic study of models (i) - (iii) and a comparison between them.

Under high concentration about $(\theta_1, \theta_2) = (0, 0)$, each of the models behaves as a bivariate normal distribution with inverse covariance matrix of the form

$$\Sigma_1^{-1} = \begin{bmatrix}\kappa_1 + \gamma_1 & -\gamma_1 \\ -\gamma_1 & \kappa_2 + \gamma_1\end{bmatrix}, \quad \Sigma_2^{-1} = \begin{bmatrix}\kappa_1 + \gamma_2 & \gamma_2 \\ \gamma_2 & \kappa_2 + \gamma_2\end{bmatrix} \quad \Sigma_3^{-1} = \begin{bmatrix}\kappa_1 & -\delta \\ -\delta & \kappa_2\end{bmatrix}. \tag{5}$$

By high concentration we mean that the relevant parameters from $\kappa_1,\ \kappa_2,\ \delta,\ \gamma_1,\ \gamma_2$ get large while remaining in constant proportion to one another, under the constraint that the inverse covariance matrix is positive definite. For the three models this constraint reduces to

$$\text{(i) } \kappa_1+\gamma_1>0,\ \kappa_2+\gamma_1>0,\ \gamma_1^2<(\kappa_1+\gamma_1)(\kappa_2+\gamma_1),$$
$$\text{(ii) } \kappa_1+\gamma_2>0,\ \kappa_2+\gamma_2>0,\ \gamma_2^2<(\kappa_1+\gamma_2)(\kappa_2+\gamma_2),$$
$$\text{(iii) } \kappa_1>0,\ \kappa_2>0,\ \delta^2<\kappa_1\kappa_2.$$

In addition, for the cosine models we require $\kappa_1 > 0,\ \kappa_2 > 0$ to ensure the global mode of f is at $(\theta_1, \theta_2) = (0, 0)$.

For each of the three models, it is possible to choose the parameters to match any positive definite inverse covariance matrix. However, the cosine model can show some unattractive multimodal behaviour if $\kappa_1 < 0$ or $\kappa_2 < 0$. Hence we restrict attention to cosine models for which $\kappa_1 > 0,\ \kappa_2 > 0$. In this case the corresponding inverse covariance matrix Σ^{-1} possesses what can be called the "dominated covariance" property, that is, $|\sigma^{12}| < \sigma^{11}$ and $|\sigma^{12}| < \sigma^{22}$.

2 Symmetry properties

By construction, each of the sine and cosine models is symmetric, $f(\theta_1, \theta_2) = f(-\theta_1, -\theta_2)$. This symmetry accommodates an elliptical pattern in the contours of constant probability density of f about the mode $(\theta_1, \theta_2) = (0, 0)$.

However, for each of the three models, f has a further symmetry since it is a continuous function on the torus,

$$f(\theta_1, \pi) = f(\theta_1, -\pi), \quad f(\pi, \theta_2) = f(-\pi, \theta_2).$$

This latter property means that an elliptical pattern in the contours of constant probability for f will generally become distorted as (θ_1, θ_2) approaches the boundary of the square on which f is defined. In particular, this distortion complicates the development of efficient simulation algorithms using a 2-dimensional envelope since the density will not necessarily be monotonically decreasing on the rays from the origin to the edge of the square.

For simplicity restrict attention to the dominated covariance situation. It turns out the positive-interaction cosine model involves the least distortion under postive correlation ($\gamma_1 > 0$) and that the negative-interaction cosine model involves the least distortion under negative correlation ($\gamma_2 > 0$). Ideally it would be nice to use a positive-interaction cosine model under positive correlation between $\sin\theta_1$ and $\sin\theta_2$ and negative-interaction cosine model under negative correlation. Unfortunately the crossover between the two models is not continuous at the independence model. Hence we consider a hybrid model to provide a smooth transition.

3 A hybrid model

For small concentration, the exact character of any departure from the uniform distribution is not too important. Hence we suggest the following hybrid model:

$$f(\theta_1, \theta_2) \propto \exp\left\{\kappa_1 \cos\theta_1 + \kappa_2 \cos\theta_2 + \beta[(\cosh\lambda - 1)\cos\theta_1 \cos\theta_2 + \sinh\lambda \sin\theta_1 \sin\theta_2]\right\} \quad (6)$$

The parameter β is a tuning parameter which we fix to the value 1 for simplicity (note that when both β and λ are free parameters, model (6) is just a reparameterization of the Rivest-Mardia model). For λ near 0 the model behaves like a sine model with $\beta\lambda \approx \delta$. For large $\lambda > 0$ (or large $-\lambda > 0$) the model behaves like a positive-interaction (or negative-interaction) cosine model with $\beta\exp(\lambda_1)/2 \approx \gamma_1$ (or $\beta\exp(-\lambda)/2 \approx \gamma_2$).

Thus for small correlation the model behaves as a sine model, and for large correlation the model behaves as a cosine model, with positive or negative interaction as appropriate.

4 Simulation

Consider the problem of simulating from the full distribution (1) or one of its subfamilies. One possibility is to use an MCMC algorithm based on the fact that the conditional distributions of $\theta_1|\theta_2$ and $\theta_2|\theta_1$ are von Mises (Mardia et al., 2008a). However, such a strategy can be overly cumbersome.

At least for the sine and cosine models there is a simpler approach. First simulate θ_1 from its marginal distribution and then simulate $\theta_2|\theta_1$ from the von Mises distribution using, e.g., the Best-Fisher algorithm (Best and Fisher, 1979). The web supplements to Mardia et al. (2007) and to Boomsma et al. (2008) discuss the empirical selection of a suitable von Mises distribution in the unimodal case (or a mixture of of two such distributions in the bimodal case) to use an an envelope in an acceptance-rejection algorithm. More recently, in unpublished work a theoretical justification has been found to confirm the appropriateness and efficiency of the von Mises envelope in the unimodal case.

5 Discussion

The geometry of the torus implies that it is not possible to get a single fully satisfactory analogue of the bivariate normal distribution. Though a complete comparison between the virtues of the cosine and sine models is not yet available, it is possible to make some interim conclusions.

(a) In most situations there is not much difference between the sine and cosine models. Further, under high concentration, using either model is equivalent to fitting a bivariate normal distribution in a tangent plane.

(b) For routine applications the sine model is somewhat easier to use, since it can be matched to any positive definite matrix Σ^{-1}, whereas the cosine models are limited to the dominated covariance case.

(c) However, if a more refined model is needed, the cosine or hybrid models may provide a better fit.

(d) For any of the models, statistical inference is intractable using the full likelihood, but becomes straightforward using a composite likelihood (sometimes called the pseudo-likelihood) obtained by taking a product of the conditional densities (Mardia et al., 2008a). Limited evidence at present suggests that the marginal angular distributions will be closer to the von Mises distribution for the cosine model than for the sine model, and that the composite likelihood estimation will be more efficient in this situation.

(e) The sine and cosine models on the bivariate torus can be easily extended a higher dimensional torus (Mardia and Patrangenaru, 2005, and Mardia et al., 2008b).

References

Best, D.J. and Fisher, N.I. (1979). Efficient simulation of the von Mises distribution. *Appl. Statist.*, **24**, 152–157.

Boomsma, W., Mardia, K.V., Taylor, C.C., Ferkinghoff-Borg, J., Krogh, A., and Hamelryck, T. (2008) A generative, probabilistic model of local protein structure. *PNAS*, in press.

Mardia, K.V. (1975a). Statistics of directional data (with discussion) *J. Royal Statist. Soc., Series B*, **37**, 349–393.

Mardia, K.V. (1975b). Characterization of directional distributions. In G.P. Patil, S. Kotz and J.K. Ord. (eds), *Statistical Distributions in Scientific Work, 3*, pp365–386. D. Reidel Publishing Co. Dordrecht (Holland) and Boston (USA).

Mardia, K.V., Hughes, G., Taylor, C.C., and Singh, H. (2008a). A multivariate von Mises distribution with applications to bioinformatics. *Canadian Journal of Statistics*, **36**, 99–109.

Mardia, K.V., Kent, J.T., Hughes, G., and Taylor, C.C. (2008b). Statistical estimation using composite likelihoods. Submitted.

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

Mardia, K.V., Taylor, C.C. and Subramaniam, G.K. (2007). Protein bioinformatics and mixtures of bivariate von Mises distributions for angular data. *Biometrics*, **63**, 505–512.

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

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