Extending Bayesian analysis of circular data to comparison of multiple groups

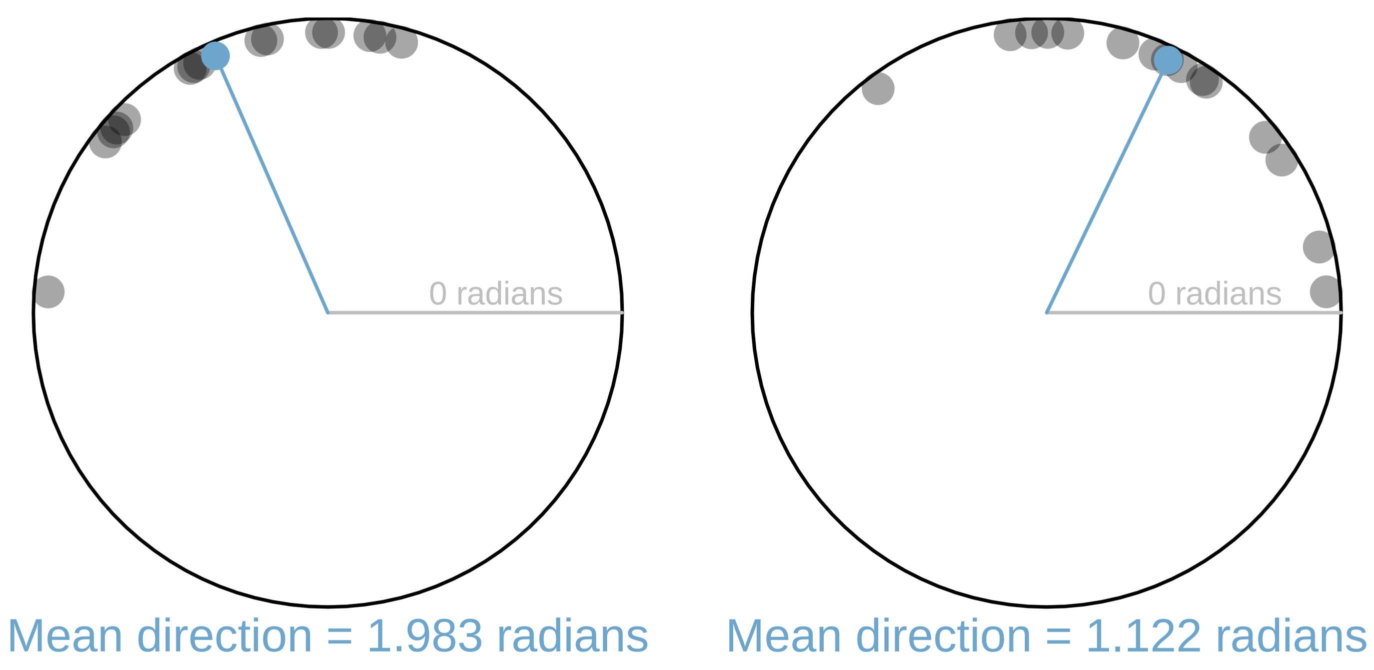
Kees Mulder
Utrecht University

What is circular data?

Circular data have a **periodical** sample space.
They may be **angles**, **directions** or **orientations**, measured in **radians**, **degrees**, or **unit vectors**.

Examples are found in a large variety of disciplines:

- The direction of animal movement.
- The orientation of fractures in rocks.
- The time of day, or day of the year.
- Measurements on a circumplex model, such as Leary's rose for interpersonal behaviour.
- Parts of the structure of proteins and DNA.



Specialized methods for circular data are necessary. Here, we focus on models with a **circular outcome**.

The problem

- Circular data are **inherently difficult to analyze**, due to the circular sample space.
- **Few methods** have been developed for circular outcomes.
- Bayesian methods, especially MCMC, offer a promising new route for circular data models.
- Focusing on the **circular ANOVA** context: There is currently no Bayesian model for testing **group mean differences**, assuming **equal variance**.
- Equal variance here means that the **concentration parameter** κ should be **equal across groups**.

Circular data approaches

- The **intrinsic** approach, directly defined on the circle, for example using the von Mises distribution
$$\mathcal{VM}(\theta|\mu,\kappa) \propto \exp\left[\kappa \cos(\theta-\mu)\right].$$
- The **wrapping** approach, where distributions in $\mathbb{R}^1$ are wrapped around the circle.
- The **embedding** approach, embedding points in $\mathbb{R}^2$ to the circle.

Here, we employ the **intrinsic approach**.

Conditionals

A conjugate prior for von Mises is available, which uses $\boldsymbol{\mu_0}, R_0, c$. For an uninformative prior, we set $R_0 = 0, c = 0$. Then, given data $\boldsymbol{\theta}$:

$$C_n = R_0 \cos\mu_0 + \sum_{i=1}^{n} \cos\theta_i, \qquad S_n = R_0 \sin\mu_0 + \sum_{i=1}^{n} \sin\theta_i,$$
$$\mu_n = \text{atan2}(S_n/C_n,)$$
$$R_n = \sqrt{C_n^2 + S_n^2}.$$

The conditionals are then:

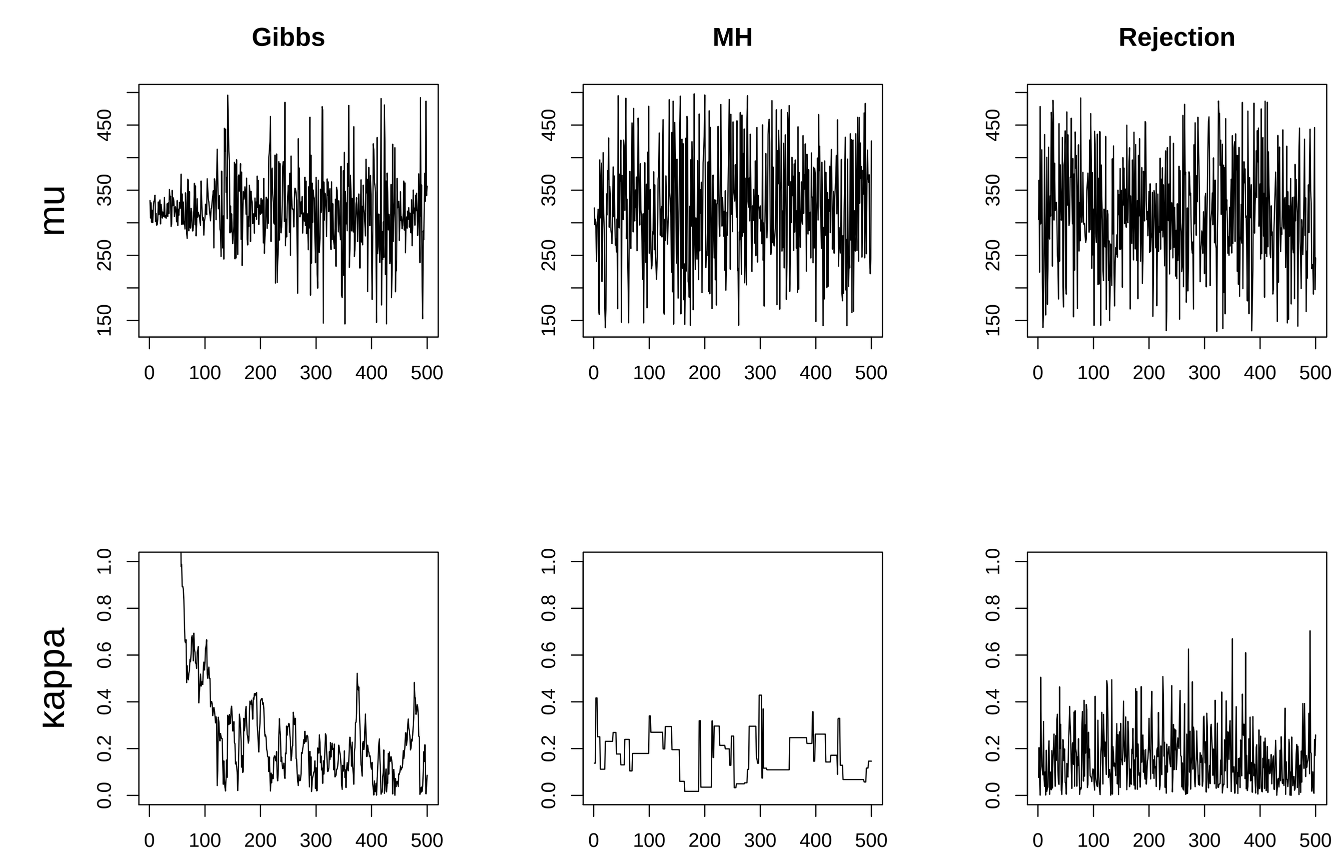
$$f(\mu_j|\kappa,\boldsymbol{\theta}) = \mathcal{VM}(\mu_{nj}, R_{nj}\kappa) \propto \exp\{R_{nj}\kappa\cos(\mu-\mu_{nj})\},$$
$$f(\kappa|\mu,\boldsymbol{\theta}) \propto \{I_0(\kappa)\}^{-(n+c)} \exp\left[R_n\kappa \sum_{j=1}^{J} \cos(\mu-\mu_{nj})\right].$$

Intrinsic methods

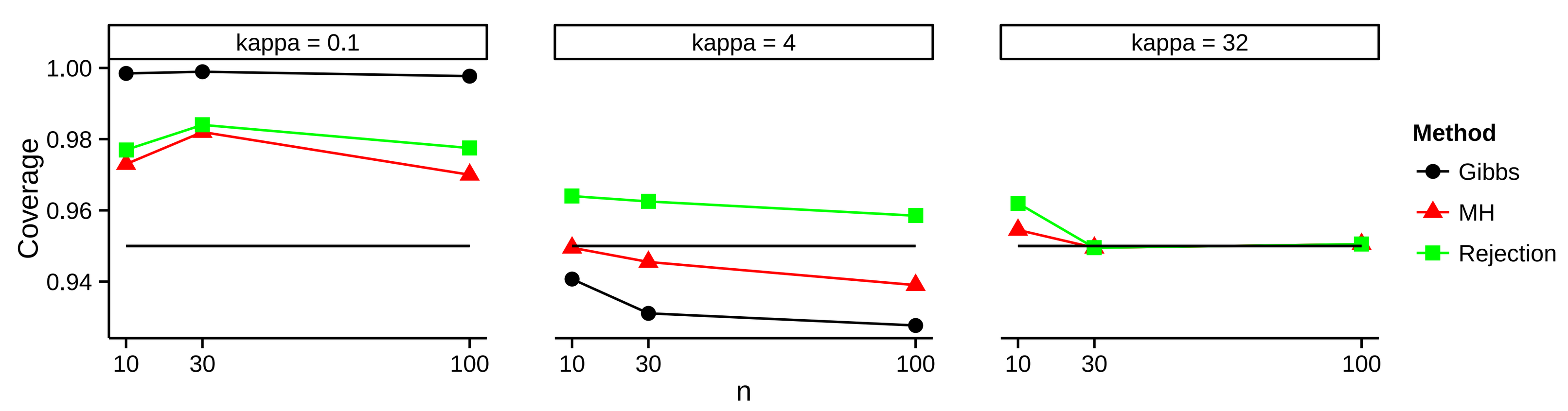
Three solutions were developed:

- A **Gibbs sampler** using auxiliary variables, extending Damien & Walker (1999).
- A **Metropolis-Hastings** method.
- A recent **rejection sampling** method, extending Forbes & Mardia (2014).

Comparison of the three methods



The Gibbs sampler encounters strong autocorrelation when roughly $\kappa > 7$.
Coverage for κ :



The MH and rejection samplers performed well, with the rejection sampler being slightly more efficient.

Contact

This poster was supported by an ISBA travel grant.
Kees Mulder & I. Klugkist
Utrecht University
k.t.mulder@uu.nl
i.klugkist@uu.nl