Assessing a Bayesian Embedding Approach to Circular Regression Models

Jolien Cremers[1], Irene Klugkist[1], Tim Mainhard[2]

Second Bayesian Young Statisticians Meeting (BAYSM 2014)
Vienna, September 18–19, 2014

[1] Utrecht University, Department of Methodology and Statistics, Utrecht, The Netherlands
j.cremers@uu.nl
i.klugkist@uu.nl

[2] Utrecht University, Department of Education, Utrecht, The Netherlands
m.t.mainhard@uu.nl

Abstract

Circular data is different from linear data and its analysis therefore also requires methods that are different from the conventional methods. In this study the Bayesian embedding approach to estimating circular regression models as presented in Nuñez-Antonio, Gutirrez-Pea and Escarela [3] is investigated, by means of five simulation studies, in terms of performance, efficiency and flexibility. In addition an empirical example of a regression model predicting teachers' scores on the interpersonal circumplex will be used throughout. The performance in terms of bias and coverage of the credible interval is reasonable in most situations and the method is deemed efficient and very flexible.
Keywords: circular data; Bayesian methods.

1 Introduction

In this study regression models for circular data are considered. Circular data is different from linear data in the sense that it contains information about directions or angles. Therefore, the analysis of circular data also requires methods that are different from the conventional methods for linear data. Mardia and Jupp [2] and Fisher [1] provide an overview of frequentist methods to estimate circular regression models. The downside of frequentist methods is that they are more complex [4], less flexible and may be less efficient than Bayesian estimation methods for circular regression models.

Only a few Bayesian methods for estimating circular regression models are available in the literature. Amongst them Nuñez-Antonio et al. [3] provide an MCMC method for circular regression based on a projected normal distribution; an embedding approach. Although Nuñez-Antonio et al. [3] introduce the mathematical model and developed a Bayesian sampler to estimate circular regression models, no extensive simulations to assess the method have been conducted so far. Therefore, in this study the Bayesian embedding approach as presented in Nuñez-Antonio et al. [3] will be investigated by means of simulation studies in terms of performance, efficiency and flexibility.

2 The Embedding Approach

As used in this paper, the embedding approach assumes that the circular outcome variable has a projected bivariate normal distribution. In regression models, the projected bivariate normal distribution has the following density [3]:

$$PN(\theta|\boldsymbol{\mu}, \boldsymbol{I}) = \frac{1}{2\pi} e^{-\frac{1}{2}||\boldsymbol{\mu}||^2} \left[1 + \frac{\boldsymbol{u}^t \boldsymbol{\mu} \Phi(\boldsymbol{u}^t \boldsymbol{\mu})}{\phi(\boldsymbol{u}^t \boldsymbol{\mu})} \right], \tag{1}$$

where:

- θ $(0 < \theta \leq 2\pi)$ is the circular random variable
- $\boldsymbol{u} = (cos\,\theta, sin\,\theta)^t$
- $\boldsymbol{\mu} = \boldsymbol{B}^t \boldsymbol{x}$
- $\boldsymbol{x}$ is a matrix with predictor variables
- $\boldsymbol{B} = [\boldsymbol{\beta}^1, \boldsymbol{\beta}^2]$, $\boldsymbol{\beta}^1$ and $\boldsymbol{\beta}^2$ are vectors with regression coefficients. Having two components this means the sine and the cosine of the outcome variable can be predicted using different regression equations.
- $\Phi(\cdot)$ is the cdf and $\phi(\cdot)$ is the pdf of the standard normal distribution.

The relation between the circular and bivariate normal outcomes is defined as $U = Y/R$ where U is a random direction, Y is a random bivariate normal vector and $R = ||Y_i||$.

2.1 Bayesian Estimation

In this study, an uninformative normal prior is specified for the two components of $\mathbf{B}$:

$$N(\boldsymbol{\beta}^j | \boldsymbol{\beta}_0^j, \boldsymbol{\Lambda}_0^j)\ \forall\, j = 1, 2, \tag{2}$$

where $\boldsymbol{\beta}_0^j = (0, 0)$ are prior values for the regression coefficients and intercept of component j and $\boldsymbol{\Lambda}_0^j = (0.0001, 0.0001)$ is the prior precision matrix.
We consider a latent variable R defined on (0,∞). This latent variable, together with the prior and likelihood results in the posterior:

$$f(\theta, r | \boldsymbol{\mu} = \boldsymbol{B}^t \boldsymbol{x}) = N_2(r\boldsymbol{u} | \boldsymbol{\mu} = \boldsymbol{B}^t \boldsymbol{x}, \boldsymbol{I})|J|, \tag{3}$$

where $|J| = r$ is the Jacobian of the transformation $\boldsymbol{y} \mapsto (\theta, r)$ and $\mathbf{u} = (\cos\,\theta, \sin\,\theta)^t$. More specifics can be found in Nuñez-Antonio et al. [3].

3 Simulations

In order to assess the performance, efficiency and flexibility of the method for estimating circular regression models illustrated above, five simulation studies are conducted. These studies estimate regression models with varying types and amounts of predictors. The parameters of interest are the regression coefficients and intercepts contained in $\boldsymbol{B}$. Manipulated parameters are the sample size, the population value for the regression parameters, the mean of the linear predictor variables and the concentrations of the circular predictor variables. The results are assesed on the basis of bias, coverage of the 95% credible interval, speed of convergence and mean computation time (the time it takes to obtain one sample from the posterior averaged over all simulated datasets).

4 Conclusions

From the results of the simulation studies we may conclude that the Bayesian embedding approach to circular regression as introduced by Nuñez-Antonio et al. [3] has a reasonable performance in sample sizes from about 50. In analyzing data with very high concentrations of the circular outcome and small sample sizes, using the approach may result in estimates that are biased. Regarding efficiency and flexibility this approach performs rather well. Estimation time and convergence are fast and with regard to types and amount of predictors that can be included the method is as flexible as possible. Lasty, circular effects can be computed from the linear regression coefficients adding extra flexibility to the method. Researchers should however take into account that the interpretation of these circular effects is different from and less straightforward as the interpretation of linear effects.

References

[1] Fisher, N. I. (1995). *Statistical analysis of circular data.* Cambridge University Press, Cambridge.

[2] Mardia, K. V., Jupp, P. E. (1999). *Directional statistics.* Wiley, New York.

[3] Nuñez-Antonio, G., Gutiérrez-Peña, E., Escarela, G. (2011). "A bayesian regression model for circular data based on the projected normal distribution." *Statistical Modelling*, **11**(3), 185–201.

[4] Ravindran, P., Ghosh, S.K. (2011). "Bayesian methods for circular regression using wrapped distributions." *Statistical Theory and Practice*, **5**(4), 547–561.