Extending Bayesian Analysis of Circular Data to comparison of Multiple Groups

Kees Mulder[1], Irene Klugkist[1]

[1] Utrecht University, Department of Methodology and Statistics, Utrecht, The Netherlands
`k.t.mulder@uu.nl`

`i.klugkist@uu.nl`

Abstract

Circular data are data measured in angles and occur in a variety of scientific disciplines. Bayesian methods promise to allow for flexible analysis of circular data, for which few methods are available. Three existing MCMC methods (Gibbs, Metropolis-Hastings, and Rejection) for a single group of circular data were extended to be used in a between-subjects design, providing a novel procedure to compare groups of circular data. Investigating the performance of the methods by simulation study, all methods were found to overestimate the concentration parameter of the posterior, while coverage was reasonable. The rejection sampler performed best. In future research, the MCMC method may be extended to include covariates, or a within-subjects design.

Keywords: MCMC methods; Gibbs; Metropolis-Hastings; rejection sampler.

1   Introduction

Circular data are data measured in angles or orientations in two-dimensional space. For example, one may imagine directions on a compass ($0° - 360°$), times of the day ($0 - 24$ hours), or directions on a circumplex model, such as Leary's Circle. Circular data are frequently encountered in many scientific disciplines, such as biology, social sciences, meteorology, astronomy, earth sciences, and medicine. Circular data can be modeled by the von Mises distribution, which is the natural analogue of the normal distribution on the circle, and given by

$$f(\theta|\mu,\kappa) = \{2\pi I_0(\kappa)\}^{-1} \exp\{\kappa \cos(\theta - \mu)\}, \qquad 0 \le \theta < 2\pi, \kappa \ge 0,$$

where θ represents the data, μ is the mean direction, κ is a concentration parameter, and $I_0(\cdot)$ is the modified Bessel function of order 0.

Due to the difficulty of working with a circular sample space, few methods have been developed in the field of analysis of circular data. Bayesian methods offer a promising new approach not only in the field of statistics at large, but also specifically in the analysis of circular data, because the flexibility of Markov chain Monte Carlo (MCMC) methods
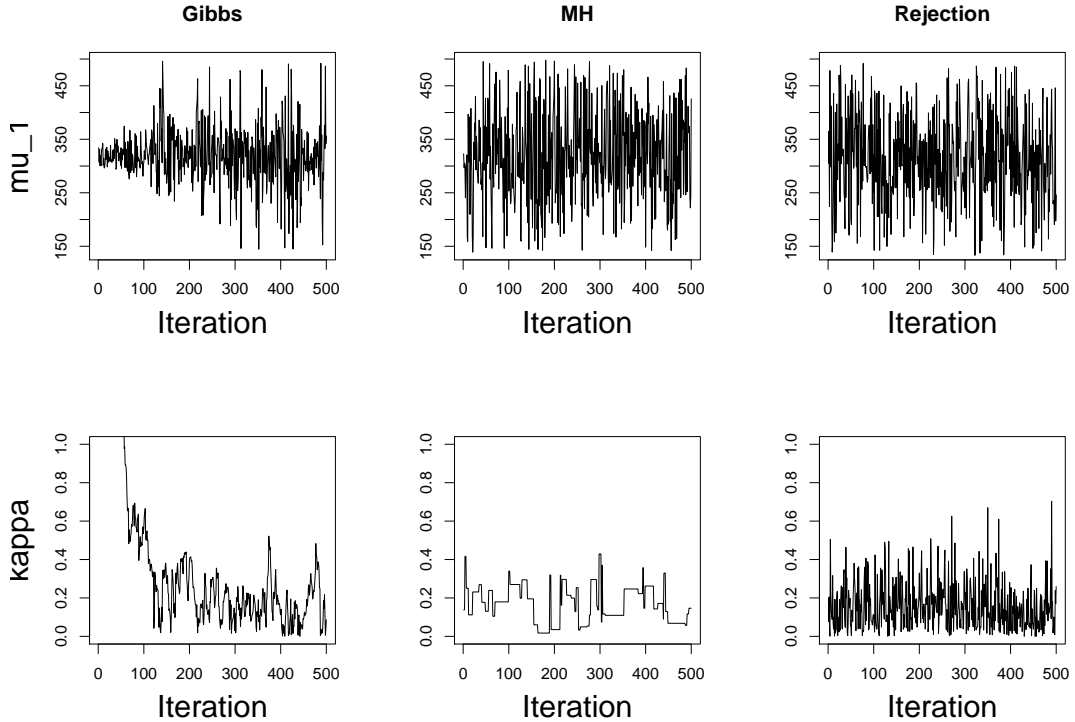
1

Figure 1: Example chains of μ (in degrees) of group 1 and common κ drawn in the first 500 iterations of each of the three methods, with no burn-in and without thinning the chain, where $J = 3$, true $\kappa = 0.1$, and $n_j = 30$.

may provide a relatively straightforward way to analyze a large variety of circular data models.

A limited number of MCMC methods for circular data have been developed. Here, a Gibbs sampler, a Metropolis-Hastings (MH) algorithm [3] and a recent rejection sampler will be examined. Many tests in between-subjects designs, such as ANOVA, assume equal variance across groups, but the current MCMC methods do allow for this assumption. The main contribution of this study was the extension of each method allowing them to sample for a model with J independent groups with a single, common concentration parameter κ.

2 Available Methods and their Performance

Figure 1 shows example chains for the three methods.

2.1 Gibbs sampler

Early work by [1] provided a Gibbs sampler that works by adding latent variables to the model. Although the relative simplicity of the Gibbs sampler usually is appealing, several problems with this approach were noted in this study.

First, it requires 'tuning' because in one step in the procedure, samples need to be drawn up to some number Z, where the most appropriate value for Z varies between datasets. Second, it is not feasible to apply this sampler to datasets with high concentrations, due to extreme autocorrelation. Third, this method showed an upward bias for

the concentration parameter when n was small. Fourth, undesirable coverages for both μ and κ were found in application of this method. Fifth, it is computationally intensive.

2.2 Metropolis-Hastings sampler

An application of the MH algorithm to multiple groups of circular data was created for this study, which may be advantageous due to the problems with the Gibbs sampler.

It performed adequately, although this method showed the same upward bias for the concentration parameter. Coverages were acceptable, however. Acceptance rates were reasonable, but they can become too low for large datasets with little concentration.

2.3 Rejection sampler

A fast rejection sampler was developed earlier this year to sample from the posterior of a von Mises distribution [2]. However, it uses a constant prior and it is applied to a single group of data. Here, it was extended to use a conjugate prior with multiple groups. It performed very well, but it does show the same upward bias in the concentration parameter found in the other methods.

3   Discussion

Using the developed methods, a Bayesian analysis comparing multiple groups of circular data may now be performed. The model developed here is still limited in terms of scope; it provides a basic between-subjects design for multiple groups of circular data, but extensions of this model such as a between-within-design or the inclusion of covariates have yet to be developed.

References

[1] Damien, P., Walker, S. (1995). *Statistical analysis of circular data full Bayesian analysis of circular data using the von Mises distribution*. Canadian Journal of Statistics 27(2): 291-298.

[2] Forbes, P.G.M., Mardia, K. (2014). *A Fast Algorithm for Sampling from the Posterior of a von Mises distribution*. arXiv preprint arXiv:1402.3569.

[3] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A.H. and Teller, E. (1953). "Equation of state calculations by fast computing machines." *The journal of chemical physics*, 21(6): 1087-1092.