Bayesian model selection between the von Mises and the Wrapped Stable distributions for Circular Data

S. Rao Jammalamadaka, R. Gatto, D. Fouskakis

Abstract Bayesian model selection between two of the more commonly used circular models, namely the von Mises distribution and the Wrapped Symmetric α-Stable distribution is considered here. Our approach is based on posterior model probabilities and the corresponding posterior model odds, which are functions of Bayes factors. Marginal likelihoods under the two models are estimated based on prior distributions for the parameters that occur in these two competing models. The proposed methodology is analyzed and assessed through an extensive simulation study and shown to perform very well.

1 Introduction

Directional data, i.e. observations on directions, arise quite frequently in many scientific fields. For example, a biologist may be measuring the direction of flight of a bird, or a geologist may be dealing with data on paleocurrent analysis of a river flow. Several concrete examples are presented in detail in books such as Fisher (1995), Mardia and Jupp (2000), Jammalamadaka and SenGupta (2001) etc. Two-dimensional directions can also be represented as angles or simply as points on the circumference of the unit circle and therefore such observations are also called

---

S. Rao Jammalamadaka
Department of Statistics and Applied Probability, University of California, Santa Barbara, CA 93106, e-mail: rao@pstat.ucsb.edu

Riccardo Gatto
Institute of Mathematical Statistics and Actuarial Science, University of Bern, Alpeneggstrasse 22, 3012 Bern, Switzerland, e-mail: riccardo.gatto@unibe.ch

Dimitrios Fouskakis
Department of Mathematics, National Technical University of Athens, Zografou Campus, Athens 15780, e-mail: fouskakis@math.ntua.gr

circular data. See for instance the books cited earlier which discuss various statistical methods for analyzing circular data.

In many applications it is reasonable to assume that the circular data of interest is symmetrically distributed and therefore probability distributions having symmetric densities are generally more appropriate. Most prominent among these is the von Mises (vM) distribution, also called the Circular Normal distribution because of its many similarities to the Normal distribution on the real line. This is often the usual first choice for modeling typical unimodal symmetric circular data. Among several alternatives to this, is the Wrapped Normal distribution which is obtained by wrapping a Normal/Gaussian distribution on the real line, around the circle. Such a Wrapped Normal distribution is a member of the symmetric Wrapped α-Stable (WSαS) class of circular distributions (see e.g. Gatto and Jammalamadaka, 2003) that are constructed by wrapping the symmetric α-stable family of distributions around the circle.

The vM distribution has been extensively studied from a frequentist perspective, see for example Chapters 3-5 of Jammalamadaka and SenGupta (2001), major parts of which cover the sampling distribution theory and inference for this distribution. However, the Bayesian literature on the vM is far less extensive. Buckle (1995) performed Bayesian computation via Markov chain Monte Carlo in order to sample from the posterior distribution of the parameters of a stable distribution, while Hans (2007) introduced an approach for Bayesian inference in the setting of stable distributions that applies a fast Fourier transformation of the characteristic function in order to approximate the likelihood function. Ravindran and Ghosh (2011) proposed a data augmentation method using slice sampling to sample from the posterior distribution of the parameters of a wrapped stable distribution. The paper by Jammalamadaka et al (1987) is concerned with Bayes predictive inferences in regression models when the error terms are spherically distributed, while the paper by Chib, Tiwari and Jammalamadaka (1988) deals with the Bayes prediction problem for linear regression models with elliptical errors. Guttorp and Lockhart (1988) studied the problem of determining the location of an emergency transmitter in a downed aircraft, by developing conjugate prior distributions for the vM distribution, which they use to compute the posterior distribution of the location. George and Ghosh (2006) presented a Bayesian approach to regress a circular variable on a linear predictor, assuming that the regression coefficients have a nonparametric distribution with a Dirichlet process prior. Camli et al (2022) consider Bayesian lasso, which is a commonly used variable selection procedure in linear regression models, in the context of circular regression.

It is well known that the vM distribution can be reasonably well approximated by a wrapped normal distribution by equating their respective first-order trigonometric moments, equivalent to equating the magnitudes and directions of their mean resultant vectors (see Section 2.2.6 of Jammalamadaka and SenGupta (2001)). The WSαS class, itself a sub-family of the even wider and flexible wrapped α-stable family, which includes the Wrapped Normal and Wrapped Cauchy distributions besides others, as special cases.

In this study we consider Bayesian model selection between the vM and the WSαS distributions. Our approach is based on posterior model probabilities and the corresponding posterior model odds, starting with prior distributions on the parameters of the two competing models and by estimating the marginal likelihoods. The prior and posterior odds so obtained, allow us to calculate the Bayes factor for our model selection problem. The proposed methodology is verified through extensive simulation studies.

This chapter is organized as follows. Section 2 provides an overview of the two models for circular data under consideration, the vM and the WSαS. Section 3 introduces the Bayesian circular model selection problem: the Bayes factor is explained in Section 3.1, prior distributions are discussed in Section 3.2, marginal likelihoods and numerical algorithms for obtaining them are discussed in Section 3.3. Section 4 provides the results of the simulation study, with concluding remarks in Section 5.

2 Two central models for circular data

This section briefly reviews the two alternative model choices that we plan to consider: the vM is considered in Section 2.1, followed by the WSαS which is discussed in Section 2.2.

2.1 The von Mises distribution

A circular random variable θ is said to have the von Mises (vM) or Circular Normal distribution if it has a probability density function given by

$$f(\theta|\mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} \mathrm{e}^{\kappa \cos(\theta-\mu)}, \;\; \forall\theta \in [0, 2\pi), \tag{1}$$

where $0 \leq \mu < 2\pi$ is a measure of location which is also the mean direction, while $\kappa \geq 0$ is a measure of concentration toward μ. Here

$$I_\nu(z) = (2\pi)^{-1} \int_0^{2\pi} \cos(\nu\theta) \exp\{z \cos\theta\} d\theta, \forall z \in \mathbb{C}$$

, is the modified Bessel function of order ν, with $\Re\nu > -1/2$; cf. 9.6.18 at p. 376 of Abramowitz and Stegun (1972). In particular, $I_0(\kappa) = (2\pi)^{-1} \int_0^{2\pi} \exp\{\kappa \cos\theta\} d\theta$ and $I_1(\kappa) = (2\pi)^{-1} \int_0^{2\pi} \cos\theta \exp\{\kappa \cos\theta\} d\theta, \forall\kappa \geq 0$. Detailed presentations of this central circular distribution can be found for instance in Section 2.2.4 of Jammalamadaka and SenGupta (2001) or in Section 3.5.4 of Mardia and Jupp (2000). The length of the first trigonometric moment can be used as a measure of concentration toward μ and it is given by

$$\rho = \rho(\kappa) = ||A(\kappa)e^{i\mu}|| = A(\kappa),$$

where $A(\kappa) = I_1(\kappa)/I_0(\kappa)$ is a (strictly) increasing function in κ. The value of ρ lies in $[0, 1)$: the closer it is to 1, the higher the concentration toward the mean direction.

2.2 Wrapped Symmetric α-Stable distribution

Any linear random variable X on the real line can be transformed to a circular random variable θ by wrapping it around a circle of unit radius i.e. reducing it modulo 2π, viz. with $\theta = X \bmod 2\pi$. Using this idea, many wrapped circular distributions have been constructed. The trigonometric moment of order p for a wrapped circular distribution corresponds to the value of the characteristic function of the unwrapped random variable at the integer value p (see e.g. Proposition 2.1 of Jammalamadaka an SenGupta(2001). The α-stable distributions are considered as important generalizations of the Wrapped Normal distribution, having the property whose location-scale families are closed under convolution. They are described by the four parameters α, β, μ and τ: the parameter $\alpha \in (0, 2]$ defines the fatness of the tails (Gaussian tails for $\alpha = 2$ and heavy tails with infinite variance for any smaller α), the parameter $\beta \in [-1, 1]$ governs skewness ($\beta = 0$ corresponds to the symmetric case), the parameter $\mu \in \mathbb{R}$ is a location and the parameter $\tau > 0$ is a scale parameter.

The WSαS class of circular distributions are constructed by using the characteristic function of the α-stable distribution with $\beta = 0$ which is given by

$$\varphi(v) = e^{-\tau^\alpha |v|^\alpha + i\mu v}, \ \forall v \in \mathbb{R}.$$

Then, the probability density function of a WSαS random variable θ taking values in $[0, 2\pi)$ is given by

$$f(\theta|\mu, \alpha, \rho) = \frac{1}{2\pi} + \frac{1}{\pi} \sum_{k=1}^{\infty} \rho^{k^\alpha} \cos(k[\theta - \mu]), \ \forall \theta \in [0, 2\pi), \tag{2}$$

where $0 \le \mu < 2\pi$ is a measure of location on the circle (the mean direction), $\rho = e^{-\tau^\alpha} \in (0, 1]$ is a measure of concentration towards μ (the closer it is to 1 the more the concentration towards the mean) and $\alpha \in (0, 2]$ is a shape parameter. For $\alpha = 2$ we get the Wrapped Normal density, while for $\alpha = 1$ we have the Wrapped Cauchy density, which has the simple form

$$f(\theta|\mu, 1, \rho) = \frac{1}{2\pi} \frac{1 - \rho^2}{1 + \rho^2 - 2\rho \cos(\theta - \mu)}, \ \forall \theta \in [0, 2\pi), \tag{3}$$

where $\rho = e^{-\tau}$. In this article we consider the value of α as known. We will however not consider the value $\alpha = 2$ or other values of α that are very close to 2, because in this case the WSαS distribution can be very close to the vM (cf. e.g. p. 45 of Jammalamadaka and SenGupta, 2001) and our test would not be able to distinguish these two models.

3 Bayesian circular model selection

After a brief overview on the Bayes factor in Section 3.1, prior distributions are discussed in Section 3.2 and marginal likelihoods and related Monte Carlo algorithms are given in Section 3.3.

3.1 Determination of the Bayes factor

Let us denote by $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$ a sample of n independent circular random variables, all of them generated from one of the two following candidate models,

$$M_1 : \text{vM}(\mu, \kappa) \quad \text{or} \quad M_2 : \text{WS}\alpha\text{S}(\mu, \alpha, \rho).$$

Let $M \in \{M_1, M_2\}$. Within the Bayesian framework, the identification of the best model between the above two competitors is equivalent to finding the model with the highest posterior model probability, defined as

$$p(M|\boldsymbol{\theta}) = \frac{f(\boldsymbol{\theta}|M)p(M)}{f(\boldsymbol{\theta}|M_1)p(M_1) + f(\boldsymbol{\theta}|M_2)p(M_2)}, \tag{4}$$

where $f(\boldsymbol{\theta}|M)$ is the marginal likelihood of model M with sample $\boldsymbol{\theta}$ and $p(M)$ is the prior model probability of model M. The marginal likelihood function (4) can be further expanded to include the effect of the model parameters as follows,

$$f(\boldsymbol{\theta}|M) = \int f(\boldsymbol{\theta}|\boldsymbol{\psi}_M, M) f(\boldsymbol{\psi}_M|M) d\boldsymbol{\psi}_M, \tag{5}$$

where $f(\boldsymbol{\theta}|\boldsymbol{\psi}_M, M)$ is the likelihood under model M with parameters $\boldsymbol{\psi}_M$ and $f(\boldsymbol{\psi}_M|M)$ is the prior distribution of model parameters, given the model M.

Closed form expression of the marginal likelihoods (5) and therefore of the posterior model probabilities (4) is available only in special cases; several approaches exist to compute the marginal likelihood; for a nice review, the reader is referred to Friel and Wyse (2012).

In the setting of Bayesian hypothesis testing, the Bayes factor B is an indicator of the support of the sample for the null hypothesis H_0 against some alternative hypothesis H_1. It is precisely the ratio of the posterior odds over the prior odds, of H_0 versus H_1, which is symbolically given by

$$B = \frac{P\,[\mathrm{H}_0|\text{sample}]}{P\,[\mathrm{H}_1|\text{sample}]}\frac{P\,[\mathrm{H}_1]}{P\,[\mathrm{H}_0]}.$$

This can obviously be re-expressed as

$$B = \frac{P\,[\text{sample}|\mathrm{H}_0]}{P\,[\text{sample}|\mathrm{H}_1]}.$$

When B is larger than 1, the sample supports H_0, and when B is smaller than 1, the sample supports H_1. In our setting of model selection, the two hypotheses H_0 and H_1 are replaced by the models M_1 and M_2 respectively. Thus the Bayes factor for model selection problem becomes

$$B = \frac{p(M_1|\boldsymbol{\theta})}{p(M_2|\boldsymbol{\theta})}\frac{p(M_2)}{p(M_1)} = \frac{f(\boldsymbol{\theta}|M_1)}{f(\boldsymbol{\theta}|M_2)}. \quad (6)$$

Table 1 provides some heuristic or practical guidelines for interpreting values of Bayes factors. They are due to Jeffreys (1939) and Kass and Raftery (1995). We refer to these references for details on these interpretations. Essentially, a Bayes factor slightly greater than 1 provides significant support for H_0. Other aspects such as sample size may also be taken into account when interpreting the magnitude of a Bayes factor.

B	**evidence for H_0 versus H_1**
< 1	negative
1 to 1.5	significant
1.5 to 5	positive
5 to 10	substantial
10 to 20	strong
> 20	decisive

Table 1 Practical interpretation of the values taken by the Bayes factor B

3.2 Choice of Prior distributions

We use the same priors for parameters that are common to both models M_1 and M_2. Let $M \in \{M_1, M_2\}$. The beta prior is used for the concentration parameter ρ of model M, i.e. $\rho|M \sim \mathrm{Beta}(a, b)$, where $a, b > 0$, with density

$$f(\rho|M) = \frac{1}{\mathrm{B}(a,b)}\rho^{a-1}(1-\rho)^{b-1}, \quad \forall \rho \in [0,1]$$

where $\mathrm{B}(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx$. The beta distribution is a commonly used prior because it provides densities with various shapes, that include the flat viz.

uniform density, and because it is the conjugate prior to binomial and negative binomial likelihoods. In our setting, although the beta prior yields neither conjugate classes nor analytical solutions to integrals, it represents a simple and flexible choice of prior for the concentration. Given the reparametrization $\kappa = A^{-1}(\rho)$ needed in M_1, namely the vM distribution, we obtain the prior density for the concentration parameter κ as

$$\begin{aligned} f(\kappa|M_1) &= \frac{1}{\mathrm{B}(a,b)} A^{a-1}(\kappa)\big\{1 - A(\kappa)\big\}^{b-1} \frac{d}{d\kappa} A(\kappa) \\ &= \frac{1}{\mathrm{B}(a,b)} A^{a-1}(\kappa)\big\{1 - A(\kappa)\big\}^{b-1} \left(1 - \frac{A(\kappa)}{\kappa} - A^2(\kappa)\right), \quad \forall \kappa \geq 0. \end{aligned} \tag{7}$$

We use a uniform or isotropic prior for the mean direction μ in M, i.e.

$$f(\mu|M) = \frac{1}{2\pi}, \ \forall \mu \in [0, 2\pi).$$

3.3 Marginal likelihoods

Analytical forms of vM and WSαS marginal likelihood are provided in Section 3.3.1 and Monte Carlo numerical algorithms for obtaining Bayes factors of our model selection problems are presented in Section 3.3.2.

3.3.1 Analytical forms of vM and WSαS marginal likelihood

• vM *marginal likelihood*
Under model M_1, the marginal likelihood according to (5) is given by

$$\begin{aligned} f(\boldsymbol{\theta}|M_1) &= \int_0^\infty \int_0^{2\pi} \prod_{i=1}^n \frac{1}{2\pi I_0(\kappa)} \mathrm{e}^{\kappa \cos(\theta_i - \mu)} f(\kappa|M_1) f(\mu|M_1) d\mu d\kappa \\ &= (2\pi)^{-n} \int_0^\infty \frac{1}{I_0^n(\kappa)} \int_0^{2\pi} \mathrm{e}^{\kappa \sum_{i=1}^n \cos(\theta_i - \mu)} f(\mu|M_1) f(\kappa|M_1) d\mu d\kappa \\ &= (2\pi)^{-n} \int_0^\infty \frac{1}{I_0^n(\kappa)} \int_0^{2\pi} \mathrm{e}^{\kappa \sum_{i=1}^n \cos(\theta_i - \bar{\theta}_n) \cos(\bar{\theta}_n - \mu)} f(\mu|M_1) f(\kappa|M_1) d\mu d\kappa \\ &= (2\pi)^{-n} \int_0^\infty \frac{I_0(\kappa R_n)}{I_0^n(\kappa)} f(\kappa|M_1) d\kappa, \end{aligned} \tag{8}$$

where the sample mean direction $\overline{\theta}_n$ and the sample resultant length are defined by

$$\sum_{i=1}^n \cos(\theta_i - \overline{\theta}_n) = R_n \ \text{ and } \ \sum_{i=1}^n \sin(\theta_i - \overline{\theta}_n) = 0.$$

They are indeed direction and length of the sample resultant vector

$$\boldsymbol{R}_n = \left(\sum_{i=1}^{n} \cos\theta_i, \sum_{i=1}^{n} \sin\theta_i \right)$$

. / • WSαS *marginal likelihood*
Under model M_2 the marginal likelihood is given by

$$f(\boldsymbol{\theta}|M_2) = (2\pi)^{-n} \int_0^1 \int_0^{2\pi} \prod_{i=1}^{n} \left(1 + 2 \sum_{k=1}^{\infty} \rho^{k^{\alpha}} \cos(k[\theta_i - \mu]) \right) \frac{1}{2\pi} f(\rho|M_2) d\mu d\rho, \tag{9}$$

when $\alpha \neq 1$, and by

$$f(\boldsymbol{\theta}|M_2) = (2\pi)^{-n} \int_0^1 \int_0^{2\pi} \prod_{i=1}^{n} \frac{1-\rho^2}{1+\rho^2-2\rho\cos(\theta_i-\mu)} \frac{1}{2\pi} f(\rho|M_2) d\mu d\rho, \tag{10}$$

when $\alpha = 1$; cf. (3).

3.3.2 Numerical evaluation of the marginal likelihoods and Bayes factor

Unfortunately, neither of these marginal likelihoods (8) or (9) can be analytically obtained. Thus we must rely on numerical methods, as we do below.

• vM *marginal likelihood*
We can evaluate $f(\boldsymbol{\theta}|M_1)$ given in (8) by numerical integration. For this purpose it would be preferable to obtain a finite integration domain, for example through the change of variables $\kappa = -\log\lambda$. This gives

$$\begin{aligned} & f(\boldsymbol{\theta}|M_1) \\ & = \frac{1}{(2\pi)^n \mathrm{B}(a,b)} \int_0^{\infty} \frac{I_0(\kappa R_n)}{I_0^n(\kappa)} A^{a-1}(\kappa) \left\{1 - A(\kappa)\right\}^{b-1} \left(1 - \frac{A(\kappa)}{\kappa} - A^2(\kappa)\right) d\kappa \\ & = \frac{1}{(2\pi)^n \mathrm{B}(a,b)} \int_0^1 \frac{I_0(-\log\{\lambda\}R_n)}{I_0^n(-\log\lambda)} A^{a-1}(-\log\lambda) \left\{1 - A(-\log\lambda)\right\}^{b-1} \\ & \qquad \cdot \left(1 + \frac{A(-\log\lambda)}{\log\lambda} - A^2(-\log\lambda)\right) \frac{d\lambda}{\lambda}. \end{aligned} \tag{11}$$

One can also use Monte Carlo integration. Generating from the density (7) in order to approximate the integral (8) appears complicated, because we have a non-standard density over an unbounded domain. We thus re-express the vM marginal likelihood as

$$f(\boldsymbol{\theta}|M_1) = \int_0^1 \frac{I_0(A^{-1}(\rho)R_n)}{\{2\pi I_0(A^{-1}(\rho))\}^n} f(\rho|M_1) d\rho,$$

in order to obtain the following Monte Carlo algorithm.

Algorithm MC-vM-post for the marginal likelihood under vM
For $j = 1, \dots, s$,
generate ρ_j from the Beta(a, b) distribution;
compute $X_j = \dfrac{I_0(A^{-1}(\rho_j)R_n)}{\{2\pi I_0(A^{-1}(\rho_j))\}^n}$.
Estimate the posterior density of model M_1 by

$$\hat{f}(\boldsymbol{\theta}|M_1) = \frac{1}{s}\sum_{j=1}^{s} X_j.$$

One other difficulty in using Algorithm vM-post is the numerical evaluation of

$$x = A^{-1}(p), \tag{12}$$

where $p \in (0, 1)$. Although we can avoid evaluating A^{-1} in the representation (11) proposed for numerical integration, this evaluation is required in the above Algorithm vM-post for the Monte Carlo integration. However and fortunately, solving (12) has received considerable attention in the literature. We first note that the function $A = I_1/I_0$ is continuous and (strictly) increasing probability distribution function and so A^{-1} is a simple quantile function. It can be well approximated by

$$A^{-1}(p) = \begin{cases} 2p + p^3 + 0.83p^5, & \text{if } 0 \le p < 0.53, \\ -0.4 + 1.39p + 0.43(1-p)^{-1}, & \text{if } 0.53 \le p < 0.85, \\ (p^3 - 4p^2 + 3p)^{-1}, & \text{if } 0.85 \le p < 1. \end{cases} \tag{13}$$

This approximation has been successfully used for the computation of the maximum likelihood estimator of κ (see e.g. p. 88 of Fisher, 1995) and for obtaining the saddlepoint approximation to the distribution of the total distance of the planar Pearson-vM random walk (see Gatto, 2017).

• WSαS *marginal likelihood*
The evaluation of $f(\boldsymbol{\theta}|M_2)$ given in (9) can be done by numerical integration of

$$f(\boldsymbol{\theta}|M_2) = \frac{1}{(2\pi)^{n+1}\mathrm{B}(a,b)} \int_0^1 \rho^{a-1}(1-\rho)^{b-1} \int_0^{2\pi} \prod_{i=1}^{n} \left(1 + 2\sum_{k=1}^{\infty} \rho^{k^\alpha} \cos(k[\theta_i - \mu])\right) d\mu d\rho. \tag{14}$$

When $\alpha = 1$, we use (3) in order to simplify (14) to

$$f(\boldsymbol{\theta}|M_2) = \frac{1}{(2\pi)^{n+1}\mathrm{B}(a,b)} \cdot \int_0^1 \rho^{a-1}(1-\rho)^{b+n-1}(1+\rho)^n \int_0^{2\pi} \left\{ \prod_{i=1}^{n} [1+\rho^2 - 2\rho\cos(\theta_i - \mu)] \right\}^{-1} d\mu d\rho. \tag{15}$$

We can also apply Monte Carlo integration: (9) and (10) give directly the following algorithm.

Algorithm MC-WSαS**-post** for the marginal likelihood under WSαS
For $j = 1, \ldots, s$,
generate μ_j from Uniform$(0, 2\pi)$;
generate ρ_j from Beta(a, b);
compute $Y_j = \prod_{i=1}^{n} \frac{1}{2\pi}\left(1 + 2\sum_{k=1}^{\infty} \rho_j^{k^\alpha} \cos(k[\theta_i - \mu_j])\right)$, if $\alpha \neq 1$, or
compute $Y_j = \prod_{i=1}^{n} \frac{1-\rho_j^2}{2\pi\{1+\rho_j^2 - 2\rho_j\cos(\theta_i - \mu_j)\}}$, if $\alpha = 1$.
Estimate the posterior density of model M_2 by

$$\hat{f}(\boldsymbol{\theta}|M_2) = \frac{1}{s}\sum_{j=1}^{s} Y_j.$$

• *Bayes factor*
Computational versions of the Bayes factor follow directly from the representation (6). It can be obtained either by numerical integration of (11) and (14) or by joining Algorithms MC-vM-post and MC-WSαS-post as follows.

Algorithm MC-Bayes-fact for the Bayes factor of model selection
Generate $X_1, \ldots, X_s$ from Algorithm MC-vM-post.
Generate $Y_1, \ldots, Y_s$ from Algorithm MC-WSαS-post.
Compute the Bayes factor

$$\hat{B} = \frac{\sum_{j=1}^{s} X_j}{\sum_{j=1}^{s} Y_j}.$$

4 A Simulation study

In this section we present a simulation study that verifies the effectiveness of the proposed model selection procedure. For this purpose, we generate samples from various vM distributions and compute Bayes factors based on underlying vM distributions compared to WSαS distributions, which are not the correct models. We thus expect Bayes factors larger than one and substantially larger when the alternative model used is a more distant or incorrect WSαS distribution.

In particular, we generate $r = 1000$ samples $\boldsymbol{\theta}$ of sizes $n = 40, 100$ from vM$(0, \kappa)$ with $\kappa = 1, 2, 4$. The Beta(a, b) prior parameters are $a = b = 2$, yielding a symmetric prior around the center $1/2$. We then compute the Bayes factors $\hat{B}$ by means of Algorithm MC-Bayes-fact. The results are shown in Table 2, in which each Bayes factor displayed is in fact the mean of the $r = 1000$ simulated Bayes factors. The standard deviation of these simulations is shown in parentheses. Simulated Bayes factors $\hat{B}$ for $n = 40$ are in Table 2 (a) and for $n = 100$ in Table 2 (b).

All Bayes factors are as expected, greater than 1 and often substantially greater than 1. We refer to Table 1 for precise interpretations of these values. We see that there is clearly more support for the vM distribution against the WSαS distribution (or for M_1 against M_2), with $n = 100$ than with $n = 40$, as expected. Also, the Bayes factors decrease with larger values of the stability index α, which is expected because as α approaches 2, the WSαS approaches the vM. Large values of κ do also lead to larger Bayes factors, because the uncertainty decreases with large κ and the distinction between the two models becomes more apparent. Only $\alpha = 3/2$ and $\kappa = 1$ give Bayes factor values close to 1, still providing significant evidence for H_0 versus H_1, according Table 1.

(a) $n = 40$

α \ κ	**1**	**2**	**4**
3/4	7.56 (1.14)	389.66 (76.24)	11465.26 (1399.98)
1	2.28 (0.18)	25.68 (2.44)	264.69 (18.78)
3/2	1.03 (0.01)	1.57 (0.04)	4.57 (0.14)

(b) $n = 100$

α \ κ	**1**	**2**	**4**
3/4	255.12 (108.94)	5349795 (1907579)	1.24×10^{12} (1.07×10^{12})
1	8.22 (0.97)	5996.47 (1329.73)	2.46×10^7 (1.63×10^7)
3/2	1.06 (0.01)	3.54 (0.20)	199.94 (38.94)

Table 2 Simulated Bayes factors $\hat{B}$ with samples $\boldsymbol{\theta}$ generated from vM$(0, \kappa)$ with $\kappa = 1, 2, 4$, for $n = 40$ (a) and $n = 100$ (b). The Beta(a, b) prior parameters are $a = b = 2$. Each Bayes factor is the mean of $r = 1000$ simulations, with standard deviation shown in parentheses.

Moreover, Figure 1 provides the boxplots of the $r = 1000$ simulated Bayes factors $\hat{B}$ of Table 2 for the cases $\kappa = 1$ (left figure), $\kappa = 2$ (right figure), for $\alpha = 3/2$ and for $n = 40, 100$. Because boxplots with $\kappa = 2$ appear right-skewed, Figure 2 provides the boxplots for the same cases of Figure 1 but in logarithmic scale, i.e. for $\log \hat{B}$.

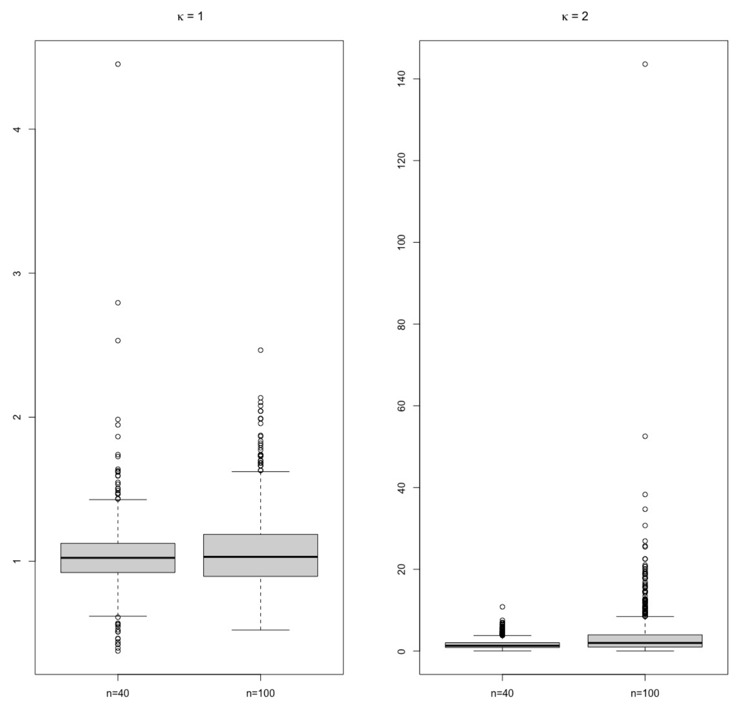

Fig. 1 Boxplots of $r = 1000$ simulated Bayes factors $\hat{B}$ with samples $\boldsymbol{\theta}$ generated from vM$(0, \kappa)$ with $\kappa = 1$ (left figure), $\kappa = 2$ (right figure) and for $n = 40$, 100. The Beta(a, b) prior parameters are $a = b = 2$ and the WSαS has $\alpha = 3/2$.

5 Summary and Conclusions

Bayesian model selection between the vM and the WSαS distributions is investigated in this chapter. Algorithms for obtaining the marginal likelihoods under the two models are suggested. These algorithms allow us to obtain the Bayes factor for the considered model selection problem. A simulation study confirms that the proposed technique can be advantageously used in scientific studies of circular data. However, we should remark here that our procedure of estimating the marginal likelihoods by sampling from the priors is a naive Monte Carlo estimator, which may become quite inefficient in some cases, especially when the prior distribution differs considerably

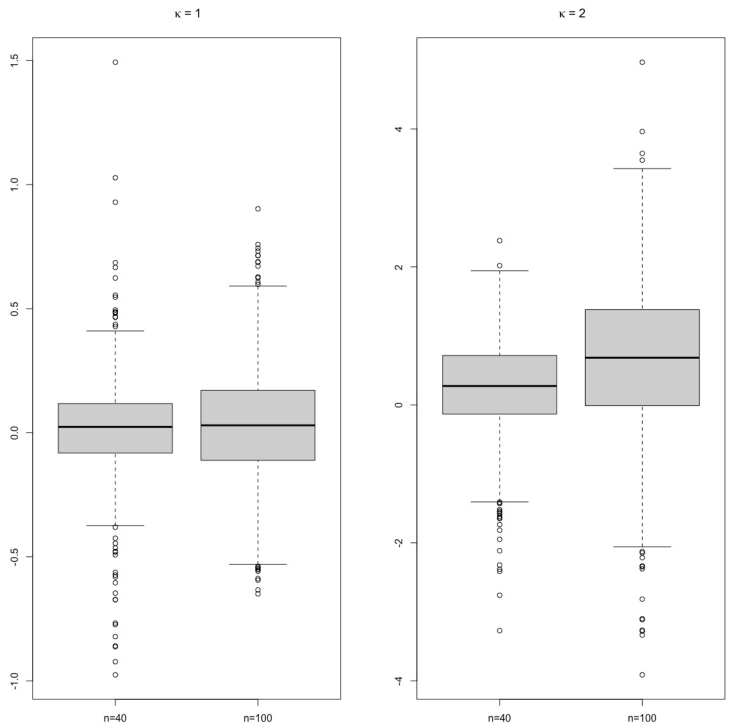

Fig. 2 Boxplots of $r = 1000$ simulated logarithmic Bayes factors $\log \hat{B}$ with samples $\boldsymbol{\theta}$ generated from vM$(0, \kappa)$ with $\kappa = 1$ (left figure), $\kappa = 2$ (right figure) and for $n = 40, 100$. The Beta(a, b) prior parameters are $a = b = 2$ and the WSαS has $\alpha = 3/2$.

from the actual posterior. In such a case, the likelihood values generated from the prior can often be zero, providing minimal contribution to the final summation. This can result in large standard errors and slow convergence. There are more subtle and methods for generating data from marginal likelihoods that could be explored in the future.

Marginal likelihoods of other models may be similarly obtained and so the proposed methodology can in principle be extended to other circular models. It may also be applied in choosing between any number of models (more than two) for a given circular data set.

Acknowledgments

The authors are thankful to the editors for the invitation to present our work and to the two reviewers for their comments. They also thank Mr. Hyeongseong Lee of the Department of Statistics and Applied Probability at the University of California, Santa Barbara for help with the numerical study.

References

1. Abramowitz, M. and Stegun, I. (1972), "Handbook of Mathematical Functions", United States Department of Commerce.
2. Buckle, D. (1995), "Bayesian inference for stable distributions", *Journal of the American Statistical Association*, 90, 605–613.
3. Camli, O., Kalaylioglu, Z., SenGupta, A. (2022), "Variable selection in linear-circular regression models", *Journal of Applied Statistics*.
4. Chib, S., Tiwari, R. and Jammalamadaka, S. Rao (1988), "Bayes prediction in regressions with elliptical errors", *Journal of Econometrics*, 38, 349–360.
5. Evans, M., Hastings, N. and Peacock, B. (2000), *Statistical Distributions*, 3rd ed., John Wiley, New York.
6. Fisher, N. I. (1995), *Statistical Analysis of Circular Data*, Cambridge University Press.
7. Friel, N. and Wyse, J. (2012), "Estimating the statistical evidence: a review", *Statistica Neerlandica*, 66, 288–308.
8. Gatto, R. (2017), "Saddlepoint approximations to the distribution of the total distance of the multivariate isotropic and von Mises–Fisher random walks", *Mathematical Methods of Statistics*, 26, 20–36.
9. Gatto, R. and Jammalamadaka, S. Rao (2003), "Inference for wrapped symmetric α-stable circular models", *Sankhya*, 65, 333–355.
10. George, B. and Ghosh, K. (2006), "A semiparametric Bayesian model for circular-linear regression", *Communications in Statistics Simulation and Computation*, 23, 911–923.
11. Guttorp, P. and Lockhart, R. (1988), "Finding the location of a signal: A Bayesian analysis", *Journal of the American Statistical Association*, 83, 322–330.
12. Hans, C. (2007), "Bayesian inference for α-stable distributions: a random walk MCMC approach", *Computational Statistics and Data Analysis*, 51, 2688-2700.
13. Jammalamadaka, S. Rao and SenGupta, A. (2001), *Topics in Circular Statistics*, World Scientific.
14. Jammalamadaka, S. Rao, Tiwari, R. and Chib, S. (1987), "Bayes prediction in the linear model with spherically symmetric errors", *Economics Letters*, 24, 39-44.
15. Jeffreys, H. (1939), *Theory of Probability*, Oxford University Press.
16. Kass, R. E. and Raftery, A. E. (1995), "Bayes factors", *Journal of the American Statistical Association*, 90, 773-795.
17. Mardia, K. V. and Jupp, P. E. (2000), *Directional Statistics*, Wiley.
18. Ravindran, P., Ghosh, S. K. (2011), "Bayesian analysis of circular data using wrapped distributions", *Journal of Statistical Theory and Practice*, 5, 547-561.