Bayesian Estimation and Hypothesis Tests for a Circular GLM

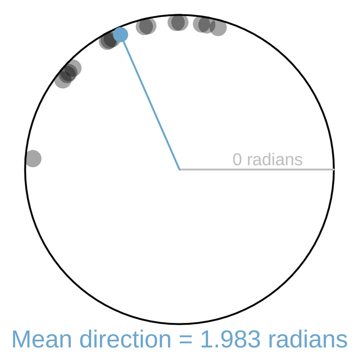
Kees Mulder & Irene Klugkist
Utrecht University

Introduction to circular data

Circular data is data that is measured in angles or directions, as degrees or radians. Circular data differ from linear data in the sense that circular data are measured in a periodical sample space. For example, an angle of $1°$ is quite close to an angle $359°$, although linear intuition suggests otherwise.
Circular data θ_i $(i = 1, \ldots, n)$ can be modeled with the von Mises distribution

$$\mathcal{M}(\theta \mid \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp\{\kappa \cos(\theta - \mu)\},$$

with mean direction μ and concentration κ, where $I_0(\cdot)$ represents the modified Bessel function of the first kind and order zero.
Our goal is to analyze von Mises based models in a Bayesian way, to include linear and dichotomous covariates, and to develop hypothesis tests.



To predict the circular outcome by linear covariates, we consider the classic model

$$\mathcal{M}(\theta \mid \mu_i, \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp\{\kappa \cos(\theta - \mu_i)\}, \quad (1)$$

$$\mu_i = \beta_0 + g(\boldsymbol{\beta}^T \boldsymbol{x}_i), \quad (2)$$

where

- β_0 is a circular intercept
- $\boldsymbol{\beta}$ is a vector of regression coefficients.
- $g(\cdot)$ is a link function, we choose the commonly used $g(x) = 2\tan^{-1} x$
- $\boldsymbol{x}$ are standardized linear covariates

We consider a Bayesian analysis of this model, and propose three extensions.

Including group differences

Including group differences as dichotomous predictors in $\boldsymbol{x}$ causes the analysis to depend on the chosen reference group.
For a dichtomous predictor d, let the coefficient be δ, the model is

$$\mu_i = \beta_0 + g(\boldsymbol{\beta}^T \boldsymbol{x}_i + d\delta),$$

which means the prediction line is 'shifted', because it is centered around $g(0)$ for $d = 0$ and $g(\delta)$ for $d = 1$.

As shown below, this means the entire analysis depends on the reference category, which is undesirable.

Instead, we propose to treat the group difference separately, so that we have

$$\mu_i = \beta_0 + \boldsymbol{\delta}^T \boldsymbol{d} + g(\boldsymbol{\beta}^T \boldsymbol{x}_i +),$$

which solves the problem.

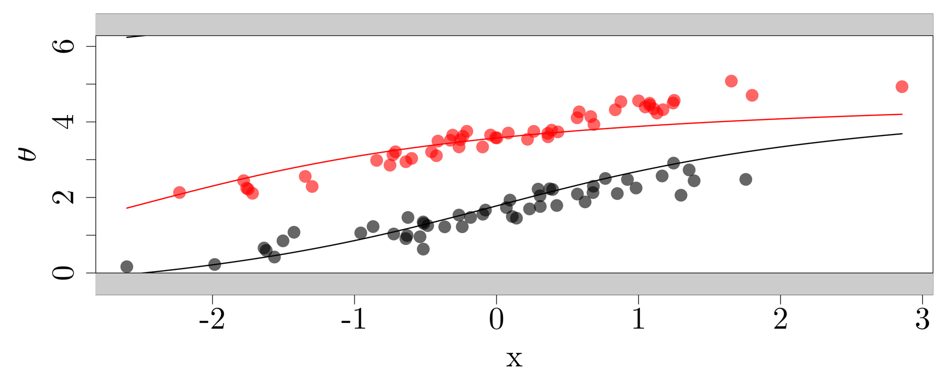


Fig. 1: Original labels

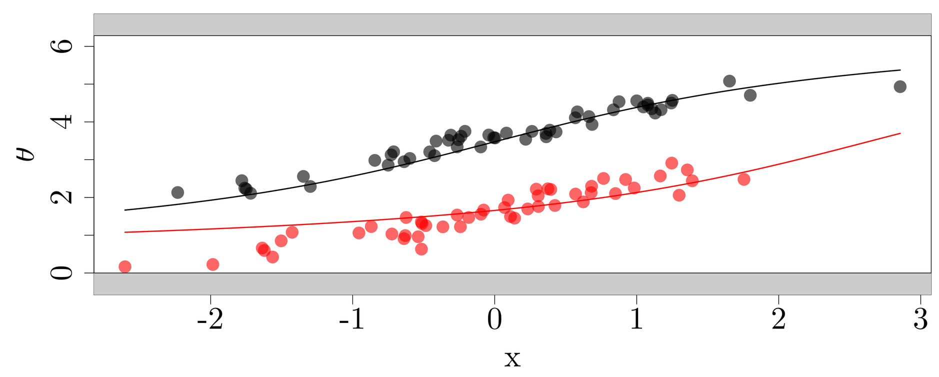


Fig. 2: Reversed labels

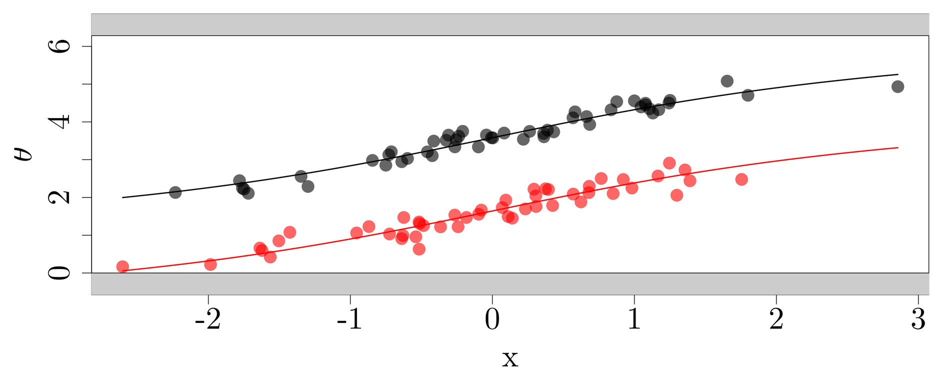


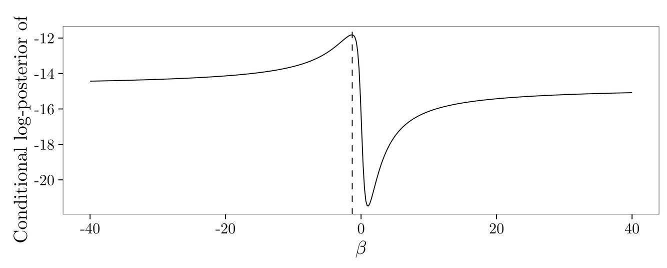
Fig. 3: $\mu_i = \beta_0 + \delta d_i + g(\beta x_i)$

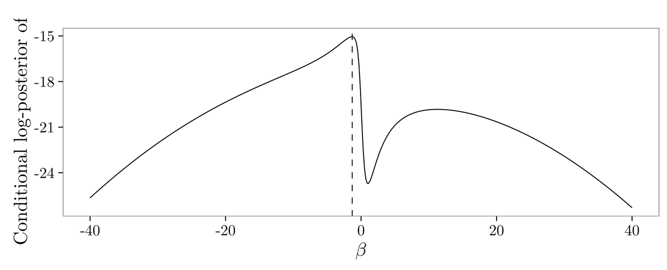
Priors

Straightforward priors are available for most of the model:

- The von Mises part of the model has a conjugate prior
  $$p(\beta_0, \kappa \mid \boldsymbol{\delta}, \boldsymbol{\beta}) \propto I_0(\kappa)^{-c} \exp[R_0 \kappa \cos(\beta_0 - \mu_0)],$$
  which is uninformative if we take $c = 0, R_0 = 0$.
- Because of the bounded nature of the circle, β_0 and δ have natural priors in the form of the circular uniform distribution.

However, the shape of the likelihood for β is not concave, and features non-zero asymptotes, as shown at the top here. This has been discussed in the literature, and is dealt with in frequentist optimization by trying many starting values and monitoring convergence.
The Bayesian paradigm provides a great solution here, by either a weakly informative prior or a subjective prior.





Our solution
If $\sum_{k=1}^{K} |\beta_k| > 1.5$, the majority of the probability mass of the data is on the semi-circle opposite of the group intercept $(\beta_0 + \boldsymbol{\delta}^T \boldsymbol{d})$, which is not likely in practice. This expectation can be translated to a weakly informative prior distribution.
So, the prior is chosen to be

$$\beta_k \sim N(0, 1) \; \forall k = 1, \ldots, K,$$

where $N(\mu, \sigma^2)$ denotes the Normal distribution with mean μ and variance σ^2. It can be seen that the problematic asymptotes are solved in the picture on the bottom. In neither case the posterior is log-concave, which might make optimization difficult, but which MCMC methods handle well.

Bayesian hypothesis tests using the Bayes factor

Equality constrained hypotheses
Consider two hypotheses about some model parameter γ,

$$H_0 : \gamma = \gamma_0, \quad H_1 : \gamma \in \Omega_\gamma, \quad (3)$$

where Ω_γ is the sample space of γ. The Bayes factor for this hypothesis is given by

$$BF_{01} = \frac{p(D \mid H_0)}{p(D \mid H_1)}. \quad (4)$$

For the Savage-Dickey density ration, we use the fact that under some conditions,

$$\frac{p(D \mid H_0)}{p(D \mid H_1)} = \frac{p(\gamma = \gamma_0 \mid D, H_1)}{p(\gamma = \gamma_0 \mid H_1)},$$

which is a ratio of the posterior and prior probability of γ_0 under model H_1.
One remark to be made is that this method is only valid if the nuisance parameters between the two hypotheses serve the same purpose.

Inequality constrained hypotheses
Researchers often have directed (one-sided) hypotheses, which may be specified by using inequality constraints.
For some model parameter γ, a simple hypothesis to evaluate could be

$$H_0 : \gamma > \gamma_0, \quad H_1 : \gamma < \gamma_0.$$

In order to quantify our belief in these hypotheses, we employ an encompassing hypothesis $H_{unc} : \gamma \in \Omega_\gamma$, from which an MCMC sample $\boldsymbol{\gamma} = \{\gamma^{(1)}, \ldots, \gamma^{(Q)}\}$ is obtained. Then, assuming the encompassing prior does not favor either hypothesis, it can be shown that the Bayes factor for H_0 versus H_1 is given by

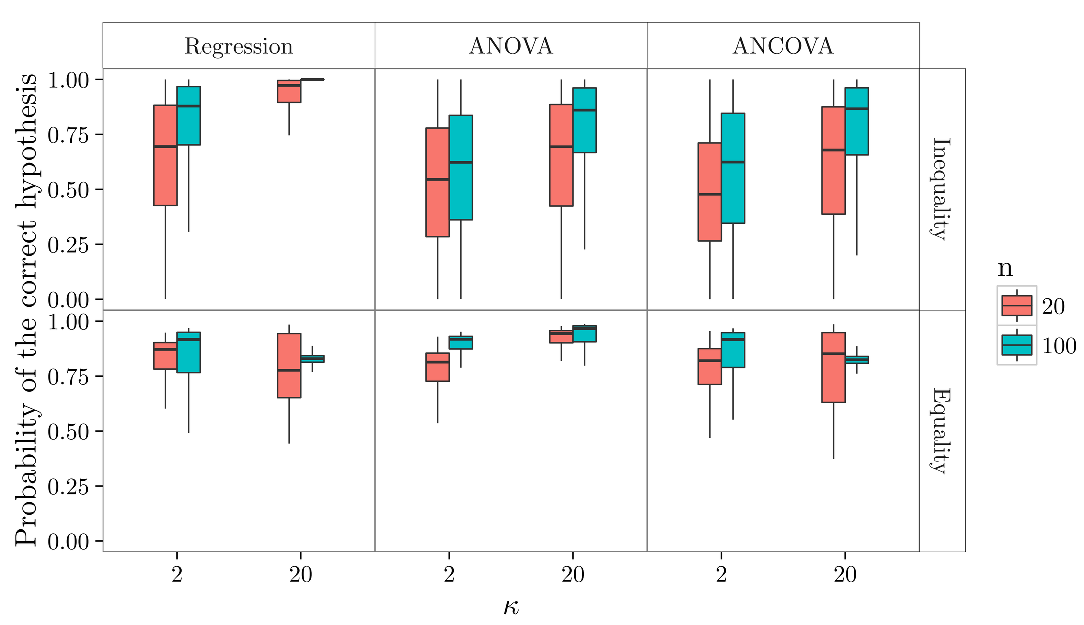
$$BF_{01} = \frac{\sum_{s=1}^{Q} I\left(\gamma^{(s)} \in \Omega_{\gamma|H_0}\right)}{\sum_{s=1}^{Q} I\left(\gamma^{(s)} \in \Omega_{\gamma|H_1}\right)}, \quad (5)$$

where $I(\cdot)$ is an indicator function, and $\Omega_{\gamma|H_s}$ is the admitted sample space for γ under hypothesis H_s. Note that for more complex models, we need to take the 'complexity' into account, which denotes the proportion of the prior in agreement with a hypothesis.
Using the inequality constrained hypothesis approach, it is easy to assess the model

$$\mu_1 > \mu_2 > \mu_3$$

versus its complement, which contains all other orderings.

Bayes factors from a simulation study for the regression model (one linear predictor), ANOVA model (two dichotomous grouping variables) and ANCOVA model (one dichotomous grouping variable, four linear predictors) are shown below.



From 5000 simulation, with true coefficients .05, the Bayes factors perform adequately.

Discussion

- The Bayesian approach provides a promising way to draw inference from circular data. Usual approaches are based on large sample or high concentration approximations or bootstrap approaches for simple models. Our approach does not need such approximations, and provides a new direction for circular data analysis of GLM-type models.
- In extensive simulations (not shown on this poster) performance was shown to be good.
- The model provides one approach to modeling circular data, but extensions to more flexible models, including additional parameters or a hierarchical structure are within reach.

Contact
Methodology & Statistics, Utrecht University
Kees Mulder
k.mulder@uu.nl
-
Irene Klugkist
i.klugkist@uu.nl