BAYESIAN INFERENCE AND MONTE CARLO METHODS FOR DIRECTIONAL DATA

Inaugural dissertation
of the Faculty of Science,
University of Bern

presented by

Sara Salvador

from Italy

Supervisor of the doctoral thesis:

Prof. Dr. Riccardo Gatto

University of Bern

BAYESIAN INFERENCE AND MONTE CARLO METHODS FOR DIRECTIONAL DATA

Inaugural dissertation
of the Faculty of Science,
University of Bern

presented by

Sara Salvador

from Italy

Supervisor of the doctoral thesis:

Prof. Dr. Riccardo Gatto

University of Bern

Accepted by the Faculty of Science

Bern, December, 16th 2022

The Dean
Prof. Dr. Marco Herwegh

Contents

Abstract

This thesis deals with directional distributions, hence distributions defined over $\mathbb{R}^p$. We will consider in details the case of circular distributions, i.e. directional distributions in $\mathbb{R}^2$. We will focus on the study of an extension of the von Mises distribution (vM), namely the generalized von Mises distribution of order two (GvM_2, or simply GvM). This distribution allows higher flexibility in terms of asymmetry and bimodality than the vM distribution. Two Bayesian tests are computed on the GvM. The first test concerns the symmetry of the GvM. Inference on symmetry is made via Bayes factors. A real circular data case is considered. The second test concerns the bimodality of the GvM. The problem is reduced to the study of the real roots of a quartic whose coefficients depend on the parameters of the model. A detailed analysis of the quartic is given and a region $\mathcal{W}$ of parameters that are associated to bimodality is obtained. Then, inference on bimodality is made via Bayes factors and via highest posterior density (HPD) credible sets. We use Markov Chain Monte Carlo methods (MCMC) to compute posterior probabilities of bimodality. Conclusions confirm that, when data are in accordance with the null hypotheses, Bayes factors are typically large. On the other hand, the HPD credible set is entirely contained inside $\mathcal{W}$ and bimodality is confirmed.

Moreover, in this thesis we consider directional distributions over the unit sphere $\mathbb{S}^{p-1}$ of $\mathbb{R}^p$. These are called spherical distributions and here we manly focus on the generalized von Mises-Fisher (GvMF) distribution. This is an extension of the von Mises-Fisher distribution (vMF). We will give two methods to generate random variables from the GvMF using a conditional acceptance-rejection method and the Metropolis Hastings algorithm.

Chapter 1

Introduction

In various scientific fields measurements can take the form of directions: the direction flight of a bird and the direction of earth's magnetic pole are two examples. These directions can be in the plane, namely in two dimensions, as in the first example, or they can be in the space, namely in three dimensions, as in the second example. These measurements are called directional data and they are of main interest in this thesis. They appear in various scientific fields: in the analysis of protein structure, cf. e.g. Kim et al. (2016), in machine learning, cf. e.g. Navarro et al. (2017), in ecology, cf. e.g. S Rao and Ulric (2006), in ornithology, cf. e.g. Schmidt-Koenig (1963), in oceanography, cf. e.g. Lin and Dong (2019), in meteorology, c.f. e.g. Zhang et al. (2018), etc. A two-dimensional direction is a point in $\mathbb{R}^2$ without magnitude, e.g. a unit vector. It can also be represented as a point on the circumference of the unit circle or as an angle, measured for example in radians, after fixing the null direction and the sense of rotation (clockwise or counter-clockwise). Because of this circular representation, observations of two-dimensional directional data are distinctively called circular data. During the last two or three decades, there has been a rise of interest for statistical methods for circular (or in general directional) data. Recent applications can be found e.g. in Ley and Verdebout (2018). Some monographs on this topic are Mardia and Jupp (2000), Jammalamadaka and SenGupta (2001), Ley and Verdebout (2017) and also Pewsey et al. (2013). A basic introduction is given in Gatto and Jammalamadaka (2014) and a recent review can be found in Pewsey and García-Portugués (2021). We refer to Fisher (1995) for the presentation of circular data from various scientific fields. The most popular circular distribution is the von Mises (vM), sometimes also called circular normal distribution. This distribution is circularly symmetric around its unique mode. Until a couple decades ago, very few asymmetric circular distributions were available, two of these can be found in Sections 15.6 and 15.7 of Batschelet (1981). In recent years, various asymmetric and multimodal circular distributions have been introduced, for example: Umbach and Jammalamadaka (2009), Kato and Jones (2015), Abe et al. (2013), Gatto and Jammalamadaka (2003) and the generalized von Mises (GvM) of

Gatto and Jammalamadaka (2007), which is of main interest in this thesis. The GvM distribution can be symmetric or asymmetric, unimodal or bimodal. It is of scientific interest to study whenever circular data are symmetric, or they are distributed around two (or more) modes, rather and an unique one. In this thesis we test these cases via some inferential approaches. The problem of testing the symmetry and the bimodality of the GvM is considered in a Bayesian framework. Some other recent applications of the GvM distributions are: Zhang et al. (2018), in meteorology, Lin and Dong (2019), in oceanography, Astfalck et al. (2018), in offshore engineering, Christmas (2014), in signal processing, and Gatto (2022) in time series analysis.

Moreover, in this thesis we study spherical distributions. These are directional distributions defined on the unit sphere $\mathbb{S}^{p-1}$ of $\mathbb{R}^p$, $p \geq 3$. We focus in particular on simulation techniques on the unit sphere. This topic has already been widely discussed in literature. For example, Kurz and Hanebeck (2015) introduce a sampling algorithm for the vMF distribution based on the inversion method and the results of Wood (1994) and Ulrich (1984). In Jammalamadaka and Terdik (2022) simulation and visualization methods in $\mathbb{R}^3$ for the generalized Fisher-Bingham distributions are discussed. In Kent et al. (2018) an acceptance rejection algorithm to simulate from the Bingham distribution is introduced. The angular central Gaussian distribution is chosen as envelope. This method is then generalized and applied to a wider class of directional distributions. In Hoff (2009) an acceptance rejection algorithm and the Gibbs sampling are proposed to generate from the matrix vMF distribution. In this thesis we report and compare two direct approaches to simulate from the GvMF distribution.

This thesis is organized as follows. In Chapter 2 circular distributions are introduced. The definitions of the vM and GvM distributions are given and some properties and optimality results are reported. In Chapter 3 we introduce spherical distributions. In particular, we focus on the vMF and the GvMF distributions on the unit sphere of $\mathbb{R}^p$. Some optimality results are reported. Moreover, some simulations results from the GvMF distribution are given. Chapter 4 deals with Bayesian tests of symmetry for GvM. Chapter 5 focuses on the study of the bimodality of the GvM distribution. Here the problem is reduced to the analysis of a quartic. A region $\mathcal{W}$ of bimodality is detected and plotted. The inference on bimodality of the GvM distribution is computed in Chapter 6. Chapter 7 reports two different approaches to simulate from the GvMF distribution. Some final remarks are given in Chapter 8. Finally, some supplementary information is reported in two appendices. Appendix A reports the proof of the characterization of symmetry for the GvM distribution. Appendix B reports the calculations of the Lagrange method used to compute the maximum of a function on the sphere.

Chapter 2

General set ups for circular distributions

The main interest of this thesis are circular distributions and random variables. Circular random variables, distribution functions (d.f.) and density functions are introduced. We mention some basic circular distributions. A special mention is given to the vM and the GvM. The second part of this chapter introduces some main properties of the GvM distribution. Here, concepts like the Kullback-Leibler information and entropy are adapted to the circular case and some optimality results are given.

As mentioned, a circular random variable can be represented as a random angle $\theta \in [0, 2\pi)$, or as the corresponding point P on the unit circle of centre O . Here we consider the representation with angle θ, that had to be intended as the angle in radiant that the unit vector $\boldsymbol{OP}$ forms with the x-axis in the positive direction and in the anti clock sense (see Figure 2.1).

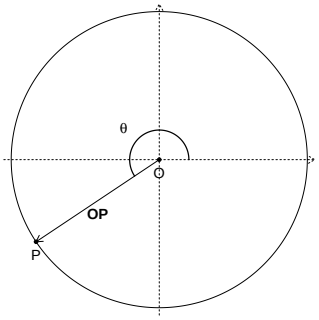


Figure 2.1: Representation of a circular random variable θ.

Let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$ be a sample of size n. Each angle θ_i corresponds to a point P_i on the unit circle of centre O, for $i = 1, \ldots, n$. The mean direction μ of $\boldsymbol{\theta}$ is defined as the direction of the resultant of the unit vectors $\boldsymbol{OP}_1, \ldots, \boldsymbol{OP}_n$. In Cartesian coordinates,

the point P_i, for $i = 1, \ldots, n$ are indicated by $(\cos\theta_i, \sin\theta_i)$, for $i = 1, \ldots, n$. Their centre of gravity is given by $(\bar{C}, \bar{S})$ where

$$\bar{C} = \frac{1}{n}\sum_{i=1}^{n}\cos\theta_i, \text{ and } \bar{S} = \frac{1}{n}\sum_{i=1}^{n}\sin\theta_i.$$

The module of the centre of gravity $\bar{R} = \left(\bar{C}^2 + \bar{S}^2\right)^{1/2}$ is called the mean resultant length. The mean direction μ is the solution of the two equations $\bar{C} = \bar{R}\cos\mu$ and $\bar{S} = \bar{R}\sin\mu$. Some optimality properties of the mean direction can be found in (Mardia, 1972, p.20).

The circular d.f. F of a random angle θ is defined as

$$F(x) = \mathbb{P}\left[0 < \theta \leq x\right], \text{ if } x \in [0, 2\pi). \tag{2.1}$$

To reflect the periodicity of a circular distribution, i.e. that the probability of obtaining a point on the unit circle within any arc of length 2π is equal to 1, we extend the definition of the d.f. given in (2.1) to the whole real line as follows: $F(x+2\pi) - F(x) = 1$ if $x \in \mathbb{R}$. It follows that

1. $F(0) = 0$, and $F(2\pi) = 1$,
2. $\mathbb{P}\left[x_1 < \theta \leq x_2\right] = F(x_2) - F(x_1) = \int_{x_1}^{x_2} \mathrm{d}F(x)$, for $x_1 \leq x_2 \leq x_1 + 2\pi$.

In particular the circular d.f. F is right continuous. In contrast with the d.f. on the real line, it holds that $F(\theta) \to \pm\infty$, as $\theta \to \pm\infty$. When F is absolutely continuous, there exists a circular density f such that

$$F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x)\mathrm{d}x, \ -\infty < x_1 \leq x_2 < +\infty.$$

A function f is a density of an absolutely continuous circular distribution if and only if the following facts hold:

1. $f(\theta) \geq 0$ almost everywhere (a.e.), $-\infty < \theta < +\infty$;
2. $f(\theta + 2\pi) = f(\theta)$ a.e., $-\infty < \theta < +\infty$, i.e. f is periodic of period 2π;
3. $\int_0^{2\pi} f(\theta)\mathrm{d}\theta = 1$.

Most of the models on the circle are obtained from transformations of standard univariate or bivariate models, or are circular analogues of important univariate characterization on the line. These models can be discrete or continuous. Here we introduce some of the most important continuous circular models. For discrete models we refer to (Mardia, 1972, p. 48 - 50).

We start by introducing the wrapped distributions on the circle. These are obtained by "wrapping" distributions on the line around the circumference of unit radius. Let

X be a random variable on the line with d.f. F. The random variable X_w of the corresponding wrapped distribution is given by

$$X_w = X \bmod 2\pi.$$

The d.f. of X_w is given by

$$F_w(\theta) = \sum_{k=-\infty}^{+\infty} \left\{F(\theta + 2\pi k) - F(2\pi k)\right\}, \quad \theta \in [0, 2\pi)$$

Some important properties can be found in (Mardia and Jupp, 2000, p.48). We mention here two cases. When $X \sim \mathcal{N}(\mu, \sigma^2)$, the corresponding wrapped distribution is called the wrapped normal distribution (Mardia and Jupp, 2000, p.49). It is denoted by $\mathrm{WN}(\mu, \rho)$ and it has density function

$$\phi_w(\theta; \mu, \rho) = \frac{1}{\sigma\sqrt{2\pi}} \sum_{k=-\infty}^{+\infty} \exp\left\{-\frac{1}{2}\frac{(\theta - \mu + 2\pi k)^2}{\sigma^2}\right\}, \quad \theta \in [0, 2\pi), \text{ where } \rho = \mathrm{e}^{-\frac{\sigma^2}{2}}.$$

When X follows the Cauchy distribution with density

$$f(x; \mu, a) = \frac{1}{\pi}\frac{a}{a^2 + (x - \mu)^2}, \quad -\infty < \mu < +\infty, \ a > 0,$$

we obtain the wrapped Cauchy distribution (Mardia and Jupp, 2000, p. 51). This is denoted by $WC(\mu, \rho)$ and it has density

$$c_w(\theta; \mu, \rho) = \sum_{k=-\infty}^{+\infty} f(\theta + 2\pi k; \mu, a) = \frac{1}{2\pi}\left(1 + 2\sum_{p=1}^{+\infty} \rho^p \cos p(\theta - \mu)\right), \quad \text{where } \rho = \mathrm{e}^{-a}.$$

Another kind of circular distributions are the offset distributions. Here we introduce the offset normal distribution. Let (X, Y) be a pair of random variables following the the bivariate normal distribution with mean vector $\boldsymbol{\mu} = (\mu_1, \mu_2)$ and covariance matrix $\boldsymbol{\Sigma}$, with density $\phi(x, y; \boldsymbol{\mu}, \boldsymbol{\Sigma})$. Moreover, let ρ be the correlations between X and Y and σ_1^2, σ_2^2 their respective variances. Let $\Phi(x)$ be the d.f. of $\mathcal{N}(0, 1)$ and $\phi(x)$ its density. The offset normal distribution of the circular random variable θ has density (Mardia, 1972, p. 52)

$$\begin{aligned} f(\theta; \mu_1, \mu_2, \sigma_1, \sigma_2, \rho) = \frac{1}{C(\theta)} &\left\{\phi(\mu_1, \mu_2; \mathbf{0}, \boldsymbol{\Sigma}) + aD(\theta)\Phi(D(\theta))\right\} \\ &\cdot \phi\left[a\left\{C(\theta)^{-\frac{1}{2}}\right\}(\mu_1 \sin\theta - \mu_2 \cos\theta)\right], \end{aligned}$$

where $a = \left[\sigma_1\sigma_2\left(1 - \rho^2\right)^{1/2}\right]^{-1}$, $C(\theta) = a^2\left(\sigma_2^2 \cos^2\theta - \rho\sigma_1\sigma_2 \sin 2\theta + \sigma_1^2 \sin^2\theta\right)$, and $D(\theta) = a^2 C(\theta)^{-1/2}\left[\mu_1\sigma_2\left(\sigma_2 \cos\theta - \rho\sigma_1 \sin\theta\right) + \mu_2\sigma_1\left(\sigma_1 \sin\theta - \rho\sigma_2 \cos\theta\right)\right]$.

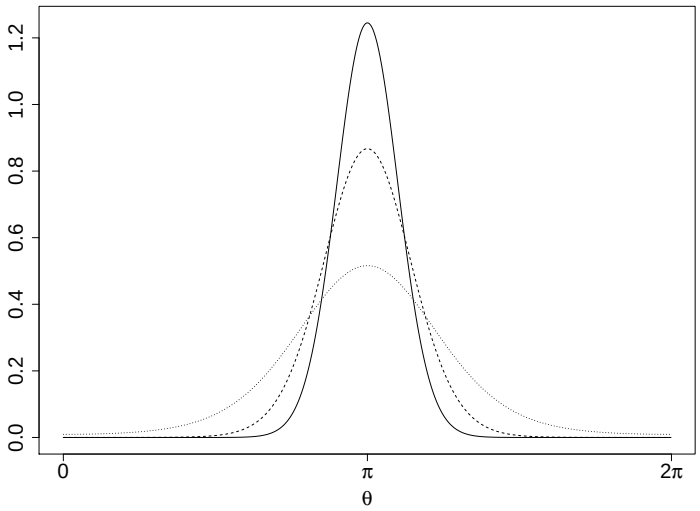


Figure 2.2: Some vM densities: vM$(\pi, 10)$ with continuous line, vM$(\pi, 5)$ with dashed line and vM$(\pi, 2)$ with dotted line.

2.1 The vM and the GvM distribution

In this thesis we are interested in some particular circular distributions that are not obtained by transformation of distributions on the real line. We start with the von Mises (vM) distribution (Mardia, 1972, p.57) The density of the vM distribution is given by

$$f(\theta|\mu,\kappa) = \frac{1}{2\pi I_0(\kappa)} \exp\left\{\kappa \cos\left(\theta-\mu\right)\right\}, \ \forall\theta \in [0, 2\pi), \tag{2.2}$$

where $\mu \in [0, 2\pi)$ is the mean direction, and $\kappa > 0$ is the concentration parameter. The parameter κ indicates the concentration around the mean direction μ. Some vM densities are displayed in Figure 2.2: we can notice how the plot of vM density recalls the "bell-shaped" form of the normal density. This is why the vM distribution is also sometimes called circular normal. The normalizing constant $I_r(z) = (2\pi)^{-1}\int_0^{2\pi} \cos r\theta \, \exp\{z\cos\theta\}\mathrm{d}\theta$, $z \in \mathbb{C}$, is the modified Bessel function I of integer order r (Abramowitz, 1974, p. 376). We denote any circular random variable with this distribution by vM(μ, κ). The generalized von Mises distribution of order k (indicated by GvM$_k$) is considered an extension of the vM distribution, and it was firstly introduced by Maksimov (1967). The density of the GvM$_k$ is given by

$$\begin{aligned} &f(\theta|\mu_1,\ldots,\mu_k,\kappa_1,\ldots,\kappa_k) \\ &\quad = \frac{1}{2\pi G_0^{(k)}(\delta_1,\ldots,\delta_{k-1},\kappa_1,\ldots,\kappa_k)} \cdot \exp\left\{\sum_{j=1}^{k} \kappa_j \cos j\left(\theta-\mu_j\right)\right\}, \end{aligned} \tag{2.3}$$

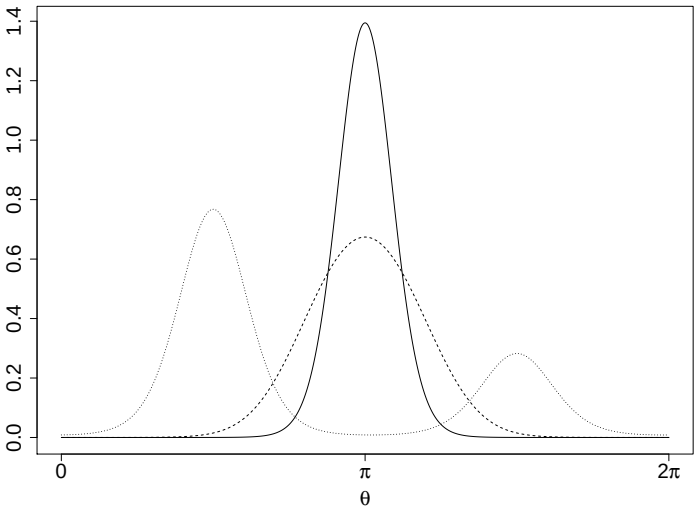


Figure 2.3: Some GvM densities: $\mathrm{GvM}(\pi, 0, 5, 2)$ with continuous line, $\mathrm{GvM}(\pi, \pi/2, 5, 0.6)$ with dashed line and $\mathrm{GvM}(\pi/2, \pi/2, 0.5, 2)$ with dotted line.

with $\kappa_j > 0$, and $\mu_j \in [0, 2\pi/j)$, for $j = 1, \ldots, k$. The normalizing constant is defined as

$$G_0^{(k)}(\delta_1, \ldots, \delta_{k-1}, \kappa_1, \ldots, \kappa_k) = \frac{1}{2\pi} \int_0^{2\pi} \exp\left\{\kappa_1 \cos\omega + \kappa_2 \cos 2(\omega + \delta_1) + \ldots + \kappa_k \cos k(\omega + \delta_{k-1})\right\} \mathrm{d}\omega,$$

and it is a generalization of the modified Bessel function I_0 given above, with $\delta_j = (\mu_1 - \mu_{j+1}) \bmod (2\pi/(j+1))$, for $j = 1, \ldots, k-1$. We denote any circular random variable with this distribution by $\mathrm{GvM}_k(\mu_1, \ldots, \mu_k, \kappa_1, \ldots, \kappa_k)$. In particular, in this thesis we focus on the case when $k = 2$. The density (2.3) reduces to

$$f(\theta|\mu_1, \mu_2, \kappa_1, \kappa_2) = \frac{1}{2\pi G_0^{(2)}(\delta, \kappa_1, \kappa_2)} \exp\left\{\kappa_1 \cos(\theta - \mu_1) + \kappa_2 \cos 2\left(\theta - \mu_2\right)\right\}, \quad (2.4)$$

with $\mu_1 \in [0, 2\pi)$, $\mu_2 \in [0, \pi)$ and $\kappa_1, \kappa_2 > 0$. The normalizing constant $G_0^{(2)}$ is simply denoted by G_0 and it is given by

$$G_0(\delta, \kappa_1, \kappa_2) = \frac{1}{2\pi} \int_0^{2\pi} \exp\left\{\kappa_1 \cos\omega + \kappa_2 \cos 2\left(\omega + \delta\right) \mathrm{d}\omega\right\}, \quad (2.5)$$

with $\delta = (\mu_1 - \mu_2) \bmod \pi$. We indicate any random random variable with this distribution by $\mathrm{GvM}(\mu_1, \mu_2, \kappa_1, \kappa_2)$, omitting the order $k = 2$. Some GvM densities are displayed in Figure 2.3.

We cite here the axial von Mises distribution. This distribution is obtained by taking $\kappa_1 = 0$ in the exponent of (2.4) and by multiplying the density by 2, yielding

$$f(\theta|\mu, \kappa) = \frac{1}{\pi I_0(\kappa)} \exp\{\kappa \cos 2(\theta - \mu)\}, \quad \forall\theta[0, \pi). \quad (2.6)$$

The GvM is more practical than the general GvM$_k$, and it is used, for example, in modeling wind directions (see Gatto and Jammalamadaka (2007)).

We mention that the GvM distribution belongs to the exponential family, after applying a proper re-parametrization. In fact the density (2.4) for a sample $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$

$$f(\boldsymbol{\theta}|\mu_1, \mu_2, \kappa_1, \kappa_2) = \frac{1}{[2\pi \cdot G_0(\delta, \kappa_1, \kappa_2)]^n} \exp\left\{ \kappa_1 \sum_{i=1}^{n} \cos(\theta_i - \mu_1) + \kappa_2 \sum_{i=1}^{n} \cos 2\left(\theta_i - \mu_2\right) \right\}. \tag{2.7}$$

can be written as

$$f\left(\boldsymbol{\theta}|\mu_1, \mu_2, \kappa_1, \kappa_2\right) = \exp\left\{ \boldsymbol{\lambda}^\top \sum_{i=1}^{n} \boldsymbol{T}\left(\theta_i\right) - nK\left(\boldsymbol{\lambda}\right) \right\},$$

where $\boldsymbol{\lambda}^\top = (\kappa_1 \cos\mu_1, \kappa_1 \sin\mu_1, \kappa_2 \cos 2\mu_2, \kappa_2 \sin 2\mu_2) \in \mathbb{R}^4$, and

$$\begin{aligned} \boldsymbol{T}(\theta_i)^\top &= (\cos\theta_i, \sin\theta_i, \cos 2\theta_i, \sin 2\theta_i) \\ K(\boldsymbol{\lambda}) &= \ln(2\pi) + \ln\left[G_0\left(\delta, \kappa_1, \kappa_2\right)\right] = \ln(2\pi) + \ln\left[G_0\left(\delta, \|\boldsymbol{\lambda}_{(1)}\|, \|\boldsymbol{\lambda}_{(2)}\|\right)\right], \end{aligned}$$

with $\boldsymbol{\lambda}^\top = (\boldsymbol{\lambda}_{(1)}^\top, \boldsymbol{\lambda}_{(2)}^\top)$, $\boldsymbol{\lambda}_{(j)}^\top = (\kappa_j \cos\mu_j, \kappa_j \sin\mu_j)$ for $j = 1, 2$, and $\delta = (\arg\{\boldsymbol{\lambda}_{(1)}\} - \arg\{\boldsymbol{\lambda}_{(2)}\}/2) \bmod \pi$.

This re-parametrization can be generalized in the case of GvM$_k$. The density (2.3) for a sample $\boldsymbol{\theta}$ of size n

$$\begin{aligned} &f(\boldsymbol{\theta}|\mu_1, \ldots, \mu_k, \kappa_1, \ldots, \kappa_k) \\ &\quad = \frac{1}{\left[2\pi G_0^{(k)}(\delta_1, \ldots, \delta_{k-1}, \kappa_1, \ldots, \kappa_k)\right]^n} \cdot \exp\left\{ \sum_{j=1}^{k} \kappa_j \sum_{i=1}^{n} \cos j\left(\theta_i - \mu_j\right) \right\}, \end{aligned}$$

can be re-expressed as an exponential family density as follows:

$$f\left(\boldsymbol{\theta}|\mu_1, \ldots, \mu_k, \kappa_1, \ldots, \kappa_k\right) = \exp\left\{ \boldsymbol{\lambda}^\top \sum_{i=1}^{n} \boldsymbol{T}\left(\theta_i\right) - nK\left(\boldsymbol{\lambda}\right) \right\}$$

where $\boldsymbol{\lambda}^\top = (\kappa_1 \cos\mu_1, \kappa_1 \sin\mu_1, \ldots, \kappa_k \cos k\mu_k, \kappa_k \sin k\mu_k) \in \mathbb{R}^{2k}$, and

$$\begin{aligned} \boldsymbol{T}(\theta_i)^\top &= (\cos\theta_i, \sin\theta_i, \cos 2\theta_i, \sin 2\theta_i, \ldots, \cos k\theta_i, \sin k\theta_i), \\ K(\boldsymbol{\lambda}) &= \ln(2\pi) + \ln\left[G_0^{(k)}\left(\delta_1, \ldots, \delta_{k-1}, \|\boldsymbol{\lambda}_{(1)}\|, \ldots, \|\boldsymbol{\lambda}_{(k)}\|\right)\right], \end{aligned}$$

with $\boldsymbol{\lambda}^\top = \left(\boldsymbol{\lambda}_{(1)}^\top, \ldots, \boldsymbol{\lambda}_{(k)}^\top\right)$, $\boldsymbol{\lambda}_{(j)}^\top = (\kappa_j \cos j\mu_j, \kappa_j \sin j\mu_j)$, for $j = 1, \ldots, k$, and $\delta_j = (\arg\{\boldsymbol{\lambda}_{(1)}\} - \arg\{\boldsymbol{\lambda}_{(j+1)}\}/j+1) \bmod \pi$ for $j = 1, \ldots, k-1$. In this form, the GvM$_k$ admits a minimal sufficient and complete statistic; cf. Section 2.1 of Gatto and Jammalamadaka (2007).

2.2 Main results

In this section we introduce some important quantities of information theory adapted to the circular case. Moreover we report some optimality results for the GvM_k.

The first quantity we introduce is due to Kullback and Leibler (1951).

Definition 2.2.1 (Circular Kullback-Leibler information)**.** Let f and g be two circular densities. The circular Kullback-Leibler (KL) information is given by

$$I(f|g) = \int_0^{2\pi} \log \frac{f(\theta)}{g(\theta)} f(\theta) \mathrm{d}\theta. \tag{2.8}$$

We assume that $0 \log 0 = 0$ and that the domain of f is contained in the domain of g, otherwise $I(f|g) = \infty$.

The KL information is the mean logarithmic likelihood ratio, or mean information per observation of f for discriminating in favour of f against g. The continuous version of Gibbs' inequality, namely

$$\int_0^{2\pi} \log f(\theta) f(\theta) \mathrm{d}\theta \geq \int_0^{2\pi} \log g(\theta) f(\theta) \mathrm{d}\theta,$$

implies that $I(f|g) \geq 0$, with equality if and only if $f = g$ (Johnson, 2004, p.4).

Another important quantity we introduce is the circular entropy (Mardia, 1972, p.6).

Definition 2.2.2 (Circular entropy)**.** Let f be a circular density. The circular entropy of f is defined as

$$H(f) = -\int_0^{2\pi} \log f(\theta) f(\theta) \mathrm{d}\theta, \tag{2.9}$$

where we assume again that $0 \log 0 = 0$.

$H(f)$ measures the uncertainty inherent in f. Equivalently, it measures the expected amount of information gained on obtaining an observation from f, based on the fact that rarer events are more informative. Moreover, we can say that the entropy $H(f)$ measures the closeness of f to the uniform density (Mardia and Jupp, 2000). This follows from the equality $H(f) = \log(2\pi) - I(f|g)$, where g is the uniform density on $[0, 2\pi]$.

In addition to the GvM_k, finite mixtures of vM distribution have been studied in literature. McVinish and Mengersen (2008) examines the use of Bayesian mixtures of triangular distributions in the context of semiparametric analysis of circular data. Moreover, Spurr and Koutbeiy (1991) proposed flexible models for circular data based on finite mixtures of vM distributions, for a comparison of different approaches to estimate such mixtures. The mixture of two vM or other distributions provides an alternative bimodal or asymmetric model to the GvM. Nevertheless, it lacks of some important theoretical properties. In fact, the mixture model does not belong to the exponential class. Moreover, while the likelihood function of GvM_k is bounded, the likelihood of a

mixture is unbounded. This can be shown as follows: Let us consider the mixture of $\mathrm{vM}(\mu_1, \kappa_1)$ and of $\mathrm{vM}(\mu_2, \kappa_2)$. As $\kappa_1 \to +\infty$, the likelihood when μ_1 is equal to any one of the sample values tends to infinity: in fact, $I_0(\kappa_1) \sim (2\pi\kappa_1)^{(-1/2)} \mathrm{e}^{\kappa_1}$, as $\kappa_1 \to +\infty$ (see Abramowitz (1974), 9.7.1 at p.377). A bounded likelihood is required in Cox and Hinkley (1979) in the consistency of the MLE, and Cheng and Traylor (1995) consider the general problem of dealing with unbounded likelihood.

In what follows, we study some properties that the GvM_k posses, but that do not hold for the mixture model. We report some optimality results with respect to (w.r.t.) the KL information (2.8) and to the entropy (2.9). Let us firstly introduce a concept that will be useful in the section.

Definition 2.2.3. Let θ be a circular random variable on $[0, 2\pi)$. The rth trigonometric moment condition is given by

$$TM_r : \int_0^{2\pi} \mathrm{e}^{ir\theta} g(\theta)\mathrm{d}\theta = \varphi_r \Leftrightarrow \int_0^{2\pi} \cos r\theta g(\theta)\mathrm{d}\theta + i\int_0^{2\pi} \sin r\theta g(\theta)\mathrm{d}\theta = \gamma_r + i\sigma_r,$$

for some φ_r such that $|\varphi_r| \leq 1$ and for some $\gamma_r, \sigma_r \in [-1, 1]$, $r = 1, 2, \ldots$.

Next theorem is an important result on the KL information, from which we derive some important results on the optimality of the GvM_k distribution.

Theorem 2.2.4.

(a) The circular density f, which satisfies TM_r for $r = i_1, \ldots, i_k$ with $1 \leq i_1 < \ldots < i_k < +\infty$, and which minimizes the KL information $I(f|f_0)$ for some circular density f_0, is given by the exponential tilt of f_0, namely

$$\begin{aligned} f(\theta) = &\left[G_0^{(i_1,\ldots,i_k)}(\delta_1^\dagger, \ldots, \delta_{k-1}^\dagger, \kappa_{i_1}, \ldots, \kappa_{i_k}; f_0)\right]^{-1} \\ &\cdot \exp\left\{\kappa_{i_1} \cos i_1 (\theta - \mu_{i_1}) + \ldots + \kappa_{i_k} \cos i_k(\theta - \mu_{i_k})\right\} f_0(\theta), \end{aligned} \tag{2.10}$$

where the normalizing constant is given by

$$\begin{aligned} &G_0^{(i_1,\ldots,i_k)}(\delta_1^\dagger, \ldots, \delta_{k-1}^\dagger, \kappa_{i_1}, \ldots, \kappa_{i_k}; f_0) \\ &= \int_0^{2\pi} \exp\left\{\kappa_{i_1} \cos i_1\theta + \kappa_{i_2} \cos i_2 \left(\theta + \delta_1^\dagger\right) + \ldots + \kappa_{i_k} \cos i_k \left(\theta + \delta_{k-1}^\dagger\right)\right\} f_0(\theta)\mathrm{d}\theta, \end{aligned}$$

with $\delta_j^\dagger = \left(\mu_{i_1} - \mu_{i_{j+1}}\right) \bmod (2\pi/i_{j+1})$, for $j = 1, \ldots, k-1$, and $\kappa_{i_1}, \ldots, \kappa_{i_k} > 0$.

(b) For any circular density f satisfying TM_r for $r = i_1, \ldots, i_k$, it holds that

$$\begin{aligned} I(f|f_0) \geq &- \log G_0^{(i_1,\ldots,i_k)}(\delta_1^\dagger, \ldots, \delta_{k-1}^\dagger, \kappa_{i_1}, \ldots, \kappa_{i_k}; f_0) \\ &+ \sum_{r=1}^{k} \kappa_{i_r} \left(\gamma_{i_r} \cos i_r\mu_{i_r} + \sigma_{i_r} \sin i_r\mu_{i_r}\right), \end{aligned}$$

with equality if and only if f satisfies (2.10).

Moreover, let us define:

$$G_r^{(i_1,\ldots,i_k)}\left(\delta_1^\dagger,\ldots,\delta_{k-1}^\dagger,\kappa_{i_1},\ldots,\kappa_{i_k};f_0\right)=\int_0^{2\pi}\cos r\theta\exp\left\{\kappa_{i_1}\cos i_1\theta+\kappa_{i_2}\cos i_2\left(\theta+\delta_1^\dagger\right)\right.$$
$$\left.+\cdots+\kappa_{i_k}\cos i_k\left(\theta+\delta_{k-1}^\dagger\right)\right\}f_0(\theta)\mathrm{d}\theta$$

$$H_r^{(i_1,\ldots,i_k)}\left(\delta_1^\dagger,\ldots,\delta_{k-1}^\dagger,\kappa_{i_1},\ldots,\kappa_{i_k};f_0\right)=\int_0^{2\pi}\sin r\theta\exp\left\{\kappa_{i_1}\cos i_1\theta+\kappa_{i_2}\cos i_2\left(\theta+\delta_1^\dagger\right)\right.$$
$$\left.+\cdots+\kappa_{i_k}\cos i_k\left(\theta+\delta_{k-1}^\dagger\right)\right\}f_0(\theta)\mathrm{d}\theta,$$

$$A_r^{(i_1,\ldots,i_k)}\left(\delta_1^\dagger,\ldots,\delta_{k-1}^\dagger,\kappa_{i_1},\ldots,\kappa_{i_k};f_0\right)=\frac{G_r^{(i_1,\ldots,i_k)}\left(\delta_1^\dagger,\ldots,\delta_{k-1}^\dagger,\kappa_{i_1},\ldots,\kappa_{i_k};f_0\right)}{G_0^{(i_1,\ldots,i_k)}\left(\delta_1^\dagger,\ldots,\delta_{k-1}^\dagger,\kappa_{i_1},\ldots,\kappa_{i_k};f_0\right)},$$

and

$$B_r^{(i_1,\ldots,i_k)}\left(\delta_1^\dagger,\ldots,\delta_{k-1}^\dagger,\kappa_{i_1},\ldots,\kappa_{i_k};f_0\right)=\frac{H_r^{(i_1,\ldots,i_k)}\left(\delta_1^\dagger,\ldots,\delta_{k-1}^\dagger,\kappa_{i_1},\ldots,\kappa_{i_k};f_0\right)}{G_0^{(i_1,\ldots,i_k)}\left(\delta_1^\dagger,\ldots,\delta_{k-1}^\dagger,\kappa_{i_1},\ldots,\kappa_{i_k};f_0\right)},$$

for $r=i_1,\ldots,i_k$, where $\delta_1^\dagger,\ldots,\delta_{k-1}^\dagger$ and $\kappa_{i_1},\ldots,\kappa_{i_k}>0$ as defined in (a). Then we have that $\mu_{i_1},\mu_{i_2}=(\mu_{i_1}-\delta_1^\dagger)\bmod(2\pi/i_2),\ldots,\mu_{i_k}=(\mu_{i_1}-\delta_{k-1}^\dagger)\bmod(2\pi/i_k)$ and $\kappa_{i_1},\ldots,\kappa_{i_k}$ are the simultaneous solutions of

$$\begin{pmatrix}\gamma_r\\ \sigma_r\end{pmatrix}=R\left(r\mu_{i_1}\right)\begin{pmatrix}A_r^{(i_1,\ldots,i_k)}\left(\delta_1^\dagger,\ldots,\delta_{k-1}^\dagger,\kappa_{i_1},\ldots,\kappa_{i_k};f_0\right)\\ B_r^{(i_1,\ldots,i_k)}\left(\delta_1^\dagger,\ldots,\delta_{k-1}^\dagger,\kappa_{i_1},\ldots,\kappa_{i_k};f_0\right)\end{pmatrix},\tag{2.11}$$

where

$$R(\alpha)=\begin{pmatrix}\cos\alpha & -\sin\alpha\\ \sin\alpha & \cos\alpha\end{pmatrix},\tag{2.12}$$

is the rotation matrix of angle $\alpha=r\mu_{i_1}$ for $r=i_1,\ldots,i_k$.

Proof. The general idea for the proof is due to Kullback (1954), and it can be found in Gatto (2009). □

Let us assume Theorem 2.2.4 with $i_1=1,\ldots,i_k=k$ and $f_0=1/2\pi$. In this case, for simplicity of notation in what follows the last argument of $G_0^{(i_1,\ldots,i_k)}$, $G_r^{(i_1,\ldots,i_k)}$, $H_r^{(i_1,\ldots,i_k)}$, $A_r^{(i_1,\ldots,i_k)}$ and $B_r^{(i_1,\ldots,i_k)}$ is omitted. Moreover, the upper right sequential index $(i_1,\ldots,i_k)$ that appears in the constants is simply the upper right index (k). We also have that $\delta_j^\dagger=\delta_j$, for $j=1,\ldots,k-1$. The following corollary is an important consequence of Theorem 2.2.4, and it states the optimality of the GvM_k w.r.t. the circular entropy (2.9).

Corollary 2.2.5.

(a) For all $k \in \{1, 2, \ldots\}$, the circular density maximizing the entropy under TM_r, $r = 1, \ldots, k$, is the $GvM_k(\mu_1, \ldots, \mu_k, \kappa_1, \ldots, \kappa_k)$.

(b) If f is a circular density satisfying TM_r, $r = 1, \ldots, k$, then

$$H(f) \leq \log(2\pi) + \log G_0^{(k)}(\delta_1, \ldots, \delta_{k-1}, \kappa_1, \ldots, \kappa_k) - \sum_{r=1}^{k} \kappa_r \left(\gamma_r \cos r\mu_r + \sigma_r \sin r\mu_r\right),$$

with equality if and only if f is the $GvM_k(\mu_1, \ldots, \mu_k, \kappa_1, \ldots, \kappa_k)$ density. The parameters above are the solutions of

$$\begin{pmatrix} \gamma_r \\ \sigma_r \end{pmatrix} = R(r\mu_1) \begin{pmatrix} A_r^{(k)}(\delta_1, \ldots, \delta_{k-1}, \kappa_1, \ldots, \kappa_k) \\ B_r^{(k)}(\delta_1, \ldots, \delta_{k-1}, \kappa_1, \ldots, \kappa_k) \end{pmatrix}, \tag{2.13}$$

where $R(r\mu_1)$ is the rotation matrix defined in (2.12) with $\alpha = r\mu_1$, for $r = 1, 2, \ldots$.

Proof.

(a) Let $f_0(\theta) = 1/(2\pi)$. It follows from (2.8) and (2.9) that $I(f|f_0) = \log(2\pi) - H(f)$. Thus, maximizing $H(f)$ under TM_r, for $r = 1, \ldots, k$ is equivalent to minimizing $I(f|f_0)$ under the same constraint. Applying Theorem 2.2.4 (a), we conclude that the minimum is obtained when f is the GvM_k density.

(b) This follows from Theorem 2.2.4 (b). The values of the parameters that satisfy (2.13) are given in the proof of Theorem 2.2.4, which can be found in Gatto (2009).

□

Part (a) of Corollary 2.2.5 is a generalization of the optimality result for the vM distribution given in Mardia (1972, p.65-66). Entropy is a principle for choosing a distribution on the basis of some partial knowledge that could be based on some prior information. In our case, the partial prior information is given by the first k trigonometric moments. Based on this principle, one should always choose distributions associated to maximal entropy, within the distributions satisfying such partial prior information. Corollary 2.2.5 states that the distribution maximizing the entropy is the GvM_k. This is widely used in Bayesian statistics, when one wants to choose the most noninformative prior which satisfies some known partial prior information. See Gatto and Jammalamadaka (2007) and Gatto (2009).

Part (b) of the corollary expresses the entropy of the $\mathrm{GvM}_k(\mu_1, \ldots, \mu_k, \kappa_1, \ldots, \kappa_k)$ distribution in terms of $A_r^{(k)}(\delta_1, \ldots, \delta_{k-1}, \kappa_1, \ldots, \kappa_k)$ and $B_r^{(k)}(\delta_1, \ldots, \delta_{k-1}, \kappa_1, \ldots, \kappa_k)$ defined in Theorem 2.2.4, with $f_0(\theta) = 1/(2\pi)$ and $\delta_j = (\mu_1 - \mu_{j+1}) \bmod (2\pi/(j+1))$, for $j = 1, \ldots, k-1$ and $r = 1, \ldots, k$. These quantities can be evaluated by numerical

integration. With some simplifications and with $\gamma_r,\ \sigma_r$ solving (2.13), we get that the entropy of the $\mathrm{GvM}_k(\mu_1,\ldots,\mu_k,\kappa_1,\ldots,\kappa_k)$ distribution is given by (Gatto, 2009):

$$\begin{aligned}H(f) =& \log(2\pi) + \log G_0^{(k)}\left(\delta_1,\ldots,\delta_{k-1},\kappa_1,\ldots,\kappa_k\right) - \kappa_1 A_1^{(k)}\left(\delta_1,\ldots,\delta_{k-1},\kappa_1,\ldots,\kappa_k\right)\\ &- \sum_{r=2}^{k}\kappa_r\left[\cos r\delta_{r-1}A_r^{(k)}\left(\delta_1,\ldots,\delta_{k-1},\kappa_1,\ldots,\kappa_k\right)\right.\\ &\left.+\sin r\delta_{r-1}B_r^{(k)}\left(\delta_1,\ldots,\delta_{k-1},\kappa_1,\ldots,\kappa_k\right)\right],\end{aligned}$$

where the summation from 2 to k simply vanishes when $k=1$.

We conclude this section with the case when the k-th trigonometric moment restrictions are the successive integers after j, and where f_0 is a GvM_j density. Theorem 2.2.4 with $i_1 = j+1,\ldots,i_k = j+k$ gives the following corollary.

Corollary 2.2.6.

(a) Let f_0 be the $GvM_j(\mu_1,\ldots,\mu_j,\kappa_1,\ldots,\kappa_j)$ density. The closest circular density f to f_0 in the KL sense under the condition TM_r, for $r = j+1,\ldots,j+k$, is the $GvM_{j+k}(\mu_1,\ldots,\mu_{j+k},\kappa_1,\ldots,\kappa_{j+k})$ density.

(b) For any f satisfying TM_r, for $r = j+1,\ldots,j+k$, the KL information satisfies

$$\begin{aligned}I\left(f \mid f_0\right) \geq& -\log G_0^{(j+1,\ldots,j+k)}\left(\delta_1^{\dagger},\ldots,\delta_{k-1}^{\dagger},\kappa_{j+1},\ldots,\kappa_{j+k};f_0\right)\\ &+\sum_{r=j+1}^{j+k}\kappa_r\left(\gamma_r\cos r\mu_r+\sigma_r\sin r\mu_r\right),\end{aligned}$$

where $\delta_l^{\dagger} = (\mu_{j+1}-\mu_{j+l+1}) \bmod (2\pi/(j+l+1))$, for $l = 1,\ldots,k-1$, with equality if and only if f is the $GvM_{j+k}(\mu_1,\ldots,\mu_{j+k},\kappa_1,\ldots,\kappa_{j+k})$ density. The parameters $\mu_{j+1},\ldots,\mu_{j+k}$ and $\kappa_{j+1},\ldots,\kappa_{j+k}$ are determined by TM_r for $r = j+1,\ldots,j+k$, namely by (2.11) with $i_1 = j+1,\ldots,i_k = j+k$.

Chapter 3

Spherical distributions

So far, we focused on circular data, i.e. directional data in $\mathbb{R}^2$. This chapter introduces directional data on the unit sphere $\mathbb{S}^{p-1} = \{\boldsymbol{x} \in \mathbb{R}^p | \langle \boldsymbol{x}, \boldsymbol{x} \rangle = 1\}$ that are of main interest in Chapter 7. Some concepts for circular distributions introduced in Chapter 2 can be extend on $\mathbb{R}^p$. Descriptive measures like the mean direction can be defined in a similar way. Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathbb{S}^{p-1}$. Their location can be described by the sample mean (Mardia and Jupp, 2000, p.163),

$$\bar{\boldsymbol{x}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i.$$

The vector $\bar{\boldsymbol{x}}$ can be expressed as

$$\bar{\boldsymbol{x}} = \bar{R} \bar{\boldsymbol{x}}_0$$

where $\boldsymbol{x}_0 \in \mathbb{S}^{p-1}$ and $\bar{R} \geq 0$. The value

$$\bar{R} = \|\bar{\boldsymbol{x}}\|,$$

is called mean resultant length, while

$$\bar{\boldsymbol{x}}_0 = \frac{\bar{\boldsymbol{x}}}{\|\bar{\boldsymbol{x}}\|},$$

is the mean direction.

A spherical random variable is represented as a point on the unit sphere $\mathbb{S}^{p-1}$. The case $p = 3$ deserves a special mention. Here, the representation can be done using spherical or Cartesian coordinates. Any unit vector $\boldsymbol{x} \in \mathbb{S}^2$ expressed in Cartesian coordinates (x_1, x_2, x_3) can be written in spherical coordinates as $\boldsymbol{x} = (1, \theta_1, \theta_2)$. The angles θ_1, θ_2 are called polar and azimuth angle, respectively. They are defined as follows.

$$\begin{cases} \theta_1 = \arccos(x_3) \in [0, \pi], \\ \theta_2 = \operatorname{arctan2}(x_2, x_1) \in (-\pi, \pi], \end{cases} \tag{3.1}$$

where arctan2(x_2, x_1), called two-argument or quadrant specific arc tangent, is defined as

$$\theta_2 = \begin{cases} \arctan\left(\frac{x_2}{x_1}\right), & \text{if } x_1 > 0, \\ \arctan\left(\frac{x_2}{x_1}\right) + \pi, & \text{if } x_1 < 0 \text{ and } x_2 \geq 0, \\ \arctan\left(\frac{x_2}{x_1}\right) - \pi, & \text{if } x_1 < 0 \text{ and } x_2 < 0, \\ +\frac{\pi}{2}, & \text{if } x_1 = 0 \text{ and } x_2 > 0, \\ -\frac{\pi}{2}, & \text{if } x_1 = 0 \text{ and } x_2 < 0, \\ \text{undefined}, & \text{if } x_1 = 0 \text{ and } x_2 = 0, \end{cases}$$

where arctan has range of usual principal value $(-\pi/2, \pi/2)$. On the other hand, it is possible to retrieve the Cartesian coordinates from the spherical coordinates using

$$\begin{cases} x_1 = \cos\theta_2 \sin\theta_1, \\ x_2 = \sin\theta_2 \sin\theta_1, \\ x_3 = \cos\theta_1 \end{cases}$$

In this thesis we manly choose the Cartesian representation. An example of representation in the Cartesian coordinate system $(X_1, X_2, X_3) = (X, Y, Z)$ can be found in Figure 3.1.

In this chapter, we introduce two important spherical distributions: the von Mises Fisher (vMF) and the generalized von Mises Fisher (GvMF). As in Chapter 2, we report some important optimality and simulation results regarding the GvMF distribution. We refer to Chapter 9 of Mardia and Jupp (2000) for some other spherical distributions.

3.1 The vMF and the GvMF

In Section 2.1 we introduced the vM and the GvM distribution. In this section we define the natural extensions of these two distributions on the unit sphere $\mathbb{S}^{p-1}$ of the generic space $\mathbb{R}^p$, for $p \geq 3$.

The density of the vMF distribution (Mardia and Jupp, 2000) is defined as

$$g(\boldsymbol{x}|\boldsymbol{\mu}, \kappa) = a_p(\kappa) \exp\left\{\kappa\left(\boldsymbol{\mu}^\top \boldsymbol{x}\right)\right\}, \tag{3.2}$$

where $\boldsymbol{\mu} \in \mathbb{S}^{p-1}$ and $\kappa > 0$, and the normalizing factor $a_p(\kappa)$ w.r.t. the spherical uniform (or isotropic) distribution U is given by

$$a_p(\kappa) = \left(\frac{\kappa}{2}\right)^{p/2-1} \frac{1}{\Gamma\left(p/2\right) I_{p/2-1}(\kappa)}, \tag{3.3}$$

where $I_\nu(z)$, $z \in \mathbb{C}$, is the modified Bessel function I of order ν, such that $\Re\nu > -1/2$; see e.g. 9.6.18 at p. 376 of Abramowitz (1974). (Note that the normalizing factor is

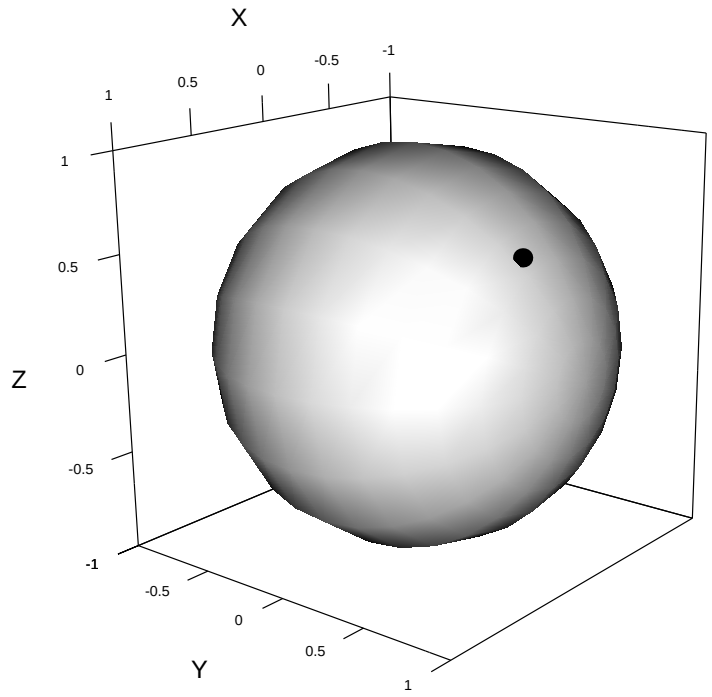


Figure 3.1: Representation of the spherical random variable $\boldsymbol{X}$ in the unit sphere $\mathbb{S}^2$ of $\mathbb{R}^3$ (bold point) in the coordinate system (X, Y, Z).

sometimes given for the spherical Lebesgue measure.) The mean direction of (3.2) is $\boldsymbol{\mu}$ and its concentration is κ. Any random direction with this distribution is denoted by $\text{vMF}(\boldsymbol{\mu}, \kappa)$.

The density of the generalized von Mises Fisher distribution of generic order k (Gatto, 2011) for a vector $\boldsymbol{x} \in \mathbb{S}^{p-1}$ w.r.t the uniform distribution U on $\mathbb{S}^{p-1}$ is given by

$$f(\boldsymbol{x}|\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k; \kappa_1, \ldots, \kappa_k; i_1, \ldots, i_k) \propto \exp\left\{\sum_{i=1}^{k} \kappa_i \left(\boldsymbol{\mu}_i^\top \boldsymbol{x}\right)^{i_k}\right\}. \tag{3.4}$$

where $\boldsymbol{\mu}_i \in \mathbb{S}^{p-1}$ and $\kappa_i > 0$ for $i = 1, \ldots, r$, and $i_1 \leq \ldots \leq i_k \in \{1, 2, \ldots\}$. Moreover, to avoid over identification, we consider that $1, \left(\boldsymbol{\mu}_1^\top \boldsymbol{x}\right)^{i_1}, \ldots, \left(\boldsymbol{\mu}_k^\top \boldsymbol{x}\right)^{i_k}$ are linearly independent functions of $\boldsymbol{x}$. We indicate every random variable with such distribution by $\text{GvMF}_k(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k; \kappa_1, \ldots, \kappa_k; i_1, \ldots, i_k)$. In particular, we notice that for $k = 1$ and $i_1 = 1$ (3.4) reduce to (3.2). Moreover, for $p = 2$ the GvMF admits a GvM reparametrization. In this thesis we focus on the GvMF_k with $k = 2$. We simply denote this by GvMF.

3.2 Main Results

In this section we present some important theoretic results for the GvMF w.r.t. to the directional KL information and the directional entropy. These two were defined on the circle in (2.8) and in (2.9), respectively. Their versions on $\mathbb{S}^{p-1}$ are given below.

Definition 3.2.1 (Directional KL information)**.** Let P and Q two probability measures on $\mathbb{S}^{p-1}$ w.r.t. the uniform distribution U. Moreover we assume that P is absolutely continuous w.r.t Q ($P \ll Q$). The KL information is given by

$$I(P|Q) = \int_{\mathbb{S}^{p-1}} \log \frac{dP}{dQ} \mathrm{d}P = \int_{\mathbb{S}^{p-1}} \log \frac{f(\boldsymbol{x})}{g(\boldsymbol{x})} f(\boldsymbol{x}) \mathrm{d}U(\boldsymbol{x}), \tag{3.5}$$

where f and g are the (directional) densities of P and Q respectively w.r.t U. As in the circular case, we assume that $0 \log 0 = 0$.

As in Section 2.2, we notice that $I(P|Q) \geq 0$ for all P and Q such that $P \ll Q$. The equality holds when $P = Q$.

In the same way we can extend the definition of the entropy given in Section 2.2 to the directional case

Definition 3.2.2 (Directional entropy)**.** The entropy of a directional probability measure P on $\mathbb{S}^{p-1}$ w.r.t the uniform distribution U is defined as

$$H(P) = -\int_{\mathbb{S}^{p-1}} \log \frac{dP}{dU} \mathrm{d}P = -\int_{\mathbb{S}^{p-1}} \log f(\boldsymbol{x}) f(\boldsymbol{x}) \mathrm{d}U(\boldsymbol{x}), \tag{3.6}$$

where, again, $0 \log 0 = 0$.

Now we give some optimality results for directional distributions on $\mathbb{R}^p$ w.r.t the directional version of Kullback-Leibler information given and entropy given in (3.5) and (3.6) respectively. Analogue to the GvM distribution in Section 2.1 the GvMF or other directional distributions having a GvMF parts, turns out to be optimal. Moreover, in Bayesian statistics these kind of distributions are optimal solutions of a constrained prior selection problem. The proofs of these results can be found in Gatto (2011).

Firstly, we introduce the directional version of the rth trigonometric moment condition TM_r given in Section 2.2.

Definition 3.2.3. We denote by DM_r the rth trigonometric moment condition on the density f as,

$$\mathrm{DM}_r : \qquad \int_{\mathbb{S}^{p-1}} \left(\boldsymbol{\mu}_r^\top \boldsymbol{x}\right)^{i_r} f(\boldsymbol{x}) \mathrm{d}U(\boldsymbol{x}) = \alpha_r$$

where $i_1 \leq, \ldots \leq i_k \in \{1, 2, \ldots\}$, $\boldsymbol{\mu}_r \in \mathbb{S}^{p-1}$ and $\alpha_r \in \mathbb{R}$ for $r = 1, \ldots, k$.

Now we introduce some important optimality results. In what follows we consider probability measure on $\mathbb{S}^{p-1}$ and w.r.t. the uniform distribution U.

Theorem 3.2.4.

(a) The directional density f which satisfies DM_r, for $r = 1, \ldots, k$ and which minimizes the KL information $I(f|f_0)$ w.r.t. a given directional density f_0, is proportional to

$$\exp\left\{\sum_{r=1}^{k} \kappa_r \left(\boldsymbol{\mu}_r^\top \boldsymbol{x}\right)^{i_r}\right\} f_0(\boldsymbol{x}), \tag{3.7}$$

for $\boldsymbol{x} \in \mathbb{S}^{p-1}$. The parameters $\kappa_1, \ldots, \kappa_k$ are the solutions in $\nu_1, \ldots, \nu_k$ of

$$\frac{\partial}{\partial \nu_r} K\left(\nu_1, \ldots, \nu_k; \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k; i_1, \ldots, i_k; f_0\right) = \alpha_r, \ r = 1, \ldots, k,$$

with $K = \log M$ and

$$M\left(\nu_1, \ldots, \nu_k; \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k; i_1, \ldots, i_k; f_0\right) = \int_{\mathbb{S}^{p-1}} \exp\left\{\sum_{r=1}^{k} \nu_r \left(\boldsymbol{\mu}_r^\top \boldsymbol{x}\right)^{i_r}\right\} f_0(\boldsymbol{x}) \mathrm{d}U(\boldsymbol{x}).$$

(b) For any directional density f satisfying DM_r, for $r = 1, \ldots, k$ we have the following lower bound on the KL information,

$$I\left(f \mid f_0\right) \geq -K\left(\kappa_1, \ldots, \kappa_k; \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k; i_1, \ldots, i_k; f_0\right) + \sum_{j=1}^{k} \kappa_j \alpha_j,$$

with equality iff (3.7) is proportional to f.

Proof. For the proofs of (a) and (b) we refer to Gatto (2011). □

Corollary 3.2.5.

(a) For any $k \in \{1, 2, \cdots\}$, the directional density f that maximizes the entropy $H(f)$ under DM_r for $r = 1, \ldots, k$ is the $GvMF_k(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k; \kappa_1, \ldots, \kappa_k; i_1, \ldots, i_k)$ density. The parameters $\kappa_1, \ldots, \kappa_k$ are the solutions in $\nu_1, \ldots, \nu_k$ of

$$\frac{\partial}{\partial \nu_r} K\left(\nu_1, \ldots, \nu_k; \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k; i_1, \ldots, i_k; 1\right) = \alpha_r,$$

for $r = 1, \ldots, k$.

(b) If f is a directional density satisfying DM_r, for $r = 1, \ldots, k$, then

$$H(f) \leq K\left(\kappa_1, \ldots, \kappa_k; \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k; i_1, \ldots, i_k; 1\right) - \sum_{r=1}^{k} \kappa_r \alpha_r,$$

with equality iff f is the $GvMF_k(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k; \kappa_1, \ldots, \kappa_k; i_1, \ldots, i_k)$ density.

Proof. (a) This is the directional version of the proof (a) of Corollary 2.2.5. The complete proof can be found in Gatto (2011).

(b) This follows from Theorem 3.2.4.

□

We conclude this section with one last property. Again all densities are given w.r.t. the uniform distribution U on $\mathbb{S}^{p-1}$. In the next corollary the restriction on the ordering $i_1 \leq \ldots \leq i_j \leq i_{j+1} \leq \ldots \leq i_{j+k}$ is relaxed.

Corollary 3.2.6.

(a) Denote by f_0 the GvMF$_j$ $(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_j; \kappa_1, \ldots, \kappa_j; i_1, \ldots, i_j)$ density. Then the closest density f to f_0, in the sense of minimizing $I(f \mid f_0)$ which satisfies DM$_r$, $r = j+1, \ldots, j+k$, is the GvMF$_{j+k}$ $(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{j+k}; \kappa_1, \ldots, \kappa_{j+k}; i_1, \ldots, i_{j+k})$ density. The parameters $\kappa_{j+1}, \ldots, \kappa_{j+k}$ are solutions in $\nu_{j+1}, \ldots, \nu_{j+k}$ of

$$\frac{\partial}{\partial \nu_r} K(\nu_{j+1}, \ldots, \nu_{j+k}; \boldsymbol{\mu}_{j+1}, \ldots, \boldsymbol{\mu}_{j+k}; i_{j+1}, \ldots, i_{j+k}; f_0) = \alpha_r,$$

for $r = j+1, \ldots, j+k$.

(b) For any density f satisfying DM$_r$, for $r = 1, \ldots, k$

$$I(f \mid f_0) \geq -K(\kappa_{j+1}, \ldots, \kappa_{j+k}; \boldsymbol{\mu}_{j+1}, \ldots, \boldsymbol{\mu}_{j+k}; i_{j+1}, \ldots, i_{j+k}; f_0) + \sum_{r=j+1}^{j+k} \kappa_r \alpha_r,$$

with equality iff f is the GvMF$_{j+k}$ density given in (a).

3.3 Other properties

In this section we provide some other properties of the GvMF$_k$. Moreover we introduce some simulations results concerning the GvMF$_k$ distribution and its relation with some other linear or circular distributions.

The first property we mention is that the GvMF$_k$ distribution is invariant under orthogonal transformations: Let $\boldsymbol{X} \sim \text{GvMF}_k(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k; \kappa_1, \ldots, \kappa_k; i_1, \ldots, i_k)$ on $\mathbb{S}^{p-1}$ and let A be a $p \times p$ orthogonal matrix (i.e. $A^\top A = I_p$). Then we have that $\boldsymbol{Y} = A\boldsymbol{X} \sim \text{GvMF}_k(A\boldsymbol{\mu}_1, \ldots, A\boldsymbol{\mu}_k; \kappa_1, \ldots, \kappa_k; i_1, \ldots, i_k)$. This is equivalent to say that the class of GvMF$_k$ distribution, with fixed $\kappa_1, \ldots, \kappa_k$ and $i_1, \ldots, i_k$, is invariant under the group of orthogonal transformations on $\mathbb{R}^p$.

The parameters $i_1, \ldots, i_k$ influence the shape of the GvMF$_k$ density as follows. If i_r is even, then the clustering around both directions $\boldsymbol{\mu}_r$ and $-\boldsymbol{\mu}_r$ increases as κ_r increases, for $r = 1, \ldots, k$. If $i_1, \ldots, i_k$ are even then $(\mathrm{d}P)/(\mathrm{d}Q)(\boldsymbol{x}) = (\mathrm{d}P)/(\mathrm{d}Q)(-\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{S}^{p-1}$. This means that the GvMF$_k$ density is invariant under the transformation $\boldsymbol{x} \mapsto -\boldsymbol{x}$, for $\boldsymbol{x} \in \mathbb{S}^{p-1}$. In general, the larger the values of $\kappa_1, \ldots, \kappa_k$, the larger the clustering around $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k$. If $\boldsymbol{\mu}_1 = \ldots = \boldsymbol{\mu}_k$ and $\kappa_1, \ldots, \kappa_k > 0$, then the density has the maximum at $\boldsymbol{\mu}_1$ and the minimum at $-\boldsymbol{\mu}_1$.

Here we provide some results on the simulation of the GvMF_k. In particular, we cite Result 3.3.1 that relates some GvMF_k distributions with the GvM, and Result 3.3.3 which shows that some GvMF_k distribution are conditional offset normals.

In what follows we will indicate with c the real positive parameter of the GvMF, to distinguish it from κ, the parameter of the GvM. The proofs are omitted and can be found in Gatto (2011).

Result 3.3.1. *Let* $\boldsymbol{X} \sim GvMF_3(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\mu}_3; c_1, c_2, c_3; i_1, i_2, i_3)$ *on* $\mathbb{S}^1$ *and define* $\theta = \arg\{\boldsymbol{X}\}$. *Then* $\theta \sim GvM_3(\mu_1, \mu_2, \mu_3; \kappa_1, \kappa_2, \kappa_3)$ *where for* $\nu_r = \arg\{\boldsymbol{\mu}_r\}$, $r = 1, 2, 3$ *and* $\arctan : \mathbb{R} \mapsto (-\pi/2, \pi/2)$ *we have*

$$\mu_1 = \left(\arctan \frac{4c_1 \sin \nu_1 + 3c_3 \sin \nu_3}{4c_1 \cos \nu_1 + 3c_3 \cos \nu_3} + \pi I \left\{ 4c_1 \cos \nu_1 + 3c_3 \cos \nu_3 < 0 \right\} \right) \bmod 2\pi, \quad (3.8)$$

with

$$\begin{aligned} &\kappa_1 = \left\{ c_1^2 + (9/16)c_3^2 + (3/2)c_1 c_3 \cos(\nu_1 - \nu_3) \right\}^{\frac{1}{2}}, \\ &\mu_2 = \nu_2, \kappa_2 = c_2/2, \mu_3 = \nu_3 \text{ and } \kappa_3 = c_3/4. \end{aligned} \quad (3.9)$$

Definition 3.3.2. A conditional offset distribution is the conditional distribution of a random vector given that it has norm equal to one.

Result 3.3.3. *Let* $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *in* $\mathbb{R}^p$, *with* $\boldsymbol{\Sigma}$ *positive-definite covariance matrix. If* $\boldsymbol{\mu} \neq \boldsymbol{0}$, *then* $\boldsymbol{X}$ *has a* $GvMF_{p+1}$ *conditional offset distribution on* $\mathbb{S}^{p-1}$. *This means that*

$$\boldsymbol{X} \mid \|\boldsymbol{X}\| = 1 \sim \text{GvMF}_{p+1}(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{p+1}; c_1, \ldots, c_{p+1}; 1, 2, \ldots, 2),$$

where, given $\boldsymbol{V} = (\boldsymbol{\nu}_{\circ 1}, \ldots, \boldsymbol{\nu}_{\circ p})$ *is non-singular such that* $\boldsymbol{\Sigma}^{-1} = \boldsymbol{V}\boldsymbol{V}^\top$, *we have that*

$$\begin{aligned} &\boldsymbol{\mu}_1 = \left\| \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \right\|^{-1} \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}, \; c_1 = \left\| \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \right\|, \\ &\boldsymbol{\mu}_{j+1} = \|\boldsymbol{\nu}_{\circ j}\|^{-1} \boldsymbol{\nu}_{\circ j}, c_{j+1} = -\|\boldsymbol{\nu}_{\circ j}\|^2/2, \quad \text{for } j = 1, \ldots, p. \end{aligned}$$

On the other hand, if $\boldsymbol{\mu} = \boldsymbol{0}$ *then*

$$\boldsymbol{X} \mid \|\boldsymbol{X}\| = 1 \sim \text{GvMF}_p(\boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_{p+1}; c_2, \ldots, c_{p+1}; 2, \ldots, 2).$$

Definition 3.3.4. Let $\boldsymbol{\Sigma}$ be a $p \times p$ semi-positive definite matrix with rank $r < p$. The generalized inverse of $\boldsymbol{\Sigma}$ is denote by $\boldsymbol{\Sigma}^-$ and is defined as the matrix that satisfies $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^-\boldsymbol{\Sigma} = \boldsymbol{\Sigma}$

Result 3.3.5. *Let* $\boldsymbol{\Sigma}$ *be a* $p \times p$ *semi-positive definite matrix with rank* $r < p$, *and let* $\lambda_1, \ldots, \lambda_r > 0$ *its non-zero eigenvalues. Let* $\boldsymbol{X} \in \mathbb{R}^p$ *have singular normal density at* $\boldsymbol{x} \in \mathbb{R}^p$ *given by*

$$(2\pi)^{-\frac{r}{2}} (\lambda_1 \cdot \ldots \cdot \lambda_r)^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^-(\boldsymbol{x} - \boldsymbol{\mu}) \right\}.$$

Then, for $\boldsymbol{\mu} \neq \boldsymbol{0}$, $\boldsymbol{X} \mid \|\boldsymbol{X}\| \sim GvMF_{r+1}(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{r+1}; c_1, \ldots, c_{r+1}; 1, 2, \ldots, 2)$ *on* $\mathbb{S}^{p-1}$, *where* $\boldsymbol{\mu}_j$ *and* c_j *for* $j = 1, \ldots, r+1$ *have the same form as in Result 3.3.3, with* $\boldsymbol{\Sigma}^-$ *instead of* $\boldsymbol{\Sigma}^{-1}$. *When* $\boldsymbol{\mu} = \boldsymbol{0}$, $\boldsymbol{X} \mid \|\boldsymbol{X}\| \sim GvMF_r(\boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_{r+1}; c_2, \ldots, c_{r+1}; 2, \ldots, 2)$ *on* $\mathbb{S}^{p-1}$.

Result 3.3.6. *Let $\boldsymbol{X} \in \mathbb{R}^p$ be a random variable with exponential spherical density proportional to*

$$\exp\left\{-\frac{1}{r}\left[(\boldsymbol{x}-\boldsymbol{\lambda})^\top \boldsymbol{W}(\boldsymbol{x}-\boldsymbol{\lambda})\right]^{\frac{r}{2}}\right\}, \tag{3.10}$$

with $\boldsymbol{W} = w\boldsymbol{I}$ for $w > 0$ and $k = r/2 \in \{1, 2, \ldots\}$. If $\boldsymbol{\lambda} \neq \boldsymbol{0}$, then

$$\boldsymbol{X} \mid \|\boldsymbol{X}\| = 1 \sim \mathrm{GvMF}_k\left(\boldsymbol{\mu}, \ldots, \boldsymbol{\mu}; c_1, \ldots, c_k; 1, \ldots, k\right),$$

where $\boldsymbol{\mu} = |\boldsymbol{\lambda}\|^{-1}\boldsymbol{\lambda}$ and

$$c_r = (-2)^{r-1}\frac{w^k}{k}\begin{pmatrix} k \\ j \end{pmatrix}\left(1 + \|\boldsymbol{\lambda}\|^2\right)^{k-r}\|\boldsymbol{\lambda}\|^r, \quad \textit{for } r = 1, \ldots, k. \tag{3.11}$$

On the other hand, when $\boldsymbol{\lambda} = \boldsymbol{0}$ then the conditional offset distribution of $\boldsymbol{X}$ is the uniform distribution on $\mathbb{S}^{p-1}$.

Result 3.3.7. *Let $\boldsymbol{X} \in \mathbb{R}^2$ with density proportional to* (3.10) *with $\boldsymbol{W} = w\boldsymbol{I}$ for $w > 0$ and $k = r/2 = 3$. Then*

$$\arg\{\boldsymbol{X}\} \mid \|\boldsymbol{X}\| = 1 \sim \mathrm{GvM}_3\left(\mu_1, \mu_2, \mu_3; \kappa_1, \kappa_2, \kappa_3\right),$$

with c_1, c_2, c_3 given in (3.11)*, μ_1 given in* (3.8)*, and κ_1 in* (3.9)*, where $\nu_1 = \nu_3 = \arg\{\boldsymbol{\lambda}\}$, $\mu_2 = \mu_3 = \arg\{\boldsymbol{\lambda}\}$, $\kappa_2 = c_2/2$ and $\kappa_3 = c_3/4$.*

Chapter 4

Bayesian tests of symmetry for the generalized von Mises distribution[†]

This chapter proposes three Bayesian tests on the symmetry of the GvM distribution. The symmetry of a circular distribution is a fundamental question and this topic has been studied in recent years. In the context of testing symmetry, one can mention: Pewsey (2002), who proposes a test of symmetry around an unknown axis based on the second sine sample moment, and Pewsey (2004), who considers the case where the symmetry is around the median axis. Both tests are frequentist and no Bayesian test of symmetry appears available in the literature. In fact, Bayesian analysis for circular data has remained underdeveloped, partly because of the lack of nice conjugate classes of distributions. Moreover, Bayesian analysis has focused on the vM model, which is symmetric. We refer to p. 278-279 of Jammalamadaka and SenGupta (2001) for a review on Bayesian analysis for circular data.

In this context, this chapter proposes Bayesian tests of symmetry for the GvM model (2.4). The first test proposed concerns the parameter δ. The null hypothesis is $\delta = 0$, that is, no shift between cosines of frequency one and two. In this case, the distribution is symmetric around the axis passing through μ_1. It is bimodal with one mode at μ_1 and the other one at $\mu_1 + \pi$, whenever $\kappa_1 < 4\kappa_2$. If $\kappa_1 \geq 4\kappa_2$, then it is unimodal with mode at μ_1. We refer to Table 1 of Gatto and Jammalamadaka (2007). The second test is on the precise characterization of axial symmetry, i.e. on $\delta = 0$ or $\delta = \pi/2$. So far $\kappa_2 > 0$ is considered and the third test is for $\kappa_2 = 0$, so that the distribution is no longer GvM but vM, which is is axially symmetric. The Bayesian tests rely on the method of probability perturbation, where the probability distribution of the null hypothesis is slightly perturbed, in order to give a positive prior probability to the null hypothesis, which would be null otherwise. It would be interesting to consider the above null hypotheses under the frequentist perspective, perhaps with the likelihood ratio approach. This topic is not studied in this chapter, in order to limit its length.

[†]Salvador and Gatto (2022b)

B_{01}	evidence for H_0
< 1	negative
1 to 1.5	not worth more than a bare mention
1.5 to 5	positive
5 to 10	substantial
10 to 20	strong
> 20	decisive

Table 4.1: Guidelines for the interpretation of Bayes factors.

The remaining part of this chapter is organized as follows. Section 4.1 gives the derivation of these Bayesian tests and their Bayes factors. Section 4.1.1 presents the approach used for these tests: Section 4.1.2 considers the test of no shift between cosines, Section 4.1.3 considers the test of symmetry and Section 4.1.4 considers the test of vM axial symmetry. Numerical results are presented in Section 4.2: Section 4.2.1 presents a Monte Carlo study of the tests of Section 4.1.1 whereas Section 4.2.4 presents the application to some real data. Final remarks are given in Section 4.3.

4.1 Bayesian tests and perturbation method for the GvM model

The proposed tests rely on Bayes factors. The Bayes factor B_{01} indicates the evidence of the null hypothesis w.r.t. the general alternative. Let us denote by $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$ the sample. Then

$$B_{01} = \frac{P[\boldsymbol{\theta}|\mathrm{H}_0]}{P[\boldsymbol{\theta}|\mathrm{H}_1]} = \frac{P[\mathrm{H}_0|\boldsymbol{\theta}]}{P[\mathrm{H}_1|\boldsymbol{\theta}]} \cdot \frac{P[\mathrm{H}_1]}{P[\mathrm{H}_0]} = \frac{R_1}{R_0}, \tag{4.1}$$

where

$$R_0 = \frac{P[\mathrm{H}_0]}{P[\mathrm{H}_1]} = \frac{P[\mathrm{H}_0]}{1 - P[\mathrm{H}_0]} \quad \text{and} \quad R_1 = \frac{P[\mathrm{H}_0|\boldsymbol{\theta}]}{P[\mathrm{H}_1|\boldsymbol{\theta}]} = \frac{P[\mathrm{H}_0|\boldsymbol{\theta}]}{1 - P[\mathrm{H}_0|\boldsymbol{\theta}]},$$

are the prior and the posterior odds, respectively. The case $B_{01} > 1$ indicates evidence for H_0. Otherwise, when $B_{01} < 1$ the sample supports H_1. Interpretations of the values of the Bayes factor can be found in Jeffreys (1939) and Kass and Raftery (1995). Our synthesis of these interpretations is given in Table 4.1, which provides a qualitative scale for the Bayes factor.

The null hypotheses are simple, in the sense that they concern only points of the parametric space. The fact that these points have probability null does not allow for the computation of Bayes factors. Therefore we use an approach with probability perturbation explained in the next section.

4.1.1 Bayesian tests of simple hypotheses

The practical relevance of a simple null hypothesis, i.e. of the type $H_0 : \xi = \xi_0$, has been widely debated in the statistical literature. According to Berger and Delampady: "it is rare, and perhaps impossible, to have a null hypothesis that can be exactly modeled as $\theta = \theta_0$". They illustrate their claim by the following example. "More common precise hypotheses such as H_0:*Vitamin C has no effect on the common cold* are clearly not meant to be though of as exact point nulls; surely vitamin C has some effects, although perhaps a very minuscule effect." A similar example involving forensic science can be found in Lindley (1977). When the parameter ξ is of continuous nature, it is usually more realistic to consider null hypotheses of the type $H_{0,\varepsilon} : |\xi - \xi_0| \leq \varepsilon/2$, for some small $\varepsilon > 0$. This solves also the problem of the vanishing prior probability of H_0, namely $P[\xi = \xi_0] = 0$. This problem is sometimes addressed by giving a positive probability to $\{\xi = \xi_0\}$. However, Berger and Sellke (1987) explain that the two approaches should be related. "It is convenient to specify a prior distribution for the testing problem as follows: let $0 < \pi_0 < 1$ denote the prior probability of $H_0 : \theta = \theta_0$ ... One might question the assignment of a positive probability to H_0, because it is rarely the case that it is thought possible for $\theta = \theta_0$ to hold exactly ... H_0 is to be understood as simply an approximation to the realistic hypothesis $H_0 : |\theta - \theta_0| \leq b$ and π_0 is to be interpreted as the prior probability that would be assigned to $\{\theta : |\theta - \theta_0| \leq b\}$." Accordingly, we assign to the original simple hypothesis $H_0 : \xi = \xi_0$ the prior probability $p_0 > 0$ of $H_{0,\varepsilon} : \xi \in [\xi_0 - \varepsilon/2, \xi_0 + \varepsilon/2]$, for some $\varepsilon > 0$. Thus, we replace the prior probability measure P by its perturbation, obtained by the assignment of the probability $p_0 > 0$ to $\{\xi_0\}$. We denote by P_0 the probability measure P with the p_0-perturbation. To summarize: the point null hypotheses is made relevant with $p_0 = P_0[\xi = \xi_0] = P[\delta \in [\xi_0 - \varepsilon/2, \xi_0 + \varepsilon/2]] > 0$.

The length ε of the neighborhood of ξ_0, which determines the prior probability p_0 of H_0 under the perturbed model, should not be too small. A significant value of p_0 for the null hypothesis is in fact coherent with the frequentist approach of hypotheses tests, where computations of rejection regions or P-values are carried over under the null hypothesis. Berger (1985), p. 149, states that ε has to be chosen such that any ξ in $(\xi_0 - \varepsilon/2, \xi_0 + \varepsilon/2)$ becomes "indistinguishable" from ξ_0, while Berger and Sellke state that ε has to be "small enough" so that $H_{0,\varepsilon}$ can be "accurately approximated" by H_0. A related reference is Berger and Delampady (1987), who studied this problem with a Gaussian model, and Berger (1985), p. 149, who obtains an upper bound for the radius $\varepsilon/2$ under a simple Gaussian model. Two other references on the practical relevance of simple null hypotheses are Jeffreys (1939) and Zellner (1984).

We end this section with some comments regarding the choice of the prior distribution of ξ. This is a generally unsolved problem of Bayesian statistics and widely discussed in the literature, see e.g. Jeffreys (1939) and Kass and Wasserman (1996). According to Berger and Delampady (1987), there is "no choice of the prior that can claim to

be objective". In this chapter we follow the directives given in Berger and Delampady (1987) and Berger and Sellke (1987), where various details on the choice of the prior are discussed and some classes of priors are analysed. According to Berger and Delampady (1987), in absence of prior information, the prior should be symmetric about ξ_0 and non-increasing w.r.t. $|\xi - \xi_0|$. Otherwise, one could find some "favoured" alternative values of ξ; cf. Berger and Sellke (1987). Our choices of priors are presented in Section 4.2: for each test of the study we compute Bayes factors under priors obtained by varying the concentration around the generic value ξ_0.

4.1.2 Test of no shift between cosines of GvM

Consider the Bayesian test on the GvM model (2.4) of the null the hypothesis

$$\mathrm{H}_0 : \delta = 0,$$

where $\delta = (\mu_1 - \mu_2) \bmod \pi$ and where the values of $\mu_1, \kappa_1, \kappa_2$ are assumed known and equal to $\mu_1^0, \kappa_1^0, \kappa_2^0$, respectively. Under the original probability measure P, the random parameter δ has an absolutely continuous prior distribution and so $P[\delta = 0] = 0$. According to Section 4.1.1 we define the perturbation of the probability measure P, denoted P_0, for which $p_0 = P_0[\delta = 0] > 0$. This perturbation is the assignment to $\{\delta = 0\}$ of the probability mass that initially lies close to that P-null set. Let $\varepsilon > 0$ and consider the set

$$\mathcal{A}_\varepsilon = \left\{ \delta \in [0, \pi) \middle| \delta \in \left[0, \frac{\varepsilon}{2}\right] \cup \left[\pi - \frac{\varepsilon}{2}, \pi\right) \right\}. \tag{4.2}$$

The complement is

$$\mathcal{A}_\varepsilon^{\mathsf{c}} = \left\{ \delta \in [0, \pi) \middle| \delta \in \left(\frac{\varepsilon}{2}, \pi - \frac{\varepsilon}{2}\right) \right\}.$$

Note that (4.2) refers to a neighborhood of the origin of the circle of circumference π. We thus assign to p_0 the value

$$p_0 = P\left[\mathcal{A}_\varepsilon\right], \tag{4.3}$$

for some suitably small $\varepsilon > 0$. The prior distribution function (d.f.) under the perturbed probability measure P_0 at any $\delta' \in [0, \pi)$ is given by

$$p_0 \Delta(\delta') + (1 - p_0) G(\delta'). \tag{4.4}$$

where G denotes the prior d.f. of δ and where Δ is the Dirac d.f., which assigns mass one to the origin. Denote by g the density of G. If $0 \notin (\delta', \delta' + \mathrm{d}\delta')$, for some $\delta' \in (0, \pi)$, where the relations $\in$ and $\notin$ are meant circularly over the circle of circumference π, then (4.4) implies

$$P_0[\delta \in (\delta', \delta' + \mathrm{d}\delta')] = (1 - p_0) g(\delta') \mathrm{d}\delta' = (1 - p_0) P[\delta \in (\delta', \delta' + \mathrm{d}\delta')]. \tag{4.5}$$

Let $\theta_1, \ldots, \theta_n$ be independent circular random variables that follow the GvM distribution (2.4). For simplicity, we denote the joint density of $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$, with the fixed values δ', μ_1^0, κ_1^0 and κ_2^0, as

$$f(\boldsymbol{\theta}|\delta') = \left\{2\pi G_0(\delta', \kappa_1^0, \kappa_2^0)\right\}^{-n} \exp\left\{\kappa_1^0 \sum_{i=1}^{n} \cos(\theta_i - \mu_1^0) + \kappa_2^0 \sum_{i=1}^{n} \cos 2(\theta_i - \mu_1^0 + \delta')\right\}. \tag{4.6}$$

When considered as a function of δ', (4.6) becomes the likelihood of δ. Then, by (4.5) the marginal density of $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$ under the perturbed probability is given by

$$\begin{aligned}
m(\boldsymbol{\theta}) &= \int_{[0,\pi)} f(\boldsymbol{\theta}|\delta') P_0[\delta \in (\delta', \delta' + \mathrm{d}\delta')] \\
&= \int_{\mathcal{A}_\varepsilon} f(\boldsymbol{\theta}|\delta') P_0[\delta \in (\delta', \delta' + \mathrm{d}\delta')] + (1 - p_0) \int_{\mathcal{A}_\varepsilon^{\mathrm{c}}} f(\boldsymbol{\theta}|\delta') g(\delta') \mathrm{d}\delta' \\
&= p_0 f(\boldsymbol{\theta}|0) + (1 - p_0) \int_{\mathcal{A}_\varepsilon} f(\boldsymbol{\theta}|\delta') g(\delta') \mathrm{d}\delta' + (1 - p_0) \int_{\mathcal{A}_\varepsilon^{\mathrm{c}}} f(\boldsymbol{\theta}|\delta') g(\delta' \mathrm{d}\delta' \\
&\sim 2p_0 f(\boldsymbol{\theta}|0) + (1 - p_0) \int_{\mathcal{A}_\varepsilon^{\mathrm{c}}} f(\boldsymbol{\theta}|\delta') g(\delta') \mathrm{d}\delta', \text{ as } \varepsilon \to 0.
\end{aligned}$$

The above asymptotic equivalence is due to

$$(1 - p_0) \int_{\mathcal{A}_\varepsilon} f(\boldsymbol{\theta}|\delta') g(\delta') \mathrm{d}\delta' = (1 - p_0) p_0 \int_{\mathcal{A}_\varepsilon} f(\boldsymbol{\theta}|\delta') \frac{g(\delta')}{p_0} \mathrm{d}\delta' \sim p_0 f(\theta|0), \text{ as } \varepsilon \to 0.$$

The posterior perturbed probability, namely the conditional perturbed probability of $\{\delta = 0\}$ given $\boldsymbol{\theta}$, can be approximated as follows,

$$\begin{aligned}
P_0[\delta = 0|\boldsymbol{\theta}] &\sim \frac{p_0 f(\boldsymbol{\theta}|0)}{2p_0 f(\boldsymbol{\theta}|0) + (1 - p_0) \int_{\mathcal{A}_\varepsilon^{\mathrm{c}}} f(\boldsymbol{\theta}|\delta') g(\delta') \mathrm{d}\delta'} \\
&= \frac{1}{2} \left(1 + \frac{1 - p_0}{p_0} \frac{\int_{\mathcal{A}_\varepsilon^{\mathrm{c}}} f(\boldsymbol{\theta}|\delta') g(\delta') \mathrm{d}\delta'}{2 f(\boldsymbol{\theta}|0)} \right)^{-1}, \quad \text{as } \varepsilon \to 0.
\end{aligned}$$

In order to compute the Bayes factor for this test, we define the prior odds $R_0 = p_0/(1-p_0)$ and the posterior odds $R_1 = P_0[\delta = 0|\boldsymbol{\theta}]/(1 - P_0[\delta = 0|\boldsymbol{\theta}])$. The Bayes factor is the posterior over the prior odds, namely $B_{01} = R_1/R_0$. Clearly $p_0 \leq P_0[\delta = 0|\boldsymbol{\theta}]$ iff $B_{01} \geq 1$ and, the larger $P_0[\delta = 0|\boldsymbol{\theta}] - p_0$ becomes, the larger B_{01} becomes: a large Bayes factor tells that the data support the null hypothesis. From the approximation

$$R_1 \sim \left[1 + \frac{1 - p_0}{p_0} \frac{\int_{\mathcal{A}_\varepsilon^{\mathrm{c}}} f(\boldsymbol{\theta}|\delta') g(\delta') \mathrm{d}\delta'}{f(\boldsymbol{\theta}|0)} \right]^{-1}$$

and from some simple algebraic manipulation, we obtain the computable approximation to the Bayes factor $B_{01} = R_1/R_0$ given by

$$B_{01} \sim \frac{f(\boldsymbol{\theta}|0)}{\int_{\mathcal{A}_\varepsilon^{\mathrm{c}}} f(\boldsymbol{\theta}|\delta') g(\delta') \mathrm{d}\delta'}, \text{ as } \varepsilon \to 0. \tag{4.7}$$

The representation of the Bayes factor (4.7) is asymptotically correct and we remind that, in the context where we approximate the null hypothesis with a neighbourhood by the point null hypothesis, the reasoning is always of asymptotic nature. A reference for this perturbation technique is Berger (1985), p. 148-150.

Regarding the large sample asymptotics of the proposed test, it is know that, for a sample of n independent random variables with common distribution with true parameter ξ_0, the posterior distribution converges to the distribution with total mass over ξ_0, as $n \to \infty$. This means that the posterior mode is a consistent estimator. We deduce that, under H_0,

$$P_0[\delta = 0|\boldsymbol{\theta}] = P[\mathcal{A}_\varepsilon|\boldsymbol{\theta}] \xrightarrow{\mathsf{P}} 1, \quad \text{as } n \to \infty.$$

Consequently, $R_1 = P_0[\delta = 0|\boldsymbol{\theta}]/(1 - P_0[\delta = 0|\boldsymbol{\theta}]) \xrightarrow{\mathsf{P}} \infty$ and $B_{01} = R_1/R_0 \xrightarrow{\mathsf{P}} \infty$, as $n \to \infty$. The Bayesian test of $\mathrm{H}_0 : \delta = 0$ is consistent in this sense.

We now give some computational remarks that are also valid for the tests of Sections 4.1.3 and 4.1.4. The integral appearing in the denominator of (4.7) can be easily evaluated by Monte Carlo integration. For a given large integer s, we generate $\delta^{(i)}$, for $i = 1, \ldots, s$, from the density g and then we compute the approximation

$$\int_{\mathcal{A}_\varepsilon^{\mathsf{c}}} f(\boldsymbol{\theta}|\delta')g(\delta')\mathrm{d}\delta' = \int_{\mathcal{A}_\varepsilon^{\mathsf{c}}} f(\boldsymbol{\theta}|\delta')P[\delta \in (\delta', \delta' + \mathrm{d}\delta')] \simeq \frac{1}{s}\sum_{i=1}^{s} f(\boldsymbol{\theta}|\delta^{(i)})\mathrm{I}\{\delta^{(i)} \in \mathcal{A}_\varepsilon^{\mathsf{c}}\}, \tag{4.8}$$

where $\mathrm{I}\{A\}$ denotes the indicator of statement or event A. For the computation normalizing constant of the GvM distribution given in (2.5) one can use the Fourier series

$$G_0(\delta, \kappa_1, \kappa_2) = I_0(\kappa_1)I_0(\kappa_2) + 2\sum_{j=1}^{\infty} I_{2j}(\kappa_1)I_j(\kappa_2)\cos 2j\delta, \tag{4.9}$$

where $\delta \in [0, \pi)$ and $\kappa_1, \kappa_2 > 0$, see e.g. Gatto and Jammalamadaka (2007).

4.1.3 Test of axial symmetry of GvM

In this Section, the symmetry of the GvMF distribution is introduced and its characterization is given. In this section we consider the Bayesian test of axial symmetry for the GvM model (2.4). A circular density g is symmetric around the angle $\alpha/2$, for some $\alpha \in [0, 2\pi)$, if $g(\theta) = g(\alpha - \theta)$, $\forall\theta \in [0, 2\pi)$. In this case we have also $g(\theta) = g((\alpha + 2\pi) - \theta)$, so that symmetry around $\alpha/2 + \pi$ holds as well: the symmetry is indeed an axial one.

Proposition 4.1.1 (Characterization of axial symmetry for the GvM distribution)**.** *The GvM distribution (2.4) is axial symmetric iff*

$$\delta = 0 \quad or \quad \delta = \frac{\pi}{2}.$$

In both cases, the axis of symmetry has angle μ_1.

The proof of Proposition 4.1.1 is given in Appendix A.

Note that δ is defined modulo π and that for $\kappa_2 = 0$ or $\kappa_1 = 0$ the GvM reduces respectively to the vM or to the axial vM, defined as vM_2 and given in (2.6). These two distributions are clearly symmetric, but Proposition 4.1.1 gives the characterization of symmetry in terms of δ since we define the GvM distribution in (2.4) with parameters $\kappa_1, \kappa_2 > 0$.

As mentioned at the beginning of the section, symmetry of a circular distribution around an angle is the symmetry around an axis. In particular, for the GvM density, adding 2π to α would not have any influence. For the GvM density, this is made explicit in (A.3) of Appendix A, where adding 2π to α would not have any influence. Figure 4.1 provides two numerical illustrations of the axial symmetry of the GvM distribution. The graph in Figure 4.1a shows the density of the $\mathrm{GvM}(\pi, \pi, 0.1, 5.5)$ distribution: $\delta = 0$ and the axis of symmetry is at angle $\mu_1 = \pi$. The graph in Figure 4.1b shows the density of the $\mathrm{GvM}(\pi/2, 0, 5.5, 0.1)$ distribution: $\delta = \pi/2$ and the axis of symmetry is at angle $\mu_1 = \pi/2$. Thus, Proposition 4.1.1 allows us to write the null hypothesis of axial symmetry as

$$\mathrm{H}_0 : \delta = 0 \text{ or } \delta = \frac{\pi}{2},$$

where the values of $\mu_1, \kappa_1, \kappa_2$ are assumed known and equal to $\mu_1^0, \kappa_1^0, \kappa_2^0$, respectively. The Bayesian test is obtained by perturbation of the probability measure P, which is denoted P_0. The probabilities

$$p_0 = P_0[\delta = 0] > 0 \text{ and } p_{\frac{\pi}{2}} = P_0\left[\delta = \frac{\pi}{2}\right] > 0$$

are the probabilities masses of $\{\delta = 0\}$ and $\{\delta = \frac{\pi}{2}\}$ of the perturbed measure, respectively. They are obtained from

$$p_0 = P\left[\delta \in \left[0, \frac{\varepsilon}{2}\right] \cup \left[\pi - \frac{\varepsilon}{2}, \pi\right)\right] \text{ and } p_{\frac{\pi}{2}} = P\left[\delta \in \left[\frac{\pi}{2} - \frac{\varepsilon}{2}, \frac{\pi}{2} + \frac{\varepsilon}{2}\right]\right],$$

for suitably small $\varepsilon > 0$. As is Section 4.1.2, the prior d.f. of δ under the perturbed probability P_0 at any $\delta' \in [0, \pi)$ is given by

$$p_0 \Delta(\delta') + p_{\frac{\pi}{2}} \Delta\left(\delta' - \frac{\pi}{2}\right) + \left\{1 - \left(p_0 + p_{\frac{\pi}{2}}\right)\right\} G(\delta'), \tag{4.10}$$

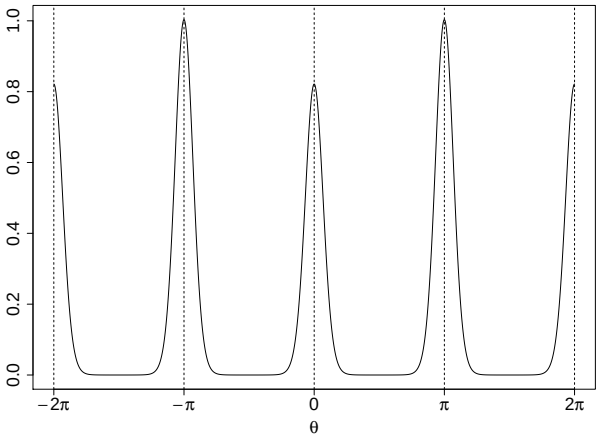
where G is the prior d.f. of δ under P. It follows from (4.10) that for $0, \pi/2 \notin (\delta', \delta' + \mathrm{d}\delta')$, for some $\delta' \in (0, \pi) \setminus \{\pi/2\}$,

$$P_0[\delta \in (\delta', \delta' + \mathrm{d}\delta')] = \left[1 - \left(p_0 + p_{\frac{\pi}{2}}\right)\right] g(\delta')\mathrm{d}\delta' = \left[1 - \left(p_0 + p_{\frac{\pi}{2}}\right)\right] P[\delta \in (\delta', \delta' + \mathrm{d}\delta')],$$

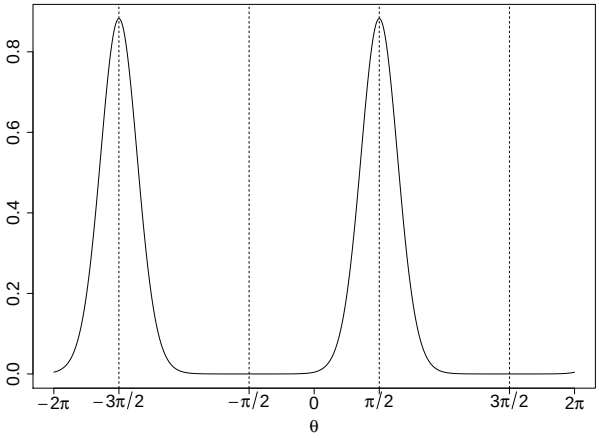
where g is the density of G.

Let

$$\mathcal{B}_{\varepsilon,0} = \left\{\delta \in [0, \pi) \middle| \delta \in \left[0, \frac{\varepsilon}{2}\right] \cup \left[\pi - \frac{\varepsilon}{2}, \pi\right)\right\}, \text{ and } \mathcal{B}_{\varepsilon,\frac{\pi}{2}} = \left\{\delta \in [0, \pi) \middle| \delta \in \left[\frac{\pi}{2} - \frac{\varepsilon}{2}, \frac{\pi}{2} + \frac{\varepsilon}{2}\right]\right\}.$$



(a) GvM$(\pi, 0, 0.1, 5.5)$ density ($\delta = 0$).



(b) GvM$(\pi/2, 0, 5.5, 0.1)$ density ($\delta = \pi/2$).

Figure 4.1: Two axial symmetric GvM densities over the interval $(-2\pi, 2\pi)$ and with their axis of symmetry at angle μ_1 shown by vertical dashed lines.

Define

$$\mathcal{B}_\varepsilon = \left\{ \delta \in [0, \pi) \middle| \delta \in \left[0, \frac{\varepsilon}{2}\right] \cup \left[\pi - \frac{\varepsilon}{2}, \pi\right) \vee \delta \in \left[\frac{\pi}{2} - \frac{\varepsilon}{2}, \frac{\pi}{2} + \frac{\varepsilon}{2}\right] \right\} = \mathcal{B}_{\varepsilon,0} \cup \mathcal{B}_{\varepsilon,\frac{\pi}{2}}.$$

Its complement is given by

$$\mathcal{B}_\varepsilon^{\mathsf{c}} = \left\{ \delta \in [0, \pi) \middle| \delta \in \left(\frac{\varepsilon}{2}, \frac{\pi}{2} - \frac{\varepsilon}{2}\right) \cup \left(\frac{\pi}{2} + \frac{\varepsilon}{2}, \pi - \frac{\varepsilon}{2}\right) \right\}.$$

The marginal density of $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$ w.r.t. the perturbed probability P_0 is given by

$$\begin{aligned}
m(\boldsymbol{\theta}) &= \int_{[0,\pi)} f(\boldsymbol{\theta}|\delta') P_0[\delta \in (\delta', \delta' + \mathrm{d}\delta')] \\
&= \int_{\mathcal{B}_\varepsilon} f(\boldsymbol{\theta}|\delta') P_0[\delta \in (\delta', \delta' + \mathrm{d}\delta')] + \left[1 - \left(p_0 + p_{\frac{\pi}{2}}\right)\right] \int_{\mathcal{B}_\varepsilon^c} f(\boldsymbol{\theta}|\delta') g(\delta') \mathrm{d}\delta' \\
&= p_0 f(\boldsymbol{\theta}|0) + p_{\frac{\pi}{2}} f\left(\boldsymbol{\theta}\middle|\frac{\pi}{2}\right) + \left[1 - \left(p_0 + p_{\frac{\pi}{2}}\right)\right] \int_{\mathcal{B}_\varepsilon} f(\boldsymbol{\theta}|\delta') g(\delta') \mathrm{d}\delta' \\
&\qquad + \left[1 - \left(p_0 + p_{\frac{\pi}{2}}\right)\right] \int_{\mathcal{B}_\varepsilon^c} f(\boldsymbol{\theta}|\delta') g(\delta') \mathrm{d}\delta' \\
&\sim 2p_0 f(\boldsymbol{\theta}|0) + 2p_{\frac{\pi}{2}} f\left(\boldsymbol{\theta}\middle|\frac{\pi}{2}\right) + \left[1 - \left(p_0 + p_{\frac{\pi}{2}}\right)\right] \int_{\mathcal{B}_\varepsilon^c} f(\boldsymbol{\theta}|\delta') g(\delta') \mathrm{d}\delta', \text{ as } \varepsilon \to 0.
\end{aligned}$$

In the asymptotic equivalence, as in Section 4.1.2, we notice that

$$\begin{aligned}
&\left[1 - \left(p_0 + p_{\frac{\pi}{2}}\right)\right] \int_{\mathcal{B}_\varepsilon} f(\boldsymbol{\theta}|\delta') g(\delta') \mathrm{d}\delta' \\
&= \left[1 - \left(p_0 + p_{\frac{\pi}{2}}\right)\right] \left[\int_{\mathcal{B}_{\varepsilon,0}} f(\boldsymbol{\theta}|\delta') g(\delta') \mathrm{d}\delta' + \int_{\mathcal{B}_{\varepsilon,\frac{\pi}{2}}} f(\boldsymbol{\theta}|\delta') g(\delta') \mathrm{d}\delta'\right] \\
&= \left[1 - \left(p_0 + p_{\frac{\pi}{2}}\right)\right] \left[p_0 \int_{\mathcal{B}_{\varepsilon,0}} f(\boldsymbol{\theta}|\delta') \frac{g(\delta')}{p_0} \mathrm{d}\delta' + p_{\frac{\pi}{2}} \int_{\mathcal{B}_{\varepsilon,\frac{\pi}{2}}} f(\boldsymbol{\theta}|\delta') \frac{g(\delta')}{p_{\frac{\pi}{2}}} \mathrm{d}\delta'\right] \\
&\sim p_0 f(\boldsymbol{\theta}|0) + p_{\frac{\pi}{2}} f\left(\boldsymbol{\theta}\middle|\frac{\pi}{2}\right), \quad \text{as } \varepsilon \to 0.
\end{aligned}$$

The posterior probability of $\{\delta = 0 \vee \delta = \pi/2\}$ under the perturbed probability measure is given by

$$\begin{aligned}
P_0\left[\delta = 0 \vee \delta = \frac{\pi}{2}\middle|\boldsymbol{\theta}\right] &\sim \frac{p_0 f(\boldsymbol{\theta}|0) + p_{\frac{\pi}{2}} f(\boldsymbol{\theta}|\frac{\pi}{2})}{2p_0 f(\boldsymbol{\theta}|0) + 2p_{\frac{\pi}{2}} f(\boldsymbol{\theta}|\frac{\pi}{2}) + [1 - (p_0 + p_{\frac{\pi}{2}})]\mathcal{I}_1(\theta)} \\
&= \frac{1}{2}\left[1 + \frac{\left[1 - (p_0 + p_{\frac{\pi}{2}})\right] \mathcal{I}_1(\boldsymbol{\theta})}{2p_0 f(\boldsymbol{\theta}|0) + 2p_{\frac{\pi}{2}} f(\boldsymbol{\theta}|\frac{\pi}{2})}\right]^{-1}, \quad \text{as } \varepsilon \to 0,
\end{aligned}$$

where

$$\mathcal{I}_1(\boldsymbol{\theta}) = \int_{\mathcal{B}_\varepsilon^c} f(\boldsymbol{\theta}|\delta') g(\delta') \mathrm{d}\delta'.$$

With this we obtain the following approximation to the posterior odds,

$$R_1 = \frac{P_0\left[\delta = 0 \vee \delta = \frac{\pi}{2}|\boldsymbol{\theta}\right]}{1 - P_0\left[\delta = 0 \vee \delta = \frac{\pi}{2}|\boldsymbol{\theta}\right]} = \left[\frac{1}{P_0\left[\delta = 0 \vee \delta = \frac{\pi}{2}|\boldsymbol{\theta}\right]} - 1\right]^{-1} \sim \left[1 + \frac{\left[1 - (p_0 + p_{\frac{\pi}{2}})\right] \mathcal{I}_1(\boldsymbol{\theta})}{p_0 f(\boldsymbol{\theta}|0) + p_{\frac{\pi}{2}} f(\boldsymbol{\theta}\middle|\frac{\pi}{2})}\right]^{-1}$$

as $\varepsilon \to 0$. With the prior odds given by

$$R_0 = \frac{p_0 + p_{\frac{\pi}{2}}}{1 - (p_0 + p_{\frac{\pi}{2}})}$$

and after algebraic manipulations, we obtain the approximation to the Bayes factor given by

$$B_{01} \sim \frac{p_0 f(\boldsymbol{\theta}|0) + p_{\frac{\pi}{2}} f(\boldsymbol{\theta}|\frac{\pi}{2})}{(p_0 + p_{\frac{\pi}{2}})\mathcal{I}_1(\boldsymbol{\theta})}, \quad \text{as } \varepsilon \to 0.$$

4.1.4 Test of vM axial symmetry

We consider the Bayesian test of the null hypothesis that the sample follows a vM distribution against the alternative that it comes from an arbitrary GvM distribution. This null hypothesis implies axial symmetry in the class of vM distributions, whereas the alternative hypothesis includes both symmetric or asymmetric GvM distributions. Precisely, we have $\mathrm{H}_0 : \kappa_2 = 0$, where μ_1, μ_2 and κ_1 are assumed known and equal to μ_1^0, μ_2^0 and κ_1^0 respectively. The GvM with $\kappa_2 = 0$ reduces to the trivially symmetric vM distribution. Formally, the GvM is defined for $\kappa_2 > 0$ only, so that the symmetry considered here is no longer within the GvM class but it is rather a vM axial symmetry. This symmetry within the GvM class should be thought as approximate for vanishing values of κ_2.

Symmetry with the GvM formula can also be obtained with $\kappa_1 = 0$, in which case the GvM formula reduces to the vM_2 given in (2.6) that is trivially symmetric. This case is not analysed. In what follows we focus on the case of vM axial symmetry.

Because $P[\kappa_2 = 0] = 0$, we construct the perturbed probability P_0 such that $p_0 = P_0[\kappa_2 = 0] > 0$, where $p_0 = P[\kappa_2 \in [0, \varepsilon]]$, for some $\varepsilon > 0$ small. The prior d.f. of κ_2 under the probability P is G, and under the perturbed probability P_0 it is $p_0\Delta(\kappa_2') + (1 - p_0)G(\kappa_2')$, $\forall \kappa_2' \geq 0$. Assume $0 \notin (\kappa_2', \kappa_2' + \mathrm{d}\kappa_2')$, then

$$P_0\left[\kappa_2 \in (\kappa_2', \kappa_2' + \mathrm{d}\kappa_2')\right] = (1 - p_0)g(\kappa_2')\mathrm{d}\kappa_2' = (1 - p_0)P\left[\kappa_2 \in (\kappa_2', \kappa_2' + \mathrm{d}\kappa_2')\right],$$

where g is the density of G. With algebraic manipulations similar to those of Section 4.1.2, one obtains the approximation of the Bayes factor B_{01} given by

$$B_{01} \sim \frac{f(\boldsymbol{\theta}|0)}{\int_{\mathcal{C}_\varepsilon^{\mathsf{c}}} f(\boldsymbol{\theta}|\kappa_2')g(\kappa_2')\mathrm{d}\kappa_2'}, \text{ as } \varepsilon \to 0,$$

where $\mathcal{C}_\varepsilon = [0, \varepsilon]$, $\mathcal{C}_\epsilon^{\mathsf{c}}$ is its complement and where the likelihood of κ_2 is

$$f(\boldsymbol{\theta}|\kappa_2) = \left\{2\pi G_0\left(\delta^0, \kappa_1^0, \kappa_2\right)\right\}^{-n} \exp\left\{\kappa_1^0 \sum_{i=1}^n \cos(\theta_i - \mu_1^0) + \kappa_2 \sum_{i=1}^n \cos 2(\theta_i - \mu_1^0 + \delta^0)\right\},$$

with $\delta^0 = (\mu_1^0 - \mu_2^0) \bmod \pi$.

4.2 Numerical studies

This section provides some numerical studies for the tests introduced in Section 4.1. The major part is Section 4.2.1, which gives a simulation or Monte Carlo study of the performance of these tests. Section 4.2.4 provides an application to real measurements of wind directions.

4.2.1 Monte Carlo study

This section presents a Monte Carlo study for the tests introduced in Section 4.1: in Section 4.2.1.1 for the test of no shift between cosines, in Section 4.2.1.2 for the test of axial symmetry and in Section 4.2.2 for the test of vM axial symmetry. The results are summarized in Section 4.2.3. We obtain Bayes factors for each one of these three tests for $r = 10^4$ generations of samples of size $n = 50$, that are generated from the GvM or the vM distributions. The Monte Carlo approximation to the integral (4.8), and to the analogue integrals of the two other tests, is computed with $s = 10^4$ generations.
This simulation scheme is repeated three times and the results are compared in order to verify convergence. Confidence intervals for the Bayes factors based on the aggregation of the three simulations (with r replications each) are provided.

The vM_2 given in (2.6) is used as a prior distribution for the parameter of shift between cosines δ. According to the remark at the end of Section 4.1.1, we choose $\varepsilon = 0.05$ for the length of the interval of H_0 and the prior densities g as follows. For the test of no shift between cosines, we choose the $\text{vM}_2(0, \tau)$ distribution for δ, which is symmetric and unimodal with mode at $\delta = 0$. For the test of axial symmetry, we choose the mixture of $\text{vM}_2(0, \tau)$ and $\text{vM}_2(\pi/2, \tau)$ for δ. Finally, for the test of vM axial symmetry, we choose an uniform distribution for κ_2 that is highly concentrated at the boundary point 0.

4.2.1.1 Test of no shift between cosines of GvM

The null hypothesis considered is H_0: $\delta = 0$, with fixed $\mu_1 = \mu_1^0, \kappa_1 = \kappa_1^0, \kappa_2 = \kappa_2^0$, where $\mu_1^0 = \pi$, $\kappa_1^0 = 0.1$, $\kappa_2^0 = 5.5$. We consider three different cases, called D1, D1' and D2.
Case D1 For $i = 1, \ldots, s$, we generate $\delta^{(i)}$ from the prior of δ, which is $\text{vM}_2(\nu, \tau)$ with values of the hyperparameters $\nu = 0$ and $\tau = 250$. We obtain $p_0 = 0.570$ as prior probability of the null hypothesis under the perturbed probability measure. We take the first r of these prior values (that are all the values, since $r = s$) and then we obtain $\mu_2^{(i)} = (\mu_1^0 - \delta^{(i)}) \bmod \pi$ and generate the elements of the vector of n sample values $\boldsymbol{\theta}^{(i)}$ independently from $\text{GvM}(\mu_1^0, \mu_2^{(i)}, \kappa_1^0, \kappa_2^0)$, for $i = 1, \ldots, r$. With these simulated data we compute the Bayes factor $B_{01}^{(i)}$ with the approximation formula (4.7). We repeat this experiment three times. The fact of generating values of δ from its prior distribution, instead of taking $\delta = 0$ fixed by null hypothesis, is a way of inserting some prior uncertainty in the generated sample. If the prior is close, in some sense, to the null hypothesis, then we should obtain the Bayes factor larger than one, but smaller than the Bayes factor that would be obtained with the fixed value $\delta = 0$.

We obtained three sequences of 10^4 Bayes factors that can be summarized as follows. Figure 4.2a displays the three boxplots of the three simulated sequences of Bayes factors: Denote by $\bar{B}_{01}^{(j)}$ the mean of the Bayes factors of the j-th sequence, for $j = 1, 2, 3$,

corresponding to left, central and right boxplot respectively. We obtained:

$$\bar{B}_{01}^{(1)} = 2.887,\ \bar{B}_{01}^{(2)} = 3.028 \text{ and } \bar{B}_{01}^{(3)} = 2.955.$$

Figure 4.2b shows the histogram of the three generated sequences of r Bayes factors. The distribution is clearly not "bell-shaped" but it is however light-tailed: the Central limit theorem applies to the mean of the simulated Bayes factors. The asymptotic normal confidence interval for the mean value of the Bayes factors at level 0.95, and based on the three generated sequences, is given by

$$(2.937, 2.976).$$

According to Table 4.1 this interval indicates positive evidence for the null hypothesis: the data have indeed increased the evidence of the null hypothesis that $\delta = 0$, however to a marginal extent only. This situation can be explained by the fact that the prior density g is (highly) concentrated around 0, circularly. This can be seen in the graph of the prior density (Figure 4.2c), where the histogram of 10^4 generated values of δ is shown together with the prior density. Moreover, the variability originating from the fact the data are simulated under different values of δ leads to weaker values of the Bayes factor.

Case D1' In this other case we consider prior values of δ less concentrated around 0, by choosing $\nu = 0$ and $\tau = 50$. The resulting prior probability of H_0 is given by $p_0 = 0.276$. For $i = 1, \ldots, r$, we generate the elements of the vector of n sample values $\boldsymbol{\theta}^{(i)}$ independently from $\text{GvM}(\mu_1^0, \mu_2^0, \kappa_1^0, \kappa_2^0)$, with $\delta = 0$, thus with $\mu_2^0 = (\mu_1^0 - \delta) \bmod \pi = 0$. With these simulated data, we compute the Bayes factor $B_{01}^{(i)}$ with the approximation formula (4.7).

We obtained three sequences of $r = 10^4$ Bayes factors with means:

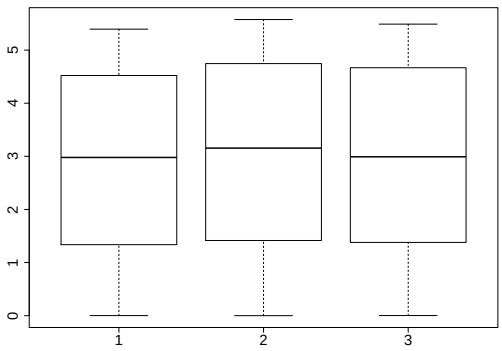
$$\bar{B}_{01}^{(1)} = 3.922,\ \bar{B}_{01}^{(2)} = 3.924 \text{ and } \bar{B}_{01}^{(3)} = 3.921.$$

The boxplots of the three respective generated sequences are shown in Figure 4.3a. The asymptotic normal confidence interval for the mean value of the Bayes factors, at level 0.95 and based on the three generated sequences, is

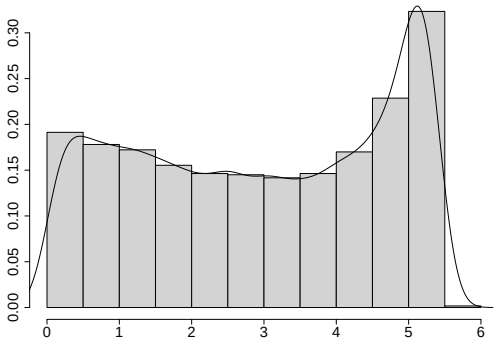
$$(3.901, 3.945).$$

As expected, the generated Bayes factors are larger than in case D1. The samples generated with $\delta = 0$ fixed have less uncertainty. We computed the posterior density of δ based on one generated sample. In Figure 4.3b we can see the graph of that posterior density, in continuous line, together with the graph of the prior density, in dashed line. The posterior is indeed more concentrated around 0, circularly.

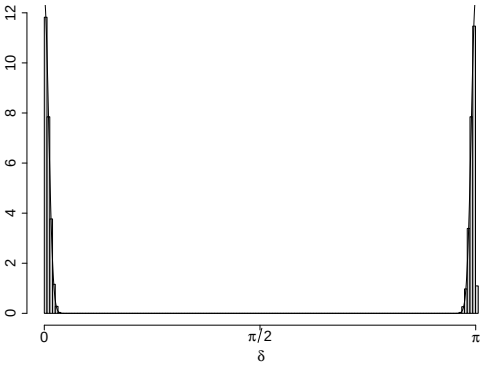
Case D2 We now further decrease the concentration of the prior of δ. The values of the hyperparameters are $\nu = 0$ and $\tau = 20$. We computed the prior probability of



(a) 3 Boxplots of the 3 sets of 10^4 simulated Bayes factors.



(b) Histogram of the sample of $3 \cdot r$ Bayes factors and graph of its estimated density (solid line).



(c) Prior density of δ with histogram of 10^4 generated values.

Figure 4.2: Results of Case D1.

the null hypothesis under perturbation $p_0 = 0.176$. We generated the samples $\boldsymbol{\theta}^{(i)}$, for $i = 1, \ldots, r$, with fixed value $\mu_2^0 = 0$.

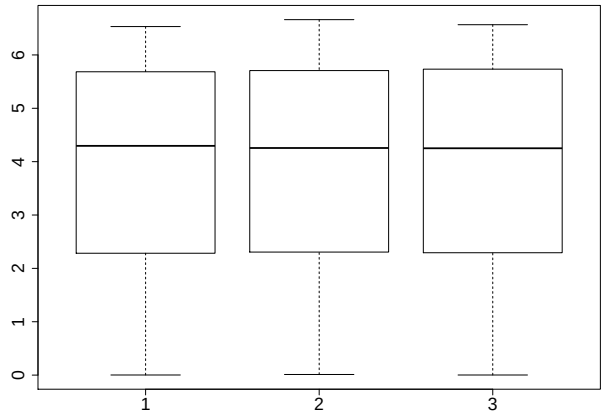
We obtained three sequences of $r = 10^4$ Bayes factors with means

$$\bar{B}_{01}^{(1)} = 5.477, \bar{B}_{01}^{(2)} = 5.539 \text{ and } \bar{B}_{01}^{(3)} = 5.511.$$

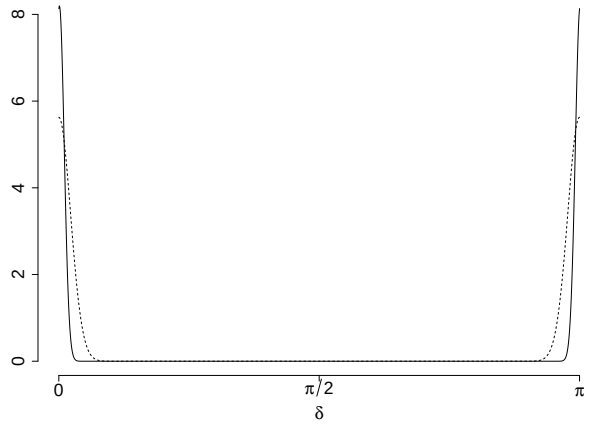
The boxplots of the three respective generated sequences are shown in Figure 4.4a. The asymptotic normal confidence interval for the mean value of the Bayes factors, at level 0.95 and based on the three generated sequences, is

$$(5.477, 5.541).$$

The Bayes factors are larger than they are in Cases D1 and D1'. Here they show substantial evidence for the null hypothesis. The prior distribution δ is less favourable to the null hypothesis and so the sample brings more additional evidence for the null hypothesis. Figure 4.4b shows the graph of the prior density, as dashed line, together



(a) 3 Boxplots of the 3 sets of 10^4 simulated Bayes factors.



(b) Prior density (dashed line) and posterior density (continuous line) of δ. The posterior is based on one generated sample.

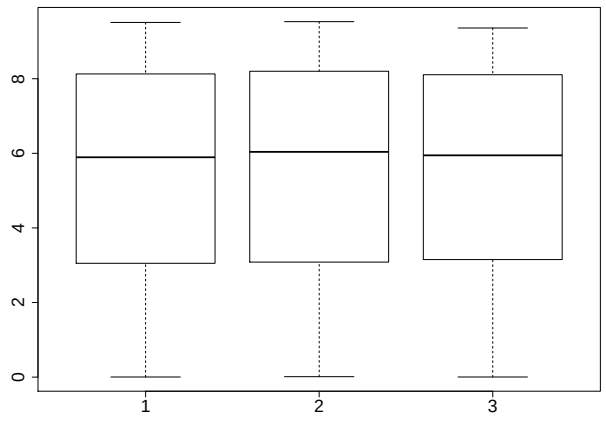
Figure 4.3: Results of Case D1'.

with the graph of a posterior density, as continuous line, for δ. The graph of the posterior density is based on one generated sample.

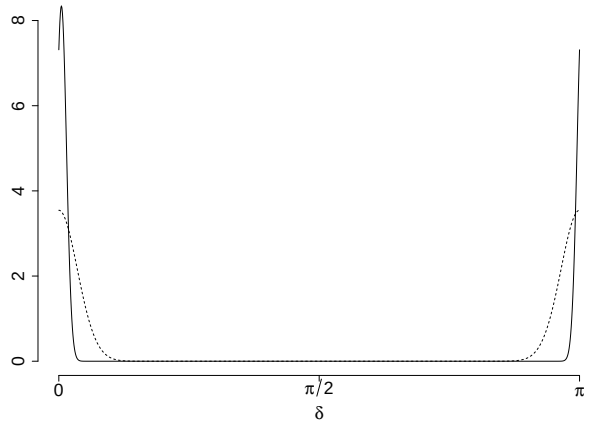
4.2.1.2 Test of axial symmetry of GvM

In this section we consider the null hypothesis of axial symmetry, viz. H_0: $\delta = 0$ or $\delta = \pi/2$, and the other parameters being fixed as follows, $\mu_1 = \mu_1^0, \kappa_1 = \kappa_1^0$ and $\kappa_2 = \kappa_2^0$. We choose as before $\mu_1^0 = \pi$, $\kappa_1^0 = 0.1$ and $\kappa_2^0 = 5.5$. We generate δ from the prior given by the mixture of vM_2 distributions $\xi\,\mathrm{vM}_2(\nu_1, \tau) + (1-\xi)\,\mathrm{vM}_2(\nu_2, \tau)$, with $\nu_1 = 0, \nu_2 = \pi/2$ and $\xi = 0.5$. We consider three different cases, called Cases S1, S2 and S3.

Case S1 We generated δ from the prior mixture with concentration parameter $\tau = 250$.



(a) 3 Boxplots of the 3 sets of 10^4 simulated Bayes factors.



(b) Prior density (dashed line) and posterior density (continuous line) of δ. The posterior is based on one generated sample.

Figure 4.4: Results of Case D2.

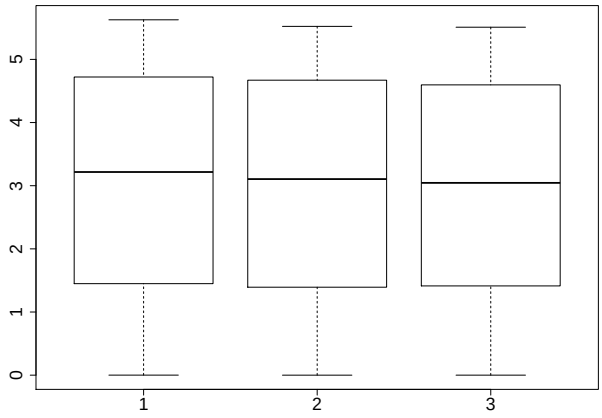
This prior distribution is close to the null distribution and Figure 4.5b displays its density, together with the histogram of 10^4 generations from it. We computed the prior probabilities of the null hypothesis under the perturbed probability measure $p_0 = p_{\pi/2} = 0.285$. We follow the principle of Case D1, where prior uncertainty is transmitted to the sample by considering generated values $\delta^{(i)}$, for $i = 1, \ldots, s$, from a prior of δ close to the null hypothesis, instead of considering the fixed values of the null hypothesis, namely $\delta = 0$ or $\pi/2$. We take the first r of these prior values and we use $\mu_2^{(i)} = (\mu_1^0 - \delta^{(i)}) \bmod \pi$ for generating $\boldsymbol{\theta}^{(i)}$, for $i = 1, \ldots, r$. Repeating this three times, we obtained the three means of the three sequences of $r = 10^4$ Bayes factors

$$\bar{B}_{01}^{(1)} = 3.044, \bar{B}_{01}^{(2)} = 2.986 \text{ and } \bar{B}_{01}^{(3)} = 2.950.$$

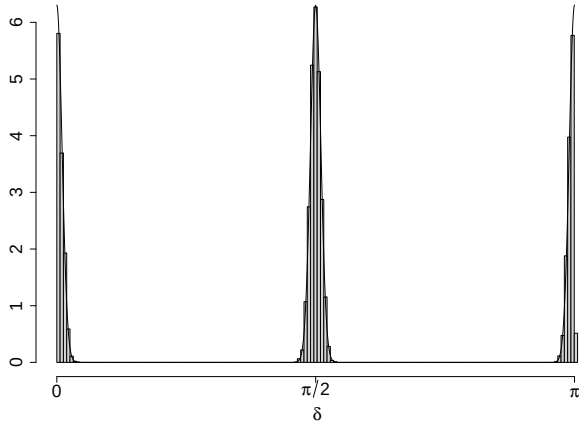
In Figure 4.5a we can find the boxplots of the three respective generated sequences. The asymptotic normal confidence interval for the mean value of the Bayes factors, at level 0.95 and based on the three generated sequences, is

$$(2.974, 3.013).$$

The conclusion is that the sample provides positive evidence of axial symmetry, even though to some smaller extent only. The same was found in Case D1.



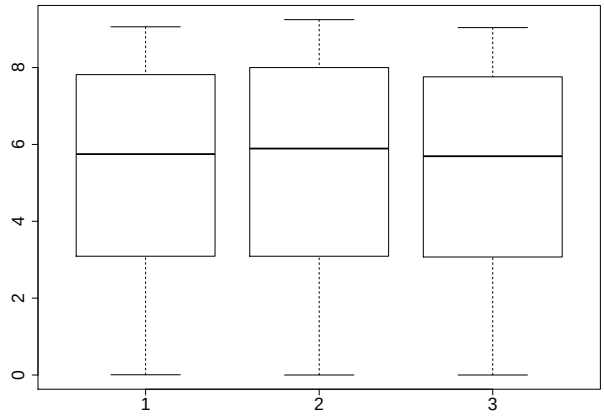
(a) 3 Boxplots of the 3 sets of 10^4 simulated Bayes factors.



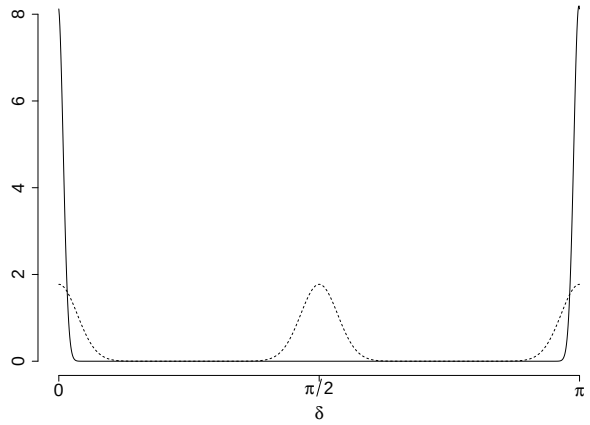
(b) Prior density of δ with histogram of 10^4 generated values.

Figure 4.5: Results of Case S1.

Case S2 We generated prior values of δ from the same mixture, however with smaller concentration hyperparameter $\tau = 20$. We found $p_0 = p_{\pi/2} = 0.088$. We generated the elements of the sample vector $\boldsymbol{\theta}^{(i)}$ with fixed value $\mu_2^0 = 0$, thus from $\text{GvM}(\mu_1^0, \mu_2^0, \kappa_1^0, \kappa_2^0)$, with $\mu_1^0 = \pi, \mu_2^0 = 0, \kappa_1^0 = 0.1, \kappa_2^0 = 5.5$, for $i = 1, \ldots, r$. We repeated this experiment



(a) 3 Boxplots of the 3 sets of 10^4 simulated Bayes factors.



(b) Prior density (dashed line) and posterior density (continuous line) of δ. The posterior is based on one generated sample.

Figure 4.6: Results of Case S2.

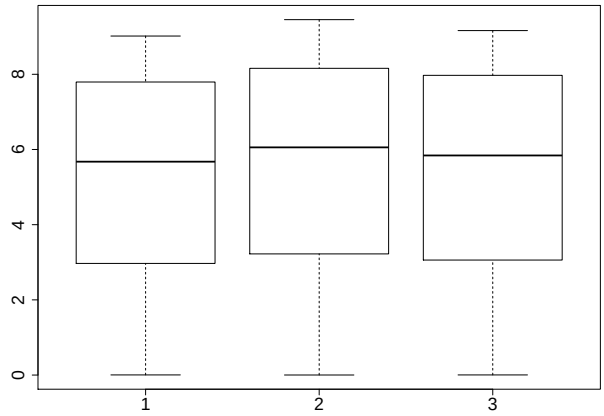
three times and obtained three sequences of Bayes factors, with respective mean values

$$\bar{B}_{01}^{(1)} = 5.322, \bar{B}_{01}^{(2)} = 5.439 \text{ and } \bar{B}_{01}^{(3)} = 5.282.$$

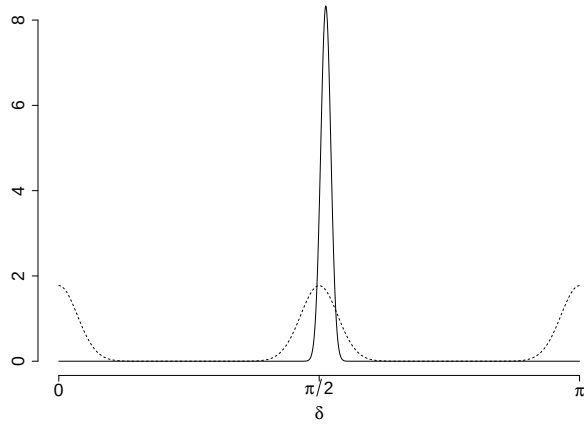
The boxplots of the three sequences of Bayes factors can be found in Figure 4.6a. After aggregating the three sequences, we obtained the asymptotic normal confidence interval at level 0.95 for the mean value of the Bayes factors given by

$$(5.317, 5.378).$$

The Bayes factor is thus larger than it was in Case S1, so that the sample has brought substantial evidence of axial symmetry. Figure (4.6b) shows the prior density of δ



(a) 3 Boxplots of the 3 sets of 10^4 simulated Bayes factors.



(b) Prior density (dashed line) and posterior density (continuous line) of δ. The posterior is based on one generated sample.

Figure 4.7: Results of Case S3.

(dashed line) and a posterior density of δ (continuous line) that is based on one of the previously generated samples. The posterior is highly concentrated around 0 and provides a stronger belief about symmetry than the prior.

Case S3 We retain the prior of δ of Case S2 but we generate samples $\boldsymbol{\theta}^{(i)}$, for $i = 1, \ldots, r$, with $\mu_1^0 = \pi, \mu_2^0 = \pi/2, \kappa_1^0 = 0.1$, and $\kappa_2^0 = 5.5$, thus from another symmetric GvM distribution. The computed values $p_0 = p_{\pi/2} = 0.088$ are the same of Case S2. We generated three sequences of $r = 10^4$ Bayes factors. The three respective boxplots of the three sequences can be found in Figure 4.7a. The three respective means of these three sequences are

$$\bar{B}_{01}^{(1)} = 5.267, \bar{B}_{01}^{(2)} = 5.553 \text{ and } \bar{B}_{01}^{(3)} = 5.395.$$

By aggregating the three sequences, we obtained the asymptotic normal confidence interval at level 0.95 for the mean of the Bayes factors given by

$$(5.374, 5.436).$$

We find substantial evidence of axial symmetry. Figure 4.7b displays the prior density of δ (dashed line) and a posterior density of δ (continuous line) that is based on one of the previously generated samples. The posterior is highly concentrated around $\pi/2$ and possesses less uncertainty about symmetry than the prior.

4.2.2 Test of vM axial symmetry of GvM

Now we have $H_0 : \kappa_2 = 0$, with fixed $\mu_1^0 = \pi$, $\mu_2^0 = \pi/2$ and $\kappa_1^0 = 0.1$. The prior distribution of κ_2 is uniform over $[0, 1/2]$ and the sample $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$ is generated from the $\text{vM}(\mu_1^0, \kappa_1^0)$ distribution. The prior probability of H_0 under the perturbation is $p_0 = 0.1$. We generated three sequences of $r = 10^4$ Bayes factors: their boxplots are shown in Figure 4.8. In these boxplots we removed a very small number of large values, in order to improve the readability. The three means of the three generated sequences are

$$B_{01}^{(1)} = 3.284, \ B_{01}^{(2)} = 3.380 \ \text{ and } B_{01}^{(3)} = 3.241,$$

where the very large values that were eliminated from the boxplots have been considered in the calculations of these means. After aggregating these three sequences, we obtained the following asymptotic normal confidence interval for the mean value of the Bayes factors at level 0.95,

$$(3.268, 3.335).$$

There is a positive evidence of symmetry although rather limited. The amount of evidence is similar to the cases D1 and S1: in all these studies, the prior is much concentrated around the null hypothesis (here $\kappa_2 = 0$), so that the data have increased the evidence of the null hypothesis only to some limited extend.

4.2.3 Summary

Table 4.2 summarizes the simulation results that we obtained for the three tests and for the various cases.

4.2.4 Application to real data

The proposed Bayesian tests have been so far applied to simulated data. This section provides the application of the test of no shift between cosines of Section 4.1.2 and of axial symmetry of Section 4.1.3 to real data obtained from the study "ArticRIMS" (A Regional, Integrated Hydrological Monitoring System for the Pan Arctic Land Mass) available

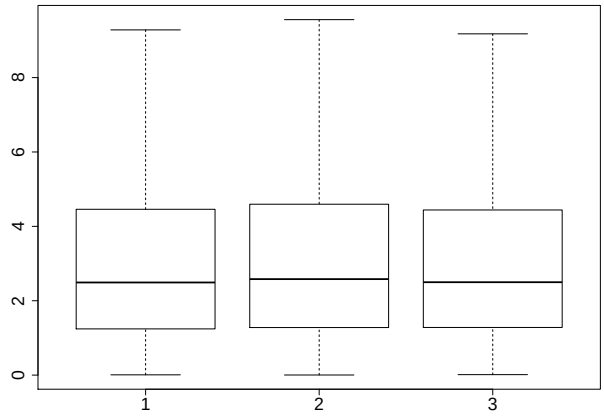


Figure 4.8: 3 boxplots of the 3 sets of 10^4 simulated Bayes factors.

H_0	case	confidence interval for B_{01}	evidence for H_0
no shift between cosines	D1	(2.937, 2.976)	positive
	D1'	(3.901, 3.945)	positive
	D2	(5.477, 5.541)	substantial
axial symmetry	S1	(2.974, 3.013)	positive
	S2	(5.317, 5.378)	substantial
	S3	(5.374, 5.436)	substantial
vM axial symmetry	–	(3.268, 3.335)	positive

Table 4.2: Summary of the simulation study.

at `http://rims.unh.edu`. The Arctic climate, its vulnerability, its relation with the terrestrial biosphere and with the recent global climate change are the subjects under investigation. For this purpose, various meteorological variables such as temperature, precipitation, humidity, radiation, vapour pressure, speed and directions of winds are measured at four different sites.

We consider wind directions measured at the site "Europe basin" and from January to December 2005. After removal of few influential measurements, the following maximum likelihood estimators are obtained: $\hat{\mu_1} = 4.095$, $\hat{\mu}_2 = 0.869$, $\hat{\kappa}_1 = 0.304$, $\hat{\kappa}_2 = 1.910$ and thus $\hat{\delta} = (\hat{\mu_1} - \hat{\mu}_2) \bmod \pi = 0.084$. The histogram of the sample together with the GvM density with theses values of the parameters are given in Figure 4.9.

For the test of no shift between cosines, the Monte Carlo integral (4.8) is computed with $s = 10^6$ values of δ generated from the prior $\mathrm{vM}_2(\nu, \tau)$, with $\nu = 0$ and $\tau = 300$. We consider $\varepsilon = 0.18$: as mentioned in Section 4.1.1, a substantial value is desirable in the practice. We obtain the Bayes factor $B_{01} = 2.550$; cf. Table 4.3.

For the test of symmetry, the prior of δ is the mixture of two vM of order two, i.e. $\xi\,\mathrm{vM}_2(\nu_1, \tau) + (1-\xi)\,\mathrm{vM}_2(\nu_2, \tau)$, with $\nu_1 = 0$, $\nu_2 = \pi/2$, $\tau = 300$ and $\xi = 0.5$. Monte

Carlo integration is done with $s = 10^6$ generations from this prior. We consider $\varepsilon = 0.18$ and obtain the Bayes factor $B_{01} = 2.252$; cf. Table 4.3.

The values of the two Bayes factors of Table 4.3 show positive evidence for the respective null hypotheses.

H_0	B_{01}	evidence for H_0
no shift between cosines	2.550	positive
axial symmetry	2.252	positive

Table 4.3: Bayes factors for wind directions data.

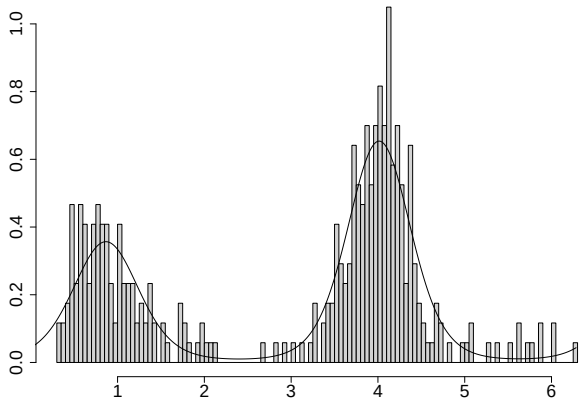


Figure 4.9: Histogram of wind directions data together with $\mathrm{GvM}(\hat{\mu}_1, \hat{\mu}_2, \hat{\kappa}_1, \hat{\kappa}_2)$ density.

4.3 Conclusion

This chapter introduces three Bayesian tests relating to the symmetry of the GvM model. The first test is about the significance of the shift parameter between the cosines of frequency one and two ($H_0 : \delta = 0$). The second test is about axial symmetry ($H_0 : \delta = 0$ or $\delta = \pi/2$). The third test is about vM symmetry ($H_0 : \kappa_2 = 0$). These tests are obtained by the technique of probability perturbation. Simulation studies show the effectiveness of these three tests, in the sense that when the sample is coherent with the null hypothesis, then the Bayes factors are typically large. Applications to real data are also shown.

Due to computational limitations, we consider null hypotheses of symmetry that concern one parameter only. The null hypotheses considered are about one or two distinct values of the parameter of interest, with all remaining parameters fixed. Composite null hypotheses that allow for unknown nuisance parameters, would require one additional dimension of Monte Carlo integration for each unknown parameter, in the computation

of the marginal distribution. The computational burden would rise substantially and the Monte Carlo study, with two levels of nested generations, would become very difficult. But the essentially simple null hypotheses considered are relevant in the practice. It can happen that nuisance parameters have been accurately estimated and the question of interest is really about the the parameter δ and axial symmetry. In the example of Section 4.2.4, we want to know if wind direction is axially symmetric within the GvM model. The values of the concentrations and of the axial direction are of secondary importance.

One could derive other Bayesian tests for the GvM model: a Bayesian test of bimodality can be found in Chapter 5. We can also note that Navarro et al. (2017) introduced an useful multivariate GvM distribution for which similar Bayesian tests could be investigated.

Chapter 5

An algebraic analysis of the bimodality of the generalized von Mises distribution[†]

The GvM circular distribution (of order 2) can be either unimodal or bimodal. This chapter provides an algebraic analysis of the uni- and bimodality of the GvM distribution. It completes the analyses initiated by Gatto and Jammalamadaka (2007), Gatto (2008) and Pfyffer and Gatto (2013). The bimodality of the GvM is formulated in terms of the number of real roots of a quartic polynomial. A detailed analysis of this quartic is presented in this chapter, in particular regarding the nature of its roots (real, in $[-1, 1]$, complex or with multiplicity). For the study of the bimodality, it is sufficient to reparametrize the GvM density in terms of δ and $\rho = \kappa_1/4\kappa_2$ alone. The main result of this chapter is the determination of the partition of the parametric set of (ρ, δ) in terms of the subset giving bimodality, denoted by $\mathcal{W}$, and its complement $\mathcal{W}^c$, giving unimodality. This partition is shown is Figure 5.4a and Figure 5.5. These results are used in Chapter 6 for Bayesian inference on the bimodality of the GvM distribution.

This chapter is organized as follows. Section 5.1 reformulates the problem of bimodality in terms of the roots of a quartic. The nature of its roots are analysed in Section 5.2. In particular, Section 5.2.1 studies the discriminant of the quartic. The algebraic results of Section 5.2.2 allows to determine the number of extrema of the GvM density. Section 5.3 summarizes the results and provides a practical graph of the parametric subset of the GvM distribution that gives bimodality. Various numerical illustrations, obtained with the software `R`, are presented.

A scheme of the structure of this chapter is given in Figure 5.1.

[†]Salvador and Gatto (2022a)

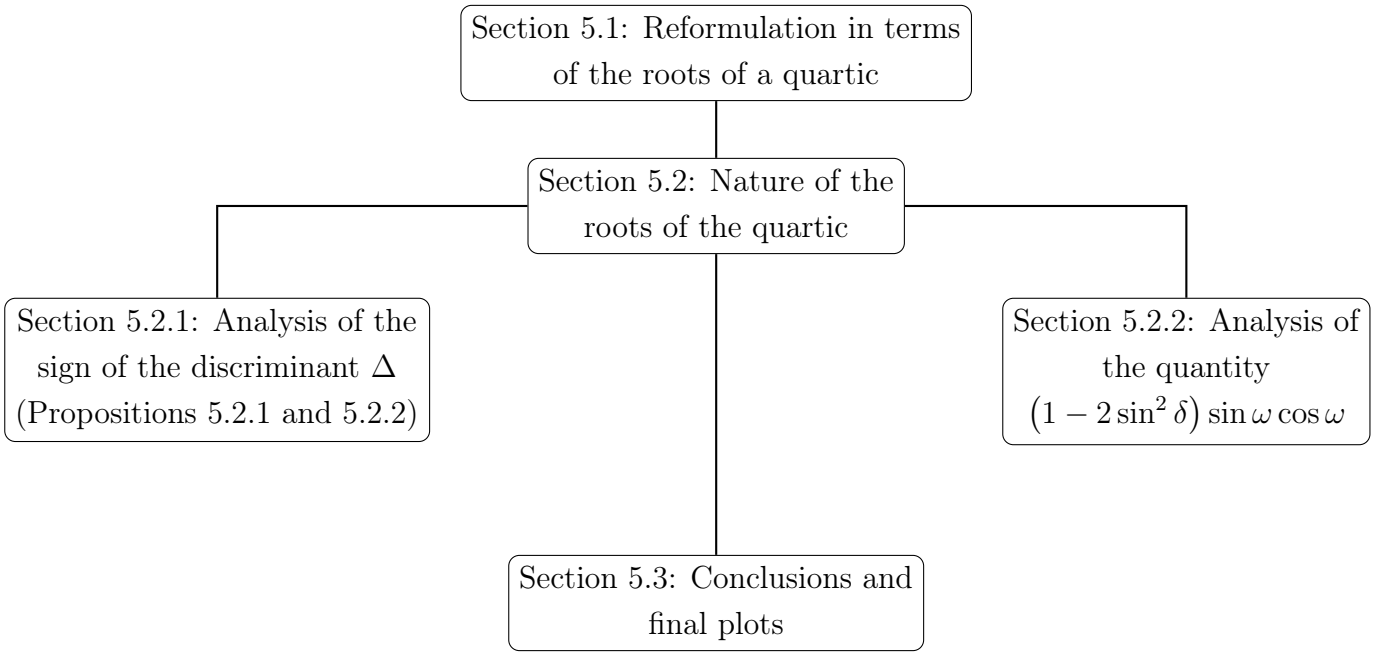


Figure 5.1: Float chart of the scheme of Chapter 5.

5.1 Reformulation in terms of roots of a quartic

Pfyffer and Gatto (2013) suggest the following approach to the study of the extrema of the GvM density. Let $\omega = \theta - \mu_1$ and $\delta = (\mu_1 - \mu_2) \bmod \pi$. The extrema points of the GvM density (2.4) are roots of $df\left(\omega|0,\delta,\kappa_1,\kappa_2\right)/d\omega$. The exponent of the GvM can be re-expressed as

$$g(\omega) = \kappa_1 \cos\omega + \kappa_2 \cos 2\left(\omega+\delta\right).$$

The points of extrema satisfy $dg(\omega)/d\omega = 0$, viz.

$$-\kappa_1 \sin\omega - 2\kappa_2 \sin 2\left(\omega+\delta\right) = 0. \tag{5.1}$$

By expanding cosine and sines, (5.1) can be re-expressed as

$$\left(1-2\sin^2\delta\right)\sin\omega\cos\omega - 2\sin\delta\cos\delta\sin^2\omega + \rho\sin\omega + \sin\delta\cos\delta = 0. \tag{5.2}$$

With the change of variable $x = \sin\omega$, (5.2) can be re-expressed as

$$\pm\left(1-2\sin^2\delta\right)x\sqrt{1-x^2} = 2\sin\delta\cos\delta x^2 - \rho x - \sin\delta\cos\delta, \tag{5.3}$$

where $\rho = \kappa_1/(4\kappa_2) > 0$. By taking the square on both sides of (5.3), we obtain that the values of x that solve (5.3) are precisely the roots of the quartic

$$q_{\rho,\delta}(x) = x^4 + b_3x^3 + b_2x^2 + b_1x + b_0, \tag{5.4}$$

where

$$b_3 = -4\rho\sin\delta\cos\delta, \quad b_2 = \rho^2 - 1, \quad b_1 = 2\rho\sin\delta\cos\delta, \text{ and } b_0 = \sin^2\delta\cos^2\delta.$$

If $\bar{x}$ is a root of $q_{\rho,\delta}(x)$, then it can be transformed back to angular variable through

$$\omega_1 = (\arcsin \bar{x}) \bmod 2\pi \quad \text{and} \quad \omega_2 = (\pi - \arcsin \bar{x}) \bmod 2\pi. \tag{5.5}$$

In general, only one of these two values is a solution of (5.2). The cases where both of these angular roots are admissible are studied in detail in Section 5.2.2. Finally, going back into the original abscissa, we obtain the extrema points of the GvM given by

$$\theta_j = (\omega_j + \mu_1) \bmod 2\pi, \quad \text{for } j = 1, 2.$$

Bimodality happens if and only if the density of the GvM (2.4) has four distinct extrema points and this holds when the quartic (5.4) has four real roots. Other particular cases can happen and they are analysed in this chapter. Because of the change of variable $x = \sin\omega$, we are interested in the roots in the interval $[-1, 1]$. We analyse the cases where the roots of $q_{\rho,\delta}(x)$ are multiple, complex and real but not in the interval $[-1, 1]$. In order to determine the nature of the roots, we summarize the results of Rees (1922) in the following theorem.

Theorem 5.1.1 (Nature of roots of quartic)**.** *Let* $q_{\rho,\delta}(x) = x^4 + b_3x^3 + b_2x^2 + b_1x + b_0$*, where* $b_0, b_1, b_2, b_3 \in \mathbb{R}$*. Define*

$$\begin{aligned}\Delta = {} & 256b_0^3 - 192b_3b_1b_0^2 - 128b_2^2b_0^2 + 144b_2b_1^2b_0 - 27b_1^4 + 144b_3^2b_2b_0^2 - 6b_3^2b_1^2b_0 \\ & - 80b_3b_2^2b_1b_0 + 18b_3b_2b_1^3 + 16b_2^4b_0 - 4b_2^3b_1^2 - 27b_3^4b_0^2 \\ & + 18b_3^3b_2b_1b_0 - 4b_3^3b_1^3 - 4b_3^2b_2^3b_0 + b_3^2b_2^2b_1^2,\end{aligned}$$

and

$$\begin{aligned}P &= 8b_2 - 3b_3^2, \\ R &= b_3^3 - 4b_3b_2 + 8b_1, \\ \Delta_0 &= b_2^2 - 3b_3b_1 + 12b_0, \\ D &= 64b_0 - 16b_2^2 + 16b_3^2b_2 - 16b_3b_1 - 3b_3^4.\end{aligned} \tag{5.6}$$

Then the roots of $q_{\rho,\delta}(x)$ *have the following nature.*

1. $\Delta < 0 \Leftrightarrow$ *two simple real roots and two complex conjugate roots.*
2. $\Delta > 0$*:*
   (a) $P < 0 \wedge D < 0 \Leftrightarrow$ *four simple real roots;*
   (b) $P > 0 \vee D > 0 \Leftrightarrow$ *two pairs of complex conjugate roots.*
3. $\Delta = 0$*:*
   (a) $P < 0 \wedge D < 0 \wedge \Delta_0 \neq 0 \Leftrightarrow$ *one double real root and two simple real roots;*

(b) $D > 0 \vee \Big(P > 0 \wedge (D \neq 0 \vee R \neq 0) \Big) \Leftrightarrow$ *one double real root and two complex conjugate roots;*

(c) $\Delta_0 = 0 \wedge D \neq 0 \Leftrightarrow$ *a triple real root and a simple real root;*

(d) $D = 0$*:*

i. $P < 0 \Leftrightarrow$ *two double real roots;*

ii. $P > 0 \wedge R \neq 0 \Leftrightarrow$ *two double complex conjugate roots;*

iii. $\Delta_0 = 0 \Leftrightarrow$ *quadruple real root equals to* $-\frac{b_3}{4}$.

These are all the possible cases.

Similar considerations on the nature of the roots of a quartic can be found in Dickson (1917, p. 45), Garver (1933) and Lazard (1988).

Theorem 5.1.2 (Roots of quartic)**.** *The roots of* $q_{\rho,\delta}(x)$ *can be obtained using the method of Ferrari as follows:*

$$\begin{aligned} \bar{x}_{1,2} &= -\frac{b_3}{4} - S \pm \frac{1}{2}\sqrt{-4S^2 - 2p + \frac{t}{S}} \quad \textit{and} \\ \bar{x}_{3,4} &= -\frac{b_3}{4} + S \pm \frac{1}{2}\sqrt{-4S^2 - 2p - \frac{t}{S}}, \end{aligned} \tag{5.7}$$

where

$$p = \frac{8b_2 - 3b_3^2}{8},\ t = \frac{b_3^3 - 4b_3 b_2 + 8b_1}{8},$$

$$S = \frac{1}{2}\sqrt{-\frac{2}{3}p + \frac{1}{3}\left(Q + \frac{\Delta_0}{Q}\right)} \quad \textit{and } Q = \sqrt[3]{\frac{\Delta_1 + \sqrt{\Delta_1^2 - 4\Delta_0^3}}{2}}, \tag{5.8}$$

with

$$\Delta_1 = 2b_2^3 - 9b_3 b_2 b_1 + 27b_3^2 b_0 + 27b_1^2 - 72b_2 b_0$$

and Δ_0 *given by* (5.6).

Alternative formulae for roots of the quartic polynomial can be found in Carpenter (1966) and Yacoub and Fraidenraich (2012).

Remark 5.1.3. *We can re-express the terms appearing in Theorem 5.1.1 directly as functions of* δ *and* ρ *as follows,*

$$\Delta = -256\cos^2\delta\sin^2\delta\left(\cos^2\delta - 1/2\right)^4\left(27\rho^4\cos^2\delta\left(1-\cos^2\delta\right) + (\rho-1)^3(\rho+1)^3\right),$$

$$P = 48\rho^2\cos^2\delta\left(\cos^2\delta - 1\right) + 8(\rho^2 - 1), \tag{5.9}$$

$$R = 64\rho^3\cos\delta\sin\delta\left(\cos^2\delta - 1/2\right)^2,$$

$$\Delta_0 = 12\left(2\rho^2+1\right)\cos^2\delta\left(1-\cos^2\delta\right) + \left(\rho^2-1\right)^2 \quad \textit{and} \tag{5.10}$$

$$D = -16\left(2\cos^2\delta - 1\right)^2\left(12\rho^4\cos^2\delta\left(\cos^2\delta - 1\right) + \left(\rho^2-1\right)^2\right). \tag{5.11}$$

Remark 5.1.4. *When $\Delta > 0$, Theorem 5.1.1, part 2, tells that $q_{\rho,\delta}(x)$ has either non-real roots or four simple real roots, depending on the signs of P and D. Consider the case of four simple real roots. It follows from*

$$\Delta_1^2 - 4\Delta_0^3 = -27\Delta < 0$$

that $\sqrt{\Delta_1^2 - 4\Delta_0^3}$ is purely imaginary. With this, (5.8) tells that Q, and therefore S, are non-real complex numbers. Hence the roots in (5.7) are indeed real, although they are expressed in terms of complex numbers.[†] *However S can be re-expressed in a purely real way as*

$$S = \frac{1}{2}\sqrt{-\frac{2}{3}p + \frac{2}{3}\sqrt{\Delta_0}\cos\frac{\psi}{3}}, \quad \textit{where } \psi = \arccos\frac{\Delta_1}{2\sqrt{\Delta_0^3}}. \tag{5.12}$$

In order to show (5.12), we first note that (5.8) tells that $z = Q^3$ can be written as $z = a + ib$, where

$$a = \frac{\Delta_1}{2} \quad \textit{and} \quad b = \frac{\sqrt{4\Delta_0^3 - \Delta_1^2}}{2}.$$

The modulus of z is

$$r = \sqrt{a^2 + b^2} = \sqrt{\frac{\Delta_1^2}{4} + \frac{4\Delta_0^3 - \Delta_1^2}{4}} = \sqrt{\Delta_0^3}$$

and the argument of z is

$$\psi = \arccos\frac{a}{r} = \arccos\frac{\Delta_1}{2\sqrt{\Delta_0^3}},$$

given that z belongs to the upper complex half plane. Thus,

$$Q = \sqrt[3]{a + ib} = \sqrt[3]{r}\left(\cos\frac{\psi}{3} + i\sin\frac{\psi}{3}\right) = \sqrt{\Delta_0}\mathrm{e}^{i\frac{\psi}{3}}$$

and

$$|Q| = \sqrt{\Delta_0}.$$

We justify (5.12) by

$$Q + \frac{\Delta_0}{Q} = \sqrt{\Delta_0}\mathrm{e}^{i\frac{\psi}{3}} + \frac{\Delta_0}{\sqrt{\Delta_0}\mathrm{e}^{i\frac{\psi}{3}}} = 2\sqrt{\Delta_0}\cos\frac{\psi}{3}.$$

5.2 The nature of the roots of the quartic

After reformulating the problem of bimodality in terms of roots of a quartic, we now study the nature of these roots. The first important quantity is the discriminant Δ and Section 5.2.1 analyses $\mathrm{sgn}\Delta$, according to $\rho \in (0, 1/2]$, $\rho \in (1/2, 1]$, and $\rho \in (1, +\infty)$. The second important quantity is $\left(1 - 2\sin^2\delta\right)\sin\omega\cos\omega$, appearing in (5.2), and Section 5.2.2 analyses the roots of this quantity.

[†]When solving equations of degree three or higher, it can happen that the roots are expressed in terms of complex numbers, although these roots are only real. This situation is referred within algebra to as "casus irriducibilis" and it was first presented in 1572 by Bombelli; cf. e.g. p. 17 of Cox (2012).

5.2.1 Analysis of Δ

Theorem 5.1.1 tells that Δ plays a major role in the study of the nature of the roots of the quartic $q_{\rho,\delta}(x)$. This section presents a study of the sign of Δ, with particular attention to the number of solutions of $\Delta = 0$.

Remark 5.1.3 allows us to rewrite Δ as

$$\Delta = 27\rho^4 \left[16\cos\delta\sin\delta\left(\cos^2\delta - 1/2\right)^2\right]^2 \cdot \Lambda$$

where

$$\Lambda = \cos^4\delta - \cos^2\delta - \frac{\left(\rho^2-1\right)^3}{27\rho^4}. \tag{5.13}$$

In particular, $27\rho^4\left[16\cos\delta\sin\delta\left(\cos^2\delta - 1/2\right)^2\right]^2 \geq 0$, with equality iff $\delta = 0, \pi/4, \pi/2, 3\pi/4$.

We now focus on the study of Λ. Using the change of variable $t = \cos^2\delta$, (5.13) becomes

$$\Lambda^* = t^2 - t - \frac{(\rho^2-1)^3}{27\rho^4}, \tag{5.14}$$

which is a quadratic polynomial w.r.t. t. The discriminant of (5.14) is given by

$$\begin{aligned}
\tilde{\Delta} &= 1 + \frac{4}{27\rho^4}\left(\rho^2-1\right)^3 \\
&= \frac{27\rho^4 + 4\rho^6 - 12\rho^4 + 12\rho^2 - 4}{27\rho^4} \\
&= \frac{4\rho^6 + 15\rho^4 + 12\rho^2 - 4}{27\rho^4} \\
&= (\rho - 1/2)\cdot \underbrace{\left(\frac{4\rho^5 + 2\rho^4 + 16\rho^3 + 8\rho^2 + 16\rho + 8}{27\rho^4}\right)}_{>0}.
\end{aligned} \tag{5.15}$$

Hence, we have the three following cases:

(a) For $\rho < 1/2$, $\tilde{\Delta} < 0$: Λ^* does not have any real roots;

(b) For $\rho = 1/2$, $\tilde{\Delta} = 0$: Λ^* has one double real root;

(c) For $\rho > 1/2$, $\tilde{\Delta} > 0$: Λ^* has two simple real roots.

Because of the change of variable $t = \cos^2\delta \geq 0$, we are only interested in the cases (b) and (c) where Λ^* has non-negative real roots.

Let us consider case (c) of two real roots t_1 and t_2. From (5.14), we have

$$t_1 + t_2 = 1 \quad \text{and} \quad t_1 \cdot t_2 = -\frac{(\rho^2-1)^3}{27\rho^4}.$$

The first equation tells us that the two roots cannot be both negative. We distinguish three situations.

(c.i) When $\rho^2 - 1 < 0$, thus here $1/2 < \rho < 1$, the second equation tells $t_1 \cdot t_2 > 0$ and thus both roots are indeed positive. We can now obtain the roots of (5.13) as follows,

$$\cos^2\delta = t_1 \Rightarrow \delta_{1,2} = \arccos(\pm\sqrt{t_1}) \quad \text{and} \quad \cos^2\delta = t_2 \Rightarrow \delta_{3,4} = \arccos(\pm\sqrt{t_2}).$$

In particular, since $t_1 + t_2 = 1$ and $t_1, t_2 > 0$, the four roots $\delta_{1,2}$ and $\delta_{3,4}$ are defined over $[0, \pi)$, as required. One can indeed show that the constant term of (5.13), i.e. $-\left(\rho^2 - 1\right)^3/(27\rho^4)$, is strictly decreasing in $(1/2, 1)$ so that $\delta_1, \delta_2, \delta_3, \delta_4$ cannot take the values $0, \pi/4, \pi/2, 3\pi/4$, that are obtained with $\rho = 1/2$ and $\rho = 1$ (see cases (b) and (c.iii) respectively, below). Thus Δ has eight simple roots.

(c.ii) When $\rho > 1$ the product of the roots of Λ^* is negative hence one and only one of t_1 or t_2 is negative. Let t_2 be the negative root, without loss of generality. This means that $t_1 = 1 - t_2 > 1$ and we cannot solve $\delta = \arccos(\pm\sqrt{t_1})$ since the function arccos is not well defined for argument grater than 1: In this case, the factor Λ is irreducible.

(c.iii) For $\rho = 1$, (5.14) reduces to $t^2 - t = t(t-1) = 0$ with roots $t_1 = 1$, $t_2 = 0$. So we obtain

$$\cos^2\delta = 1 \Rightarrow \cos\delta = \pm 1 \Rightarrow \delta = 0, \pi,$$
$$\cos^2\delta = 0 \Rightarrow \cos\delta = 0 \Rightarrow \delta = \pi/2.$$

But $\delta = \pi$ is not admissible and so the roots are $\delta = 0, \pi/2$.

Now we analyse case (b) of one double real root. We simplify (5.14) to

$$t^2 - t + 1/4 = (t - 1/2)^2 = 0$$

hence $t = 1/2$ is the (double) real root of (5.14) and by the change of variables we obtain

$$\cos^2\delta = 1/2 \Rightarrow \cos\delta = \pm\sqrt{2}/2,$$

and to conclude that $\delta = \pi/4, 3\pi/4$ are roots of $\Lambda = 0$.

We have thus shown Proposition 5.2.1.

Proposition 5.2.1 (Roots of Δ)**.** *The roots of the discriminant Δ are the following.*

1. *For $\rho \leq 1/2 \vee \rho \geq 1$, Δ has four simple roots, namely $\delta = 0, \pi/4, \pi/2, 3\pi/4$.*
2. *For $1/2 < \rho < 1$, Δ has four other simple roots in addition to $\delta = 0, \pi/4, \pi/2, 3\pi/4$, that are*

$$\delta_1 = \arccos(\sqrt{t_1}) \quad \textit{and} \quad \delta_2 = \arccos(-\sqrt{t_1}) = \pi - \delta_1,$$
$$\delta_3 = \arccos(\sqrt{t_2}) \quad \textit{and} \quad \delta_4 = \arccos(-\sqrt{t_2}) = \pi - \delta_3,$$

with

$$t_{1,2} = \frac{1 \pm \sqrt{\tilde{\Delta}}}{2} \quad \textit{and} \quad \tilde{\Delta} = (\rho - 1/2) \cdot \left(\frac{4\rho^5 + 2\rho^4 + 16\rho^3 + 8\rho^2 + 16\rho + 8}{27\rho^4} \right).$$

Moreover we have $0 < \delta_1 < \frac{\pi}{4} < \delta_3 < \frac{\pi}{2} < \delta_4 < \frac{3\pi}{4} < \delta_2 < \pi$.

We can now show the following results regarding the sign of Δ.

Proposition 5.2.2 (Sign of Δ)**.** *We can distinguish the following cases for the sign of* Δ.

(1) For $\rho \leq 1/2$, $\Delta \geq 0$ *for all* $\delta \in [0, \pi)$.

(2) For $\rho \geq 1$, $\Delta \leq 0$ *for all* $\delta \in [0, \pi)$.

(3) For $1/2 < \rho < 1$, $\Delta \geq 0$ *for all* $\delta \in [0, \delta_1] \cup [\delta_3, \delta_4] \cup [\delta_2, \pi)$.

Proof of Proposition 5.2.2.

(1) For $\rho = 1/2$, we have $\Lambda = (\cos^2 \delta - 1/2)^2 \geq 0$, for all $\delta \in [0, \pi)$. On the other hand for $\rho < 1/2$, $\tilde{\Delta} < 0$ and Λ is irreducible. Hence it has constant sign for all values of δ. In particular for $\delta = 0$ we have $\Lambda = -(\rho^2 - 1)^3 / (27\rho^4) > 0$. We conclude that $\Lambda > 0$ for all $\delta \in [0, \pi)$, thus $\Delta \geq 0$ for all $\delta \in [0, \pi)$.

(2) For $\rho = 1$ we have $\Lambda = \cos^2 \delta(\cos^2 \delta - 1) \leq 0$ while for $\rho > 1$, we have seen in (c.ii) that Λ is irreducible, hence it has constant sign. For $\delta = 0$, $\Lambda < 0$ which means that $\Lambda < 0$ for all $\delta \in [0, \pi)$. Hence we conclude that $\Delta \leq 0$ for all $\delta \in [0, \pi)$.

(3) We get from (c.i) that $\Lambda^* \geq 0$ iff $t \leq t_2 \vee t \geq t_1$. With the change of variables $t = \cos^2 \delta$ and with the fact that arccos is strictly decreasing, we get that $\Lambda \geq 0$ for all $\delta \in [0, \delta_1] \cup [\delta_3, \delta_4] \cup [\delta_2, \pi)$. Hence we conclude $\Delta \geq 0$ for all $\delta \in [0, \delta_1] \cup [\delta_3, \delta_4] \cup [\delta_2, \pi)$.

□

Analysis of $\Delta \neq 0$

We can now analyse the case $\Delta \neq 0$. We have $\delta \neq 0, \pi/4, \pi/2, 3\pi/4$ and depending of the values on ρ, $\delta \neq \delta_1, \delta_2, \delta_3, \delta_4$.

(i) For $\rho \leq 1/2$, we have from Proposition 5.2.2 that $\Delta > 0$. It follows from (5.9), using $\cos^2 \delta < 1$ and $\rho^2 - 1 < 0$ for all $\rho \leq 1/2$, that $P < 0$. Moreover, we have that $D < 0$ for all $\rho \leq 1/2$.

This can be proved as follows. Let us re-write (5.11) as $D = -16\left(2\cos^2\delta - 1\right)^2 \cdot \Lambda'$ with $\Lambda' = 12\rho^4\cos^2\delta\left(\cos^2\delta - 1\right) + \left(\rho^2 - 1\right)^2$. Now let us study the sign of Λ' for $\rho \leq 1/2$: $\cos^2\delta\left(\cos^2\delta - 1\right)$ has a minimum at $\cos^2\delta = 1/2$ hence when $\delta = \pi/4, 3\pi/4$, where it assumes value $-1/4$. Hence we can say that

$$\Lambda' > -\frac{12}{4}\rho^4 + \left(\rho^2 - 1\right)^2 = -2\rho^2\left(\rho^2 + 1\right) + 1. \tag{5.16}$$

The inequality in (5.16) is strict because we are under the hypothesis $\Delta \neq 0$, thus $\delta \neq \pi/4, 3\pi/4$. The right hand side of (5.16) is a decreasing function and hence we have that

$$-2\rho^2\left(\rho^2 + 1\right) + 1 \geq -2\left(\frac{1}{2}\right)^2\left[\left(\frac{1}{2}\right)^2 + 1\right] + 1 = \frac{3}{8}, \forall\rho \leq 1/2.$$

In conclusion, $\Lambda' > 0, \forall\rho \leq 1/2$ and thus $D < 0, \forall\rho \leq 1/2$.

Hence from Theorem 5.1.1 we know what $q_{\rho,\delta}(x)$ has four real distinct roots. In order to check that all these roots are in $[-1, 1]$, we apply Sturm's theorem (due to Jacques Charles François Sturm in 1829). This theorem expresses the number of real roots of a polynomial q located in an interval, in terms of the sign changes that the Sturm's sequence of q assumes at the upper and lower bounds of the interval. A Sturm's sequence of a polynomial q is a sequence of polynomials associated to q and its first derivative q', by a variant of the Euclid's algorithm for polynomials (see Basu et al., p. 14). A reference on Sturm's theorem and sequence is Basu et al., p. 52. With the help of a computer program based on Sturm's theorem, one checks that the four real roots are in $[-1, 1]$ and that these are all cases of bimodality.

(ii) For $\rho \geq 1$, we have from Proposition 5.2.2 that $\Delta < 0$ and from Theorem 5.1.1 we know that $q_{\rho,\delta}(x)$ has two simple real roots. Each of these roots is transformed through (5.5) into one acceptable angular value. This is explained in detail in Section 5.2.2. Hence the GvM has two extrema points, and we conclude that these cases give unimodality.

(iii) For $1/2 < \rho < 1$, we have from Proposition 5.2.2 that $\Delta > 0$ for all $\delta \in (0, \delta_1) \cup (\delta_3, \delta_4) \cup (\delta_2, \pi)$, given that $\delta \neq \pi/2$ in our initial assumptions. By using (5.9) it is not hard to check that $P < 0$ for all values of δ. The sign of D in this case follows from the following general proposition.

Proposition 5.2.3. *It holds that $D < 0$ for all $\rho \in (1/2, 1)$ and for all $\delta \in [0, \delta_1] \cup [\delta_3, \delta_4] \cup [\delta_2, \pi)$.*

Proof of Proposition 5.2.3. We can re-write D as in (5.11) as follows:

$$D = -16\left(12\rho^4\right)\left(2\cos^2\delta - 1\right)^2 \cdot \Lambda_1,$$

with

$$\Lambda_1 = \cos^4\delta - \cos^2\delta + \frac{\left(\rho^2-1\right)^2}{12\rho^4}.$$

In particular $-16\left(12\rho^4\right)\left(2\cos^2\delta-1\right)^2 < 0$ for all $\delta \in [0,\pi)$ with equality for $\delta = \pi/4, 3\pi/4$. Using the change of variable $\cos^2\delta = t$ (as in Proposition 5.2.1) we obtain

$$\Lambda_1^* = t^2 - t + \frac{\left(\rho^2-1\right)^2}{12\rho^4}. \tag{5.17}$$

Let us study the sign of (5.17). The discriminant of (5.17) is

$$\tilde{\Delta}' = 1 - 4\frac{(\rho^2-1)^2}{12\rho^4} = \frac{2\rho^4+2\rho^2-1}{3\rho^4}.$$

With the change of variable $y = \rho^2 > 0$ we find that $\tilde{\Delta}' > 0$ for all $\rho \in (\rho^*, 1)$ where $\rho^* = \left(\left(-1+\sqrt{3}\right)/2\right)^{1/2} \simeq 0.6$. So we can say that

1. For $\rho \in (1/2, \rho^*)$, $\tilde{\Delta}' < 0$ and thus $\Lambda_1 > 0, \forall\delta \in [0,\pi)$. In conclusion, $D \leq 0, \forall\delta \in [0,\pi)$.
2. For $\rho = \rho^*$, $\tilde{\Delta}' = 0$ and $\Lambda_1^* = (t-t_1')^2$ where $t_1' = 1/2$. Thus we can say that $\Lambda_1 \geq 0$ for all $\delta \in [0,\pi)$. Hence $D \leq 0$ for all $\delta \in [0,\pi)$.
3. For $\rho \in (\rho^*, 1)$, $\tilde{\Delta}' > 0$ and the roots of (5.17) are $t_{1,2}' = \left(1 \pm \sqrt{\tilde{\Delta}'}\right)/2$. In particular, from (5.17) we notice that the product of these roots is (strictly) positive and their sum is one. Hence they are both acceptable. With similar considerations made in Proposition 5.2.2 we have $\Lambda_1 \geq 0$ for all $\rho \in (\rho^*, 1)$ and for all $\delta \in [0,\delta_1'] \cup [\delta_3', \delta_4'] \cup [\delta_2', \pi]$, where

$$\begin{aligned}
&\delta_1' = \arccos(\sqrt{t_1'}) \quad \text{and} \quad \delta_2' = \arccos(-\sqrt{t_1'}) = \pi - \delta_1', \\
&\delta_3' = \arccos(\sqrt{t_2'}) \quad \text{and} \quad \delta_4' = \arccos(-\sqrt{t_2'}) = \pi - \delta_3',
\end{aligned}$$

with

$$t_{1,2}' = \frac{1 \pm \sqrt{\tilde{\Delta}'}}{2} \quad \text{and} \quad \tilde{\Delta}' = \frac{2\rho^4+2\rho^2-1}{3\rho^4}.$$

In particular with similar considerations made in (c.i) of Section 5.2.1 we can say that $\delta_i' \neq \pi/4, 3\pi/4$ for $i = 1,2,3,4$.

Consider now δ_i for $i = 1,2,3,4$ introduced in Proposition 5.2.1. In cases 1. and 2. above, the equality $D = 0$ holds for $\delta = \pi/4, 3\pi/4$. Now again from Proposition 5.2.1 we have $\pi/4, 3\pi/4 \notin [0,\delta_1] \cup [\delta_3,\delta_4] \cup [\delta_2,\pi)$. Hence we conclude $D < 0$ for all $\rho \in (1/2, \rho^*]$ and for all $\delta \in [0,\delta_1] \cup [\delta_3,\delta_4] \cup [\delta_2,\pi)$.

Consider case 3, hence $\rho \in (\rho^*, 1)$. From what written above, we have $D \leq 0$, for all $\delta \in [0,\delta_1'] \cup [\delta_3',\delta_4'] \cup [\delta_2',\pi) \cup \{\pi/4, 3\pi/4\}$. We want to compare the values δ_i' with

δ_i for $i = 1, 2, 3, 4$. It is not hard to check that $\tilde{\Delta} > \tilde{\Delta}'$, where $\tilde{\Delta}$ is given in (5.15). In fact

$$\begin{aligned}&(\rho - 1/2) \cdot \left(\frac{4\rho^5 + 2\rho^4 + 16\rho^3 + 8\rho^2 + 16\rho + 8}{27\rho^4} \right) > \frac{2\rho^4 + 2\rho^2 - 1}{3\rho^4} \\ &\iff h(\rho) = 4\rho^6 - 3\rho^4 - 6\rho^2 + 5 > 0.\end{aligned}$$

The function $h(\rho)$ is strictly positive. In fact $\lim_{\rho \to 0^+} h(\rho) = 5$ and $\lim_{\rho \to +\infty} h(\rho) = +\infty$. Moreover, a study of the sign of the first derivative tells us that there exist only one minimum $\rho_{min} = \left(\left(1 + \sqrt{3}\right)/4 \right)^{1/2}$, such that $h(\rho_{min}) > 0$. From the fact that $\tilde{\Delta} > \tilde{\Delta}'$ we can say that $t_1 > t_1'$ and $t_2 < t_2'$, where t_1, t_2 are given in Propositions 5.2.1. Now, since arccos is decreasing, we have that

$$\begin{aligned}&\delta_1 = \arccos\sqrt{t_1} < \arccos\sqrt{t_1'} = \delta_1', \quad \text{and } \delta_2 = \pi - \delta_1 > \pi - \delta_2' = \delta_2' \\ &\delta_3 = \arccos\sqrt{t_2} > \arccos\sqrt{t_2'} = \delta_3', \quad \text{and } \delta_4 = \pi - \delta_3 < \pi - \delta_3' = \delta_4'.\end{aligned} \tag{5.18}$$

From (5.18) and from the fact that $\delta_1 < \delta_3 < \delta_4 < \delta_4$ and in the same way that $\delta_1' < \delta_3' < \delta_4' < \delta_4'$, we can say that $[0, \delta_1] \cup [\delta_3, \delta_4] \cup [\delta_2, \pi)$ is a proper subset of $[0, \delta_1'] \cup [\delta_3', \delta_4'] \cup [\delta_2', \pi)$. This is visually confirmed in Figure 5.2. Moreover from Proposition 5.2.1 we can say $\pi/4, 3\pi/4 \notin [0, \delta_1] \cup [\delta_3, \delta_4] \cup [\delta_2, \pi)$. Thus $D \leq 0$ for all $\rho \in (\rho^*, 1)$ and for all $\delta \in [0, \delta_1'] \cup [\delta_3', \delta_4'] \cup [\delta_2', \pi] \cup \{\pi/4, 3\pi/4\}$ implies that $D < 0$ for all $\rho \in (\rho^*, 1)$ and for all $\delta \in [0, \delta_1] \cup [\delta_3, \delta_4] \cup [\delta_2, \pi)$.

We conclude that $D < 0$ for all $\rho \in (1/2, 1)$ and for all $\delta \in [0, \delta_1] \cup [\delta_3, \delta_4] \cup [\delta_2, \pi)$. □

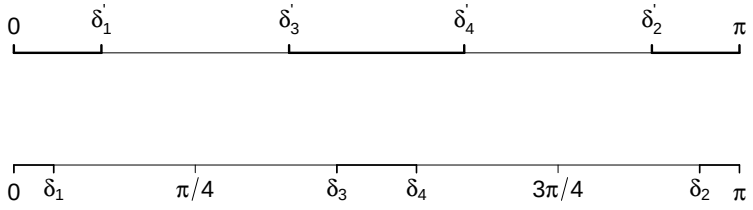


Figure 5.2: In bold line the set $[0, \delta_1] \cup [\delta_3, \delta_4] \cup [\delta_2, \pi]$ (lower panel); in thicker line the set $[0, \delta_1'] \cup [\delta_3', \delta_4'] \cup [\delta_2', \pi]$ (upper panel).

In particular, given that we are in the case $\Delta \neq 0$, it follows that $D < 0$ for all $\rho \in (1/2, 1)$ and for all $\delta \in (0, \delta_1) \cup (\delta_3, \delta_4) \cup (\delta_2, \pi)$.

Since $P, D < 0$ and Theorem 5.1.1, we know that $q_{\rho,\delta}(x)$ has four real distinct roots. Again upon applying Sturm's theorem, one can check numerically that these four roots are in $[-1, 1]$. Hence we have cases of bimodality.

(iv) From Proposition 5.2.2 we obtain $\Delta < 0$, for $\rho \in (1/2, 1)$ and $\delta \in (\delta_1, \delta_3) \cup (\delta_4, \delta_2)$, namely from δ in the complementary set given under (iii), given that $\delta \neq \pi/4, 3\pi/4$ in our initial hypothesis. With similar considerations as in (ii), we have cases of unimodality.

The cases when $\Delta = 0$ deserve a special mention and they are analysed in what follows.

As seen in Section 5.1, the number of solutions of (5.2) is important in the study of bimodality of the GvM. In the next section we analyse in detail the cases where both angular values obtain in (5.5) are or are not admissible solutions of (5.2).

5.2.2 Analysis of the term $(1 - 2\sin^2\delta)\sin\omega\cos\omega$

Let us return to the original problem: Bimodality of $f(\cdot|\mu_1, \mu_2, \kappa_1, \kappa_2)$ means four extrema points, that, in turn, means that (5.2) has four simple solutions. These extrema are analysed through the roots $\bar{x}$ of $q_{\rho,\delta}(x)$ and the transformation (5.5) tells us that we are only interested in real roots and within $[-1, 1]$.

In particular, the number of acceptable angular values given by (5.5) is crucial. The term $(1 - 2\sin^2\delta)\sin\omega\cos\omega$ in (5.2) plays an important role in the analysis of the bimodality: in general, if this term is different than zero, then it follows from $\cos\omega_1 = -\cos\omega_2$ that only one of the angular values ω_1 and ω_2 can be solution of (5.2). Hence, only one of the two angular values can be an extremum of the GvM density. Thus it is useful to analyse in detail the term $(1 - 2\sin^2\delta)\sin\omega\cos\omega$ and the cases where it vanishes.

The equation $(1 - 2\sin^2\delta)\sin\omega\cos\omega = 0$ holds in the following situations.

(i) $1 - 2\sin^2\delta = 0$, hence $\delta = \pi/4, 3\pi/4$.

(ii) $\sin\omega = 0$, hence $\omega = 0, \pi$. From (5.5) this happens when $\bar{x} = 0$ is one of the roots of $q_{\rho,\delta}(x)$, which means $\sin\delta\cos\delta = 0$ and hence $\delta = 0, \pi/2$.

(iii) $\cos\omega = 0$, hence $\omega = \pi/2, 3\pi/2$.
Let us analyse case (iii). From (5.5) the case $\omega = \pi/2$ happens when $\bar{x}_1 = 1$ is one of the roots of $q_{\rho,\delta}(x)$. On the other hand, $\omega = 3\pi/2$ holds when one of the roots is $\bar{x} = -1$. If $\bar{x} = 1$ is one of the roots of $q_{\rho,\delta}(x)$, then

$$1 + b_3 + b_2 + b_1 + b_0 = 0,$$

which leads to

$$\rho^2 - 2\rho\sin\delta\cos\delta + \sin^2\delta\cos^2\delta = 0,$$

which gives $\rho = \sin\delta\cos\delta$. Because $\rho > 0$, we restrict δ to the interval $(0, \pi/2)$. One can easily verify that the function $h_1(\delta) = \sin\delta\cos\delta$ has a maximum at

$\delta = \pi/4$ and at this point it assumes value 0.5. Hence $\rho = \sin\delta\cos\delta \leq 0.5,\ \forall\delta \in (0, \pi/2)$. On the other hand, if $\bar{x} = -1$ is one of the roots of $q_{\rho,\delta}(x)$ then

$$1 - b_3 + b_2 - b_1 + b_0 = 0,$$

hence

$$\rho^2 + 2\rho\sin\delta\cos\delta + \sin^2\delta\cos^2\delta = 0,$$

which leads to $\rho = -\sin\delta\cos\delta$ and $\rho > 0$ implies $\delta \in (\pi/2, \pi)$. With similar considerations as above, but this time for the function $h_2(\delta) = -\sin\delta\cos\delta$, we can say that $\rho = -\sin\delta\cos\delta \leq 0.5$. Case (i) of Section 5.2.1 tells us these are all cases of bimodality.

As mentioned, case (iii) will always give bimodality. On the other hand, it is interesting to study in detail the cases (i) and (ii). In fact, when $\delta = 0, \pi/4, \pi/2, 3\pi/4$ it holds that $\Delta = 0$ for all values of $\rho > 0$, as seen in Section 5.2.1. Moreover, Theorem 5.1.1 tells us that $\Delta = 0$ deserves a particular mention. In this case $q_{\rho,\delta}(x)$ can have multiple real roots, and this might influence the number of angular values ω, hence the number of extrema of the GvM density. We now analyse in detail cases (i) and (ii).

(i) Let us study the case $\delta = \pi/4$. The study of $\delta = 3\pi/4$ is analogous and it is omitted. When $\delta = \pi/4$, we have that $b_0 = 1/4$, $b_1 = \rho$, $b_2 = \rho^2 - 1$, $b_3 = -2\rho$ and so the quartic reduces to

$$q_{\rho,\pi/4}(x) = x^4 - 2\rho x^3 + (\rho^2 - 1)x^2 + \rho x + \frac{1}{4}.$$

Following the analysis of Δ in Section 5.2.1, we know that $\Delta = 0\ \forall\rho > 0$, and using D and P as in Remark 5.1.3 we can easily obtain

$$D = 0 \ \text{ and } \ P = -4\rho^2 - 8 < 0, \forall\rho > 0.$$

Thus according to Theorem 5.1.1 this is the only case where $q_{\rho,\delta}(x)$ has two double real roots $\bar{x}_1$ and $\bar{x}_2$. According to (5.7), they are

$$\bar{x}_1 = -\frac{b_3}{4} - S = \frac{\rho}{2} - S, \text{ and } \bar{x}_2 = -\frac{b_3}{4} + S = \frac{\rho}{2} + S, \tag{5.19}$$

with Q that reduces to

$$Q = \sqrt[3]{\frac{\Delta_1}{2}},$$

since $\Delta = 0$. In this case we have that

$$p = -\frac{\rho^2 + 2}{2}, \quad \Delta_1 = 2\left(\rho^2 + 2\right)^3,$$
$$Q = \left(\rho^2 + 2\right), \quad \Delta_0 = \left(\rho^2 + 2\right)^2.$$

Thus

$$S = \frac{1}{2}\sqrt{-\frac{2}{3} + \frac{1}{3}\left(Q + \frac{\Delta_0}{Q}\right)} = \frac{1}{2}\sqrt{-\frac{2}{3}\left[-\frac{\rho^2+2}{2} + \frac{1}{3}\left(2\left(\rho^2+2\right)\right)\right]}$$
$$= \frac{1}{2}\sqrt{\frac{1}{3}\left(3\left(\rho^2+2\right)\right)} = \frac{1}{2}\sqrt{\rho^2+2}.$$

We can re-write (5.19) as

$$\bar{x}_1 = \frac{\rho}{2} - \frac{1}{2}\sqrt{\rho^2+2}, \quad \text{and} \quad \bar{x}_2 = \frac{\rho}{2} + \frac{1}{2}\sqrt{\rho^2+2}.$$

We have that $\bar{x}_1 \in [-1,1]$, $\forall \rho > 0$, while $\bar{x}_2 \in [-1,1]$, $\forall \rho \in (0, 1/2]$. When $\bar{x}_1, \bar{x}_2 \in [-1,1]$, transforming in angular values leads to

$$\begin{aligned} \omega_1 &= (\arcsin \bar{x}_1) \bmod 2\pi, \qquad & \omega_2 &= (\pi - \arcsin \bar{x}_1) \bmod 2\pi, \\ \omega_3 &= (\arcsin \bar{x}_2) \bmod 2\pi, \qquad & \omega_4 &= (\pi - \arcsin \bar{x}_2) \bmod 2\pi, \end{aligned} \tag{5.20}$$

and since the term $(1 - 2\sin^2\delta)\sin\omega\cos\omega$ in (5.2) is equal to zero, all four angular values in (5.20) are distinct solutions of (5.2): we are in a case of bimodality. But for $\delta = \pi/4$ we have cases of unimodality as well. In fact, in Figure 5.3 the two roots $\bar{x}_1$ and $\bar{x}_2$ used in (5.20) are displayed: for $\rho > 1/2$, $\bar{x}_2$ is not in the interval $[-1,1]$, thus only $\bar{x}_1$ can be transformed in two angular roots, and in this case we only have two extrema points for the GvM. This is a case of unimodality.

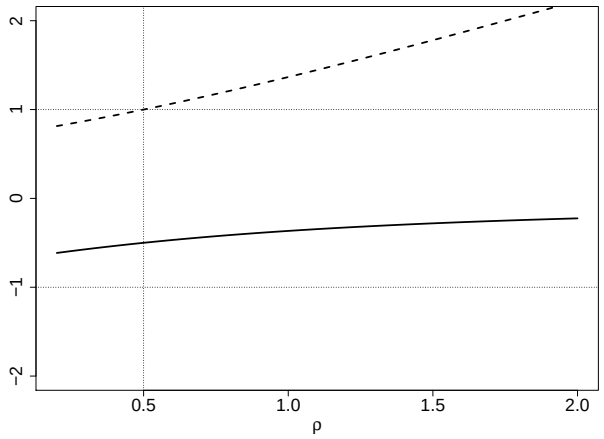


Figure 5.3: Plot of the two roots of $q_{\rho,\delta}(x)$, $\bar{x}_1$ (solid line) and $\bar{x}_2$ (dashed line), for $\delta = \frac{\pi}{4}$: for $\rho > 1/2$, $\bar{x}_2 \notin [-1,1]$.

(ii) Now we study the case $\delta = \pi/2$, while $\delta = 0$ is analogue and omitted. In this case, $b_0 = b_1 = b_3 = 0$, $b_2 = \rho^2 - 1$ and the quartic reduces to

$$q_{\rho,\pi/2}(x) = x^4 + (\rho^2 - 1)x^2,$$

which has a double real root $\bar{x}_1 = 0$, for all $\rho > 0$. Factoring out x^2 in $q_{\rho,\pi/2}(x)$, we get the following three possible cases:

1. For $\rho > 1$, $q_{\rho,\pi/2}(x)$ has two complex conjugate roots, plus the real double root $\bar{x}_1 = 0$,
2. For $\rho < 1$, $q_{\rho,\pi/2}(x)$ has two simple real roots $\bar{x}_2 = \sqrt{1-\rho^2}$ and $\bar{x}_3 = -\sqrt{1-\rho^2}$ in $[-1, 1]$, plus the real double root $\bar{x}_1 = 0$,
3. For $\rho = 1$, $q_{\rho,\pi/2}(x)$ has the unique root $\bar{x}_1 = 0$, with multiplicity four.

In case 1 the only real root $\bar{x}_1$ is transformed in angular roots

$$\begin{aligned} \omega_{1,1} &= \arcsin(\bar{x}_1) \bmod 2\pi = 0, \\ \omega_{1,2} &= \pi - \arcsin(\bar{x}_1) \bmod 2\pi = \pi. \end{aligned} \tag{5.21}$$

Since the term $(1 - 2\sin^2\delta)\sin\omega\cos\omega$ in (5.2) vanishes, both $\omega_{1,1}$ and $\omega_{1,2}$ are acceptable, thus the GvM has only two extrema points and this is counted as a case of unimodality. In case 2 the roots $\bar{x}_2$ and $\bar{x}_3$ can be transformed to four angular roots according to (5.5), but only two are acceptable. Hence the GvM has in total four extrema points and this is a case of bimodality. In case 3, $q_{\rho,\pi/2}(x)$ has only one root $\bar{x}_1 = 0$, which is transformed in two angular roots as in (5.21) of case 1. Thus it is a case of unimodality.

We conclude this analysis with the study of the roots of $q_{\rho,\delta}(x)$ for $1/2 < \rho < 1$ and $\delta = \delta_1, \delta_2, \delta_3, \delta_4$. We remind that, from the considerations made in (c.i) of Section 5.2.1, $\delta_i \neq 0, \pi/4, \pi/2, 3\pi/4$, for $i = 1, 2, 3, 4$. From (5.9) it is simple to check that $P < 0$. From Proposition 5.2.3 follows that $D < 0$. Moreover, Δ_0 as in 5.10 is sum of two strictly positive terms. Thus $\Delta_0 \neq 0$. Hence we are in case (3a) of Theorem 5.1.1: $q_{\rho,\delta}(x)$ has three real roots counted with multiplicity, namely $\bar{x}_1$, $\bar{x}_2$ (single real roots) and $\bar{x}_3$ (double real root). Each root is transformed into two angular values through (5.5) but, since the term $(1 - 2\sin^2\delta)\cos\omega\sin\omega$ is not zero, only one is accepted. Thus there are three critical points, namely ω_1, ω_2 and ω_3. Since the GvM is a differential periodic function, it must have the same number of maxima and minima, or in other words an even number of extrema. This means that a case with two maxima and one minimum, or with two minima and one maximum cannot happen. Hence, we deduce that one of the three critical points is a turning point, precisely a stationary inflection point. We verified numerically our claim for $\delta = \delta_1$ and for values of ρ uniformly generated in $(0.5, 1)$. Let ω_3 be the angular value associated to the double root $\bar{x}_3$. The second derivative of the exponent of the GvM divided by $4\kappa_2$, namely

$$\frac{1}{4\kappa_2} \cdot \frac{\mathrm{d}^2 g(\omega)}{\mathrm{d}\omega^2} = -\rho\cos\omega - \cos 2(\omega + \delta),$$

vanishes at ω_3. We conclude that the extrema of the GvM associated to ω_3 is indeed a stationary inflection point. Thus the GvM density has only two extrema and this is a case of unimodality.

Remark 5.2.4. In the light of Section 5.2.2, it is simple to check that for $\rho \in (1/2, 1)$ and $\delta = \delta_1, \delta_2, \delta_3, \delta_4$, we have $(1 - 2\sin^2\delta)\cos\omega\sin\omega \neq 0$. In fact, $\delta \neq 0, \pi/4, \pi/2, 3\pi/4$ and it is easy to check that $\rho = \sin\delta\cos\delta$ has no solutions for $\rho \in (1/2, 1)$. We analyse, without loss of generality, the case $\delta = \delta_1$. We get $\rho^2 = \cos^2\delta\,(1 - \cos^2\delta) = t_1(1 - t_1)$ that leads to $t_1^2 - t_1 + \rho^2 = 0$ and that has no real roots for $\rho > 1/2$. In the same way $\rho = -\sin\delta\cos\delta$ has no solutions for $\rho \in (1/2, 1)$.

5.3 Conclusion

The analysis of the quartic is summarized with a plot that visualizes the incidence of ρ and δ on bimodality and unimodality. Figure 5.4a shows the decomposition of the (ρ, δ) - plane in region of bimodality $\mathcal{W}$, in grey, and of unimodality $\mathcal{W}^c$, in white.[†] In particular, we fix nine points A-I belonging to $\mathcal{W}$ or $\mathcal{W}^c$ and we plot the corresponding GvM densities. The bimodality and unimodality of these points is confirmed by the densities in Figure 5.4b .

Figure 5.5 illustrates the analysis of Sections 5.2.1 and 5.2.2: A vertical line at ρ intersects only $\mathcal{W}$, $\mathcal{W}^c$ or both $\mathcal{W}$ and $\mathcal{W}^c$, as follows:

- For $\rho \leq 1/2$, a vertical line at ρ intersects only $\mathcal{W}$,
- For $\rho \geq 1$, a vertical line at ρ intersects only $\mathcal{W}^c$,
- For $1/2 < \rho < 1$, a vertical line at ρ intersects both $\mathcal{W}$ and $\mathcal{W}^c$.

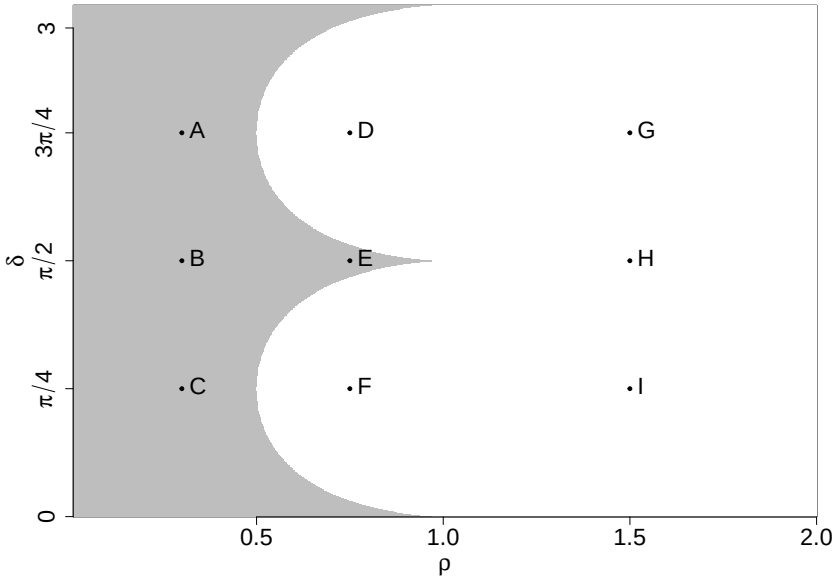
This confirms the algebraic study of Δ and of the cases $\delta = 0, \pi/4, \pi/2, 3\pi/4$. For example, the dashed vertical lines at $\rho = 0.25, 0.66$ and $\rho = 1.5$ shown in Figure 5.5 do follow this rule.

We can finally conclude that:

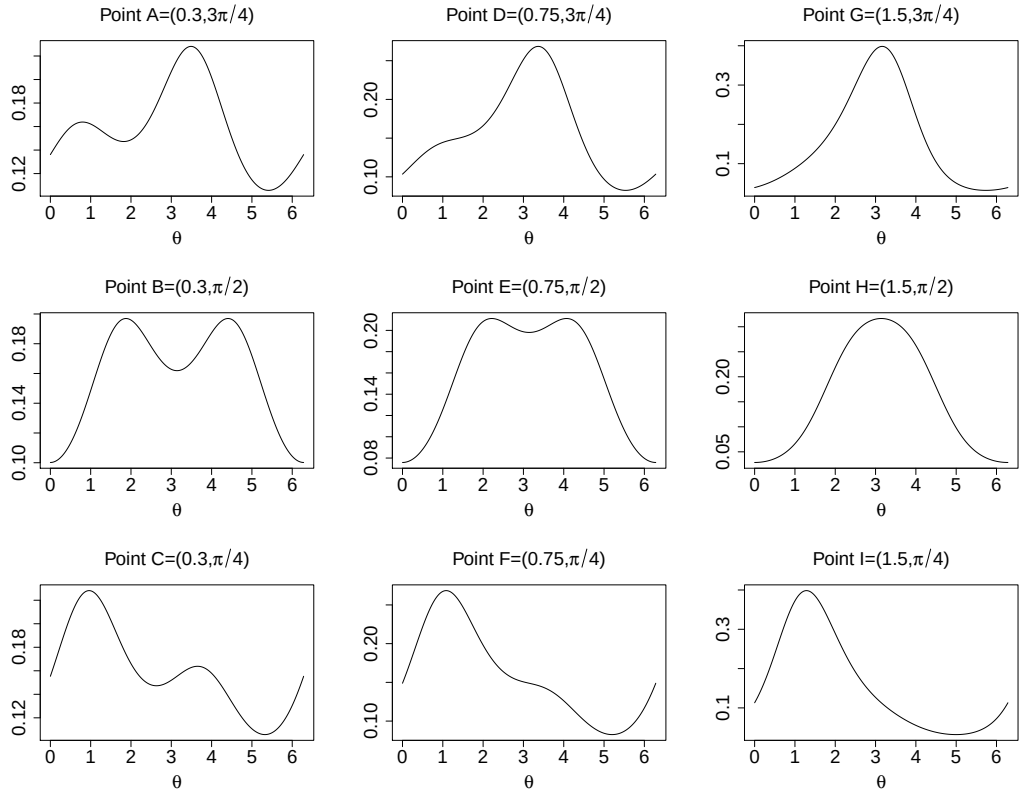
- For $\rho \leq 1/2$ and for all $\delta \in [0, \pi)$, we have cases of bimodality,
- For $\rho \geq 1$ and for all $\delta \in [0, \pi)$, we have cases of unimodality,
- For $1/2 < \rho < 1$ and for $\delta \in [0, \delta_1) \cup (\delta_3, \delta_4) \cup (\delta_2, \pi)$, we are in cases of bimodality.

The study of inflection points of the GvM is far from simple because it involves a polynomial of degree 8. We refer to Pfyffer and Gatto (2013). It is strictly related to the study of the change of sign of $dg(\omega)/d\omega$. This is subject of future research.

[†]Note that Figure 5.4a is a corrected version of Figure 2 of Pfyffer and Gatto (2013).



(a) Plane (ρ, δ) with region of bimodality $\mathcal{W}$, in grey, and of unimodality $\mathcal{W}^c$, in white.



(b) GvM densities corresponding to the points A - I of (a)

Figure 5.4: A point (ρ, δ) in $\mathcal{W}$ corresponds to a bimodal GvM; a point (ρ, δ) in $\mathcal{W}^c$, corresponds to a unimodal GvM.

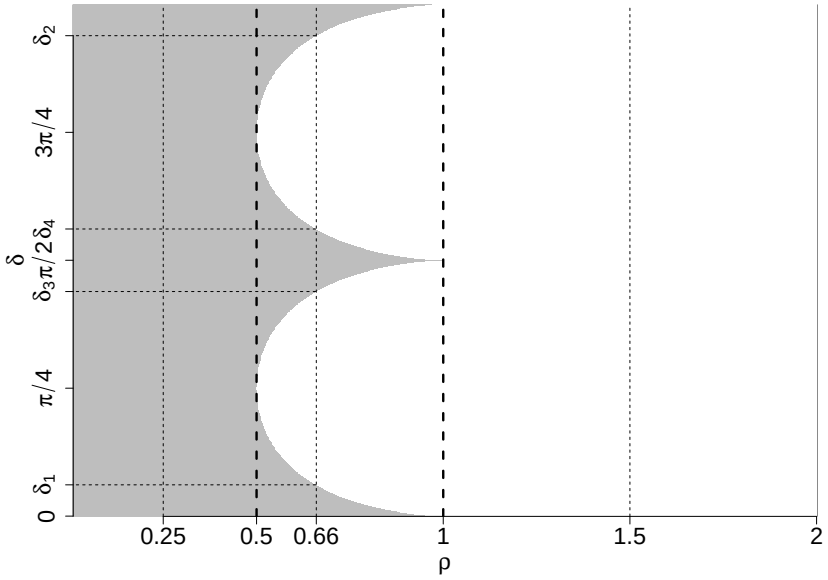


Figure 5.5: Region of bimodality $\mathcal{W}$ (grey) and of unimodality $\mathcal{W}^c$ (white), with focus on different values of ρ. In dashed lines the value at $\rho = 0.66$ and corresponding values $\delta_1, \delta_2, \delta_3, \delta_4$. Additional dashed lines at $\rho = 0.25$ and $\rho = 1.5$. Thicker dashed lines at $\rho = 0.5$ and at $\rho = 1$.

Chapter 6

Bayesian inference on the bimodality of the generalized von Mises [†]

This chapter introduces Bayesian inference on the bimodality of the generalized von Mises distribution for planar directions (Gatto and Jammalamadaka, 2007). As already mentioned, the GvM distribution is a flexible model that can be axial symmetric or asymmetric, unimodal or bimodal. The number of modes for a circular distribution has practical relevance in various fields. It is the case for example in meteorology, precisely in the context of the analysis of directions of winds, and in biology, precisely in the study of directions taken by various animals. In this chapter two approaches of Bayesian inference are considered. The first concerns test of hypotheses, where the null hypothesis H_0 is the bimodality of the GvM and the alternative hypothesis H_1 is unimodality. This test was motivated by Basu and Jammalamadaka (2000), who proposed a similar test for mixtures of vM distributions. The Bayes factor of this test is obtained. The second approach concerns HPD credible set. Two parameters relevant for bimodality are obtained and the HPD credible set for those parameters is compared with the subset of all parameters yielding bimodality.

We can summarize the Bayesian test as follows. Define by Ψ the set of all vectors of parameters $\boldsymbol{\psi} = (\kappa_1, \kappa_2, \mu_1, \mu_2)$ of the GvM distribution with density (2.7) that lead to bimodality and denote the prior probability of bimodality by $p = P[\boldsymbol{\psi} \in \boldsymbol{\Psi}]$. Since we assume that the elements of $\boldsymbol{\psi}$ are independent, the computation of p can be done by simple Monte Carlo. Assume that the components of the sample $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$ are independent and identically distributed (i.i.d.), with GvM distribution (2.7). Let $P[B|\boldsymbol{\theta}]$ be the conditional probability given $\boldsymbol{\theta}$, viz. the sample, for any measurable set B. The Bayesian test requires the posterior probability of bimodality $P[\boldsymbol{\psi} \in \boldsymbol{\Psi}|\boldsymbol{\theta}]$.

In the posterior distribution, the components of $\boldsymbol{\psi}$ are generally not independent. The computation of the posterior probability is done by Markov Chain Monte Carlo (MCMC). We remind that MCMC is a class of iterative simulation techniques that is

[†]Gatto and Salvador (2022)

based on three main algorithms: the Metropolis, due to Metropolis et al. (1953), the Metropolis-Hastings, due to Hastings (1970) and the Gibbs sampler due to Geman and Geman (1984). MCMC algorithms allow to generate a Markov chain whose equilibrium distribution coincides with the posterior distribution. The states of the Markov chain are recorded only after a certain initial period, usually called burn-in period: after this period the generated values of the Markov chain become very close to generations from the posterior distribution.

Precisely, we propose the following nested simulation scheme for the posterior distribution of $\boldsymbol{\psi}$. The main algorithm is the Gibbs sampler: we generate from the full conditional distributions of each parameter. The full conditional of one element of $\boldsymbol{\psi}$ is the posterior conditional distribution of that element given all other parameters fixed. The generation from each one of the full conditionals is done with the Metropolis-Hastings (MH) algorithm. Here, the sampling or instrumental density is a piecewise function that interpolates the full conditional.

The simulations from prior and posterior distributions are used to compute prior and posterior probability of bimodality. These probabilities allow us to infer on the bimodality by means of Bayes factors and HPD credible sets.

The next sections of the chapter present the following topics. Section 6.1 presents two Bayesian approaches for the bimodality of the GvM distribution: the testing approach, with the Bayes factor, and the HPD approach. It shows how to obtain posterior and full conditional distributions, for some given prior distributions. Then, a nested simulation algorithm from the posterior distribution is explained. It uses the Gibbs sampler and a MH algorithm. Section 6.2 presents the results of the simulation study. Bayes factors for unimodal and bimodal samples are obtained. HPD credible intervals for single parameters are presented. A bivariate HPD credible set for two parameters that determine bimodality of the GvM is given. Gelman's convergence diagnostic is presented. This method consists in studying the variance between, the variance within and the shrink factor of m simultaneous chains. We refer to (Gelman et al., 1992) for this. Simulation results under a more general prior are given at the end of the numerical study. Final remarks are given in Section 6.3.

6.1 Bayesian inference and MCMC

This section presents the methodology for Bayes inference on bimodality for the GvM, precisely the derivation of Bayes factors and HPD credible sets and the related MCMC. Section 6.1.1 provides the derivation of the posterior distribution. Section 6.1.2 concerns simulation from the posterior distribution. Its multidimensional and complicated form requires MCMC. We suggest using the Gibbs sampling from the posterior distribution, in conjunction with the MH algorithm for sampling from the one-dimensional full condi-

tionals. Inferential aspects related to Bayes factors and HPD credible sets are explained in Section 6.1.3.

6.1.1 Likelihood, posterior computations and bimodality

This section provides likelihood of the sample $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$ of independent angles from the GvM$(\mu_1, \mu_2, \kappa_1, \kappa_2)$ distribution and the posterior distribution of the parameters, for some given prior distributions.

The prior distribution of κ_j is the uniform distribution over $(0, \bar{\kappa}_j)$, for some $\bar{\kappa}_j > 0$, for $j = 1, 2$. Note that although the major part of the simulation study that follows uses these uniform priors, the more general beta prior is also considered at the end of the simulation study (in Section 6.2.3). The prior distribution of μ_1 is the vM(ν_1, τ_1) distribution with density given by (2.2). Thus

$$f(\mu_1|\nu_1, \tau_1) = \frac{1}{2\pi I_0(\tau_1)} \exp\left\{\tau_1 \cos\left(\mu_1 - \nu_1\right)\right\}, \quad \forall \mu_1 \in [0, 2\pi), \tag{6.1}$$

where $\tau_1 > 0$ and $\nu_1 \in [0, 2\pi)$. The prior distribution of μ_2 is the axial vM with density

$$f(\mu_2|\nu_2, \tau_2) = \frac{1}{\pi I_0(\tau_2)} \exp\left\{\tau_2 \cos 2\left(\mu_2 - \nu_2\right)\right\}, \quad \forall \mu_2 \in [0, \pi), \tag{6.2}$$

where $\tau_2 > 0$ and $\nu_2 \in [0, \pi)$ and $\tau_2 > 0$. We use the notation $\mu_2 \sim \text{vM}_2(\nu_2, \tau_2)$.

The likelihood of the sample $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$ is given by

$$\begin{aligned} l(\mu_1, \mu_2, \kappa_1, \kappa_2|\boldsymbol{\theta}) &\propto f(\boldsymbol{\theta}|\mu_1, \mu_2, \kappa_1, \kappa_2) \\ &= \frac{1}{\left[2\pi \cdot G_0\left(\delta, \kappa_1, \kappa_2\right)\right]^n} \exp\left\{\kappa_1 \sum_{i=1}^{n} \cos\left(\theta_i - \mu_1\right) + \kappa_2 \sum_{i=1}^{n} \cos 2\left(\theta_i - \mu_2\right)\right\}. \end{aligned} \tag{6.3}$$

The priors (6.1) and (6.2) and the likelihood (6.3), yields the joint posterior density of $(\kappa_1, \kappa_2, \mu_1, \mu_2)$ given by

$$f(\mu_1, \mu_2, \kappa_1, \kappa_2|\boldsymbol{\theta}) \propto f(\boldsymbol{\theta}|\mu_1, \mu_2, \kappa_1, \kappa_2) \cdot f(\mu_1|\nu_1, \tau_1) \cdot f(\mu_2|\nu_2, \tau_2) \cdot \mathrm{I}_{(0,\bar{\kappa}_1)}(\kappa_1)\mathrm{I}_{(0,\bar{\kappa}_2)}(\kappa_2), \tag{6.4}$$

where I denotes the indicator function. We note in particular that, although $(\kappa_1, \kappa_2, \mu_1, \mu_2)$ are a priori independent, they are a posteriori dependent.

The number of modes of the GvM$(\mu_1, \mu_2, \kappa_1, \kappa_2)$ distribution is determined by the nature of the roots of the quartic

$$q_{\rho,\delta}(x) = x^4 + b_3 x^3 + b_2 x^2 + b_1 x + b_0, \tag{6.5}$$

where $b_3 = -4\rho \sin\delta \cos\delta$, $b_2 = \rho^2 - 1$, $b_1 = 2\rho \sin\delta \cos\delta$ and $b_0 = \sin^2\delta \cos^2\delta$, with $\rho = \kappa_1/(4\kappa_2)$ and $\delta = (\mu_1 - \mu_2) \mathrm{mod} \pi$, cf. Gatto and Jammalamadaka (2007). the

algebraic analysis that can be found in Chapter 5 shows that the null hypothesis of bimodality admits the analytical expression

$$\mathrm{H}_0 : (\rho, \delta) \in \mathcal{W},$$

where

$$\mathcal{W} = \{(\rho, \delta) \in (0, \infty) \times [0, \pi) : \ q_{\rho,\delta} \text{ possesses four real roots over } [-1, 1]\}. \qquad (6.6)$$

Note that repeated roots are admissible. Unfortunately, the boundary of $\mathcal{W}$ does not admit a closed form expression. We can however determine it numerically and a graphical representation of $\mathcal{W}$ is given in Figure 6.7a. The alternative hypothesis H_1 is the general one, namely that the roots are of any other possible nature.

6.1.2 Computational aspects

Bayesian inference on the bimodality of the GvM distribution requires simulation from prior and posterior distributions. Prior simulations are straightforward and posterior simulations are done by MCMC.

Besides Monte Carlo, we mention that we compute the normalizing constant of the GvM distribution by (4.9).

6.1.2.1 Prior simulation

It follows from the assumed a priori independence of the parameters, that their simulation is elementary. For the sake of completeness, we describe here their generation. Simulating the prior probability of bimodality is accomplished with the following simple Monte Carlo algorithm. Let $\boldsymbol{\psi} = (\kappa_1, \kappa_2, \mu_1, \mu_2)$.

Computation of prior probability of bimodality by simulation

1. Generate $\boldsymbol{\psi}^{(t)}$, for $t = 1, \ldots, T$, where each element is generated independently from its one-dimensional prior distribution.
2. Compute the prior probability of bimodality by

$$\frac{1}{T} \sum_{t=1}^{T} \mathrm{I}\{\boldsymbol{\psi}^{(t)} \in \mathcal{W}\}.$$

Simulating κ_1 and κ_2 is straightforward, as they are assumed uniformly distributed. The simulation of μ_1 can be done with the program `rvonmises` of the package `Directional` in `R`. Concerning the simulation of μ_2, from the axial vM distribution, we note that $\theta \sim \mathrm{vM}(2\nu_2, \tau_2) \Rightarrow \theta/2 \sim \mathrm{vM}_2(\nu_2, \tau_2)$. Thus, generations from $\mathrm{vM}_2(\nu_2, \tau_2)$ are the generations from $\mathrm{vM}(2\nu_2, \tau_2)$ (with `rvonmises`) that are then divided by 2.

6.1.2.2 Posterior simulation via MCMC

The computation of the posterior probability of bimodality is more complicated. We compute the posterior distribution $\boldsymbol{\psi}$ by MCMC and precisely by Gibbs sampling (cf. Geman and Geman (1984)). By iterative generations from the full conditionals, we obtain a Markov chain whose stationary distribution is the posterior (6.4) (cf. e.g. Robert and Casella (2013), p. 371-424).

The full conditional of κ_1, namely the conditional density of κ_1 given μ_1, μ_2, κ_2 and $\boldsymbol{\theta}$, is given by

$$f_1(\kappa_1|\mu_1, \mu_2, \kappa_2, \boldsymbol{\theta}) \propto \frac{1}{G_0^n(\delta, \kappa_1, \kappa_2)} \exp\left\{\kappa_1 \sum_{i=1}^{n} \cos(\mu_1 - \theta_i)\right\} \cdot \mathrm{I}_{(0,\bar{\kappa}_1)}(\kappa_1).$$

The full conditional of κ_2 is given by

$$f_2(\kappa_2|\mu_1, \mu_2, \kappa_1, \boldsymbol{\theta}) \propto \frac{1}{G_0^n(\delta, \kappa_1, \kappa_2)} \exp\left\{\kappa_2 \sum_{i=1}^{n} \cos 2(\mu_2 - \theta_i)\right\} \cdot \mathrm{I}_{(0,\bar{\kappa}_2)}(\kappa_2).$$

The full conditional of μ_1 is given by

$$\begin{aligned} f_3(\mu_1|\mu_2, \kappa_1, \kappa_2, \boldsymbol{\theta}) &\propto \frac{1}{G_0^n(\delta, \kappa_1, \kappa_2)} \exp\Bigg\{\tau_1 \cos\mu_1 \cos\nu_1 + \tau_1 \sin\mu_1 \sin\nu_1 \\ &\qquad + \kappa_1 \cos\mu_1 \sum_{i=1}^{n} \cos\theta_i + \kappa_1 \sin\mu_1 \sum_{i=1}^{n} \sin\theta_i\Bigg\} \\ &\propto \frac{1}{G_0^n(\delta, \kappa_1, \kappa_2)} \exp\left\{\tau_1 \cos(\mu_1 - \nu_1) + \kappa_1 \sum_{i=1}^{n} \cos(\mu_1 - \theta_i)\right\}. \end{aligned}$$

The last full condition is for μ_2 and it is given by

$$\begin{aligned} f_4(\mu_2|\mu_1, \kappa_1, \kappa_2, \boldsymbol{\theta}) &\propto \frac{1}{G_0^n(\delta, \kappa_1, \kappa_2)} \exp\Bigg\{\tau_2 \cos 2\nu_2 \cos 2\mu_2 + \tau_2 \sin 2\nu_2 \sin 2\mu_2 \\ &\qquad + \kappa_2 \cos 2\mu_2 \sum_{i=1}^{n} \cos 2\theta_i + \kappa_2 \sin 2\mu_2 \sum_{i=1}^{n} \sin 2\theta_i\Bigg\} \\ &\propto \frac{1}{G_0^n(\delta, \kappa_1, \kappa_2)} \exp\left\{\tau_2 \cos 2(\mu_2 - \nu_2) + \kappa_2 \sum_{i=1}^{n} \cos 2(\mu_2 - \theta_i)\right\}. \end{aligned}$$

Given the above full conditionals, the Gibbs sampling scheme is as follows.

Posterior Gibbs sampling of $\boldsymbol{\psi}$

1. Select an arbitrary starting value $\boldsymbol{\psi}^{(0)} = \left(\kappa_1^{(0)}, \kappa_2^{(0)}, \mu_1^{(0)}, \mu_2^{(0)}\right)$.
2. For $t = 0, \ldots, n_s - 1$, generate iteratively:
   - $\kappa_1^{(t+1)}$ from $f_1\left(\kappa_1 \middle| \mu_1^{(t)}, \mu_2^{(t)}, \kappa_2^{(t)}, \boldsymbol{\theta}\right)$,

- $\kappa_2^{(t+1)}$ from $f_2\left(\kappa_2 \middle| \mu_1^{(t)}, \mu_2^{(t)}, \kappa_1^{(t+1)}, \boldsymbol{\theta}\right)$,
- $\mu_1^{(t+1)}$ from $f_3\left(\mu_1 \middle| \mu_2^{(t)}, \kappa_1^{(t+1)}, \kappa_2^{(t+1)}, \boldsymbol{\theta}\right)$,
- $\mu_2^{(t+1)}$ from $f_4\left(\mu_2 \middle| \mu_1^{(t+1)}, \kappa_1^{(t+1)}, \kappa_2^{(t+1)}, \boldsymbol{\theta}\right)$.

3. For a selected positive integer $n_b < n_s$, called burn-in period length, discard $\{\boldsymbol{\psi}^{(t)}\}_{t=0,\ldots,n_b}$. Re-index the $T = n_s - n_b$ retained simulations from 1 to T.

However, the generation from the full conditionals with one of basic methods for univariate distributions (such as inversion, acceptance-rejection) seems neither simple nor efficient. The complication arises from the part $1/G_0^n(\delta, \kappa_1, \kappa_2)$. Either by the expansion (4.9) or by numerical integration of (2.5), its evaluation is relatively complicated in the context of simulation and it tends to take very small values when n is large. We apply the MH algorithm presented in Robert and Casella (2013), p. 276. This algorithm requires a density that is close to the full conditional of interest and from which the simulation is simple. We call this surrogate density as instrumental density and we present it just after the following MH algorithm.

Univariate MH simulation from full conditionals of $\boldsymbol{\psi}$

Redenote by ψ any one of the parameters $\kappa_1, \kappa_2, \mu_1, \mu_2$. and by $f(\psi)$ its full conditional.

1. Select an arbitrary starting value $\psi^{(0)}$.
2. For $t = 0, \ldots, n_s - 1$:
   - generate $\tilde{\psi}_t$ from the instrumental density g,
   - set
   $$\psi^{(t+1)} = \begin{cases} \tilde{\psi}_t, & \text{with probability } \rho\left(\psi^{(t)}, \tilde{\psi}_t\right), \\ \psi^{(t)}, & \text{with probability } 1 - \rho\left(\psi^{(t)}, \tilde{\psi}_t\right), \end{cases}$$
   where
   $$\rho(\psi, \tilde{\psi}) = \min\left\{\frac{f(\tilde{\psi})}{f(\psi)}\frac{g(\psi)}{g(\tilde{\psi})}, 1\right\}.$$
3. For a selected burn-in $n_b < n_s$, discard $\{\psi^{(t)}\}_{t=0,\ldots,n_b}$.
   Re-index the $T = n_s - n_b$ retained simulations from 1 to T.

Thus, our posterior simulation algorithm has two levels of MCMC.

The instrumental density g in the above MH algorithm is the normalized piecewise linear interpolation of the full conditional. When the knots are $(\psi_0, y_0), (\psi_1, y_1), \ldots, (\psi_k, y_k)$,

it is given by

$$g(\psi) \propto \begin{cases} y_0 + d_0(\psi - \psi_0), & \text{if } \psi \in [\psi_0, \psi_1), \\ y_1 + d_1(\psi - \psi_1), & \text{if } \psi \in [\psi_1, \psi_2), \\ \quad \vdots & \quad \vdots \\ y_i + d_i(\psi - \psi_i), & \text{if } \psi \in [\psi_i, \psi_{i+1}), \\ \quad \vdots & \quad \vdots \\ y_{k-1} + d_{k-1}(\psi - \psi_{k-1}), & \text{if } \psi \in (\psi_{k-1}, \psi_k], \end{cases}$$

where $d_i = (y_{i+1} - y_i)/(\psi_{i+1} - \psi_i)$, for $i = 0, \ldots, k-1$.

Let us explain the simulation from the instrumental density of κ_1, for example. The interpolation of the full conditional $f_1(\kappa_1|\mu_1, \mu_2, \kappa_2, \boldsymbol{\theta})$ is taken with 5 interpolation points, A-E in Figure 6.1, that are chosen as follows. We fix the first and last point with abscissae values $\varepsilon = 0.2$ and $\bar{\kappa}_1$, respectively. The remaining points are generated uniformly in the interval $(\varepsilon, \bar{\kappa}_1)$. The instrumental density g is the normalized interpolation function. We generate from g with the inverse transform method. Figure 6.1 shows the full conditional $f_1(\kappa_1|\mu_1, \mu_2, \kappa_2, \boldsymbol{\theta})$ and its piecewise interpolation function.

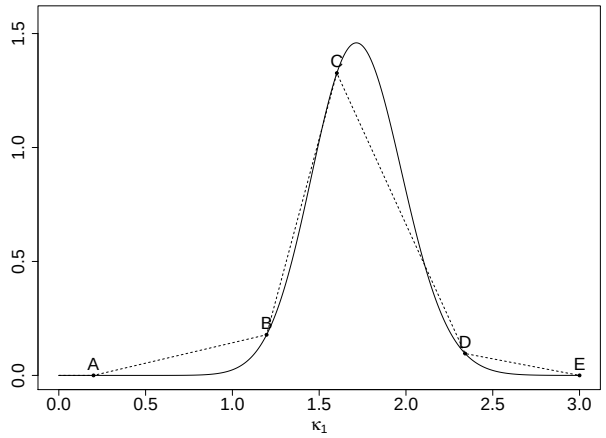


Figure 6.1: Full conditional of κ_1 (solid line) and interpolation function (dashed line) with 5 interpolation points A-E.

6.1.3 Bayes factor, HPrD and HPD credible sets

As mentioned, this chapter concerns inference of bimodality of the GvM distribution. This is done via Bayes factors introduced in Section 4.1. In our setting, the Bayes factor B_{01} in (4.1) is an indicator of the support of the sample for the null hypothesis H_0 of bimodality of the GvM, against the alternative hypothesis H_1 of unimodality.

Alternatively to the Bayes factor, Bayesian inference can rely on the HPD credible set. We firstly introduce a general concept. Let f be a density over $\mathbb{R}$. The highest

density region (HDR) at level $1-\alpha$ is the set

$$R(f_\alpha) = \{x \in \mathbb{R} : f(x) \geq f_\alpha\},$$

where f_α is the largest constant such that $P[R(f_\alpha)] \geq 1-\alpha$, for some (typically small) $\alpha \in (0,1)$; cf. e.g. Hyndman (1996). In particular it follows that, amongst all regions with coverage probability $1-\alpha$, the HDR has the smallest volume. Computational aspects of HDR are studied in Wright (1986), Hyndman (1990) and Hyndman (1996).

In the Bayesian setting where f is the posterior density, the HDR is called HPD credible set; cf. e.g. Box and Tiao (1973). When f is the prior density, we call the HDR as highest prior density (HPrD) credible set. We refer to Chen and Shao (1999) for computational aspects of HPD credible sets. Bimodality of the GvM is related to the parameters ρ and δ, as explained in Section 6.1.1. Thus Bayesian inference on bimodality can rely on bivariate HPD credible sets for (ρ, δ).

6.2 Simulation study

In this section we apply the methods that we proposed for the inference on the bimodality of the GvM: the MCMC algorithm of Section 6.1.2.2, the convergence diagnostics by Gelman et al. (1992) and the Bayes factors and HPD of Section 6.1.3.

We use various simulated samples $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$. The sample values are generated independently from the following distributions.

- A sample $\boldsymbol{\theta}$ of $n = 100$ values is generated from the unimodal GvM$(\pi, \pi/2, 1.2, 0.2)$ in Section 6.2.1.1, for the Bayes factor. This sample is also used in Section 6.2.3.1, for the Bayes factor with an alternative prior.

- A sample $\boldsymbol{\theta}$ of $n = 100$ values is generated from the bimodal GvM$(\pi, \pi/2, 0.16, 0.2)$ in Section 6.2.1.2, for the Bayes factor. This sample is also used in Section 6.2.3.3, for the Bayes factor with an alternative prior.

- A sample $\boldsymbol{\theta}$ of $n = 100$ values is generated from the bimodal GvM$(\pi, \pi/2, 1.5, 1.5)$ in Section 6.2.2, for the HPD credible set, and also used at the end of Section 6.2.2, for convergence diagnostic.

- Two samples $\boldsymbol{\theta}$ of $n = 10$ and $n = 200$ values are additionally generated from the same bimodal GvM$(\pi, \pi/2, 1.5, 1.5)$ in Section 6.2.2, for the study of the influence of the sample size n on the HPD credible intervals.

- A sample $\boldsymbol{\theta}$ of $n = 200$ values is generated from the unimodal GvM$(\pi, \pi/2, 1.2, 0.2)$ in Section 6.2.3.2, for the Bayes factor with an alternative prior.

All these generated samples are displayed through their kernel density estimations.

The hyperparameters for the prior specifications are chosen as follows:

$$\bar{\kappa}_1 = \bar{\kappa}_2 = 3,\ \nu_1 = \pi,\ \tau_1 = 10,\ \nu_2 = \frac{\pi}{2},\ \tau_2 = 10.$$

Prior and posterior probabilities of bimodality are obtained by simple Monte Carlo and Gibbs sampling, respectively, as explained in Section 6.1.2. The number of simulations is $T = 4 \cdot 10^4$ for the prior and for the posterior we have $n_s = 5 \cdot 10^4$ minus $n_b = 10^4$ retained simulations, namely total simulation minus the burn-in simulations. We choose $\boldsymbol{\psi}^{(0)} = (\kappa_1^{(0)}, \kappa_2^{(0)}, \mu_1^{(0)}, \mu_2^{(0)}) = (1.5, 1.5, \pi, \pi/2)$ as starting point for all Markov chain. Our Bayes factors and HPD credible sets are based on one single Markov chain. However, for the purpose of Gelman's convergence diagnostic, $m = 5$ Markov chains are generated with the same starting point $\boldsymbol{\psi}^{(0)}$.

6.2.1 Bayes factor

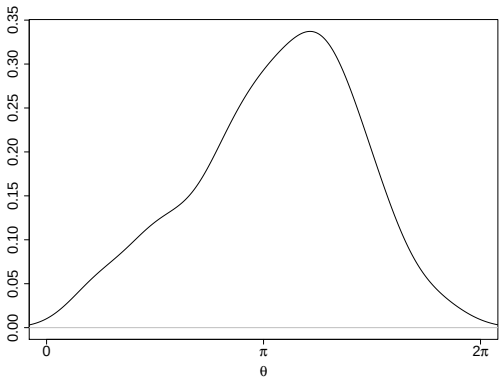
This section presents simulation results for the Bayes factor of the test of bimodality of the GvM. For the given sample $\boldsymbol{\theta}$ from a GvM, we compute prior and posterior probabilities of bimodality by simulating $\{\boldsymbol{\psi}^{(t)}\}_{t=1,\ldots,T}$ from the prior and posterior distributions respectively, with the algorithms of Section 6.1.2. These prior and posterior simulations are iterated $N = 500$ times and the retained Bayes factor is the average of N simulated Bayes factors. The section is divided in two parts: in part 6.2.1.1 we generate the sample $\boldsymbol{\theta}$ from the unimodal GvM whereas in part 6.2.1.2 we generate the sample $\boldsymbol{\theta}$ from the bimodal GvM.

6.2.1.1 Unimodal GvM sample

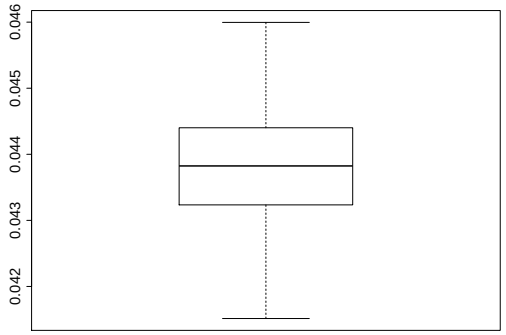
We generate the sample $\boldsymbol{\theta}$ of size $n = 100$ from the unimodal GvM$(\pi, \pi/2, 1.2, 0.2)$. Unimodality is formally justified by the fact that the point $(\rho, \delta) = (1.5, \pi/2)$ is really in the region $\mathcal{W}^{\mathrm{c}}$ of unimodality. This region is shown in white in Figure 6.7a only for values of ρ smaller than 1.25, but for values larger than 1.25 the region does remain white. The kernel density estimation of this sample is given in Figure 6.2a, which confirms unimodality.

We compute the representative Bayes factor $\bar{\mathrm{B}}_{01}$ by the mean of the generated Bayes factors $\mathrm{B}_i,\ i = 1, \ldots, N$. The $N = 500$ prior probabilities of bimodality are obtained by simple Monte Carlo; cf. Section 6.1.2.1. The average of these N values is $\bar{P}[\mathrm{H}_0] = 0.802$: there is a priori support for the null hypothesis of bimodality. The posterior probability of bimodality $\bar{P}[\mathrm{H}_0|\boldsymbol{\theta}] = 0.150$ is also obtained by averaging N values that are simulated by the MCMC algorithm of Section 6.1.2.2. The Bayes factor obtained with this study is given by

$$\bar{\mathrm{B}}_{01} = \frac{\bar{P}[\mathrm{H}_0|\boldsymbol{\theta}]}{1 - \bar{P}[\mathrm{H}_0|\,\boldsymbol{\theta}]} \frac{1 - \bar{P}[\mathrm{H}_0]}{\bar{P}[\mathrm{H}_0]} = 0.04381.$$



(a) Density estimation of unimodal sample $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{100})$.



(b) Boxplot of $N = 500$ simulated Bayes factors with unimodal sample of size $n = 100$.

Figure 6.2: Unimodal sample $\boldsymbol{\theta}$.

The aim of the process of averaging $N = 500$ values is the control of the simulation variability. The boxplot of the generated N values of B_{01} can be found in Figure 6.2b. The asymptotic normal confidence interval at level 0.95 for the Bayes factor is given by $\mathrm{I} = \bar{\mathrm{B}}_{01} \pm 1.96 N^{-1/2} \left\{ N^{-1} \sum_{i=1}^{N} (\bar{\mathrm{B}}_{01} - \mathrm{B}_i)^2 \right\}^{1/2}$. Obviously, although the N Bayes factors may not be normally distributed, their mean is asymptotically normal. We thus obtain the confidence interval

$$(0.04373, 0.04389).$$

As the sample $\boldsymbol{\theta}$ is generated from a unimodal GvM, our prior evidence of bimodality is no longer supported by the sample and thus $\bar{\mathrm{B}}_{01}$ does not support bimodality. The Bayes factor gives *negative* evidence of bimodality according to Table 4.1. Note that very few large values have been discarded from Figure 6.2b but they are nevertheless considered in the computation of the average Bayes factor and of the confidence interval.

6.2.1.2 Bimodal GvM sample

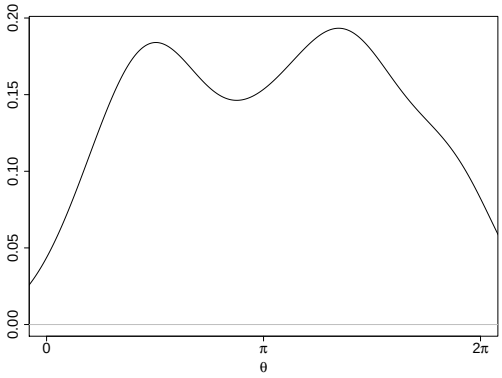
We now consider a sample of size $n = 100$ $\boldsymbol{\theta}$ generated from a bimodal GvM, precisely from $\mathrm{GvM}(\pi, \pi/2, 0.16, 0.2)$. We compute $(\rho, \delta) = (0.2, \pi/2)$, which belongs to the region of bimodality $\mathcal{W}$ shown is grey in Figure 6.7a. The kernel density estimation of $\boldsymbol{\theta}$ shown in Figure 6.3a is also indicating bimodality. As before, we compute the representative Bayes factor by the mean of $N = 500$ values. The prior probability is the same as in 6.2.1.1 and it is $\bar{P}[\mathrm{H}_0] = 0.802$. The posterior probability of bimodality $\bar{P}[\mathrm{H}_0|\boldsymbol{\theta}] = 0.992$ and thus we obtain

$$\bar{\mathrm{B}}_{01} = 34.935.$$

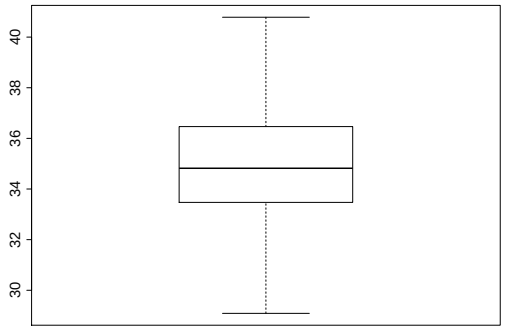
As in Section 6.2.1.1 we compute the asymptotic normal confidence interval for the Bayes factor at level 0.95. It is given by

$$(34.731, 35.139).$$

In this case the Bayes factor gives decisive evidence of bimodality. Figure 6.3b shows the boxplot of the Bayes factors. Here also, very few large values are discarded from the boxplot but are considered for the average of the Bayes factors and for the confidence interval.



(a) Density estimation of bimodal sample $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{100})$.



(b) Boxplot of $N = 500$ simulated Bayes factors with bimodal sample of size $n = 100$.

Figure 6.3: Bimodal sample $\boldsymbol{\theta}$.

Remark: Non-GvM sample

In this remark we study the behaviour of this test when the sample does not follow the assumed GvM distribution and thus when the given likelihood is inappropriate. We generate a sample of size $n = 100$ from the mixture of three vM distributions given by

$$\xi_1 \mathrm{vM}(\eta_1, \sigma_1) + \xi_2 \mathrm{vM}(\eta_2, \sigma_2) + \xi_3 \mathrm{vM}(\eta_3, \sigma_3),$$

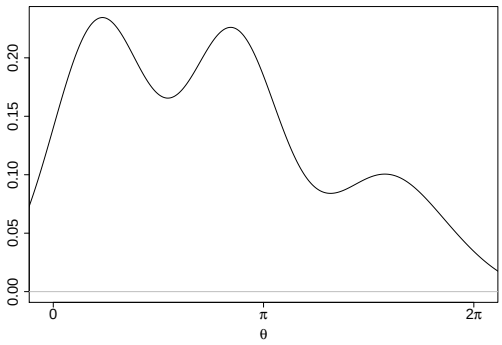
with

$$\eta_1 = \frac{\pi}{4},\ \eta_2 = \frac{7\pi}{8},\ \eta_3 = \frac{8\pi}{5},\ \sigma_1 = \sigma_2 = 12,\ \sigma_3 = 1 \ \text{ and } \ \xi_1 = \xi_2 = \xi_3 = \frac{1}{3}.$$

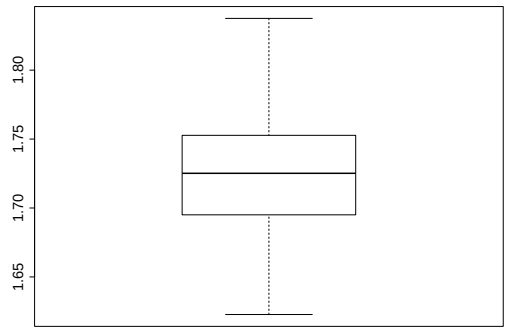
The density estimation of this non-GvM sample is given in Figure 6.4a and we can observe some clear evidence of trimodality. Figure 6.4b shows the boxplot of $N = 200$ simulated Bayes factors. Their mean value is

$$\bar{\mathrm{B}}_{01} = 1.724.$$

This value lies just within the range of positive evidence of bimodality of Table 4.1. Thus, two of the three modes have been associated to a single one, but it appears difficult to tell which ones they are. Perhaps the two modes of similar height (at η_1 and η_2) have been confounded or perhaps the lower mode (at η_3) has been neglected. Although we cannot provide an unambiguous explanation, we note that the amount of



(a) Density estimation of the non-GvM sample $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{100})$.



(b) Boxplot of $N = 200$ simulated Bayes factors with non-GvM sample of size $n = 100$.

Figure 6.4: Non-GvM sample $\boldsymbol{\theta}$. One outlier value is considered in the calculation of the representative Bayes factor, but not plotted in the boxplot.

positive evidence for bimodality is minor. The behaviour of the test in this erroneous situation seems reasonable.

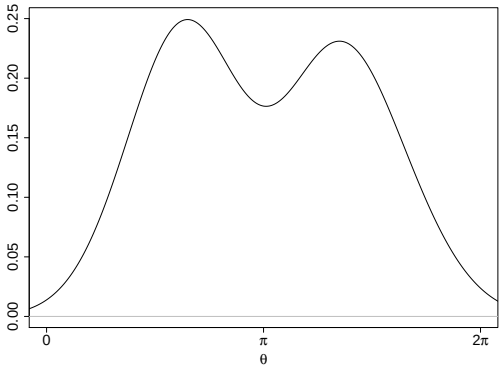
This test of bimodality is indeed constructed around the GvM distribution. The null hypothesis is characterized by the region of bimodality of the GvM, $\mathcal{W}$ given in (6.6), which is obtained from the nature of the roots of the quartic (6.5). The adequacy of the sample with the GvM model could be checked with the goodness-of-fit tests of Section 7.2 of Jammalamadaka and SenGupta (2001). Let us remind that various types of tests are available: Kuiper's test (the circular version of Kolmogorov-Smirnov test), Watson's test (the circular version of Cramer-von Mises test), Ajne's test (the circular version of chi-square test) and Rao's test (based on spacings between sample angles). The GvM distribution function is required for some of these tests and it can be found in Section 3.3 of Gatto (2022).

6.2.2 HPD and HPrD credible sets

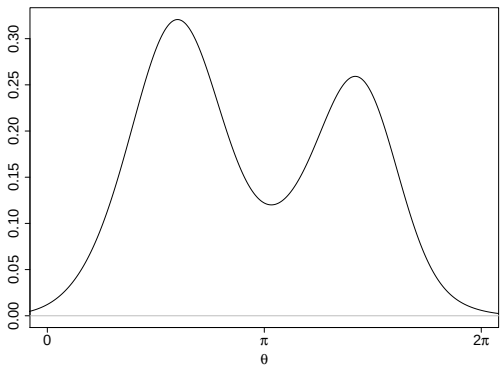
This section provides a simulation of the HPD credible set of (ρ, δ), which allows to determine the a posteriori evidence of uni- or bimodality. It also provides the simulation of HPrD and HPD credible intervals for $\kappa_1, \kappa_2, \mu_1, \mu_2$. One sample $\boldsymbol{\theta}$ of size $n = 100$ is considered and it is generated under the bimodal GvM$(\pi, \pi/2, 1.5, 1.5)$. Its kernel density estimation is shown in Figure 6.5b and we can see that bimodality is preserved.

We begin with the presentation of the HPrD and HPD credible intervals. These intervals are given in Table 6.1. We see that all HPD credible intervals are narrower than the HPrD credible intervals and that they contain their true value. The level is 0.95. Then we illustrate the effect of the sample size n on the HPD credible interval by considering two additional samples of sizes $n = 10$ and 200 that are generated from the same GvM$(\pi, \pi/2, 1.5, 1.5)$. Their density estimation is given in Figures 6.5a and

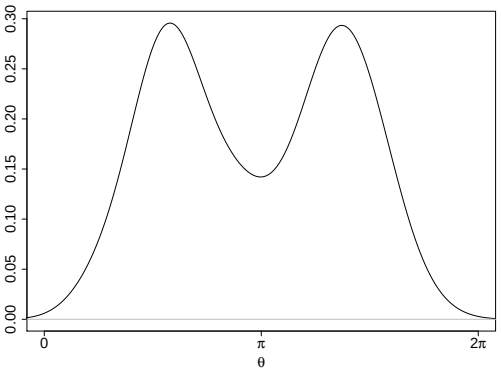
6.5c, respectively, and we see that bimodality is preserved. We restrict the analysis to the parameter μ_1. The results given in Table 6.2 shows that, as n increases, the HPD credible intervals at level 0.95 become narrower around the true value. A graph of this can be found in Figure 6.6, where the posterior densities, HPrD and HPD credible intervals are given. The grey regions have area 0.95 and they are based over the HPD credible intervals. The HPrD credible intervals are displayed with a bold line along the abscissae. As the sample size n increases, the HPD credible intervals become narrower around the true value of the parameter. Computations are done using `emp.hpd` of the `R`-package `TeachingDemos`, with $\alpha = 0.05$.



(a) Density estimation of bimodal sample $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{10})$.



(b) Density estimation of bimodal sample $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{100})$.



(c) Density estimation of bimodal sample $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{200})$.

Figure 6.5: Density estimations of the three samples used for HPrD and HPD credible intervals.

As mentioned in Section 6.1.1, the bimodality of the GvM is determined by the number of real roots in $[-1, 1]$ of the quartic $q_{\rho,\delta}$ in (6.5). We can infer on the bimodality of the GvM through the bivariate HPD credible set of (ρ, δ). We use the same sample $\boldsymbol{\theta}$ of size $n = 100$. The level of the HPD credible set is 0.95 and we compute the HPD

parameter	HPrD credible interval	HPD credible interval
κ_1	(0.120, 2.969)	(0.718, 1.907)
κ_2	(0.0007, 2.845)	(0.803, 1.671)
μ_1	(2.504, 3.782)	(3.033, 3.454)
μ_2	(1.248, 1.890)	(1.403, 1.644)

Table 6.1: HPrD and HPD credible intervals for parameters of GvM, at level 0.95.

with the function `HPDregionplot` of the `R`-package `emdbook`. Figure 6.7a shows that the HPD credible set obtained from the simulation is completely included in $\mathcal{W}$, the region of bimodality of the set of the parameters (ρ, δ) given in (6.6). This region of bimodality is shown in grey. The graph with the region of bimodality is due to Salvador and Gatto (2022a) and Pfyffer and Gatto (2013). It can also be found in Chapter 5 in Figure 5.4a. Thus, the HPD credible set gives us a clear confirmation of the bimodality. Figure 6.7b shows the posterior of (ρ, δ) and indicates that the HPD of Figure 6.7a should maintain the connected form at different levels.

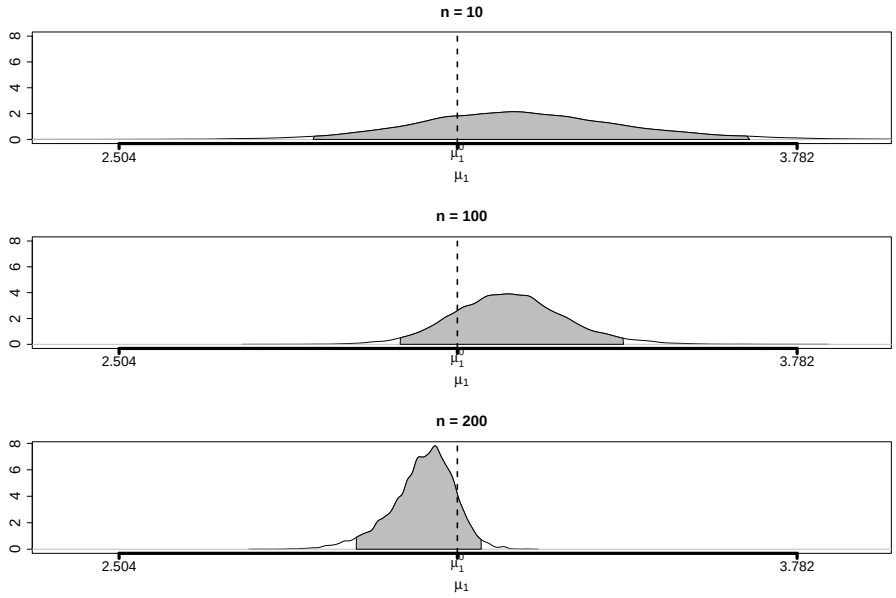


Figure 6.6: Comparison between HPrD credible interval (bold line) and HPD credible interval (base of grey area) for μ_1 with $n = 10, 100, 200$. True value of the parameter μ_1 is $\mu_1^0 = \pi$ (dashed line).

We now verify the convergence of the simulations. We consider only the sample $\boldsymbol{\theta}$ of size $n = 100$, which is also used in the first part of this section. In order to apply Gelman's diagnostic, we generate $m = 5$ chains. Let $g_{ij} = g(\boldsymbol{\psi}_i^{(j)})$ for $i = 1, \ldots, m$ and $j + 1, \ldots, T$. In our case the function g extracts the components of $\boldsymbol{\psi}$. For example, to check the converge of the first component we set $g(\boldsymbol{\psi}_i^{(j)}) = \kappa_{1,ij}$. The variance between

sample size n	**HPD credible interval for μ_1**
10	(2.869, 3.692)
100	(3.033, 3.454)
200	(2.951, 3.186)

Table 6.2: HPD credible intervals for μ_1, for different sample sizes n.

b, the variance within w and the shrink factor r of the chains are given by, respectively,

$$b = \frac{1}{m}\sum_{i=1}^{m}(\bar{g}_{i\cdot} - \bar{g}_{\cdot\cdot})^2, \quad w = \frac{1}{m}\sum_{i}^{m} s_i^2, \quad r := \sqrt{\frac{b+w}{w}},$$

where

$$\bar{g}_{i\cdot} = \frac{1}{T}\sum_{j=1}^{T} g_{ij}, \quad \bar{g}_{\cdot\cdot} = \frac{1}{m}\sum_{i=1}^{m}\bar{g}_{i\cdot} \text{ and } s_i^2 = \frac{1}{T}\sum_{j=1}^{T}(g_{ij} - \bar{g}_{i\cdot})^2.$$

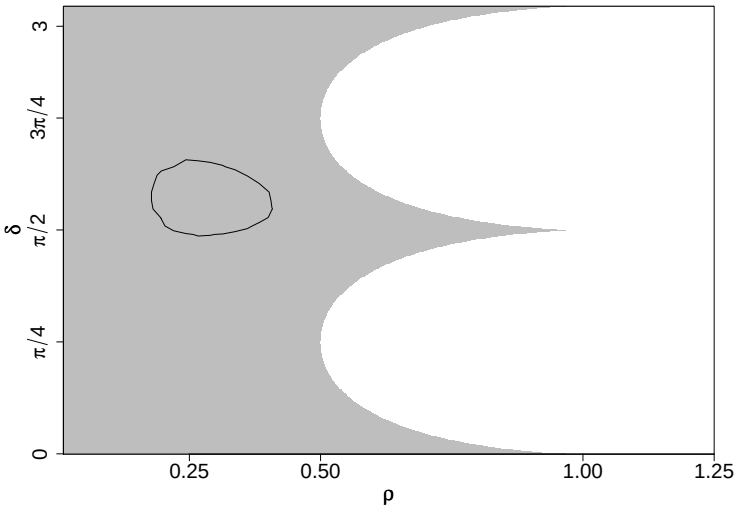
According to Gelman et al. (1992), convergence is achieved when r is close to 1. Results in Table 6.3 confirm convergence for each one of the four components of the chain. Thus, convergence of the whole Markov chain $\{\boldsymbol{\psi}^{(t)}\}_{t=1,\dots,T}$ is achieved. A visual

parameter	$\boldsymbol{b}$	$\boldsymbol{w}$	$\boldsymbol{r}$
κ_1	$9.391 \cdot 10^{-6}$	0.0135	1.00034
κ_2	$1.043 \cdot 10^{-5}$	0.0460	1.00011
μ_1	$8.492 \cdot 10^{-7}$	0.0027	1.00015
μ_2	$7.404 \cdot 10^{-7}$	0.0020	1.00018

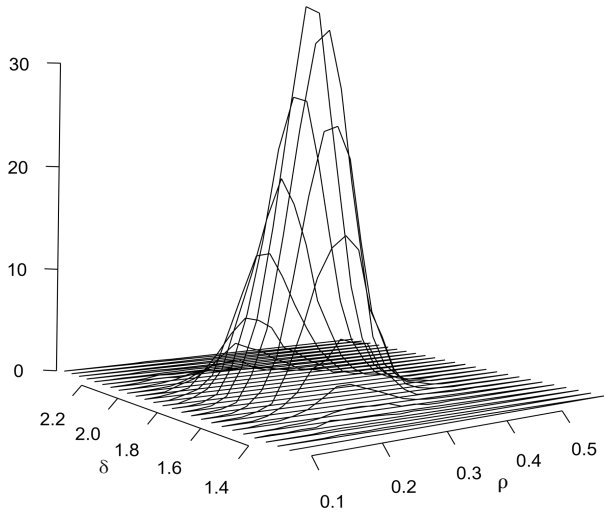
Table 6.3: Gelman diagnostics for $\kappa_1, \kappa_2, \mu_1, \mu_2$, with $m = 5$ chains and with $T = 4 \cdot 10^4$ retained simulations.

representation of the convergence is done using the command `gelman.plot` of the package `coda` that plots the shrink factor r evaluated at different instant of the simulation. The Markov chains are divided into bins according to the parameters `bin.width` and `max.bins`, which are the number of simulations per segment and the maximum number of bins, respectively. The shrink factor is repeatedly calculated. The first shrink factor is calculated with 50 simulations, the second with $50 +$ `bin.width` simulations, the third with $50 + 2 \cdot$ `bin.width` simulations, and so on. In Figure 6.8 we can see how the shrink factor of each parameter κ_1, κ_2, μ_1, μ_2 evolves with respect to the iterations of the chain: the convergence of the chains associated to κ_1 and κ_2 appears slower, since at the beginning the shrink factors fluctuate. But then they stabilize around the value 1. Thus, we can consider the test passed and convergence is achieved.

Visuals results for the convergence of the single parameters can be found in Figure 6.9. The four histograms of the simulations are close to four graphs of the full condi-



(a) Bivariate posterior HPD credible set at level 0.95 for (ρ, δ), $\mathcal{W}$ region of bimodality of the GvM distribution (grey) and $\mathcal{W}^{\mathsf{c}}$ region of unimodality (white).



(b) Posterior density of (ρ, δ).

Figure 6.7: (a) Uni- and bimodality regions of GvM together with HPD credible set of (ρ, δ) at level 0.95. (b) Posterior density of (ρ, δ) corresponding to the HPD credible set given in (a).

tionals (solid line), indicating that each one of the four components of the Markov chain converges to the respective full conditional.

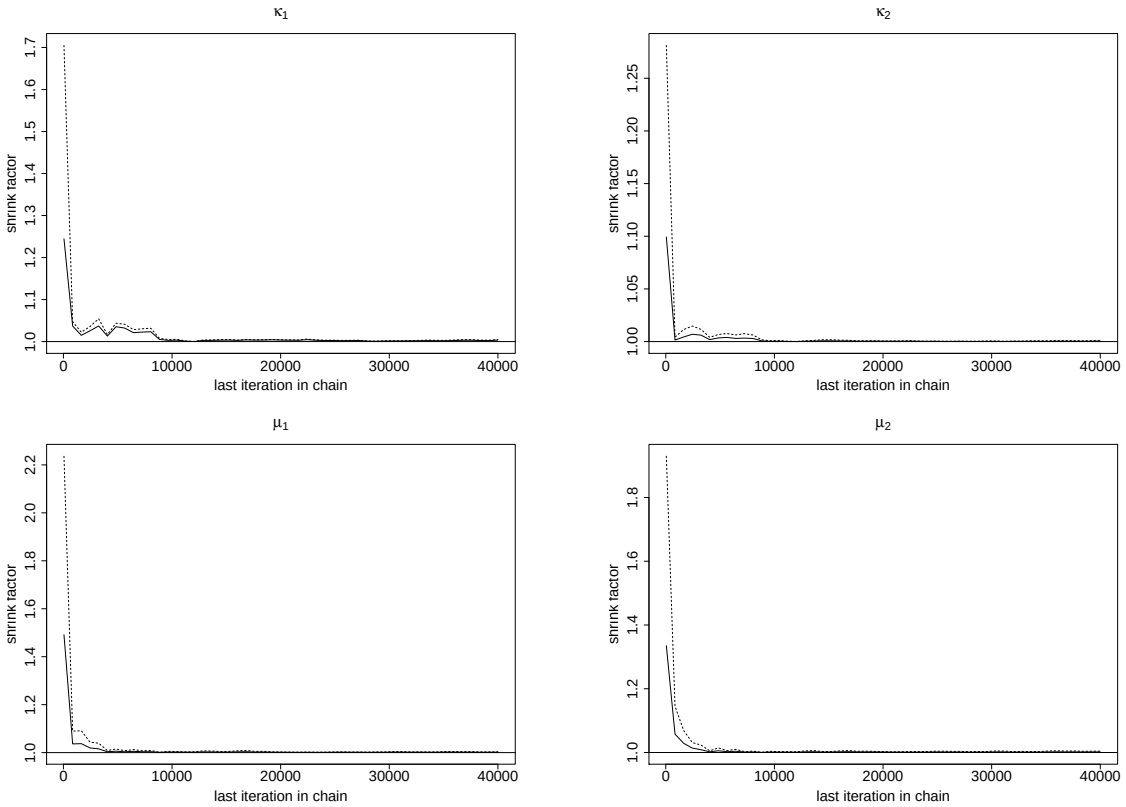


Figure 6.8: Median (solid line) and 97.5%-upper confidence limit (dashed line) of shrink factors r for $\kappa_1, \kappa_2, \mu_1, \mu_2$ with number of iterations. We kept the default values `bin.width` $= 10$ and `max.bins` $= 50$.

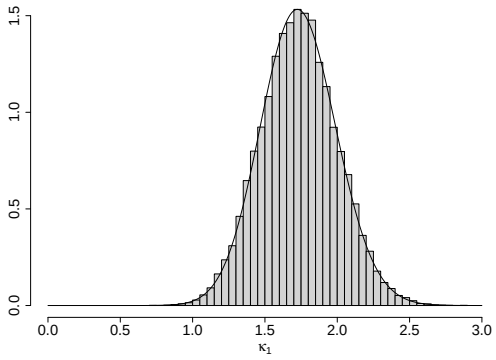
6.2.3 More general prior

So far, the prior distribution of κ_j has been the uniform distribution over $(0, \bar{\kappa}_j)$, for some $\bar{\kappa}_j > 0$, for $j = 1, 2$. This choice may not always be convenient mainly because of the abrupt jump of the densities at $\bar{\kappa}_1$ and $\bar{\kappa}_2$, making the choice of these values rather consequent. Since we may want a prior that decays slowly on the right tail, we can consider the two beta priors with densities proportional to $(\kappa_j/\bar{\kappa}_j)^{\alpha_j - 1} (1 - \kappa_j/\bar{\kappa}_j)^{\beta_j - 1}$, $\forall \kappa_j \in [0, \bar{\kappa}_j]$, with $\alpha_j, \beta_j > 0$, for $j = 1, 2$. Thus the uniform priors used so far and given in Section 6.1.2.1 are retrieved with $\alpha_j = \beta_j = 1$, for $j = 1, 2$.

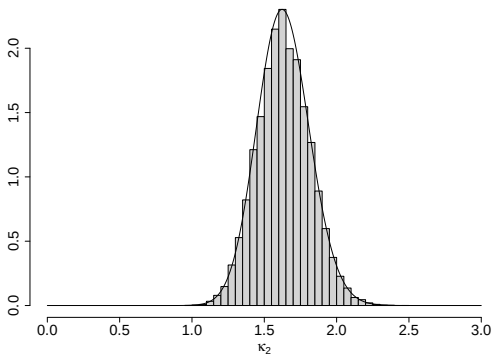
In the simulation study, we select $\alpha_j = \beta_j = 2$ and, as before, $\bar{\kappa}_j = 3$, for $j = 1, 2$. The priors for μ_1, μ_2 are taken as in Section 6.1.2.2. All hyperparameters are as in Section 6.2.

Thus the full conditional of κ_1 is given by

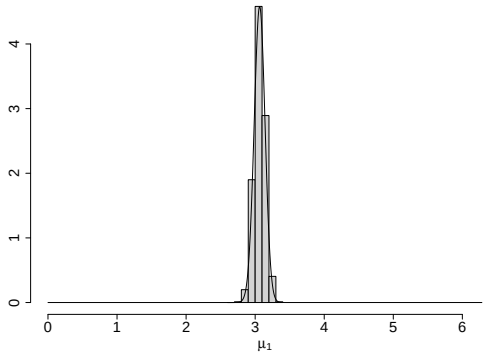
$$f_1(\kappa_1 | \mu_1, \mu_2, \kappa_2, \boldsymbol{\theta}) \propto \frac{1}{G_0^n(\delta, \kappa_1, \kappa_2)} \exp\left\{ \kappa_1 \sum_{i=1}^{n} \cos(\mu_1 - \theta_i) \right\} \cdot \left(\frac{\kappa_1}{\bar{\kappa}_1} \right) \left(1 - \frac{\kappa_1}{\bar{\kappa}_1} \right)$$



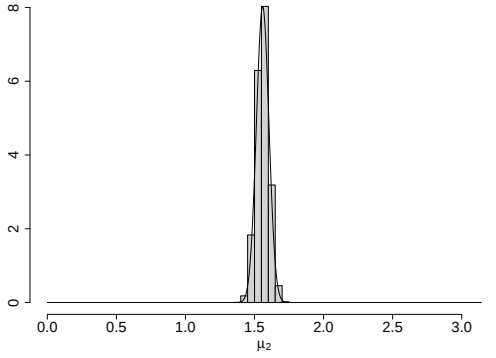
(a) Full conditional and histogram of generations for κ_1.



(b) Full conditional and histogram of generations for κ_2.



(c) Full conditional and histogram of generations for μ_1.



(d) Full conditional and histogram of generations for μ_2.

Figure 6.9: Visual test of convergence of generation by MH from full conditionals with $T = 4 \cdot 10^4$ simulated values.

and the full conditional of κ_2 is

$$f_2(\kappa_2|\mu_1, \mu_2, \kappa_1, \boldsymbol{\theta}) \propto \frac{1}{G_0^n(\delta, \kappa_1, \kappa_2)} \exp\left\{\kappa_2 \sum_{i=1}^{n} \cos 2\left(\mu_2 - \theta_i\right)\right\} \cdot \left(\frac{\kappa_2}{\bar{\kappa}_2}\right)\left(1 - \frac{\kappa_2}{\bar{\kappa}_2}\right).$$

The full conditionals of μ_1 and μ_2 are as in Section 6.1.2.2.

We use the unimodal sample of size $n = 100$ and the bimodal sample of size $n = 100$ that are used in Sections 6.2.1.1 and 6.2.1.2, respectively. Moreover we consider a third additional unimodal sample with larger size $n = 200$.

6.2.3.1 Unimodal sample of size $n = 100$

We first consider the same unimodal sample of size $n = 100$ of Section 6.2.1.1, which is generated from the GvM$(\pi, \pi/2, 1.2, 0.2)$ distribution. The density estimation with this sample is given in Figure 6.2a. Prior and posterior simulations are iterated $N = 500$ times. These simulations allow us to compute the prior probability of bimodality

$\bar{P}[\mathrm{H}_0] = 0.882$ and the posterior analogue $\bar{P}[\mathrm{H}_0|\boldsymbol{\theta}] = 0.784$. The mean value of the $N = 500$ generated Bayes factors is

$$\bar{B}_{01} = 0.487,$$

and the asymptotic normal confidence interval at level 0.95 for the true Bayes factor is

$$(0.486, 0.488).$$

The boxplot of the resulting Bayes factors is given in Figure 6.10a. The posterior probability of bimodality $\bar{P}[\mathrm{H}_0|\boldsymbol{\theta}]$ is indeed smaller than its prior $\bar{P}[\mathrm{H}_0]$, but not much smaller. The mean Bayes factor shows negative evidence of bimodality, but its value is quite larger than the value obtained in Section 6.2.1.1.

6.2.3.2 Unimodal sample of size $n = 200$

In order to clarify the question of the size of the previous Bayes factor, we increase the sample size to $n = 200$. A new sample $\boldsymbol{\theta}$ of size $n = 200$ is generated from $\mathrm{GvM}(\pi, \pi/2, 1.2, 0.2)$. Its kernel density estimation is given in Figure 6.10c. The posterior probability is now given by $\bar{P}[\mathrm{H}_0|\boldsymbol{\theta}] = 0.0519$. The previous prior probability $\bar{P}[\mathrm{H}_0] = 0.882$ is considered. The mean of $N = 500$ simulated Bayes factors is

$$\bar{B}_{01} = 0.00732,$$

and the asymptotic normal confidence interval at level 0.95 is

$$I = (0.00730, 0.00734).$$

The Bayes factor gives again negative evidence of bimodality, but it is now more substantial. The boxplot of the simulated Bayes factors can be found in Figure 6.10c. The Bayes factor and the posterior probability are significantly lower than with the sample of size $n = 100$. This is what we would have expected. The prior probability of bimodality $\bar{P}[\mathrm{H}_0] = 0.882$ is bigger than the one of Section 6.2.1.1. Thus we need an unimodal sample of larger size in order to obtain a sufficiently small posterior probability of bimodality.

6.2.3.3 Bimodal sample of size $n = 100$

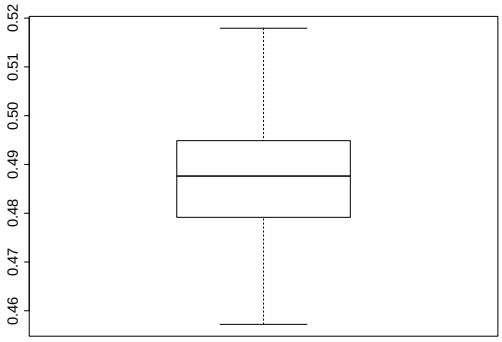
We now consider the bimodal sample used in Section 6.2.1.2, of $n = 100$ replications generated from $\mathrm{GvM}(\pi, \pi/2, 0.16, 0.2)$. The density kernel estimation of the sample can be found in Figure 6.3a. The prior probability of bimodality does not change: it is given by $\bar{P}[\mathrm{H}_0] = 0.882$. The posterior probability of bimodality in this obtained by $N = 500$ simulations is $\bar{P}[\mathrm{H}_0|\boldsymbol{\theta}] = 0.997$. This leads to the mean of $N = 500$ Bayes factors

$$\bar{B}_{01} = 66.348$$

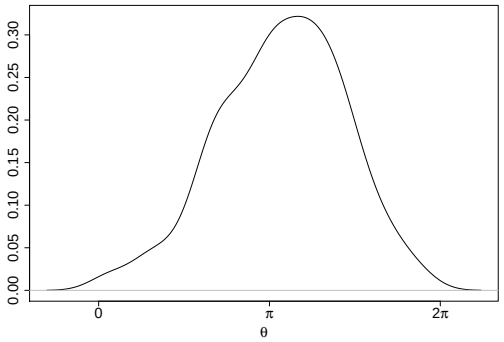
and to the asymptotic normal confidence interval at level 0.95

$$I = (65.545, 67.151).$$

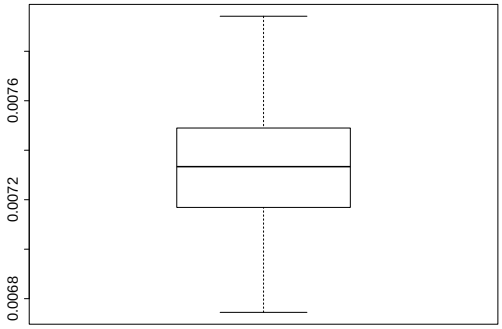
The Bayes factor gives thus decisive evidence of bimodality. The boxplots of the resulting Bayes factors can be found in Figure 6.10d.



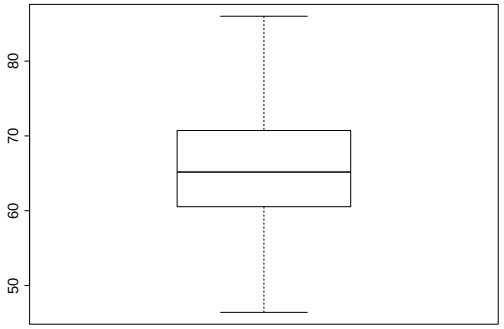
(a) Boxplot of $N = 500$ simulated Bayes factors with unimodal sample of size $n = 100$.



(b) Density estimation of unimodal sample $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{200})$.



(c) Boxplot of $N = 500$ simulated Bayes factors with unimodal sample of size $n = 200$.



(d) Boxplot of $N = 500$ simulated Bayes factors with bimodal sample of size $n = 100$.

Figure 6.10: Inference on bimodality via Bayes factors using Beta priors. Some extreme values are considered in the calculations but omitted in the boxplots.

6.3 Conclusion

This chapter studies Bayesian inference on the bimodality of the GvM distribution: MCMC algorithms for computing Bayes factors and HPD regions are proposed and tested by simulation. Thus, it extends the list of available inferential methods for the GvM model: trigonometric method of moments estimation, cf. Gatto (2008), and Bayesian test of symmetry (Chapter 4).

At the beginning of the chapter, various interesting properties of the GvM model are mentioned together with the importance of bimodality in various applied fields. For example, data on wind directions and graphs given in Gatto and Jammalamadaka (2007) show bimodality. Bimodality is indeed more frequent and important with circular than it is with linear data. In this context, the proposed Bayesian tests are practically relevant.

One may extend this study in two ways. First by proposing a frequentist test of bimodality. One could write a likelihood ratio test and exploit the partition of the parametric set in terms of unimodality and bimodality which is shown in Figure 6.7a. A second and perhaps more challenging way of extending this study would be by considering prior distributions with dependent GvM parameters. One would thus generalize the full conditionals and the MCMC algorithm of Section 6.1.2.2. We would then obtain circular-circular, circular-linear or other unusual types of joint distributions. A difficulty would be to obtain specific simulation methods from these joint distributions. Another difficulty would be the quantification of the dependence of the joint distributions on these manifolds. We could use or adapt the circular-circular and circular-linear correlation coefficients that are given e.g. in Sections 8.2-8.5 of Jammalamadaka and SenGupta (2001).

Chapter 7

Generation of random directions from the generalized von Mises-Fisher distribution[†]

This chapter introduces two generation algorithms for the generalized von Mises-Fisher (GvMF) distribution: one based on conditional simulation and one based on the MH algorithm. The GvMF distribution (3.4) is defined over the sphere and it is called directional distribution, because it provides a probability distributions for spatial directions. As previously mentioned in Chapter 2 and Chapter 3, data representing directions in $\mathbb{R}^p$, for $p = 2, 3, \ldots$, can be represented by points on the surface of the unit hypersphere $\mathbb{S}^{p-1}$. Most practical situations are with circular data, where $p = 2$, and with spherical data, where $p = 3$. Although some results of this chapter are given for the general case $p \geq 3$, because it is conceptually similar to the case of $p = 3$, the focus (and the numerical studies) are for the most practical situation $p = 3$. Spherical data arise in various scientific fields: physics, astronomy, earth sciences, biology, etc. The probability distribution of a random unit vector in $\mathbb{R}^p$ is generally called directional probability distribution. It is called circular distribution if $p = 2$ and spherical distribution if $p = 3$. This chapter proposes two simulation algorithms for GvMF$(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2; \kappa_1, \kappa_2; 1, 2)$ random directions in $\mathbb{S}^{p-1}$. The first algorithm is from conditional simulation and the second one from MCMC. Both algorithms apply the acceptance-rejection algorithm proposed by Ulrich (1984).

A brief survey of the coming sections is as follows. Section 7.1 introduces a method of conditional simulation for the GvMF. Section 7.2 provides MCMC algorithm. It is a MH algorithm for the GvMF distribution, for which the instrumental distribution is a mixture of vMF. Numerical results are given in Section 7.3 and some comparisons in Section 7.3.1 show that the two proposed generation algorithms yield similar samples. Note that although various results of Sections 7.1 and 7.2, are given for $\mathbb{S}^{p-1}$ with general

[†]Salvador and Gatto (2022c)

$p \geq 3$, numerical results are presented in $\mathbb{S}^2$ which is the case of most practical case. Some concluding remarks are given in Section 7.4.

7.1 Conditional simulation for GvMF

This section presents an algorithm of conditional simulation for GvMF$(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2; \kappa_1, \kappa_2; 1, 2)$ directions. Because this algorithm requires the normalizing constant of the density (3.4), Section 7.1.1 proposes a simple Monte Carlo method for obtaining this constant. The novel conditional simulation algorithm is given in Section 7.1.2.

The algorithm uses the acceptance-rejection algorithm for the vMF distribution of Ulrich (1984), which generates vMF directions with fixed mean direction $\boldsymbol{\mu} = \boldsymbol{e}_p = (0, \ldots, 0, 1)^\top$. This method is based on the following fact: a p-vector $\boldsymbol{X}$ has vMF distribution with mean direction $\boldsymbol{e}_p = (0, 0, \ldots, 1)^\top$ if and only if $\boldsymbol{X}^\top = ((1 - W^2)^{\frac{1}{2}} \boldsymbol{V}^\top, W)$, where $\boldsymbol{V}$ a $(p-1)$-unit vector which has uniform distribution, and W is a scalar random variable on $[-1, 1]$ with density

$$f(w) = c_{p,\kappa}^{-1} (1 - w^2)^{(p-3)/2} \exp\{\kappa w\}, \ \forall w \in [-1, 1]. \tag{7.1}$$

where

$$c_{p,\kappa} = \pi^{\frac{1}{2}} (\kappa/2)^\nu I_\nu(\kappa) \Gamma\{(2\nu + 1)/2\},$$

is the normalizing constant, with $\nu = (p-2)/2$, and Γ is the gamma function. Ulrich (1984) proposed the following envelope proportional to the density (7.1)

$$e(x, b) = d_{p,b} (1 - x^2)^{(p-3)/2} \left\{(1 + b) - (1 - b)x\right\}^{-(p-1)}, \ \forall x \in [-1, 1] \tag{7.2}$$

with

$$d_{p,b} = \left[\Gamma\left\{(p-1)/2\right\}\right]^2 b^{-(p-1)/2} \left\{2\Gamma(p-1)\right\}^{-1}.$$

It is easy to generate form (7.2) since it is a simple transformation of a symmetric beta random variable. Ulrich (1984) proposed the value

$$b = \left[-2\kappa + \left\{4\kappa^2 + (p-1)^2\right\}^{1/2}\right] / (p-1), \tag{7.3}$$

in order to maximize the acceptance ratio. However Wood (1994) noted some issues and proposed the following corrected version.

A1 Acceptance-rejection algorithm for vMF on $\mathbb{S}^{p-1}$

Choose n large. For $j = 1, \ldots, n$:

1. Calculate b as in (7.3), and set $x_0 = (1-b)/(1+b)$ and $c = \kappa x_0 + (p-1)\log(1 - x_0^2)$.

2. Generate $Z_j \sim \text{Beta}\left\{(p-1)/2, (p-1)/2\right\}$, U_j from the uniform distribution on $(0, 1)$, and calculate $W_j = \left\{1 - (1+b)Z_j\right\} / \left\{1 - (1-b)Z_j\right\}$.

3. If $\kappa W_j + (p-1)\log(1 - x_0 W_j) - c < \log(U_j)$, then go to 2.

4. Generate a uniform $(p-1)$-dimensional unit vector $\boldsymbol{V}_j$, and return $\boldsymbol{X}_j^\top = ((1 - W_j^2)^{1/2}\boldsymbol{V}_j^\top, W_j)$.

Then $\boldsymbol{X}$ has vMF distribution with mean direction $\boldsymbol{e}_p$ and concentration parameter $\kappa > 0$.

The following additional step allows to generate with arbitrary mean direction $\boldsymbol{\mu} \in \mathbb{S}^{p-1}$. This consists in finding a rotational matrix $\boldsymbol{R} \in \mathrm{SO}(p) = \{\boldsymbol{A} \in \mathbb{R}^{p\times p} \mid \boldsymbol{A}^\top\boldsymbol{A} = \boldsymbol{I}_p \wedge \det \boldsymbol{A} = 1\}$ (the group of all rotation matrices of $\mathbb{R}^p$), that rotates $\boldsymbol{e}_p$ into $\boldsymbol{\mu}$.

We notice that every orthogonal matrix with determinant equal to 1 and last column equal to $\boldsymbol{\mu}$ can be chosen. We can proceed as follows.

A2 Rotation algorithm on $\mathbb{S}^{p-1}$

1. Compute iteratively $(p-1)$ orthonormal vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{p-1}$ that form with $\boldsymbol{\mu}$ an orthonormal basis of $\mathbb{R}^p$, as follows. Let $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)^\top$.

   1.a The first orthonormal vector $\boldsymbol{u}_1 = (x_1, \ldots, x_p)^\top$ is a solution of $\sum_{i=1}^p x_i\mu_i = 0$.

   1.b For $j = 2, \ldots, p-1$, $\boldsymbol{u}_j = (x_1, \ldots, x_p)^\top$ is a solution of the system

$$\begin{cases} \sum_{i=1}^p x_i\mu_i = 0 \\ \sum_{i=1}^p x_i u_{j-1,i} = 0, \quad \text{for } j = 2, \ldots, p-1, \end{cases}$$

   where $\boldsymbol{u}_{j-1} = (u_{j-1,1}, \ldots, u_{j-1,p})^\top$, for $j = 2, \ldots, p-1$.

2. If the orthogonal matrix $(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{p-1}, \boldsymbol{\mu})$ has determinant 1, then set $\boldsymbol{R} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{p-1}, \boldsymbol{\mu})$. Otherwise, obtain $\boldsymbol{R}$ by permutating any two columns among the first $(p-1)$ columns of $(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{p-1}, \boldsymbol{\mu})$.

Thus, with Algorithm A1 for generating $\boldsymbol{X} \sim \mathrm{vMF}(\boldsymbol{e}_p, \kappa)$, we can obtain $\boldsymbol{Y} \sim \mathrm{vMF}(\boldsymbol{\mu}, \kappa)$ by finding the rotational matrix $\boldsymbol{R} \in \mathrm{SO}(p)$ that rotates $\boldsymbol{e}_p = (0, \ldots, 0, 1)^\top$ to $\boldsymbol{\mu}$ and then set $\boldsymbol{Y} = \boldsymbol{R}\boldsymbol{X}$.

Rodrigues (1840) provides a technique for constructing a matrix that rotates any vector $\boldsymbol{x}$ of $\mathbb{S}^2$ into another vector $\boldsymbol{y}$.

A3 Rodrigues rotation algorithm on $\mathbb{S}^2$

Assume that the vector $\boldsymbol{x} \in \mathbb{S}^2$ is rotated to $\boldsymbol{y} \in \mathbb{S}^2$. Define $\boldsymbol{v} = \boldsymbol{x} \times \boldsymbol{y}$ and $s = \langle \boldsymbol{x}, \boldsymbol{y} \rangle$. Then the matrix $\boldsymbol{R} \in \mathrm{SO}(3)$ that rotates $\boldsymbol{x}$ into $\boldsymbol{y}$ is obtained from

$$\boldsymbol{R} = \boldsymbol{I} + [\boldsymbol{v}]_\times + [\boldsymbol{v}]_\times^2 \frac{1-s}{\|\boldsymbol{v}\|^2},$$

where $[\boldsymbol{v}]_\times$ is the skew-symmetric cross-product matrix of $\boldsymbol{v} = (v_1, v_2, v_3)$ given by

$$[\boldsymbol{v}]_\times = \begin{pmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{pmatrix}.$$

The matrix $\boldsymbol{R}$ is called Rodrigues rotational matrix. We can apply Algorithm A3 to $\boldsymbol{x} = \boldsymbol{e}_3$ and $\boldsymbol{\mu} = \boldsymbol{y}$. This construction is restricted to $\mathbb{S}^2$, which is however the most practical case.

7.1.1 Monte Carlo computation of GvMF normalizing constant

No closed-form expression for the normalizing constant of the GvMF density (3.4) is available, if we exclude the vMF case where $k = 1$ and $i_1 = 1$. This section presents a Monte Carlo algorithm for obtaining the normalizing constant of the GvMF$(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2; \kappa_1, \kappa_2; 1, 2)$ distribution over $\mathbb{S}^2$. This algorithm is then used in Sections 7.1.2 and 7.2 for generating random directions from this distribution. Consider this distribution generally over $\mathbb{S}^{p-1}$, for $p \geq 3$, and define

$$\tilde{f}(\boldsymbol{x}) = \exp\left\{\kappa_1 \langle \boldsymbol{\mu}_1, \boldsymbol{x}\rangle + \kappa_2 \langle \boldsymbol{\mu}_2, \boldsymbol{x}\rangle^2\right\}, \ \forall \boldsymbol{x} \in \mathbb{S}^{p-1}, \tag{7.4}$$

which is thus proportional to the density f of the GvMF$_2(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2; \kappa_1, \kappa_2; 1, 2)$ distribution. Define the normalizing constant $B = \int_{\mathbb{S}^{p-1}} \tilde{f}(\boldsymbol{x}) dU(\boldsymbol{x})$. Define then

$$\tilde{g}(x) = \exp\left\{\kappa_1 \langle \boldsymbol{\mu}_1, \boldsymbol{x}\rangle\right\}, \ \forall \boldsymbol{x} \in \mathbb{S}^{p-1}, \tag{7.5}$$

which is thus proportional to the density g of the vMF$(\boldsymbol{\mu}_1, \kappa_1)$ distribution given in (3.2). Denote by A the inverse of the normalizing factor of vMF$(\boldsymbol{\mu}_1, \kappa_1)$, as given in (3.3) with $\kappa = \kappa_1$. We then have

$$\begin{aligned} B &= \int_{\mathbb{S}^{p-1}} \tilde{f}(\boldsymbol{x}) \mathrm{d}U(\boldsymbol{x}), \\ &= \int_{\mathbb{S}^{p-1}} \frac{\tilde{f}(\boldsymbol{x})}{A^{-1}\tilde{g}(\boldsymbol{x})} A^{-1} \tilde{g}(\boldsymbol{x}) \mathrm{d}U(\boldsymbol{x}) \\ &= A E_g\left[\frac{\tilde{f}(\boldsymbol{X})}{\tilde{g}(\boldsymbol{X})}\right] \\ &= A E_g\left[\exp\left\{\kappa_2 \langle \boldsymbol{\mu}_2, \boldsymbol{X}\rangle^2\right\}\right]. \end{aligned} \tag{7.6}$$

This last expression leads directly to a simple Monte Carlo estimator of B. However we consider a further simplification. Denote by $\boldsymbol{R}(\boldsymbol{x})$ the matrix of rotation of $\boldsymbol{x}$ to $\boldsymbol{e}_p$. This is the inverse of the matrix we compute with algorithm A2. We obtain from (7.6)

and from radial symmetry of the isotropic distribution $U^{\dagger}$ that

$$\begin{aligned} B &= AA^{-1} \int_{\mathbb{S}^{p-1}} \exp\left\{\kappa_1 \langle \boldsymbol{\mu}_1, \boldsymbol{x}\rangle + \kappa_2 \langle \boldsymbol{\mu}_2, \boldsymbol{x}\rangle^2\right\} \mathrm{d}U(\boldsymbol{x}) \\ &= AA^{-1} \int_{\mathbb{S}^{p-1}} \exp\left\{\kappa_1 \langle \boldsymbol{e}_p, \boldsymbol{R}(\boldsymbol{\mu}_1)\boldsymbol{x}\rangle + \kappa_2 \langle \boldsymbol{R}(\boldsymbol{\mu}_1)\boldsymbol{\mu}_2, \boldsymbol{R}(\boldsymbol{\mu}_1)\boldsymbol{x}\rangle^2\right\} \mathrm{d}U(\boldsymbol{x}) \\ &= AA^{-1} \int_{\mathbb{S}^{p-1}} \exp\left\{\kappa_1 \langle \boldsymbol{e}_p, \boldsymbol{x}\rangle + \kappa_2 \langle \boldsymbol{R}(\boldsymbol{\mu}_1)\boldsymbol{\mu}_2, \boldsymbol{x}\rangle^2\right\} \mathrm{d}U(\boldsymbol{x}) \\ &= AE_{g_1}\left[\exp\left\{\kappa_2 \langle \boldsymbol{R}(\boldsymbol{\mu}_1)\boldsymbol{\mu}_2, \boldsymbol{X}\rangle^2\right\}\right], \end{aligned} \tag{7.7}$$

where g_1 is the vMF$(\boldsymbol{e}_p, \kappa_1)$ density.

This leads to the following algorithm.

A4 Monte Carlo algorithm for normalizing constant of GvMF on $\mathbb{S}^{p-1}$

1. Compute $\boldsymbol{R}(\boldsymbol{\mu}_1)$, matrix of rotation of $\boldsymbol{\mu}_1$ to $\boldsymbol{e}_p$, with Algorithm A2.
2. For n large, generate $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ independently from vMF$(\boldsymbol{e}_p, \kappa_1)$, with the acceptance-rejection Algorithm A1.
3. Compute the sample mean estimator of B given by

$$\hat{B}_n = \frac{A}{n} \sum_{j=1}^{n} \exp\{\kappa_2 \langle \boldsymbol{R}(\boldsymbol{\mu}_1)\boldsymbol{\mu}_2, \boldsymbol{X}_j\rangle^2\}. \tag{7.8}$$

Note that the normalizing constant of the GvMF$(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2; \kappa_1, \kappa_2; 1, i_2)$, for $i_2 = 1, 3, \ldots$, can also be evaluated by Algorithm A4, after replacing 2 by i_2 in (7.8).

7.1.2 Algorithm of conditional simulation for GvMF

This section provides an algorithm of conditional simulation for generating directions from the GvMF$(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2; \kappa_1, \kappa_2; 1, 2)$ distribution. It uses acceptance-rejection and therefore the numerical efficiency depends on the closeness of the envelope to the target density, proportional to (7.4). The following developments hold generally on $\mathbb{S}^{p-1}$.

Precisely, two different envelopes for $\tilde{f}$ are used, according to sign of the scalar product $\langle \boldsymbol{\mu}_2, \boldsymbol{X}\rangle$, where $\boldsymbol{X}$ follows the GvMF$(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2; \kappa_1, \kappa_2; 1, 2)$ distribution. Define the random variable

$$Y = I\left\{\langle \boldsymbol{\mu}_2, \boldsymbol{X}\rangle \geq 0\right\}.$$

Then the GvMF$_2(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2; \kappa_1, \kappa_2; 1, 2)$ density f can be decomposed as

$$f(\boldsymbol{x}) = f(\boldsymbol{x}|0) \cdot P\left[Y = 0\right] + f(\boldsymbol{x}|1) \cdot P\left[Y = 1\right], \ \forall \boldsymbol{x} \in \mathbb{S}^{p-1},$$

†The random vector $\boldsymbol{X}$ is radially symmetric if $\boldsymbol{X} \sim R\boldsymbol{X}$, for any orthogonal matrix R. The distribution is radially symmetric in this case.

where $f(\cdot|j)$ denotes the conditional GvMF$(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2; \kappa_1, \kappa_2; 1, 2)$ density given $Y = j$, for $j = 0, 1$. Conditional simulation consists in generating first Y and then $\boldsymbol{X}$ from the conditional distribution given Y.

We note the following inequalities.

- When $Y = 1$, we obtain the inequality

$$\langle \boldsymbol{\mu}_2, \boldsymbol{X} \rangle^2 \leq \langle \boldsymbol{\mu}_2, \boldsymbol{X} \rangle.$$

  It leads to

$$\kappa_1 \langle \boldsymbol{\mu}_1, \boldsymbol{X} \rangle + \kappa_2 \langle \boldsymbol{\mu}_2, \boldsymbol{X} \rangle^2 \leq \kappa_1 \langle \boldsymbol{\mu}_1, \boldsymbol{X} \rangle + \kappa_2 \langle \boldsymbol{\mu}_2, \boldsymbol{X} \rangle = \kappa_1 \kappa_2 \beta_1 \langle \boldsymbol{\nu}_1, \boldsymbol{X} \rangle, \tag{7.9}$$

  where $\beta_1 = \|\kappa_2^{-1} \boldsymbol{\mu}_1 + \kappa_1^{-1} \boldsymbol{\mu}_2\|$ and $\boldsymbol{\nu}_1 = (\kappa_2^{-1} \boldsymbol{\mu}_1 + \kappa_1^{-1} \boldsymbol{\mu}_2)/\beta_1 \in \mathbb{S}^{p-1}$.

  We have that

$$\tilde{f}(\boldsymbol{x}) \leq \tilde{l}(\boldsymbol{x}) = \exp\{\kappa_1 \kappa_2 \beta_1 \langle \boldsymbol{\nu}_1, \boldsymbol{x} \rangle\}, \tag{7.10}$$

  for all $\boldsymbol{x} \in \mathbb{S}^{p-1}$ such that $\langle \boldsymbol{\mu}_2, \boldsymbol{x} \rangle \geq 0$.

  Thus, $\tilde{l}$ is an efficient envelope for the function (7.4), to be used in the acceptance-rejection Algorithm A1. Conditional on $Y = 1$, we can generate $\boldsymbol{X}$ by acceptance rejection of generations from vMF$(\boldsymbol{\nu}_1, \kappa_1 \kappa_2 \beta_1)$.

- When $Y = 0$, we obtain

$$\langle \boldsymbol{\mu}_2, \boldsymbol{X} \rangle^2 \leq -\langle \boldsymbol{\mu}_2, \boldsymbol{X} \rangle.$$

  In this case we have

$$\kappa_1 \langle \boldsymbol{\mu}_1, \boldsymbol{X} \rangle + \kappa_2 \langle \boldsymbol{\mu}_2, \boldsymbol{X} \rangle^2 \leq \kappa_1 \langle \boldsymbol{\mu}_1, \boldsymbol{X} \rangle - \kappa_2 \langle \boldsymbol{\mu}_2, \boldsymbol{X} \rangle = \kappa_1 \kappa_2 \beta_0 \langle \boldsymbol{\nu}_0, \boldsymbol{X} \rangle, \tag{7.11}$$

  where $\beta_0 = \|\kappa_2^{-1} \boldsymbol{\mu}_1 - \kappa_1^{-1} \boldsymbol{\mu}_2\|$ and $\boldsymbol{\nu}_0 = (\kappa_2^{-1} \boldsymbol{\mu}_1 - \kappa_1^{-1} \boldsymbol{\mu}_2)/\beta_0 \in \mathbb{S}^{p-1}$. In this case an efficient envelope for the function (7.4) is given by

$$\tilde{h}(\boldsymbol{x}) = \exp\{\kappa_1 \kappa_2 \beta_0 \langle \boldsymbol{\nu}_0, \boldsymbol{x} \rangle\}, \tag{7.12}$$

  $\forall \boldsymbol{x} \in \mathbb{S}^{p-1}$ such that $\langle \boldsymbol{\mu}_2, \boldsymbol{x} \rangle < 0$, which is proportional to the density of the vMF$(\boldsymbol{\nu}_0, \kappa_1 \kappa_2 \beta_0)$. Conditional on $Y = 0$, we can generate from this distribution with the acceptance-rejection Algorithm A1.

The probability $P[Y = 1]$ can be obtained via simple Monte Carlo as follows. As before, f is the GvMF$(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2; \kappa_1, \kappa_2; 1, 2)$ density, B is its normalizing constant, g is the vMF$(\boldsymbol{\mu}_1, \kappa_1)$ density, A is its normalizing constant and g_1 is the vMF$(\boldsymbol{e}_p, \kappa_1)$ density. The non-normalized GvMF and vM densities $\tilde{f}$ and $\tilde{g}$ are respectively defined in (7.4)

and (7.5). Then

$$\begin{aligned}
P[Y=1] &= E_f\left[I\left\{\langle \boldsymbol{\mu}_2, \boldsymbol{X}\rangle \geq 0\right\}\right] \\
&= \int_{\mathbb{S}^{p-1}} I\left\{\langle \boldsymbol{\mu}_2, \boldsymbol{X}\rangle \geq 0\right\} \cdot B^{-1}\tilde{f}(\boldsymbol{x})\mathrm{d}U(\boldsymbol{x}) \\
&= \int_{\mathbb{S}^{p-1}} I\left\{\langle \boldsymbol{\mu}_2, \boldsymbol{X}\rangle \geq 0\right\} \cdot \frac{B^{-1}\tilde{f}(\boldsymbol{x})}{A^{-1}\tilde{g}(\boldsymbol{x})} \cdot A^{-1}\tilde{g}(\boldsymbol{x})\mathrm{d}U(\boldsymbol{x}) \\
&= \frac{A}{B}\int_{\mathbb{S}^{p-1}} I\left\{\langle \boldsymbol{\mu}_2, \boldsymbol{x}\rangle \geq 0\right\} \cdot \exp\{\kappa_2\langle \boldsymbol{\mu}_2, \boldsymbol{x}\rangle^2\} g(\boldsymbol{x})\mathrm{d}U(\boldsymbol{x}) \\
&= \frac{A}{B}E_g\left[I\left\{\langle \boldsymbol{\mu}_2, \boldsymbol{X}\rangle \geq 0\right\} \cdot \exp\{\kappa_2\langle \boldsymbol{\mu}_2, \boldsymbol{X}\rangle^2\}\right] \\
&= \frac{E_g\left[I\left\{\langle \boldsymbol{\mu}_2, \boldsymbol{X}\rangle \geq 0\right\} \cdot \exp\{\kappa_2\langle \boldsymbol{\mu}_2, \boldsymbol{X}\rangle^2\}\right]}{E_g\left[\exp\{\kappa_2\langle \boldsymbol{\mu}_2, \boldsymbol{X}\rangle^2\}\right]} \\
&= \frac{E_{g_1}\left[I\left\{\langle \boldsymbol{R}(\boldsymbol{\mu}_1)\boldsymbol{\mu}_2, \boldsymbol{X}\rangle \geq 0\right\} \cdot \exp\{\kappa_2\langle \boldsymbol{R}(\boldsymbol{\mu}_1)\boldsymbol{\mu}_2, \boldsymbol{X}\rangle^2\}\right]}{E_{g_1}\left[\exp\{\kappa_2\langle \boldsymbol{R}(\boldsymbol{\mu}_1)\boldsymbol{\mu}_2, \boldsymbol{X}\rangle^2\}\right]}.
\end{aligned}$$

The last two equalities follow from (7.6) and (7.7) respectively . Hence we have the following algorithm for computing the probability that $Y = 1$.

A5 Monte Carlo algorithm for $P[Y = 1]$ on $\mathbb{S}^{p-1}$

1. Compute $\boldsymbol{R}(\boldsymbol{\mu}_1)$, matrix of rotation of $\boldsymbol{\mu}_1$ to $\boldsymbol{e}_p$, with Algorithm A2.
2. For n large, generate $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ independently from vMF$(\boldsymbol{e}_p, \kappa_1)$, with the acceptance-rejection Algorithm A1.
3. Estimate the desired probability by

$$\hat{p} = \frac{\sum_{j=1}^{n} I\left\{\langle \boldsymbol{R}(\boldsymbol{\mu}_1)\boldsymbol{\mu}_2, \boldsymbol{X}_j\rangle \geq 0\right\} \cdot \exp\{\kappa_2\langle \boldsymbol{R}(\boldsymbol{\mu}_1)\boldsymbol{\mu}_2, \boldsymbol{X}_j\rangle^2\}}{\sum_{j=1}^{n} \exp\{\kappa_2\langle \boldsymbol{R}(\boldsymbol{\mu}_1)\boldsymbol{\mu}_2, \boldsymbol{X}_j\rangle^2\}}.$$

We are now ready to give the conditional simulation algorithm.

A6 Conditional simulation algorithm for GvMF on $\mathbb{S}^{p-1}$

1. Estimate $P[Y = 1]$ by Algorithm A5 and denote $\hat{p}$ the estimation.
2. Compute $\boldsymbol{R}(\boldsymbol{\nu}_j)$, matrix of rotation of $\boldsymbol{\nu}_j$ in (7.9) and (7.11) to $\boldsymbol{e}_p$ with Algorithm A2, for $j = 0, 1$.
3. Choose n large and repeat for $j = 1, \ldots, n$:
   - 3.a generate Y_j from the Bernoulli distribution with parameter $\hat{p}$;
   - 3.b if $Y_j = 1$, then apply acceptance-rejection algorithm to generate from $f(\boldsymbol{x}|1)$:

3.b.1 generate $\boldsymbol{X}_j$ from vMF$(\boldsymbol{e}_p, \kappa_1\kappa_2\beta_1)$ with the acceptance-rejection Algorithm A1. Generate U_j from Uniform$(0,1)$, independent from $\boldsymbol{X}_j$;

3.b.2 if $U_j \leq \tilde{f}(\boldsymbol{R}^{-1}(\boldsymbol{\nu}_1)\boldsymbol{X}_j)/\tilde{l}(\boldsymbol{R}^{-1}(\boldsymbol{\nu}_1)\boldsymbol{X}_j)$, with $\tilde{l}$ defined in (7.10), accept $\boldsymbol{R}^{-1}(\boldsymbol{\nu}_1)\boldsymbol{X}_j$;

3.b.3 otherwise reject and return to 3.b.1.

3.c Else if $Y_j = 0$, then apply acceptance-rejection algorithm to generate from $f(\boldsymbol{x}|0)$:

3.c.1 generate $\boldsymbol{X}_j$ from vMF$(\boldsymbol{e}_p, \kappa_1\kappa_2\beta_0)$ with the acceptance-rejection Algorithm A1. Generate U_j from Uniform$(0,1)$, independent from $\boldsymbol{X}_j$;

3.c.2 if $U_j \leq \tilde{f}(\boldsymbol{R}^{-1}(\boldsymbol{\nu}_0)\boldsymbol{X}_j)/\tilde{h}(\boldsymbol{R}^{-1}(\boldsymbol{\nu}_0)\boldsymbol{X}_j)$, with $\tilde{h}$ defined in (7.12), accept $\boldsymbol{R}^{-1}(\boldsymbol{\nu}_0)\boldsymbol{X}_j$;

3.c.3 otherwise reject and return to 3.c.1.

In Algorithm A6, the acceptance-rejection part when $Y = 1$ uses the envelope $M\tilde{l}$ of $\tilde{f}$, where $M > 0$ is a constant such that $\tilde{f}(\boldsymbol{x}) \leq M\tilde{l}(\boldsymbol{x})$, $\forall \boldsymbol{x} \in \mathbb{S}^{p-1}$. A small value of M leads to a small rejection rate of generations from the density l, which is the vM density obtained by normalizing $\tilde{l}$. The best admissible value of M is $\sup_{\boldsymbol{x}\in\mathbb{S}^{p-1}} \tilde{f}(\boldsymbol{x})/\tilde{l}(\boldsymbol{x})$ and we now show that it is equal to 1. We thus choose $M = 1$ in the part conditional on $Y = 1$ of Algorithm A6. The same choice is considered in the part conditional on $Y = 0$ of Algorithm A6.

Let us show the optimality of $M = 1$ in the part conditional on $Y = 1$. We have from (7.9) that

$$r(\boldsymbol{x}) = \frac{\tilde{f}(\boldsymbol{x})}{\tilde{l}(\boldsymbol{x})} = \frac{\exp\{\kappa_1\langle\boldsymbol{\mu}_1, \boldsymbol{x}\rangle + \kappa_2\langle\boldsymbol{\mu}_2, \boldsymbol{x}\rangle^2\}}{\exp\{\kappa_1\langle\boldsymbol{\mu}_1, \boldsymbol{x}\rangle + \kappa_2\langle\boldsymbol{\mu}_2, \boldsymbol{x}\rangle\}} = \exp\left\{\kappa_2\langle\boldsymbol{\mu}_2, \boldsymbol{x}\rangle\left[\langle\boldsymbol{\mu}_2, \boldsymbol{x}\rangle - 1\right]\right\} \leq 1, \tag{7.13}$$

$\forall \boldsymbol{x} \in \mathbb{S}^{p-1}$ such that $\langle\boldsymbol{\mu}_2, \boldsymbol{x}\rangle \geq 0$. We need to maximize the ratio r over the given domain. According to Lagrange's multiplier method, we need to solve

$$\nabla\Lambda(\boldsymbol{x}, \lambda) = \boldsymbol{0}, \tag{7.14}$$

where $\Lambda(\boldsymbol{x}, \lambda) = r(\boldsymbol{x}) - \lambda K(\boldsymbol{x})$ is the Lagrangian, $K(\boldsymbol{x}) = ||\boldsymbol{x}||^2 - 1$ is the spherical constraint, $\lambda \in \mathbb{R}$ is Lagrange's multiplier and ∇ denotes the gradient operator. By solving (7.14), we obtain that r has two extrema points over $\mathbb{S}^{p-1}$, namely $\boldsymbol{\mu}_2$ and $-\boldsymbol{\mu}_2$. The only acceptable solution under $Y = 1$ is $\boldsymbol{\mu}_2$. The algebraic details can be found in Appendix B. Thus $r(\boldsymbol{x}) \leq r(\boldsymbol{\mu}_2) = 1$ and we select $M = 1$.

In the case $Y = 0$, we have to maximize the ratio

$$r(\boldsymbol{x}) = \frac{\tilde{f}(\boldsymbol{x})}{\tilde{h}(\boldsymbol{x})} = \exp\left\{\kappa_2\langle\boldsymbol{\mu}_2, \boldsymbol{x}\rangle\left[\langle\boldsymbol{\mu}_2, \boldsymbol{x}\rangle + 1\right]\right\} \leq 1,$$

$\forall \boldsymbol{x} \in \mathbb{S}^{p-1}$ such that $\langle \boldsymbol{\mu}_2, \boldsymbol{x} \rangle < 0$. Similar considerations as above leads to one maximum, namely $-\boldsymbol{\mu}_2$; cf. details in Appendix B. Thus we find again $r(\boldsymbol{x}) \leq r(-\boldsymbol{\mu}_2) = 1$, for all $\boldsymbol{x}$ of the considered domain, hence $M = 1$.

7.2 Metropolis-Hastings generation algorithm for GvMF distribution

This section proposes an algorithm of sequential generation of GvMF directions over $\mathbb{S}^{p-1}$, with $i_1 = 1$ and $i_2 = 2$, as before. It is the MH algorithm of MCMC introduced by Hastings (1970). In MCMC, a Markov chain with unique stationary distribution is generated. The stationary distribution is the target GvMF distribution. The generations are done from the transition distribution, which is a mixture of a Dirac distribution and a jumping, also called instrumental, distribution. Thus, the MH algorithm accepts or rejects generations from the instrumental distribution according to an acceptance probability that depends on the target and on the instrumental distributions. A reference for MCMC is Chapter XIII of Asmussen and Glynn (2007), for example. Note that the normalizing constant of the target distribution need not be known. We consider the independent version of the MH algorithm, see e.g. Robert and Casella (2013, p. 276), which is described as follows.

A7 MH algorithm for GvMF on $\mathbb{S}^{p-1}$

1. Let f be the density function of the GvMF$(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2; \kappa_1, \kappa_2; 1, 2)$. Define the instrumental density g as the:
   - vMF$(\boldsymbol{\mu}_1, \kappa_1)$ density, if $\kappa_1 \geq \kappa_2$,
   - $1/2 \cdot \text{vMF}(\boldsymbol{\mu}_2, \kappa_2) + 1/2 \cdot \text{vMF}(-\boldsymbol{\mu}_2, \kappa_2)$ mixture density, otherwise.
2. Select an arbitrary starting value over $\mathbb{S}^{p-1}$, denoted $\boldsymbol{X}^{(0)}$.
3. Repeat for $k = 0, \ldots, n_s - 1$:

   3.a generate $\boldsymbol{Y}_k$ from the density g;

   3.b set

$$\boldsymbol{X}^{(k+1)} = \begin{cases} \boldsymbol{Y}_k, & \text{with probability } \rho\left(\boldsymbol{X}^{(k)}, \boldsymbol{Y}_k\right), \\ \boldsymbol{X}^{(k)}, & \text{with probability } 1 - \rho\left(\boldsymbol{X}^{(k)}, \boldsymbol{Y}_k\right), \end{cases}$$

   where

$$\rho(\boldsymbol{x}, \boldsymbol{y}) = \min\left\{ \frac{f(\boldsymbol{y})}{f(\boldsymbol{x})} \frac{g(\boldsymbol{x})}{g(\boldsymbol{y})}, 1 \right\}.$$

Generation from the vMF distribution is done according to Algorithm A1, along with Algorithm A2.

4. For a selected burn-in length $n_b < n_s$, discard $\{\boldsymbol{X}_k\}_{k=0,\dots,n_b}$.
   Re-index the $n = n_s - n_b$ retained generations from 1 to n.

The form of the density g given in Algorithm A7 follows from the simple observation that if $\kappa_1 \geq \kappa_2$, then the GvMF$(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2; \kappa_1, \kappa_2; 1, 2)$ distribution is mainly concentrated around the direction of $\boldsymbol{\mu}_1$, otherwise if $\kappa_2 > \kappa_1$, then the concentration is along the axis of $\boldsymbol{\mu}_2$, namely around $\boldsymbol{\mu}_2$ and around $-\boldsymbol{\mu}_2$. This was mentioned in Section 7.1.2. Numerical results show that this simple form of g is satisfactory, although an optimal form might be obtained.

Numerical results and comparisons with Algorithm A6 of conditional simulation are presented in Section 7.3.

7.3 Numerical Results

In this section we report the results obtained using the acceptance-rejection with conditional approach and the MH algorithms. We consider only the case in $\mathbb{S}^2$ which is of most practical interest. Here the rotational matrices are computed with Algorithm A3. In order to simplify notation, let us call the conditional acceptance-rejection Algorithm A6 and the MH Algorithm A7 with $\kappa_1 = 8$, $\kappa_2 = 1$, CAR1 and MH1 respectively. In the same way let us call the acceptance-rejection and the MH with $\kappa_1 = 1, \kappa_2 = 8$, CAR2 and MH2 respectively.

Numerical results for CAR1 are now presented. The $n = 10^3$ simulations from the GvMF are plotted on the unit sphere $\mathbb{S}^2$ in the coordinate system $(X_1, X_2, X_3) = (X, Y, Z)$. The directions are chosen as follows: let $\boldsymbol{v}_1 = (3, 12, 9)$ and $\boldsymbol{v}_2 = (1, 0, 1)$. Then $\boldsymbol{\mu}_i = \boldsymbol{v}_i / \|\boldsymbol{v}_i\|$, for $i = 1, 2$. Figure 7.1 shows the results.

Similar results are obtained using CAR2, with same directions $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ chosen above. Results are presented in Figure 7.2. In particular the $n = 10^3$ points are more concentrated not only around the direction of $\boldsymbol{\mu}_2$ as seen in Figure 7.2a, but also around $-\boldsymbol{\mu}_2$, as shown in Figure 7.2b. This phenomenon depends on the term $\kappa_2 \langle \boldsymbol{x}, \boldsymbol{\mu}_2 \rangle^2$ of the GvMF distribution.

We present the results obtained for MH1. The numerical results are given using the same directions as for the conditional approaches above. We generate $n = 10^3$ effective random variables. These are the result of $n_s = 1.5 \cdot 10^3$ simulations, and $n_b = 5 \cdot 10^2$ initial burn in. In Figure 7.3 the 3 dimensional plot of n simulations from the GvMF distribution with parameters $\kappa_1 = 8$ and $\kappa_2 = 1$. The starting value of the chain is $\boldsymbol{\mu}_1$.

The results using MH2 are plotted in 3 dimensions in Figure 7.4. Again here we can notice the points are concentrated around the direction of $\boldsymbol{\mu}_2$ in Figure 7.4a, and around the direction of $-\boldsymbol{\mu}_2$ in Figure 7.4b. The starting value of the chain is $\boldsymbol{\mu}_2$.

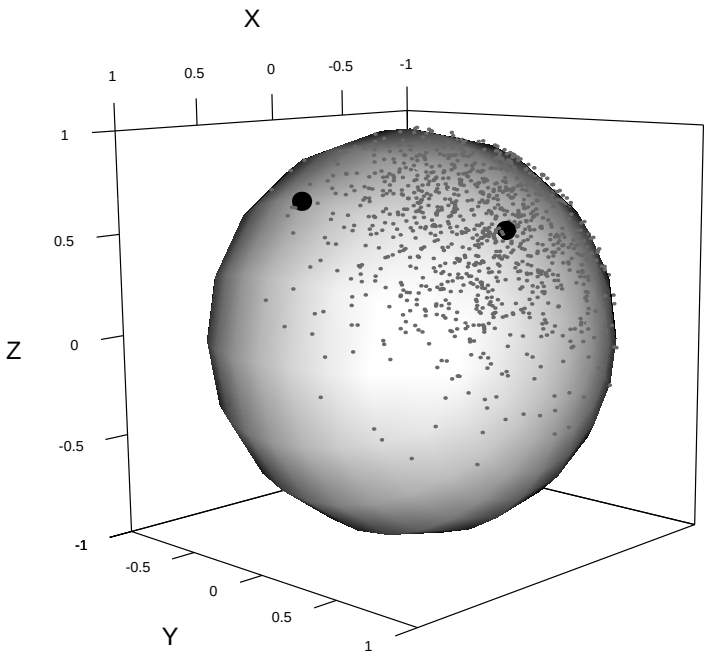
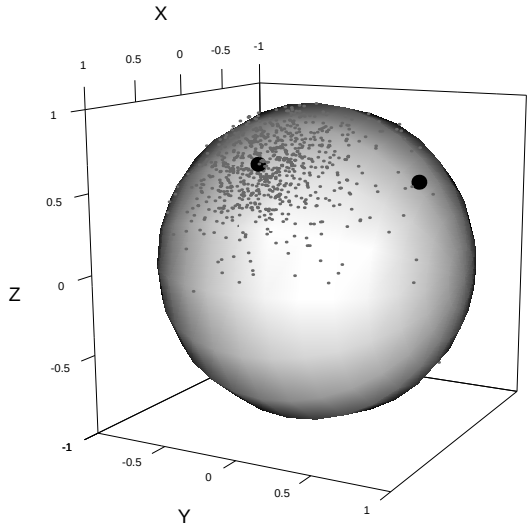


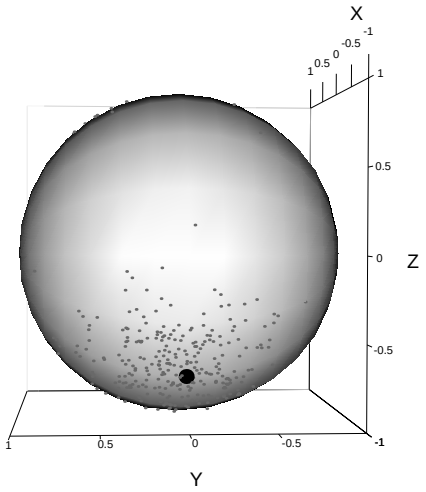
Figure 7.1: The points are the $n = 10^3$ simulations from the GvMF($\boldsymbol{\mu}_1, \boldsymbol{\mu}_2; \kappa_1, \kappa_2; 1, 2$), with $\kappa_1 = 8$ and $\kappa_2 = 1$. Right bold point is $\boldsymbol{\mu}_1$. Left bold point is $\boldsymbol{\mu}_2$.

7.3.1 Numerical comparisons: spherical coordinates

In this section we want to compare the conditional approach in Section 7.1.2 and MH in Section 7.2. To do so we compare the densities of the spherical coordinates that are given in (3.1) of the simulation results obtained with the two methods, and the efficiency in terms of the computing time and number of rejections. In order to simplify the notation, let $\boldsymbol{S}_{C1}$ and $\boldsymbol{S}_{C2}$ be the samples of size $n = 10^3$ of Section 7.1.2 simulated from the GvMF with CAR1 and CAR2, respectively. Let $\boldsymbol{S}_{MH1}$ and $\boldsymbol{S}_{MH2}$ be the samples of size $n = 10^3$ of Section 7.2 simulated with MH1 and MH2, respectively. The kernel density estimations (and the smoothing parameter or bandwidth h) of the resulting θ_1 and θ_2 are presented in Figure 7.5. The kernel density estimation of θ_1 for $\boldsymbol{S}_{C1}$ ($h = 0.075$) and $\boldsymbol{S}_{MH1}$ ($h = 0.077$) can be found in Figure 7.5a. Figure 7.5b shows the kernel density estimation of θ_2 for $\boldsymbol{S}_{C1}$ ($h = 0.107$) and $\boldsymbol{S}_{MH1}$ ($h = 0.113$). Same is done for $\boldsymbol{S}_{C2}$ and $\boldsymbol{S}_{MH2}$. The kernel density estimation of θ_1 for $\boldsymbol{S}_{C2}$ ($h = 0.1601$) and $\boldsymbol{S}_{MH2}$ ($h = 0.1612$) is shown in Figure 7.5c. Finally, in Figure 7.5d the kernel density estimation of θ_2 for $\boldsymbol{S}_{C2}$ ($h = 0.139$) and $\boldsymbol{S}_{MH2}$ ($h = 0.129$).



(a) Points concentrated around the direction of $\boldsymbol{\mu}_2$ (left bold point). Right bold point is $\boldsymbol{\mu}_1$.



(b) Points concentrated around -$\boldsymbol{\mu}_2$ (bold point).

Figure 7.2: The points are the $n = 10^3$ simulations from the GvMF($\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \kappa_1, \kappa_2; 1, 2$), with $\kappa_1 = 1$ and $\kappa_2 = 8$.

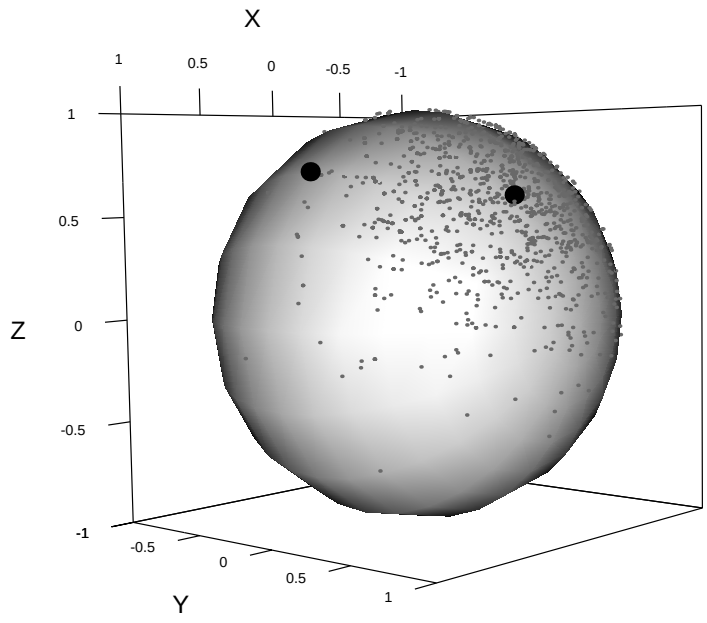


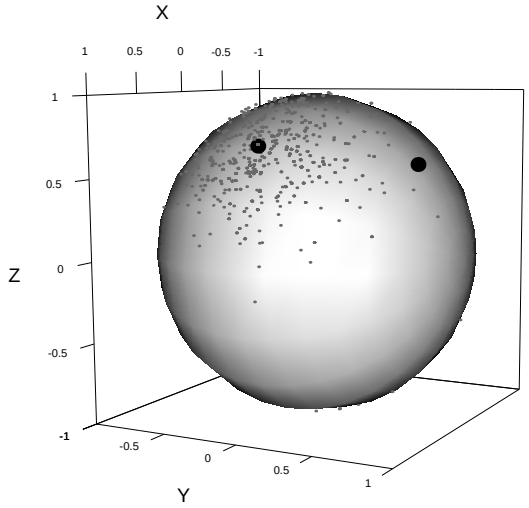
Figure 7.3: The points are the $n = 1e3$ simulations from the GvMF($\boldsymbol{\mu}_1, \boldsymbol{\mu}_2; \kappa_1, \kappa_2; 1, 2$) using MH, with $\kappa_1 = 8$ and $\kappa_2 = 1$. Right bold point is $\boldsymbol{\mu}_1$. Left bold point is $\boldsymbol{\mu}_2$.

7.3.2 Numerical comparison: rejection rates

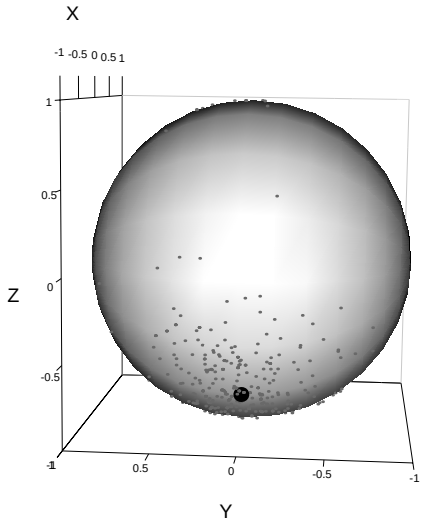
Another comparison can be done considering the rejection rates of the two approaches. The comparison of the spherical coordinates in Figure 7.5 and of the rejections rates in Table 7.1, suggests that the two approaches are both valid for the simulation. The functions to generate from the vMF, and the algorithm using acceptance-rejection with conditional approach and the MH are written in `R`.

	CAR1	MH1	CAR2	MH2
rejections	15%	15%	45%	40%

Table 7.1: Comparison of the rejections rates of the two methods.

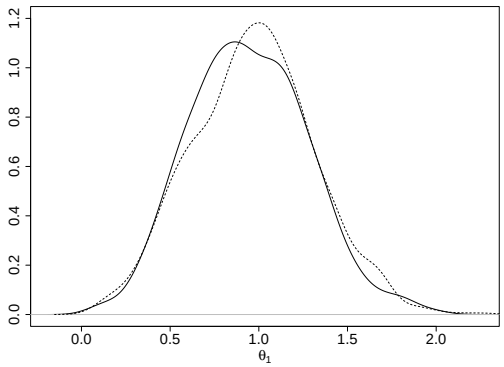


(a) Points concentrated around the direction of $\boldsymbol{\mu}_2$ (left bold point). Right bold point is $\boldsymbol{\mu}_1$.

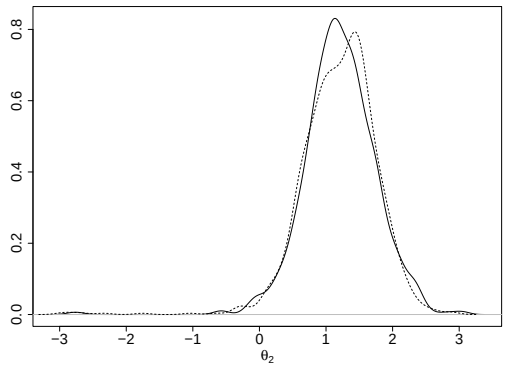


(b) Points concentrated around $-\boldsymbol{\mu}_2$ (bold point).

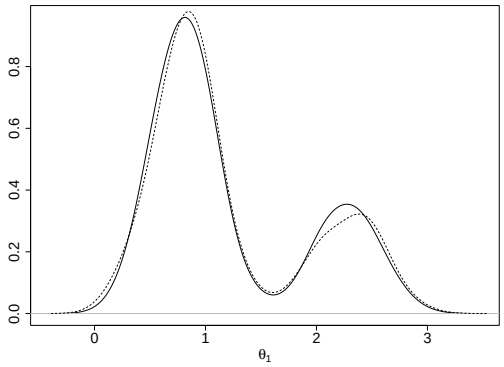
Figure 7.4: Points are the $n = 10^3$ simulations from the GvMF$(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2; \kappa_1, \kappa_2; 1, 2)$ using MH, with $\kappa_1 = 1$ and $\kappa_2 = 8$.



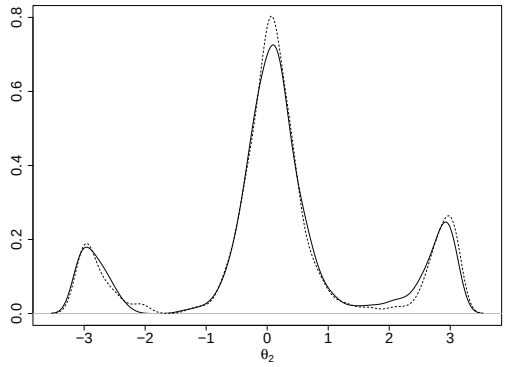
(a) Kernel density estimation of θ_1 for $\boldsymbol{S}_{C1}$ (solid line), and for $\boldsymbol{S}_{MH1}$ (dashed line).



(b) Kernel density estimation of θ_2 for $\boldsymbol{S}_{C1}$ (solid line), and for $\boldsymbol{S}_{MH1}$ (dashed line).



(c) Kernel density estimation of θ_1 for $\boldsymbol{S}_{C2}$ (solid line), and for $\boldsymbol{S}_{MH2}$ (dashed line).



(d) Kernel density estimation of θ_2 for $\boldsymbol{S}_{C2}$ (solid line), and for $\boldsymbol{S}_{MH2}$ (dashed line).

Figure 7.5: Density kernel estimations (for linear data) of the spherical coordinates of the simulations generated via acceptance-rejection with conditional approach and MH.

7.4 Conclusion

Two novel algorithms of random direction generation from the GvMF$(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2; \kappa_1, \kappa_2; 1, 2)$ distribution are proposed. This distribution seems the most interesting within the GvMF class, besides the vMF. As a by-product, we obtain a Monte Carlo estimator of the normalizing constant of the GvMF$(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2; \kappa_1, \kappa_2; 1, 2)$ distribution. By using either conditional simulation or MCMC, it should be possible to obtain generation algorithms for some other GvMF distributions or for some of the other directional exponential distributions that are given e.g. in Section 9.3.3 of Mardia and Jupp (2000).

Chapter 8

Final remarks

In this thesis we have discussed directional distributions. In particular, the circular case in $\mathbb{R}^2$ has been analysed. Here the GvM distribution have been widely discussed and some tests to test its symmetry and bimodality have been proposed. The Bayes factor and HPD regions have been considered. Results are presented, and they confirm the nature of the data. Most of the data in this thesis have been generated in `R`. A real data case is considered when testing the symmetry. Future work could involve investigating the bimodality of real data. Moreover, as mentioned, the analysis computed in Chapter 5 could be completed with the study of the inflection points of the GvM distribution.

Moreover, the general case of directional data in the unit sphere of $\mathbb{R}^p$ has been considered in this thesis. Here we computed two methods to simulate from the GvMF distribution. Plots in 3d have been reported in the most practical case of $\mathbb{R}^3$.

In this thesis, the simulations and plots are computed with the software `R`. In particular, we mention the use of the main computer cluster UBELIX of the University of Bern. This was particularly useful for the computations in Chapter 4 and Chapter 6. All the computer programs of this thesis can be found at `http://www.stat.unibe.ch`.

Appendix A

Proof of Proposition 4.1.1

Proof. The definition of axial symmetry given at the beginning of Section 4.1.3 tells that the GvM distribution is symmetric around $\alpha/2$ (or $\alpha/2+\pi$), for some $\alpha \in [0, 2\pi)$, iff

$$f(\theta|\mu_1, \mu_2, \kappa_1, \kappa_2) = f(\alpha - \theta|\mu_1, \mu_2, \kappa_1, \kappa_2), \quad \forall\theta \in [0, 2\pi).$$

This means

$$\begin{aligned} \kappa_1 \cos(\theta - \mu_1) + \kappa_2 \cos 2(\theta - \mu_2) &= \kappa_1 \cos[(\alpha - \theta) - \mu_1] + \kappa_2 \cos 2[(\alpha - \theta) - \mu_2] \\ &= \kappa_1 \cos[\theta - (\alpha - \mu_1)] + \kappa_2 \cos 2[\theta - (\alpha - \mu_2)], \end{aligned} \tag{A.1}$$

$\forall\theta \in [0, 2\pi)$. By using the cosine addition formula, (A.1) can be re-expressed as

$$\begin{aligned} &\kappa_1 \cos\theta \cos\mu_1 + \kappa_1 \sin\theta \sin\mu_1 + \kappa_2 \cos 2\theta \cos 2\mu_2 + \kappa_2 \sin 2\theta \sin 2\mu_2 = \\ &\kappa_1 \cos\theta \cos(\alpha - \mu_1) + \kappa_1 \sin\theta \sin(\alpha - \mu_1) + \kappa_2 \cos 2\theta \cos 2(\alpha - \mu_2) + \kappa_2 \sin 2\theta \sin 2(\alpha - \mu_2), \end{aligned}$$

$\forall\theta \in [0, 2\pi)$. This is equivalent to the equation

$$\begin{aligned} &\kappa_1[\cos\mu_1 - \cos(\alpha - \mu_1)] \cos\theta + \kappa_2[\cos 2\mu_2 - \cos 2(\alpha - \mu_2)] \cos 2\theta \\ &\quad + \kappa_1[\sin\mu_1 - \sin(\alpha - \mu_1)] \sin\theta + \kappa_2[\sin 2\mu_2 - \sin 2(\alpha - \mu_2)] \sin 2\theta = 0, \end{aligned}$$

$\forall\theta \in [0, 2\pi)$. It is convenient to re-express this last equation in terms of a trigonometric polynomial of degree $N = 2$, precisely as

$$p(\theta) = \sum_{j=1}^{N} (a_j \cos j\theta + b_j \sin j\theta) = 0, \ \forall\theta \in [0, 2\pi), \tag{A.2}$$

whose coefficients are given by

$$a_j = \kappa_j[\cos j\mu_j - \cos j(\alpha - \mu_j)] \ \text{ and } \ b_j = \kappa_j[\sin j\mu_j - \sin j(\alpha - \mu_j)], \ \text{for } j = 1, 2.$$

A trigonometric polynomial of degree N has maximum $2N$ roots in $[0, 2\pi)$, unless it is the null polynomial; see e.g. p. 150 of Powell et al. (1981). With this, (A.2) implies

that $p(\theta)$ is the null polynomial, which means that $a_j = b_j = 0$, for $j = 1, 2$. These four equalities give the system of equations

$$\begin{cases} \cos\mu_1 = \cos(\alpha - \mu_1), \\ \sin\mu_1 = \sin(\alpha - \mu_1), \\ \cos 2\mu_2 = \cos 2(\alpha - \mu_2), \\ \sin 2\mu_2 = \sin 2(\alpha - \mu_2), \end{cases}$$

which, in terms of $\delta = (\mu_1 - \mu_2) \bmod \pi$, simplifies to

$$\begin{cases} \alpha = 2\mu_1 + 2k\pi, \\ \alpha = 2\big[(\mu_1 - \delta) \bmod \pi\big] + k_1\pi, \end{cases} \qquad \text{for some } k, k_1 \in \mathbb{Z}.$$

One can eliminate the congruence symbol mod and obtain

$$\begin{cases} \alpha = 2\mu_1 + 2k\pi, \\ \alpha = 2\mu_1 - 2\delta + (k_1 + 2k_2)\pi, \end{cases} \qquad \text{for some } k, k_1, k_2 \in \mathbb{Z}. \tag{A.3}$$

This system of simultaneous equation admits solutions iff 2δ is a multiple of π, i.e. $2\delta = 0 \bmod \pi$. Since $\delta \in [0, \pi)$, we have found the desired symmetry characterization. □

Appendix B

Lagrange multiplier method

We provide the calculations of the Lagrange multiplier method to maximize the ratio (7.13). Let

$$r(\boldsymbol{x}) = \exp\left\{A(\boldsymbol{x})\right\},$$

with $A(\boldsymbol{x}) = \kappa_2\langle\boldsymbol{\mu}_2, \boldsymbol{x}\rangle\left[\langle\boldsymbol{\mu}_2, \boldsymbol{x}\rangle - 1\right]$. In order to find the extrema points of r with the Lagrange multiplier method, we need to solve the system of equations

$$\begin{cases} \frac{\partial\Lambda(\boldsymbol{x})}{\partial x_j} = 0, \text{ for } j = 1, \ldots, p, \\ \frac{\partial\Lambda(\boldsymbol{x})}{\partial\lambda} = 0, \end{cases} \tag{B.1}$$

where $\Lambda(\boldsymbol{x}) = r(\boldsymbol{x}) + \lambda K(\boldsymbol{x})$ is the Lagrangian, $K(\boldsymbol{x}) = (\|\boldsymbol{x}\|^2 - 1)$, $\lambda \neq 0$ and $\boldsymbol{x} = (x_1, \ldots, x_p)$. With some calculations, from (B.1) we have

$$\begin{cases} B(\boldsymbol{x})\mu_{2,j} - 2\lambda x_j = 0, \text{ for } j = 1, \ldots, p, \\ x_1^2 + \ldots + x_p^2 - 1 = 0, \end{cases} \tag{B.2}$$

where $B(\boldsymbol{x}) = \exp\left\{A(\boldsymbol{x})\right\}\kappa_2\left[2\langle\boldsymbol{\mu}_2, \boldsymbol{x}\rangle - 1\right]$ and $\boldsymbol{\mu}_2 = (\mu_{2,1}, \ldots, \mu_{2,p})$. We notice in particular that for $B(\boldsymbol{x}) = 0$, (B.2) has no solutions. From (B.2), for $x_1 \neq 0$ and $\mu_{2,1} \neq 0$ we get

$$\begin{cases} 2\lambda = \frac{1}{x_1}B(\boldsymbol{x})\mu_{2,1}, \\ x_j = \frac{\mu_{2,j}}{\mu_{2,1}}x_1, \text{ for } j = 2, \ldots, p, \\ x_1^2 + \ldots + x_p^2 - 1 = 0. \end{cases} \tag{B.3}$$

Solving (B.3) we obtain two extrema points, namely $\boldsymbol{\mu}_2$ and $-\boldsymbol{\mu}_2$. If $x_1 = 0$ and $\mu_{2,1} = 0$ similar considerations lead to the same solutions $\boldsymbol{\mu}_2$ and $-\boldsymbol{\mu}_2$. Since $\langle\boldsymbol{\mu}_2, \pm\boldsymbol{\mu}_2\rangle = \pm 1$, and that we are considering the case $Y = 1$ hence $\langle\boldsymbol{\mu}_2, \boldsymbol{x}\rangle \geq 0$, the only acceptable solution of (B.1) is $\boldsymbol{x} = \boldsymbol{\mu}_2$.

The case $Y = 0$ is the restriction of $\mathbb{S}^{p-1}$ to $\langle\boldsymbol{\mu}_2, \boldsymbol{x}\rangle < 0$. We search the extrema points of $r(\boldsymbol{x}) = \exp\{A(\boldsymbol{x})\}$ where in this case $A(\boldsymbol{x}) = \kappa_2\langle\boldsymbol{\mu}_2, \boldsymbol{x}\rangle\left[\langle\boldsymbol{\mu}_2, \boldsymbol{x}\rangle + 1\right]$. Again we need to solve (B.1), which leads to (B.2), with $B(\boldsymbol{x}) = \exp\left\{A(\boldsymbol{x})\right\}\kappa_2\left[2\langle\boldsymbol{\mu}_2, \boldsymbol{x}\rangle + 1\right]$ in this case. In the same way from (B.3) and with similar consideration as above, we

obtain two solutions namely $\boldsymbol{\mu}_2$ and $-\boldsymbol{\mu}_2$. Since in this case we are considering the case $Y = 0$ hence $\langle \boldsymbol{\mu}_2, \boldsymbol{x} \rangle < 0$, the only acceptable solution is $\boldsymbol{x} = -\boldsymbol{\mu}_2$.

Bibliography

Abe, T., A. Pewsey, and K. Shimizu (2013). Extending circular distributions through transformation of argument. *Annals of the Institute of Statistical Mathematics 65*(5), 833–858.

Abramowitz, M. (1974). *Handbook of Mathematical Functions, With Formulas, Graphs, and Mathematical Tables*. USA: Dover Publications, Inc.

Asmussen, S. and P. W. Glynn (2007). *Stochastic simulation: and analysis*, Volume 57. Springer.

Astfalck, L., E. Cripps, J. Gosling, M. Hodkiewicz, and I. Milne (2018). Expert elicitation of directional metocean parameters. *Ocean Engineering 161*, 268–276.

Basu, S. and S. Jammalamadaka (2000). Unimodality in circular data: a Bayes test. *Advances on Methodological and Applied Aspects of Probability and Statistics 1*, 4–153.

Basu, S., R. Pollack, and M. F. Roy (2003). *Algorithms in Real Algebraic Geometry*. Springer-Verlag Berlin Heidelberg.

Batschelet, E. (1981). *Circular statistics in biology*. Academic Press, New York.

Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media.

Berger, J. O. and M. Delampady (1987). Testing precise hypotheses. *Statistical Science 2*(3), 317–335.

Berger, J. O. and T. Sellke (1987). Testing a point null hypothesis: The irreconcilability of p values and evidence. *Journal of the American statistical Association 82*(397), 112–122.

Bombelli, R. (1966). *Algebra. 1572*. Milà: Feltrinelli Editore.

Box, G. E. and G. C. Tiao (1973). *Bayesian inference in statistical analysis*. Wiley.

Carpenter, W. F. (1966). On the solution of the real quartic. *Mathematics Magazine 39*(1), 28–30.

Chen, M.-H. and Q. M. Shao (1999). Monte Carlo estimation of Bayesian credible and HPD intervals. *Journal of Computational and Graphical Statistics 8*(1), 69–92.

Cheng, R. and L. Traylor (1995). Non-regular maximum likelihood problems. *Journal of the Royal Statistical Society: Series B (Methodological) 57*(1), 3–24.

Christmas, J. (2014). Bayesian spectral analysis with Student-t noise. *IEEE Transactions on Signal Processing 62*(11), 2871–2878.

Cox, D. A. (2012). *Galois theory*, Volume 61. John Wiley & Sons.

Cox, D. R. and D. V. Hinkley (1979). *Theoretical statistics*. CRC Press.

Dickson, L. E. (1917). *Elementary theory of equations*. J. Wiley.

Fisher, N. I. (1995). *Statistical analysis of circular data*. Cambridge University Press.

Garver, R. (1933). On the nature of the roots of a quartic equation. *Mathematics News Letter 7*(4), 6–8.

Gatto, R. (2008). Some computational aspects of the generalized von Mises distribution. *Statistics and Computing 18*(3), 321–331.

Gatto, R. (2009). Information theoretic results for circular distributions. *Statistics 43*(4), 409–421.

Gatto, R. (2011). The Generalized von Mises Fisher Distribution. In *Advances in Directional and Linear Statistics: A Festschrift for Sreenivasa Rao Jammalamadaka*, pp. 51–68. Physica-Verlag.

Gatto, R. (2022). Information theoretic results for stationary time series and the Gaussian-generalized von Mises time series. *Selected Papers for the Bicentennial Birth Anniversary of F. Nightingale editors B. Arnold and A. SenGupta, Springer, to appear*.

Gatto, R. and S. R. Jammalamadaka (2003). Inference for wrapped symmetric α-stable circular models. *Sankhyā: The Indian Journal of Statistics 65*(2), 333–355.

Gatto, R. and S. R. Jammalamadaka (2007). The generalized von Mises distribution. *Statistical Methodology 4*(3), 341–353.

Gatto, R. and S. R. Jammalamadaka (2014). Directional statistics: Introduction. In *Wiley StatsRef: Statistics Reference Online*, pp. 1–8. Wiley Online Library.

Gatto, R. and S. Salvador (2022). Bayesian inference on the bimodality of the generalized von Mises distribution. *Journal of Theory and Practice 16*(2), 32.

Gelman, A., D. B. Rubin, et al. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science 7*(4), 457–472.

Geman, S. and D. Geman (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence PAMI-6*(6), 721–741.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika 57*(1), 97–109.

Hoff, P. D. (2009). Simulation of the matrix Bingham-von Mises-Fisher distribution, with applications to multivariate and relational data. *Journal of Computational and Graphical Statistics 18*(2), 438–456.

Hyndman, R. J. (1990). *An algorithm for constructing highest density regions*. University of Melbourne, Department of Statistics.

Hyndman, R. J. (1996). Computing and graphing highest density regions. *The American Statistician 50*(2), 120–126.

Jammalamadaka, S. R. and A. SenGupta (2001). *Topics in circular statistics*, Volume 5. World Scientific.

Jammalamadaka, S. R. and G. Terdik (2022). Simulation and visualization of 3d-spherical distributions. In A. SenGupta and B. C. Arnold (Eds.), *Directional Statistics for Innovative Applications: A Bicentennial Tribute to Florence Nightingale*, Singapore, pp. 119–145. Springer Nature Singapore.

Jeffreys, H. (1939). *Theory of probability*. Oxford University Press, London.

Johnson, O. (2004). *Information theory and the central limit theorem*. World Scientific.

Kass, R. E. and A. E. Raftery (1995). Bayes factors. *Journal of the American Statistical Association 90*(430), 773–795.

Kass, R. E. and L. Wasserman (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association 91*(435), 1343–1370.

Kato, S. and M. Jones (2015). A tractable and interpretable four-parameter family of unimodal distributions on the circle. *Biometrika 102*(1), 181–190.

Kent, J. T., A. M. Ganeiber, and K. V. Mardia (2018). A new unified approach for the simulation of a wide class of directional distributions. *Journal of Computational and Graphical Statistics 27*(2), 291–301.

Kim, S., A. SenGupta, and B. C. Arnold (2016). A multivariate circular distribution with applications to the protein structure prediction problem. *Journal of Multivariate Analysis 143*, 374–382.

Kullback, S. (1954). Certain inequalities in information theory and the Cramer-Rao inequality. *The Annals of Mathematical Statistics 25*(4), 745–751.

Kullback, S. and R. A. Leibler (1951). On information and sufficiency. *The annals of mathematical statistics 22*(1), 79–86.

Kurz, G. and U. Hanebeck (2015). Stochastic sampling of the hyperspherical von Mises–Fisher distribution without rejection methods. In *2015 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, pp. 1–6.

Lazard, D. (1988). Quantifier elimination: optimal solution for two classical examples. *Journal of Symbolic Computation 5*(1-2), 261–266.

Ley, C. and T. Verdebout (2017). *Modern directional statistics*. CRC Press.

Ley, C. and T. Verdebout (2018). *Applied directional statistics: Modern methods and case studies*. CRC Press.

Lin, Y. and S. Dong (2019). Wave energy assessment based on trivariate distribution of significant wave height, mean period and direction. *Applied Ocean Research 87*, 47–63.

Lindley, D. V. (1977). A problem in forensic science. *Biometrika 64*(2), 207–213.

Maksimov, V. M. (1967). Necessary and sufficient statistics for the family of shifts of probability distributions on continuous bicompact groups. *Theory of Probability & Its Applications 12*(2), 267–280.

Mardia, K. V. (1972). *Statistics of directional data*. Academic press.

Mardia, K. V. and P. E. Jupp (2000). *Directional statistics*, Volume 494. John Wiley & Sons.

McVinish, R. and K. Mengersen (2008). Semiparametric Bayesian circular statistics. *Computational Statistics & Data Analysis, 52*(10), 4722–4730.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics 21*(6), 1087–1092.

Navarro, A. K. W., J. Frellsen, and R. E. Turner (2017). The multivariate generalised von Mises distribution: inference and applications. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, AAAI'17, pp. 2394–2400. AAAI Press.

Pewsey, A. (2002). Testing circular symmetry. *Canadian Journal of Statistics 30*(4), 591–600.

Pewsey, A. (2004, 06). Testing for circular reflective symmetry about a known median axis. *Journal of Applied Statistics 31*, 575–585.

Pewsey, A. and E. García-Portugués (2021). Recent advances in directional statistics. *TEST 30*(1), 1–58.

Pewsey, A., M. Neuhäuser, and G. D. Ruxton (2013). *Circular statistics in R*. Oxford University Press.

Pfyffer, S. and R. Gatto (2013). An efficient simulation algorithm for the generalized von Mises distribution of order two. *Computational Statistics 28*(1), 255–268.

Powell, M. J. D. et al. (1981). *Approximation theory and methods*. Cambridge University press.

Rees, E. (1922). Graphical discussion of the roots of a quartic equation. *The American Mathematical Monthly 29*(2), 51–55.

Robert, C. and G. Casella (2013). *Monte Carlo statistical methods*. Springer Science & Business Media.

Rodrigues (1840). Des lois géométriques qui régissent les déplacements d'un système solide dans l'espace, et de la variation des coordonnées provenant de ces déplacements considérés indépendamment des causes qui peuvent les produire. *Journal de Mathématiques Pures et Appliquées 1*(5), 380–440.

S Rao, J. and L. Ulric (2006). The effect of wind direction on Ozone levels- a case study. *Environmental and Ecological Statistics 13*, 287–298.

Salvador, S. and R. Gatto (2022a). An algebraic analysis of the bimodality of the generalized von Mises distribution. *Preprint, Institute of mathematical statistics and actuarial science, Universty of Bern*.

Salvador, S. and R. Gatto (2022b). Bayesian tests of symmetry for the generalized von Mises distribution. *Computational Statistics 37*, 947–974. Available online at https://doi.org/10.1007/s00180-021-01147-7.

Salvador, S. and R. Gatto (2022c). Generation of random directions from the generalized von Mises-Fisher distribution. *Preprint, Institute of mathematical statistics and actuarial science, Universty of Bern*.

Schmidt-Koenig, K. (1963). On the role of loft, the distance and site of release in pigeon homing (the "cross-loft experiment"). *Biol. Bull. 125*, 154–164.

Spurr, B. D. and M. A. Koutbeiy (1991). A comparison of various methods for estimating the parameters in mixtures of von Mises distributions. *Communications in Statistics-Simulation and Computation 20*(2-3), 725–741.

Ulrich, G. (1984). Computer generation of distributions on the m-sphere. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 33*(2), 158–163.

Umbach, D. and S. R. Jammalamadaka (2009). Building asymmetry into circular distributions. *Statistics & Probability Letters 79*(5), 659–663.

Wood, A. T. (1994). Simulation of the von Mises Fisher distribution. *Communications in Statistics-Simulation and Computation 23*(1), 157–164.

Wright, D. (1986). A note on the construction of highest posterior density intervals. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 35*(1), 49–53.

Yacoub, M. D. and G. Fraidenraich (2012). 96.33 A solution to the quartic equation. *The Mathematical Gazette 96*(536), 271–275.

Zellner, A. (1984). Posterior odds ratios for regression hypotheses: General considerations and some specific results. *Basic Issues in Econometrics (A.Zellner, ed) 16*(1), 275–305.

Zhang, L., Q. Li, Y. Guo, Z. Yang, and L. Zhang (2018). An investigation of wind direction and speed in a featured wind farm using joint probability distribution methods. *Sustainability 10*(12), 4338.

Declaration of consent

on the basis of Article 18 of the PromR Phil.-nat. 19

Name/First Name: Salvador Sara

Registration Number: 18 - 129 - 288

Study program: PhD in Statistics, Faculty of Science

Bachelor ☐ Master ☐ Dissertation ☑

Title of the thesis: Bayesian inference and Monte Carlo methods for directional data

Supervisor: Prof. Dr. Riccardo Gatto

I declare herewith that this thesis is my own work and that I have not used any sources other than those stated. I have indicated the adoption of quotations as well as thoughts taken from other authors as such in the thesis. I am aware that the Senate pursuant to Article 36 paragraph 1 litera r of the University Act of September 5th, 1996 and Article 69 of the University Statute of June 7th, 2011 is authorized to revoke the doctoral degree awarded on the basis of this thesis.
For the purposes of evaluation and verification of compliance with the declaration of originality and the regulations governing plagiarism, I hereby grant the University of Bern the right to process my personal data and to perform the acts of use this requires, in particular, to reproduce the written thesis and to store it permanently in a database, and to use said database, or to make said database available, to enable comparison with theses submitted by others.

Bern, 21.10.2022

Place/Date

Signature