

## 0.1 Introduction

Circular data are measured in angles or directions, and are frequently encountered in scientific fields as diverse as life sciences (?), behavioural biology (?), cognitive psychology (?), bioinformatics (?), political science (?) and environmental sciences (???). In psychology, circular data occur often in motor behaviour research (????), as well as in the application of circumplex models (???). Circular data differ from linear data in the sense that circular data are measured in a periodical sample space. For example, an angle of  $1^\circ$  is quite close to an angle  $359^\circ$ , although linear intuition suggests otherwise.

Therefore, linear models may not properly describe the process that has generated the circular data of interest. Circular data analysis has been developed to deal with this, although attention to this type of analysis has been limited. Only slightly more than a handful of in-depth books on circular data analysis have been published (????), and in general, statistical methods for circular data are somewhat limited.

Here, attention is turned to analysis of datasets with a circular outcome, predicted by covariates that can be continuous (linear) or categorical. This leads to a structure similar to the Generalized Linear Model (GLM), which has both multiple regression and ANCOVA as special cases.

Three main approaches to circular data analysis might be distinguished. First, the intrinsic approach employs distributions directly defined on the circle (?). Second, the wrapping approach ‘wraps’ a univariate distribution around the circle by taking the modulus of data on the real line (?). Third, the embedding approach projects points from a bivariate distribution to the circle (????). While the wrapping and embedding approach provide promising avenues of study in their own right, here attention is restricted to the intrinsic approach, as it might provide the most natural analysis of circular data.

Within the intrinsic approach, the circular analogue to the Normal distribution is the von Mises distribution (?). This symmetric unimodal distribution is given by

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

where  $\theta \in (-\pi, \pi)$  represents an angular measurement,  $\mu \in (-\pi, \pi)$  represents the mean direction,  $\kappa \in \mathbb{R}^+$  is a concentration parameter, and  $I_0(\cdot)$  represents the modified Bessel function of the first kind and order zero. Some examples of frequentist methods that employ the von Mises distribution are a circular ANOVA (?), circular ANCOVA (?) and circular regression (?). Here, a Bayesian analysis of such models will be developed.

Early approaches to Markov chain Monte Carlo (MCMC) sampling for the von Mises distribution provide a method for sampling  $\mu$  when  $\kappa$  is known (?) and sampling both parameters for a single group of data (?). ? present a conjugate prior for the von Mises model. Recent theoretical work has much improved the efficiency of the sampling of the concentration parameter of the von Mises distribution (?).

Some development has also taken place in the field of semiparametric inference for circular data models, often using Dirichlet process priors (????). In particular, ? provide Bayes factors for the simple hypothesis test of equality of two means. However, these methods are generally complex, which makes it hard to extend these models, for example to include covariates. Therefore, we will focus on parametric models, with residuals following the von Mises distribution.

A Bayesian circular regression analysis has been developed by ?, using starting values from a frequentist iterative reweighted least squares (IRLS) algorithm, which is similar to that used by ?. ? note that the likelihood function of the regression coefficients from their model is not globally logarithmically concave, which might cause the algorithm to converge to a local maximum. To combat this, ? advise careful inspection of the likelihood surface of the regression coefficients. Drawbacks of the approach taken by ? are that a prior is not specified, the algorithm is slow, categorical predictors are not treated separately and for larger models it may be unclear whether the regression coefficients have converged to the global maximum.

Recent work has provided a multivariate extension of the von Mises distribution (?), which offers a promising new way of thinking about circular covariate models. The multivariate von Mises was applied in this context by ? within a Generalized Linear Model (GLM) setting, applying MCMC likelihood approximation as in ? to compute maximum likelihood estimates. This approach is not Bayesian, but it is a promising approach because of its flexibility, allowing both the mean and concentration to be dependent on an arbitrary set of covariates, as well as allowing observations to be dependent.

$$(a) \mu_i = \beta_0 + g(\beta x_i + \delta d_i)$$

$$(b) \mu_i = \beta_0 + \delta d_i + g(\beta x_i)$$

Figure 1: Prediction lines from two different models, which were fitted to a dataset with  $n = 100$ , and true parameters  $\delta = 2, \beta_0 = \pi/2, \beta = 0.4, \kappa = 20$ . The two models have (a) dichotomous predictors placed in the link function, and (b) dichotomous predictors treated separately.

There are three main drawbacks of the circular GLM approach to circular data analysis currently. First, the GLM approach is not free from the lack of concavity as described in ?, although this has not yet been investigated in detail. Second, the current approach does not have separate parameters for differences in group mean direction, which precludes the popular ANCOVA model to some extent. Third, Bayesian hypothesis tests for this model are not available, which limits its applicability.

The structure of this paper is as follows. The circular data GLM model is developed in a fully Bayesian setting in Section 0.2. The lack of concavity in the likelihood function will be examined, and suggestions will be formulated on how to deal with this issue. Details on the MCMC sampler are provided in Section 0.3. Section 0.4 outlines Bayesian hypothesis tests for this model, both for equality and inequality constrained hypotheses. Then, a simulation study for the method is provided in Section 0.5. Section 0.6 provides an application of our method to empirical data from cognitive psychology. Finally, Section 0.7 provides a short discussion.

## 0.2 Bayesian circular GLM

Consider a dataset  $\{\theta_i, \mathbf{x}_i, \mathbf{d}_i\}, (i = 1, \dots, n)$ , where  $\theta_i$  is a circular outcome variable,  $\mathbf{x}_i \in \mathbb{R}^K$  is a column vector of continuous linear covariates which are assumed to be standardized, and  $\mathbf{d}_i \in \{0, 1\}^J$  is a column vector of dichotomous variables indicating group membership. Assume that each observed angle  $\theta_i$  is generated independently from a von Mises distribution  $\mathcal{M}(\theta_i | \mu_i, \kappa)$ . Then,  $\mu_i$  is chosen to be

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

where  $\beta_0 \in [-\pi, \pi)$  is an offset parameter which serves as a circular intercept,  $\boldsymbol{\delta} \in [-\pi, \pi)^J$  is a column vector of circular group difference parameters,  $g(\cdot) : \mathbb{R} \rightarrow (-\pi, \pi)$  is a twice differentiable link function, and  $\boldsymbol{\beta} \in \mathbb{R}^K$  is a column vector of regression coefficients. ? and ? discuss the choice of the link function. A common and natural choice for the link function is  $g(x) = 2 \tan^{-1}(x)$ , which we will focus on here.

This model specification differs from the usual approach to circular regression models, as these generally set  $\mu_i = \beta_0 + g(\boldsymbol{\beta}^T \mathbf{x}_i)$  (???). However, we view this model as unsatisfactory when including dichotomous predictors in  $\mathbf{x}$ , which we will illustrate in Figure 1. Consider a single dichotomous predictor  $d$  added to a model with a single continuous predictor  $x$ . The dichotomous predictor might be added into the model as  $\mu = \beta_0 + g(\beta x + \delta d)$ . Adding  $\delta$  in the link function shifts the location of the prediction line, but also its shape. Therefore, the shape for  $d = 0$  is fixed, but for  $d = 1$  the shape is dependent on a free parameter,  $\delta$ . This makes the shape of the prediction line (and therefore the analysis) depend on the arbitrary choice of reference group, which can be seen in Figure 1a. To solve this, we advocate setting  $\mu = \beta_0 + \delta d + g(\beta x)$ , the resulting prediction lines of which are shown in Figure 1b.

A comparable approach is taken in ?, where a separate intercept is estimated for each group. However, having a separate intercept for each group means that a factorial design with main effects only can not be specified. In many applications, especially in psychology, this is problematic. The approach here is more flexible in that it allows a researcher to either fit a model with main effects only, to fit a model with specific interactions, or to compare these models. In addition, ? also describes a non-parallel case where the regression parameters are estimated separately for each group. This model can be obtained as a special case of the model provided here by including appropriate interaction terms in the model.

### 0.2.1 Likelihood

Denote the set of parameters by  $\phi = \{\beta_0, \kappa, \boldsymbol{\delta}, \boldsymbol{\beta}\}$ . The joint likelihood for the GLM-type model is then given by

$$f(\boldsymbol{\theta}, \mathbf{X}, \mathbf{d} \mid \phi) = \prod_{i=1}^n \mathcal{M}(\theta_i \mid \mu_i, \kappa) \quad (3)$$

$$= \{2\pi I_0(\kappa)\}^{-n} \exp \left\{ \kappa \sum_{i=1}^n \cos \left[ \theta_i - \left( \beta_0 + \boldsymbol{\delta}^T \mathbf{d}_i + g(\boldsymbol{\beta}^T \mathbf{x}_i) \right) \right] \right\}. \quad (4)$$

If we let  $\psi_i = \theta_i - \boldsymbol{\delta}^T \mathbf{d}_i - g(\boldsymbol{\beta}^T \mathbf{x}_i)$ , ( $i = 1, \dots, n$ ), then we can recognize the conditional likelihood  $f(\beta_0, \kappa \mid \boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{X}, \mathbf{d})$  as the likelihood of the parameters of a von Mises distribution with mean direction  $\beta_0$  and concentration  $\kappa$ , as shown in Appendix .1.

The conditional distribution of  $\beta_0$  is  $\mathcal{M}(\bar{\psi}, R_\psi \kappa)$ , where  $\bar{\psi}$  and  $R_\psi$  are the mean direction and resultant length of the vector  $\boldsymbol{\psi}$ , given by

$$\bar{\psi} = \left( \sum_{i=1}^n \sin \psi_i, \sum_{i=1}^n \cos \psi_i \right), \quad R_\psi = \sqrt{\left( \sum_{i=1}^n \cos \psi_i \right)^2 + \left( \sum_{i=1}^n \sin \psi_i \right)^2}.$$

Conditionals for  $\kappa$ ,  $\boldsymbol{\beta}$  and  $\boldsymbol{\delta}$  are not of simple form and require special attention.

### 0.2.2 Priors

The next step in the model specification is setting prior distributions for the parameters  $\phi$ . We will focus on uninformative, default priors where possible. The joint prior is factored as

$$p(\phi) \propto p(\beta_0, \kappa \mid \boldsymbol{\delta}, \boldsymbol{\beta}) p(\boldsymbol{\delta}) p(\boldsymbol{\beta}) \quad (5)$$

so that  $\boldsymbol{\beta}$  and  $\boldsymbol{\delta}$  are independent. Furthermore,

$$p(\boldsymbol{\delta}) \propto \prod_{j=1}^J p(\delta_j), \quad p(\boldsymbol{\beta}) \propto \prod_{k=1}^K p(\beta_k). \quad (6)$$

Next, the choice of each of these priors is discussed.

$p(\delta_j)$

For each  $\delta_j$ , the circular uniform distribution is a natural and uninformative default prior, so that

$$p(\delta_j) = \frac{1}{2\pi}, \quad \forall j = 1, \dots, J. \quad (7)$$

This prior indicates that given a mean direction for some reference group, there is no knowledge on the mean direction of group  $j$ .

$p(\boldsymbol{\beta})$

For each  $\beta_k$  there is no natural uninformative prior. A constant prior  $p(\beta_k) \propto 1$  could be employed. However, as noted by ? and ?, this leads to a posterior of irregular form, including local maxima and non-zero asymptotes, as shown in Figure 2a.

However, because the linear predictors are standardized, the interpretation of the size of  $\boldsymbol{\beta}$  is equal across studies. Therefore, we can determine a priori which values of  $\beta_k$  would be probable in practical research

$$\begin{aligned}
\text{(a) } p(\beta_k) &\propto 1 & \text{(b) } p(\beta_k) &\propto N(0, 1/5) \text{ (dashed)} \\
& & p(\beta_k) &\propto N(0, 1) \text{ (solid)} \\
& & p(\beta_k) &\propto N(0, 5) \text{ (dotted)}
\end{aligned}$$

Figure 2: Comparison of the conditional log-posterior of  $\beta_k$  when using (a) a constant prior, which means the log-posterior is equal to the log-likelihood, and (b) a Normal prior with three different values for  $\sigma^2$ .

scenarios. In this case, 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  $(\beta_0 + \delta^T \mathbf{d})$ , which is not likely in practice. This expectation can be translated to a weakly informative prior distribution. Here, this was done by setting the prior as

$$\beta_k \sim N(0, \sigma^2), \forall k = 1, \dots, K, \quad (8)$$

where  $N(\mu, \sigma^2)$  denotes the Normal distribution with mean  $\mu$  and variance  $\sigma^2$ . For a Normal prior with any finite  $\sigma^2$ , there are no non-zero asymptotes in the conditional posterior of  $\beta$  for any values of  $\beta_0, \kappa, \delta$ , because  $\log f_N(x | 0, \sigma^2) \rightarrow -\infty$  as  $|x| \rightarrow \infty$ . As  $\sigma^2 \rightarrow 0$ , the prior becomes more informative and the posterior for  $\beta_k$  centers on 0. As  $\sigma^2 \rightarrow \infty$ , the prior becomes less informative, but the posterior becomes more irregular, with large plateaus and more local maxima. By default, we choose  $\sigma^2 = 1$  so that the prior is the standard Normal distribution, which represents the weakly informative prior mentioned previously, for which values of  $|\beta_k| > 1.5$  are a priori unlikely.

To illustrate this, the resulting posterior is compared to the posterior resulting from the constant prior in Figure 2. The posterior is based on a synthetic data set of 7 observations with a single predictor  $\mathbf{x} = -3, \dots, 3$ , which is standardized and the outcome is then computed as  $\theta_i = 2 \tan^{-1}(x_i) + \varepsilon_i$  where  $\varepsilon_i \sim N(0, 1/10)$ , so the true  $\beta$  is 1. In Figure 2, the conditional posterior of  $\beta$  is displayed given  $\beta_0 = 0, \kappa = 1$ . Figure 2a gives a zoomed-out view of the posterior resulting from the constant prior, where the asymptotes are clearly visible. Figure 2b shows a zoomed-in view of the resulting conditional posterior from a prior with  $\sigma^2 = 1/5$  (dashed),  $\sigma^2 = 1$  (solid),  $\sigma^2 = 5$  (dotted). It can be seen that the  $N(0, 5)$  prior takes a shape not unlike the one seen in Figure 2a, although the asymptote is avoided. The  $N(0, 1/5)$  prior can be seen to have a strong influence on the posterior estimate for this data where the true  $\beta = 1$ . The  $N(0, 1)$  prior represents a balance for which the asymptotes are solved, but the posterior estimates are very close to the maximum likelihood estimates. In practical settings generally  $|\beta| \ll 1$ , so that the influence of the prior will often be minimal. For these priors the posterior is not necessarily logarithmically concave, which might make optimization difficult, but which MCMC methods handle well.

$$p(\beta_0, \kappa)$$

For the von Mises part of the model we follow the conjugate prior provided by ?, given by

$$p(\beta_0, \kappa | \delta, \beta) \propto I_0(\kappa)^{-c} \exp[R_0 \kappa \cos(\beta_0 - \mu_0)]. \quad (9)$$

The prior hyperparameters  $\{c, R_0, \mu_0\}$  can be interpreted as the prior sample size  $c$ , prior resultant length  $R_0$  and prior mean direction  $\mu_0$  respectively of a hypothetical set of angles  $\psi = \theta - \delta^T \mathbf{d} - g(\beta^T \mathbf{x})$ . Setting informative prior expectations for the parameters of the distribution of a random angle  $\psi$  might be difficult, because conditioning on  $\beta$  and  $\delta$  makes  $\psi$  hard to interpret. However, an uninformative prior is easily obtained by setting  $c = 0, R_0 = 0$ , which is the approach taken here. Note that this does induce an improper (constant) prior on  $\kappa$ .

### 0.3 MCMC sampling

In this section, details will be discussed for the MCMC sampling procedure, given below as Algorithm 1. Usually, the algorithm converges fast and mixes rapidly, at least for smaller models. The following sections provide further details on sampling  $\beta_0, \kappa, \beta$  and  $\delta$ .

---

**Algorithm 1** MCMC algorithm for circular GLM

---

Set  $\phi^{(1)} \leftarrow \{\beta_0^{(1)}, \kappa^{(1)}, \beta^{(1)}, \delta^{(1)}\}$ , which are the given starting values.

**for**  $q = 2, \dots, Q$  **do**

$\psi_i \leftarrow \theta_i - \delta^T \mathbf{d}_i - g(\beta^T \mathbf{x}_i), \forall i = 1, \dots, n.$

$R_\psi \leftarrow \sqrt{(\sum_{i=1}^n \cos \psi_i)^2 + (\sum_{i=1}^n \sin \psi_i)^2}.$

$\bar{\psi} \leftarrow [\sum_{i=1}^n \sin \psi_i, \sum_{i=1}^n \cos \psi_i].$

Sample  $\beta_0 \sim \mathcal{M}(\bar{\psi}, R_\psi \kappa).$

$\zeta \leftarrow -n^{-1} R_\psi \cos(\beta_0 - \bar{\psi}).$

Sample  $\kappa$  with `sampleKappa`( $n, \zeta$ ) as in ?.

**for**  $j = 1, \dots, J$  **do**

Sample a candidate  $\delta_j^* \sim \mathcal{M}(\delta_j, R_\psi \kappa).$

$\alpha_{\delta_j} \leftarrow \log p(\delta_j^*, \phi_{(-\delta_j)} \mid \boldsymbol{\theta}, \mathbf{X}, \mathbf{d}) - \log p(\delta_j, \phi_{(-\delta_j)} \mid \boldsymbol{\theta}, \mathbf{X}, \mathbf{d}).$

Sample  $u_1 \sim U[0, 1].$

**if**  $\alpha_{\delta_j} > \log u_1$  **then**

$\delta_j \leftarrow \delta_j^*$

**end if**

**end for**

**for**  $k = 1, \dots, K$  **do**

Sample  $u_2 \sim U[-w, w]$  and  $u_3 \sim U[0, 1].$

$\beta_k^* \leftarrow (\beta_k + \tan(u_2 \pi / 2)) / (1 - \beta_k \tan(u_2 \pi / 2)).$

$\alpha_{\beta_k} \leftarrow \log p(\beta_k^*, \phi_{(-\beta_k)} \mid \boldsymbol{\theta}, \mathbf{X}, \mathbf{d}) + \log f_{tc}(\beta_k^* \mid w) -$

$\log p(\beta_k, \phi_{(-\beta_k)} \mid \boldsymbol{\theta}, \mathbf{X}, \mathbf{d}) - \log f_{tc}(\beta_k \mid w),$

where  $f_{tc}(x \mid w) = 1 / (w\pi [1 + x^2]).$

**if**  $\alpha_{\beta_k} > \log u_3$  **then**

$\beta_k \leftarrow \beta_k^*$

**end if**

**end for**

$\phi^{(q)} \leftarrow \phi$

**end for**

---

### 0.3.1 Sampling $\beta_0$

Using the likelihood discussed in Section 0.2.1 and the uninformative prior from Section 0.2.2, it can be seen that the conditional posterior distribution of  $\beta_0$  is  $\mathcal{M}(\bar{\psi}, R_\psi \kappa)$ . To draw from this distribution, a new vector  $\psi$  is computed in each iteration, using the current values of  $\{\beta, \delta\}$ . Then, the corresponding values of  $\bar{\psi}$ , and  $R_\psi$  are computed. In this case, a Gibbs step can be applied, because it is straightforward to sample from the von Mises distribution, for example as in ?.

### 0.3.2 Sampling $\kappa$

Sampling  $\kappa$  is performed by employing a fast rejection sampler described by ?. This algorithm takes inputs  $\{m, \zeta\}$  and returns a new value from the conditional distribution of  $\kappa$ . Here, with an uninformative prior on the von Mises model,  $m = n$  and  $\zeta = -R_\psi \cos(\beta_0 - \bar{\psi})/n$ . For further details, see ?.

### 0.3.3 Sampling $\beta$

Sampling  $\beta$  is performed by a Metropolis-Hastings step (??). However, because of the irregular shape of the posterior, a random walk on  $\beta_k$  may cause slow convergence. If the current value for  $\beta_k$  is further from zero, we might prefer to propose candidates that are further away from the current value.

Therefore, motivated by the circular nature of the parameter space, candidates are generated by

$$\beta_k^* = \frac{\beta_k^{(cur)} + \tan(u\pi/2)}{1 - \beta_k^{(cur)} \tan(u\pi/2)}, \quad (10)$$

where  $u$  is a random variate from the uniform distribution  $U(-w, w)$ , with  $w$  a tuning parameter. Here, we choose  $w = .05$ . This procedure can be shown to be equivalent to drawing a proposal from the truncated Cauchy distribution  $f(\beta_k^* | \beta_k^{(cur)}, w) = 1 / (w\pi [1 + \beta_k^{*2}])$  with bounds

$$\left[ \frac{\beta_k^{(cur)} + \tan(-\pi w/2)}{1 - \beta_k^{(cur)} \tan(-\pi w/2)}, \frac{\beta_k^{(cur)} + \tan(\pi w/2)}{1 - \beta_k^{(cur)} \tan(\pi w/2)} \right].$$

Note that this proposal is not symmetric, although the Metropolis-Hastings ratio corrects for this lack of symmetry in the usual way.

### 0.3.4 Sampling $\delta$

It can be seen that the conditional posterior of each  $\delta_j$  is a convolution of two von Mises distributions, which itself is not von Mises. ?, p. 44 provide an approximation for such a convolution. Here, a slightly simpler approach is taking employing another Metropolis-Hastings step with von Mises proposals such that

$$\delta_j^* \sim \mathcal{M}(\delta_j^{(cur)}, R_\psi^{(cur)} \kappa^{(cur)}). \quad (11)$$

## 0.4 Hypothesis tests

In order to make decisions on a researcher's hypotheses, it is useful to consider hypothesis testing. A Bayesian approach to testing two discrete hypotheses against each other is by updating the prior odds of the hypotheses by multiplying them by the Bayes factor (??), in order to produce the posterior odds of the two hypotheses. From the posterior odds, we can obtain the posterior model probability of  $H_1$  compared to  $H_0$ ,  $p(H_1 | D)$ , where  $D$  represents the data, in this case  $\{\theta, \mathbf{X}, \mathbf{d}\}$ . This is an intuitive probability of interest, because it represents our current belief that  $H_1$  is true rather than  $H_0$ , and this probability might further be used to make decisions.

Two types of hypothesis tests are considered here. First, we consider traditional equality constrained hypotheses, comparing a null hypothesis to an alternative. Second, we consider inequality constrained hypotheses (??), where we test whether a parameter is larger than some other parameter, which can be a function of the parameters in the model or a fixed value.

#### 0.4.1 Equality constrained hypotheses

Consider two hypotheses about some model parameter  $\gamma$ ,

$$H_0 : \gamma = \gamma_0, \quad H_1 : \gamma \in \Omega_\gamma, \quad (12)$$

where  $\Omega_\gamma$  is the sample space of  $\gamma$ . The associated Bayes factor is given by

$$BF_{01} = \frac{p(D | H_0)}{p(D | H_1)}. \quad (13)$$

To obtain the Bayes factor, we must compute the quantity

$$p(D | H_s) = \int p(D, \phi_s | H_s) d\phi_s, \quad (14)$$

where  $\phi_s$  denotes the set of parameters in the model for hypothesis  $H_s$ . In general, this integral is not easy to compute, although special cases admit closed-form solutions. Here we will apply the Savage-Dickey method (??), following ?. This method is based upon the result that, under some assumptions,

$$\frac{p(D | H_0)}{p(D | H_1)} = \frac{p(\gamma = \gamma_0 | D, H_1)}{p(\gamma = \gamma_0 | H_1)}, \quad (15)$$

which is a ratio of the posterior and prior probability of  $\gamma_0$  under model  $H_1$ . In practice, this means that the probability of  $\gamma_0$  under  $H_1$  must be evaluated, both in the prior and the posterior. Although trivial in some conjugate situations, for the circular GLM this needs to be addressed separately for each parameter to which we wish to compute the Bayes factor.

In this study, this type of hypothesis test will be applied to both  $\delta$  and  $\beta$ . For some  $\beta_k$ , the hypotheses under evaluation are

$$H_0 : \beta_k = 0, \quad H_1 : \beta_k \in \mathbb{R}. \quad (16)$$

Taking a Normal prior on  $\beta_k$  as before,  $p(\beta_k = 0 | H_1) = p_N(0 | \mu = 0, \sigma^2 = 1) \approx .399$ .

For some  $\delta_j$ , the hypothesis under evaluation is

$$H_0 : \delta_j = 0, \quad H_1 : \delta_j \in [-\pi, \pi). \quad (17)$$

With a uniform prior on  $\delta_j$ ,  $p(\delta_j = 0 | H_1) = 1/2\pi \approx .159$ .

For  $p(\beta_k = 0 | D, H_1)$  and  $p(\delta_j = 0 | D, H_1)$ , a simple estimate is obtained by computing the height of the histogram bar that would contain 10% of the observations and which would have  $\gamma_0$  as its midpoint. A more sophisticated approach could employ log-spline distributions on the real line (?) and on the circle (?).

We note that if  $\gamma_0$  is far from the posterior samples of  $\gamma$ , this estimate will not be stable. However, ? note that evidence for  $H_1$  is overwhelming by that point, which means that accuracy is much less important.

Another remark to be made is that this method is only valid if the nuisance parameters between the two hypotheses serve the same purpose. For a discussion, see ?.

#### 0.4.2 Inequality constrained hypotheses

In practice, researchers often have directed (one-sided) hypotheses, which may be specified by using inequality constraints. Bayesian analysis of inequality constrained hypotheses has been studied by ? and ?.

For some model parameter  $\gamma$ , 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  $\gamma = \{\gamma^{(1)}, \dots, \gamma^{(Q)}\}$  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

$$BF_{01} = \frac{p(D | H_0)}{p(D | H_1)} = \frac{p(D | H_0)/p(D | H_{unc})}{p(D | H_1)/p(D | H_{unc})} = \frac{\sum_{s=1}^Q I(\gamma^{(s)} \in \Omega_{\gamma|H_0})}{\sum_{s=1}^Q I(\gamma^{(s)} \in \Omega_{\gamma|H_1})}, \quad (18)$$

where  $I(\cdot)$  is an indicator function,  $\gamma^{(s)}$  is a sample from the unconstrained model  $H_{unc}$ , and  $\Omega_{\gamma|H_s}$  is the admitted sample space for  $\gamma$  under hypothesis  $H_s$ .

This is a flexible approach, because it allows evaluation of any combination of inequality constrained hypotheses against each other. For example, consider a model with three groups, where we denote the mean directions by  $\{\mu_1, \mu_2, \mu_3\}$ . Then, a major advantage of the inequality constrained hypothesis approach is that it becomes easy to assess the model

$$\mu_1 > \mu_2 > \mu_3 \quad (19)$$

## 0.5 Simulation study

A simulation study was performed to assess the effectiveness of the proposed method. The sampler was implemented in Rcpp (?), and analyzed in R (?). Generally, the method converges fast and mixes well for each cell in the simulation, so that a burn-in of 1000 and a number of iterations of 20000 was deemed sufficient, with no thinning. Three different models are considered. First, a circular regression scenario with a single linear predictor. Second, a  $2 \times 2$  factorial ANOVA model with main effects only. Third, an ANCOVA model with a single grouping variable and four linear covariates. For all models, the artificial data featured (total) sample sizes  $n = \{20, 100\}$ , concentrations  $\kappa = \{2, 10\}$  and circular intercept  $\beta_0 = \pi/2$ . Additional simulations were performed with  $n = 50$ ,  $\kappa = 5$  and  $\beta, \delta = 0.2$ , which are not shown here for brevity's sake, as they provided similar results to the other scenarios.

For each scenario, 5000 datasets were generated and subsequently analyzed. Point estimates obtained from the MCMC sampler are  $\hat{\beta}_0$ , the posterior mean direction of  $\beta_0$ ,  $\hat{\kappa}$ , the posterior mode of  $\kappa$ ,  $\hat{\beta}$ , the posterior mean of  $\beta$ , and  $\hat{\delta}$ , the posterior mean direction of  $\delta$ . In addition, credible intervals are obtained from the posterior samples as well, by taking the circular quantiles of  $\beta_0$ , the Highest Posterior Density (HPD) interval of  $\kappa$ , the regular quantiles of  $\beta$ , and the circular quantiles of  $\delta$ . Circular quantiles of a set of angles  $\theta$  are obtained by computing the set of angles  $(\theta + \bar{\theta} - \pi)$ , obtaining the linear quantiles, and finally subtracting  $(\bar{\theta} - \pi)$  from the computed lower and upper bound.

In order to assess bias, point estimates were averaged over the datasets. For  $\beta_0$ , and  $\delta$ , this means computation of the mean direction, for other values this refers to the regular linear mean. In addition, a coverage was obtained for each parameter by computing the proportion of the appropriate credible intervals that contained the true value. Finally, acceptance probabilities refer to the proportion of proposals for this parameter that were accepted by the Metropolis-Hastings step, which is applicable only for  $\delta$  and  $\beta$ .

The three different scenarios for the predictors will be discussed separately in the following sections. The section will be concluded with a discussion of the behavior of the Bayes factors for these three scenarios.

### 0.5.1 Simple regression

For the simple regression data were simulated by first generating a vector  $\mathbf{x}_r$  independently from the standard normal distribution  $N(0, 1)$ . This vector is then standardized by  $\mathbf{x} = (\mathbf{x}_r - \bar{\mathbf{x}}_r)/\text{var}(\mathbf{x}_r)$ . Then, the circular outcome is computed by

$$\theta_i = \pi/2 + g(\beta_1 x_i) + \varepsilon_i, \quad (20)$$

where  $\beta$  takes values  $\{0.05, 0.8\}$ , and  $\varepsilon_i \sim \mathcal{M}(0, \kappa)$ .



Table 1: Results of the simulation study for the simple regression scenario. 'Cov.' denotes the 95% coverage for a specific parameter, while 'Acc.' denotes the acceptance probability. MCT denotes the mean computation time in seconds.

True			$\beta_0$		$\kappa$		$\beta_1$			MCT
$\beta$	$\kappa$	n	Bias	Cov.	$\hat{\kappa}$	Cov.	$\hat{\beta}_1$	Cov.	Acc.	
0.05	2	20	-0.00	0.95	2.09	0.97	0.05	0.94	0.84	0.58
		100	-0.00	0.95	1.99	0.98	0.05	0.95	0.66	2.26
	20	20	-0.00	0.94	22.70	0.95	0.05	0.94	0.47	0.59
		100	-0.00	0.95	20.36	0.95	0.05	0.94	0.23	2.37
	0.80	20	-0.00	0.94	2.08	0.97	0.81	0.95	0.86	0.58
		100	0.00	0.96	2.00	0.98	0.81	0.95	0.72	2.33
	20	20	-0.00	0.94	22.37	0.95	0.80	0.95	0.55	0.60
		100	0.00	0.95	20.37	0.95	0.80	0.94	0.29	2.44

Table 2: Results of the simulation study for the factorial ANOVA scenario. 'Cov.' denotes the 95% coverage for a specific parameter, while 'Acc.' denotes the acceptance probability. MCT denotes the mean computation time in seconds.

True			$\beta_0$		$\kappa$		$\delta_1$			MCT
$\delta$	$\kappa$	n	Bias	Cov.	$\hat{\kappa}$	Cov.	$\hat{\delta}_1$	Cov.	Acc.	
0.05	2	20	-0.00	0.95	2.10	0.97	0.06	0.95	0.78	0.70
		100	-0.00	0.95	1.99	0.98	0.05	0.95	0.78	2.63
	20	20	-0.00	0.93	22.75	0.95	0.05	0.93	0.79	0.71
		100	0.00	0.95	20.35	0.95	0.05	0.94	0.78	2.71
	0.80	20	-0.00	0.95	2.10	0.97	0.80	0.96	0.79	0.69
		100	0.00	0.95	2.00	0.98	0.80	0.95	0.78	2.62
	20	20	-0.00	0.94	22.65	0.95	0.80	0.93	0.78	0.71
		100	-0.00	0.95	20.35	0.95	0.80	0.95	0.78	2.71

Table 1 shows the results of this simulation study. Performance is quite good, providing unbiased estimates and great coverage, which shows that the weakly informative prior on  $\beta_1$  does not strongly harm the frequency properties of our Bayesian estimation procedure. Acceptance rates decline slightly for more concentrated data, which could be ameliorated in practice by tuning the proposals. The sampler runs quite fast, as the longest time a single analysis took was 2.44 seconds.

### 0.5.2 Factorial ANOVA

For the factorial ANOVA scenario, data were simulated by first generating two vectors,  $\mathbf{d}_1$  and  $\mathbf{d}_2$ , by randomly drawing either 0 or 1 with probability .5. This means that we generally create unbalanced designs, and the number of subjects in each group differs between the simulated datasets. This reflects a realistically broad range of possible research designs. The outcome is computed by

$$\theta_i = \pi/2 + \delta_1 d_{1i} + \delta_2 d_{2i} + \varepsilon_i, \quad (21)$$

where  $\delta_1 = \delta_2$  takes values  $\{0.05, 0.8\}$ , and  $\varepsilon_i \sim \mathcal{M}(0, \kappa)$ .

Table 2 shows the results of the simulation study. Once again, the sampler performs well in all situations. In contrast with  $\beta$  in the regression model, the acceptance rate of  $\delta$  does not depend on  $n, \kappa$ , or true  $\delta$ , because its proposal adapts itself to these parameters.

Table 3: Results of the simulation study for the ANCOVA scenario. ‘Cov.’ denotes the 95% coverage for a specific parameter, while ‘Acc.’ denotes the acceptance probability. MCT denotes the mean computation time in seconds.

True			$\beta_0$		$\kappa$		$\delta_1$			$\beta_1$			MCT
$\beta, \delta$	$\kappa$	n	Bias	Cov.	$\hat{\kappa}$	Cov.	$\hat{\delta}_1$	Cov.	Acc.	$\hat{\beta}_1$	Cov.	Acc.	
0.05	2	20	0.01	0.95	1.96	0.94	0.04	0.96	0.79	0.01	0.96	0.85	1.24
		100	-0.00	0.95	1.99	0.98	0.05	0.95	0.78	0.05	0.95	0.67	4.11
	20	20	0.00	0.93	22.88	0.96	0.05	0.94	0.79	0.05	0.94	0.47	1.21
		100	0.00	0.95	20.36	0.95	0.05	0.95	0.78	0.05	0.95	0.23	4.11
	0.80	20	-0.01	0.95	1.95	0.94	0.73	0.96	0.79	0.58	0.95	0.86	1.24
		100	0.00	0.87	1.78	0.88	0.79	0.82	0.78	0.60	0.88	0.73	4.32
	20	20	0.00	0.93	17.43	0.92	0.80	0.74	0.79	0.69	0.93	0.56	1.25
	100	100	-0.00	0.70	13.18	0.69	0.80	0.82	0.78	0.40	0.70	0.46	4.31

### 0.5.3 ANCOVA with four covariates

Here, a single grouping variable was generated as in Section 0.5.2, while the linear covariates were generated as in Section 0.5.1. Then, the outcome is computed by

$$\theta_i = \pi/2 + \delta_1 d_{1i} + g(\beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4}) + \varepsilon_i, \quad (22)$$

where we have chosen the true values  $\delta_1 = \beta_1 = \beta_2 = \beta_3 = \beta_4$  in all simulations, taking values  $\{0.05, 0.8\}$ , and  $\varepsilon_i \sim \mathcal{M}(0, \kappa)$ .

Table 3 shows the results of the simulation study. Only the results for  $\beta_1$  and  $\delta_1$  are shown, because results the other regression parameters are almost identical. For most scenarios, the results are once again adequate. It can be seen that the scenario with strong effects ( $\beta = \delta = 0.8$ ), estimates are often unsatisfactory. It should be noted that in these scenarios, the data show high variance. In general, if  $\sum_{k=1}^K |\beta_k| > 1.5$ , the data are spread all over the circle, such that many estimates of  $\beta$  are somewhat plausible, which results in an irregular posterior. Therefore, the sampler performs quite badly. Note that this is a property of the GLM approach to circular regression, rather than this specific model or implementation. In practice, however, we expect that this kind of dataset will almost never occur, although it might be advisable to monitor obtained estimates for this situation.

### 0.5.4 Bayes factors and posterior model probabilities

For hypothesis testing, Bayes factors and posterior model probabilities were obtained as detailed in Section 0.4. Figure 3 depicts boxplots of the posterior probability of the correct model for the inequality and equality hypotheses, for the three different models, where in each case all true  $\beta = \delta = 0.05$ . This means that in all cases  $H_1$  is true, so values close to 1 indicate that in a given scenario greatly prefers the correct model. The hypotheses given are listed below.

- Regression
  - Test for equality:  $H_0 : \beta_1 = 0$  vs.  $H_1 : \beta_1 \neq 0$ ,
  - Test for inequality:  $H_0 : \beta_1 < 0$  vs.  $H_1 : \beta_1 > 0$ ,
- ANOVA and ANCOVA
  - Test for equality:  $H_0 : \delta_1 = 0$  vs.  $H_1 : \delta_1 \neq 0$ ,
  - Test for inequality:  $H_0 : \delta_1 < 0$  vs.  $H_1 : \delta_1 > 0$ ,

Figure 3: Boxplot of the posterior model probability assigned to the correct model in 5000 simulations with  $\beta = \delta = 0.05$  for the inequality (top) and equality (bottom) hypotheses. For the regression scenario, inequality tests  $\beta_1 > 0$  vs.  $\beta_1 < 0$  and equality tests  $\beta_1 \neq 0$  vs.  $\beta_1 = 0$ . For the ANOVA and ANCOVA models inequality tests  $\delta_1 > 0$  vs.  $\delta_1 < 0$  and equality tests  $\delta_1 \neq 0$  vs.  $\delta_1 = 0$ .

Table 4: Summary statistics of the mean age (years), mean education (years) and mean direction of deviation (degrees).

	Age		Education		Deviation
	Mean	SD	Mean	SD	
Deaf	41.20	13.48	16.60	1.55	14.73
Interpreter	38.44	8.60	16.88	1.50	25.22
Control	44.75	9.73	17.06	1.44	28.82

Generally, the correct hypothesis becomes favored as the sample size increases, as expected. In addition, there is less simulation variability when  $n$  increases, shown by a smaller range in the boxplot. Compared to the inequality hypothesis, the equality hypothesis is more prone to pick up group differences in  $\delta$  (ANOVA and ANCOVA model), as well as in the regression model when  $\kappa = 2$ .

## 0.6 Example

In this section, our method will be applied to data from ?. In this study, an experiment was conducted to assess whether deafness enhances haptic perception. Haptic perception is assessed here by means of a haptic parallel setting task, where subjects are required to set two bars parallel, so that errors are measured in an angular difference to the reference direction. In this task, errors generally fall in the counterclockwise direction, which produces a positive score on the deviation from the target. Therefore, groups that are less apt at this task are expected to have stronger positive deviations, counterclockwise from the reference direction.

Three groups are distinguished: deaf subjects, sign language interpreters and a control group. Table 4 shows some summary statistics of background variables for the three groups, as well as the main outcome, deviation. Note that the original study examines two different conditions in a repeated measures design, an "immediate" and a "delayed" condition, of which we only show the "immediate" outcome for illustration purposes. Therefore, the data under consideration are made independent.

The analysis will proceed as follows. First, a basic ANOVA model will be fitted to this dataset. Then, an ANCOVA model is examined. Lastly, informative inequality constrained hypotheses based on theory are evaluated.

### 0.6.1 ANOVA model

The goal in the ANOVA model is to assess whether the three groups differ. For this and all following models, the control group will be used as the reference group. The outcome in the ANOVA model is given by  $\theta_i \sim \mathcal{M}(\mu_i, \kappa)$ , where  $\mu_i = \beta_0 + \delta_{df}d_{df} + \delta_{in}d_{in}$ , where  $d_{df}$  and  $d_{in}$  are dummy variables for the deaf and interpreter group, respectively, and each  $\delta$  is labeled appropriately. The main hypothesis is that the deaf group performs better than the control group, which could manifest itself as  $\delta_{df} < 0$ . Second, interpreters might also outperform the control group, which would mean  $\delta_{in} < 0$ .

A burn-in of 1000 was used, and the MCMC sampler was run for 100000 iterations. Figure 4 shows convergence plots for the four model parameters, where it can be seen that the sampler converges well.

Results are shown in Table 5. Estimates are given by the posterior mean direction of  $\beta_0, \delta_{df}, \delta_{in}$  and the posterior mode of  $\kappa$ . The credible interval of  $\delta_{df}$ , the difference between the deaf group mean direction and

Table 5: Results for the ANOVA model. LB and UB respectively represent lower and upper bound of the 95% credible interval of the given parameter.

	Estimate	LB	UB
$\beta_0$	0.51	0.38	0.63
$\kappa$	17.77	11.26	26.16
$\delta_{df}$	-0.25	-0.42	-0.08
$\delta_{in}$	-0.07	-0.24	0.11

Figure 4: Convergence plots for the ANOVA model.

the control group mean direction, is given by  $(-0.42, -0.08)$ , which can be seen as evidence for a non-zero group difference, as zero is not in this interval. The credible interval for the interpreter group mean direction is  $(-0.24, 0.11)$ , which can be seen as evidence against a non-zero group difference.

A more sophisticated approach is to employ the hypothesis tests that were developed. Employing the equality constrained hypotheses, a mild amount of support was found in favor of the hypothesis that deaf participants differ from the controls ( $BF_{\mu_{cn} \neq \mu_{df} : \mu_{cn} = \mu_{df}} = 2.69$ ), while a mild amount of support was found against the hypothesis that deaf participants differ from sign language interpreters ( $BF_{\mu_{in} \neq \mu_{df} : \mu_{in} = \mu_{df}} = 0.34$ ).

This highlights that this method is conservative in supporting alternative hypotheses such as  $H_1 : \mu_{in} \neq \mu_{df}$ . This is a result of the circular uniform prior on  $\delta$ , which suggests that more subjective approach could be less conservative and more likely to pick up on group differences. For example, a von Mises prior on  $\delta$  with  $\kappa > 0$  could be used. This subjective prior represents the knowledge that mean directions of different groups are usually somewhat close together. In this case, selection of the  $\kappa$  to be used becomes a core issue, which represents a trade-off between the amount additional power for the hypothesis test and the amount of (potentially unwanted) prior information included in the analysis.

The inequality hypothesis tests show a large amount of support for the hypothesis that deaf participants perform better than the controls ( $BF_{\mu_{cn} > \mu_{df} : \mu_{cn} < \mu_{df}} = 267.82$ ), and for the hypothesis that deaf participants perform better than sign language interpreters ( $BF_{\mu_{in} > \mu_{df} : \mu_{in} < \mu_{df}} = 52.28$ ).

### 0.6.2 ANCOVA model

In order to properly assess the effects found, it is useful to take into account theoretically relevant covariates. Here, we include age and handedness (that is, strength of hand preference) as covariates, and investigate the effect this has on previous conclusions. Table 6 shows the output of the ANCOVA model, with age and handedness as linear covariates.

From the results, it seems that the covariates do not have an effect on performance. Under equal prior odds, the Bayes factors for the predictors indeed indicate 29.35 times more support for the hypothesis

Table 6: Results for the ANCOVA model.

	Estimate	LB	UB
$\beta_0$	0.50	0.38	0.62
$\kappa$	18.30	11.54	27.19
$\delta_{df}$	-0.21	-0.39	-0.04
$\delta_{in}$	-0.08	-0.24	0.09
$\beta_{age}$	-0.02	-0.06	0.01
$\beta_{hand}$	0.03	-0.01	0.06

$\beta_{age} = 0$  compared to the alternative  $\beta_{age} \neq 0$ , and 19.08 times more support for the hypothesis  $\beta_{hand} = 0$  than the alternative  $\beta_{hand} \neq 0$ . While controlling for covariates, the evidence for a negative  $\delta_{df}$  is still substantive, indicating superior performance of deaf individuals over controls ( $BF^{\mu_{df} < \mu_{cn} : \mu_{df} > \mu_{cn}} = 96.75$ ). As there is little support for the inclusion of covariates, they are omitted in subsequent analyses.

### 0.6.3 Inequality constrained hypothesis

Although ? do not evaluate inequality constrained hypotheses directly, the theories stated by the authors can be interpreted as such. First, they state:

”On the basis of a greater proneness in visuospatial processing, we could expect a better developed haptic orientation processing ability in deaf individuals.”

Then, with regards to sign language interpreters, they state:

”The relative positions of the signer’s hands are used to map spatial relations in the real world. We may speculate here that experienced signers can also do the reverse more easily: interpret the hand positions forced by inspecting the bars in the parallel setting task in absolute world reference frames. If so we would expect both deaf and hearing signers to outperform non-signing hearing controls but not to differ from each other.”

These expectations can be mapped to hypotheses about the mean directions of the three groups, which are given by

$$(\text{Deaf}) \mu_{df} = \beta_0, \quad (\text{Interpreter}) \mu_{in} = \beta_0 + \delta_1, \quad (\text{Control}) \mu_{cn} = \beta_0 + \delta_2.$$

Following for example ? and ?, it is important to specify inequality constraints on circular data as either isotropic, or non-isotropic. Isotropic orderings are defined on the circle, and denote in which order the parameters are encountered as we move around the circle, relative to one another. Non-isotropic orderings are orderings relative to a fixed point on the circle. In this case, a type of non-isotropic orderings are considered where the hypothesis states that one parameter lies in the semi-circle counterclockwise of another parameter. In this case, we have chosen to translate the expectations to the following hypotheses:

$$H_1 : \mu_{df} < (\mu_{in}, \mu_{cn}) < \mu_{df} + \pi, \quad H_2 : \mu_{cn} - \pi < (\mu_{df}, \mu_{in}) < \mu_{cn}. \quad (23)$$

Using the inequality constrained framework as described in Section 0.4.2, the support for either hypothesis can be quantified.

Following this method, we find that  $H_1$  is true in 97.9% of the MCMC iterations, while  $H_2$  occurs in 78% of the iterations, so both hypotheses are likely. From this, we find no conclusive evidence for either  $H_1$  or  $H_2$  ( $BF_{H_1:H_2} = 1.25$ ). This means that although the study provides useful insight in the performance of deaf subjects, there is not enough evidence yet to decide on these two competing hypotheses.

## 0.7 Discussion

We developed a Bayesian circular GLM, with appropriate priors, proper treatment of dichotomous variables, and Bayesian hypothesis tests. Our method forms a middle ground between two veins of research into Bayesian analysis of circular data. On one hand, analysis of complex data shapes (???) provides modeling for a broad class of datasets but few possibilities for prediction and covariate models. On the other hand, circular regression models (???) provide prediction and covariates, but encounter problems with likelihood shapes and a lack of available hypothesis tests. The GLM approach is promising because of its flexibility to allow for many different kinds of models, while also allowing straightforward extensions. Our method brings three main contributions to the literature.

First, we have shown that the irregular log-likelihood surface of the regression parameters  $\beta$  in a circular GLM can be dealt with naturally by employing a weakly informative prior that encapsulates our actual belief

that extreme values for  $\beta$  are unlikely in applied research, while we still let the data overpower the prior. This is analogous to the widely accepted idea that very large effect sizes, say, Cohen’s  $d > 1$ , are improbable in most scientific disciplines where empirical research is necessary, in particular the social sciences. If the method would indicate support for such large values of  $|\beta|$ , a researcher would not believe that the model is correct, and reassess it. Our prior simply represents the lack of belief in large values of  $|\beta|$ .

Second, we have separated the group difference parameters from linear covariates, in order to allow modeling of a large array of ANOVA designs, including factorial and ANCOVA designs. This provides researchers with a straightforward way to map their hypotheses to a design. In addition, model comparison can easily be made possible through the DIC or WAIC (?, Ch. 7).

Third, we have developed Bayesian hypothesis tests based on the Bayes factor for the circular data case. The tests employed here are phased on the Savage-Dickey method advocated by ? and the inequality constrained approach of ?. Many Bayesian approaches to circular data analysis lack any form of hypothesis testing, which we view as limiting their ability to be applied in practice. In order to create statistical methods that are employed in practice, we must accomodate the desire for hypothesis testing, and compute posterior model probabilities. Therefore, we have taken a step in this direction as well, showing how Bayesian hypothesis tests can be developed easily in the circular data context by using MCMC output.

Although the computational methods employed here are stable and allow for useful inferences, further consideration of useful hypotheses and their associated Bayes factors will be important for the applicability of the Bayesian paradigm to circular data analysis, in particular in behavioural research. Here, we have not provided Bayes factors based on estimation of the marginal likelihood as in ? and subsequent work in this field, although this approach might be more flexible than the methods applied here. Another approach might be to attempt to develop priors that allow for closed-form Bayes factors in a similar vein as the  $g$ -prior in the linear case (??). The computational simplicity of such Bayes factors is useful in many scenarios, although the complexity of the designs for such Bayes factors are usually limited.

In the broader scope of circular data analysis, our method can be seen as a Bayesian extension of the approaches of ? and ?. Further extensions of this model might ease the assumption of i.i.d. observations taken here by applying properties of the multivariate von Mises distribution (??), as in ?.

In sum, 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.

## 0.8 Acknowledgements

This work was supported by a Vidi grant awarded to I. Klugkist from NWO, the Dutch Organization for Scientific Research (NWO 452-12-010).

The authors are grateful to two reviewers for helpful comments.

The authors are grateful to A. Postma for providing the illustrative data.

User-friendly code for the main analyses of the paper can be found in a GitHub package at <https://github.com/keesmulder/CircGLMBayes>.

All code for both the statistical tools, the simulation study and the paper is available online at <https://github.com/keesmulder/BayesMultCircCovariates>.

## .1 Conditional distribution of $\beta_0$

Here, it will be shown that  $\beta_0$  conditionally has the von Mises distribution, that is

$$L(\beta_0 \mid \kappa, \boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{X}, \mathbf{d}) \propto \mathcal{VM}(\beta_0 \mid \bar{\psi}, R_\psi \kappa). \quad (24)$$

The proof for the conditional distribution of  $\beta_0$  in the GLM closely follows the derivation for the distribution of the mean direction  $\mu$  of the von Mises distribution, which shows that  $L(\mu \mid \kappa, \boldsymbol{\theta}) \propto \mathcal{VM}(\mu \mid \bar{\theta}, R_\theta \kappa)$ .

The conditional likelihood of  $\beta_0$  is given by

$$\begin{aligned} L(\beta_0 \mid \kappa, \boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{X}, \mathbf{d}) &\propto \exp \left\{ \kappa \sum_{i=1}^n \cos \left[ \theta_i - (\beta_0 + \boldsymbol{\delta}^T \mathbf{d}_i + g(\boldsymbol{\beta}^T \mathbf{x}_i)) \right] \right\} \\ &= \exp \left\{ \kappa \sum_{i=1}^n \cos \left[ \beta_0 - (\theta_i - \boldsymbol{\delta}^T \mathbf{d}_i - g(\boldsymbol{\beta}^T \mathbf{x}_i)) \right] \right\}. \end{aligned}$$

We know that for any angle  $\psi_i, i = 1, \dots, n$ ,

$$\begin{aligned} C_\psi &= \sum_{i=1}^n \cos(\psi_i), \quad S_\psi = \sum_{i=1}^n \sin(\psi_i), \quad R_\psi = \sqrt{C_\psi^2 + S_\psi^2}, \\ \text{and } \frac{C_\psi}{R_\psi} &= \cos \bar{\psi}, \quad \frac{S_\psi}{R_\psi} = \sin \bar{\psi}, \quad \text{where } \bar{\psi} = (S_\psi, C_\psi). \end{aligned}$$

Thus, setting angle  $\psi_i = \theta_i - \boldsymbol{\delta}^T \mathbf{d}_i - g(\boldsymbol{\beta}^T \mathbf{x}_i)$ ,

$$\begin{aligned} L(\beta_0 \mid \kappa, \boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{X}, \mathbf{d}) &\propto \exp \left\{ \kappa \sum_{i=1}^n \cos(\beta_0 - \psi_i) \right\} \\ &= \exp \left\{ \kappa \left[ \cos \beta_0 \sum_{i=1}^n \cos \psi_i + \sin \beta_0 \sum_{i=1}^n \sin \psi_i \right] \right\} \\ &= \exp \left\{ R_\psi \kappa \left[ \cos \beta_0 \frac{C_\psi}{R_\psi} + \sin \beta_0 \frac{S_\psi}{R_\psi} \right] \right\} \\ &= \exp \left\{ R_\psi \kappa [\cos \beta_0 \cos \bar{\psi} + \sin \beta_0 \sin \bar{\psi}] \right\} \\ &= \exp \left\{ R_\psi \kappa \cos(\beta_0 - \bar{\psi}) \right\} \\ &\propto \mathcal{VM}(\beta_0 \mid \bar{\psi}, R_\psi \kappa). \end{aligned}$$