UC Santa Barbara
UC Santa Barbara Electronic Theses and Dissertations

Title
Some Contributions to Circular and Linear Statistics

Permalink
https://escholarship.org/uc/item/08z1c4ct

Author
Jin, Qianyu

Publication Date
2020

Peer reviewed|Thesis/dissertation

University of California
Santa Barbara

Some Contributions to Circular and Linear Statistics

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy

in

Statistics & Applied

Probability

by

Qianyu Jin

Committee in charge:

Professor S. Rao Jammalamadaka, Chair
Professor John Hsu
Professor Wendy Meiring

March 2020

The Dissertation of Qianyu Jin is approved.

_____

Professor John Hsu

_____

Professor Wendy Meiring

_____

Professor S. Rao Jammalamadaka, Committee Chair

March 2020

Some Contributions to Circular and Linear Statistics

Acknowledgements

I would like to express my deepest gratitude to my advisor, Distinguished Professor S. Rao Jammalamadaka. His kind and gentle guidance, support, patience and encouragement are undoubtedly the strongest propelling power for me to finish my dissertation and will continue to be my inspiration in the future. I would like to sincerely thank my committee members Professors John Hsu and Wendy Meiring for their accommodating and friendly advises and help while serving on my committee. I am really grateful to Professor Simos G. Meintanis, Department of Economics, National and Kapodistrian University of Athens, Greece, for the opportunity to participate in the collaborative research presented in Chapter 2, as well as his valuable guidance and suggestions during the process. I am thankful to Dr. Brian Wainwright for his wonderful and diligent work as a collaborator in the research presented in Chapter 3. I am thankful to Dr. Michael Nava for his great help and practical suggestions regarding the research presented in Chapter 6.

Furthermore, I would like to thank Professors Yuedong Wang, Andrew Carter, and Sang-Yun Oh for their excellent teaching and instructions that I received from them during my stay as a graduate student. I would like to thank my friends Fuqin, Yin, Mark and Liangchen for their heartwarming support during my time at USCB.

Last but not least, I would like to thank my wonderful parents Qi'an Jin and Ling Yang for all their love and support over all these years.

Curriculum Vitæ
Qianyu Jin

Education

2020	Ph.D. in Statistics, University of California, Santa Barbara.
2012	B.S. in Statistics, University of Science and Technology of China

Experience

2018	Summer Associate, Model Governance Group – Risk and Finance, JPMorgan Chase & Co., New York.
2017	Summer Associate, Statistical Modeling and Development Team, Barclays Investment Bank, New York
2012-2018	Teaching Assistant, Department of Statistics and Applied Probability, University of California, Santa Barbara

Abstract

Some Contributions to Circular and Linear Statistics

by

Qianyu Jin

This dissertation focuses mainly on directional data in two dimensions, called "circular data," because such two-dimensional directions can be represented as points on the circumference of a unit circle. Such data, collected and analyzed by researchers in many scientific fields, needs special modeling and analysis. The thesis contains several somewhat independent results on the circular models and their analysis. First, a goodness-of-fit test for checking if a given dataset follows the wrapped stable distribution family is presented based on the empirical characteristic function. Then two dissimilarity measures for comparing any pair of curves around the circle are introduced and their use are explored in clustering such curves. This is followed by proving a result showing that wrapping a convolution of any number of linear components, yields the convolution of the corresponding wrapped distributions. Testing symmetry within the family of sine-skewed von Mises distributions is considered and compared with an existing test. The final result is a departure from the directional domain, and presents a Bayesian test for the number of modes in a two-component Gaussian mixture.

Contents

Chapter 1

Introduction

1.1 Introduction to Circular Data

Directional data deals with measurement of directions in two, three, or higher dimensions, which are collected and analyzed by researchers in many scientific fields. Some common examples of directional data are the directions in real world, such as those dealing with bird migration, ocean currents, or magnetic pole direction of the earth.

Two-dimensional directions taken on the plane, can be represented as points on the circumference of a unit circle and are referred to as circular data. In general, there are two ways to represent circular data. A two-dimensional direction may be represented as an angle on $[0, 2\pi)$. However, to ensure the angles are meaningful and unique, one needs to specify a zero direction and a sense of rotation i.e. whether we are measuring angles going clockwise or counterclockwise. Alternatively, such circular data may be represented as points on the circumference of the unit circle centered at the origin, or equivalently as unit vectors within this coordinate system. Even here, to ensure the uniqueness of a point and its vector representation, one needs to specify the directions of the axes of the coordinate system.

Circular data, when represented by angles, have properties that are quite different from linear data defined on the real line. For example, circular data has periodicity with a period of 2π since rotating a direction by 2π results in the same direction. Another point is that one can not order circular data by magnitude since they depend on the choice of zero direction and the sense of rotation. Also the mean of a set of angles is not computed by taking their arithmetic mean. A straightforward way to visualize the circular mean is to convert circular data to unit vectors, then the circular mean is the direction of the resultant or sum of all these unit vectors. The mean direction can be undefined if all the unit vectors sum to a zero vector. Due to various unique properties of circular data, we need to be careful when analyzing circular data and make sure the correct methodology is used. See Jammalamadaka and SenGupta (2001) [20] for more details.

1.2   Review of the Thesis

In Chapter 2, we present a goodness-of-fit test for wrapped stable distribution family, where the test statistic is based on comparing the model-based characteristic function with the empirical (data-based) characteristic function. We evaluate the proposed new test with respect to other known tests for goodness-of-fit, including Kuiper's Test and Watson's Test, via a simulation study with samples drawn from various alternative circular distribution families. An example of real-data application is also presented.

In Chapter 3, we introduce a hierarchical clustering procedure for periodic curves around the circle, by approximating them as a mixture of an appropriate number of von Mises (vM) distributions. We choose the L^2 distance to measure dissimilarities between such curves after deriving explicit formulae for the L^2 distance. The dissimilarity matrix is used as the basis upon which the clustering hierarchy is constructed. We illustrate the

effectiveness of this clustering procedure with a simulation study, where the true clusters are known, as well as a real data application where the true clusters are unknown.

In Chapter 4, we present some general results on wrapped distributions, which are obtained by wrapping a linear distribution around a circle. It was shown earlier in Jammalamadaka and Kozubowski (2017) [18] that wrapping a mixture distribution gives the same result as mixing the corresponding wrapped components. We now show that a similar result holds for convolutions viz. wrapping a convolution of any number of linear components, yields the convolution of the corresponding wrapped distributions. As an illustration, we use this approach for deriving a wrapped Quasi Lindley distribution, and a convolution of two wrapped Quasi Lindley distributions.

In Chapter 5, we consider the family of sine-skewed von Mises (SvM) distributions, which provides a generalization of the classical vM model, incorporating possible asymmetry. Given a random sample from such a general family, we consider the problem of testing for symmetry within this class. The powers of likelihood ratio test and an earlier test proposed by Batschelet (1965) [5] are computed and compared.

Chapter 6 is a bit of departure from all the rest of the work on circular data, and tackles a somewhat classical but not fully resolved problem of checking for the number of modes in a two-component Gaussian mixture on the real line. We use a Bayesian approach and the test statistic uses the Bayes Factor with the null hypothesis of unimodality versus the alternative hypothesis representing bimodality. Value of the test statistic is computed numerically using Markov Chain Monte Carlo method and a simulation study shows the effectiveness of the test. We conclude with an application on a real dataset of adult human heights.

Chapter 2

Goodness-of-Fit Test for Wrapped Stable Distributions Based on the Characteristic Function

2.1 Introduction

The Wrapped Stable (WS) distribution is one of the most flexible models for circular data. Specifically this four–parameter family of distributions includes symmetric as well as asymmetric members, with varying tail features ranging from the medium–tailed wrapped Normal distribution to the heavy–tailed wrapped Cauchy distribution, and many others.

A convenient way to describe the WS distribution is by means of its characteristic function (ChF) which is given by

$$C_{\varphi}(r) = \begin{cases} e^{i\mu r - \tau^{\gamma}|r|^{\gamma}\{1 - i\delta \text{sgn}(r)\tan(\pi\gamma/2)\}}, & \gamma \neq 1, \\ e^{i\mu r - \tau|r|\{1 + i\delta \frac{2}{\pi}\text{sgn}(r)\log|r|\}}, & \gamma = 1, \end{cases} \tag{2.1}$$

4

where (μ, τ) are location and scale parameters, and (γ, δ) are shape and skewness parameters, respectively. We will write $\boldsymbol{\varphi} = (\gamma, \delta, \tau, \mu)$ for the entire vector of parameters, with the parameter space specified by $(\mu, \tau) \in [0, 2\pi) \times (0, \infty)$ and $(\gamma, \delta) \in (0, 2] \times [-1, 1]$. The parameter γ is often called the characteristic exponent and regulates tail–behavior. Specifically smaller values of γ progressively lead from lighter to heavier tails. On the other hand δ is a skewness parameter, with $\delta = 0$ corresponding to a distribution that is symmetric, while with increasing $|\delta|$, and as δ approaches $+1$ (resp. -1), the density becomes asymmetric to the right (resp. to the left). In this connection note that as $\gamma \to 2$, the parameter δ loses its significance with $\gamma = 2$ rendering the wrapped Normal distribution. The other well known member of WS distribution family is the wrapped Cauchy distribution for $(\gamma, \delta) = (1, 0)$, which together with the wrapped Normal, is one of the three isolated cases with a closed form density, with all other WS distributions admitting only series representations of densities. See Jammalamadaka and SenGupta (2001) [20], pp.44-48 for a brief discussion of these distributions for modeling circular data.

In this chapter we suggest a class of Goodness-of-Fit (GoF) tests for the family of WS distributions which utilizes the ChF of these distributions. Specifically let Θ be an arbitrary circular random variable. Then on the basis of independent and identically distributed (i.i.d.) random samples $\vartheta_1, ..., \vartheta_n$ on Θ we are interested in testing the composite null hypothesis,

$$\mathcal{H}_0 : \Theta \text{ follows } \mathcal{S}_{\boldsymbol{\varphi}}, \text{ for some } \boldsymbol{\varphi} \in \Phi, \tag{2.2}$$

against general alternatives, where $\mathcal{S}_{\boldsymbol{\varphi}} = \{S(\cdot; \boldsymbol{\varphi}), \boldsymbol{\varphi} \in \Phi\}$ denotes the family of WS distributions with cumulative distribution function $S(\cdot; \boldsymbol{\varphi})$, and Φ stands for the corresponding parameter space.

We note that the ChF of an arbitrary linear random variable at a given integer

5

argument r is equal to the trigonometric moment of order r of the corresponding distribution wrapped around the unit circle (see Prop. 2.1 in Jammalamadaka and SenGupta (2001) [20]). The use of this tool for performing statistical inference in the circular context has been hitherto mostly confined to the case of testing for uniformity or symmetry. We refer to the early work of Beran (1969) [7], and to the more recent contributions of Pycke (2010) [29] and Meintanis and Verdebout (2019) [26] and references therein. This is despite the fact that earlier ChF–based GoF methods, for conventional (linear) stable distributions for example, have proved to be more convenient to apply and to compete well with other GoF methods; see Csörgő (1987) [9], Koutrouvelis and Meintanis (1999) [22], Matsui and Takemura (2008) [23], Meintanis (2005) [24] and Meintanis et al. (2015) [25].

The chapter is organized as follows. In Section 2.2 we introduce the new testing method, while in Section 2.3 testing for the symmetric WS distribution against general alternatives is studied in detail. In Section 2.4 the finite–sample properties of an appropriate resampling version of the test are illustrated by means of a Monte Carlo study. In Section 2.5 we consider empirical applications, and finally we end in Section 2.6 with conclusions and discussion.

2.2   Tests Based on the Characteristic Function

In order to test the null hypothesis $\mathcal{H}_0$ specified in (2.2) we propose to use a distance metric between the ChF of the WS distribution and the empirical ChF based on the data, defined by

$$C_n(r) = \frac{1}{n} \sum_{j=1}^{n} e^{\mathrm{i}r\vartheta_j} := \alpha_n(r) + \mathrm{i}\beta_n(r), \mathrm{i} = \sqrt{-1}, \tag{2.3}$$

where

$$\alpha_n(r) = \frac{1}{n} \sum_{j=1}^{n} \cos(r\vartheta_j), \quad \beta_n(r) = \frac{1}{n} \sum_{j=1}^{n} \sin(r\vartheta_j), \tag{2.4}$$

are the Cartesian coordinates of the empirical ChF. For integer r, $\alpha_n(r)$ and $\beta_n(r)$ are also called the empirical trigonometric moments.

In fact since for circular distributions the ChF needs to be evaluated only at the integers (Jammalamadaka and SenGupta (2001),[20] §2.2), and taking into account the symmetry property of the ChF and the empirical ChF, our test statistic is formulated only on the basis of theoretical and empirical trigonometric moments. Specifically we suggest to reject the null hypothesis $\mathcal{H}_0$ for large values of the test criterion

$$T_{n,f} = n \sum_{r \geq 0} |C_n(r) - C_{\widehat{\varphi}}(r)|^2 \, f(r), \tag{2.5}$$

where $\widehat{\varphi}$ is a suitable estimator of the parameter φ, and $f(\cdot)$ denotes a "weight function" which we take to be a probability function over the non–negative integers.

By straightforward algebra we have from (2.5)

$$T_{n,f} = n \sum_{r \geq 0} \left\{ |C_n(r)|^2 + |C_{\widehat{\varphi}}(r)|^2 - 2 \left(\alpha_n(r)\alpha_{\widehat{\varphi}}(r) + \beta_n(r)\beta_{\widehat{\varphi}}(r) \right) \right\} f(r), \tag{2.6}$$

where $\alpha_{\varphi}(r)$ and $\beta_{\varphi}(r)$ denote the population trigonometric moments of the WS distribution, and $|z|^2$ stands for the modulus of a complex number z.

2.3   Tests for the Symmetric WS Distribution

2.3.1   The Test Criterion and Consistency

The case of testing for the symmetric wrapped stable (SWS) distributions is considerably simpler. Specifically by proper location shift we can simply test the null hypothesis that $\mathcal{H}_0^{(s)} : C \equiv C_{\gamma,\tau}$ where $C_\Theta(r) = \mathbb{E}(e^{ir\Theta})$ is the ChF of Θ and $C_{\gamma,\tau}(\cdot)$ denotes the ChF of the zero–location SWS distribution that results from (2.1) by replacing $\boldsymbol{\varphi}$ by $(\gamma, 0, \tau, 0)$. In fact the test statistic figuring in (2.5) may conveniently be reparameterized as

$$T_{n,f} = n \sum_{r \geq 0} \left| \widehat{C}_n(r) - \widehat{\varrho}^{r^{\widehat{\gamma}}} \right|^2 f(r), \tag{2.7}$$

where $\widehat{C}_n(r)$ is the empirical ChF resulting from (2.3) by replacing ϑ_j by

$$\widehat{\vartheta}_j = \vartheta_j - \widehat{\mu}, j = 1, ..., n,$$

and where $\widehat{\gamma}$, $\widehat{\mu}$ and $\widehat{\varrho}$ denote consistent estimators of the parameters γ, μ and $\varrho := \exp\{-\tau^\gamma\}$, respectively (discussed later in Section 2.3.3).

Before going any further, we will investigate the consistency of the test based on $T_{n,f}$ against all fixed alternative circular distributions. To this end assume that the estimator $\widehat{\boldsymbol{\varphi}}^{(s)} := (\widehat{\gamma}, \widehat{\varrho}, \widehat{\mu})$ attains a strong probability limit, say $\boldsymbol{\varphi}_A^{(s)} := (\gamma_a, \varrho_a, \mu_a)$, even under alternatives. Also suppose that $f(r) > 0$.

Now from (2.7) write $T_{n,f} = n\Delta_{n,f}^{(s)}$ and notice that $\left| \widehat{C}_n(r) - \widehat{\varrho}^{r^{\widehat{\gamma}}} \right|^2 \leq 4$, so by invoking Lebesgue's theorem of dominated convergence we have

$$\Delta_{n,f}^{(s)} \longrightarrow \sum_{r \geq 0} \left| e^{-i\mu_a r} C_\Theta(r) - \varrho_a^{r^{\gamma_a}} \right|^2 f(r) := \Delta_f, \text{ a.s. as } n \to \infty, \tag{2.8}$$

by means of the strong pointwise consistency of the empirical ChF (see Feuerverger and Mureika, 1977 [13]). To proceed further recall the definition of the parameter ϱ and thus replace $\varrho_{\mathrm{a}}{}^{r^{\gamma_{\mathrm{a}}}}$ by $C_{\gamma_{\mathrm{a}},\tau_{\mathrm{a}}}(r) = e^{-(\tau_{\mathrm{a}} r)^{\gamma_{\mathrm{a}}}}$, where $\tau_{\mathrm{a}} = (-\log \varrho_{\mathrm{a}})^{1/\gamma_{\mathrm{a}}}$. Clearly then Δ_f is positive, unless $C_{\Theta}(r) = e^{\mathrm{i}\mu_{\mathrm{a}} r - (\tau_{\mathrm{a}} r)^{\gamma_{\mathrm{a}}}}$, identically in r, which by the uniqueness of the ChF would imply that Θ follows a SWS distribution indexed by $\varphi_{\mathrm{A}}^{(\mathrm{s})}$, and leads to the strong consistency of the test that rejects the null hypothesis $\mathcal{H}_0^{(\mathrm{s})}$ for large values of $T_{n,f}$.

2.3.2   Computations

Our point of departure for computations is Equation (2.6). In this connection recall from the previous section that under the null hypothesis $\mathcal{H}_0^{(\mathrm{s})}$, we have that $\beta_{\boldsymbol{\varphi}}(\cdot) \equiv 0$, and hence the test criterion figuring in (2.6) reduces to

$$T_{n,f} = \Sigma_1 + \Sigma_2 - 2\Sigma_3, \tag{2.9}$$

with

$$\Sigma_1 = \frac{1}{n} \sum_{j,k=1}^{n} \mathcal{E}_1(\widehat{\vartheta}_j - \widehat{\vartheta}_k), \tag{2.10}$$

$$\Sigma_2 = n\mathcal{E}_2(\widehat{\gamma}, \widehat{\tau}), \tag{2.11}$$

and

$$\Sigma_3 = \sum_{j=1}^{n} \mathcal{E}_3(\widehat{\vartheta}_j; \widehat{\gamma}, \widehat{\tau}), \tag{2.12}$$

where the series figuring in (2.10)-(2.12) are defined by

$$\mathcal{E}_1(\theta) = \sum_{r \geq 0} \cos(\theta r) f(r),$$

9

$$\mathcal{E}_2(\gamma, \tau) = \sum_{r \geq 0} e^{-2\tau^\gamma r^\gamma} f(r),$$

and

$$\mathcal{E}_3(\theta; \gamma, \tau) = \sum_{r \geq 0} \cos(\theta r) e^{-\tau^\gamma r^\gamma} f(r).$$

Now recall that $f(r)$, $r \geq 0$, is a probability function and therefore $\mathcal{E}_k$, $k = 1, 2, 3$, may be viewed as expectations of corresponding quantities taken with respect to the probability mass function (PMF) $f(r)$. Then after some extra manipulations it follows that these three series may be written as

$$\mathcal{E}_1(\theta) = \mathbb{E}_r[\cos(\theta r)], \tag{2.13}$$

$$\mathcal{E}_2(\gamma, \tau) = \mathbb{E}_r \mathbb{E}_\vartheta[\cos(2^{1/\gamma} \vartheta r)], \tag{2.14}$$

and

$$\mathcal{E}_3(\theta; \gamma, \tau) = \frac{1}{2}\left\{ \mathbb{E}_r \mathbb{E}_\vartheta[\cos((\vartheta - \theta)r)] + \mathbb{E}_r \mathbb{E}_\vartheta[\cos((\vartheta + \theta)r)] \right\}, \tag{2.15}$$

where $\mathbb{E}_r[\cdot]$ and $\mathbb{E}_\vartheta[\cdot]$ are meant as expectations taken with respect to the PMF $f(r)$ and with respect to the SWS distribution with parameter $\boldsymbol{\varphi} = (\gamma, 0, \tau, 0)$, respectively.

While the expectations figuring in (2.13)–(2.15) are generally not easy to compute analytically, they nevertheless allow for some simplification if the PMF $f(r)$ corresponding to the weights is properly chosen. Specifically letting $f(r)$ be a Poisson distribution with parameter λ, we have

$$\mathcal{E}_1(\theta) := \mathcal{E}_1(\theta; \lambda) = \cos(\lambda \sin \theta) e^{\lambda(\cos \theta - 1)}. \tag{2.16}$$

Moreover since $\mathcal{E}_k$, $k = 2, 3$, are absolutely convergent, by application of Fubini's theorem

10

we have $\mathbb{E}_r\mathbb{E}_\vartheta(\cdot) = \mathbb{E}_\vartheta\mathbb{E}_r(\cdot)$, so that the expectations in (2.14)–(2.15) may be Monte–Carlo approximated by the quantities

$$\mathcal{E}_{2,M}(\gamma,\tau) = \frac{1}{M}\sum_{m=1}^{M}\mathcal{E}_1(2^{1/\gamma}\theta_m),$$

and

$$\mathcal{E}_{3,M}(\theta;\gamma,\tau) = \frac{1}{2}\frac{1}{M}\sum_{m=1}^{M}\left(\mathcal{E}_1(\theta_m - \theta) + \mathcal{E}_1(\theta_m + \theta)\right),$$

respectively, where M is a large positive integer, and θ_m, $m = 1, ..., M$, denote i.i.d. observations from the SWS distribution with parameter $\boldsymbol{\varphi} = (\gamma, 0, \tau, 0)$.

Alternatively, since we have rapid convergence at least for $\gamma \geq 1$, one can compute the test statistic $T_{n,f}$ given in (2.7) by means of direct numerical computation of only a few terms of the series defining $\mathcal{E}_2$ and $\mathcal{E}_3$.

2.3.3 Estimation of Parameters

From Equation (2.7) it is clear that the test criterion requires prior estimation of the stable distribution parameters. As estimators of the parameters (γ, ϱ, μ) of a SWS distribution we suggest the moment estimator given by the following equations (see §4.6 of Jammalamadaka and SenGupta (2001) [20]) :

$$\widehat{\mu} = \arctan^*\left(\frac{\beta_n(1)}{\alpha_n(1)}\right), \quad \widehat{\varrho} = \sqrt{\alpha_n^2(1) + \beta_n^2(1)}, \quad \widehat{\gamma} = \frac{1}{\log 2}\log\left(\frac{\log(\sqrt{\alpha_n^2(2) + \beta_n^2(2)})}{\log(\sqrt{\alpha_n^2(1) + \beta_n^2(1)})}\right),$$
$$\tag{2.17}$$

where $\alpha_n(\cdot), \beta_n(\cdot)$ are the trigonometric moments defined in (2.4), and $\arctan^*(\cdot)$ denotes the principal inverse of $\tan(\cdot)$, which maps the tangent of an angle on $[0, 2\pi)$ into the correct quadrant.

We note that moment estimation of parameters yields an interesting limit for the test

statistic $T_{n,f}$ figuring in (2.7). To this end notice that $C_n(0) = C(0) = 1$, and hence the first term in $T_{n,f}$ vanishes regardless of the distribution being tested and for any function $f(r)$ used as probability function. In addition the second term involves the quantity $|\widehat{C}_n(1) - \widehat{\varrho}|$ which also vanishes on the account of the first two equations in (2.17) and because $\widehat{\beta}_n(1) = 0$. Then write $T_{n,\lambda}$ for the criterion in (2.7) with $f(r)$ being the Poisson probability function with parameter λ, so that

$$T_{n,\lambda} = n \, e^{-\lambda} \left(\left| \widehat{C}_n(2) - \widehat{\varrho}^{2\widehat{\gamma}} \right|^2 \frac{\lambda^2}{2} + \mathrm{o}(\lambda^2) \right), \ \lambda \to 0,$$

which leads to

$$\lim_{\lambda \to 0} \frac{2T_{n,\lambda}}{n\lambda^2} = \left| \widehat{C}_n(2) - \widehat{\varrho}^{2\widehat{\gamma}} \right|^2 := T_{n,0}. \tag{2.18}$$

Clearly the limit statistic $T_{n,0}$ simply compares the trigonometric moment computed from the sample with the trigonometric moment of the SWS distribution, both of order $r = 2$. (In this connection note that $T_{n,0}$ vanishes as $n \to \infty$ under the null hypothesis). On the other hand the test statistic $T_{n,\lambda}$ employs an infinite weighted sum in which the empirical trigonometric moments of all integer orders $r \geq 0$ are accounted for, and thus the probability function $f(r)$ plays the role of a weight function. Consequently if this function is decreasing with r (which is typically the case at least for large r), $f(r)$ downweights higher order trigonometric moments that are more prone to periodic behavior; see for instance Epps (1993) [11].

2.4 Finite-Sample Comparisons

This section summarizes the results of a simulation study designed to evaluate the performance of the proposed test and compare it with other existing tests. As competitors we include the Kuiper's test (K) and the Watson's test (W) for which there

exist computationally convenient formulae (Jammalamadaka and SenGupta, [20] §7.2.1). Specifically recall the notation $S(\cdot; \boldsymbol{\varphi})$ for the distribution function of the WS distribution, let $U_j = S(\vartheta_j; \widehat{\boldsymbol{\varphi}})$ and write $U_{(j)}$, $j = 1, ..., n$, for the corresponding order statistics. Then we have

$$\mathrm{K} = \max_{1 \leq j \leq n} \left\{ U_{(j)} - \frac{j-1}{n} \right\} + \max_{1 \leq j \leq n} \left\{ \frac{j}{n} - U_{(j)} \right\}.$$

$$\mathrm{W} = \frac{1}{12n} + \sum_{j=1}^{n} \left(\left(U_{(j)} - \frac{2j-1}{2n} \right) - \left(\overline{U} - \frac{1}{2} \right) \right)^2,$$

where $\overline{U} = n^{-1} \sum_{j=1}^{n} U_j$.

It is well known however that the null distribution of goodness–of–fit statistics such as $T_{n,f}$, K and W, is complicated and depends on several unknown quantities. Therefore here we implement all tests based on parametric bootstrap resampling, which is an automatic computer–based procedure for performing this task. Next we outline the steps of the parametric bootstrap procedure within a fairly general setting of testing the null hypothesis $\mathcal{H}_0$ in (2.2) by means of an arbitrary criterion $T := T(\vartheta_1, \ldots, \vartheta_n)$. More specifically write $\widehat{T} := T(\vartheta_1, \ldots, \vartheta_n; \widehat{\boldsymbol{\varphi}})$ for this test criterion involving the original observations as well as the resulting parameter estimate $\widehat{\boldsymbol{\varphi}} := \widehat{\boldsymbol{\varphi}}(\vartheta_1, ..., \vartheta_n)$. Then parametric bootstrap critical points are computed as follows:

(i) Generate i.i.d. observations, $\{\vartheta_j^*, 1 \leq j \leq n\}$ from $S(\cdot; \widehat{\boldsymbol{\varphi}})$.

(ii) Using these observations obtain the bootstrap estimate $\widehat{\boldsymbol{\varphi}}^* := \widehat{\boldsymbol{\varphi}}(\vartheta_1^*, ..., \vartheta_n^*)$ of $\boldsymbol{\varphi}$.

(iii) Calculate the bootstrap test statistic, say $T^* := T(\vartheta_1^*, ..., \vartheta_n^*; \widehat{\boldsymbol{\varphi}}^*)$.

(iv) Repeat steps (i) to (iii) a number of times, say B, and obtain $\{T_b^*\}_{b=1}^{\mathrm{B}}$.

(v) Calculate the critical point of a test of size α as the order $(1-\alpha)$ empirical quantile $T_{1-\alpha}^*$ of $T_{(b)}^*$, $(b = 1, ..., \mathrm{B})$, where $T_{(1)}^* \leq T_{(2)}^* \leq \ldots \leq T_{(\mathrm{B})}^*$, are the corresponding order statistics.

(vi) Reject the null hypothesis if $\widehat{T} > T^*_{1-\alpha}$.

As the parametric bootstrap is somewhat time–consuming, we employ instead its warp–speed version under which for each Monte Carlo sample we draw just one bootstrap resample and compute the test statistic for this resample. Then the critical point is computed as in step (v) above as the $(1 - \alpha)$ empirical quantile of $T^*_{(m)}$, $(m = 1, ..., \mathrm{MC})$, where T^*_m is the bootstrap statistic resulting from the mth Monte Carlo sample and MC denotes the number of Monte Carlo samples; see Giacomini et al. (2013) [15].

In this simulation study, we choose six distributions from the family of SWS distribution (under the null hypothesis) and eight distributions from alternative hypotheses, including: (i) von Mises (vM) distributions, (ii) 2-component mixtures of vM distributions, (iii) asymmetric WS (AWS) distributions, (iv) 2-component mixtures of a symmetric WS distribution with an AWS distribution, (v) generalized von Mises (GvM) distributions (introduced by Gatto and Jammalamadaka, 2007 [14]). Table 2.1 displays the parameters for these distributions and Figure 2.1 illustrates the density functions of the alternative hypothesis distributions. The computations were carried out in R, utilizing CircStats package to generate random samples for WS and vM distributions. We choose $\lambda = 0.3, 0.5, 0.7, 0.9$, with $f(r)$ being the Poisson probability function with mean λ, and write T_λ for our test statistic. The proportion of rejections of the null hypothesis for each distribution is computed from applying the tests to MC=5,000 samples, with the critical points estimated using the warp–speed method. The results for sample size $n = 100, 50$ and $n = 30$ are summarized in Table 2.2, 2.3, and 2.4, respectively, at nominal level $\alpha = 0.05$ and $\alpha = 0.10$. On the basis of these simulations results we make the following observations: (1) The empirical level of our test is close to the nominal level for Null 1 (which is at the boundary of the parameter space), and Null 2, 3 and 6, while

it is higher than α for Null 4 and 5. (This kind of inaccuracy was also observed in the Monte Carlo results of Koutrouvelis and Meintanis, 1999 [22]). (2) The empirical power of our test is high for Alt. 1, 2, 3, 6 and 8, while it is very low (almost indistinguishable from level of the test) for Alt. 4 and 5. Intuitively, this is probably due to the fact that distributions Alt. 4 and Alt. 5 are unimodal and symmetrical, and thus they can be approximated sufficiently well by a SWS distribution. (3) Compared with our test, both the Kuiper's and the Watson's tests exhibit lower percentage of rejection both under the null as well as under alternatives, except for Null 1. Generally, these tests perform much worse or no better than our test in most cases, with occasional exceptions like Null 4 with $n = 100$ and Alt. 3 with $n = 30$. In particular, for Alt. 8, both tests have almost no power while our test has high power when the sample size is large. (4) The test statistic T_λ is robust with respect to λ and thus the value of λ does not affect significantly the empirical level/power of our test. In other testing situations though finding a good value for λ turns out to be important; see for instance Allison and Santana (2015) [2] for more details.

Figure 2.1: Density functions under alternative hypotheses

Table 2.1: Null and Alternative hypotheses used for data simulation

Null 1	$\text{SWS}(2, 0.5, \pi)$
Null 2	$\text{SWS}(1, 0.5, \pi)$
Null 3	$\text{SWS}(1, 1, \pi)$
Null 4	$\text{SWS}(1.9, 1, \pi)$
Null 5	$\text{SWS}(1.75, 1, \pi)$
Null 6	$\text{SWS}(1.5, 1, \pi)$
Alternative 1	$0.8\text{vM}(\pi, 5) + 0.2\text{vM}(\pi/2, 5)$
Alternative 2	$0.65\text{vM}(\pi, 5) + 0.35\text{vM}(\pi/2, 5)$
Alternative 3	$0.5\text{vM}(\pi, 5) + 0.5\text{vM}(\pi/2, 5)$
Alternative 4	$(1/3)\text{vM}(\pi, 8) + (2/3)\text{vM}(\pi, 0.1)$
Alternative 5	$\text{vM}(\pi/2, 2)$
Alternative 6	$\text{AWS}(1.2, -1, 0.5, \pi)$
Alternative 7	$0.5\text{WS}(2, 0, 0.75, 0) + 0.5\text{WS}(1.5, -1, 0.5, \pi)$
Alternative 8	$\text{GvM}(\pi, 3\pi/4, 1, 5)$

Table 2.2: Observed proportion of rejection at nominal level α for 5000 Monte Carlo samples of size $n = 100$. T_λ represents proposed test with $Poisson(\lambda)$ weight.

Hypothesis	α	K	W	$T_{0.3}$	$T_{0.5}$	$T_{0.7}$	$T_{0.9}$	T_1
Null 1	0.05	0.0490	0.0488	0.0540	0.0554	0.0590	0.0592	0.0606
	0.10	0.0976	0.1070	0.1076	0.1078	0.1070	0.1080	0.1078
Null 2	0.05	0.0100	0.0136	0.0470	0.0468	0.0454	0.0446	0.0418
	0.10	0.0378	0.0432	0.0962	0.0960	0.0940	0.0920	0.0924
Null 3	0.05	0.0080	0.0144	0.0332	0.0332	0.0312	0.0310	0.0310
	0.10	0.0344	0.0414	0.0728	0.0714	0.0708	0.0672	0.0670
Null 4	0.05	0.0440	0.0594	0.1008	0.1008	0.0984	0.0980	0.0972
	0.10	0.0922	0.1170	0.1834	0.1796	0.1754	0.1722	0.1688
Null 5	0.05	0.0314	0.0474	0.0926	0.0890	0.0872	0.0832	0.0838
	0.10	0.0740	0.1016	0.1576	0.1558	0.1524	0.1478	0.1448
Null 6	0.05	0.0224	0.0308	0.0628	0.0624	0.0606	0.0582	0.0558
	0.10	0.0620	0.0708	0.1122	0.1118	0.1106	0.1096	0.1082
Alt. 1	0.05	0.7288	0.9072	0.9842	0.9860	0.9868	0.9864	0.9862
	0.10	0.8612	0.9588	0.9946	0.9952	0.9952	0.9944	0.9946
Alt. 2	0.05	0.9334	0.9868	0.9824	0.9862	0.9888	0.9898	0.9904
	0.10	0.9720	0.9944	0.9944	0.9950	0.9958	0.9960	0.9962
Alt. 3	0.05	0.9104	0.9784	0.9554	0.9660	0.9704	0.9758	0.9762
	0.10	0.9598	0.9918	0.9834	0.9874	0.9890	0.9904	0.9912
Alt. 4	0.05	0.0000	0.0004	0.0378	0.0356	0.0334	0.0328	0.0318
	0.10	0.0028	0.0084	0.0904	0.0862	0.0824	0.0792	0.0772
Alt. 5	0.05	0.0394	0.0418	0.0474	0.0482	0.0464	0.0466	0.0448
	0.10	0.0908	0.0892	0.0954	0.0954	0.0944	0.0950	0.0950
Alt. 6	0.05	0.2942	0.3170	0.6962	0.7148	0.7252	0.7298	0.7308
	0.10	0.5256	0.5788	0.8158	0.8340	0.8416	0.8442	0.8436
Alt. 7	0.05	0.0026	0.0376	0.5182	0.4958	0.4736	0.4476	0.4366
	0.10	0.0146	0.1096	0.6366	0.6152	0.6002	0.5828	0.5720
Alt. 8	0.05	0.0000	0.0002	0.9470	0.9332	0.9098	0.8722	0.8484
	0.10	0.0002	0.0010	0.9816	0.9750	0.9672	0.9522	0.9430

Table 2.3: Observed proportion of rejection at nominal level α for 5000 Monte Carlo samples of size $n = 50$. T_λ represents proposed test with $Poisson(\lambda)$ weight.

Hypothesis	α	K	W	$T_{0.3}$	$T_{0.5}$	$T_{0.7}$	$T_{0.9}$	T_1
Null 1	0.05	0.0492	0.0528	0.0506	0.0478	0.0478	0.0490	0.0490
	0.10	0.0930	0.0966	0.0992	0.1038	0.1074	0.1084	0.1070
Null 2	0.05	0.0048	0.0106	0.0346	0.0358	0.0356	0.0344	0.0336
	0.10	0.0314	0.0382	0.0776	0.0756	0.0742	0.0766	0.0760
Null 3	0.05	0.0032	0.0078	0.0314	0.0308	0.0290	0.0276	0.0274
	0.10	0.0180	0.0266	0.0706	0.0684	0.0650	0.0642	0.0636
Null 4	0.05	0.0146	0.0352	0.1104	0.1090	0.1060	0.1024	0.1026
	0.10	0.0486	0.0804	0.1898	0.1848	0.1808	0.1754	0.1742
Null 5	0.05	0.0154	0.0292	0.0854	0.0838	0.0802	0.0774	0.0762
	0.10	0.0462	0.0748	0.1520	0.1534	0.1516	0.1494	0.1480
Null 6	0.05	0.0094	0.0172	0.0634	0.0636	0.0632	0.0602	0.0584
	0.10	0.0358	0.0456	0.1278	0.1246	0.1188	0.1156	0.1138
Alt. 1	0.05	0.2980	0.5064	0.7880	0.7864	0.7860	0.7790	0.7762
	0.10	0.4702	0.6666	0.8756	0.8752	0.8770	0.8726	0.8666
Alt. 2	0.05	0.5884	0.7778	0.7698	0.7962	0.8134	0.8284	0.8312
	0.10	0.7292	0.8654	0.8892	0.9008	0.9088	0.9098	0.9106
Alt. 3	0.05	0.5860	0.7630	0.6566	0.7036	0.7324	0.7516	0.7640
	0.10	0.7248	0.8714	0.8170	0.8412	0.8594	0.8734	0.8758
Alt. 4	0.05	0.0010	0.0032	0.0494	0.0486	0.0452	0.0442	0.0422
	0.10	0.0096	0.0158	0.0924	0.0902	0.0870	0.0842	0.0826
Alt. 5	0.05	0.0370	0.0342	0.0434	0.0422	0.0418	0.0380	0.0392
	0.10	0.0834	0.0820	0.0884	0.0836	0.0860	0.0844	0.0838
Alt. 6	0.05	0.0378	0.0350	0.3272	0.3422	0.3434	0.3458	0.3476
	0.10	0.1446	0.1362	0.4980	0.5080	0.5100	0.5184	0.5212
Alt. 7	0.05	0.0036	0.0120	0.2524	0.2406	0.2238	0.2128	0.2048
	0.10	0.0190	0.0622	0.3958	0.3816	0.3670	0.3522	0.3424
Alt. 8	0.05	0.0000	0.0010	0.6582	0.5946	0.5378	0.4674	0.4234
	0.10	0.0000	0.0046	0.7940	0.7544	0.7052	0.6556	0.6198

Table 2.4: Observed proportion of rejection at nominal level α for 5000 Monte Carlo samples of size $n = 30$. T_λ represents proposed test with $Poisson(\lambda)$ weight.

Hypothesis	α	K	W	$T_{0.3}$	$T_{0.5}$	$T_{0.7}$	$T_{0.9}$	T_1
Null 1	0.05	0.0410	0.0490	0.0434	0.0472	0.0500	0.0498	0.0508
	0.10	0.0928	0.0906	0.1060	0.1062	0.1036	0.1044	0.1016
Null 2	0.05	0.0036	0.0060	0.0326	0.0328	0.0298	0.0280	0.0290
	0.10	0.0226	0.0314	0.0808	0.0796	0.0814	0.0762	0.0758
Null 3	0.05	0.0034	0.0074	0.0322	0.0330	0.0324	0.0306	0.0294
	0.10	0.0206	0.0266	0.0818	0.0806	0.0778	0.0718	0.0720
Null 4	0.05	0.0054	0.0144	0.0838	0.0816	0.0772	0.0746	0.0738
	0.10	0.0348	0.0524	0.1496	0.1484	0.1474	0.1426	0.1402
Null 5	0.05	0.0076	0.0158	0.0808	0.0792	0.0770	0.0760	0.0754
	0.10	0.0328	0.0528	0.1462	0.1468	0.1410	0.1370	0.1366
Null 6	0.05	0.0052	0.0108	0.0606	0.0590	0.0560	0.0530	0.0524
	0.10	0.0250	0.0386	0.1124	0.1098	0.1054	0.1042	0.1048
Alt. 1	0.05	0.0814	0.1570	0.4998	0.5120	0.5088	0.5040	0.4994
	0.10	0.2314	0.3570	0.6342	0.6402	0.6384	0.6272	0.6240
Alt. 2	0.05	0.2622	0.4296	0.5218	0.5398	0.5588	0.5656	0.5724
	0.10	0.4366	0.6104	0.6954	0.7150	0.7274	0.7288	0.7288
Alt. 3	0.05	0.2924	0.4466	0.3868	0.4352	0.4822	0.5088	0.5164
	0.10	0.4696	0.6230	0.5976	0.6226	0.6424	0.6546	0.6608
Alt. 4	0.05	0.0030	0.0046	0.0560	0.0546	0.0518	0.0502	0.0474
	0.10	0.0132	0.0216	0.1078	0.1052	0.1010	0.0990	0.0980
Alt. 5	0.05	0.0198	0.0182	0.0412	0.0400	0.0396	0.0394	0.0398
	0.10	0.0658	0.0594	0.0912	0.0924	0.0888	0.0870	0.0866
Alt. 6	0.05	0.0154	0.0158	0.1774	0.1826	0.1832	0.1830	0.1840
	0.10	0.0702	0.0660	0.3138	0.3146	0.3192	0.3224	0.3212
Alt. 7	0.05	0.0036	0.0094	0.1626	0.1542	0.1446	0.1358	0.1290
	0.10	0.0220	0.0434	0.2606	0.2582	0.2502	0.2388	0.2320
Alt. 8	0.05	0.0000	0.0020	0.4076	0.3774	0.3432	0.2756	0.2512
	0.10	0.0000	0.0058	0.5652	0.5356	0.4782	0.4268	0.4050

2.5   Real-Data Application

This section shows the application of our proposed test on a couple of real data sets. Taylor and Burns (2016) [34] collected data sets for the radial distributions of mistletoes and epiphytes from 5 different species, and discovered that they are highly directional

and related to the availability of light and humidity. We consider two of their data sets: Data set 1 consists of $n = 67$ observations on *peraxilla colensoi* and Data set 2 consists of $n = 65$ observations on *asplenium flaccidum*. Figure 2.2 shows the histograms of two data sets. The corresponding fitted symmetric WS distribution density functions are shown in black lines. Table 2.5 gives the estimated parameters and p-values of the tests for Data Set 1 and 2. At $\alpha = 0.05$, all tests fail to reject the null hypothesis for Data Set 1 and reject the null hypothesis for Data Set 2.

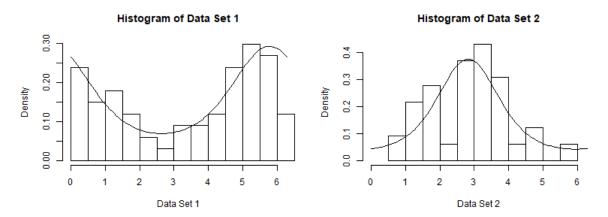Figure 2.2: Histograms of real data sets together with density functions of fitted symmetric WS distributions

Table 2.5: Parameter estimates, and p-values of the tests for Data Sets 1 and 2

Data set	Estimated Parameters	K	W	$C_{0.3}$	$C_{0.5}$	$C_{0.7}$	$C_{0.9}$	C_1
1	$\widehat{\gamma} = 1.32, \widehat{\tau} = 1.06, \widehat{\mu} = 5.78$	0.502	0.586	0.316	0.344	0.370	0.412	0.434
2	$\widehat{\gamma} = 1.40, \widehat{\tau} = 0.79, \widehat{\mu} = 2.82$	0.036	0.012	0.028	0.020	0.014	0.008	0.006

2.6 Discussion

We suggest a goodness–of–fit test for family of SWS distributions with unknown parameters. The proposed test statistic is based on the characteristic function of this family which unlike the corresponding density may be written in closed form. Furthermore the

methods used for estimating the stable distribution parameters also utilize the characteristic function, and thus avoid numerically complicated likelihood–based procedures. The suggested test criterion, which is be expressed as a weighted L^2–type distance between empirical trigonometric moments and the corresponding theoretical quantities, is shown to be consistent against general alternatives. The findings of a Monte Carlo study show that the new test criteria compete well with classical procedures, while a couple of real–data examples further illustrate the applicability of the new procedures.

Chapter 3

Distance-Based Clustering for von Mises Mixtures

3.1   Introduction

This chapter introduces a hierarchical clustering procedure for periodic curves around a circle. Such curves are generated by measuring the thickness of the neuroretinal rim (NRR) obtained in Optical Coherence Tomography (OCT). We develop a model-based clustering by fitting a mixture of vM densities for each curve after proper scaling. Suppose we have n such circular curves presented in a dataset. Our goal is to cluster these curves based on the similarity of their underlying distributions. For doing this, we need to define a metric measuring the distance between each pair of curves, and then use hierarchical clustering method. The general steps would be:

1. Draw samples from the corresponding circular density of each curve that are obtained by scaling the curve,

2. Estimate vM mixture distribution for each curve from its sample,

3. Compute the dissimilarity or distance matrix between the estimated curves,

4. Apply a hierarchical clustering method to this distance matrix and obtain the hierarchy of clusters,

5. Choose the proper number of clusters and group the curves accordingly.

Specifically, in Step 1 we choose to approximate each curve by a vM mixture due to its flexibility. In fact, analogous to the result that any distribution on the real line can be approximated by a normal mixture with countable number of components (see Teicher(1960) [35] and Ferguson(1983) [12]), one may show that any circular distribution can be approximated by a vM mixture with countable number of components.

In Step 2, we would like to use L^2 distance, a widely used measure for the dissimilarity between two curves. Section 3.2 and 3.3 will give the explicit form of L^2 distance for vM mixtures with arbitrary number of components.

3.2 L^2 Distance Between Two vM Distributions

L^2 distance is a metric for measuring the dissimilarity between two probability distributions. The definition of L^2 distance for two circular distributions with probability density functions (PDF) $f(\alpha)$ and $g(\alpha)$, where $\alpha \in [0, 2\pi)$, is presented as follows.

$$L^2(f,g) = \int_0^{2\pi} (f(\alpha) - g(\alpha))^2 d\alpha \qquad (3.1)$$

When both f and g are vM distributions, the explicit form of their L^2 distance can be derived with the help of the following lemma.

Lemma 3.2.1 *(See Jammalamadaka and SenGupta (2001) [20], p.40) For any two vM*

distributions $f \sim vM(\mu_1, \kappa_1)$ and $g \sim vM(\mu_2, \kappa_2)$,

$$\int_0^{2\pi} f(\alpha)g(\alpha)d\alpha = \frac{I_0(\kappa)}{2\pi I_0(\kappa_1)I_0(\kappa_2)} \tag{3.2}$$

where

$$\kappa = \sqrt{\kappa_1^2 + \kappa_2^2 + 2\kappa_1\kappa_2\cos(\mu_1 - \mu_2)}$$

Proof: By following straightforward algebra,

$$\int_0^{2\pi} f(\alpha)g(\alpha)d\alpha = \frac{1}{(2\pi)^2 I_0(\kappa_1)I_0(\kappa_2)} \int_0^{2\pi} e^{\kappa_1\cos(\alpha-\mu_1)+\kappa_2\cos(\alpha-\mu_2)}d\alpha$$

$$= \frac{1}{(2\pi)^2 I_0(\kappa_1)I_0(\kappa_2)} \int_0^{2\pi} e^{(\kappa_1\cos\mu_1+\kappa_2\cos\mu_2)\cos\alpha+(\kappa_1\sin\mu_1+\kappa_2\sin\mu_2)\sin\alpha}d\alpha$$

$$= \frac{1}{(2\pi)^2 I_0(\kappa_1)I_0(\kappa_2)} \int_0^{2\pi} e^{A\cos\alpha+B\sin\alpha}d\alpha$$

under reparameterization, where

$$A = \kappa_1\cos\mu_1 + \kappa_2\cos\mu_2,$$

$$B = \kappa_1\sin\mu_1 + \kappa_2\sin\mu_2.$$

We can rewrite them as $A = \kappa\cos\mu, B = \kappa\sin\mu$, where μ and κ are given by

$$\mu = \arctan\frac{B}{A},$$

and

$$\kappa^2 = A^2 + B^2$$

$$= \kappa_1^2 \cos^2 \mu_1 + \kappa_2^2 \cos^2 \mu_2 + 2\kappa_1\kappa_2 \cos \mu_1 \cos \mu_2$$

$$+ \kappa_1^2 \sin^2 \mu_1 + \kappa_2^2 \sin^2 \mu_2 + 2\kappa_1\kappa_2 \sin \mu_1 \sin \mu_2$$

$$= \kappa_1^2 + \kappa_2^2 + 2\kappa_1\kappa_2 \cos(\mu_1 - \mu_2).$$

Under this representation, there is

$$\int_0^{2\pi} f(\alpha)g(\alpha)d\alpha = \frac{1}{(2\pi)^2 I_0(\kappa_1)I_0(\kappa_2)} \int_0^{2\pi} e^{\kappa \cos \mu \cos \alpha + \kappa \sin \mu \sin \alpha}d\alpha$$

$$= \frac{1}{(2\pi)^2 I_0(\kappa_1)I_0(\kappa_2)} \int_0^{2\pi} e^{\kappa \cos(\alpha - \mu)}d\alpha$$

$$= \frac{I_0(\kappa)}{2\pi I_0(\kappa_1)I_0(\kappa_2)}.$$

∎

Proposition 3.2.1 (L^2 **distance between two vM distributions**) *For any two vM distributions* $f \sim vM(\mu_1, \kappa_1)$ *and* $g \sim vM(\mu_2, \kappa_2)$, *the* L^2 *distance is*

$$L^2(f, g) = \frac{1}{2\pi} \left(\frac{I_0(2\kappa_1)}{I_0(\kappa_1)^2} + \frac{I_0(2\kappa_2)}{I_0(\kappa_2)^2} - \frac{2I_0(\kappa)}{I_0(\kappa_1)I_0(\kappa_2)} \right) \tag{3.3}$$

where

$$\kappa = \sqrt{\kappa_1^2 + \kappa_2^2 + 2\kappa_1\kappa_2 \cos(\mu_1 - \mu_2)}$$

Proof: By Lemma 3.2.1,

$$\int_0^{2\pi} f^2(\alpha)d\alpha = \frac{I_0(\sqrt{\kappa_1^2 + \kappa_1^2 + 2\kappa_1\kappa_1 \cos(\mu_1 - \mu_1)})}{2\pi I_0(\kappa_1)I_0(\kappa_1)}$$

$$= \frac{I_0(2\kappa_1)}{2\pi (I_0(\kappa_1))^2},$$

26

and similarly,

$$\int_0^{2\pi} g^2(\alpha)d\alpha = \frac{I_0(2\kappa_2)}{2\pi(I_0(\kappa_2))^2}.$$

Thus,

$$\begin{aligned}
L^2(f,g) &= \int_0^{2\pi} (f(\alpha) - g(\alpha))^2 d\alpha \\
&= \int_0^{2\pi} f^2(\alpha)d\alpha + \int_0^{2\pi} g^2(\alpha)d\alpha - 2\int_0^{2\pi} f(\alpha)g(\alpha)d\alpha \\
&= \frac{1}{2\pi}\left(\frac{I_0(2\kappa_1)}{(I_0(\kappa_1))^2} + \frac{I_0(2\kappa_2)}{(I_0(\kappa_2))^2} - \frac{2I_0(\kappa)}{I_0(\kappa_1)I_0(\kappa_2)} \right)
\end{aligned}$$

■

3.2.1 L^2 Distance Between Two vM Mixtures

Now consider two vM mixtures, h_1 and h_2, with PDFs given by

$$h_1(\alpha) = \sum_{i=1}^k p_i f_i(\alpha), \quad h_2(\alpha) = \sum_{j=1}^l q_j g_j(\alpha), \tag{3.4}$$

where f_i and g_j are vM PDFs and the mixture proportions satisfy the restriction $\sum_{i=1}^k p_i = \sum_{j=1}^l q_j = 1$. Then, by definition, the L^2 distance is

$$\begin{aligned}
L_2(h_1, h_2) &= \int_0^{2\pi} \left(\sum_{i=1}^k p_i f_i(\alpha) - \sum_{j=1}^l q_j g_j(\alpha) \right)^2 d\alpha \\
&= \sum_{i=1}^k \sum_{i'=1}^k p_i p_{i'} \int_0^{2\pi} f_i(\alpha)f_{i'}(\alpha)d\alpha + \sum_{j=1}^l \sum_{j'=1}^l q_j q_{j'} \int_0^{2\pi} g_j(\alpha)g_{j'}(\alpha)d\alpha \\
&\quad - 2\sum_{i=1}^k \sum_{j=1}^l p_i q_j \int_0^{2\pi} f_i(\alpha)g_j(\alpha)d\alpha. \tag{3.5}
\end{aligned}$$

And the integrals are computed using Lemma 3.2.1.

3.3 Another Dissimilarity Measure—Symmetric Kullback-Liebler Divergence

We would like to compare the L^2 distance with another dissimilarity measure—Symmetric Kullback-Liebler (SKL) divergence. The SKL divergence is a symmetrized version of the KL divergence. For two circular distributions with PDFs f and g, the KL divergence is defined as

$$KL(f||g) = \int_0^{2\pi} \log\left(\frac{f(\alpha)}{g(\alpha)}\right) f(\alpha) d\alpha.$$

KL divergence is not symmetric since in general $KL(f||g) \neq KL(g||f)$. This problem can be rectified using the SKL divergence defined as

$$SKL(f,g) = KL(f||g) + KL(g||f).$$

Wainwright (2019) [37] gives the explicit formula for SKL between two vM distributions, and suggested a numerical method for computing the SKL between two vM mixture. The results are presented as follows.

For any two vM distributions $f \sim vM(\mu_1, \kappa_1)$ and $g \sim vM(\mu_2, \kappa_2)$, the SKL divergence is given by

$$SKL(f,g) = \kappa_1 A(\kappa_1) + \kappa_2 A(\kappa_2) - \cos(\mu_1 - \mu_2)\left[\kappa_2 A(\kappa_1) + \kappa_1 A(\kappa_2)\right]. \tag{3.6}$$

And for any two vM mixtures h_1, h_2 defined in (3.4), the SKL divergence is given by

$$SKL(h_1, h_2) = \int_0^{2\pi} \log\left(\frac{h_1(\alpha)}{h_2(\alpha)}\right) h_1(\alpha)d\alpha + \int_0^{2\pi} \log\left(\frac{h_2(\alpha)}{h_1(\alpha)}\right) h_2(\alpha)d\alpha. \tag{3.7}$$

In general (3.7) cannot be computed analytically and one may use numerical integration methods to evaluate these integrals quite accurately.

3.3.1   A Visual Comparison of the L^2 and SKL Measures for vM Distributions

Figure 3.1 and 3.2 illustrate the two measures side-by-side in terms of how they change by varying $|\mu_1 - \mu_2|$, or $\kappa_2 - \kappa_1$, or both. Although the magnitudes of the two measures are different—for instance, the SKL divergence is about 10 times the L^2 distance for the same pair of vM mixtures, yet they would provide comparable clustering results when used as dissimilarity measurements, judging by the similarities in the shapes of the line plots and 3D plots between the two.

3.4   Clustering Circular Curves—A Simulation Study

The simulation study aims at assessing the effectiveness of our proposed clustering procedure, and will focus on two aspects: (1) whether the density estimates are accurate, and (2) whether the clusters obtained using the samples can identify the similarities between the true distributions that generated them.

3.4.1   Evaluating Estimations of the Curves

As explained in Section 3.1, we use vM mixtures to approximate the true circular distributions. Parameter estimation for vM mixture can be accomplished in R using
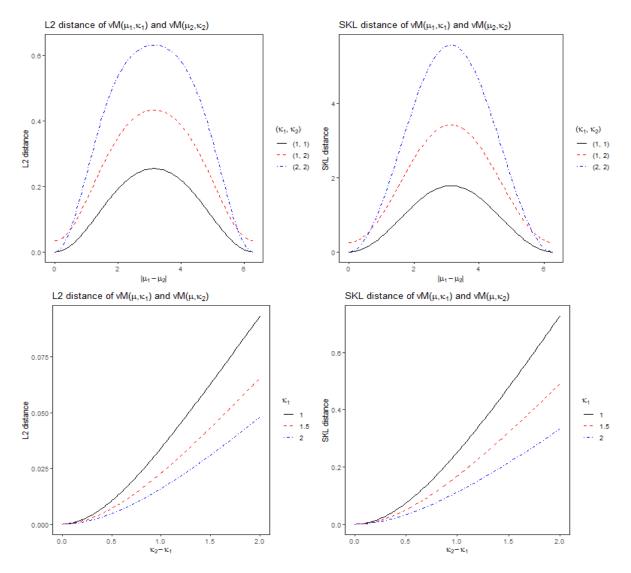
Figure 3.1: Top: L^2 distance (left) and SKL distance (right) vs $|\mu_1 - \mu_2|$, for fixed κ_1, κ_2. Bottom: L^2 distance (left) and SKL distance (right) vs $\kappa_2 - \kappa_1$, for fixed $\mu_1 = \mu_2 = \mu$ and κ_1. Note that the y-axis scales are different for L^2 and SKL distance.

the `movMF` package developed by Hornik and Grün [17]. This package implements EM algorithm variants to estimate parameters for mixture of von Mises-Fisher distribution, a high-dimensional spherical distribution that reduces to vM mixture when the distribution lies in $\mathbb{R}^2$. Note that this algorithm requires number of mixture components k to be specified. For real data, k is often unknown, and our strategy is to fit the mixture model
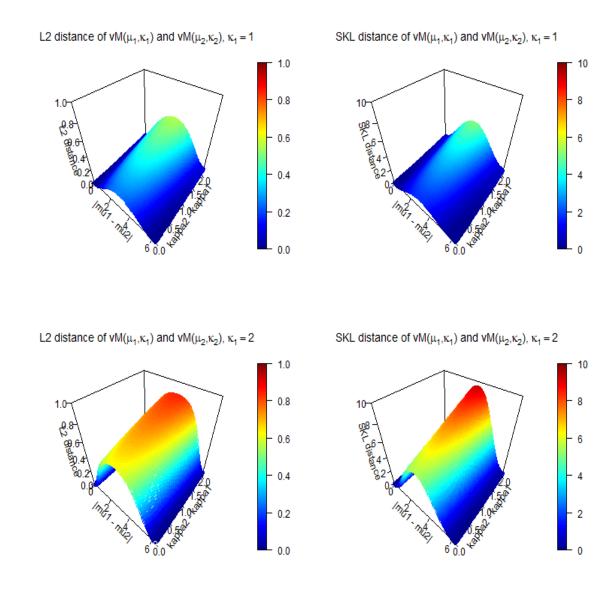
Figure 3.2: 3D plots for L^2 distance (left) and SKL distance (right) vs $|\mu_1 - \mu_2|$ and $\kappa_2 - \kappa_1$ for fixed κ_1. Note that the color scales are different for L^2 and SKL distance.

for a series of k values and select the one with minimum BIC value.

We would like to explore whether the algorithm can give desirable vM mixture estimates in 4 simulated case studies. For each case study, two 2-component vM mixtures and one 3-component vM mixture are selected as the true distributions, and 5 samples

are generated from each distribution. The choices of parameters for true distributions are given in Table 3.1. Case 1 contains vM mixtures that have equal mixture weights p, moderately high κ's, and with differences between μ's that are large enough such that the number of components k equals the number of modes. Cases 2, 3 and 4 are all variants of Case 1. Case 2 has smaller $\kappa's$, Case 3 has smaller μ differences such that its 2-component vM mixtures is indeed unimodal, and Case 4 has unequal p's and thus unequal heights for each mode. These cases aim at showing how sample-based parameter estimation and component selection results are affected by the underlying mixture distributions.

Figure 3.3 compares the estimated densities with the corresponding true densities. In the grid of plots, the rows represent case studies and the columns represent the three different vM mixture distributions within each case. In any particular plot, the black dashed line is the true density, while the colored solid lines are the five associated estimated densities. Case 1 shows good consistency between true and estimated densities, except that the red curve in vM mixture 3 is too flat and only has $k = 1$ component. Case 2 shows that smaller κ values can impair this consistency, as wrong k (number of components) are often chosen. Case 3 shows that the selected k values match with the number of modes, if not at their true values. Nevertheless, the true and estimated density curves are consistent when it is not near the modes. Case 4 show that the estimation procedure sometimes fails to identify those smaller modes, which correspond to the components with smaller p's.

From the results above we can summarize that, in general this algorithm gives better parameter estimation results for vM mixture when (1) the concentration parameters of vM components are large, (2) the mixing probabilities do not have a small value, and (3) the number of modes match with the number of vM components since the modes are well separated. Although if condition (3) is not met and the algorithm cannot recover the correct number of mixing components, the estimated density curve is still reasonably

Case	Mixture	Component	$\boldsymbol{\mu}$ (rad.)	$\boldsymbol{\kappa}$	$\boldsymbol{p}$
Case 1	vM-mix 1	$f_1(\alpha; \mu, \kappa)$	0	4.0	0.5
		$f_2(\alpha; \mu, \kappa)$	$2\pi/3$ (≈ 2.09)	4.0	0.5
	vM-mix 2	$f_1(\alpha; \mu, \kappa)$	π (≈ 3.14)	3.0	0.5
		$f_2(\alpha; \mu, \kappa)$	$5\pi/3$ (≈ 5.24)	3.0	0.5
	vM-mix 3	$f_1(\alpha; \mu, \kappa)$	0	5.0	0.333
		$f_2(\alpha; \mu, \kappa)$	$2\pi/3$ (≈ 2.09)	5.0	0.333
		$f_3(\alpha; \mu, \kappa)$	$4\pi/3$ (≈ 4.19)	5.0	0.333
Case 2	vM-mix 1	$f_1(\alpha; \mu, \kappa)$	0	2.0	0.5
		$f_2(\alpha; \mu, \kappa)$	$2\pi/3$ (≈ 2.09)	2.0	0.5
	vM-mix 2	$f_1(\alpha; \mu, \kappa)$	π (≈ 3.14)	2.0	0.5
		$f_2(\alpha; \mu, \kappa)$	$5\pi/3$ (≈ 5.24)	2.0	0.5
	vM-mix 3	$f_1(\alpha; \mu, \kappa)$	0	4.0	0.333
		$f_2(\alpha; \mu, \kappa)$	$2\pi/3$ (≈ 2.09)	4.0	0.333
		$f_3(\alpha; \mu, \kappa)$	$4\pi/3$ (≈ 4.19)	4.0	0.333
Case 3	vM-mix 1	$f_1(\alpha; \mu, \kappa)$	0	4.0	0.5
		$f_2(\alpha; \mu, \kappa)$	$\pi/3$ (≈ 1.05)	4.0	0.5
	vM-mix 2	$f_1(\alpha; \mu, \kappa)$	π (≈ 3.14)	3.0	0.5
		$f_2(\alpha; \mu, \kappa)$	$4\pi/3$ (≈ 4.19)	3.0	0.5
	vM-mix 3	$f_1(\alpha; \mu, \kappa)$	0	5.0	0.333
		$f_2(\alpha; \mu, \kappa)$	$2\pi/3$ (≈ 2.09)	5.0	0.333
		$f_3(\alpha; \mu, \kappa)$	$4\pi/3$ (≈ 4.19)	5.0	0.333
Case 4	vM-mix 1	$f_1(\alpha; \mu, \kappa)$	0	4.0	0.75
		$f_2(\alpha; \mu, \kappa)$	$2\pi/3$ (≈ 2.09)	4.0	0.25
	vM-mix 2	$f_1(\alpha; \mu, \kappa)$	π (≈ 3.14)	3.0	0.25
		$f_2(\alpha; \mu, \kappa)$	$5\pi/3$ (≈ 5.24)	3.0	0.75
	vM-mix 3	$f_1(\alpha; \mu, \kappa)$	0	5.0	0.20
		$f_2(\alpha; \mu, \kappa)$	$2\pi/3$ (≈ 2.09)	5.0	0.60
		$f_3(\alpha; \mu, \kappa)$	$4\pi/3$ (≈ 4.19)	5.0	0.20

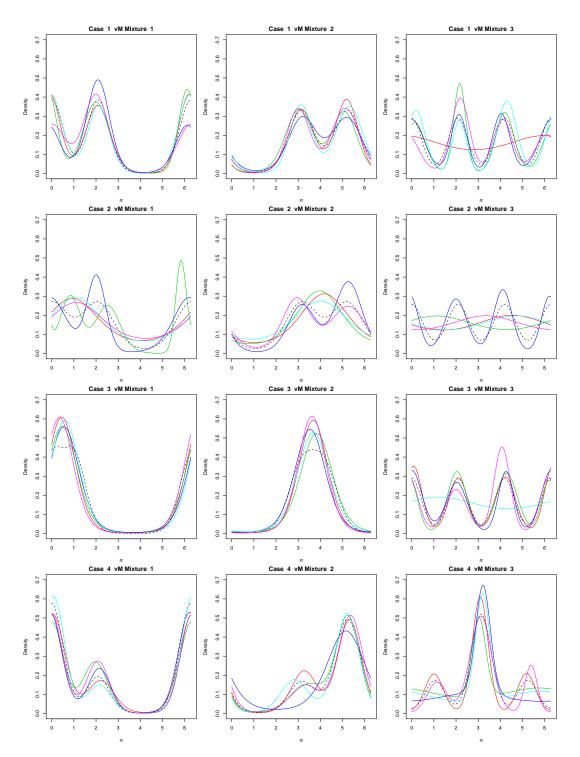Table 3.1: Simulated Data Parameter Values

Figure 3.3: Estimated Density Curves vs. True Density Curves. The three plots in each row corresponds to the three vM mixtures in each case. Black dashed lines are the true density curves and colored solid lines are estimated density curves.
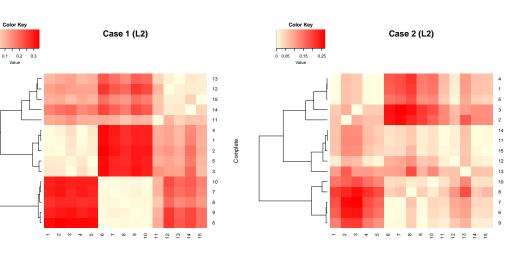
close to the true density curve.

3.4.2   Evaluation of Clustering Results

After estimating the vM mixture densities for each sample we are able to use formula (3.5) to compute pairwise L^2 distances and construct the distance matrix. Then a hierarchical clustering method can be applied in order to build a hierarchy of clusters, which can be represented by a dendrogram. Each level of the dendrogram shows the clustering result for certain number of clusters.

In each case study, if our proposed clustering procedure is ideal, we should be able to see a complete separation at 3 clusters, i.e. each of the 3 clusters should only contain the samples generated from the same true density. Table 3.2 show that L^2-based clustering is able to recover the correct cluster membership for all the samples. Figure 3.4 display row end dendrograms and heatmaps for these two clustering metrics. For Case 1, 3, and 4, the between-cluster L^2 distances are much larger than within-cluster L^2 distances. The heatmap patterns suggest that, for arbitrary samples denoted by A, B, C and D, the distance $L^2(A, B) \approx L^2(C, D)$ as long as A, C are in the same cluster and B, D are in the same cluster. For Case 2, the heatmap pattern is more fuzzy, mainly due to the less desirable consistency between estimated and true density curves as discussed earlier.

	vM-mix 1					vM-mix 2					vM-mix 3				
	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15
Case 1	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3
Case 2	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3
Case 3	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3
Case 4	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3

Table 3.2: L^2-based Cluster Membership

Figure 3.4: L^2 Clustering Heatmap (Simulated Data)

3.5 A Real-Data Application

The data of interest [30] contains periodic curves that measure the thickness of the neuroretinal rim (NRR), which are made with Optical Coherence Tomography (OCT), a non-invasive imaging technique that can generate high-resolution images of the retina, retinal nerve fiber layer and optic nerve head. For each eye, the NRR is a ring-like region between the margin of the optic disc and optic cup, and its center is the same as the

center of the optic cup. The measurements of NRR thickness can be made at any angle with respect to the center and thus form a continuous circular curve.

In the data set we use there are 100 NRR curves $S^{(i)}, i = 1, ..., 100$, each observed at $\alpha_j = (j-1)\pi/90, j = 1, ..., 180$, the 180 equally-spaced angles on $[0, 2\pi)$. Suppose we are interested in clustering the curves only based on their shape characteristics and not their magnitudes, then we can scale each NRR curve to a circular density

$$f^{(i)}(\alpha) = \frac{S^{(i)}(\alpha)}{\int_0^{2\pi} S^{(i)}(\alpha)d\alpha}, \alpha \in [0, 2\pi) \tag{3.8}$$

and cluster these densities. The values of $f^{(i)}$ evaluated at all the α_j's can be approximated by the discrete scaled measurements

$$\hat{f}^{(i)}(\alpha_j) = \frac{S^{(i)}(\alpha_j)}{\sum_{j=1}^{180} S^{(i)}(\alpha_j)}, \text{ for } j = 1, ..., 180 \tag{3.9}$$

Our proposed clustering procedure requires samples from circular densities rather than the density values around the circumference. Although random samples can be generated based on the circular densities (3.9), we decide to artificially construct samples for better accuracy. For each set of scaled measurements $\{\hat{f}^{(i)}(\alpha_j)\}_{j=1}^{180}$ we construct a sample of size n by repeating the value α_j for $[n\hat{f}^{(i)}(\alpha_j)]$ times in the sample, where $[.]$ denotes the rounding function. This method ensures the scaled measurement can be sufficiently approximated by the empirical distribution of the artificially constructed sample when the sample size is large. Figure 3.5 illustrates that the histogram of the constructed sample for a randomly chosen NRR curve matches nicely with the corresponding scaled measurement curve when $n = 1000$.
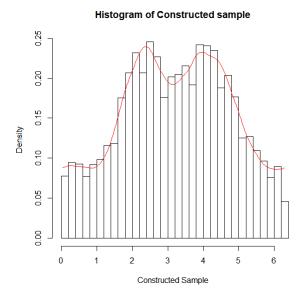
Figure 3.5: Histogram of the constructed sample for a randomly chosen NRR curve. The red line represents the scaled measurements. The sample size is $n = 1000$

By constructing the samples for all the scaled NRR curves and applying the L^2-based clustering procedure, the hierarchy of clusters is obtained and illustrated by the dendrogram in Figure 3.6. To choose the proper number of clusters, we consider two criterion—Average Silhouette Width (ASW) and Dunn Index (DI)— for evaluating the goodness of clustering. ASW is a measurement for the similarity of each subject to its own cluster compared to other clusters (see Rousseeu (1987) [31]), and DI is a measurement that compares the within-cluster distances with the inter-cluster distances (see Dunn (1974) [10]). A larger value for either criterion is an indication of better clustering. Figure 3.7 shows that among the number of clusters $k = 2, 3, ..., 10$, both criterion is maximized when $k = 3$. Thus, the optimal clustering comes from cutting the dendrogram at 3 clusters, which are enclosed by red boxes in Figure 3.6.

Therefore, our conclusion is that under our proposed hierarchical clustering framework, the NRR curves can most reasonably be clustered into three groups, based on the similarity of the circular density the curves represent under proper scaling. It provides a

foundation for Ophthalmologists to further research if the cluster assignment are related to the other characteristics of the patients.

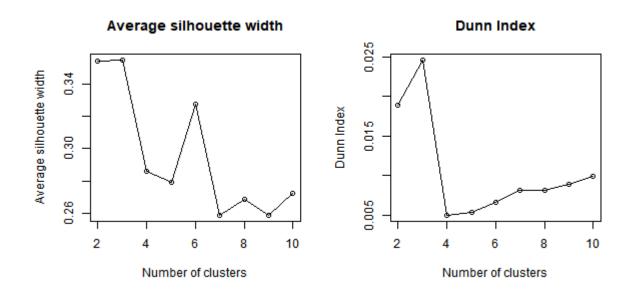Figure 3.6: L^2 clustering dendrogram for real data

Figure 3.7: Averaged Silhouette Width (left) and Dunn Index (right) vs. number of clusters.

Chapter 4

Some Properties of Wrapped Distributions

4.1 Introduction

A circular distribution can be obtained by wrapping a linear distribution around a unit circle, besides several other methods (See [20], Section 2.1.1). Such wrapped distributions enjoy some interesting properties. For instance Jammalamadaka and Kozubowski (2017) [18] showed that the wrapped distribution obtained from a mixture of linear components, is the same as the corresponding mixture of individually wrapped linear components i.e. that mixing and wrapping commute. In this chapter, first in Section 4.2, we will show that a similar property holds in terms of convolutions viz. wrapping a convolution of an arbitrary number of linear components corresponds to the convolution of the corresponding wrapped distributions. Then using the property that mixtures and wrapping commute, as shown in [18], we produce a Wrapped Quasi Lindley (WQL) distribution which happens to be a mixture of two linear components, instead of working with the original density of Quasi Lindley model. We will also show that the convolution

and mixing also commutes, and will derive the convolution of WQL distributions as an example.

4.2   The Commutative Property of Convolution and Wrapping

We use X and X^w to denote the random variables that follow an arbitrary linear distribution and its corresponding wrapped distribution, which is obtained by wrapping X around a unit circle. By definition,

$$X^w = X \bmod 2\pi. \tag{4.1}$$

Suppose $f_X(x)$ is the PDF of X, the PDF of X^w is given by

$$f_X^w(\gamma) = \sum_{k=-\infty}^{+\infty} f_X(\gamma + 2\pi k), \gamma \in [0, 2\pi). \tag{4.2}$$

The following theorem shows the commutative property of the two operations, namely convoluting and wrapping.

Theorem 4.2.1 *Suppose X and Y are two arbitrary r.v.s defined on the real line, then wrapping their convolution is equivalent to the convolution of their corresponding wrapped distributions, i.e.*

$$(X + Y)^w = X^w + Y^w$$

Proof: The PDF of their convolution $Z = X + Y$ is given by

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z - x) dx \tag{4.3}$$

42

By applying (4.2) to the equation above, we can write the PDF of wrapped Z as

$$f_Z^w(\gamma) = \sum_{k=-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_X(x) f_Y(\gamma + 2k\pi - x) dx \tag{4.4}$$

$$= \int_{-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} f_Y(\gamma + 2k\pi - x) f_X(x) dx \tag{4.5}$$

$$= \int_{-\infty}^{+\infty} f_Y^w(\gamma - x) f_X(x) dx \tag{4.6}$$

$$= \sum_{j=-\infty}^{+\infty} \int_0^{2\pi} f_Y^w(\gamma - x - 2j\pi) f_X(x + 2j\pi) dx \tag{4.7}$$

$$= \sum_{j=-\infty}^{+\infty} \int_0^{2\pi} f_Y^w(\gamma - x) f_X(x + 2j\pi) dx \tag{4.8}$$

$$= \int_0^{2\pi} \sum_{j=-\infty}^{+\infty} f_X(x + 2j\pi) f_Y^w(\gamma - x) dx \tag{4.9}$$

$$= \int_0^{2\pi} f_X^w(x) f_Y^w(\gamma - x) dx. \tag{4.10}$$

■

Alternatively, this theorem can be proved using the ChF of the two random variables.

Proof: Still let $Z = X + Y$. Z^w can be written as

$$Z^w = Z \bmod 2\pi \tag{4.11}$$

$$= \left(\sum_{k=-\infty}^{+\infty} Z \cdot I\{2\pi k \le Z < 2\pi(k+1)\} \right) \bmod 2\pi \tag{4.12}$$

$$= \sum_{k=-\infty}^{+\infty} (Z - 2\pi k) \cdot I\{2\pi k \le Z < 2\pi(k+1)\} \tag{4.13}$$

According to Proposition 2.1 of Jammalamadaka and SenGupta (2001) [20], the ChF of wrapped distribution, when evaluated at $p \in \mathbb{Z}$, is identical to the ChF of unwrapped distribution, i.e. $\phi_{X^w}(p) = \phi_X(p), \phi_{Y^w}(p) = \phi_Y(p)$, and $\phi_{Z^w}(p) = \phi_Z(p)$. With these and

Eqn. (4.13), we are able to prove the ChF of $(X + Y)^w$ equals the ChF $X^w + Y^w$, and the details are as follows.

$$\phi_{(X+Y)^w}(p) = \phi_{Z^w}(p) \tag{4.14}$$

$$= E[\exp(ipZ^w)] \tag{4.15}$$

$$= E\left\{\exp\left[ip\left(\sum_{k=-\infty}^{+\infty}(Z - 2\pi k)\cdot I\{2\pi k \leq Z < 2\pi(k+1)\}\right)\right]\right\} \tag{4.16}$$

$$= E\left\{\prod_{k=-\infty}^{+\infty}\exp[ip(Z - 2\pi k)]\cdot I\{2\pi k \leq Z < 2\pi(k+1)\}\right\} \tag{4.17}$$

$$= E\left\{\prod_{k=-\infty}^{+\infty}\exp(ipZ)\cdot I\{2\pi k \leq Z < 2\pi(k+1)\}\right\} \tag{4.18}$$

$$= E\left\{\exp\left[ipZ\cdot\left(\sum_{k=-\infty}^{+\infty}I\{2\pi k \leq Z < 2\pi(k+1)\}\right)\right]\right\} \tag{4.19}$$

$$= E[\exp(ipZ)] \tag{4.20}$$

$$= \phi_Z(p) \tag{4.21}$$

$$= \phi_X(p)\cdot\phi_Y(p) \tag{4.22}$$

$$= \phi_{X^w}(p)\cdot\phi_{Y^w}(p) \tag{4.23}$$

$$= \phi_{X^w+Y^w}(p) \tag{4.24}$$

Since ChF uniquely determines the distribution if its PDF is continuous, it can be concluded that $(X + Y)^w = X^w + Y^w$.

■

4.2.1   The Commutative Property of Mixing and Convoluting

First, consider the case where the convolution occurs between two distributions, each of which is a 2-component mixture. Suppose the mixture distributions are represented by

$Z_1 = pX_1 + (1-p)X_2$ and $Z_2 = qY_1 + (1-q)Y_2$, and we would like to find the distribution of their convolution $S = Z_1 + Z_2$. By the definition of convolution,

$$f_S(s) = (f_{Z_1} * f_{Z_2})(s) = \int_\Omega f_{Z_1}(t)f_{Z_2}(s-t)dt. \tag{4.25}$$

Since Z_1, Z_2 are mixtures, their PDFs are given by

$$f_{Z_1}(z) = pf_{X_1}(z) + (1-p)f_{X_2}(z), \tag{4.26}$$

$$f_{Z_2}(z) = qf_{Y_1}(z) + (1-q)f_{Y_2}(z). \tag{4.27}$$

Then

$$f_S(s) = \int_\Omega [pf_{X_1}(t) + (1-p)f_{X_2}(t)][qf_{Y_1}(s-t) + (1-q)f_{Y_2}(s-t)]dt \tag{4.28}$$

$$= pq \int_\Omega f_{X_1}(t)f_{Y_1}(s-t)dt + p(1-q) \int_\Omega f_{X_1}(t)f_{Y_2}(s-t)dt$$

$$+ (1-p)q \int_\Omega f_{X_2}(t)f_{Y_1}(s-t)dt + (1-p)(1-q) \int_\Omega f_{X_2}(t)f_{Y_2}(s-t)dt \tag{4.29}$$

$$= pq(f_{X_1} * f_{Y_1})(s) + p(1-q)(f_{X_1} * f_{Y_2})(s)$$

$$+ (1-p)q(f_{X_2} * f_{Y_1})(s) + (1-p)(1-q)(f_{X_2} * f_{Y_2})(s) \tag{4.30}$$

Therefore, S is a 4-component mixture. By induction, this result can be extended to the convolution of an arbitrary number of distributions, each of which is a mixture of an arbitrary number of components.

4.3 Wrapped Quasi Lindley Distribution

The Quasi Lindley Distribution $QLD(\theta, \alpha)$ is a two-parameter probability distribution defined on $\mathbb{R}^+$ and its PDF is given by

$$f_{QLD}(x; \theta, \alpha) = \frac{\theta}{\alpha + 1}(\alpha + \theta x)e^{-\theta x}, x \in \mathbb{R}^+, \tag{4.31}$$

where $\theta > 0$ and $\alpha \geq 0$. See Shanker and Mishra (2013) [33]. The QLD has the following mixture expression.

Proposition 4.3.1 *A QLD can be expressed as a mixture of an exponential distribution and a Gamma distribution i.e.*

$$f_{QLD}(x; \theta, \alpha) = \frac{\alpha}{1 + \alpha}f_{Exp}(x; \theta) + \frac{1}{1 + \alpha}f_G(x; 2, \theta), \tag{4.32}$$

where $f_{Exp}(x; \theta)$ is the PDF of exponential distribution with rate parameter θ, and $f_G(x; 2, \theta)$ is the PDF of Gamma distribution with shape parameter equals 2 and rate parameter θ.

The Wrapped Quasi Lindley (WQL) distribution, denoted by $WQL(\theta, \alpha)$, can be obtained by wrapping $QLD(\theta, \alpha)$ around a unit circle. Since $QLD(\theta, \alpha)$ can be expressed as a mixture of exponential and Gamma distributions, by exchanging the order of wrapping and mixing, we can conclude that WQL is also a mixture of wrapped exponential distribution and wrapped Gamma distribution.

Corollary 4.3.1 *A WQL distribution can be expressed as a mixture of wrapped exponential distribution and wrapped Gamma distribution.*

$$f_{WQL}(\beta; \theta, \alpha) = \frac{\alpha}{1 + \alpha}f_{WE}(\beta; \theta) + \frac{1}{1 + \alpha}f_{WG}(\beta; 2, \theta), \tag{4.33}$$

46

*where $f_{WE}(\beta; \theta)$ is the PDF of wrapped exponential distribution with rate parameter θ,
and $f_{WG}(\beta; 2, \theta)$ is the PDF of wrapped Gamma distribution with shape parameter equals
2 and rate parameter θ.*

The mixture representation of the WQL distribution can be used to derive the following
result.

Corollary 4.3.2 *The PDF of $WQL(\theta, \alpha)$ is given by*

$$f_{WQL}(\beta; \theta, \alpha) = \frac{\theta e^{-\theta\beta}}{\alpha + 1} \left(\frac{\alpha + \theta\beta}{1 - e^{-2\pi\theta}} + \frac{2\pi\theta e^{-2\pi\theta}}{(1 - e^{-2\pi\theta})^2} \right). \tag{4.34}$$

Proof: From Jammalamadaka and Kozubowski (2004) [19], the PDF of $WE(\theta)$ is

$$f_{WE}(\beta) = \frac{\theta e^{-\theta\beta}}{1 - e^{-2\pi\theta}}, \tag{4.35}$$

The WG distribution can be obtained from a mixture of truncated Gamma distributions.
The number of mixture components is equal to the shape parameter when it is an integer
(see Coelho (2007) [8]). Therefore, the PDF of $WG(2, \theta)$ is given by

$$f_{WG}(\beta; 2, \theta) = p_0 f_{tG0}(\beta; 1, \theta) + p_1 f_{tG1}(\beta; 2, \theta), \tag{4.36}$$

where f_{tG0}, f_{tG1} are the PDFs of truncated Gamma distributions $tG(1, \theta, [0, 2\pi))$ and
$tG(2, \theta, [0, 2\pi))$ and p_0, p_1 are mixing probabilities. The values of p_0, p_1 are given by

$$\begin{aligned}
p_0 &= \gamma^*(1; 2\pi\theta) \cdot \frac{2\pi\theta}{1!} \cdot \Phi^*(e^{-2\pi\theta}, 1) \\
&= (1 - e^{-2\pi\theta}) \cdot 2\pi\theta \cdot \frac{e^{-2\pi\theta}}{(1 - e^{-2\pi\theta})^2} \\
&= \frac{2\pi\theta e^{-2\pi\theta}}{1 - e^{-2\pi\theta}},
\end{aligned} \tag{4.37}$$

$$p_1 = 1 - p_0 = \frac{1 - e^{-2\pi\theta} - 2\pi\theta e^{-2\pi\theta}}{1 - e^{-2\pi\theta}}. \tag{4.38}$$

The truncated Gamma distribution $tG(k, \theta, [0, 2\pi)$ is the Gamma distribution $Gamma(k, \theta)$ restricted on $[0, 2\pi)$. Thus f_{tG0} and f_{tG1} are given by

$$f_{tG0}(\beta; 1, \theta) = \frac{e^{-\theta\beta}}{\int_0^{2\pi} e^{-\theta\beta} d\beta} = \frac{\theta e^{-\theta\beta}}{1 - e^{-2\pi\theta}}, \tag{4.39}$$

$$f_{tG1}(\beta; 2, \theta) = \frac{\beta e^{-\theta\beta}}{\int_0^{2\pi} \beta e^{-\theta\beta} d\beta} = \frac{\theta^2 \beta e^{-\theta\beta}}{1 - e^{-2\pi\theta} - 2\pi e^{-2\pi\theta}}. \tag{4.40}$$

Therefore,

$$\begin{aligned}
f_{WG}(\beta; 2, \theta) &= \frac{2\pi\theta e^{-2\pi\theta}}{1 - e^{-2\pi\theta}} \cdot \frac{\theta e^{-\theta\beta}}{1 - e^{-2\pi\theta}} + \frac{1 - e^{-2\pi\theta} - 2\pi\theta e^{-2\pi\theta}}{1 - e^{-2\pi\theta}} \cdot \frac{\theta^2 \beta e^{-\theta\beta}}{1 - e^{-2\pi\theta} - 2\pi e^{-2\pi\theta}} \\
&= \left[\frac{2\pi\theta^2 e^{-2\pi\theta}}{(1 - e^{-2\pi\theta})^2} + \frac{\theta^2 \beta}{1 - e^{-2\pi\theta}} \right] e^{-\theta\beta}. \tag{4.41}
\end{aligned}$$

Then Eqn. (4.33) becomes

$$\begin{aligned}
f_{WQL}(\beta; \theta, \alpha) &= \frac{\alpha}{1 + \alpha} \cdot \frac{\theta e^{-\theta\beta}}{1 - e^{-2\pi\theta}} + \frac{1}{1 + \alpha} \cdot \left[\frac{2\pi\theta^2 e^{-2\pi\theta}}{(1 - e^{-2\pi\theta})^2} + \frac{\theta^2 \beta}{1 - e^{-2\pi\theta}} \right] e^{-\theta\beta} \\
&= \frac{\theta e^{-\theta\beta}}{\alpha + 1} \left(\frac{\alpha + \theta\beta}{1 - e^{-2\pi\theta}} + \frac{2\pi\theta e^{-2\pi\theta}}{(1 - e^{-2\pi\theta})^2} \right). \tag{4.42}
\end{aligned}$$

∎

4.4   The Convolution of Two WQL Distributions

The distribution of the convolution of two independent WQL distributions with the same shape parameters can be derived by using the commutative property of wrapping, mixing and convoluting operations. The details are shown below.

Suppose X_1 and X_2 are two independent r.v.s with $X_1 \sim QLD(\theta, \alpha_1)$ and $X_2 \sim$

$QLD(\theta, \alpha_2)$. By wrapping X_1 and X_2 around a unit circle, we can generate X_1^w and

X_2^w, two independent circular r.v.s with $X_1^w \sim WQL(\theta, \alpha_1)$ and $X_2^w \sim WQL(\theta, \alpha_2)$. Let

$S = X_1 + X_2$ denote the convolution of X_1 and X_2. By exchanging the order of wrapping

and convolution, there is

$$S^w = (X_1 + X_2)^w = X_1^w + X_2^w, \tag{4.43}$$

i.e. the convolution of the two WQL r.v.s X_1^w and X_2^w can be obtained by wrapping S.

The distribution of S can be identified by the following corollary.

Corollary 4.4.1 *The convolution of $QLD(\theta, \alpha_1)$ and $QLD(\theta, \alpha_2)$ is distributed as a*

mixture of Gamma distributions

$$\frac{\alpha_1\alpha_2}{(1+\alpha_1)(1+\alpha_2)}Gamma(2,\theta)+\frac{2\alpha_1+2\alpha_2}{(1+\alpha_1)(1+\alpha_2)}Gamma(3,\theta)+\frac{1}{(1+\alpha_1)(1+\alpha_2)}Gamma(4,\theta),$$
$$\tag{4.44}$$

where $Gamma(k,\theta)$ is the Gamma distribution with shape parameter k and rate param-

eter θ.

Proof: Based on (4.31) and the commutative property of mixing and convoluting,

$$f_{X_1+X_2}(x) = \left(\frac{\alpha_1}{1+\alpha_1}f_{Exp}(x;\theta) + \frac{1}{1+\alpha_1}f_G(x;2,\theta)\right) * \left(\frac{\alpha_2}{1+\alpha_2}f_{Exp}(x;\theta) + \frac{1}{1+\alpha_2}f_G(x;2,\theta)\right)$$
$$= \frac{\alpha_1\alpha_2}{(1+\alpha_1)(1+\alpha_2)}f_{Exp}(x;\theta) * f_{Exp}(x;\theta) + \frac{2\alpha_1+2\alpha_2}{(1+\alpha_1)(1+\alpha_2)}f_{Exp}(x;\theta) * f_G(x;2,\theta)$$
$$+ \frac{1}{(1+\alpha_1)(1+\alpha_2)}f_G(x;2,\theta) * f_G(x;2,\theta). \tag{4.45}$$

Since $Exp(\theta)$ is equivalent to $Gamma(1,\theta)$ and the convolution of Gamma distributions

follows the rule: $Gamma(k_1,\theta) + Gamma(k_2,\theta) \sim Gamma(k_1+k_2,\theta)$, the convolutions

49

in Eqn. (4.45) becomes

$$f_{Exp}(x; \theta) * f_{Exp}(x; \theta) = f_G(x; 2, \theta) \tag{4.46}$$

$$f_{Exp}(x; \theta) * f_G(x; 2, \theta) = f_G(x; 3, \theta) \tag{4.47}$$

$$f_G(x; 2, \theta) * f_G(x; 2, \theta) = f_G(x; 4, \theta) \tag{4.48}$$

Thus expression (4.44) is verified.   ■

By applying the wrapping operation to S, we can show that a similar result exists for WQL convolution.

Corollary 4.4.2 *The convolution of $WQL(\theta, \alpha_1)$ and $WQL(\theta, \alpha_2)$ is distributed as a mixture of wrapped Gamma distributions*

$$\frac{\alpha_1 \alpha_2}{(1 + \alpha_1)(1 + \alpha_2)} WG(2, \theta) + \frac{2\alpha_1 + 2\alpha_2}{(1 + \alpha_1)(1 + \alpha_2)} WG(3, \theta) + \frac{1}{(1 + \alpha_1)(1 + \alpha_2)} WG(4, \theta), \tag{4.49}$$

where $WG(k, \theta)$ is the wrapped Gamma distribution with shape parameter k and rate parameter θ.

Chapter 5

Testing Symmetry in Sine-Skewed von Mises Distributions

5.1 Sine-Skewed Circular Distribution

Umbach and Jammalamadaka (2009) [36] adapted the idea of Azzalini (2005) [3] for linear distributions and proposed a class of asymmetric variations for any given circular distribution. The proposed distribution class takes on the form described below.

Theorem 5.1.1 (Umbach and Jammalamadaka (2009) [36]) *Suppose f and g are circular densities defined on $[-\pi, \pi)$ that are symmetric about 0 and $G(\theta) = \int_{-\pi}^{\theta} g(\gamma)d\gamma$. If w is an odd function with $|w(\theta)| \leq \pi$ and periodic with $w(\theta) = w(\theta + 2\pi k)$ for all $k \in \mathbb{Z}$. Then*

$$f(\theta|\mu) = 2f(\theta - \mu)G(w(\theta - \mu)) \tag{5.1}$$

is a circular density.

A sine-skewed circular distribution (studied in Abe and Pewsey (2011) [1]) is a special case of above distribution class with $G(\theta) = (\pi + \theta)/(2\pi)$ and $w(\theta) = \lambda\pi\sin(\theta)$. For any

51

given circular distribution $f_0(\theta)$ that is symmetrical around 0, its sine-skewed circular distribution has the PDF

$$f(\theta|\mu, \lambda) = f_0(\theta - \mu)[1 + \lambda \sin(\theta - \mu)] \tag{5.2}$$

where $\lambda \in [-1, 1]$ is the skewness parameter.

The sine-skewed distribution is in general not symmetrical unless $\lambda = 0$, and is usually unimodal if the initial distribution f_0 is unimodal. Therefore it is an effective method to obtain a asymmetrical variation of a given circular distribution. In this chapter we focus on the sine-skewed vM (SvM) distribution, which assumes f_0 is the PDF of $vM(0, \kappa)$. Then the PDF of $SvM(\mu, \kappa, \lambda)$ is given by

$$\begin{aligned}
f(\theta|\mu, \kappa, \lambda) &= f_0(\theta - \mu|\kappa)[1 + \lambda \sin(\theta - \mu)] \\
&= \frac{e^{\kappa \cos(\theta - \mu)}}{2\pi I_0(\kappa)}[1 + \lambda \sin(\theta - \mu)]
\end{aligned} \tag{5.3}$$

We use the following notations in this chapter: α_p, β_p are the p-th cosine and sine moment and a_p, b_p are the p-th sample cosine and sine moments, $\bar{\alpha}_p, \bar{\beta}_p, \bar{a}_p, \bar{b}_p$ are the respective centered moments that are centered around mean μ or sample mean $\bar{\theta}$. ρ is the mean resultant length and $\bar{R}$ is the sample mean resultant length. $\arctan^*(\cdot)$ is the quadrant-specific inverse tangent that maps the tangent of an angle in $[0, 2\pi)$ into the correct quadrant. The formulae of these quantities are given by

$$\alpha_p = \mathbb{E}(\cos p\theta), \beta_p = \mathbb{E}(\cos p\theta),$$

$$\bar{\alpha}_p = \mathbb{E}\{\cos p(\theta - \mu)\}, \bar{\beta}_p = \mathbb{E}\{\sin p(\theta - \mu)\},$$

$$\rho = \sqrt{\alpha_1^2 + \beta_1^2} = \sqrt{\bar{\alpha}_1^2 + \bar{\beta}_1^2}$$

$$a_p = \frac{1}{n} \sum_{i=1}^{n} \cos p\theta_i, b_p = \frac{1}{n} \sum_{i=1}^{n} \sin p\theta_i,$$

$$\bar{a}_p = \frac{1}{n} \sum_{i=1}^{n} \cos p(\theta_i - \bar{\theta}), \bar{b}_p = \frac{1}{n} \sum_{i=1}^{n} \sin p(\theta_i - \bar{\theta}),$$

$$\bar{\theta} = \arctan^*(b_1/a_1), \bar{R} = \sqrt{a_1^2 + b_1^2}.$$

Based on the vM trigonometric moment formulae given in Jammalamadaka and Sen-Gupta [20] Section 2.2.4, it is easy to derive that for $vM(\mu, \kappa)$, the trigonometric moments centered around μ are

$$\bar{\alpha}_p^{vM} = \frac{I_p(\kappa)}{I_0(\kappa)}, \quad \bar{\beta}_p^{vM} = 0 \tag{5.4}$$

for any $p \in \mathbb{Z}$. With this, the trigonometric moments of SvM can be deducted as follows.

Proposition 5.1.1 *For $SvM(\mu, \kappa, \lambda)$, the trigonometric moments centered around μ are given by*

$$\bar{\alpha}_p = \frac{I_p(\kappa)}{I_0(\kappa)} \tag{5.5}$$

$$\bar{\beta}_p = \frac{\lambda}{2} \frac{I_{p-1}(\kappa) - I_{p+1}(\kappa)}{I_0(\kappa)} \tag{5.6}$$

and the trigonometric moments are given by

$$\alpha_p = \frac{I_p(\kappa)}{I_0(\kappa)} \cos p\mu - \frac{\lambda}{2} \frac{I_{p-1}(\kappa) - I_{p+1}(\kappa)}{I_0(\kappa)} \sin p\mu \tag{5.7}$$

$$\beta_p = \frac{\lambda}{2} \frac{I_{p-1}(\kappa) - I_{p+1}(\kappa)}{I_0(\kappa)} \cos p\mu + \frac{I_p(\kappa)}{I_0(\kappa)} \sin p\mu \tag{5.8}$$

Proof: For trigonometric moments centered around μ.

$$\bar{\alpha}_p = \int_0^{2\pi} \cos p(\theta - \mu) f_{SvM}(\theta) d\theta$$

$$= \int_0^{2\pi} \cos p(\theta - \mu)[1 + \lambda \sin(\theta - \mu)] f_{vM}(\theta) d\theta$$

$$= \int_0^{2\pi} \cos p(\theta - \mu) f_{vM}(\theta) d\theta + \lambda \int_0^{2\pi} \cos p(\theta - \mu) \sin(\theta - \mu) f_{vM}(\theta) d\theta$$

$$= \int_0^{2\pi} \cos p(\theta - \mu) f_{vM}(\theta) d\theta + \lambda \int_0^{2\pi} \frac{\sin[(p+1)(\theta - \mu)] - \sin[(p-1)(\theta - \mu)]}{2} f_{vM}(\theta) d\theta$$

$$= \bar{\alpha}_p^{vM} + \frac{\lambda}{2}(\bar{\beta}_{p+1}^{vM} - \bar{\beta}_{p-1}^{vM})$$

$$= \frac{I_p(\kappa)}{I_0(\kappa)}$$

$$\bar{\beta}_p = \int_0^{2\pi} \sin p(\theta - \mu) f_{SvM}(\theta) d\theta$$

$$= \int_0^{2\pi} \sin p(\theta - \mu)[1 + \lambda \sin(\theta - \mu)] f_{vM}(\theta) d\theta$$

$$= \int_0^{2\pi} \sin p(\theta - \mu) f_{vM}(\theta) d\theta + \lambda \int_0^{2\pi} \sin p(\theta - \mu) \sin(\theta - \mu) f_{vM}(\theta) d\theta$$

$$= \int_0^{2\pi} \sin p(\theta - \mu) f_{vM}(\theta) d\theta + \lambda \int_0^{2\pi} \frac{\cos[(p-1)(\theta - \mu)] - \cos[(p+1)(\theta - \mu)]}{2} f_{vM}(\theta) d\theta$$

$$= \bar{\beta}_p^{vM} + \frac{\lambda}{2}(\bar{\alpha}_{p-1}^{vM} - \bar{\alpha}_{p+1}^{vM})$$

$$= \frac{\lambda}{2} \frac{I_{p-1}(\kappa) - I_{p+1}(\kappa)}{I_0(\kappa)}$$

Then, for (non-centered) trigonometric moments,

$$\alpha_p = \int_0^{2\pi} \cos p(\theta - \mu + \mu) f_{SvM}(\theta) d\theta$$

$$= \int_0^{2\pi} [\cos p(\theta - \mu) \cos p\mu - \sin p(\theta - \mu) \sin p\mu] f_{SvM}(\theta) d\theta$$

$$= \bar{\alpha}_p \cos p\mu - \bar{\beta}_p \sin p\mu$$

$$\beta_p = \int_0^{2\pi} \sin p(\theta - \mu + \mu) f_{SvM}(\theta) d\theta$$

54

$$= \int_0^{2\pi} [\sin p(\theta - \mu) \cos p\mu + \cos p(\theta - \mu) \sin p\mu] f_{SvM}(\theta) d\theta$$

$$= \bar{\beta}_p \cos p\mu + \bar{\alpha}_p \sin p\mu$$

Eqn. (5.7) and (5.8) are obtained by substituting $\bar{\alpha}_p$ and $\bar{\beta}_p$ in above equations with Eqn. (5.5) and (5.6) respectively. ∎

5.2 Tests for Circular Symmetry for Sine-skewed vM Distribution

Let $x_1, ..., x_n$ be a i.i.d. random sample of size n from $SvM(\mu, \kappa, \lambda)$. A test on whether $SvM(\mu, \kappa, \lambda)$ is symmetrical can be formulated as testing $\mathcal{H}_0 : \lambda = 0$ against $\mathcal{H}_1 : \lambda \neq 0$. Since $SvM(\mu, \kappa, 0)$ is equivalent to $vM(\mu, \kappa)$, the hypothesis can also be written as $\mathcal{H}_0 : f \sim vM(\mu, \kappa)$ for arbitrary $\mu \in [0, 2\pi)$ and $\kappa > 0$, and $\mathcal{H}_1 : f \sim SvM(\mu, \kappa, \lambda)$ for arbitrary $\mu \in [0, 2\pi)$, $\kappa > 0$ and $\lambda \in [-1, 0) \cup (0, 1]$.

5.2.1 Likelihood Ratio Test

A likelihood ratio test has the test statistic

$$\Lambda_n = -2 \left(\sup_{(\mu, \kappa, \lambda) \in \Omega_0} l(\mu, \kappa, \lambda | \theta_1, ..., \theta_n) - \sup_{(\mu, \kappa, \lambda) \in \Omega} l(\mu, \kappa, \lambda | \theta_1, ..., \theta_n) \right) \quad (5.9)$$

where

$$l(\mu, \kappa, \lambda | \theta_1, ..., \theta_n) = \sum_{i=1}^n \log f(\theta_i | \mu, \kappa, \lambda)$$

is the log-likelihood function of the observations. $\Omega_0 = [0, 2\pi) \times (0, +\infty) \times \{0\}$ and $\Omega = [0, 2\pi) \times (0, +\infty) \times [-1, 1]$ represent the parameter space for (μ, κ, λ) restricted under $\mathcal{H}_0$ and without restriction, respectively.

The test statistic follows χ_1^2 asymptotically with n approaches $+\infty$. Therefore, the test rejects $\mathcal{H}_0$ at level α when $\Lambda_n > \chi_1^2(\frac{\alpha}{2})$, the upper $\frac{\alpha}{2}$ quantile of χ_1^2.

5.2.2 Batschelet's Test

$\bar{b}_2$ is a measure of circular skewness proposed by Batschelet (1965) [5]. When the underlying circular distribution has non-zero mean resultant length ρ and its μ exists and unique, Pewsey (2002) [28] gives the asymptotic mean and variance of $\bar{b}_2$:

$$\mathbb{E}(\bar{b}_2) = \bar{\beta}_2 + \frac{1}{n}\left(-\frac{\bar{\beta}_3}{\rho} - \frac{\bar{\beta}_2}{\rho^2} + \frac{2\bar{\alpha}_2\bar{\beta}_2}{\rho^4}\right) + O(n^{-3/2}) \tag{5.10}$$

$$var(\bar{\beta}_2) = \frac{1}{n}\left[\frac{1-\bar{\alpha}_4}{2} - 2\bar{\alpha}_2 - \bar{\beta}_2^2 + \frac{2\bar{\alpha}_2}{\rho}\left\{\bar{\alpha}_3 + \frac{\bar{\alpha}_2(1-\bar{\alpha}_2)}{\rho}\right\}\right] + O(n^{-3/2}) \tag{5.11}$$

Thus the asymptotic distribution of $\bar{b}_2$ as $n \to \infty$ is

$$\frac{\bar{b}_2 - \mathbb{E}(\bar{b}_2)}{\sqrt{var(\bar{b}_2)}} \xrightarrow{D} N(0,1) \tag{5.12}$$

where the higher order terms of $\mathbb{E}(\bar{b}_2)$ and $var(\bar{\beta}_2)$ are omitted.

The distribution under $\mathcal{H}_0$ is $vM(\mu, \kappa)$ and the trigonometric moments formula (5.4) apply. Then (5.10), (5.11) becomes

$$\mathbb{E}(\bar{b}_2^{vM}) = 0 \tag{5.13}$$

$$\begin{aligned}var(\bar{\beta}_2^{vM}) &= \frac{1}{n}\left[\frac{1-\bar{\alpha}_4^{vM}}{2} - 2\bar{\alpha}_2^{vM} + \frac{2\bar{\alpha}_2^{vM}}{\bar{\alpha}_1^{vM}}\left\{\bar{\alpha}_3^{vM} + \frac{\bar{\alpha}_2^{vM}(1-\bar{\alpha}_2^{vM})}{\bar{\alpha}_1^{vM}}\right\}\right] \\ &= \frac{1}{n}\left[\frac{I_0(\kappa)-I_4(\kappa)}{2I_0(\kappa)} - 2\frac{I_2(\kappa)}{I_0(\kappa)} + \frac{2I_2(\kappa)}{I_1(\kappa)}\left\{\frac{I_3(\kappa)}{I_0(\kappa)} + \frac{I_2(\kappa)(1-I_2(\kappa))}{I_1(\kappa)I_0(\kappa)}\right\}\right]\end{aligned} \tag{5.14}$$
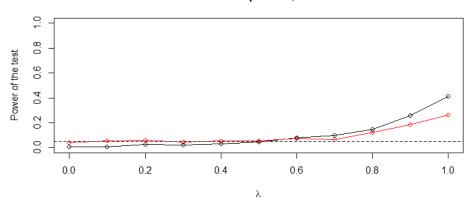
The test based on $\bar{b}_2$ has the test statistic

$$B_n = \frac{\bar{b}_2}{\sqrt{var(\bar{b}_2^{vM})}} \tag{5.15}$$

At level α, the test should reject the null hypothesis when $B_n > z_{\alpha/2}$
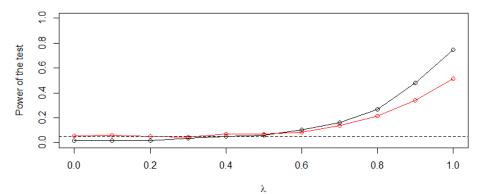
5.2.3   Power Comparison with Simulated Data

To compare the powers of the likelihood ratio test and Batschelet's test, a simulation study is conducted on samples from distribution $SvM(\mu = \pi, \kappa = 1, \lambda)$, where λ take on the values of $0, 0.1, 0.2, ..., 1$. For each distribution, 1000 samples of size $n = 50, 100$ or 200 are drawn and the two tests are applied to each sample. The empirical power of a test is computed by the proportion of times the test is rejected at significance level 0.05. The power curves are displayed in Figure 5.1. In each graph, the red curve corresponds to Batschelet's test and the black curve corresponds to likelihood ratio test. The horizontal dashed line indicates the power of the tests, which is $\alpha = 0.05$. It can be observed from the graphs that, when $\lambda = 0$, i.e. under the null hypothesis of symmetry, the level of likelihood ratio test is less than α while the level of Batschelet's test is almost identical to α. The power of the likelihood ratio test start to exceed that of Batschelet's test around $\lambda = 0.5$ and the difference between the two continue to grow as λ gets larger. For $n = 200$, the power of likelihood ratio test almost reaches 1 when λ reaches the maximum value of 1. For $n = 100$ and 50, the maximum power of the likelihood ratio test are close to 0.8 and 0.5, respectively. Therefore, the likelihood ratio test is in general more powerful than Batschelet's test when the SvM alternative is considered.
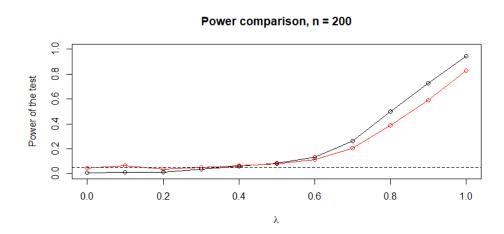
Figure 5.1: Power of likelihood ratio test and Batschelet's test with SvM as alternative distributions. Red curve represents likelihood ratio test and black curve represents Batschelet's test.

Chapter 6

A Bayesian Test for the Number of Modes in a Gaussian Mixture

6.1   Introduction

In this chapter we start with an investigation as to when a mixture of two Gaussian distributions is unimodal or bimodal. Certain subspace of the parameter space leads to unimodality of the mixture, and using Bayesian arguments, we find the posterior probability of the unimodality for a given data set, and the Bayes factor.

6.2   Mixture of Two Gaussian Distributions

A two-component mixture of Gaussian distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ has the form $pN(\mu_1, \sigma_1^2) + (1-p)N(\mu_2, \sigma_2^2)$, with the PDF

$$f(x|\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, p) = pf_1(x|\mu_1, \sigma_1^2) + (1-p)f_2(x|\mu_2, \sigma_2^2),$$

where f_1 and f_2 are the PDFs of $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$. The parameter vector $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, p)$ lies in the parameter space, $\Omega = \mathbb{R} \times \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R}^+ \times (0, 1)$. Depending on different parameter values, such a Gaussian mixture can be either unimodal or bimodal, so that the parameter space can be partitioned into two disjoint parts: an Ω_0 where the mixture distribution is unimodal and an Ω_1 where it is bimodal.

The necessary and sufficient conditions for determining whether a certain parameter vector gives unimodal or bimodal distribution has not yet been fully explored for the general 2-component Gaussian mixture. However, for the special case where $\sigma_1^2 = \sigma_2^2 = \sigma^2$, the conditions have been discussed in Behboodian (1970) [6], which are given in the theorem below.

Theorem 6.2.1 (Behboodian (1970) [6]) *The 2-component Gaussian mixture with equal variances $pN(\mu_1, \sigma^2) + (1 - p)N(\mu_2, \sigma^2)$ is unimodal if and only if either of the following conditions is satisfied:*

(a) $D^2 \leq 1$,

(b) $D^2 > 1$ and $\left| \log \frac{p}{1-p} \right| \geq 2 \log(D - \sqrt{D^2 - 1}) + 2D\sqrt{D^2 - 1}$,

where $D = \frac{|\mu_1 - \mu_2|}{2\sigma}$.

Let $\theta = (\mu_1, \mu_2, \sigma^2, p)$ denote the parameter vector and the subspace corresponding to unimodality by Ω_0, and its complement where the PDF becomes bimodal, by Ω_1.

A visual illustration of the boundary between Ω_0 and Ω_1 is given in Figure 6.2. For any given σ value, Ω_1 is the region on the right of the colored line (borderline) while Ω_0 is the region on the left.
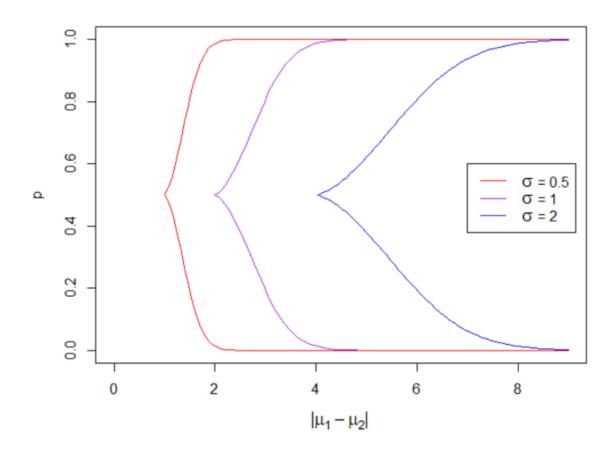
Figure 6.1: The borderline separating Ω_0 and Ω_1, the unimodal and bimodal parameter space for $pN(\mu_1, \sigma^2) + (1-p)N(\mu_2, \sigma^2)$.

6.3   A Bayesian Test Procedure

For a sample from a two-component Gaussian mixture with equal variances with unknown parameters, suppose we want to test for the unimodality of the underlying distribution. A parametric likelihood ratio test was proposed and studied in Holzmann and Vollmer (2008) [16]. Alternatively, in this section we introduce a Bayesian test procedure similar to what Jammalamadaka and Basu [4] proposed for two-component von Mises mixture.

Let $\mathbf{X} = (X_1, X_2, ..., X_n)$ be i.i.d. random variables from the Gaussian mixture with PDF

$$f(x|\theta) = pf_1(x|\mu_1, \sigma^2) + (1-p)f_2(x|\mu_2, \sigma^2), \tag{6.1}$$

where f_1, f_2 are the PDFs of $N(\mu_1, \sigma^2), N(\mu_1, \sigma^2)$, and $\mathbf{x} = (x_1, x_2, ..., x_n)$ is a random sample for $\mathbf{X}$. Suppose a Bayesian test is to be performed for testing $\mathcal{H}_0 : \theta \in \Omega_0$ vs. $\mathcal{H}_1 : \theta \in \Omega_1$. Let $\pi(\theta)$ denote the joint prior on θ. We compute the joint posterior of θ as $\pi(\theta|\mathbf{x})$ and proceed with the following steps.

1. Set up priors for each parameter μ_1, μ_2, σ, p, and compute the joint prior $\pi(\theta)$ as well as joint posterior $\pi(\theta|\mathbf{x})$

2. Calculate prior probability of unimodality

$$\mathbb{P}(\mathcal{H}_0) = \int_{\Omega_0} \pi(\theta)d\theta$$

using Monte Carlo method, and then the prior probability of bimodality is given by $\mathbb{P}(\mathcal{H}_1) = 1 - \mathbb{P}(\mathcal{H}_0)$.

3. Calculate posterior probability of unimodality

$$\mathbb{P}(\mathcal{H}_0|\mathbf{x}) = \int_{\Omega_0} \pi(\theta|\mathbf{x})d\theta$$

using Gibbs sampling and Monte Carlo method, and posterior probability of bimodality is given by $\mathbb{P}(\mathcal{H}_1|\mathbf{x}) = 1 - \mathbb{P}(\mathcal{H}_0|\mathbf{x})$.

4. Calculate the Bayes factor

$$B_{10} = \frac{\mathbb{P}(\mathcal{H}_1|\mathbf{x})\mathbb{P}(\mathcal{H}_0)}{\mathbb{P}(\mathcal{H}_0|\mathbf{x})\mathbb{P}(\mathcal{H}_1)},$$

then compare it with the suggested scale for interpretation as in Kass and Raftery [21].

6.4 Monte Carlo Method for Sampling from Prior and Posterior Distributions

6.4.1 Prior and Posterior Distributions for 2-component Gaussian Mixture

The first step in performing a Bayesian test is to set up priors for all parameters in the 2-component Gaussian mixture and compute the corresponding conditional posterior. Ideally, one would like to use conjugate priors so that the conditional posteriors are in the same family as the priors and have nice analytical forms. This task is made possible by considering the following framework called indicator Gaussian mixture model.

Suppose $\mathbf{x} = (x_1, ..., x_n)$ is an i.i.d random sample of size n from the Gaussian mixture (6.1). Each observation can be interpreted as being drawn from f_1 with probability p and being drawn from f_2 with probability $1 - p$. Define the indicators

$$z_i = I\{x_i \text{ is drawn from } f_1\},$$

for $i = 1, ..., n$, then there are

$$1 - z_i = I\{x_i \text{ is drawn from } f_2\},$$

In general, $\mathbf{z} = (z_1, ..., z_n)$ is not observable. However, conditionally given $\mathbf{z}$, the conditional distribution of $\mathbf{x}$ on $\mathbf{z}$ is

$$f(\mathbf{x}|\mathbf{z}, \theta) = \prod_{i=1}^{n} [f_1(x_i)]^{z_i} \cdot [f_2(x_i)]^{1-z_i}.$$

Using the fact that the mixture parameter is p, $z_i|\theta \sim Bernoulli(p)$ for $i = 1, ..., n$, and the joint distribution of $\mathbf{x}, \mathbf{z}$ is

$$f(\mathbf{x}, \mathbf{z}|\theta) = \prod_{i=1}^{n} [pf_1(x_i)]^{z_i} \cdot [(1-p)f_2(x_i)]^{1-z_i}. \tag{6.2}$$

This setup expresses the joint distribution as a product, and we chose appropriate conjugate priors as follows:

$$\sigma^2 \sim InverseGamma(\frac{\nu}{2}, \frac{s^2}{2}), \tag{6.3}$$

$$\mu_1|\sigma^2 \sim N(\xi_1, \frac{\sigma^2}{m_1}), \tag{6.4}$$

$$\mu_2|\sigma^2 \sim N(\xi_2, \frac{\sigma^2}{m_2}), \tag{6.5}$$

$$p \sim Uniform(0, 1), \tag{6.6}$$

where p is independent of μ_1, μ_2, or σ^2; and $m_1, m_2, \xi_1, \xi_2, \nu, s^2$ are predetermined hyperparameters.

Theorem 6.4.1 *Given the sample joint distribution 6.2 and the priors defined in (6.3)–*

(6.6), the conditional posterior distributions are given by

$$\sigma^2|\mathbf{x}, \mathbf{z} \sim InverseGamma\left(\frac{n+\nu}{2}, \frac{s^2 + \sum_{i=1}^{n} x_i^2 + m_1\xi_1^2 + m_2\xi_2^2 - C_1 - C_2}{2}\right), \quad (6.7)$$

$$\mu_1|\sigma^2, \mathbf{x}, \mathbf{z} \sim N\left(\frac{\sum_{i=1}^{n} z_i x_i + m_1\xi_1}{\sum_{i=1}^{n} z_i + m_1}, \frac{\sigma^2}{\sum_{i=1}^{n} z_i + m_1}\right), \quad (6.8)$$

$$\mu_2|\sigma^2, \mathbf{x}, \mathbf{z} \sim N\left(\frac{\sum_{i=1}^{n}(1-z_i)x_i + m_2\xi_2}{n - \sum_{i=1}^{n} z_i + m_2}, \frac{\sigma^2}{n - \sum_{i=1}^{n} z_i + m_2}\right), \quad (6.9)$$

$$p|\mathbf{x}, \mathbf{z} \sim Beta\left(\sum_{i=1}^{n} z_i + 1, n - \sum_{i=1}^{n} z_i + 1\right). \quad (6.10)$$

where

$$C_1 = \frac{(\sum_{i=1}^{n} z_i x_i + m_1\xi_1)^2}{m_1 + \sum_{i=1}^{n} z_i},$$

$$C_2 = \frac{(\sum_{i=1}^{n}(1-z_i)x_i + m_2\xi_2)^2}{m_2 + n - \sum_{i=1}^{n} z_i}.$$

Proof: Using the notation $\pi(\cdot)$ to denote the joint prior and $\pi(\cdot|\mathbf{x}, \mathbf{z})$ to denote the joint posterior. The joint posterior is given by

$$\pi(\theta|\mathbf{x}, \mathbf{z}) = \frac{f(\mathbf{x}, \mathbf{z}|\theta)\pi(\theta)}{\int_\Omega f(\mathbf{x}, \mathbf{z}|\theta)\pi(\theta)d\theta} \quad (6.11)$$

where $\pi(\theta)$ is the joint prior. Further decomposition of the joint prior and joint posterior distributions are given by

$$\pi(\theta) = \pi(\mu_1|\sigma^2) \cdot \pi(\mu_2|\sigma^2) \cdot \pi(\sigma^2) \cdot \pi(p),$$

$$\pi(\theta|\mathbf{x}, \mathbf{z}) = \pi(\mu_1|\sigma^2, \mathbf{x}, \mathbf{z}) \cdot \pi(\mu_2|\sigma^2, \mathbf{x}, \mathbf{z}) \cdot \pi(\sigma^2|\mathbf{x}, \mathbf{z}) \cdot \pi(p|\mathbf{x}, \mathbf{z}).$$

Therefore, the conditional posterior of a single parameter can be obtained by integrating the joint posterior with respect to other parameters over their respective parameter space.

For example,

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{0}^{+\infty} \pi(\theta|\mathbf{x}, \mathbf{z})d\mu_1 d\mu_2 d(\sigma^2) = \int_{-\infty}^{+\infty} \pi(\mu_1|\sigma^2, \mathbf{x}, \mathbf{z})d\mu_1 \cdot \int_{-\infty}^{+\infty} \pi(\mu_2|\sigma^2, \mathbf{x}, \mathbf{z})d\mu_2$$

$$\cdot \int_{0}^{+\infty} \pi(\sigma^2|\mathbf{x}, \mathbf{z})d(\sigma^2) \cdot \pi(p|\mathbf{x}, \mathbf{z})$$

$$= \pi(p|\mathbf{x}, \mathbf{z})$$

Since the denominator of (6.11) is only a normalizing constant, we have

$$\pi(\mu_1|\sigma^2, \mathbf{x}, \mathbf{z}) \cdot \pi(\mu_2|\sigma^2, \mathbf{x}, \mathbf{z}) \cdot \pi(\sigma^2|\mathbf{x}, \mathbf{z}) \cdot \pi(p|\mathbf{x}, \mathbf{z}) \propto f(\mathbf{x}, \mathbf{z}|\theta)\pi(\theta),$$

meaning that the conditional posterior distributions can be found by decomposing $f(\mathbf{x}, \mathbf{z}|\theta)\pi(\theta)$ into a product of kernels for $\mu_1, \mu_2, \sigma^2, p$, each kernel corresponds to a distribution family. The details of this operation are shown below.

$$f(\mathbf{x}, \mathbf{z}|\theta)\pi(\theta) = f(\mathbf{x}, \mathbf{z}|\theta) \cdot \pi(\mu_1|\sigma^2) \cdot \pi(\mu_2|\sigma^2) \cdot \pi(\sigma^2) \cdot \pi(p)$$

$$= p^{\sum_{i=1}^{n} z_i} \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{\sum_{i=1}^{n} z_i} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} z_i(x_i - \mu_1)^2\right\}$$

$$\cdot (1-p)^{\sum_{i=1}^{n}(1-z_i)} \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{\sum_{i=1}^{n}(1-z_i)} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(1-z_i)(x_i - \mu_2)^2\right\}$$

$$\cdot \frac{\sqrt{m_1}}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{m_1}{2\sigma^2}(\mu_1 - \xi_1)^2\right\} \cdot \frac{\sqrt{m_2}}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{m_2}{2\sigma^2}(\mu_2 - \xi_2)^2\right\}$$

$$\cdot \frac{(s^2/2)^{\nu/2}}{\Gamma(\nu/2)}(\sigma^2)^{-\nu/2-1} \exp\left\{-\frac{s^2}{2\sigma^2}\right\} \cdot 1$$

$$\propto T_1(\mu_1, \sigma^2) \cdot T_2(\mu_2, \sigma^2) \cdot T_3(\sigma^2) \cdot T_4(p)$$

Where T_1 through T_4 are the kernels for each parameter and have the following forms:

$$T_1(\mu_1, \sigma^2) = \frac{1}{\sigma} \exp\left\{ -\frac{1}{2\sigma^2} \left[\sum_{i=1}^{n} z_i(x_i - \mu_1)^2 + m_1(\mu_1 - \xi_1)^2 \right] \right\}$$

$$\propto \frac{1}{\sigma} \exp\left\{ -\frac{1}{2\sigma^2} \left[\left(\sum_{i=1}^{n} z_i + m_1 \right) \mu_1^2 - 2 \left(\sum_{i=1}^{n} z_i x_i + m_1 \xi_1 \right) \mu_1 + C_1 \right] \right\}$$

$$\propto N\left(\frac{\sum_{i=1}^{n} z_i x_i + m_1 \xi_1}{\sum_{i=1}^{n} z_i + m_1}, \frac{\sigma^2}{\sum_{i=1}^{n} z_i + m_1} \right)$$

$$T_2(\mu_2, \sigma^2) = \frac{1}{\sigma} \exp\left\{ -\frac{1}{2\sigma^2} \left[\sum_{i=1}^{n} (1 - z_i)(x_i - \mu_2)^2 + m_2(\mu_2 - \xi_2)^2 \right] \right\}$$

$$\propto \frac{1}{\sigma} \exp\left\{ -\frac{1}{2\sigma^2} \left[\left(n - \sum_{i=1}^{n} z_i + m_2 \right) \mu_2^2 - 2 \left(\sum_{i=1}^{n} (1 - z_i)x_i + m_2 \xi_2 \right) \mu_2 + C_2 \right] \right\}$$

$$\propto N\left(\frac{\sum_{i=1}^{n} (1 - z_i)x_i + m_2 \xi_2}{n - \sum_{i=1}^{n} z_i + m_2}, \frac{\sigma^2}{n - \sum_{i=1}^{n} z_i + m_2} \right)$$

$$T_3(\sigma^2) = (\sigma^2)^{-\frac{n+\nu}{2} - 1} \exp\left\{ -\frac{s^2 + \sum_{i=1}^{n} x_i^2 + m_1 \xi_1^2 + m_2 \xi_2^2 - C_1 - C_2}{2\sigma^2} \right\}$$

$$\propto InverseGamma\left(\frac{n + \nu}{2}, \frac{s^2 + \sum_{i=1}^{n} x_i^2 + m_1 \xi_1^2 + m_2 \xi_2^2 - C_1 - C_2}{2} \right)$$

$$T_4(p) = p^{\sum_{i=1}^{n} z_i} (1 - p)^{\sum_{i=1}^{n} (1 - z_i)}$$

$$\propto Beta\left(\sum_{i=1}^{n} z_i + 1, n - \sum_{i=1}^{n} z_i + 1 \right)$$

where

$$C_1 = \frac{\left(\sum_{i=1}^{n} z_i x_i + m_1 \xi_1 \right)^2}{m_1 + \sum_{i=1}^{n} z_i},$$

$$C_2 = \frac{\left(\sum_{i=1}^{n} (1 - z_i)x_i + m_2 \xi_2 \right)^2}{m_2 + n - \sum_{i=1}^{n} z_i}.$$

The distributions that correspond to the 4 kernels are the conditional posterior distributions for each parameter, which correspond to Eqn. (6.7)–(6.10). ∎

6.4.2   Computation of the Prior Probabilities of Unimodality and Bimodality

After setting up the priors and finding the conditional posteriors, the next step is to calculate the prior probability of unimodality and bimodality. To calculate the prior probability of unimodality

$$\mathbb{P}(\mathcal{H}_0) = \int_{\Omega_0} \pi(\theta) d\theta,$$

one need to integrate the joint priors over the unimodal parameter space. Since the criterion determining the boundary of Ω_0 is complicated, it is difficult to find an analytical solution to the integral. Instead, a Monte Carlo method is employed to get its value numerically.

Algorithm 1 (Monte Carlo method for calculating prior probabilities) *To numerically compute the prior probability of unimodality and bimodality—$\mathbb{P}(\mathcal{H}_0)$ and $\mathbb{P}(\mathcal{H}_1)$—these steps are followed.*

1. *Determine the values for prior hyperparameters $m_1, m_2, \xi_1, \xi_2, \nu, s^2$.*

2. *Generate N parameter vectors $\theta^{(1)}, ..., \theta^{(N)}$ from prior distributions (6.3)–(6.6). The order of parameter generation is $\sigma^2 \rightarrow \mu_1 \rightarrow \mu_2 \rightarrow p$.*

3. *For each $\theta^{(i)}, i = 1, ..., N$, check if it belongs in Ω_0 using the conditions in Theorem 6.2.1 and obtain the values of indicators $I\left\{\theta^{(i)} \in \Omega_0\right\}$.*

4. *Compute $\mathbb{P}(\mathcal{H}_0) = \frac{1}{N} \sum_{i=1}^{N} I\left\{\theta^{(i)} \in \Omega_0\right\}$ and $\mathbb{P}(\mathcal{H}_1) = 1 - \mathbb{P}(\mathcal{H}_0)$.*

6.4.3 Computation of the Posterior Probabilities of Unimodality and Bimodality

Similar to $\mathbb{P}(\mathcal{H}_0)$, the posterior probability of unimodality

$$\mathbb{P}(\mathcal{H}_0|\mathbf{x}) = \int_{\Omega_0} \pi(\theta|\mathbf{x})d\theta$$

needs to be computed numerically using Monte Carlo method. The conditional posteriors given in Theorem 6.4.1 are conditional on latent variables $\mathbf{z}$ that are not observable from the samples. To mitigate this problem, we would like to treat $\mathbf{z}$ as unknown parameters and use Gibbs sampling method to generate samples from θ and $\mathbf{z}$ simultaneously.

Algorithm 2 (Gibbs sampling) *To generate θ and $\mathbf{z}$ from conditional distributions, follow these steps:*

1. *Set up initial values $\theta^{(0)}$. Then, at the k-th iteration,*

2. *Generate $\mathbf{z}^{(k)}$ from $f(\mathbf{z}^{(k)}|\theta^{(i-1)}, \mathbf{x})$.*

3. *Generate $\theta^{(k)}$ from $\pi(\theta^{(k)}|\mathbf{x}, \mathbf{z}^{(k)})$.*

4. *Repeat Step 2 and 3 for a total of N times until convergence.*

To find the conditional distribution of step 2 in Algorithm 2, we have the following results.

Lemma 6.4.1 *The conditional distribution of each z_i in $\mathbf{z}$ is given by*

$$z_i|\theta, x_i \sim Bernoulli\left(\frac{pf_1(x_i|\mu_1, \sigma^2)}{pf_1(x_i|\mu_1, \sigma^2) + (1-p)f_2(x_i|\mu_2, \sigma^2)}\right).$$

Proof: From (6.1) and (6.2) and since x_i's are i.i.d.,

$$f(\mathbf{z}|\theta, \mathbf{x}) = \frac{f(\mathbf{x}, \mathbf{z}|\theta)}{f(\mathbf{x}|\theta)}$$
$$= \frac{\prod_{i=1}^{n} [pf_1(x_i|\mu_1, \sigma^2)]^{z_i} \cdot [(1-p)f_2(x_i|\mu_1, \sigma^2)]^{1-z_i}}{\prod_{i=1}^{n} [pf_1(x_i|\mu_1, \sigma^2) + (1-p)f_2(x_i|\mu_2, \sigma^2)]}$$
$$= \prod_{i=1}^{n} P_i^{z_i}(1-P_i)^{1-z_i}$$

where each

$$P_i = \frac{pf_1(x_i|\mu_1, \sigma^2)}{pf_1(x_i|\mu_1, \sigma^2) + (1-p)f_2(x_i|\mu_2, \sigma^2)}$$

Since each z_i is only dependent on x_i, and x_i's are independent, z_i's are independent of each other as well. It is easy to see

$$f(z_i|\theta, x_i) = P_i^{z_i}(1-P_i)^{1-z_i}$$

for $z_i \in \{0, 1\}, i = 1, ..., n$. This is the PMF of $Bernoulli(P_i)$. ∎

Given the Gibbs sampling procedure, the Monte Carlo calculation is outlined below.

Algorithm 3 (Monte Carlo method for posteriors) *To numerically compute the posterior probability of unimodality and bimodality—$\mathbb{P}(\mathcal{H}_0|\mathbf{x})$ and $\mathbb{P}(\mathcal{H}_1|\mathbf{x})$—these steps are followed.*

1. *Generate N parameter vectors $\theta^{(1)}, ..., \theta^{(N)}$ for conditional posterior distributions (6.7) - (6.10) by utilizing Gibbs sampling procedure in Algorithm 2. The order of parameter generation is $\sigma^2 \rightarrow \mu_1 \rightarrow \mu_2 \rightarrow p$.*

2. *Make sure the Markov chain generated in Step 1 is convergent after a burn-in period of length K.*

3. *For each $\theta^{(i)}, i = K + 1, ..., N$, check if it belongs in Ω_0 using the conditions in Theorem 6.2.1 and obtain the values of indicators $I\left\{\theta^{(i)} \in \Omega_0\right\}$*

4. *Compute $\mathbb{P}(\mathcal{H}_0|\mathbf{x}) = \frac{1}{N-K} \sum_{i=K+1}^{N} I\left\{\theta^{(i)} \in \Omega_0\right\}$ and $\mathbb{P}(\mathcal{H}_1|\mathbf{x}) = 1 - \mathbb{P}(\mathcal{H}_0|\mathbf{x})$.*

Remark: A mixture distribution has two equivalent parameter representations, namely $(\mu_1, \mu_2, \sigma^2, p)$ and $(\mu_2, \mu_1, \sigma^2, 1-p)$. To differentiate between these two parameterizations and uniquely define the mixture, we restrict $p \in (0, 0.5]$, so that the prior of p becomes

$$p \sim Uniform(0, 0.5] \tag{6.12}$$

and the conditional posterior of p becomes

$$p|\mathbf{x}, \mathbf{z} \sim Beta\left(\sum_{i=1}^{n} z_i + 1, n - \sum_{i=1}^{n} z_i + 1\right) \text{ restricted on } (0, 0.5]. \tag{6.13}$$

Note that when $p = 0.5$ the two parameterizations become the same.

6.4.4   Judging the Bayes Factor

Kass and Raftery (1995) [21] discuss the Bayes factor for testing $\mathcal{H}_1$ against $\mathcal{H}_0$ which is defined as

$$B_{10} = \frac{\text{posterior odds}}{\text{prior odds}} = \frac{\mathbb{P}(\mathcal{H}_1|\mathbf{x})\mathbb{P}(\mathcal{H}_0)}{\mathbb{P}(\mathcal{H}_0|\mathbf{x})\mathbb{P}(\mathcal{H}_1)}. \tag{6.14}$$

It is used as a summary of the evidence provided by data in favor of $\mathcal{H}_1$ against $\mathcal{H}_0$. In general, larger value of Bayes factor indicates stronger evidence in favor of $\mathcal{H}_1$. Kass and Raftery (1995) [21] suggest using Table 6.1 as a reasonable scale for interpreting B_{10} and $\log_{10}(B_{10})$ values. Although not shown in the table, it should be noted that $0 < B_{10} < 1$ (or equivalently $\log_{10} B_{10} < 0$) indicates no evidence against $\mathcal{H}_0$ at all.

$\log_{10} B_{10}$	B_{10}	Evidence against $\mathcal{H}_0$
0 to 1/2	1 to 3.2	Not worth more than a bare mention
1/2 to 1	3.2 to 10	Substantial
1 to 2	10 to 100	Strong
>2	>100	Decisive

Table 6.1: Suggested scale for interpreting B_{10} and $\log_{10}(B_{10})$ values

6.5 A Simulation Study

To perform simulation studies for the Bayesian test procedure proposed above, first we select 2 sets of Gaussian mixture distributions to generate data. Each set of distributions contains 7 distributions, among which μ_1, σ^2 and p are fixed and μ_2 are different, resulting in parameter combinations from both unimodal and bimodal parameter space. The actual parameter values and their corresponding modality are presented in Table 6.2. The mixing probabilities are even for distributions from the first set and uneven for distributions from the second set. Figure 6.2 shows the density plots for mixture distribution within each set. It is clear that as μ_2 moves away from μ_1, the density curve gradually changes from being unimodal to being bimodal.

	μ_1	σ^2	p	μ_2	modality
Distribution Set 1	1.0	4	0.5	4.0, 4.5, 5.0	unimodal
				5.5, 6.0, 6.5, 7.0	bimodal
Distribution Set 2	1.0	4	0.2	5.0, 5.5, 6.0, 6.5	unimodal
				7.0, 7.5, 8.0	bimodal

Table 6.2: Choice of parameters for simulated data.

For each distribution in the two distribution sets, we generate 50 samples of size $n = 25$ and 50 samples of size $n = 50$, then conduct Bayesian test for each sample assuming the same prior hyperparameter values ($\nu = 4, \xi_1 = 0, \xi_2 = 10, s^2 = 30, m_1 = 10$ and $m_2 = 10$). The Bayes factors are obtained by running Algorithms 1 and 3 with $N = 100000$ and burn-in period of 20000 and plugging their results into (6.14). Restricted
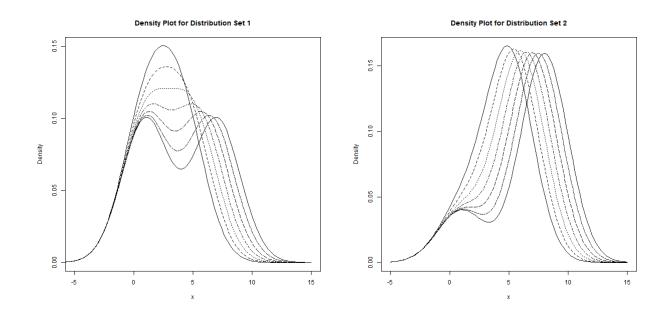
Figure 6.2: Density plots for Gaussian mixtures in Distribution Set 1 and 2, respectively.

prior (6.12) and conditional posterior (6.13) for p are used to avoid identifiability issues in θ simulation. The initial values of the parameters are $\mu_1 = 0, \mu_2 = 10, \sigma^2 = 100, p = 0.5$ in posterior simulation for all samples.

Figure 6.3 displays the Bayes factors in boxplots for each of the μ_2 values. The plots in the same row correspond to the same sample sizes and the plots in the same column correspond to the same distribution set. Table 6.3 and 6.4 show the proportion of times that the Bayes factor exceeds 3.2. From these figures and tables, we see that (i) within each distribution set, the Bayes factor is more likely to have larger values when μ_2 is larger; (ii) given a certain bimodal distribution, the Bayes factor is more likely to have larger values when n is larger. Both these observations (i) and (ii) show that the Bayes factor is more likely to have larger values when there is a stronger evidence that the sample came from a bimodal distribution.

To illustrate the properties of the Markov chain generated in simulation of posterior distributions, an example Markov chain that corresponds to a sample from Distribution
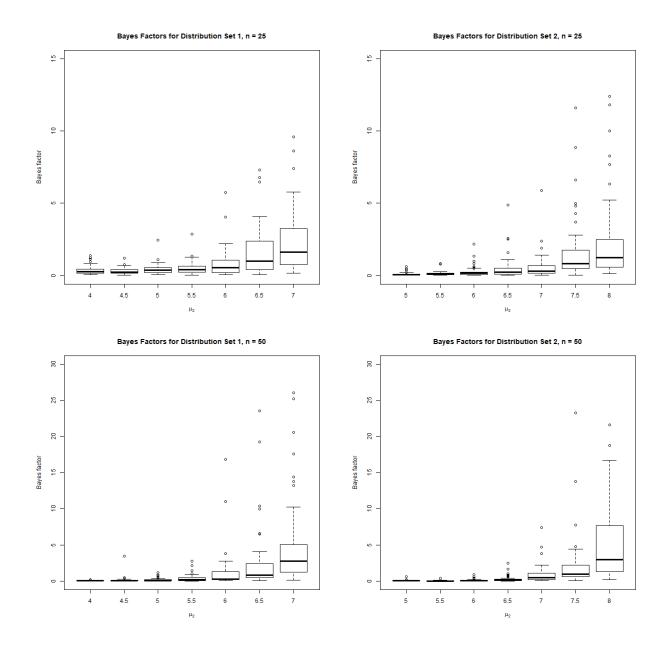
Figure 6.3: Boxplots for Bayes factors for different μ_2 values.

μ_2	4.0	4.5	5.0	5.5	6.0	6.5	7.0
Distribution Set 1, $n = 25$	0.00	0.00	0.00	0.00	0.06	0.12	0.18
Distribution Set 1, $n = 50$	0.00	0.02	0.00	0.00	0.06	0.20	0.46

Table 6.3: Proportion of times that Bayes factors > 3.2 for Distribution Set 1.

μ_2	5.0	5.5	6.0	6.5	7.0	7.5	8.0
Distribution Set 2, $n = 25$	0.00	0.00	0.00	0.02	0.02	0.14	0.20
Distribution Set 2, $n = 50$	0.00	0.00	0.00	0.00	0.06	0.16	0.44

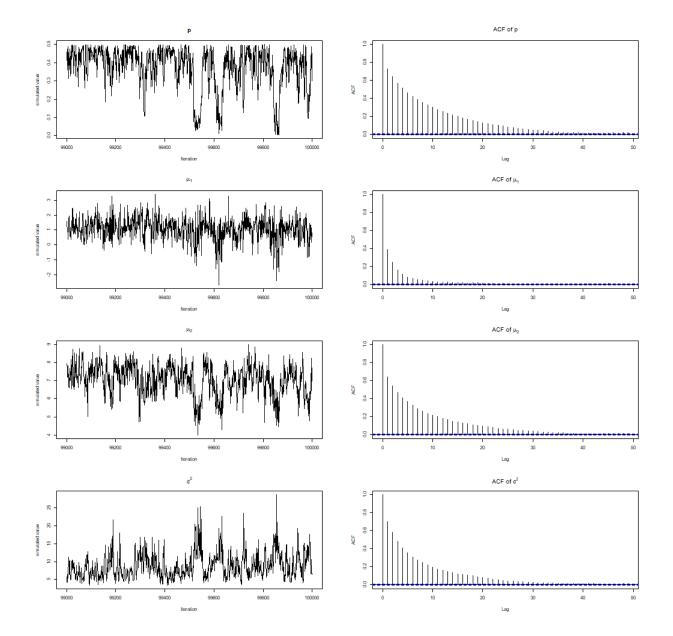Table 6.4: Proportion of times that Bayes factors > 3.2 for Distribution Set 2.

Figure 6.4: Path plot and ACF plot for each parameter in a Markov chain in posterior simulation. The sample is from Distribution Set 1, $\mu_2 = 8, n = 50$.

Set 1, $\mu_2 = 8, n = 50$ is selected, and its path plot and ACF plot for each parameter are shown in Figure 6.4. Note that only the last 10000 iterations are plotted in the path plot and only the iterations after the burn-in period (> 20000) are used to compute the ACFs. It can be seen that the paths of p and σ^2 are able to traverse a wide range within their parameter space and there is no sign of identifiability issue for μ_1 and μ_2. The number of lags required for ACFs to diminish is pretty large (≈ 30) for all parameters.

6.6   Real Data Application—Adult Height Data

In this section we present an application of the modality test on human heights data. There have been discussions on whether the combination of men and women heights will give a unimodal or bimodal distribution, for example see Schilling et al. (2002) [32]. To investigate this problem, we source the data from 2013–2014 National Health and Nutrition Examination Survey (NHANES). NHANES [27] is a program of studies that aim at assessing the health and nutritional status of adults and children in the United States, and it is conducted by National Center for Health Statistics (NCHS), a part of the Centers for Disease Control and Prevention (CDC). To select the data for testing, we took the variables of gender and age from the Demographic Variables and Sample Weights dataset with the measurement of standing height in centimeters from the Body Measures dataset, matched the records on the unique identifier (respondent sequence number), and filtered out the records with age less than 18 or have any missing values. An overview of the filtered dataset is given in Table 6.5 and the histogram is shown in Figure 6.5. Although male and female adult heights have significantly different standard deviations, we would still assume they are the same when applying the Bayesian test.

The data indicates that the sample distribution of adult heights is unimodal when modeled by a Gaussian mixture model $\hat{p}N(\hat{\mu}_1, \hat{\sigma}^2) + (1 - \hat{p})N(\hat{\mu}_2, \hat{\sigma}^2)$ where the mix-
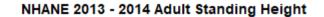
	# of records	% of records	Average height (cm)	SD of height (cm)
Overall	5857	100%	167.09	7.508 (pooled)
Male	2795	48.07%	173.13	7.834
Female	3062	51.9%	159.67	7.194

Table 6.5: Overview of NHANE 2013–2014 adult standing height data

	Prior hyperparameters						Prior bimodal prob.	Bayes factor	Conclusion
	ξ_1	ξ_2	σ^2	m_1	m_2	ν	$\mathbb{P}(\mathcal{H}_0)$	B_{10}	
Setting 1						8	0.1663	2.24×10^{-5}	Unimodal
Setting 2	150	180	400	10	10	5	0.4175	1.79×10^{-5}	Unimodal
Setting 3						3	0.6129	7.71×10^{-5}	Unimodal
Setting 4						1.5	0.8710	2.53×10^{-4}	Unimodal

Table 6.6: Prior parameters and Bayes factor results for NHANE 2013–2014 adult height data

ing components are estimated from male and female heights data. The histogram also supports this claim. However, are we going to reach the same conclusion without the information about respondents' gender? To check this, we performed the Bayesian test with $N = 100000$ Markov chain iterations and a burn-in period of 20000 under several different choices of prior distribution parameters. Table 6.6 shows that the Bayes factor is close to 0 in all cases, strongly suggesting the underlying distribution is unimodal. Figure 6.6 shows the path plot for and the ACF plot for the posterior Markov chain after the burn-in period. It can be seen that the parameters simulated from the posterior distribution converge and the number of lags for autocorrelation to diminish can be 10–30 for different parameters.
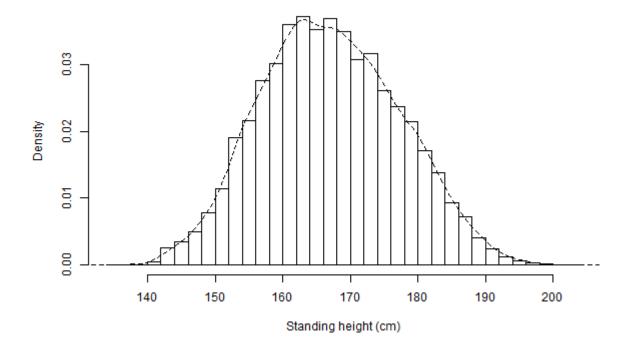
Figure 6.5: Histogram of NHANE 2013–2014 adult standing height data. The dashed line represents kernel density estimation of the histogram
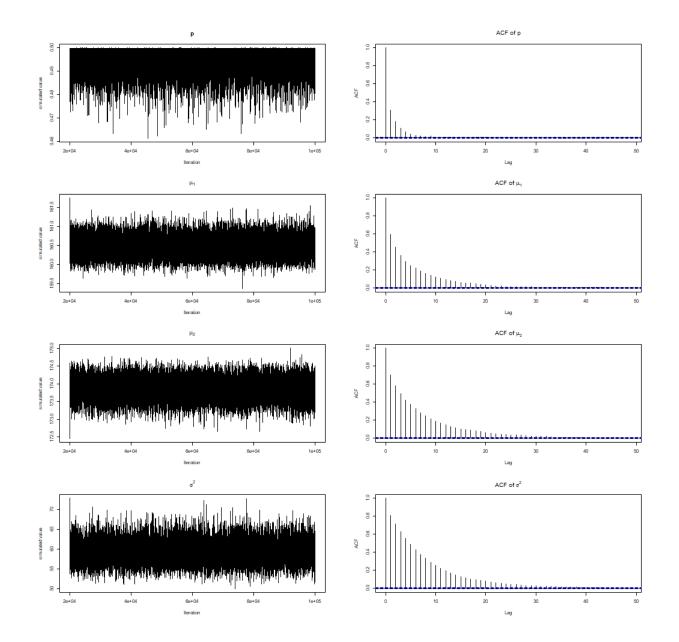
Figure 6.6: Path plot and ACF plot for testing NHANE 2013–2014 adult height data.

Appendix A

Some Computer Programs

A.1 Code for Testing for Symmetric WS Distribution (Chapter 2)

```
require(circular)


### 1. Functions for symmetric wrapped stable distribution
# 1.1 Generate symmetric wrapped stable r.v. with shift
# rwrpstab is a function in package "circular"
rwrpstab.sym = function(n, gamma, tau, mu){
  (rwrpstab(n, gamma, 0, tau) + mu)%%(2*pi)
}


# 1.2 Compute PDF of symmetric wrapped stable
# The infinite sum in wrapped stable PDF is approximated by a finite sum.
# den.n.term is the number of summands used in the finite sum
dwrpstab.sym = function(x, gamma, tau, mu, den.n.term = 100){
  summand.part1.vec = exp(-(tau*(1:den.n.term))^gamma)
  summand.part2.mat = outer(1:den.n.term, x - mu,  function(x, y){cos(x * y
```

```
    )})

  return(c((summand.part1.vec %*% summand.part2.mat + 0.5)/pi))

}


# 1.3 Compute CDF of symmetric wrapped stable
pwrpstab.sym = function(q, gamma, tau, mu, den.n.term = 100){

  den.fun = function(x){

    dwrpstab.sym(x, gamma = gamma, tau = tau, mu = mu, den.n.term = den.n.

      term)

  }

  F.fun = function(x){

    integrate(den.fun, 0, min(x, (2 * pi)), stop.on.error = FALSE)$value

  }

  return(sapply(q, F.fun))

}


# 1.4 Estimate parameters for wrapped stable sample
# gamma.min.val sets a lower bound for estimated values of parameter gamma;

    if gamma is too small there will be numerical problem
wrpstab.param.est = function(dat, gamma.min.val = 0.1){

  mu = (atan2(mean(sin(dat)), mean(cos(dat))))%%(2*pi)


  theta.hat = dat - mu

  R1 = sqrt(mean(cos(theta.hat))^2 + mean(sin(theta.hat))^2)

  R2 = sqrt(mean(cos(2 * theta.hat))^2 + mean(sin(2 * theta.hat))^2)


  gamma = log(log(R2)/log(R1)) / log(2)


  gamma = max(min(gamma, 2), gamma.min.val)


  tau = (-log(R1))^(1/gamma)
```

81

```
    return(list(gamma = gamma, tau = tau, mu = mu, rho = R1))

}


### 2. Functions for ChF-based goodness-of-fit test
# 2.1 These functions computes Epsilon_1, _2 and _3 for test statistic
# Infinite sums in Epsilon_2 and _3 are approximated by finite sums with
    number of summands specified by n.terms
test.stat.Eps1 = function(theta, lambda){
  cos(lambda * sin(theta)) * exp(lambda * (cos(theta) - 1))
}


test.stat.Eps2 = function(gamma, tau, lambda, n.terms = 20){
  sum(exp(-2 * (tau * (0:(n.terms - 1)))^gamma) * dpois(0:(n.terms - 1),
      lambda))
}


test.stat.Eps3 = function(theta, gamma, tau, lambda, n.terms = 20){
  sum(cos(theta * (0:(n.terms - 1))) * exp(-(tau * (0:(n.terms - 1)))^gamma
      ) * dpois(0:(n.terms - 1), lambda))
}


# 2.2 Compute test statistic (denoted by T)
# If the estimated parameters are known, they can be supplies as a list (
    gamma, tau, mu, rho) to argument param.est
T.test.stat = function(dat, lambda, n.terms = 20, param.est = NULL){
  n = length(dat)
  theta.hat = dat - (atan2(mean(sin(dat)), mean(cos(dat))))%%(2*pi)
  if(is.null(param.est)){
    param = wrpstab.param.est(dat)
  } else {
    param = param.est
```

82

```
  }
  Sigma1 = sum(sapply(outer(theta.hat, theta.hat, "-"), test.stat.Eps1,
      lambda = lambda))/n
  Sigma2 = n * test.stat.Eps2(param$gamma, param$tau, lambda, n.terms)
  Sigma3 = sum(sapply(theta.hat, test.stat.Eps3, gamma = param$gamma, tau =
      param$tau, lambda = lambda, n.terms = n.terms))
  return(Sigma1 + Sigma2 - 2 * Sigma3)
}


### 3. Functions for other goodness-of-fit tests (Kuiper, Watson)


# 3.1 Kuiper test
kuiper.test.stat = function(dat){
  n = length(dat)
  sorted.dat = sort(dat)
  return(max(sorted.dat - (0:(n-1))/n, (1:n)/n - sorted.dat))
}


# 3.2 Watson test
watson.test.stat = function(dat){
  n = length(dat)
  sorted.dat = sort(dat)
  return(1/12/n + sum((sorted.dat - (2*(1:n)-1)/2/n - (atan2(mean(sin(
      sorted.dat)), mean(cos(sorted.dat))))%%(2*pi) + 1/2)^2))
}


# 3.3 This function computes the test statistics for both tests.
# If the estimated parameters are known, they can be supplied as a list (
    gamma, tau, mu, rho) to argument param.est
kuiper.and.watson.test.stat = function(dat, param.est = NULL){
  n = length(dat)
```

83

```r
  if(is.null(param.est)){
    param = wrpstab.param.est(dat)
  } else {
    param = param.est
  }
  sorted.U = pwrpstab.sym(sort(dat), param$gamma, param$tau, param$mu)
  K = max(sorted.U - (0:(n-1))/n, (1:n)/n - sorted.U)
  W = 1/12/n + sum((sorted.U - (2*(1:n)-1)/2/n - (atan2(mean(sin(sorted.U))
      , mean(cos(sorted.U))))%%(2*pi) + 1/2)^2)
  return(c(K = K, W = W))
}



### 4. Main function for computing (1 - power) for all tests


# 4.1 This function computes (1 - power) of these 7 tests: our test with 5
    different lambda values (0.3, 0.5, 0.7, 0.9, 1),
#Kuiper, and Watson's test.
# B is the number of of times each test are performed. (Also the sample
    size for wrap-speed bootstrapping)
get.all.test.results = function(dat, B = 100, n.terms = 20){
  param = wrpstab.param.est(dat)
  T.1 = T.test.stat(dat, lambda = 0.3, n.terms = n.terms, param.est = param
      )
  T.2 = T.test.stat(dat, lambda = 0.5, n.terms = n.terms, param.est = param
      )
  T.3 = T.test.stat(dat, lambda = 0.7, n.terms = n.terms, param.est = param
      )
  T.4 = T.test.stat(dat, lambda = 0.9, n.terms = n.terms, param.est = param
      )
  T.5 = T.test.stat(dat, lambda = 1, n.terms = n.terms, param.est = param)
  K.and.W = kuiper.and.watson.test.stat(dat, param.est = param)
```

84

```
K = c(K.and.W["K"])
W = c(K.and.W["W"])


n = length(dat)
bs.sample = matrix(rwrpstab.sym(B*n, param$gamma, param$tau, param$mu), B
    , n)
T.bs.test.stat1 = apply(bs.sample, 1, T.test.stat, lambda = 0.3, n.terms
    = n.terms)
T.bs.test.stat2 = apply(bs.sample, 1, T.test.stat, lambda = 0.5, n.terms
    = n.terms)
T.bs.test.stat3 = apply(bs.sample, 1, T.test.stat, lambda = 0.7, n.terms
    = n.terms)
T.bs.test.stat4 = apply(bs.sample, 1, T.test.stat, lambda = 0.9, n.terms
    = n.terms)
T.bs.test.stat5 = apply(bs.sample, 1, T.test.stat, lambda = 1, n.terms =
    n.terms)
kuiper.and.watson.bs.test.stat = apply(bs.sample, 1, kuiper.and.watson.
    test.stat)
kuiper.bs.test.stat = kuiper.and.watson.bs.test.stat["K", ]
watson.bs.test.stat = kuiper.and.watson.bs.test.stat["W", ]


return(c(T1.p = sum(T.bs.test.stat1 < T.1), T2.p = sum(T.bs.test.stat2 <
    T.2), T3.p = sum(T.bs.test.stat3 < T.3),
  T4.p = sum(T.bs.test.stat4 < T.4), T5.p = sum(T.bs.test.stat5 < T.5),
  K.p = sum(kuiper.bs.test.stat < K), W.p = sum(watson.bs.test.stat < W))
      / B)
}



# 4.2 This function applies all tests to each of the dataset in the dat.
  list
```

85

```
test.procedure = function(dat.list, B = 100, n.terms = 20){

  p.mat = matrix(NA, length(dat.list), 7)

  for(i in 1:length(dat.list)){

    p.mat[i,] = get.all.test.results(arb.data.list[[i]], B = B, n.terms = n

        .terms)

  }

  return(p.mat)

}



### 5. Test using real data

# Read data

# arb.data.list is the list of all dataset we want to test.

arb.data.list = ...



# Compute (1-power) for all tests from real data

arb.p.res = test.procedure(arb.data.list, 500)



# Print power of the tests

1 - arb.p.res
```

A.2   Code for Bayesian Test for Unimodality of Gaussian Mixture (Chapter 6)

```
### 1. Functions for certain distribution families used in prior or

    posterior distributions

# 1.1 Sample from inverse Gamma

rinvgamma = function (n, shape, scale = 1){

  return(1/rgamma(n = n, shape = shape, rate = scale))

}
```

```
# 1.2 Compute PDF of inverse Gamma

dinvgamma = function (x, shape, scale = 1)

{

  return(scale^shape/gamma(shape)*x^(-shape-1)*exp(-scale/x))

}


# 1.3 Find the maximum of beta distribution within interval (0,0.5]

# Parameters a, b must be > 1

# This function is used in

beta.trunc.max = function(a, b){

  if(a < 1 || b < 1) stop ("Beta distribution parameters are not greater

      than 1.\n")

  beta.mode = (a - 1) / (a + b - 2)

  return(dbeta(min(beta.mode, 0.5), a, b))

}


# 1.4 Sampling for truncated beta distribution (defined on (0,0.5])

rbeta.trunc = function(a, b){

  maxi = beta.trunc.max(a, b)

  while(1){

    x = runif(1, 0, 0.5)

    d = runif(1, 0, maxi)

    if(d <= dbeta(x, a, b)) return(x)

  }

}


### 2. Functions for 2-component Gaussian mixture

# 2.1 Check if a 2-component Gaussian mixture is bimodal for given

    parameter values, return a string outcome

check.bimodal = function(mu10, mu20, sigsq0, pi0){

  D0<-abs((mu10-mu20)/2/sqrt(sigsq0))
```

87

```
  Dsq0<-D0^2

  if (Dsq0 > 1){

    if (abs(log(pi0/(1-pi0))) < 2 * log (D0 - sqrt(Dsq0 - 1)) + 2 * D0 *
        sqrt(Dsq0 - 1)){

      return("bimodal")

    } else {

      return("unimodal")

    }

  } else {

    return("unimodal")

  }

}


# Check if a 2-component Gaussian mixture is bimodal for given parameter
    values, return a boolean outcome.
# 2.2 The parameters can be vectors if multiply combinations of parameters
    are to be checked
check.bimodal.binary = function(mu10, mu20, sigsq0, pi0){

  D0<-abs((mu10-mu20)/2/sqrt(sigsq0))

  Dsq0<-D0^2

  return((Dsq0 > 1) & (abs(log(pi0/(1-pi0))) < 2 * log (D0 - sqrt(pmax(Dsq0
      - 1, 0))) + 2 * D0 * sqrt(pmax(Dsq0 - 1, 0))) )

}


# 2.3 Sample from Gaussian mixture for given parameter values
generate.sample = function(n, mu10, mu20, sigsq0, pi0){

  indic<-rbinom(n,1,pi0)

  sampl.1<-rnorm(n,mu10,sqrt(sigsq0))

  sampl.2<-rnorm(n,mu20,sqrt(sigsq0))

  return(ifelse(indic, sampl.1, sampl.2))

}
```

88

```
### 3. Generate Markov chains for prior and posterior distributions
# 3.1 Gibbs sampler for prior distribution for given hyperparameter values
# iter.max is the number of iterations
gibbs.prior = function(v, ssq, ksi1, ksi2, m1, m2, iter.max = 1e5){
  pi.v = mu1.v = mu2.v = sigsq.v = rep(0, iter.max)
  iter = 1
  while(iter <= iter.max){

    pi = runif(1, 0, 0.5)
    sigsq = rinvgamma(1, v/2, ssq/2)
    mu1 = rnorm(1,ksi1,sqrt(sigsq/m1))
    mu2 = rnorm(1,ksi2,sqrt(sigsq/m2))

    pi.v[iter] = pi
    sigsq.v[iter] = sigsq
    mu1.v[iter] = mu1
    mu2.v[iter] = mu2

    iter = iter + 1
  }
  return(data.frame(pi = pi.v, mu1 = mu1.v, mu2 = mu2.v, sigsq = sigsq.v))
}


# 3.2 MCMC for posterior distribution
# The arguments mu1, mu2, sigsq, pi should be supplied the initial values
    for these parameters
MCMC.posterior = function(sampl, ssq, ksi1, ksi2, m1, m2, mu1, mu2, sigsq,
    pi, iter.max = 1e5){
  sampl.ssq <- sum(sampl^2)
```

89

```
pi.v = mu1.v = mu2.v = sigsq.v = rep(0, iter.max)

sig = sqrt(sigsq)

n = length(sampl)

iter = 1

while(iter <= iter.max){

  #vector of parameters of Bernoulli for z2

  pz <- pi * dnorm(sampl,mu1,sig) / ((1 - pi) * dnorm(sampl,mu2,sig) + pi
      * dnorm(sampl,mu1,sig))

  z1 <- sapply(1:n, function(i){rbinom(1,1,pz[i])})


  z1.sum <- sum(z1) #number of samples from N(mu2, sigma)

  z2 <- 1 - z1 #vector of z1

  z2.sum <- sum(z2)

  z1sampl.sum <- sum (z1 * sampl)

  z2sampl.sum <- sum (z2 * sampl)


  # pi has to be <= 0.5

  pi = rbeta.trunc(z1.sum+1, z2.sum+1)


  #Calculate second parameter of invgamma (posterior of sigsq)

  invg.para2 <- (ssq + sampl.ssq + m1 * ksi1^2 + m2 * ksi2^2

                 - (z1sampl.sum + m1 * ksi1)^2 / (m1 + z1.sum)

                 - (z2sampl.sum + m2 * ksi2)^2 / (m2 + z2.sum)) / 2


  sigsq <- rinvgamma(1, (n + v)/2, invg.para2)

  sig <- sqrt(sigsq)


  mu1.mu <- (z1sampl.sum + m1 * ksi1)/(z1.sum + m1)

  mu1.sigsq <- sigsq / (z1.sum + m1)


  mu2.mu <- (z2sampl.sum + m2 * ksi2)/(z2.sum + m2)
```

90

```
    mu2.sigsq <- sigsq / (z2.sum + m2)


    mu1 <- rnorm(1,mu1.mu,sqrt(mu1.sigsq))
    mu2 <- rnorm(1,mu2.mu,sqrt(mu2.sigsq))


    pi.v[iter] = pi
    sigsq.v[iter] = sigsq
    mu1.v[iter] = mu1
    mu2.v[iter] = mu2


    iter = iter + 1
  }
  return(data.frame(pi = pi.v, mu1 = mu1.v, mu2 = mu2.v, sigsq = sigsq.v))
}


### 4. Compute probability of bimodality
# 4.1 Calculate P(bimodal) for prior
calculate.bimod.prior = function(v, ssq, ksi1, ksi2, m1, m2, len = 1e5){
  gbs.prior = gibbs.prior(v, ssq, ksi1, ksi2, m1, m2, iter.max = len)
  is.bimod.prior = check.bimodal.binary(gbs.prior$mu1, gbs.prior$mu2, gbs.
      prior$sigsq, gbs.prior$pi)
  return(mean(is.bimod.prior))
}


# 4.2 Calculate P(bimodal | data) for posterior
calculate.bimod.post = function(MCMC.sampl, burn.in.prop = 0.2){
  size2 = nrow(MCMC.sampl)
  gbs.post = MCMC.sampl[(floor(burn.in.prop * size2)+1):size2,]
  is.bimod.post = check.bimodal.binary(gbs.post$mu1, gbs.post$mu2, gbs.
      post$sigsq, gbs.post$pi)
```

```
  return(mean(is.bimod.post))

}


### 5. Real data application
# Read real data
height_dat = ...
n = length(height_dat)


# Set hyperparameters for the prior, and the number of iterations
v = 5
ksi1 = 150
ksi2 = 180
ssq = 400
m1 = 10
m2 = 10
iter.max0 = 1e5


# Calculate P(bimodal) for prior distribution
prior.prob = calculate.bimod.prior(v, ssq, ksi1, ksi2, m1, m2, len = iter.
    max0)


# Generate Markov chain for posterior distribution
height_dat.gibbs = MCMC.posterior(sampl = height_dat, ssq, ksi1, ksi2, m1,
    m2,
              mu1 = 140, mu2 = 180, sigsq = 100, pi = 0.5, iter.max = iter.
                 max0)


# Calculate P(bimodal | data) for posterior distribution
height_dat.post.prob = calculate.bimod.post(height_dat.gibbs)


# Calculate Bayes factor
```

92

```
height_dat.Bayes.factor =  height_dat.post.prob / (1 - height_dat.post.prob
    ) * (1-prior.prob) / prior.prob

print(height_dat.Bayes.factor)
```

Bibliography

[1] Toshihiro Abe and Arthur Pewsey. "Sine-skewed circular distributions". In: *Statistical Papers* 52.3 (2011), pp. 683–707.

[2] J. Allison and L. Santana. "On a data-dependent choice of the tuning parameter appearing in certain goodness-of-fit tests". In: *J. Statist. Comput. Simul.* 85 (2015), pp. 3276–3288.

[3] Adelchi Azzalini. "The skew-normal distribution and related multivariate families". In: *Scandinavian Journal of Statistics* 32.2 (2005), pp. 159–188.

[4] Sanjib Basu and S. Rao Jammalamadaka. "Unimodality in Circular Data: A Bayes Test". In: vol. Advances on Methodological and Applied Aspects of Probability and Statistics, (Ed. Balakrishnan, N. et al). Taylor & Francis, 2002, pp. 141–153.

[5] Edward Batschelet. *Statistical Methods for the Analysis of Problems in Animal Orientation and Certain Biological Rhythms*. Washington D.C.: American Institute of Biological Sciences, 1965.

[6] Javad Behboodian. "On a mixture of normal distributions". In: *Biometrika* 57.1 (Apr. 1970), pp. 215–217.

[7] R.J. Beran. "Asymptotic theory of a class of tests for uniformity of a circular distribution". In: *Ann. Math. Statist.* 40 (1969), pp. 1196–1206.

[8] Carlos Coelho. "The Wrapped Gamma Distribution and Wrapped Sums and Linear Combinations of Independent Gamma and Laplace Distributions". In: *Journal of Statistical Theory and Practice* 1 (Mar. 2007), pp. 1–29.

[9] S. Csörgő. "Testing for stability". In: *Goodness–of–fit, Colloquia Mathematica Societatis Janos Bolyai* Vol 45 (1987), pp. 101–132.

[10] J. C. Dunn. "Well-Separated Clusters and Optimal Fuzzy Partitions". In: *Journal of Cybernetics* 4.1 (1974), pp. 95–104.

[11] T.W. Epps. "Characteristic functions and their empirical counterparts: Geometric interpretations and applications to statistical inference". In: *Amer. Statistician* 47 (1993), pp. 33–38.

[12]   T. S. Ferguson. "Bayesian Density Estimation by Mixtures of Normal Distributions". In: *Recent Advances in Statistics (M. Rizvi, J. Rustagi, and D. Siegmund, eds.)* 24 (1983), pp. 287–302.

[13]   A. Feuerverger and R. Mureika. "The empirical characteristic function and its applications". In: *Ann. Statist.* 5 (1977), pp. 88–97.

[14]   Riccardo Gatto and S. Rao Jammalamadaka. "The generalized von Mises distribution". In: *Statistical Methodology* 4 (July 2007), pp. 341–353.

[15]   R. Giacomini, D.N. Politis, and H. White. "A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators". In: *Econometr. Theor.* 29 (2013), pp. 567–589.

[16]   Hajo Holzmann and Sebastian Vollmer. "A likelihood ratio test for bimodality in two-component mixtures with application to regional income distribution in the EU". In: *AStA Advances in Statistical Analysis* 92.1 (2008), pp. 57–69.

[17]   K. Hornik and B. Grün. "movMF: An R Package for Fitting Mixtures of von Mises-Fisher Distributions". In: *Journal of Statistical Software, Articles* 58.10 (2014), pp. 1–31.

[18]   S. Rao Jammalamadaka and T. J. Kozubowski. "A General Approach for Obtaining Wrapped Circular Distributions via Mixtures". In: *Sankhya A* 79 (Apr. 2017).

[19]   S. Rao Jammalamadaka and T. J. Kozubowski. "New Families of Wrapped Distributions for Modeling Skew Circular Data". In: *Communications in Statistics - Theory and Methods* 33.9 (2004), pp. 2059–2074.

[20]   S. Rao Jammalamadaka and A. SenGupta. *Topics in Circular Statistics*. Singapore: World Scientific Press, 2001.

[21]   Robert E. Kass and Adrian E. Raftery. "Bayes factors". In: *Jour. of Amer. Statist. Assoc* 90.430 (1995), pp. 773–795.

[22]   I.A. Koutrouvelis and S.G. Meintanis. "Testing for stability based on the empirical characteristic funstion with applications to financial data". In: *Statist. Comput. Simul.* 64 (1999), pp. 275–300.

[23]   M. Matsui and A. Takemura. "Goodness-of-fit tests for symmetric stable distributions–Empirical characteristic function approach". In: *Test* 17 (2008), pp. 546–566.

[24]   S.G. Meintanis. "Consistent tests for symmetric stability with finite mean based on the empirical characteristic function". In: *J. Statist. Plann. Infer.* 128 (2005), pp. 373–380.

[25]   S.G. Meintanis, J. Ngatchou-Wandji, and E. Taufer. "Goodness-of-fit tests for multivariate stable distributions based on the empirical characteristic function". In: *J. Multivar. Anal.* 140 (2015), pp. 171–192.

[26] S.G. Meintanis and T. Verdebout. "Le Cam maxmin tests for symmetry of circular data based on the characteristic function". In: *Statist. Sinica* 29 (2019), pp. 1301–1320.

[27] *National Health and Nutrition Examination Survey.* `https://www.cdc.gov/nchs/nhanes/index.htm`.

[28] Arthur Pewsey. "Testing Circular Symmetry". In: *The Canadian Journal of Statistics / La Revue Canadienne de Statistique* 30.4 (2002), pp. 591–600.

[29] J.R. Pycke. "Some tests of uniformity of circular distributions powerful against multimodal alternatives". In: *Canad. J. Statist.* 38 (2010), pp. 80–96.

[30] S. Pyne. Personal communication. School of Public Health, Univ of Pittsburgh.

[31] P. J. Rousseau. "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis". In: *Journal of Computational and Applied Mathematics* 20 (1987), pp. 53–65.

[32] Mark F. Schilling, Ann E. Watkins, and William Watkins. "Is Human Height Bimodal?" In: *The American Statistician* 56.3 (2002), pp. 223–229.

[33] R. Shanker and A Mishra. "A quasi Lindley distribution". In: *African Journal of Mathematics and Computer Science Research* 6.4 (2013), pp. 64–71.

[34] A. Taylor and K. Burns. "Radial distributions of air plants: a comparison between epiphytes and mistletoes". In: *Ecology* 97 (2016), pp. 819–825.

[35] H. Teicher. "On the Mixture of Distributions". In: *Annals of Math. Statist,* 31.1 (1960), pp. 55–73.

[36] Dale Umbach and S. Rao Jammalamadaka. "Building asymmetry into circular distributions". In: *Statistics & probability letters* 79.5 (2009), pp. 659–663.

[37] Brian Wainright. "Contributions to Directional Statistics Based Clustering Methods". PhD thesis. University of California, Santa Barbara, Sept. 2019.