UNIVERSITY OF CALIFORNIA
Santa Barbara

A Change-point Problem and Preliminary Test Estimation in Circular Statistics

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

Doctor of Philosophy

in

Statistics & Applied Probability

by

Michael Marcelino Nava

Committee in Charge:

Professor Sreenivasa Rao Jammalamadaka, Chair

Professor John Hsu

Professor Wendy Meiring

September 2015

The Dissertation of
Michael Marcelino Nava is approved:

Professor John Hsu

Professor Wendy Meiring

Professor Sreenivasa Rao Jammalamadaka, Committee Chairperson

June 2015

A Change-point Problem and Preliminary Test Estimation in Circular Statistics

by

Michael Marcelino Nava

For my wife. We made it!

Acknowledgements

I am most thankful to Professor Sreenivasa Rao Jammalamadaka for his patience, constant encouragement, and for generously sharing his wealth of knowledge in various branches of statistics. It was a true honor to learn from one of the best statisticians of our time.

I would also wish to thank Professors Wendy Meiring and John Hsu for their support and for serving on my thesis committee. My very special thanks go to Professors Magnus Ekstrom, Stephane Guerrier, and Kaushik Ghosh for their willingness to join and participate in our research group. In the short-time of their visits to our department, they made a lasting impact on my research and professional development.

Another special thank you goes to my fellow graduate students, especially Jose Ochoa, Ekaterina Shatskikh, and Mark Dela for the good laughs they shared when needed and for being my crutch in difficult times.

And finally, I wish to express my gratitude to all the faculty and staff of our great department for making our program the most rewarding that it can be.

Curriculum Vitæ

Michael Marcelino Nava

Education

2015	Ph.D. in Statistics & Applied Probability, University of California, Santa Barbara.
2012	M.A. in Mathematical Statistics, University of California, Santa Barbara.
2009	B.S. in Mathematics, California State University of Channel Islands, Camarillo.

Experience

2011-2015	Teaching Assistant, Department of Statistics & Applied Probability, University of California, Santa Barbara.
2013-2015	Teaching Associate, Department of Statistics & Applied Probability, University of California, Santa Barbara.
2011-2013	Head/Lead Teaching Assistant, Department of Statistics & Applied Probability, University of California, Santa Barbara.
2014-2015	Research Mentor, California Institute for Regenerative Medicine, University of California, Santa Barbara.

Awards

2014-15	UCSB GSA Teaching Award (Teaching Associate Category) Honorable mention
2013-14	UCSB GSA Teaching Award (STEM TA Category) Honorable mention
2009-2011	UC Bridge to Doctorate Fellowship

Abstract

A Change-point Problem and Preliminary Test Estimation in Circular Statistics

Michael Marcelino Nava

This thesis investigates two different problems relating to circular data. One relates to change-point problems. Tests in this context are meant to detect the point in time at which a sample of observations changes the probability distribution from which they came. Suppose one has a set of independent vectors of measurements, observed in a time-ordered or space-ordered sequence. In our set-up, these observations are circular data and we are interested as to which point in time does the distribution change from having one mode to having more than one mode. In this work we model unimodality or bimodality with a mixture of two Circular Normal distributions, which admits both possibilities, albeit for different parameter values. Tests for detecting the change-point are derived using the generalized likelihood ratio method. We obtain simulated distributions and critical values for the appropriate test statistics in finite samples, as well as provide the asymptotic distributions, under some regularity conditions. We also tackle this problem from a Bayesian perspective. In the second part, the goal is to estimate the concentration parameter of a Circular Normal distribution when the mean direction is unknown. We present two alternate approaches that incorporate prior knowledge on the mean direction (i) via a preliminary test on the mean direction, the so-called "preliminary test estimators" and (ii) through an assumed prior distribution on the mean direction as

one does in Bayes procedures. We compare such alternate estimators with the standard maximum likelihood estimator and explore when one method is superior to the other.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In cell biology, scientists are interested in studying various characteristics of the cell, such as its morphology, size, cell cycle phase, DNA content, and the presence or absence of specific proteins on the cell surface or in the cytoplasm. These characteristics are useful for research in cell biology as well as in medical diagnostics for a wide range of diseases such as cancer and AIDS.

Flow cytometry is among the most widely used platforms in biomedical research and clinical labs. It is used for investigation of a wide variety of biological problems at single cell level. Classical applications of flow cytometry include quantitative measurements of DNA content and cell cycle progression (Darzynkiewicz *et al.* , 2004). It is also one of the key platforms for studying dynamic cellular properties such as differentiation, proliferation and apoptosis, especially in the context of stem cells and cancer ((Eds.)Krishan *et al.* , 2011).

Given a group of cells, flow cytometry records characteristics, in a similar way our ink jet printers work. The method applies an antibody with a florescent dye to the cells, then sends the cells through a laser. Different types of substances are used to study different characteristics needed by the scientist. Fluorescent Activated Cell Sorting (FACS) can then sort the cells into two or more groups.

We refer to (Pyne *et al.* , 2009), (Ho *et al.* , 2012), and (Aghaeepour *et al.* , 2013) as a few recent works which attempt to model the cell cycle. There are also studies that model cell transitions over time. An important aspect of this transition is the observed fact that the cell cycle of a stem cell before it transforms into a specialized cell, can be modeled by a unimodal circular distribution, followed by a multimodal distribution after transformation into a specialized cell. A biologist would be very interested to identify the time point at which this change occurs. This change in modality is an indication that the stem cell has become a specialized cell and that change-point is an crucial piece of information to the cell biologist.

Given a set of independent vectors of circular observations, $\underset{\sim}{\alpha}_1, \underset{\sim}{\alpha}_2, \ldots, \underset{\sim}{\alpha}_T$ that are time-ordered as in our case, we are interested to find the point in time in which the observations change from having a unimodal distribution to a multimodal distribution. Here, $\underset{\sim}{\alpha}_j = (\alpha_{j1}, \alpha_{j2}, \ldots, \alpha_{jm})$, is a vector of independent observations observed at time j with length of m. For simplicity, we will assume that each vector of observations is of the same length m, although this can be generalized.

Specifically, we assume there is some unknown but fixed k, $(1 \leq k \leq T-1)$ such that $\underset{\sim}{\alpha}_1, \ldots, \underset{\sim}{\alpha}_k$ have unimodal densities with pdf's in $\{f_1\}$ and $\underset{\sim}{\alpha}_{k+1}, \ldots, \underset{\sim}{\alpha}_T$ have multimodal densities with pdf's in $\{f_2\}$. The point k, is considered the "change-point" of the observed data.

In our approach we use the Generalized Likelihood Ratio Test (GLRT) to test for the presence of a change-point i.e., $H_0 : k = n$ versus $H_1 : 1 \leq k \leq T-1$. The null hypotheses corresponds to no change in distribution over the sequence of observations until perhaps the end i.e. all the given data have a unimodal distribution. The alternative hypothesis corresponds to there being exactly one change in modality of distribution at some k^{th} step in-between, before the end of the sequence.

This problem and the test involve two-dimensional directional data, also called circular data, which requires quite a different treatment and methodologies than traditional linear statistical methods. So, before we continue in describing our change-point test, we will give a brief introduction to circular statistics and methods.

1.1 An Overview of Circular Statistics

Whether scientists are studying the direction of the earth's magnetic pole, the direction of flight of migrating birds, or to see if there is a preferred direction in which a cricket or baseball player hits the ball, the type of data studied is referred to as directional data. In each scenario that we consider in this thesis the scientist is only interested in direction and not the magnitude, so the data can be represented as points on the circumference

of the unit circle, which gives arise to the name "circular data", or in the case of 3-dimensional directions on the surface of the unit sphere, making for spherical data. Now, one may be tempted to use linear statistical methods for their analysis but we will give simple examples as to why linear statistical methods are inadequate and indeed wrong.

In the cricket example, suppose we record the hit direction of two batted balls by one of the best cricket players of all time, Sachin Tendulkar. The opposing team would like to see if Tendulkar has a preferred hitting direction based off his two previous hits. After we have selected a suitable "zero direction" we have two measurements of $\alpha_1 = 15\,^\circ$ and $\alpha_2 = 345\,^\circ$. The usual summary statistic of interest would be the "mean" direction of batted balls. If we used the usual arithmetic mean, the mean direction would be $\bar{\alpha} = 180\,^\circ$. Figure 1.1 illustrates that the arithmetic mean direction points in the opposite direction of the observed batted ball angles. The opposing team would make Tendulkar's day much easier if they shifted their outfielders to this, albeit wrong, average direction. The opposing coach may lose her/his job over such a call. This implies the need for different methodologies for analysis of circular data.

Supposing there are two other players on the team conducting their own statistical analyses. One chooses true North as zero direction and the other chooses true East as the zero direction. The first records his data as $\alpha_1 = 30\,^\circ$ and $\alpha_2 = 60\,^\circ$, while the second records $\alpha_1 = 170\,^\circ$ and $\alpha_2 = 200\,^\circ$. Their preferred mean directions are $\bar{\alpha} = 45\,^\circ$ and $185\,^\circ$ respectively. The three preferred means are spread around the circle and the linear mean is dependent on the selected zero direction. The circular mean will give the same location

on the circle for all three cases. This motivates the requirement that methodologies in circular statistics be rotationally invariant, i.e. independent of the chosen zero direction.

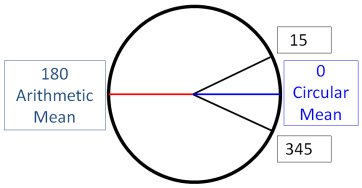

Figure 1.1: Circular vs. Linear Mean

Since the observations can be represented as points on the circumference of the unit circle, rectangular coordinates provide an appropriate representation of $\alpha_i = (x_i, y_i)$. Using polar coordinates, $x_i = r\cos(\alpha_i)$ and $y_i = r\sin(\alpha_i)$, with $r = 1$. Note that (x_i, y_i) are not really bivariate data since $x_i^2 + y_i^2 = 1$ for all points on the unit circle which has area 0 on the plane. With this representation, the first step in computing the circular mean is to compute the resultant vector for the n observations:

$$\mathbf{R} = \left(\sum_{j=1}^{n} \cos\alpha_j, \sum_{j=1}^{n} \sin\alpha_j\right) = (C, S). \tag{1.1}$$

The resultant vector is the result of the sum of two or more observations in unit vector form. The direction of this resultant vector is proposed as the mean direction, denoted as $\bar{\alpha}_0$. Next, the following trigonometric function provides the mean direction:

$$\bar{\alpha}_0 = \arctan\left(\frac{S}{C}\right) \tag{1.2}$$

Of course, we need to pay attention to the quadrant-specific inverse tangent function in (1.2). When the the circular mean method is applied to our example we get the circular mean $\bar{\alpha}_0 = 0\,^\circ$. As seen in Figure 1.1, this gives the true summary statistic for the data. Other relevant and interesting properties of this mean direction $\bar{\alpha}_0$, include:

$$\sum_{i=1}^{n} \sin(\alpha_i - \bar{\alpha}_0) = 0 \tag{1.3}$$

and

$$\sum_{i=1}^{n} \cos(\alpha_i - \bar{\alpha}_0) = R. \tag{1.4}$$

Suppose the sample is drawn from a population with mean direction μ:

$$\sum_{i=1}^{n} \cos(\alpha_i - \mu) = V_0, \tag{1.5}$$

where in (Jammalamadaka & SenGupta, 2001), they prove $V_0 \leq R$, with equality if and only if $\bar{\alpha}_0 = \mu$.

Thus using $\bar{\alpha}_0$, the opposing team's coach can now place his outfielders towards Tendulkar's true preferred hitting direction. Note that Tendulkar could make the opposing team's day more difficult if his hit directions were recorded as $\alpha_1 = 0\,^\circ$ and $\alpha_2 = 180\,^\circ$. We see C in (1.2) would be 0 and the inverse tangent function is undefined. The interpretation of this would be that Tendulkar does not have a preferred (or mean) direction of hitting. So the placement of the outfielders for his next at-bat could be anyone's guess. This interpretation should make sense as the batted balls were hit in two opposite direc-

tions. Whether or not there exists a preferred or mean direction for a population given a sample of circular data, as opposed to isotropy, is a basic question in circular statistics.

We also need to define probability density functions with properties required for circular data. Examples include the Circular Normal (CN, from now on) and mixture of two Circular Normal probability density functions, which play an important role in our work.

1.1.1 Circular Probability Densities

The total probability is concentrated on the circumference of the unit circle for a circular probability distribution. If the probability density function, say $f(\alpha)$, exists then it satisfies the properties:

(i) $f(\alpha) \geq 0$;

(ii) $\int_0^{2\pi} f(\alpha)\mathrm{d}\alpha = 1$;

(iii) $f(\alpha) = f(\alpha + k2\pi)$ for any integer k (i.e., f is periodic with period 2π).

The circular distributions share similarities to the linear distributions but we require the pdf's to be periodic. Next we present the most commonly used and popular distribution among the circular probability distributions, called the Circular Normal distribution.

1.1.2 The Circular Normal Distribution

In circular statistics, the Circular Normal Distribution (CND, from now on) is considered the analog to the Normal distribution in linear statistics. A Circular Normal random variable, α, has density function:

$$\frac{1}{2\pi I_0(\kappa)} \exp(\kappa \cos(\alpha - \mu)), \quad 0 \leq \alpha < 2\pi \tag{1.6}$$

The parameters of the distribution are the mean direction μ, and the concentration parameter, κ, where $0 \leq \mu < 2\pi$ and $\kappa \geq 0$. The modified Bessel function of the first kind and order zero is involved in the normalizing constant for the CND and is given by,

$$I_0(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} \exp(\kappa \cos \alpha) d\alpha = \sum_{r=0}^{\infty} \left(\frac{\kappa}{2}\right)^{2r} \left(\frac{1}{r!}\right)^2. \tag{1.7}$$

For more useful properties of the modified Bessel function refer to (Jammalamadaka & SenGupta, 2001) and (Mardia & Jupp, 1999). The distribution is also called the von Mises distribution after Richard von Mises (von Mises, 1918) who introduced this statistical model.

One refers to this model as the Circular Normal (CN) to emphasize the similarities it shares with the Normal distribution on the line. The CN is the most extensively studied circular distribution and is most popular choice for most data analyses.

Figure 1.2 has images of CND drawn on the line $[-\pi, \pi)$ with mean directions of 0 and concentration parameter values of 0.01, 1.5, 3. We see as κ becomes larger the

distribution becomes more concentrated around the mean direction. Also, for all values of κ the distribution is symmetric about the mean direction.

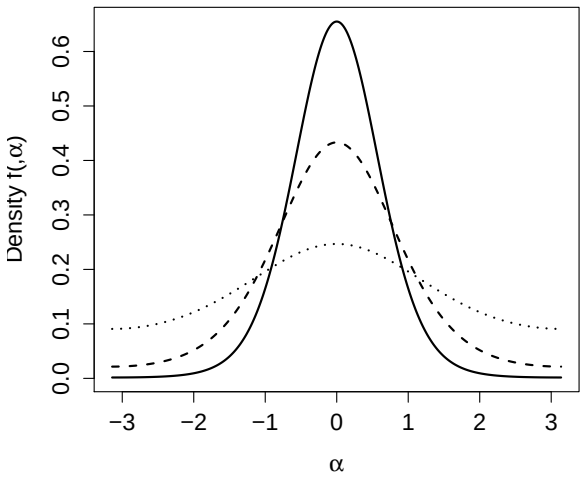

Figure 1.2: The Circular Normal Curve for κ=.01 ($\cdots$), 1.5 (- - -), 3 (—)

We list below some properties of the CND:

A Symmetry: By symmetry of $\cos(\alpha - \mu)$, the distribution is symmetric about the mean direction, μ, and $\mu + \pi$.

B Mode at μ: Since $\cos(\alpha - \mu)$ has maximum at $\alpha = \mu$, the Circular Normal density has maximum at $\alpha = \mu$. So μ is the modal direction with maximum value.

$$f(\mu) = \frac{e^{\kappa}}{2\pi I_0(\kappa)}. \tag{1.8}$$

C Antimode at $\mu \pm \pi$: $\cos(\alpha - \mu)$: has minimum for $\alpha = \mu \pm \pi$, then $\mu \pm \pi$ is called the anti-modal direction since the density is minimum at that direction.

$$f(\mu \pm \pi) = \frac{e^{-\kappa}}{2\pi I_0(\kappa)}. \tag{1.9}$$

D Role of κ: From equations (8) and (9) we have:

$$\frac{f(\mu)}{f(\mu \pm \pi)} = e^{2\kappa}. \tag{1.10}$$

A larger value for κ will increase the ratio in (1.10). The increase results in a high concentration towards the population mean and hence κ measures the concentration towards the mean direction μ.

Given a random sample $\alpha_1, \ldots, \alpha_n$ from a CND(μ, κ), it can be checked that the MLEs for μ and κ are given by:

$$\bar{\alpha}_0 = \arctan\left(\frac{\sum_{i=1}^{n} \sin(\alpha_i)}{\sum_{i=1}^{n} \cos(\alpha_i)}\right) \tag{1.11}$$

and

$$\hat{\kappa}_{\text{MLE}} \text{ is the solution to: } \quad \frac{I_1(\kappa)}{I_0(\kappa)} = \frac{1}{n}\sum_{i=1}^{n} \cos(\alpha_i - \bar{\alpha}_0) = \frac{R}{n} \tag{1.12}$$

since $\sum_{i=1}^{n} \cos(\alpha_i - \bar{\alpha}_0) = R$.

The MLE for the mean direction is independent of κ but the MLE for κ depends on the estimate for the mean direction. When the mean direction is known, then the MLE for κ is obtained by substituting the value μ in place of $\bar{\alpha}_0$ in 1.12. The MLEs carry asymptotic properties.

If we write

$$A(\kappa) = \frac{I_1(\kappa)}{I_0(\kappa)},$$

then:

- $0 \leq A(\kappa) \leq 1$
- $A(\kappa) \to 0$ as $\kappa \to 0$ and $A(\kappa) \to 1$ as $\kappa \to \infty$
- $A'(\kappa) = (1 - \frac{A(\kappa)}{\kappa} - A^2(\kappa)) \geq 0.$

The fact that $A(\kappa)$ is a strictly monotonically increasing function of κ, guarantees a unique solution for the MLE. Checking the determinant of the Hessian matrix evaluated at our MLEs we are reassured that the estimates for μ and κ are maximum critical points on the joint parameter space.

The Fisher Information matrix in this case is given by,

$$I = \begin{bmatrix} A(\kappa)/\kappa & 0 \\ 0 & 1 - \frac{A(\kappa)}{\kappa} - A^2(\kappa) \end{bmatrix}.$$

Thus the asymptotic variance-covariance matrix for the MLE's evaluated at $(\bar{\alpha}_0, \hat{\kappa}_{\text{MLE}})$:

$$V = \begin{bmatrix} 1/R\hat{\kappa} & 0 \\ 0 & \frac{1}{n(1-\bar{R}/\hat{\kappa}-\bar{R}^2)} \end{bmatrix}. \tag{1.13}$$

In the aforementioned properties we find some similarities to the Normal Distribution as well as some major differences. In linear statistics, a univariate density f has a single mode if f is non-decreasing up to a point M and non-increasing thereafter. The lack of well-defined left and right-end points in circular statistics (i.e. -∞ and ∞ on the real line) leads the definition of the mode to also require an antimode A. A circular probability density $f(\alpha)$ is unimodal with mode at M if there exists an antimode A such that $f(\alpha)$ is non-decreasing for $A \leq \alpha \leq M$ and non-increasing for $M \leq \alpha \leq A$.

While the CND approaches a Circular Uniform distribution for small values of κ, we see below that the CND will approach the linear Normal distribution for very large values of κ.

Proposition 1 *As* $\kappa \to \infty$,

$$\beta = \sqrt{\kappa}(\alpha - \mu) \xrightarrow{d} N(0, 1),$$

where $\alpha \sim CN(\mu, \kappa)$.

Proof 1 *Recall the CND*

$$f(\alpha) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(\alpha - \mu)}, \ \ 0 \leq \alpha < 2\pi$$

.

Let $\beta = \sqrt{\kappa}(\alpha - \mu)$. *Then for large* κ *we have small* $\frac{\beta}{\sqrt{\kappa}}$,

$$\begin{aligned} \cos(\alpha - \mu) &= \cos\left(\frac{\beta}{\sqrt{\kappa}}\right) \\ &\simeq 1 - \frac{\beta^2}{2\kappa}. \end{aligned}$$

Here we used the Taylor series expansion for $\cos(\alpha)$ *and a change of variable. Suppose* $g(\beta)$ *is the pdf for* β *and we will use the fact that for large* κ, $I_0(\kappa) \simeq \exp(\kappa)/\sqrt{2\pi\kappa}$ *to get,*

$$\begin{aligned} g(\beta) &\simeq \frac{\exp\left(\kappa \cos\left(\frac{\beta}{\sqrt{\kappa}}\right)\right)}{2\pi I_0(\kappa)} \frac{1}{\sqrt{\kappa}} \\ &\simeq \frac{\exp\left(\kappa \cos\left(\frac{\beta}{\sqrt{\kappa}}\right)\right)}{2\pi \frac{\exp(\kappa)}{\sqrt{2\pi\kappa}}} \frac{1}{\sqrt{\kappa}} \\ &\simeq \frac{\exp\left(\kappa\left(1 - \frac{\beta^2}{2\kappa}\right)\right)}{e^{\kappa}\sqrt{2\pi}} \frac{1}{\sqrt{\kappa}} \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\beta^2}{2}\right). \end{aligned}$$

Therefore, $\beta \xrightarrow{d} N(0,1)$.

For concentration parameter, $\kappa = 0$, we have:

$$CN(\alpha|\mu, 0) = \frac{1}{2\pi}, \quad 0 \leq \alpha < 2\pi,$$

where the CND becomes the Circular Uniform density, a density that does not have a mean, or preferred direction. Note that the modified Bessel function in (1.7) equals to one in this case.

The CN distribution is an extensively studied density in circular statistics with many properties similar to that of the Normal Distribution in linear statistics. Note that the CND can only have up to one mode. Referring back to our introduction, we are interested in a circular density that can model one or more modes so a natural choice would be to use a mixture of CND.

1.2 Mixture of two Circular Normal Distributions

We model multimodality with a mixture of CNDs. The resulting family of distributions can be unimodal or bimodal, symmetric/asymmetric, and can take on many shapes, therefore mixtures of CNDs are good candidates for our parametric model. A circular random variable is said to have distribution being a mixture of CNDs with j components if the variable has density:

$$f(\alpha) = \sum_{i=1}^{j} \frac{p_i}{2\pi I_0(\kappa_i)} \exp(\kappa_i \cos(\alpha - \mu_i)) = \sum_{i=1}^{j} p_i CN(\alpha|\mu_i, \kappa_i), \quad 0 \leq \alpha < 2\pi \qquad (1.14)$$

where the mean direction of each component $0 \leq \mu_i < 2\pi$, concentration parameters $\kappa_i > 0$, and $I_0(\kappa_i)$ is the modified Bessel function as defined in 1.7. Also, p_i is the assigned weight of the ith single component to the mixture, with $0 \leq p_i \leq 1$, $\forall i \in \{1, \ldots, j\}$ and $\sum_{i=1}^{j} p_i = 1$. A mixture of j components can have up to j modes. The number of modes depends on the values of the parameters of mixture component mean directions, concentrations, and weights. Next we illustrate various shapes of the mixture of two CNDs (mixCN from now on) with equal concentration parameters.

In Figure 1.3 we examine how the mixture distribution changes shape as a single parameter is increased or decreased. In the graph on the left we begin with a mix$CN(\delta = 0, \kappa = 3, p = 0.05)$, where $\delta \triangleq \mu_2 - \mu_1$, which is the unimodal curve with the solid line. We increase δ to values of $(\pi/4, 3\pi/8, 3\pi/4)$ and plot each pdf curve. As the difference in the mean direction increases, the density curve become bimodal. In the center plot we examine the pdf curve for changes in the mixing proportion p, where $p_1 = p$ and $p_2 = 1 - p$ as in 1.14. In the graph we begin with a mix$CN(\delta = \pi/2, \kappa = 3, p = 0.85)$ which is the curve with a solid line. As we decrease p to values of $(0.7, 0.6, 0.5)$ the curve changes from unimodal to bimodal. For the right panel of Figure 1.3 we examine changes in the curve for different concentration parameter values of κ. In the right panel we begin with a mix$CN(\delta = \pi/2, \kappa = 4, p = 0.05)$ which is the bimodal curve with the solid line. We decrease the values of κ to $(3, 2, 1)$ and the density becomes unimodal when $\kappa = 1$.

In our work to come, we note the importance of being able to maximize the parameters over the restricted unimodal parameter space. The restricted parameter space can be

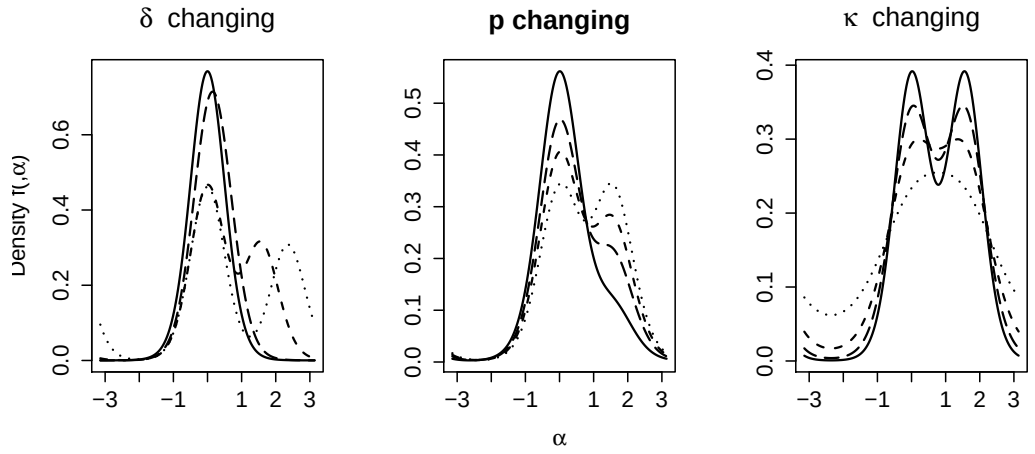

Figure 1.3: The Various Shapes of The Mixture of Two CNDs: Left Panel - $\delta = \{\pi/4(---), 3\pi/8(\text{- - -}), 3\pi/4(\cdots)\}$. Mid Panel - $p = \{(0.7(---), 0.6(\text{- - -}), 0.5(\cdots)\}$. Right Panel - $\kappa = \{3(---), 2(\text{- - -}), 1(\cdots)\}$.

difficult to compute in our case as the restricted parameter space is non-linear and the boundaries are also non-linear. To the best of our knowledge, no work has been done for finding the unimodal parameter space for a mixture with three or more components. (Mardia & Sutton, 1975), find the unimodal parameter space for a mixCNDs. Using their results, we use the mixCNDs as our parametric model for multimodality. Note in this case, we can model data with up to two modes under the mixCND aasumption.

The Mardia-Sutton condition gives the parameter space for unimodality and bimodality for a two component mixCND. The pdf for a mixCND can be written:

$$f(\alpha) = pCN(\alpha|\mu_1, \kappa_1) + (1-p)CN(\alpha|\mu_2, \kappa_2), \quad (0 \leq \alpha < 2\pi). \tag{1.15}$$

Without loss of generality we let $\mu_1 = 0$ and $\delta = \mu_2 - \mu_1$ for the mean direction parameters. The resulting density is given by,

$$f(\alpha) = pCN(\alpha|0,\kappa_1) + (1-p)CN(\alpha|\delta,\kappa_2), \quad (0 \leq \alpha < 2\pi). \tag{1.16}$$

The modes of the mixture (1.15) are solutions of:

$$f'(\alpha) = p\kappa_1 CN(\alpha|0,\kappa_1)\sin\alpha + (1-p)\kappa_2 CN(\alpha|\delta,\kappa_2)\sin(\alpha-\delta) = 0 \tag{1.17}$$

Table 1.1 provides a listing of the parameter values for unimodality (Ω_0) and bimodality (Ω_1) as presented by (Mardia & Sutton, 1975). Note that the parameter space may be expressed as a union, $\Omega = \Omega_0 \bigcup \Omega_1$, where Ω_0 corresponds to parameters for a unimodal distribution while Ω_1 corresponds to parameters for a bimodal distribution. Table 1.1 provides a list of the parameter space for bimodality that was presented by (Mardia & Sutton, 1975).

Table 1.1: Parameter Subspaces determine Modality for Mixture of 2 Circular Normals

Case	δ	Range of p	Type
(i)	0	$0 \leq p \leq 1$	Unimodal
(ii)	π	$\{1+\kappa^*\exp(\kappa_1+\kappa_2)\}^{-1} \leq p \leq \{1+\kappa^*\exp(-\kappa_1-\kappa_2)\}^{-1}$	Bimodal
		$0 \leq p < \{1+\kappa^*\exp(\kappa_1+\kappa_2)\}^{-1}$	Unimodal
		$\{1+\kappa^*\exp(-\kappa_1-\kappa_2)\}^{-1} < p \leq 1$	Unimodal
(iiia)	$0<\delta<\pi$ $\sin(\delta) > h(\alpha^*)$	$0 \leq p \leq 1$	Unimodal
(iiib)	$0<\delta<\pi$ $\sin(\delta) \leq d(\alpha^*)$	$\{1-\kappa^*/u(\alpha_1)\}^{-1} \leq p \leq \{1-\kappa^*/u(\alpha_2)\}^{-1}$	Bimodal
		$0 \leq p < \{1-\kappa^*/u(\alpha_1)\}^{-1}$	Unimodal
		$\{1-\kappa^*/u(\alpha_2)\}^{-1} < p \leq 1$	Unimodal

Here $\kappa^* = \{\kappa_1 I_0(\kappa_2)/\kappa_2 I_0(\kappa_1)\}$ and $u(\alpha) = \{\sin(\alpha-\delta)/\sin(\alpha)\}\exp(\kappa_2\cos(\alpha-\delta) - \kappa_1\cos(\alpha))$ where $0 < \alpha_1 < \alpha_2 < \delta$ are the two solutions of $d(\alpha) = \sin(\delta)$. Also, $d(\alpha) = \sin(\alpha)\sin(\alpha-\delta)\{\kappa_2\sin(\alpha\delta) - \kappa_1\sin(\alpha)\}$ and α^* maximizes $d(\alpha)$ for $0 < \alpha < \delta$.

Case (i) is the obvious case for unimodal due the mean directions of the mixture being equal. Case (ii) is another boundary case where the mean directions point in opposite directions. Cases of (iii) are the more prevalent cases where the difference in mean directions is between 0 and π.

In either case the range of the mixing proportion p leading to bimodality depends on the mean direction δ, as well as the concentration parameters κ_1 and κ_2. In case (ii), as κ_1 and κ_2 increase, the range of p becomes larger for the bimodal case. For two distributions with mean directions pointing in the opposite directions, the results indicate that an increase in concentration parameters will make the individual component distributions more concentrated around their mean direction in order to maintain bimodality.

Now that we have developed some tools for modeling unimodal as well as bimodal circular data with a common distribution such as the mixCN, in Chapters 3 and 4 we return to the question of testing for presence of a change-point in the number of modes from a time-ordered series of mixCN data.

Chapter 2

Bayes and Preliminary Test Estimators for the Concentration Parameter

In this chapter we investigate how prior knowledge about the mean direction via a preliminary test or a prior distribution will help improve the efficiency in estimating the concentration parameter. Suppose we observe $\alpha_1, \alpha_2, \ldots, \alpha_n$ from a CND with unknown mean direction and concentration parameter. Suppose our goal is to efficiently estimating the concentration parameter, while considering the mean direction as a nuisance parameter. In this study we introduce a preliminary hypothesis test on the value of the mean, and then exploit the knowledge so gained, to improve the estimation of the concentration — the so-called Preliminary Test Estimator (PTE). See for instance (Saleh, 2006) for a

review in the linear statistics setting. We compare such a PTE to the MLE in 1.12 and the MLE of the Bayes-derived likelihood. These comparisons are made through the mean square error (MSE) of the different estimators, obtained through simulation.

2.1 Introduction

(Saleh, 2006), provides an introduction and thorough review on PTEs and Stein-type estimators for various linear models. In statistical inference, the use of prior information on other parameters in a statistical model, usually leads to improved inference on the parameter of interest. Prior information may be (i) known and deterministic which is then incorporated into the model in the form of constraints on the parameter space, leading to a restricted model, or (ii) uncertain and specified in the form of a prior distribution or a null hypothesis. In (ii), choosing certain restricted estimators may be justified when the prior information can be quantified i.e. comes with a specified confidence.

In some statistical models, certain parameters are of primary interest while other parameters may be considered as nuisance parameters. One procedure to mitigate the presence of nuisance parameters is to assess what value(s) such nuisance parameter(s) take, by a preliminary test with a null hypothesis restricting the nuisance parameter values. The null hypothesized value(s) of the nuisance parameter are either used or not, depending on whether the observed preliminary test statistic falls in the acceptance or rejection region of the hypothesis. That is, our final estimator for the parameter of interest is thus a linear combination, conditional on whether the preliminary test

statistic is in the acceptance or rejection region of the test, and is called a Preliminary Test Estimator (PTE).

(Bancroft, 1944, 1964), and (Bancroft, 1965) were among the first to implement the idea of preliminary test estimation (PTE) in an analysis of variance (ANOVA) framework to analyze the effect of the preliminary test on the estimation of variance. The idea goes back to a suggestion in (Snedecor, 1938), which considers testing differences between two means after testing for the equality of variances; then using the usual t-test with the pooled estimate for variance, if the variance test shows equality; otherwise, it falls into the category of Behren's Fisher problem. In these problems it became clear that the performance of the PTE depended heavily on the significance level of the preliminary test. (Han & Bancroft, 1968) were the first to attempt to find an optimum size of significance level for the preliminary test for this two-sample problem.

(Stein *et al.* , 1955), (Stein *et al.* , 1956) followed by (James & Stein, 1961) pointed out a paradoxical situation (the Stein-type estimator) that showed the sample mean vector to be inadmissible under the quadratic loss function for the mean vector of a p-dimensional multivariate normal distribution for $p \geq 3$. This runs counter to the long held belief that the sample mean is the "best" to estimate the population mean and that no other estimation rule is uniformly better than the sample mean. The paradoxical aspect of Stein's work is that it contradicts this idea, in higher dimensions.

All Stein-type estimators involve appropriate test statistics for testing the adequacy of uncertain prior information on the parameter space, which is incorporated into the actual

formulation of the estimator. Stein-type estimators adjust the unrestricted estimator by an amount of the difference between unrestricted and restricted estimators scaled by the adjusted test statistics for the uncertain prior information. Usually, the test statistics are the normalized distance between the unrestricted and restricted estimators and follow a noncentral chi-square or an F-distribution. The risk or the MSE of Stein-type estimators depends on the non-centrality parameter, which represents the distance between the full model and restricted model. The PTE may be considered a precursor of the Stein-type estimator. A simple replacement of the indicator function that we will see in the PTE with a multiple of the test statistic, leads to a Stein-type estimator.

2.2 Other Estimators

The CND is the most widely used circular distribution in circular statistics. It plays as central role as the Normal distribution does in usual 'linear' statistics. Recall that the probability density for a CN random variable, α is:

$$\frac{1}{2\pi I_0(\kappa)} \exp(\kappa \cos(\alpha - \mu)), \quad 0 \leq \alpha < 2\pi \tag{2.1}$$

The mean direction is also referred to as the 'preferred' direction and the concentration parameter can be thought of as the inverse of variance as it is a measure of concentration around the mean direction. A larger value for κ implies that observations are more concentrated around the mean direction, while a value of κ close to 0 implies there may

not be a strongly preferred direction. When estimating the parameters of the distribution it is important to have reliable estimates for both μ and κ parameters. We will now provide the maximum likelihood estimates (MLEs) for the parameter κ in a classical and Bayesian setting.

2.2.1 Maximum Likelihood Estimate for Concentration Parameter

As stated in the introduction, given a random sample $\alpha_1, \ldots, \alpha_n$ from a CND(μ, κ), the MLE for κ when μ in unknown is given by:

$$\hat{\kappa}_{\text{MLE}} \text{ is the solution to: } \frac{1}{n}\sum_{i=1}^{n}\cos(\alpha_i - \bar{\alpha}_0) = \frac{I_1(\kappa)}{I_0(\kappa)} \tag{2.2}$$

When the mean direction μ is known, then the MLE for κ is obtained by substituting μ in place of $\bar{\alpha}_0$ in 2.2. Since the estimation of concentration parameter is of main interest here, we will denote $\hat{\kappa}_{\text{MLE}}$ and $\hat{\kappa}_{\mu}$ the MLEs for κ when sample mean direction is used (if μ unknown), and when the mean direction μ is known, respectively. In both cases the MLEs carry the usual asymptotic properties. Analogous to the case of a linear Normal distribution, $\hat{\kappa}_{\mu}$ is superior (has smaller MSE) than $\hat{\kappa}_{\text{MLE}}$, (Jammalamadaka & SenGupta, 2001).

If the sample comes from a population with population mean μ then by 1.5 and 1.13, the $\text{MSE}(\hat{\kappa}_\mu) < \text{MSE}(\hat{\kappa}_{\text{MLE}})$:

$$\frac{1}{n(1-\bar{V}_0/\hat{\kappa}_\mu-\bar{V}_0^{\,2})} \leq \frac{1}{n(1-\bar{R}/\hat{\kappa}_{\text{MLE}}-\bar{R}^2)} \tag{2.3}$$

where we have inequality if and only if $\mu = \bar{\alpha}_0$. Referring to 1.4 and 1.5 we denote $\bar{R} = R/n$ and $\bar{V}_0 = V_0/n$. This raises the question whether we can do somewhere in between if we have partial information on μ.

2.2.2 MLE for κ when there is a prior on μ

In this semi-Bayesian setting we will place a prior on the nuisance mean direction μ, a convenient choice being a CN:

$$\pi(\mu) = \frac{1}{2\pi I_0(\tau)} \exp(\tau \cos(\mu - \mu_0)), \quad 0 \leq \mu < 2\pi \tag{2.4}$$

where μ_0 and τ are the mean direction and concentration parameters for the prior. The value for τ measures confidence in the prior mean direction μ_0. A larger value of τ makes the prior distribution have higher concentration around μ_0. A value of $\tau = 0$ implies a uniform prior on $[0, 2\pi)$ for μ.

In this context, the parameter μ has a prior distribution, while the parameter κ is an unknown parameter as in the classical setting. The parameter κ is of interest, while μ is

the nuisance parameter. We thus blend together classical and Bayesian methods to get an estimate for κ.

We begin with the usual likelihood given the data $(\alpha_1, \ldots, \alpha_n)$ independent and identically distributed:

$$\mathrm{L}(\mu, \kappa|\underset{\sim}{\alpha}) = \left(\frac{1}{2\pi I_0(\kappa)}\right)^n \exp\left(\kappa \sum_{i=1}^{n} \cos(\alpha_i - \mu)\right), \quad 0 \leq \alpha_i < 2\pi \tag{2.5}$$

Given the prior distribution on μ, we wish to estimate the concentration parameter κ. We derive the likelihood function for κ by first averaging out our prior knowledge on μ. The result is the likelihood for κ given by:

$$\int_0^{2\pi} L(\mu, \kappa|\alpha_1, \ldots, \alpha_n)\pi(\mu)d\mu = L(\kappa|\alpha_1, \ldots, \alpha_n) \tag{2.6}$$

In 2.6, we begin with joint likelihood for the μ and κ which is just the joint density of the data. We then derive marginal distribution for the observations by integrating with respect to μ. After incorporating our prior knowledge on μ and integrating with respect to μ, we obtain a valid likelihood $L(\kappa|\alpha_1, \ldots, \alpha_n)$ for κ which we want to maximize with respect to κ.

$L(\kappa|\alpha_1, \ldots, \alpha_n)$

$$= \int_0^{2\pi} \frac{\exp\left(\kappa(\sum \cos(\alpha_i)\cos(\mu) + \sum \sin(\alpha_i)\sin(\mu))\right) + \tau(\cos(\mu)\cos(\mu_0) + \sin(\mu)\sin(\mu_0)))}{(2\pi)^n (I_o(\kappa))^n 2\pi I_o(\tau)} d\mu \tag{2.7}$$

where setting $\sum \cos(\alpha_i) = R\cos(\bar{\alpha})$ and $\sum \sin(\alpha_i) = R\sin(\bar{\alpha})$ in 2.7 gives,

$$= \int_0^{2\pi} \frac{\exp((\kappa R\cos(\bar{\alpha}) + \tau\cos(\mu_0))\cos(\mu) + (\kappa R\sin(\bar{\alpha}) + \tau\sin(\mu_0))\sin(\mu))}{(2\pi)^n (I_0(\kappa))^n 2\pi I_0(\tau)} d\mu. \tag{2.8}$$

Putting $\kappa R\cos(\bar{\alpha}) + \tau\cos(\mu_0) = \gamma\cos(\alpha^*)$ and $\kappa R\sin(\bar{\alpha}) + \tau\sin(\mu_0) = \gamma\sin(\alpha^*)$, and by the definition of the Bessel function $I_0(x)$, the resulting integral in 2.8 is our likelihood for κ which is given by:

$$L(\kappa|\alpha_1, \ldots, \alpha_n) = \frac{2\pi I_0\left(\sqrt{\kappa^2 R^2 + \tau^2 + 2\kappa R\tau\cos(\bar{\alpha}_0 - \mu_0)}\right)}{(2\pi)^n (I_0(\kappa))^n 2\pi I_0(\tau)}. \tag{2.9}$$

The likelihood is a ratio of Bessel functions as given in 1.7. Given the likelihood, prior distribution on μ, and data we can find the MLE for κ. There is not a simple analytical solution for the MLE, so numerical methods are required for the maximization of 2.9 with respect to κ leading to the semi-Bayesian MLE $\hat{\kappa}_{\text{Bay}}$.

One interesting comparison would be of the frequentist MLE for κ as in 2.2 with the semi-Bayesian MLE obtained from 2.9, using a circular uniform prior distribution on μ in the latter, i.e. setting $\tau = 0$. In some cases, placing uniform priors result in Bayes estimates that are similar to classical MLEs. Using a circular uniform prior distribution on μ in 2.9, we derive the Fisher Information to find the variance of our semi-Bayesian MLE. From 2.9, with a circular uniform prior , the log-likelihood is,

$$\ell = \ln I_0(\kappa R) - n\ln 2\pi - n\ln I_0(\kappa), \tag{2.10}$$

and the semi-Bayesian MLE for κ is the solution to setting $\dot{\ell} = 0$ where $\dot{\ell} = \frac{\partial \ell}{\partial \kappa}$,

$$\dot{\ell} = R\frac{I_1(\kappa R)}{I_0(\kappa R)} - n\frac{I_1(\kappa)}{I_0(\kappa)} \tag{2.11}$$

The solution for MLE κ in this case is found by,

$$\hat{\kappa}_{\text{Bay}} \text{ is the solution to: } \frac{R}{n} = \frac{A(\kappa)}{A(\kappa R)}. \tag{2.12}$$

Immediately we notice a difference when comparing $\hat{\kappa}_{\text{MLE}}$ in 2.2. Taking another derivative of 2.11 we have obtained the Hessian where $\ddot{\ell} = \frac{\partial^2 \ell}{\partial \kappa^2}$,

$$\ddot{\ell} = R^2\frac{I_0(\kappa R)I_2(\kappa R) - I_1^2(\kappa R)}{I_0^2(\kappa R)} + n\frac{I_1^2(\kappa) - I_0(\kappa)I_2(\kappa)}{I_0^2(\kappa)} \tag{2.13}$$

$$= R^2\left(\frac{I_2(\kappa)}{I_0(\kappa)} - A^2(\kappa R)\right) + n\left(A^2(\kappa) - \frac{I_2(\kappa)}{I_0(\kappa)}\right) \tag{2.14}$$

$$= \frac{I_2(\kappa)}{I_0(\kappa)}(R^2 - n) + nA^2(\kappa) - R^2A^2(\kappa R). \tag{2.15}$$

Then the Fisher Information (I) is given by,

$$I = \frac{I_2(\kappa)}{I_0(\kappa)}(n - R^2) - nA^2(\kappa) + R^2A^2(\kappa R), \tag{2.16}$$

where substituting the semi-Bayes MLE, the asymptotic variance (V) of $\hat{\kappa}_{\text{Bay}}$:

$$V = \left(\frac{I_2(\hat{\kappa}_{\text{Bay}})}{I_0(\hat{\kappa}_{\text{Bay}})}(R^2 - n) + nA^2(\hat{\kappa}_{\text{Bay}}) - R^2A^2(\hat{\kappa}_{\text{Bay}}R)\right)^{-1} \tag{2.17}$$

Therefore the asymptotic variance of the MLE can be found using 2.17. Next, we can compare the two MLE's via their respective large-sample confidence intervals. The $(1-\gamma)\times 100\%$ confidence interval for κ is given by,

$$\left(\hat{\kappa}_{\text{MLE}} \pm Z_{(\gamma/2)}\sqrt{\frac{1}{n(1-\bar{R}/\hat{\kappa}_{\text{MLE}}-\bar{R}^2)}}\right)$$

$$\left(\hat{\kappa}_{\text{Bay}} \pm Z_{\gamma/2}\sqrt{\left(\frac{I_2(\hat{\kappa}_{\text{Bay}})}{I_0(\hat{\kappa}_{\text{Bay}})}(R^2-n)+nA^2(\hat{\kappa}_{\text{Bay}})-R^2A^2(\hat{\kappa}_{\text{Bay}}R)\right)^{-1}}\right)$$

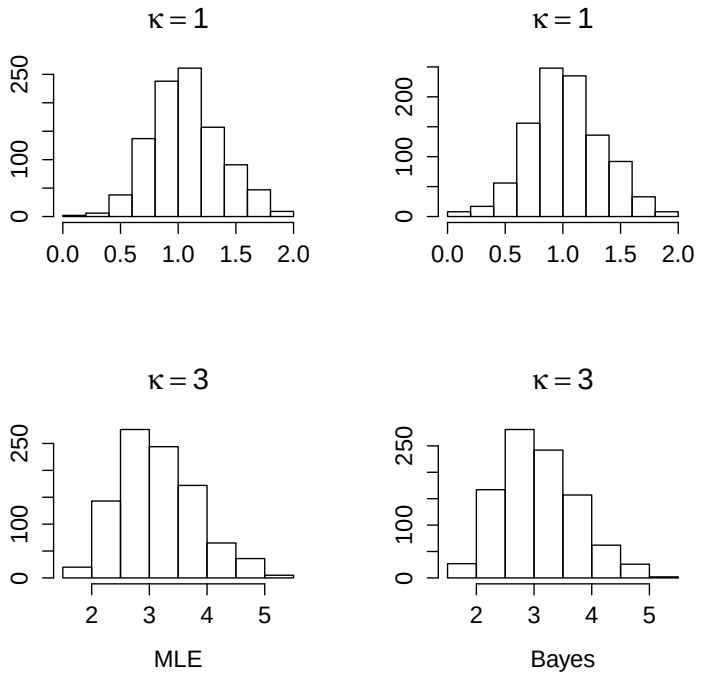

Figure 2.1: Histograms of $\hat{\kappa}_{\text{MLE}}$ (MLE) and $\hat{\kappa}_{\text{Bay}}$ (Bayes) with circular uniform prior for 1000 simulations from CND(μ, κ).

Figure 2.1 displays histograms for $\hat{\kappa}_{\text{MLE}}$ and $\hat{\kappa}_{\text{Bay}}$ based on 1000 simulations from the CND with $\kappa = 1, 3$, using sample size of $n = 30$. In each κ setting the histograms of estimated values are nearly identical for $\hat{\kappa}_{\text{MLE}}$ and $\hat{\kappa}_{\text{Bay}}$.

2.3 Preliminary Test Estimators

A preliminary test estimator (PTE) is a method of estimation that introduces sample-based prior information via a hypothesis test on the nuisance parameter to aid in estimating the parameter of interest (Saleh, 2006). If we fail to reject the null, then we use an estimator evaluated using the null hypothesis value. If we reject the null hypothesis, we use an estimator based directly on the sample, the usual MLE. The parameter value in the null hypothesis represents our prior knowledge. The idea is when the true parameter value is in or near the null hypothesis value, the PTE will provide a better estimator in terms of mean squared error (MSE), or any other risk function.

We observe data from a CND with unknown mean direction and concentration parameter. We are interested in estimating the concentration parameter, with the mean direction being a nuisance parameter. Our preliminary test has null hypothesis of mean direction equal to a pre-specified direction, versus a two-sided alternative.

Our PTE for the concentration performs better than the usual MLE and Bayesian estimates for the parameter. The result is similar to the linear case where we have a normal distribution with unknown mean and variance, (Ohtani, 1991). This methodology can be used to improve the estimation accuracy in many existing applications since the CND is one of the most commonly used distributions in circular statistics.

2.3.1 Test for assumed Mean Direction

Suppose we have observations $\alpha_1, \ldots, \alpha_n$ from a CND with both mean direction and concentration parameter unknown. We want to test:

$$H_0 : \mu = \mu_0 \text{ vs. } H_1 : \mu \neq \mu_0 \tag{2.18}$$

In the linear case with data from a Normal distribution, this is parallel to the standard Student's t-test. In (Jammalamadaka & SenGupta, 2001), the Likelihood Ratio Test (LRT) is based on the test statistic:

$$V_0 = \sum_{i=1}^{n} \cos(\alpha_i - \mu_0) = R\cos(\bar{\alpha}_0 - \mu_0) \text{ or } \frac{V_0}{R} = \cos(\bar{\alpha}_0 - \mu_0) \tag{2.19}$$

where we reject the null hypothesis for small values of the test statistic. Note the distribution for V_0 and $\frac{V_0}{R}$ depend on the nuisance parameter κ. However, the exact conditional test for the mean direction of the CN can be obtained by using the conditional distribution of $R|V_0$, which is independent of κ. V_0 is the length of the projection of sample resultant vector, R, towards the null hypothesized mean direction, $\mu_0 = (\cos(\mu_0), \sin(\mu_0))$. In the conditional test we reject null if V_0 is too small for a given R, or equivalently, we reject the null if R is too large for a given V_0.

To illustrate the geometry of the test, suppose we have polar vector given by the null hypothesis, $(\cos(\mu_0), \sin(\mu_0))$. Next, we have n observations and we calculate the length of projection, c, of the sample resultant vector on the polar vector. Conditioning on the value of c, we find the probability of observing our sample resultant vector, R, and larger values when the null direction is true, conditional on the observed value of $V_0 = v$.

The space consists of sample resultant vectors that have projection length, c, on the polar vector. Suppose R_1 and R_2 are two resultant vector with equal projection length and $R_1 > R_2$. Then the direction of R_1 is further away from μ_0, than R_2's direction.

For significance level γ, we find the rejection region via the exact conditional distribution of $R|V_0$. That is, r_0 is the solution to the equation that satisfies:

$$\mathbb{P}(R > r_0|V_0 = v) = \alpha. \tag{2.20}$$

As shown in (Jammalamadaka & SenGupta, 2001), this critical point r_0 is the solution to:

$$\int_{r_0}^{n} f(r|v)dr = \int_{r_0}^{n} \frac{r\Psi_n(r)}{\sqrt{r^2 - v^2}f_0(v)\pi}dr = \alpha. \tag{2.21}$$

where we solve for r_0, for a given v and n. Equations for $\Psi_n(r)$ and $f_0(v)$ can be found in (Jammalamadaka & SenGupta, 2001). There is no analytical solution for r_0 in this case, and (Stephens, 1962) provides a table of rejection regions for various values of V_0. To simplify our hypothesis test we use results in (Upton, 1986), where approximate confidence intervals for the mean direction are provided. Our test statistic derived from the approximate LRT is broken into two cases:

(i) For $\bar{R} \leq 0.9$, we reject H_0 if:

$$R^2 > V_0^2 + \frac{1}{4n}(2n^2 - V_0^2)Z_\gamma \tag{2.22}$$

where $\bar{R} = \frac{R}{n}$ and Z_γ is the upper quantile of the standard Normal distribution.

(ii) For $\bar{R} > 0.9$, we reject H_0 if:

$$n \log\left(\frac{n^2 - V_0^2}{n^2 - R^2}\right) > \chi^2_{1,\gamma} \tag{2.23}$$

These approximations hold well for even small sample sizes when the concentration is high.

2.3.2 The PTE for the concentration parameter

Now we introduce our PTE for estimating the concentration parameter, where the mean direction is a nuisance parameter. Given observations, $\alpha_1, \ldots, \alpha_n$, with unknown mean direction and concentration parameter we test our null hypothesized mean direction via the aforementioned hypothesis test. Our PTE is given by:

(i) For $\bar{R} \leq 0.9$,

$$\hat{\kappa}_{\text{PTE}} = \hat{\kappa}_{\text{MLE}} \mathbb{1}(Z > Z_\gamma) + \hat{\kappa}_{\mu_0} \mathbb{1}(Z < Z_\gamma) \tag{2.24}$$

where Z is found by using 2.22 and solving for Z_γ.

(ii) For $\bar{R} > 0.9$,

$$\hat{\kappa}_{\text{PTE}} = \hat{\kappa}_{\text{MLE}} \mathbb{1}(\chi^2 > \chi^2_{1,\gamma}) + \hat{\kappa}_{\mu_0} \mathbb{1}(\chi^2 < \chi^2_{1,\gamma}) \tag{2.25}$$

where χ^2 is found by using 2.23.

we break the estimator into the two cases according to our hypothesis test. The PTE in either case selects only one of the two estimators according to the result of the hypothesis test. The performance of the PTE depends on the level of the test and the proximity of the true mean direction to the null hypothesized value. We measure performance in terms of mean squared error (MSE) of our estimator over different significance levels γ, and different true differences between the mean directions $\delta = \mu - \mu_0$.

In Figure 2.2, we observe the simulation-based MSE of the PTE and MLE for the concentration parameter. We perform 1000 simulations of $(\alpha_1, \ldots, \alpha_{20}) \sim CN(\delta_j, \kappa)$, for $j = 1, \ldots, 50$. Here $\delta_1, \ldots, \delta_{50}$ represent the 50 equally spaced points between 0 and π. For each δ_j, we record

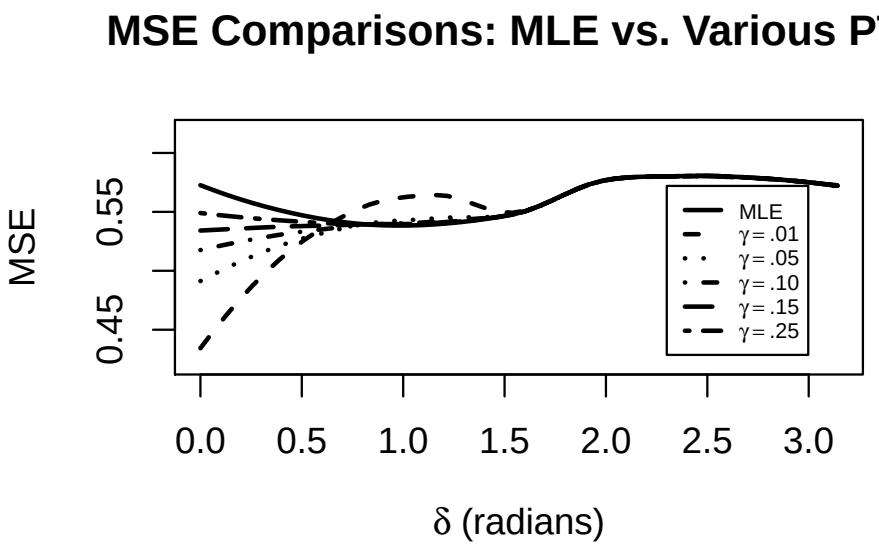

Figure 2.2: Simulation-based MSE of MLE and PTE for different significance levels γ.

the MSE. Each line represents MSE of an estimator over values of $\delta = \mu - \mu_0$, where δ represents the true difference between the population mean direction and the null hypothesized mean direction.

For significance levels $\gamma = .10, .15, .25$, the PTE performs at least as good as the MLE, and performs better when the true mean direction is closer to the null hypothesized value. For larger significance levels the test requires less evidence to reject the null hypothesis, and when we reject the null the PTE is equivalent to the MLE, $\hat{\kappa}_{\text{MLE}}$. In Figure 2.2, we observe that as the significance level increases the PTE is more likely to use $\hat{\kappa}_{\text{MLE}}$ for smaller values of δ. To show the vast improvement in our PTE, we examine the mean-square relative efficiency (MRE) of the 2 estimators PTE and MLE, defined by

$$e\left(\hat{\kappa}_{\text{PTE}}, \hat{\kappa}_{\text{MLE}}\right) = \frac{\text{MSE}(\hat{\kappa}_{\text{MLE}})}{\text{MSE}(\hat{\kappa}_{\text{PTE}})} \tag{2.26}$$

Values larger than unity imply that the PTE performs better than the MLE. In Figure 2.3, we have the MRE of the MLE and PTE with $\gamma = 0.01$ across all values of δ. The relative efficiency is greater than 1 for all δ less than approximately 0.65 radians. In this example, the PTE can reduce the MSE by 20% when the true difference in mean directions is small. For $.65 < \delta < 1.5$, the MRE is less than 1 implying the MLE has the smaller MLE. This is due to our preliminary test failing to reject the null hypothesis. For $\delta > 1.5$, the preliminary test will almost always reject the null hypothesis value and the PTE will be the same as the MLE resulting in the MRE being equal to one.

In Equation 2.3, the MRE is maximum for $\delta = 0$, and when the PTE will almost always reject the null for large enough δ the MRE is equal to one. For $0 < \delta < \pi$ the PTE may reject or fail to reject the null hypothesis depending on the sample observed. In the case it fails to reject, $V_0 = \sum_{i=1}^{n} \cos(\alpha_i - \mu_0)$ is no longer minimized at μ_0 since μ is the population mean (Recall $\delta = \mu - \mu_0$). Therefore $V_0 < R$ if μ_0 is closer to μ than $\bar{\alpha}_0$, and $V_0 > R$ if $\bar{\alpha}_0$ is closer to μ than μ_0. If the latter case appears more often than the former case for some intermediate values of δ, then the MRE will be less than 1.

In Figure 2.2, we compare the MLE and PTE with $\gamma = 0.01$ from Figure 2.2 .The PTE with $\gamma = 0.01$ has the best results for smaller values of δ, but could perform worse than the MLE for intermediate values of δ. PTEs with $\gamma = 0.1, 0.15, 0.25$ perform at least as good as the MLE. Now, we illustrate other possibilities that can occur and the performance of the PTE.

In Figure 2.4, we simulate from four different realities and examine the performance of our PTE for the same significance levels as used in Figure 2.2. Note that the lines have same labels as in Figure 2.2. For each plot we have simulated-based MSEs for each line. We perform 1000 simulations of $(\alpha_1, \ldots, \alpha_n) \sim CN(\delta_j, \kappa)$, for $j = 1, \ldots, 50$. Here $\delta_1, \ldots, \delta_{50}$ represent the 50

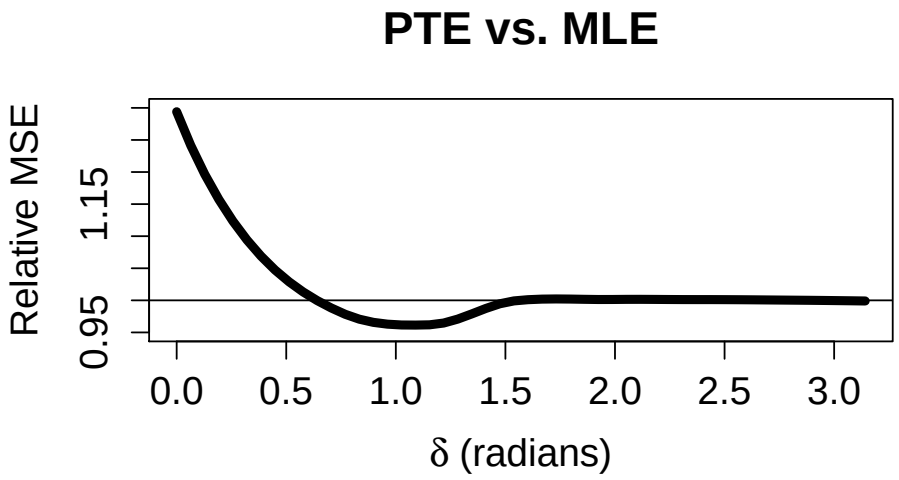

Figure 2.3: Mean-Squared Error Relative Efficiency of MLE and PTE with $\gamma = 0.01$.

equally spaced points between 0 and π. For each δ_j, we record the MSE which creates our MSE curve over δ for each scenario.

First examine that in all scenarios, the PTE with significance level $\gamma = 0.01$ performs the best when the true difference in mean direction is null or small. In the top-left plot we have $n = 50$ simulated observations from $CN(\delta, \kappa = 0.5)$; top-right plot we have $n = 10$ simulated observations from $CN(\delta, \kappa = 0.5)$; bottom-left plot we have $n = 40$ simulated observations from $CN(\delta, \kappa = 3)$; bottom-right plot we have $n = 10$ simulated observations from $CN(\delta, \kappa = 2.5)$.

In the top-right plot all of the PTE's in this simulation performed uniformly better (over δ) than $\hat{\kappa}_{\text{MLE}}$. In the remaining three plots there are values of δ where the $\hat{\kappa}_{\text{MLE}}$ has better performance. This occurs when our preliminary test fails to reject the null hypothesis for intermediate values of δ. The difference becomes more obvious when we have a large sample size and the value of κ is small as in the top-left plot. Here the PTE's MSE increases for intermediate values of δ, for relatively smaller significance levels. There is a similar pattern

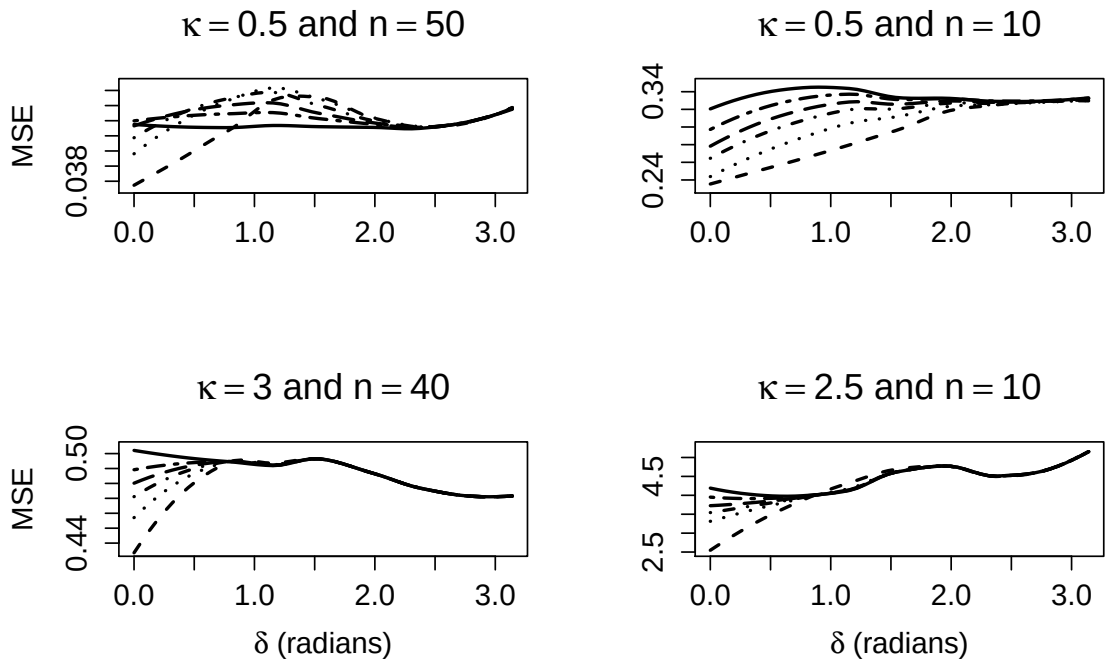

Figure 2.4: Simulation-based Comparison of PTE Performances for Sample Sizes $n = 10, 40, 50$ and Concentration Parameters $\kappa \in \{0.5, 2.5, 3\}$. Lines are labeled as in Figure 2.2.

in the bottom two plots. This pattern is to be expected, since smaller significance level will require more evidence to reject the null hypothesis of the preliminary test.

In applications, the values of δ and κ are unknown. So how do we select the optimal significance level given n observations from $CN(\mu, \kappa)$? Following the work of (Saleh, 2006), we create tables to find a PTE with minimum and maximum MREs.

Tables were constructed through simulations. Given a sample size n and value for κ, we generate values from a $CN(\delta, \kappa)$ distribution to estimate the MRE over a grid of α and δ values, where $0 \leq \delta \leq \pi$. For each α, we compute the maximum MRE, $E_{\max}$, minimum MRE, $E_{\min}$, over all δ, and record the δ where $E_{\min}$ is located, $\Delta_{\min}$. For almost all cases the location of the maximum MRE is located at $\delta = 0$ and the function MRE(δ) is monotone decreasing from

$\delta = 0$ to $\delta = E_{\min}$. For values $\delta > E_{\min}$, the function MRE(δ) increases back to unity since the PTE will reject the null hypothesized values for larger δ. We then repeated this procedure for different parameter values for κ.

The mean resultant vector is the normalized length of 1.1 since $0 < \bar{R} < 1$ and is a measure of concentration for a sample of observations. A value close to 1 implies high concentration and a value close to 0 implies little to no concentration around any single direction. This estimate does not depend on the knowledge of κ or of the mean μ of the distribution. For the CND, there is a one-to-one correspondence between statistic $\bar{R}$ and the concentration parameter κ. Given a sample size n and κ, we observe the average $\bar{R}$ over our simulations and use the average as an indication of strength of concentration. In practice, we advise the user to find the sample observed $\bar{R}$ of the n observations, and then use the column of the table with the nearest $\bar{R}$ value.

In Table 2.1, we provide a list of potential PTEs for $n = 5$. The rows list various significance levels γ for the PTE ranging from 1% to 50%. The columns list the different observed values for $\bar{R}$. Suppose we have a sample size of 5 observations and observe $\bar{R}$ close to 0.751. Following the procedure in (Saleh, 2006), we then decide the minimum MRE preferred is $E_{\min} = 0.977$. Then using the Table 2.1, the optimal PTE corresponds to using $\alpha = 0.15$. In the appendix we provide tables for various sample sizes, where the tables require only knowledge of sample size, $\bar{R}$, and the predetermined $E_{\min}$.

2.3.3 Comparison of the PTE and Bayes Estimators

Both the PTE and Bayes estimators in 2.12, use prior information on the the mean direction μ to aid in estimation of the concentration parameter. A smaller significance level for the PTE

requires stronger evidence to reject the null hypothesized value μ_0. A smaller significance level may be chosen to coincide with a stronger belief in the mean direction μ_0. In the previously mentioned Bayesian setting of this chapter, a larger value for the concentration parameter τ focuses our prior distribution around the mean direction μ_0. A larger value in parameter τ represents a stronger belief in prior mean direction μ_0.

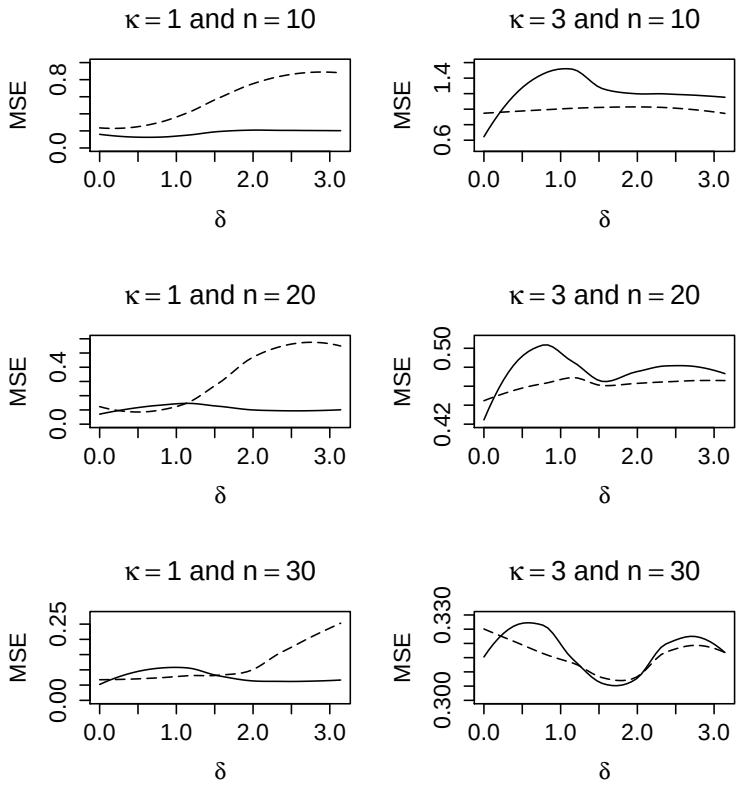

Figure 2.5: MSE of PTE and Bayes Estimators over δ: $\hat{\kappa}_{\text{PTE}}$ (—) and $\hat{\kappa}_{\text{Bay}}$ (- - -)

In Figue 2.5 we make a comparison of the MSE of our PTE with significance level of 1% with the Bayes estimator with CN prior centered around the null hypothesis value μ_0 and with $\tau = 4$. We plot the MSE curve of each estimator over values of δ.

In each plot, the solid line is MSE curve for the PTE and the dashed line is the MSE curve for the Bayes estimator. For $\kappa = 1$, $\hat{\kappa}_{\text{PTE}}$ performs better overall for all sample sizes. For $n = 10$, $\hat{\kappa}_{\text{PTE}}$ performs uniformly better than $\hat{\kappa}_{\text{bay}}$. For $n = 20$ and $n = 30$ the estimators have similar performances for small values of δ, but the MSE for $\hat{\kappa}_{\text{bay}}$ is much larger for large values of δ.

If $\kappa = 3$, we have different results when comparing the MSEs. In all sample sizes of $n = 10, 20, 30$, the MSE of $\hat{\kappa}_{\text{PTE}}$ is best for small values of δ. Also, for all sample sizes, $\hat{\kappa}_{\text{bay}}$ has the smaller MSE for the larger values of δ. In this case for large value of κ, $\hat{\kappa}_{\text{bay}}$ would be the preferred estimator since the performance is better overall.

In reality we do not know the value of κ, so need a data driven way to select $\hat{\kappa}_{\text{bay}}$ versus $\hat{\kappa}_{\text{MLE}}$. If we suspect a high conentration then we suggest to use $\hat{\kappa}_{\text{bay}}$, and for suspect a weak concentration then use $\hat{\kappa}_{\text{MLE}}$. If given a sample size n, go to the corresponding PTE table for the same sample size. In the table, go to the 7th column which gives the expected $\bar{R}$ under $\kappa = 3$ simulations. From your observed sample of size n, calculate $\bar{R}$ in column 7, and compare to the value from the PTE table. If less than the PTE table value, then use $\hat{\kappa}_{\text{MLE}}$, otherwise use $\hat{\kappa}_{\text{bay}}$.

2.4 PTE Tables

Here we have constructed PTE tables for sample sizes of $n = 5, 10, 15, 20, 30, 40$ and 50. The rows list several significance levels ranging between 1% to 50% for the PTE, while the columns are organized by sample $\bar{R}$ values. For each pair (α, $\bar{R}$), we list $E_{\max}$, $E_{\min}$, and $\Delta_{\min}$.

Table 2.1: $n = 5$: Maximum and Minimum Guaranteed Efficiencies for the PTE

		$\bar{R}$									
γ		0.403	0.400	0.402	0.560	0.668	0.751	0.802	0.845	0.890	0.916
0.01	$E_{\max}$	0.997	0.997	1.044	1.855	3.126	3.28	3.265	4.788	9.047	12.751
	$E_{\min}$	0.992	0.984	1.007	1.002	0.972	0.969	0.954	0.917	0.894	0.871
	$\Delta_{\min}$	2.244	0	3.142	2.885	2.629	2.436	1.667	1.603	1.282	1.218
0.02	$E_{\max}$	1.031	1.036	1.202	2.313	2.483	2.554	3.513	4.492	5.608	5.43
	$E_{\min}$	1.022	1.027	1.027	0.998	0.978	0.977	0.938	0.938	0.919	0.919
	$\Delta_{\min}$	3.142	2.052	3.142	2.821	2.5	1.795	1.603	1.539	1.218	1.218
0.05	$E_{\max}$	1.137	1.183	1.592	1.939	1.885	2.352	2.677	2.676	2.679	2.512
	$E_{\min}$	1.131	1.128	1.052	0.993	0.986	0.958	0.962	0.956	0.961	0.963
	$\Delta_{\min}$	0	3.142	3.142	2.629	2.116	1.603	1.603	1.282	1.218	1.218
0.1	$E_{\max}$	1.165	1.217	1.515	1.56	1.601	1.884	1.894	1.776	1.785	1.656
	$E_{\min}$	1.157	1.142	1.042	0.993	0.972	0.968	0.975	0.976	0.98	0.982
	$\Delta_{\min}$	0	3.142	3.142	2.5	1.667	1.603	1.282	1.218	1.218	1.218
0.15	$E_{\max}$	1.124	1.157	1.353	1.368	1.39	1.621	1.538	1.512	1.459	1.383
	$E_{\min}$	1.113	1.107	1.034	0.994	0.974	0.977	0.983	0.986	0.988	0.989
	$\Delta_{\min}$	0	3.142	3.013	2.436	1.603	1.346	1.282	1.218	1.218	1.218
0.2	$E_{\max}$	1.099	1.119	1.245	1.238	1.27	1.449	1.346	1.34	1.301	1.261
	$E_{\min}$	1.082	1.071	1.024	0.993	0.98	0.983	0.989	0.991	0.992	0.993
	$\Delta_{\min}$	0	3.142	2.949	2.052	1.603	1.346	1.282	1.218	1.218	1.218
0.25	$E_{\max}$	1.075	1.089	1.176	1.148	1.184	1.327	1.245	1.232	1.213	1.19
	$E_{\min}$	1.065	1.051	1.018	0.991	0.985	0.989	0.992	0.993	0.995	0.995
	$\Delta_{\min}$	0	3.142	2.885	1.731	1.539	1.282	1.282	1.218	1.218	1.218
0.3	$E_{\max}$	1.057	1.064	1.125	1.097	1.132	1.234	1.18	1.152	1.168	1.137
	$E_{\min}$	1.049	1.039	1.011	0.989	0.987	0.991	0.994	0.996	0.996	0.996
	$\Delta_{\min}$	0	3.142	3.142	1.667	1.346	1.282	1.282	1.218	1.218	1.218
0.35	$E_{\max}$	1.041	1.046	1.096	1.067	1.09	1.173	1.136	1.114	1.12	1.102
	$E_{\min}$	1.035	1.029	1.008	0.988	0.99	0.993	0.996	0.996	0.997	0.997
	$\Delta_{\min}$	0	3.142	3.142	1.603	1.346	1.282	1.218	1.218	1.218	1.218
0.4	$E_{\max}$	1.03	1.033	1.069	1.048	1.059	1.134	1.099	1.077	1.094	1.078
	$E_{\min}$	1.026	1.022	1.006	0.989	0.992	0.996	0.997	0.997	0.998	0.998
	$\Delta_{\min}$	0	3.142	3.142	1.603	1.282	1.282	1.282	1.218	1.218	1.218
0.45	$E_{\max}$	1.022	1.024	1.047	1.032	1.038	1.098	1.07	1.053	1.062	1.059
	$E_{\min}$	1.016	1.017	1.004	0.991	0.994	0.997	0.998	0.998	0.998	0.998
	$\Delta_{\min}$	0	3.077	3.142	1.603	1.346	1.282	1.218	1.218	1.218	1.218
0.5	$E_{\max}$	1.016	1.018	1.035	1.017	1.031	1.062	1.054	1.037	1.049	1.039
	$E_{\min}$	1.012	1.012	1.002	0.992	0.995	0.998	0.999	0.999	0.999	0.999
	$\Delta_{\min}$	0	3.142	3.142	1.539	1.282	1.282	1.282	1.218	1.218	1.218

Table 2.2: $n = 10$: Maximum and Minimum Guaranteed Efficiencies for the PTE

		$\bar{R}$									
γ		0.278	0.279	0.286	0.498	0.627	0.722	0.784	0.827	0.876	0.904
0.01	$E_{\max}$	1.009	1.017	1.338	1.632	1.552	1.751	1.883	1.959	1.758	1.616
	$E_{\min}$	1.002	1.008	1.009	0.982	0.746	0.663	0.714	0.799	0.876	0.897
	$\Delta_{\min}$	0	2.18	3.077	2.18	1.218	1.154	1.154	1.09	0.898	0.833
0.02	$E_{\max}$	1.05	1.066	1.594	1.408	1.416	1.546	1.623	1.722	1.587	1.481
	$E_{\min}$	1.042	1.048	1.015	0.963	0.74	0.736	0.812	0.89	0.938	0.948
	$\Delta_{\min}$	1.988	2.436	3.077	1.731	1.218	1.154	1.09	1.09	0.898	0.833
0.05	$E_{\max}$	1.134	1.199	1.744	1.273	1.21	1.318	1.405	1.506	1.414	1.355
	$E_{\min}$	1.128	1.127	1.014	0.922	0.788	0.839	0.924	0.964	0.978	0.98
	$\Delta_{\min}$	1.731	3.142	2.821	1.539	1.154	1.026	1.154	1.154	0.962	0.898
0.1	$E_{\max}$	1.147	1.227	1.428	1.152	1.094	1.173	1.269	1.39	1.297	1.22
	$E_{\min}$	1.134	1.126	1.008	0.91	0.843	0.923	0.971	0.983	0.99	0.991
	$\Delta_{\min}$	0	2.757	2.757	1.282	1.026	0.898	1.154	1.218	0.962	0.898
0.15	$E_{\max}$	1.107	1.16	1.294	1.077	1.049	1.118	1.197	1.314	1.2	1.16
	$E_{\min}$	1.101	1.091	1.003	0.911	0.885	0.96	0.985	0.989	0.995	0.994
	$\Delta_{\min}$	0	2.949	2.693	1.218	0.962	0.898	1.154	1.218	1.218	0.898
0.2	$E_{\max}$	1.086	1.12	1.209	1.039	1.026	1.09	1.14	1.229	1.141	1.122
	$E_{\min}$	1.079	1.071	1	0.923	0.922	0.975	0.992	0.992	0.997	0.995
	$\Delta_{\min}$	0	2.757	2.629	1.154	0.898	0.898	1.218	1.218	1.218	0.898
0.25	$E_{\max}$	1.064	1.093	1.151	1.014	1.01	1.071	1.104	1.185	1.099	1.094
	$E_{\min}$	1.062	1.054	0.999	0.93	0.95	0.983	0.995	0.994	0.998	0.996
	$\Delta_{\min}$	3.142	2.757	2.565	1.154	0.833	0.898	1.154	1.218	1.218	0.898
0.3	$E_{\max}$	1.049	1.073	1.114	0.997	1.001	1.053	1.083	1.147	1.072	1.071
	$E_{\min}$	1.047	1.039	0.998	0.938	0.968	0.989	0.997	0.995	0.998	0.997
	$\Delta_{\min}$	3.142	2.757	2.5	1.09	0.769	0.898	1.154	1.218	0.962	0.898
0.35	$E_{\max}$	1.035	1.054	1.082	0.997	1.001	1.041	1.064	1.11	1.053	1.051
	$E_{\min}$	1.032	1.028	0.997	0.943	0.977	0.994	0.998	0.997	0.999	0.998
	$\Delta_{\min}$	3.142	2.821	2.436	1.026	0.769	0.898	1.218	1.218	0.962	0.898
0.4	$E_{\max}$	1.026	1.036	1.057	0.996	1.001	1.027	1.047	1.081	1.04	1.038
	$E_{\min}$	1.023	1.02	0.995	0.95	0.985	0.997	0.998	0.998	0.999	0.998
	$\Delta_{\min}$	3.142	2.693	2.18	0.962	0.641	0.898	1.218	1.218	0.962	0.898
0.45	$E_{\max}$	1.019	1.026	1.037	0.997	1	1.02	1.036	1.052	1.024	1.03
	$E_{\min}$	1.017	1.015	0.994	0.959	0.99	0.998	0.999	0.998	0.999	0.999
	$\Delta_{\min}$	3.142	2.436	2.116	0.769	0.577	0.962	1.218	1.218	0.962	0.898
0.5	$E_{\max}$	1.016	1.021	1.021	0.998	1	1.014	1.023	1.04	1.017	1.021
	$E_{\min}$	1.014	1.012	0.994	0.97	0.993	0.999	0.999	0.999	1	0.999
	$\Delta_{\min}$	1.154	2.436	2.052	0.705	0.513	0.962	1.218	1.218	0.962	0.898

Table 2.3: $n = 15$: Maximum and Minimum Guaranteed Efficiencies for the PTE

						$\bar{R}$					
γ		0.226	0.227	0.232	0.483	0.620	0.717	0.777	0.820	0.870	0.900
0.01	$E_{\max}$	1.016	1.021	1.534	1.46	1.448	1.315	1.241	1.262	1.139	1.075
	$E_{\min}$	1.011	1.012	1.002	0.832	0.614	0.747	0.868	0.906	0.906	0.905
	$\Delta_{\min}$	2.693	0	2.885	1.346	1.09	0.833	0.833	0.833	0.769	0.769
0.02	$E_{\max}$	1.054	1.079	1.701	1.371	1.281	1.159	1.185	1.241	1.128	1.072
	$E_{\min}$	1.048	1.044	1.002	0.796	0.685	0.847	0.924	0.955	0.949	0.943
	$\Delta_{\min}$	3.142	3.142	2.821	1.218	1.026	0.833	0.833	0.833	0.769	0.769
0.05	$E_{\max}$	1.132	1.212	1.467	1.185	1.097	1.074	1.126	1.224	1.113	1.079
	$E_{\min}$	1.12	1.103	0.997	0.784	0.822	0.943	0.977	0.989	0.981	0.98
	$\Delta_{\min}$	3.142	3.142	2.693	1.154	0.833	0.833	0.898	1.026	0.833	0.769
0.1	$E_{\max}$	1.147	1.234	1.298	1.049	1.025	1.051	1.085	1.168	1.094	1.075
	$E_{\min}$	1.132	1.098	0.996	0.814	0.925	0.978	0.991	0.994	0.993	0.992
	$\Delta_{\min}$	3.142	3.142	2.629	1.09	0.769	0.833	0.898	1.218	0.833	0.833
0.15	$E_{\max}$	1.107	1.176	1.203	1.009	1.013	1.034	1.062	1.125	1.076	1.06
	$E_{\min}$	1.097	1.074	0.995	0.841	0.96	0.987	0.995	0.996	0.996	0.995
	$\Delta_{\min}$	3.142	3.142	2.565	0.962	0.769	0.833	0.898	1.218	0.898	0.833
0.2	$E_{\max}$	1.085	1.138	1.146	1.004	1.011	1.027	1.042	1.093	1.064	1.051
	$E_{\min}$	1.077	1.058	0.991	0.876	0.978	0.993	0.997	0.998	0.997	0.997
	$\Delta_{\min}$	2.244	3.142	2.052	0.898	0.769	0.833	0.898	1.218	0.898	0.898
0.25	$E_{\max}$	1.065	1.1	1.101	1.003	1.007	1.019	1.029	1.068	1.056	1.035
	$E_{\min}$	1.059	1.044	0.99	0.912	0.987	0.996	0.998	0.998	0.998	0.998
	$\Delta_{\min}$	2.436	3.142	1.731	0.833	0.769	0.833	0.898	0.962	0.898	0.833
0.3	$E_{\max}$	1.049	1.071	1.069	1.002	1.002	1.018	1.022	1.047	1.041	1.032
	$E_{\min}$	1.045	1.033	0.986	0.938	0.991	0.997	0.999	0.998	0.998	0.998
	$\Delta_{\min}$	2.436	3.142	1.603	0.769	0.769	0.833	0.898	0.898	0.898	0.898
0.35	$E_{\max}$	1.035	1.052	1.041	1.001	1.001	1.008	1.017	1.034	1.033	1.024
	$E_{\min}$	1.032	1.023	0.983	0.956	0.995	0.999	0.999	0.999	0.999	0.999
	$\Delta_{\min}$	0	3.142	1.603	0.641	0.769	0.833	0.962	0.962	0.898	0.898
0.4	$E_{\max}$	1.026	1.039	1.027	1.001	1	1.004	1.014	1.03	1.023	1.017
	$E_{\min}$	1.022	1.016	0.982	0.969	0.996	0.999	0.999	0.999	0.999	0.999
	$\Delta_{\min}$	0	3.142	1.539	0.577	0.769	0.833	0.898	0.898	0.898	0.898
0.45	$E_{\max}$	1.019	1.029	1.016	1.001	1	1.003	1.011	1.022	1.015	1.012
	$E_{\min}$	1.016	1.012	0.983	0.978	0.998	0.999	1	0.999	0.999	0.999
	$\Delta_{\min}$	0	2.885	1.346	0.385	0.577	0.769	0.962	0.962	0.898	0.898
0.5	$E_{\max}$	1.015	1.02	1.005	1.002	1	1.002	1.008	1.017	1.012	1.01
	$E_{\min}$	1.012	1.01	0.984	0.987	0.997	0.999	1	0.999	1	0.999
	$\Delta_{\min}$	0	2.885	1.346	0.256	0	0.769	0.962	0.898	0.898	0.898

Table 2.4: $n = 20$: Maximum and Minimum Guaranteed Efficiencies for the PTE

						$\bar{R}$					
γ		0.196	0.198	0.202	0.471	0.617	0.710	0.779	0.818	0.871	0.900
0.01	$E_{\max}$	1.016	1.027	1.676	1.38	1.244	1.075	1.081	1.116	1.033	1.018
	$E_{\min}$	1.008	1.015	0.995	0.688	0.695	0.869	0.928	0.951	0.936	0.915
	$\Delta_{\min}$	0	2.949	2.757	1.154	0.833	0.769	0.769	0.833	0.705	0.449
0.02	$E_{\max}$	1.055	1.103	1.555	1.231	1.105	1.049	1.07	1.118	1.048	1.012
	$E_{\min}$	1.043	1.046	0.993	0.689	0.807	0.921	0.961	0.98	0.968	0.95
	$\Delta_{\min}$	0	3.142	2.629	1.09	0.769	0.769	0.769	0.833	0.769	0.449
0.05	$E_{\max}$	1.127	1.228	1.313	1.08	1.02	1.031	1.06	1.111	1.061	1.005
	$E_{\min}$	1.116	1.102	0.991	0.743	0.924	0.968	0.988	0.995	0.988	0.983
	$\Delta_{\min}$	0	3.142	2.5	0.962	0.769	0.769	0.833	0.898	0.833	0.513
0.1	$E_{\max}$	1.138	1.246	1.202	1.012	1.004	1.017	1.042	1.08	1.051	1.006
	$E_{\min}$	1.134	1.103	0.984	0.825	0.964	0.988	0.994	0.997	0.995	0.994
	$\Delta_{\min}$	0	3.142	1.795	0.833	0.705	0.769	0.833	0.898	0.833	0.641
0.15	$E_{\max}$	1.103	1.195	1.139	1.007	1.002	1.02	1.031	1.066	1.044	1.007
	$E_{\min}$	1.099	1.076	0.975	0.889	0.979	0.994	0.997	0.997	0.997	0.997
	$\Delta_{\min}$	3.142	3.142	1.667	0.769	0.705	0.833	0.833	0.898	0.898	0.769
0.2	$E_{\max}$	1.084	1.144	1.085	1.005	1.001	1.015	1.026	1.054	1.031	1.007
	$E_{\min}$	1.076	1.06	0.968	0.935	0.987	0.996	0.998	0.998	0.998	0.998
	$\Delta_{\min}$	1.988	3.077	1.603	0.641	0.641	0.833	0.898	0.898	0.898	0.769
0.25	$E_{\max}$	1.064	1.111	1.045	1.005	1	1.017	1.018	1.046	1.023	1.006
	$E_{\min}$	1.059	1.046	0.965	0.96	0.991	0.997	0.998	0.998	0.999	0.999
	$\Delta_{\min}$	0	3.077	1.346	0.385	0.449	0.833	0.833	0.898	0.898	0.833
0.3	$E_{\max}$	1.049	1.08	1.015	1.006	1	1.013	1.016	1.029	1.02	1.007
	$E_{\min}$	1.046	1.032	0.963	0.971	0.995	0.998	0.999	0.999	0.999	0.999
	$\Delta_{\min}$	0	3.142	1.346	0.128	0.577	0.833	0.833	0.898	0.898	0.833
0.35	$E_{\max}$	1.036	1.059	1.002	1.004	1	1.01	1.012	1.02	1.015	1.006
	$E_{\min}$	1.033	1.024	0.965	0.979	0.996	0.999	0.999	0.999	0.999	1
	$\Delta_{\min}$	1.859	3.142	1.282	0	0.449	0.833	0.898	0.962	0.898	0.833
0.4	$E_{\max}$	1.026	1.044	0.992	1.002	1	1.008	1.008	1.013	1.012	1.007
	$E_{\min}$	1.024	1.016	0.965	0.982	0.996	0.999	1	1	0.999	1
	$\Delta_{\min}$	0	3.142	1.154	0	0	0.898	0.898	0.962	0.898	0.898
0.45	$E_{\max}$	1.019	1.031	0.99	1.001	1	1.007	1.008	1.009	1.01	1.005
	$E_{\min}$	1.017	1.011	0.969	0.987	0.998	1	1	1	0.999	1
	$\Delta_{\min}$	0	3.142	1.154	0	0	0.898	0.962	0.898	0.898	0.898
0.5	$E_{\max}$	1.015	1.023	0.993	1.001	1	1.005	1.003	1.005	1.006	1.003
	$E_{\min}$	1.013	1.009	0.971	0.99	0.999	1	1	1	1	1
	$\Delta_{\min}$	0	3.142	0.962	0	0.449	0.898	0.898	0.898	0.898	1.218

Table 2.5: $n = 30$: Maximum and Minimum Guaranteed Efficiencies for the PTE

		$\bar{R}$									
γ		0.165	0.164	0.163	0.460	0.604	0.706	0.772	0.814	0.867	0.897
0.01	$E_{\max}$	1.019	1.037	1.506	1.24	1.02	1.007	1.017	1.041	1.009	1.016
	$E_{\min}$	1.014	1.01	0.984	0.621	0.881	0.942	0.969	0.977	0.965	0.906
	$\Delta_{\min}$	0	3.142	2.436	0.962	0.705	0.641	0.705	0.769	0.449	0
0.02	$E_{\max}$	1.059	1.119	1.344	1.116	1.011	1.006	1.026	1.046	1.014	1.01
	$E_{\min}$	1.048	1.048	0.977	0.705	0.927	0.965	0.982	0.988	0.985	0.939
	$\Delta_{\min}$	3.142	3.142	2.052	0.833	0.641	0.641	0.769	0.833	0.705	0
0.05	$E_{\max}$	1.125	1.26	1.244	1.017	1.004	1.006	1.03	1.043	1.034	1.004
	$E_{\min}$	1.12	1.085	0.942	0.862	0.963	0.983	0.994	0.995	0.995	0.972
	$\Delta_{\min}$	2.372	3.142	1.603	0.769	0.513	0.641	0.833	0.833	0.833	0
0.1	$E_{\max}$	1.14	1.276	1.118	1.011	1.001	1.008	1.023	1.037	1.042	1.002
	$E_{\min}$	1.133	1.079	0.921	0.938	0.981	0.993	0.998	0.997	0.997	0.989
	$\Delta_{\min}$	0.833	3.142	1.346	0.513	0.321	0.769	0.833	0.833	0.898	0
0.15	$E_{\max}$	1.105	1.212	1.055	1.005	1.001	1.009	1.016	1.03	1.041	1.001
	$E_{\min}$	1.096	1.06	0.914	0.963	0.989	0.996	0.998	0.999	0.998	0.997
	$\Delta_{\min}$	0.833	3.142	1.218	0.449	0.256	0.769	0.833	0.898	0.898	0.385
0.2	$E_{\max}$	1.084	1.155	1.007	1.003	1.001	1.008	1.02	1.024	1.03	1.001
	$E_{\min}$	1.077	1.048	0.919	0.977	0.991	0.998	0.999	0.999	0.999	0.999
	$\Delta_{\min}$	0.833	3.142	1.154	0.385	0	0.833	0.898	0.898	0.898	0.513
0.25	$E_{\max}$	1.064	1.119	0.996	1.002	1.001	1.006	1.016	1.016	1.022	1.003
	$E_{\min}$	1.059	1.04	0.926	0.984	0.994	0.999	0.999	0.999	0.999	0.999
	$\Delta_{\min}$	0.769	3.142	1.09	0.192	0	0.833	0.898	0.898	0.898	0.769
0.3	$E_{\max}$	1.048	1.087	0.994	1.001	1.001	1.003	1.012	1.014	1.021	1.004
	$E_{\min}$	1.045	1.028	0.928	0.988	0.993	1	1	0.999	0.999	0.999
	$\Delta_{\min}$	0.769	3.142	0.962	0	0	0.898	0.962	0.898	0.898	0.833
0.35	$E_{\max}$	1.035	1.06	0.994	1.001	1	1.003	1.008	1.012	1.018	1.004
	$E_{\min}$	1.033	1.021	0.93	0.991	0.994	1	1	0.999	0.999	1
	$\Delta_{\min}$	0.641	3.013	0.833	0	0	0.898	0.962	0.898	0.898	0.833
0.4	$E_{\max}$	1.026	1.041	0.993	1	1	1.001	1.006	1.01	1.015	1.005
	$E_{\min}$	1.023	1.014	0.936	0.994	0.997	1	1	1	0.999	1
	$\Delta_{\min}$	0.705	2.949	0.705	0	0	0.833	0.898	0.898	0.898	0.833
0.45	$E_{\max}$	1.019	1.03	0.995	1	1	1.001	1.005	1.006	1.008	1.004
	$E_{\min}$	1.017	1.011	0.945	0.996	0.997	1	1	1	1	1
	$\Delta_{\min}$	0.641	2.949	0.641	0	0	0.833	0.833	0.898	0.898	0.833
0.5	$E_{\max}$	1.015	1.022	0.996	1	1.001	1.001	1.003	1.004	1.005	1.002
	$E_{\min}$	1.014	1.008	0.959	0.996	0.999	1	1	1	1	1
	$\Delta_{\min}$	0.769	3.142	0.513	0	0.641	0.898	0.833	0.898	0.898	0.833

Table 2.6: $n = 40$: Maximum and Minimum Guaranteed Efficiencies for the PTE

						$\bar{R}$					
γ		0.141	0.142	0.142	0.461	0.600	0.703	0.769	0.814	0.865	0.896
0.01	$E_{\max}$	1.019	1.046	1.337	1.082	1.008	1.005	1.01	1.016	1.006	1.012
	$E_{\min}$	1.014	1.019	0.928	0.744	0.925	0.966	0.981	0.989	0.981	0.879
	$\Delta_{\min}$	3.142	3.142	1.603	0.769	0.513	0.513	0.705	0.769	0.449	0
0.02	$E_{\max}$	1.055	1.135	1.287	1.032	1.003	1.004	1.008	1.019	1.011	1.007
	$E_{\min}$	1.05	1.043	0.895	0.844	0.953	0.979	0.989	0.994	0.992	0.913
	$\Delta_{\min}$	1.923	3.142	1.539	0.705	0.449	0.513	0.705	0.769	0.705	0
0.05	$E_{\max}$	1.124	1.29	1.169	1.011	1.002	1.002	1.015	1.018	1.041	1.003
	$E_{\min}$	1.115	1.087	0.855	0.923	0.976	0.989	0.996	0.998	0.997	0.955
	$\Delta_{\min}$	1.731	3.142	1.282	0.577	0.321	0.385	0.833	0.833	0.833	0
0.1	$E_{\max}$	1.136	1.298	1.054	1.004	1.002	1.001	1.014	1.017	1.046	1.001
	$E_{\min}$	1.129	1.078	0.847	0.957	0.984	0.995	0.999	0.999	0.998	0.976
	$\Delta_{\min}$	1.475	3.142	1.154	0.385	0	0.385	0.898	0.898	0.898	0
0.15	$E_{\max}$	1.102	1.218	0.998	1.002	1.001	1.001	1.009	1.014	1.037	1.001
	$E_{\min}$	1.097	1.06	0.857	0.97	0.992	0.997	0.999	0.999	0.998	0.989
	$\Delta_{\min}$	1.731	3.142	1.09	0.128	0	0.385	0.898	0.898	0.898	0
0.2	$E_{\max}$	1.081	1.163	0.997	1.001	1	1	1.007	1.012	1.032	1
	$E_{\min}$	1.076	1.047	0.869	0.975	0.996	0.997	1	0.999	0.998	0.989
	$\Delta_{\min}$	1.603	3.142	0.962	0	0	0	0.898	0.898	0.898	0
0.25	$E_{\max}$	1.062	1.122	0.997	1.001	1	1.001	1.004	1.009	1.028	1
	$E_{\min}$	1.058	1.035	0.884	0.982	0.997	0.999	1	0.999	0.999	0.99
	$\Delta_{\min}$	1.539	3.142	0.833	0	0	0.705	0.898	0.833	0.898	0
0.3	$E_{\max}$	1.046	1.088	0.997	1	1	1	1.002	1.008	1.023	1
	$E_{\min}$	1.044	1.026	0.906	0.986	0.998	1	1	0.999	0.999	0.993
	$\Delta_{\min}$	1.539	2.949	0.769	0	0	0.705	0.898	0.833	0.898	0
0.35	$E_{\max}$	1.034	1.065	0.998	1.001	1	1.001	1.003	1.005	1.016	1
	$E_{\min}$	1.031	1.019	0.929	0.987	0.999	1	1	1	0.999	0.996
	$\Delta_{\min}$	1.539	2.821	0.641	0	0	1.218	0.962	0.833	0.898	0
0.4	$E_{\max}$	1.024	1.046	0.999	1.001	1	1	1.003	1.005	1.014	1
	$E_{\min}$	1.023	1.013	0.947	0.99	1	1	1	1	0.999	0.996
	$\Delta_{\min}$	1.154	2.757	0.513	0	0.641	0.641	0.898	0.833	0.898	0
0.45	$E_{\max}$	1.018	1.035	1	1	1.002	1	1.003	1.005	1.009	1
	$E_{\min}$	1.017	1.01	0.96	0.993	1	0.999	1	1	1	0.998
	$\Delta_{\min}$	0.833	2.693	0.321	0	0.833	0	0.898	0.833	0.898	0
0.5	$E_{\max}$	1.014	1.025	1.001	1	1	1	1.002	1.005	1.007	1
	$E_{\min}$	1.013	1.007	0.97	0.994	1	0.999	1	1	1	0.999
	$\Delta_{\min}$	3.142	2.693	0.128	0	0.769	0	0.898	0.898	0.898	0

Table 2.7: $n = 50$: Maximum and Minimum Guaranteed Efficiencies for the PTE

						$\bar{R}$					
γ		0.126	0.124	0.124	0.455	0.604	0.705	0.767	0.814	0.866	0.896
0.01	$E_{\max}$	1.024	1.056	1.335	1.028	1.004	1.004	1.004	1.004	1.013	1.008
	$E_{\min}$	1.016	1.021	0.842	0.837	0.949	0.976	0.979	0.987	0.994	0.907
	$\Delta_{\min}$	0.962	3.142	1.346	0.641	0.385	0.449	0.128	0.385	0.769	0
0.02	$E_{\max}$	1.062	1.155	1.245	1.015	1.004	1.003	1.002	1.007	1.032	1.004
	$E_{\min}$	1.052	1.052	0.801	0.893	0.962	0.986	0.99	0.995	0.997	0.936
	$\Delta_{\min}$	1.154	3.142	1.218	0.513	0.128	0.449	0.385	0.769	0.833	0
0.05	$E_{\max}$	1.134	1.327	1.111	1.005	1.003	1.002	1.001	1.015	1.04	1.002
	$E_{\min}$	1.114	1.091	0.778	0.942	0.975	0.993	0.997	0.998	0.998	0.97
	$\Delta_{\min}$	1.154	3.142	1.154	0.385	0	0.385	0.513	0.833	0.898	0
0.1	$E_{\max}$	1.15	1.335	1.006	1.002	1.001	1.004	1.002	1.013	1.029	1.001
	$E_{\min}$	1.126	1.065	0.802	0.964	0.99	0.998	0.999	0.999	0.999	0.984
	$\Delta_{\min}$	0.833	3.142	1.026	0.128	0.256	0.769	0.769	0.898	0.898	0
0.15	$E_{\max}$	1.114	1.252	1.005	1.001	1.001	1.007	1	1.011	1.022	1
	$E_{\min}$	1.097	1.051	0.828	0.972	0.994	0.999	0.999	1	0.999	0.99
	$\Delta_{\min}$	3.142	3.142	0.898	0	0	0.833	0.513	0.898	0.898	0
0.2	$E_{\max}$	1.085	1.183	1.005	1.001	1	1.006	1.002	1.01	1.013	1
	$E_{\min}$	1.078	1.043	0.864	0.982	0.995	0.999	1	1	0.999	0.993
	$\Delta_{\min}$	3.142	3.142	0.833	0	0	0.833	0.769	0.898	0.898	0
0.25	$E_{\max}$	1.069	1.132	1.004	1.001	1	1.007	1.001	1.007	1.012	1
	$E_{\min}$	1.061	1.03	0.902	0.984	0.997	0.999	1	1	0.999	0.995
	$\Delta_{\min}$	0.833	3.142	0.769	0	0	0.833	0.705	0.898	0.898	0
0.3	$E_{\max}$	1.051	1.098	1.004	1	1	1.006	1.002	1.005	1.011	1
	$E_{\min}$	1.046	1.021	0.931	0.987	0.998	1	1	1	0.999	0.997
	$\Delta_{\min}$	0.833	3.142	0.641	0	0	0.833	0.833	0.898	0.898	0
0.35	$E_{\max}$	1.037	1.072	1.004	1	1	1.004	1	1.003	1.01	1
	$E_{\min}$	1.033	1.015	0.952	0.991	0.997	1	0.999	1	0.999	0.998
	$\Delta_{\min}$	2.629	3.142	0.513	0	0	0.833	0	0.833	0.898	0
0.4	$E_{\max}$	1.025	1.048	1.004	1	1	1.003	1	1.003	1.009	1
	$E_{\min}$	1.024	1.01	0.974	0.994	0.997	1	0.999	1	1	0.999
	$\Delta_{\min}$	2.693	3.142	0.385	0	0	0.833	0	0.833	0.898	0
0.45	$E_{\max}$	1.018	1.033	1.005	1	1	1.002	1	1.003	1.007	1
	$E_{\min}$	1.017	1.006	0.986	0.996	0.998	1	0.998	1	1	0.998
	$\Delta_{\min}$	0.898	3.142	0.385	0	0	0.833	0	0.833	0.898	0
0.5	$E_{\max}$	1.014	1.025	1.005	1	1	1.002	1	1.003	1.006	1
	$E_{\min}$	1.014	1.005	0.99	0.999	0.999	1	0.999	1	1	1
	$\Delta_{\min}$	0.833	3.142	0.128	0	0	0.833	0	0.898	0.898	0.769

Chapter 3

Detecting Change in the number of Modes

3.1 Previous Work

Previous work in circular statistics regarding change-point problems included detecting change in the mean direction in a time-ordered sequence of observations, see (Ghosh *et al.* , 1999). Likelihood ratio based tests were derived when both the concentration parameter is known and unknown. Other examples of change-point problems can be found in (Jammalamadaka & SenGupta, 2001).

Other work that relates to our current problem, considered testing for unimodality of circular data. In (Basu & Jammalamadaka, 2000) a Bayesian test is derived to test whether a group of circular observations is unimodal or not. Also relevant to our current work is the paper by (Holzmann & Vollmer, 2008) who consider the asymptotic distribution of the likelihood ratio

test statistic in dealing with the mixture of two CN densities. We first review the papers before we embark on our investigation of detecting change in the number of modes.

3.1.1 Test for Change in Mean Direction

In (Ghosh *et al.* , 1999) a parametric test for the change-point problem is considered which deals with the change in mean direction within a group of time-ordered circular observations. An application of this for instance, is when a change in wind direction carries pollution from a big city into a neighboring small town. Given a set of time-ordered wind directions a researcher will be interested if there is a change in wind direction over a given period of time, for example a single day, a week, or a month. This is useful in a context like the one studied in (Nava & Jammalamadaka, 2008) and (Jammalamadaka & Lund, 2006) where the relationship between ozone levels and wind direction is considered.

(Ghosh *et al.* , 1999) consider a set of univariate independent time-ordered observations $(\alpha_1, \ldots, \alpha_n)$ and asks if there is a point of change k, such that $(\alpha_1, \ldots, \alpha_k)$ have a CND with mean direction μ_1, while the succeeding observations $(\alpha_{k+1}, \ldots, \alpha_n)$ have a CND with a different mean direction μ_2. We assume the change occurs at some unknown point k, $(1 \leq k \leq n-1)$. The Generalized Likelihood Ratio Test (GLRT) method is used to derive a test for the null hypothesis of no change-point (i.e. all the n observations have the same distribution), i.e. $H_0 : k = n$ versus $H_1 : 1 \leq k \leq n-1$.

Let $\theta = (k, \mu_1, \mu_2)$ denote the parameter vector and for illustrative purposes we will assume the concentration parameter κ is known, as in one of the cases presented in (Ghosh *et al.* , 1999). The parameter space for this problem becomes $\Omega = \{1, \ldots, n\} \times [-\pi, \pi) \times [-\pi, \pi)$. The null hypothesis corresponds to no-change or that the change if any is at n corresponding to

the subspace $\Omega_0 = \Omega \cap H_0 = \{n\} \times [-\pi, \pi) \times [-\pi, \pi)$ of Ω. The subspace corresponding to a change at a specific k, $1 \leq k \leq n-1$ is given by $\Omega_k = \{k\} \times [-\pi, \pi) \times [-\pi, \pi)$. For the change-point problem the likelihood ratio is computed for each possible k where the change can occur, resulting in the likelihood ratio for a single instance of $k \in \{1, \ldots, n-1\}$,

$$\begin{aligned}
\lambda_k &= \frac{\sup_{\theta \in \Omega_0} \prod_{i=1}^{n} \frac{1}{2\pi I_0(\kappa)} \exp(\kappa \cos(\alpha_i - \mu_1))}{\sup_{\theta \in \Omega_0 \cup \Omega_k} \prod_{i=1}^{k} \frac{1}{2\pi I_0(\kappa)} \exp(\kappa \cos(\alpha_i - \mu_1)) \prod_{i=k+1}^{n} \frac{1}{2\pi I_0(\kappa)} \exp(\kappa \cos(\alpha_i - \mu_2))} \\
&= \frac{\sup_{\theta \in \Omega_0} \exp(\kappa \sum_{i=1}^{n} \cos(\alpha_i - \mu_1))}{\sup_{\theta \in \Omega_0 \cup \Omega_k} \exp(\kappa \sum_{i=1}^{k} \cos(\alpha_i - \mu_1) + \kappa \sum_{i=k+1}^{n} \cos(\alpha_i - \mu_2))}
\end{aligned}$$

Note, if $k = n$ then $\lambda_k = 1$. In both the numerator and the denominator we compute the MLEs for the parameters μ_1 and μ_2. As stated in 1.11, there is an analytic solution for the MLE of the mean direction parameter. Specifically under H_0, the MLE for μ_1 is the solution to:

$$\sum_{i=1}^{n} \sin(\alpha_i - \mu_1) = 0 \tag{3.1}$$

Under H_k, the alternative for a given $k \in \{1, \ldots, n-1\}$, the MLEs for μ_1 and μ_2 are solutions to:

$$\sum_{i=1}^{k} \sin(\alpha_i - \mu_1) = 0 \quad \text{and} \quad \sum_{i=k+1}^{n} \sin(\alpha_i - \mu_2) = 0 \tag{3.2}$$

We denote the MLEs in (3.1) as $\bar{\alpha}_0$ and in (3.2) as $\bar{\alpha}_{1k}$ and $\bar{\alpha}_{2k}$. The MLEs are computed using the formula for mean direction in (1.2). The likelihood ratio for a single $k \in \{1, \ldots, n-1\}$ becomes:

$$\begin{aligned} \lambda_k &= \exp\left[\kappa\left(\sum_{i=1}^{n}\cos(\alpha_i - \bar{\alpha}_0) - \sum_{i=1}^{k}\cos(\alpha_i - \bar{\alpha}_{1k}) - \sum_{i=k+1}^{n}\cos(\alpha_i - \bar{\alpha}_{2k})\right)\right] \\ &= \exp\left[\kappa(R - R_{1k} - R_{2k})\right] \end{aligned}$$

As stated in (1.4), each summation of cosine functions is equal to the length of the resultant vector in (1.1) which are denoted by R, R_{1k}, and R_{2k}. In these models, the length of the resultant vector, represents how concentrated the data are around the mean direction.

Since the location of the change-point is unknown, we compute the likelihood ratio for each possible value of k, then choose the minimum λ_k over all possible k values. Or, alternatively we take a Bayesian approach and assume a uniform prior distribution on possible change-points $\{1, \ldots, n\}$. In this case, the average of $-\ln(\lambda_k)$'s forms the test statistic. In each case we reject H_0 for large values of:

$$\lambda_{max} = \sup_{k \in 1,\ldots,n} (R_{1k} + R_{2k}) - R > c_1; \tag{3.3}$$

$$\lambda_{avg} = \frac{1}{n}\sum_{k=1}^{n}(R_{1k} + R_{2k}) - R > c_2; \tag{3.4}$$

where the critical points c_1 and c_2 are determined by some pre-determined significance level. The sampling distributions of the test statistics for CN are complicated and thus simulations

were used to determine the cut-off points. See (Ghosh *et al.* , 1999) for further details and for the case when κ is unknown.

3.2 Some Asymptotic Results for a Likelihood Ratio Test

(Holzmann & Vollmer, 2008) propose a test for bimodality on a set of independent and identically distributed sample, based on the likelihood ratio test by using two-component mixtures. As in (Basu & Jammalamadaka, 2000), their discussion does not involve any change-point but just whether or not a group of independent and identically distributed circular observations come from a unimodal distribution or not. The results of the paper can be applied to any mixture of two densities that satisfy certain assumptions outlined below. For example, their results are applicable to mixtures of two Normals, or two CNDs, and we will focus on the latter.

In the general set-up let $f(\alpha|\theta)$, $\theta \in \Theta \subset \mathbb{R}^d$, be a parametric family of densities with two component mixtures:

$$g(\alpha, \theta_1, \theta_2, p) = pf(\alpha|\theta_1) + (1-p)f(\alpha|\theta_2), \tag{3.5}$$

where

$$(\theta_1, \theta_2, p) \in \Theta \times \Theta \times [0,1] = \Theta_{mix} \subset \mathbb{R}^{2d+1}.$$

They consider the mixtures to have equal variances or concentration parameters i.e., the subset $E_{mix} \subset \Theta_{mix}$ such that $E_{mix} \subset \mathbb{R}^q$ where $q \leq 2d+1$. Since the mixture can have

up to two modes we can split E_{mix} disjointly into $E_{unim} \cup E_{bim}$ where E_{unim} is the unimodal parameter set and E_{bim} is the bimodal parameter set. Also let ∂E_{unim} denote the boundary between E_{unim} and E_{bim}. Given observations $\alpha_1, \ldots, \alpha_n$ i.i.d. from (3.5) the log-likelihood function is given by,

$$\ell_n(\theta_1, \theta_2, p) = \sum_{i=1}^{n} \log f(\alpha_i; \theta_1, \theta_2, p) \tag{3.6}$$

The following assumptions are now made:

Assumption 1. The partial derivatives of $\log f(\alpha_i; \theta_1, \theta_2, p)$ of order 3 with respect to θ_1, θ_2 and p exists a.s., at least in a neighborhood of N of the true value $(\theta_1^0, \theta_2^0, p^0)$.

Assumption 2. For $(\theta_1, \theta_2, p) \in N$, the first and second order partial derivatives of (3.5) are uniformly bounded in absolute value by a function $F(\alpha)$ with finite integral, and the third order partial derivatives of $\log f(\alpha_i; \theta_1, \theta_2, p)$ are uniformly bounded in absolute value by a function $H(\alpha)$ with $\mathbb{E}H(\alpha) < \infty$.

Assumption 3. The expectation of the matrix of the second order derivatives of $\log f(\alpha_i; \theta_1, \theta_2, p)$ is finite and positive definite for $(\theta_1, \theta_2, p) \in N$.

Theorem 1 *Suppose that the true parameter vector $(\theta_1^0, \theta_2^0, p^0)$ of the mixture density lies on the boundary ∂E_{unim}, and locally around $(\theta_1^0, \theta_2^0, p^0)$, ∂E_{unim} is a smooth $(q-1)$-dimensional surface in $\mathbb{R}^q$. Further, if Assumptions 1-3 hold, then we have*

$$R_n := 2\left(\sup_{(\theta_1,\theta_2,p)\in E_{mix}} \ell_n(\theta_1, \theta_2, p) - \sup_{(\theta_1,\theta_2,p)\in E_{unim}} \ell_n(\theta_1, \theta_2, p) \right) \xrightarrow{d} \left(\chi_0^2 + \chi_1^2\right)/2, \tag{3.7}$$

where χ^2_0 is the point measure at 0 and χ^2_1 is the Chi-squared distribution with 1 degree of freedom.

The theorem utilizes regularity conditions given in (Chernoff, 1954) and asymptotic properties of MLE's in non-regular likelihood ratio settings discussed in (Self & Liang, 1987).

Assuming the concentration parameters to be equal and using (1.16) to represent the mixture of two CNDs, we can check the Assumptions 1-3 and hence Theorem 1 holds in our case. We refer back to Table 1.1 for the special case of $\kappa_1 = \kappa_2$. We note that all the cases where the parameters $(\delta, \kappa, p) \in \Omega_0$ give a unimodal distribution (see (Mardia & Sutton, 1975)) form a continuous set. For example, for $\sin(\delta) = 2\kappa \sin^3(\delta/2)$, the mixture is unimodal for all $(0 < p < 1)$. Also, as $\delta \to \pi$, then $t(\alpha)/(1 - t(\alpha)) \to (1 + \exp(2\kappa))^{-1}$ and we see how case (ii) and case(iii) merge together in Table 1.1.

In Figure 3.1 we provide a visual representation of the negative log-likelihood via a contour plot over all possible values of δ and p for a specific value of κ. The plot is obtained by simulating one-thousand observations from a mixCND with $(\delta = \pi/2, \kappa = 4, p = 0.5)$. The bold wine glass shaped curve is the boundary between the unimodal and bimodal parameter subspaces. Inside the wine glass is the bimodal parameter space and outside the wine glass is the unimodal parameter space. The distribution we simulated was bimodal and the bimodal parameter space contains the MLEs for δ and p as expected.

For any fixed p, Theorem 1 applies directly for any combination of δ and κ. As p varies there will occur a singularity on the boundary of the set of unimodal parameters at $p = 0.50$ if $\sin(\delta) = 2\kappa \sin^3(\delta/2)$, but the test asymptotically preserves the critical values.

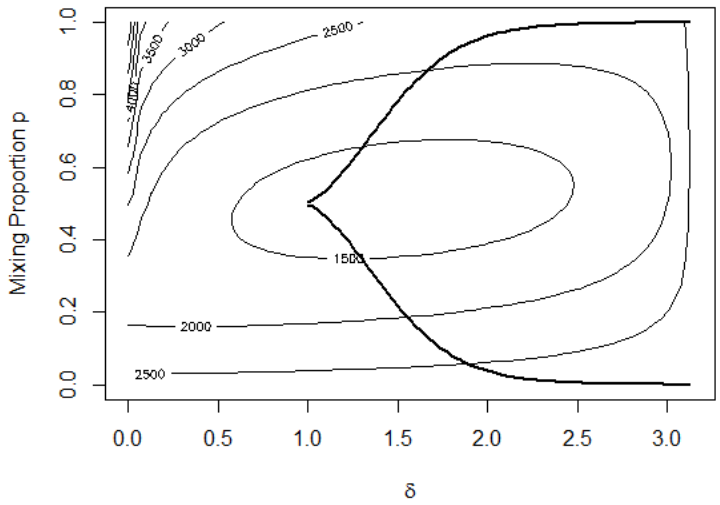

Figure 3.1: The likelihood space for δ and p, given $\kappa = 4$ and simulation-bases data from $(\alpha_1, \ldots, \alpha_{1000}) \sim \text{mixCN}(\delta = \pi/2, \kappa = 4, p = 0.5)$

3.3 The Generalized Likelihood Ratio Test

Recall we have a set of independent, time-ordered vectors of circular observations, $\underset{\sim}{\alpha}_1, \underset{\sim}{\alpha}_2, \ldots, \underset{\sim}{\alpha}_n$ and are interested to find the point when the observations change from having a unimodal distribution to a multimodal distribution. Here, $\underset{\sim}{\alpha}_j = (\alpha_{j1}, \alpha_{j2}, \ldots, \alpha_{jm})$, is a vector of i.i.d. observations at time j, where $j \in \{1, \ldots, n\}$. We assume each of these is of the same length m for simplicity, although this can be relaxed.

Specifically, we assume there is some unknown but fixed k, $(1 \leq k \leq n-1)$ such that $\underset{\sim}{\alpha}_1, \ldots, \underset{\sim}{\alpha}_k$ have unimodal densities with pdf's say $\{f_1\}$ and $\underset{\sim}{\alpha}_{k+1}, \ldots, \underset{\sim}{\alpha}_n$ have bimodal densities with pdf's say $\{f_2\}$. Note, the the unimodal/bimodal vectors of observations are not required to be identically distributed, for example two unimodal vectors could be centered around a

different direction. The point k is considered the "change-point" of the observed data, which is unknown.

Let $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_n, k) \subset \Omega$ be the (3n+1)-dimensional parameter space where $\boldsymbol{\theta}_j = (\delta_j, \kappa_j, p_j)$ is the the parameter vector for the j^{th} vector of data. Here, f_j is our parametric model that can be unimodal or bimodal, depending on the parameter values $\boldsymbol{\theta}_j$ at the j^{th} stage. For a given $\boldsymbol{\theta}_j = (\mu_j, \kappa_j, p_j)$ and f_j, let ω_{0j} represent the unimodal subspace as before, while $\omega_{1j} = \omega - \omega_{0j}$ is the bimodal subset. Also for the full parameter set $\boldsymbol{\Theta}$ let $\Omega_k \subset \Omega$ be the subset of $\boldsymbol{\Theta}$ values for which the change from unimodality to bimodality occurs at step k, while Ω_0 represents no change. Since we assume there is at most one change and the sequence begins with a unimodal density, we re-state the null and alternative hypothesis in terms of the parameter sets as follows:

- H_0 : no change, i.e. the data continues to be unimodal. This corresponds to the unimodal parameter space.

$$\boldsymbol{\Theta} \in \Omega_0 = \underbrace{(\omega_{01} \times \omega_{02} \times \cdots \times \omega_{0n})}_{\text{n times}}$$

.

- H_k : change at $k+1$, i.e. unimodal until k, and multimodal from $k+1$.

$$\boldsymbol{\Theta} \in \Omega_k = \underbrace{(\omega_{01} \times \cdots \times \omega_{0k})}_{\text{k times}} \times \underbrace{(\omega_{1(k+1)} \times \cdots \times \omega_{1n})}_{\text{n-k times}}$$

- $H_A = \bigcup_{k=1}^{n-1} H_k$: we consider all possible change points k.

Below is an illustration of the parameter space Ω_k for different values of k.

$$
\begin{aligned}
\Omega_1 &= \underbrace{(\omega_{01})}_{\text{1 time}} \times \underbrace{(\omega_{12} \times \cdots \times \omega_{1n})}_{\text{n-1 times}} \\
\Omega_2 &= \underbrace{(\omega_{01} \times \omega_{02})}_{\text{2 times}} \times \underbrace{(\omega_{13} \times \cdots \times \omega_{1n})}_{\text{n-2 times}} \\
&\vdots \\
\Omega_{n-1} &= \underbrace{(\omega_{01} \times \ldots \times \omega_{0(n-1))}}_{\text{n-1 times}} \times \underbrace{(\omega_{1n})}_{\text{1 time}}
\end{aligned}
$$

Parameter space Ω_1 corresponds to the change-point occurring at $k = 1$. Therefore the first vector of time corresponds to the unimodal parameter space ω_{01} and the remaining vectors after $k = 1$ correspond to the multimodal parameter spaces $\omega_{12}, \ldots, \omega_{1n}$. Therefore $\underset{\sim}{\alpha}_1$ are drawn from $\{f_1\}$, and the remaining $\underset{\sim}{\alpha}_2, \ldots, \underset{\sim}{\alpha}_n$ come from $\{f_2\}$. After considering all possible change-points, the alternate hypothesis parameter space consists of the union over all the possible change-point values thus, $\Omega_A = \bigcup_{k=1}^{n-1} \Omega_k$.

We proceed to derive the GLRT for our change-point problem, which for a given value $k \in 1, \ldots, n-1$ takes the form:

$$
\lambda_k := \frac{\sup_{\Theta \in \Omega_0} \prod_{j=1}^{n} \prod_{i=1}^{m} f_j(\alpha_{ji}, \boldsymbol{\theta}_j)}{\sup_{\Theta \in \Omega_0 \cup \Omega_k} \prod_{j=1}^{k} \prod_{i=1}^{m} f_j(\alpha_{ji}, \boldsymbol{\theta}_j) \prod_{j=k+1}^{n} \prod_{i=1}^{m} f_j(\alpha_{ji}, \boldsymbol{\theta}_j)} \tag{3.8}
$$

Note that $\lambda_k = 1$ for $k = n$. Writing $L(\boldsymbol{\theta}_j | \underset{\sim}{\alpha}_j) = \prod_{i=1}^{m} f_j(\alpha_{ji}, \boldsymbol{\theta}_j)$ for the likelihood at time j, the likelihood ratio for a given k can be written as:

$$\lambda_k = \frac{\sup_{\boldsymbol{\Theta}\in\Omega_0}\prod_{j=1}^{n} L(\boldsymbol{\theta}_j|\underset{\sim}{\alpha}_j)}{\sup_{\boldsymbol{\Theta}\in\Omega_0\cup\Omega_k}\prod_{j=1}^{k} L(\boldsymbol{\theta}_j|\underset{\sim}{\alpha}_j)\prod_{j=k+1}^{n} L(\boldsymbol{\theta}_j|\underset{\sim}{\alpha}_j)} \tag{3.9}$$

$$= \frac{\prod_{j=1}^{n}\sup_{\omega_0\in\Omega_0} L(\boldsymbol{\theta}_j|\underset{\sim}{\alpha}_j)}{\prod_{j=1}^{k}\sup_{\omega_0\in\Omega_0} L(\boldsymbol{\theta}_j|\underset{\sim}{\alpha}_j)\prod_{j=k+1}^{n}\sup_{\omega\in\Omega} L(\boldsymbol{\theta}_j|\underset{\sim}{\alpha}_j)} \tag{3.10}$$

$$= \frac{\prod_{j=1}^{k}\sup_{\omega_0\in\Omega_0} L(\boldsymbol{\theta}_j|\underset{\sim}{\alpha}_j)\prod_{j=k+1}^{n}\sup_{\omega_0\in\Omega_0} L(\boldsymbol{\theta}_j|\underset{\sim}{\alpha}_j)}{\prod_{j=1}^{k}\sup_{\omega_0\in\Omega_0} L(\boldsymbol{\theta}_j|\underset{\sim}{\alpha}_j)\prod_{j=k+1}^{n}\sup_{\omega\in\Omega} L(\boldsymbol{\theta}_j|\underset{\sim}{\alpha}_j)} \tag{3.11}$$

$$= \prod_{j=k+1}^{n}\frac{\sup_{\omega_0\in\Omega_0} L(\boldsymbol{\theta}_j|\underset{\sim}{\alpha}_j)}{\sup_{\omega\in\Omega} L(\boldsymbol{\theta}_j|\underset{\sim}{\alpha}_j)} \tag{3.12}$$

Here $\lambda_n = 1$ since in the denominator of 3.9, $\Omega_0 \cup \Omega_{k=n} = \Omega_0$. In 3.11 we use independence of the vectors, $(\underset{\sim}{\alpha}_1, \ldots, \underset{\sim}{\alpha}_n)$, as well as independence of observations $(\alpha_{j1}, \ldots, \alpha_{jm})$ within each vector for $j = 1, \ldots, n$. From 3.11 to 3.12 we have:

$$\Omega_0 \cup \Omega_k = \underbrace{(\omega_{01} \times \omega_{02} \times \cdots \times \omega_{0n})}_{\text{n times}} \cup \left(\underbrace{(\omega_{01} \times \cdots \times \omega_{0k})}_{\text{k times}} \times \underbrace{(\omega_{1(k+1)} \times \cdots \times \omega_{1n})}_{\text{n-k times}} \right) \tag{3.13}$$

$$= \underbrace{(\omega_{01} \times \cdots \times \omega_{0k})}_{\text{k times}} \times \underbrace{(\omega_{k+1} \times \cdots \times \omega_n)}_{\text{n-k times}} \tag{3.14}$$

where $\omega_j = \omega_{1j} \cup \omega_{0j}$, for $j = 1, \ldots, n$, is the unrestricted parameter space. As a result, the likelihood ratio reduces into a ratio involving only the data that occurs after the change-point k. Since we have $0 < \lambda_k < 1$ for $k \in \{1, \ldots, n-1\}$, we can re-express 3.11:

$$\lambda_k = \begin{cases} \prod_{j=k+1}^{n} \frac{\sup_{\omega_0 \in \Omega_0} L(\boldsymbol{\theta}_j|\mathcal{Q}_j)}{\sup_{\omega \in \Omega} L(\boldsymbol{\theta}_j|\mathcal{Q}_j)}, & \text{if } \prod_{j=k+1}^{n} \sup_{\omega \in \Omega} L(\boldsymbol{\theta}_j|\mathcal{Q}_j) > \prod_{j=k+1}^{n} \sup_{\omega_0 \in \Omega_0} L(\boldsymbol{\theta}_j|\mathcal{Q}_j) \\ 1, & \text{o.w.} \end{cases}$$

The maximum value of unity represents the case where the MLEs both lie in the unimodal parameter space for $k \in \{1, \ldots, n\}$, representing no change at the single vector of time.

For change-point tests, λ_k can be computed over each instance where a change can occur, i.e. for $k = 1, \ldots, n-1$. Since k is an unknown constant, the $n-1$ sub-test-statistics are usually combined in one of two ways to form the overall test statistic for detecting change anywhere. As discussed before, one uses either the supremum of the log-likelihood ratios over k (3.15) or average the log-likelihood ratios over k (3.16), which agrees with a Bayesian setting assuming a discrete uniform prior distribution on the change-point k. In either case, we would reject the null hypothesis for large value of the test statistic where the critical regions, d_1 and d_2 are set by some predetermined significance level.

$$\lambda_{\text{sup}} := \sup_{k=1,\ldots,n-1} -2 \log \lambda_k > d_1 \tag{3.15}$$

$$\lambda_{\text{avg}} := \frac{1}{n-1} \sum_{k=1}^{n-1} -2 \log \lambda_k > d_2 \tag{3.16}$$

Because of our set-up, the two test statistics reduce to considerably simpler form. Consider first the logarithm of the likelihood for a single k:

$$\begin{aligned} -2\log\lambda_k &= -2\log\left(\prod_{j=k+1}^{n}\frac{\sup_{\omega_0\in\Omega_0}L(\boldsymbol{\theta}_j|\underset{\sim}{\alpha}_j)}{\sup_{\omega\in\Omega}L(\boldsymbol{\theta}_j|\underset{\sim}{\alpha}_j)}\right) \\ &= -2\left(\sum_{j=k+1}^{n}\sup_{\omega_0\in\Omega_0}\ell(\boldsymbol{\theta}_j|\underset{\sim}{\alpha}_j)-\sum_{j=k+1}^{n}\sup_{\omega\in\Omega}\ell(\boldsymbol{\theta}_j|\underset{\sim}{\alpha}_j)\right) \\ &= \sum_{j=k+1}^{n}2\left(\sup_{\omega\in\Omega}\ell(\boldsymbol{\theta}_j|\underset{\sim}{\alpha}_j)-\sup_{\omega_0\in\Omega_0}\ell(\boldsymbol{\theta}_j|\underset{\sim}{\alpha}_j)\right) \end{aligned}$$

The result is a sum of independent quantities because they are functions of independent observations. Also, each term in the summation is non-negative thus $-2\log\lambda_1 \geq -2\log\lambda_2 \geq \ldots \geq -2\log\lambda_{n-1}$. This leads to the conclusion of the supremum of the likelihoods over all k is λ_1, i.e. the case where change-point occurs at time-point one. Also, an examination of the average test statistic corresponds to a weighted sum of these independent quantities.

$$\lambda_{\text{sup}} = -2\log\lambda_1 = \sum_{j=2}^{n}2\left(\sup_{\omega\in\Omega}\ell(\boldsymbol{\theta}_j|\underset{\sim}{\alpha}_j)-\sup_{\omega_0\in\Omega_0}\ell(\boldsymbol{\theta}_j|\underset{\sim}{\alpha}_j)\right) \tag{3.17}$$

$$\lambda_{\text{avg}} = \frac{1}{n-1}\sum_{j=2}^{n}(j-1)2\left(\sup_{\omega\in\Omega}\ell(\boldsymbol{\theta}_j|\underset{\sim}{\alpha}_j)-\sup_{\omega_0\in\Omega_0}\ell(\boldsymbol{\theta}_j|\underset{\sim}{\alpha}_j)\right) \tag{3.18}$$

In each case the test statistic reduces to a sum of independent terms. In 3.17, the j^{th} term in the summation is the evidence against unimodality for each $j \in \{2,\ldots,n\}$ (single time point of data). A zero value for the j^{th} term indicates the unimodal and unrestricted MLEs are the same at that time-point, indicating no evidence against unimodality at that time-point. A value greater than zero implies the unrestricted MLE's are better than the unimodal (restricted)

MLEs, which indicates that the MLEs correspond to a multimodal distribution form. We reject the null hypothesis for large values of the supremum test statistic.

In 3.18, the average test statistic reduces down into a weighted average of independent terms. The terms are the same as in the supremum test 3.17, but the weight of each term in the averaging process increases by an increment of $\frac{1}{n-1}$ for every increase in a single unit time-point. The last time-point of data has that largest weight of $\frac{n-1}{n-1}$ and time-point two has smallest weight at $\frac{1}{n-1}$. The intuition behind this statistic is that change-points early in the observation time period can be difficult to detect. The change may not be instantaneous and could fluctuate between modes before the transition completes. Across all possibilities of timings of a change-point, the last time-point vector of data would be the best indicator of the occurrence of the change in distribution.

In either case, both 3.17 and 3.18 reduce down to a form of sum of independent terms. Each term is the log-likelihood (or a weighted log-likelihood) for a single time-point of data. Though the exact distributions of test statistics are very complex, this reduction promises potential for obtaining the asymptotic distributions.

Thus far the discussion has been general and and holds for any unimodal or multimodal densities. We now choose a specific model $f_j(\alpha_{ji}, \theta_j)$ for each $i \in \{1, \ldots, m\}$ as in (3.8), the mixture of two CNDs which allows for unimodality as well as bimodality.

A mixCND has a $(3n+1)$-dimensional parameter $\boldsymbol{\Theta}$ with parameter space $\Omega = [-\pi, \pi)^n \times (0, \infty)^n \times [0, 1]^n \times \{1, \ldots, n\}$. Under H_0, the parameter space becomes $\Omega_0 = \Omega \cap H_0 = \{\boldsymbol{\theta}_j \in \omega_{0j} \ \ \forall j\} \times \{n\}$. Refer to Table 1.1 to view unimodal and bimodal parameter spaces.

Under the mixCND special case, our likelihood ratio for single $k \in \{1, \ldots, n-1\}$, (3.8), becomes:

$$\lambda_k = \prod_{j=k+1}^{n} \frac{\sup_{\omega_0 \in \Omega_0} pCN(0, \kappa|\underset{\sim}{\alpha}_j) + (1-p)CN(\delta, \kappa|\underset{\sim}{\alpha}_j)}{\sup_{\omega \in \Omega} pCN(0, \kappa|\underset{\sim}{\alpha}_j) + (1-p)CN(\delta, \kappa|\underset{\sim}{\alpha}_j)} \tag{3.19}$$

In the numerator of (3.19) we find the MLEs for (δ, κ, p) under the unimodal parameter space for each vector of data.

In the denominator of (3.19) , we compute the unrestricted MLEs for (δ, κ, p) based on each vector of data from the $(k+1)^{st}$ to the n^{th}. There is not an analytic solution for the MLEs for the parameters of a mixCNDs. To find the MLEs we use the function `nlimnb()` in R which uses a numerical minimization similar to Gauss-Newton algorithm for a non-linear parameter space. See (Gay, 1990) for details.

With the assumptions made, combined with the simplified form of our test statistics we can use the results in (Holzmann & Vollmer, 2008) to find the asymptotic distribution of the test statistics. We begin with the results for the supremum statistic. In 3.17 the test statistic reduced to a sum of independent vectors of data. Each term in the sum has the same form as in 3.6 of (Holzmann & Vollmer, 2008). As long as there are a large number of observations for each vector then each term converges in distribution to a mixture of chi-squared distributions. The direct result is a convolution of independent mixtures of chi-square distributions. To see how this convolution evolves, we examine a simple case where $n = 3$.

For $n = 3$ the test statistic for the supremum will be the sum for two vectors of data. Following notations as in (Holzmann & Vollmer, 2008):

$$\lambda_{\text{suprem}} = R_{2m} + R_{3m}, \tag{3.20}$$

where $R_{2m} \xrightarrow{d} \left(\chi_0^2 + \chi_1^2\right)/2$ and $R_{3m} \xrightarrow{d} \left(\chi_0^2 + \chi_1^2\right)/2$ for sufficiently large m.

The convolution of two distributions is defined as:

$$f_{X+Y}(x) = \int f_X(x-y) g_Y(y) dy \tag{3.21}$$

Using the definition we have:

$$\begin{aligned} f_{X+Y}(x) &= \int \frac{1}{2}\left(\chi_0^2(x-y) + \chi_1^2(x-y)\right)\frac{1}{2}\left(\chi_0^2(y) + \chi_1^2(y)\right) dy \\ &= \frac{1}{4}\int \mathbb{1}(x-y=0)\mathbb{1}(y=0)dy + \frac{1}{4}\int \mathbb{1}(x-y=0)\chi_1^2(y)dy \\ &+ \frac{1}{4}\int \mathbb{1}(y=0)\chi_1^2(x-y)dy + \frac{1}{4}\int \chi_1^2(x-y)\chi_1^2(y)dy. \end{aligned}$$

The resulting integration can be broken down into cases:

$$f_{X+Y}(x) = \begin{cases} 0, & \text{if } x = y = 0 \\ \frac{1}{4}\chi_1^2(x), & \text{if } x = y \\ \frac{1}{4}\chi_1^2(x), & \text{if } y = 0 \\ \frac{1}{4}\chi_2^2(x), & \text{if } x > y. \end{cases}$$

We can express as $f(x) = \frac{1}{4}\chi_0^2(x) + \frac{1}{2}\chi_1^2(x) + \frac{1}{4}\chi_2^2(x)$, a mixture of chi-square distributions. This distribution is for the case of three independent vectors of observations. We now continue

into the next case of 4 independent vectors.

$$\lambda_{\text{suprem}} = (R_{2m} + R_{3m}) + R_{4m}$$

Note the parenthesis since we proceed iteratively. Using the definition of convolution once again, we have:

$$\begin{aligned} f_{X+Y}(x) &= \int \frac{1}{2}\left(\chi_0^2(x-y) + \chi_1^2(x-y)\right)\frac{1}{4}\left(\chi_0^2(y) + \frac{1}{2}\chi_1^2(y) + \chi_2^2(y)\right)dy \\ &= \frac{1}{8}\int \mathbb{1}(x-y=0)\mathbb{1}(y=0)dy + \frac{1}{4}\int \mathbb{1}(x-y=0)\chi_1^2(y)dy \\ &+ \frac{1}{8}\int \mathbb{1}(x-y=0)\chi_2^2(y)dy + \frac{1}{8}\int \chi_1^2(x-y)\mathbb{1}(x-y=0)dy \\ &+ \frac{1}{4}\chi_1^2(x-y)\chi_1^2(y)dy + \frac{1}{8}\int \chi_1^2(x-y)\chi_2^2(y)dy, \end{aligned}$$

where the integral is once again represented into different cases:

$$f_{X+Y}(x) = \begin{cases} 0, & \text{if } x = y = 0 \\ \frac{1}{4}\chi_1^2(x), & \text{if } x = y, \\ \frac{1}{8}\chi_2^2(x), & \text{if } x = y, \\ \frac{1}{8}\chi_1^2(x), & \text{if } y = 0, \\ \frac{1}{4}\chi_2^2(x), & \text{if } x > y, \\ \frac{1}{8}\chi_3^2(x), & \text{if } x > y. \end{cases}$$

The convolution is expressed as

$$f_{X+Y}(x) = \frac{1}{8}\chi_0^2(x) + \frac{3}{8}\chi_1^2(x) + \frac{3}{8}\chi_2^2 x + \frac{1}{8}\chi_3^2(x). \tag{3.22}$$

For Figure 3.2, we simulated 1000 test statistics of the supremum under the reality of the null hypothesis. We have $n = 4$ time-points and examine the improvement of the λ_{sup} approximate asymptotic distribution as the length of each vector (m) increases from 10, 30, to 50. In Figure 3.2, there is a point mass at 0 and then a Chi-square curve from there on. For $m = 50$, we plot the asymptotic distribution in 3.22 over the histogram and notice a good approximation.

In Figure 3.3, we simulated 1000 test statistics of the average under the reality of the null hypothesis and the results are similar to our asymptotic expectations. As m increases the simulated distribution approaches a chi-square mixture distribution. The simulations support our approximate distributions for $m \geq 20$.

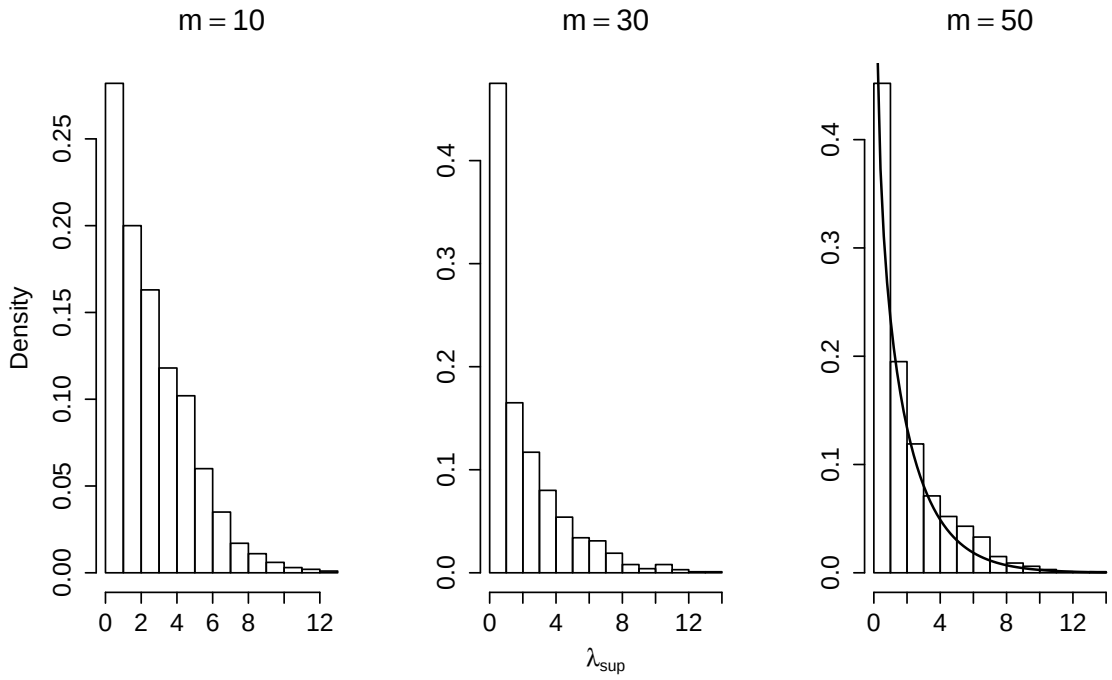

Figure 3.2: Simulated-based distribution of λ_{sup} for $m = 10, 30, 50$. Comparison of simulated-based distribution to the asymptotic distribution in $m = 50$.

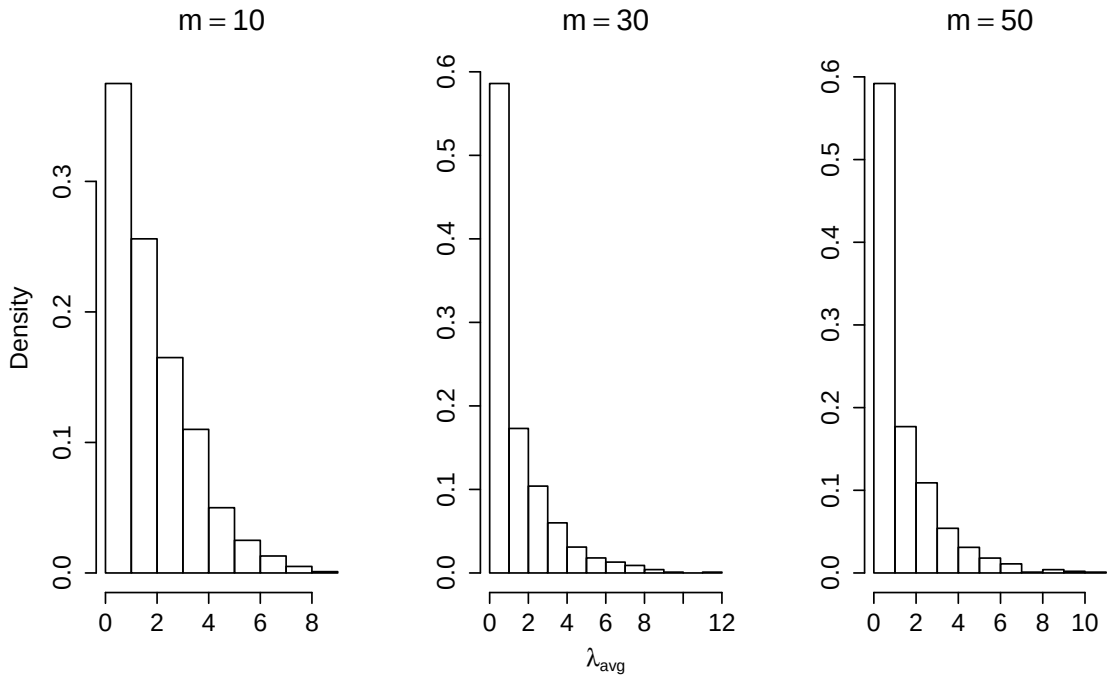

Figure 3.3: Simulated-based distribution of λ_{avg} $m = 10, 30, 50$. Illustrate as m increases an aymptotic distribution is obtained

The convolutions result in a mixture of Chi-square distributions with a growing number of terms as the number of vectors increase. As the number of vectors increase there are a larger number of terms to sum over which lead to Central Limit Theorem results.

Proposition 2 *A general result for positive integers k and d, $m \geq 20$. Then the asymptotic distribution for the supremum test statistic is mixture of Chi-square distribution with weights, number of terms, and degrees of freedom given by the following:*

$$\lambda_{suprem} = \sum_{i=2}^{n} R_{im} \xrightarrow{d} \sum_{i=0}^{n-1} \binom{n-1}{i} \left(\frac{1}{2}\right)^{n-1} \chi_i^2(x) \tag{3.23}$$

These results work well for sufficiently large amount of data m from each vector, in our case $m \geq 20$. The mixture can become quite involved for larger number of time-points n. Referring back to 3.17 and 3.18 are a sum of independent terms, where each term has asymptotic mixture Chi-square distribution. One would suspect that as n increases the Central Limit Theorem would give an approximate Normal distribution. The result would make the calculation of rejection region much more user friendly.

In Figure 3.4, as n increases the distribution of λ_{sup} approaches a Normal distribution. In each plot $m = 50$ and simulate the supremum test statistic for 1000 simulations. We then examine the distribution as we increase the number of time-points n from 8,10, to 12. As n increases the histogram of simulations becomes more and more symmetric and closer to a bell curve shape. For $n = 4$ in Figure 3.2, the distribution is clearly skewed since there is much weight on the point mass at 0 and we have chi-square distributions with 1 and 2 degrees of freedom in the mixture. For statistics with $n = 8$ and $n = 10$ the histogram is shifting away from skewness and becoming more symmetric. This is due to the fact of less weight on the point mass at 0 and an increase in the number of Chi-square terms with increasing degrees of freedom. For $n = 12$ there is a clear bell shaped curve and seems appropriate for a Normal approximation via the Central Limit Theorem. We plot the asymptotic distribution and the fit is close, but we recommend $n \geq 15$.

Proposition 3 *For sufficiently large m, $R_{2m}, \ldots, R_{nm}$ are independent and identically distributed with $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$ with mean 1 and variance 2. Then as $n \to \infty$,*

$$\frac{(\lambda_{sup} - (n-1))}{\sqrt{2(n-1)}} \xrightarrow{d} \Phi(z), \tag{3.24}$$

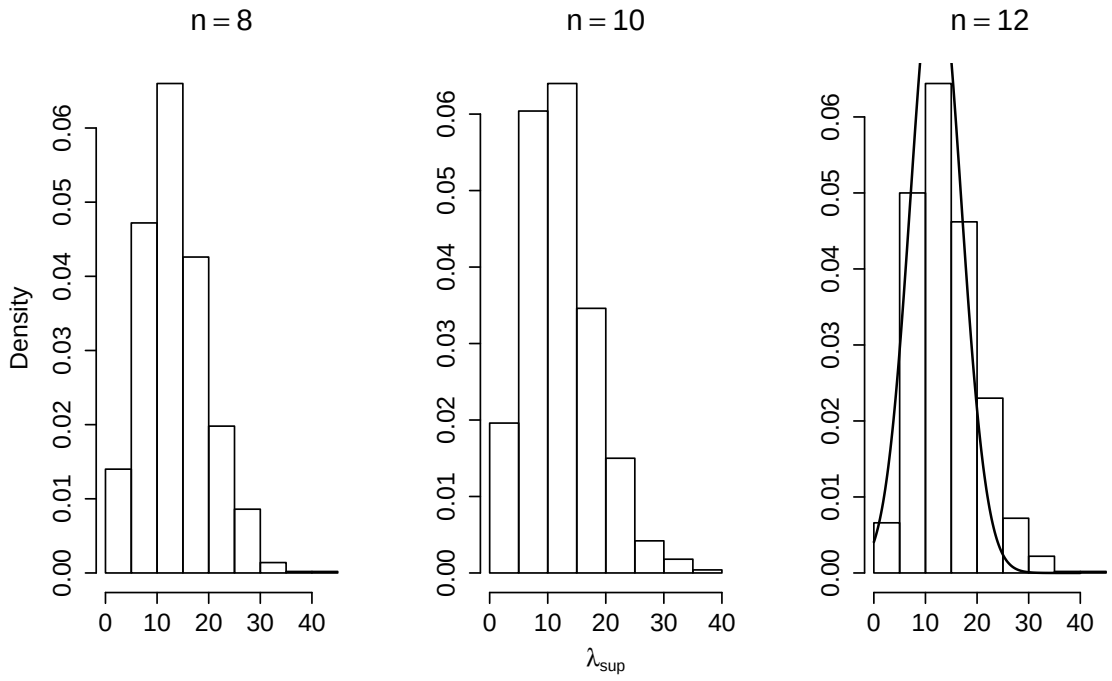

Figure 3.4: Simulation-based distribution for $n = 8, 10, 12$ versus the CLT asymptotic distribution of λ_{sup}

where $\Phi(z)$ *is the Standard Normal cumulative distribution function.*

We assume the process begins as a unimodal distribution, so we exclude the first time-point of data. For the case of $n = 12$, the asymptotic distribution is $N(11, 22)$ which close to the behavior of the simulations. In Figure 3.5, we examine the simulation-based distribution of λ_{avg} for $n = 8, 10, 12$ and note the density is approaching a Normal distribution shape. In each plot $m = 50$ and simulate the average test statistic for 1000 simulations.

For $m \geq 20$, we approximate the test statistics by a mixture of Chi-square distributions. And if both m and n are sufficiently large, we can simplify our approximation into a single Normal distribution by Central Limit theorem. If m is not large enough then we recommend to

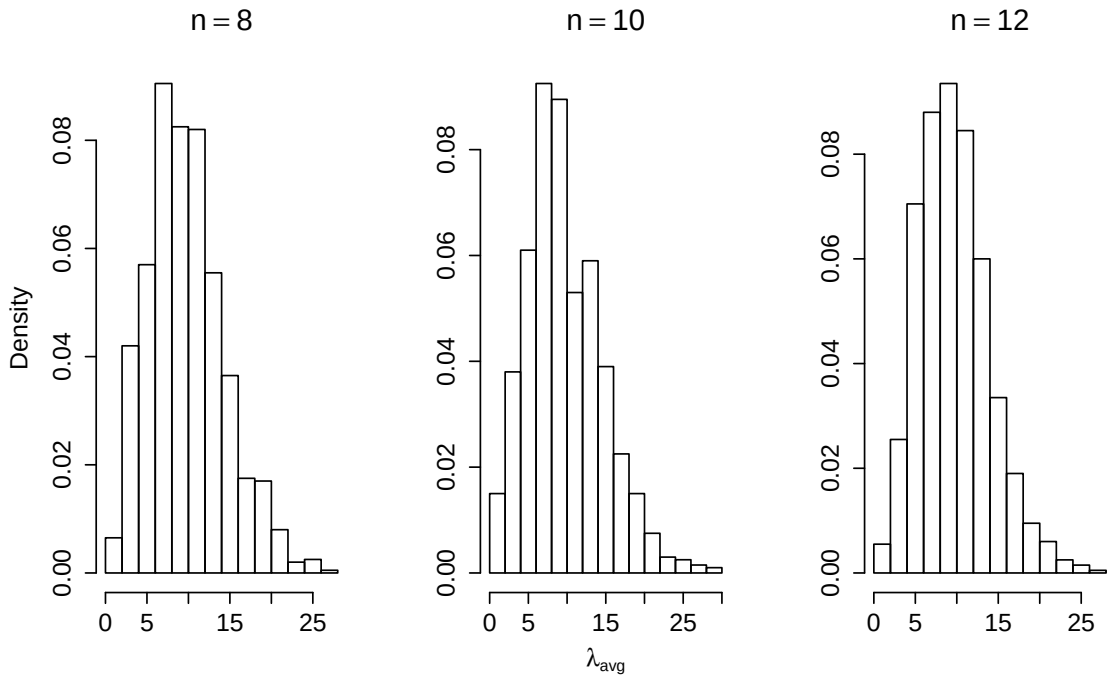

Figure 3.5: Simulation-based distribution for $n = 8, 10, 12$ of λ_{avg}. As n increases the density is approaching a Normal distribution.

find critical values through simulations. Due to the the set-up of our problem and the parameter spaces, the simulations must be performed for each set of data observed.

The exact distribution of test statistics (3.15) and (3.16) are unknown for small m, so the critical values are found through parametric bootstrap simulations. That is, given the data we find the MLEs for each vector of data under the unimodal parameter space. Then we use those MLEs as the parameters for our simulations for each vector of data. In a simple example in Table 3.3, we have a set of four independent vectors $\underset{\sim}{\alpha}_1, \underset{\sim}{\alpha}_2, \underset{\sim}{\alpha}_3, \underset{\sim}{\alpha}_4$, each with length 10 ($m = 10$), and in this case, the data consists of ten observations at each of the four time-points.

In Table 3.3, the critical values for three different significance levels for each of the average and supremum test statistics. The critical values were found by using the quantiles of simula-

Table 3.1: Parametric bootstrap critical values for λ_{sup} and λ_{avg} statistics for data in Table 3.3.

	Significance Level : γ		
	0.10	0.05	0.01
Test Statistic			
λ_{avg}	0.530	0.729	1.701
λ_{sup}	0.753	1.023	2.496

tions of the test statistics under the null hypothesis for two thousand iterations. In `R`, we use the function `rmixedvm()` found in the `circular` package to simulate a mixCN random variables. The first two vectors of data $(\underset{\sim}{\alpha}_1, \underset{\sim}{\alpha}_2)$ come from a unimodal distribution and the last two $(\underset{\sim}{\alpha}_3, \underset{\sim}{\alpha}_4)$ times come from a bimodal distribution. From Table 3.3, we see the first two vectors of data are centered around 2π. For the third vector there are two separate concentrations around 2π and $\pi/2$. For the last vector in time there are two separate concentrations around 2π and π.

Table 3.2: Small Sample Results with $n = 4$ and $m = 10$

	n			
	1	2	3	4
$m = 10$				
	0.655	0.880	1.340	1.333
	5.832	0.241	1.874	6.109
	0.151	0.867	1.699	0.002
	1.172	4.869	5.481	3.391
	5.540	5.380	0.580	3.913
	0.996	5.761	2.025	3.226
	5.683	0.107	2.090	2.936
	6.169	0.587	2.064	2.770
	0.792	5.790	1.239	3.977
	5.797	0.452	0.052	5.625

The observed test statistics for the data set were $\lambda_{sup} = 1.910$ and $\lambda_{avg} = 1.865$. For the average test would reject the null hypothesis for the significance levels $\gamma = 0.01, 0.05, 0.10$. The

supremum test rejects the null hypothesis for significance levels $\gamma = 0.10, 0.05$ with exception to $\gamma = 0.01$. This parametric bootstrap produces fast and reliable results for small sample sizes. Note that our test only tests for the existence of a change-point, whereas one also may concerned with the location of the change-point k. We attempt to address this issue via Bayesian methods in our next section.

Here is an outline on how to use GLRT for the change-point:

- If $m < 20$ and for any n, then estimate null distribution by simulating test statistic under the null hypothesis using parametric bootstrap methods. That is, simulate the test statistic according to the unimodal MLEs for each vector. Reject the null hypothesis if the test statistic is larger than the upper γ^{th} percentile of simulations, where γ is the significance level.

- If $m \geq 20$ and $n < 15$, then reject null hypothesis if $\lambda_{suprem} > \lambda_{suprem}(\gamma)$, where $\lambda_{suprem}(\gamma)$ is the upper γ^{th} percentile of the distribution in 3.23.

- If $m \geq 20$ and $n \geq 15$, then reject null hypothesis if $\lambda_{suprem} > \lambda_{suprem}(\gamma)$, where $\lambda_{suprem}(\gamma)$ is the upper γ^{th} percentile of the distribution in 3.24.

Chapter 4

A Bayesian Approach to detecting change

In contrast to what was discussed in the earlier Chapter 3 relying on the GLRT, we now take a completely different approach to detecting the change-point location k. Specifically, we present a purely Bayesian approach that relies on computational tools such as the Markov Chain Monte Carlo (MCMC). We specify prior distributions on each the parameters, including k, and find the posterior distribution of the change-point k via the Metropolis-Hastings algorithm. In Section 4.1, we first review a Bayes test for bimodality from (Basu & Jammalamadaka, 2000). We then describe MCMC in Section 4.2 and in 4.3 we briefly review the Metropolis-Hastings (M-H) algorithm. We briefly review the Metropolis-Hastings algorithm followed by an application of the general theory to our situation. A good introduction to Metropolis-Hastings algorithm can be found in (Chib & Greenberg, 1995). In Section 4.4 we apply M-H algorithm to estimate the location of the change-point.

4.1 A Bayes Test for Unimodality

As stated before, our goal is to consider a change-point problem but where the change-point occurs in the number of modes of the true distribution from which the data is drawn. In (Basu & Jammalamadaka, 2000), a Bayes test for bimodality for circular data is presented, which we now review.

Let $\alpha_1, \alpha_2, \ldots, \alpha_n$ be i.i.d. observations from the circular density $f(\alpha)$. We want to test $H_0 : f(\alpha)$ is unimodal versus $H_1 : f(\alpha)$ is not unimodal. A Bayes test is proposed based on the observed data, which are assumed to come from a mixCND as in (1.15). The test uses a prior distribution on the parameters. Independent priors are assumed for the parameters $\mu_1, \mu_2, \kappa_1, \kappa_2$, and p. The posterior probability of $f(\alpha)$ being unimodal is then compared to the prior probability of $f(\alpha)$ being unimodal. The probabilities are computed using Markov Chain Monte Carlo sampling where the simulations used the densities described below.

The model structure for this procedure proceeds as follows:

- We observe circular observations $\alpha_1, \ldots, \alpha_n$ i.i.d. from density

$$f(\alpha) = pCN(\alpha|\mu_1, \kappa_1) + (1-p)CN(\alpha|\mu_2, \kappa_2)$$

- The likelihood for the observed data is:

$$L(p, \mu_1, \mu_2, \kappa_1, \kappa_2|\underset{\sim}{\alpha}) = \prod_{i=1}^{n} pCN(\alpha_i|\mu_1, \kappa_1) + (1-p)CN(\alpha_i|\mu_2, \kappa_2)$$

- The prior for μ_j is $p(\mu_j) = CN(\nu_j, \tau_j)$, $j = 1, 2$.

- The prior for κ_j is $p(\kappa_j) = \text{Gamma}(\nu_j, \tau_j)$, $j = 1, 2$.

- The mixing proportion p has a Uniform[0,1] prior distribution.

The choice of these prior distributions for mean directions and concentration parameters are commonly chosen. The CN serves as a conjugate prior for the mean direction since the CND is a member of the exponential family. There is no conjugate prior for the concentration parameter κ, but since it takes non-negative values on the real line, a Gamma density provides a good and flexible prior. As for the the mixing proportion, the Uniform prior density is used to reflect maximum uncertainty. We also assume that these prior distributions are independent.

From the data and prior distributions a test is derived to determine which model best fits the data. The Bayes factor is used to test H_1 against H_0 given the data and is defined as:

$$B_{10} = \frac{\text{Posterior Odds}}{\text{Prior Odds}} = \frac{\mathbb{P}(H_1|\text{data})\mathbb{P}(H_0)}{\mathbb{P}(H_0|\text{data})\mathbb{P}(H_1)} \tag{4.1}$$

We would reject H_0 for large values of the Bayes factor in (4.1). A table which lists the strength of evidence against H_1 is found in (Basu & Jammalamadaka, 2000).

The prior probability of unimodality is the integral of the joint prior density $f(\mu_1, \mu_2, \kappa_1, \kappa_2, p)$ over the region where the parameter space gives a unimodal density for the mixture of two CNDs. The region of the parameter space Ω_0 over which unimodality holds, is given earlier in Table 1.1. Specifically,

$$\mathbb{P}(H_0) = \int_{\Omega_0} f(\mu_1, \mu_2, \kappa_1, \kappa_2, p)\mathrm{d}\omega \tag{4.2}$$

Since the prior densities are assumed independent, the joint prior distribution is the product of five prior densities. However, the integral over the unimodal parameter space makes the integral analytically intractable. Instead a Monte Carlo method was used to estimate $\mathbb{P}(H_0)$ and the method is outlined below:

(i) Let $\underset{\sim}{\phi} = f(\mu_1, \mu_2, \kappa_1, \kappa_2, p)$, then generate i.i.d. samples $\{\underset{\sim}{\phi}^{(t)} : t = 1, \ldots, T_1\}$ from the joint prior distribution of $\underset{\sim}{\phi}$.

(ii) For each generated $\underset{\sim}{\phi}^{(t)}$, we check if the mixture density is unimodal using the Mardia-Sutton condition.

(iii) An estimate for $\mathbb{P}(H_0)$ equals {Number of generated samples $\underset{\sim}{\phi}^{(t)}$ in which $f(\alpha)$ was unimodal}$/T_1$.

Due to independence, simulating from the joint prior density reduces to simply simulate component-wise densities at each step.

Next, we outline the calculation for the posterior probability of unimodality:

$$\mathbb{P}(H_0|\text{data}) = \int_{\Omega_0} f(\mu_1, \mu_2, \kappa_1, \kappa_2, p|\text{data})\text{d}\omega, \tag{4.3}$$

namely the posterior joint distribution integrated over the unimodal parameter space. The procedure to estimating this probability would be the analogous to before in steps (i) and (ii). However, direct simulation from $f(\mu_1, \mu_2, \kappa_1, \kappa_2, p|\text{data})$ is difficult since the joint distribution is analytically intractable. So steps (ii) and (iii) remain the same as before, but step (i) of the Monte Carlo sampling, we will use MCMC sampling for step (i). We will not outline the Gibbs Sampling technique but the gist of the algorithm is to simulate alternately and iteratively for the conditional posterior distributions of each unobservable given the data and other observables. See for instance (Muralidharan & Parikh, 2012) for one example of the Gibbs Sampler. There is no common agreement on how to derive the full conditionals needed for the Gibbs sampler in the mixCN context so instead of Gibbs sampling, we use a Metropolis-Hastings approach.

4.2 Markov Chain Monte Carlo Simulation

The usual approach to Markov chain theory on a continuous state space is to start with a transition kernel $P(x, A)$ for $x \in \mathbb{R}^d$ and $A \in \mathscr{B}$, where $\mathscr{B}$ is the Borel σ-field on $\mathbb{R}^d$. The transition kernel is a conditional distribution function that represents the probability of moving from x to a point on the set A. This transition kernel is a probability distribution function such that $P(x, \mathbb{R}^d) = 1$ and $P(x, x)$ is not necessarily zero.

In Markov chain theory, we consider conditions needed to show the existence of an invariant distribution π^* and conditions for the iterations of the transition kernel to converge to the invariant distribution. The invariant distribution satisfies

$$\pi^*(dy) = \int_{\mathbb{R}^d} P(x, dy)\pi(x)dx \tag{4.4}$$

where π is the density with respect to the Lebesgue measure of π^*. The nth iterate is given by $P^{(n)}(x, A) = \int_{\mathbb{R}^d} P^{(n-1)}(x, dy)P(y, A)$, where $P^{(1)}(x, dy) = P(x, dy)$. Under some regularity conditions, it can be shown that the nth iterate converges to the invariant distribution as $n \to \infty$.

MCMC methods work in the opposite direction using the known invariant distribution (perhaps up to a constant multiple). Here $\pi(.)$, is the target density from which samples are desired. To generate samples from $\pi(.)$, MCMC methods find and utilize a transition kernel $P(x, dy)$ whose nth iterate converges to $\pi(.)$ for large n. The process is then started for some value of x and iterated a large number of times. After a sufficient amount of iterations the generated observations approximates the target distribution.

Suppose that the transition kernel, for some function $p(x, y)$ can be expressed as

$$P(x, dy) = p(x, y)dy + r(x)\delta_x(dy) \tag{4.5}$$

where $p(x, x) = 0$, $\delta_x(dy) = 1$ if $x \in dy$ and 0 otherwise, and $r(x) = 1 - \int_{\mathbb{R}^d} p(x, y)dy$ is the probability that the chain remains at x. Note that since the chain can remain at x, the integral of $p(x, y)$ with respect to y does not necessarily equate to 1. Now we assume that $p(x, y)$ in (4.5) satisfies the reversibility condition:

$$\pi(x)p(x, y) = \pi(y)p(y, x) \tag{4.6}$$

with this property we say $\pi(.)$ is the invariant density of $P(x, .)$. The verification of the result is given by,

$$\begin{aligned}
\int P(x, A)\pi(x)dx &= \int \left[\int_A p(x, y)dy\right] \pi(x)dx \\
&+ \int r(x)\delta_x(A)\pi(x)dx \\
&= \int \left[\int_A p(x, y)\pi(x)dx\right] dy \\
&+ \int_A r(x)\pi(x)dx \\
&= \int \left[\int_A p(y, x)\pi(y)dx\right] dy \\
&+ \int_A r(x)\pi(x)dx \\
&= \int_A (1 - r(y))\pi(y)dy + \int_A r(x)\pi(x)dx \\
&= \int_A \pi(y)dy.
\end{aligned}$$

In the reversibility equation, 4.6, $p(x, y)$ is the unconditional probability to move from x to y, when x is generated from $\pi(.)$. Also, $p(y, x)$ is the unconditional probability to move from x

to y, when y is generated from $\pi(.)$. By reversibility the two sides are equal and thus π^* is the invariant distribution for $P(.,.)$. This result provides a sufficient condition for $p(x,y)$ and next we will demonstrate how a specific Metropolis-Hastings algorithm finds such a $p(x,y)$.

4.3 Metropolis-Hastings Algorithm

The Metropolis-Hastings (M-H) algorithm is akin to the Acceptance-Rejection sampling method for generating independent samples, but since how we are simulating dependent Markov chains, the density will depend on the prior point in the chain's state. We begin with the candidate-generating density denoted $q(x,y)$, where $\int q(x,y) = 1$. The density generates a value y when the process is currently at point x. If $q(x,y)$ satisfies 4.6 for all x and y, then our search is complete. This is the unlikely outcome and we may have for some x, y:

$$\pi(x)q(x,y) > \pi(y)q(y,x). \tag{4.7}$$

.

In this special case, the process will move from x to y too often, and from y to x too rarely. An easy correction to reduce the over abundance of moves from x to y is to introduce a probability $\alpha(x,y) < 1$ that the move is made. If a move is not made the process remains at point x. Moves from x to y are determined by:

$$p_{MH}(x,y) = q(x,y)\alpha(x,y) \quad x \neq y.$$

If 4.7 holds, $\alpha(y,x)$ is set to equal one since moves from y to x are not made often enough. And $\alpha(x,y)$ is determined such that $p_{MH}(x,y)$ satisfies the aforementioned reversibility condition i.e.,

$$\pi(x)q(x,y)\alpha(x,y) = \pi(y)q(y,x)\alpha(y,x)$$
$$= \pi(y)q(y,x)$$

From this result $\alpha(x,y) = \pi(y)q(y,x)\alpha(y,x)/\left(\pi(x)q(x,y)\right)$, in order for $p_{MH}(x,y)$ to satisfy the reversibility condition. Therefore,

$$\alpha(x,y) = \min\left[\frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}, 1\right], \qquad \text{if } \pi(x)q(x,y) > 0, \tag{4.8}$$

$$= 1, \qquad \text{otherwise.} \tag{4.9}$$

Next we consider the possibility of the process to remain at point x. From our definitions, we have the probability to remain at point x given by,

$$r(x) = 1 - \int_{\mathbb{R}^d} q(x,y)\alpha(x,y)dy.$$

Then the transition kernel of the M-H chain, $P_{MH}(x,dy)$, is defined as,

$$P_{MH}(x,dy) = q(x,y)\alpha(x,y)dy + \left[1 - \int_{\mathbb{R}^d} q(x,y)\alpha(x,y)dy\right]\delta_x(dy),$$

This is a particular case of 4.5 and since $p_{MH}(x,y)$ is reversible by construction, the M-H kernel has $\pi(x)$ as its invariant density. (Note, we can use similar proof as in the previous general case). Note, if the candidate value is rejected, then the current value remains as the process continues into the next step. Also, if the candidate-generating density is symmetric, i.e. $q(x,y) = q(y,x)$, then the probability of a move becomes $\pi(y)/\pi(x)$. Another useful fact is the

calculation of $\alpha(x, y)$ does not require the normalizing constant of $\pi(.)$ because this normalizing constant appears in both numerator and denominator.

As with any MCMC method, the chain takes time until the the transient (burn-in) stage is passed. Meaning initial values generated from $\pi(x)$ are discarded until the process has converged to the invariant distribution. Under mild regularity conditions the chain is guaranteed to converge if run infinitely long, but the selection of $q(x, y)$ will determine the rate of convergence. In (Chib & Greenberg, 1995), presents are five methods in selecting $q(x, y)$, but we mention the one used for our study that was first introduced by (Metropolis *et al.* , 1953).

The family of generating densities we use are specified such that $q(x, y) = q_1(y - x)$, where candidate y is drawn according to process $y = x + z$. The candidate y is equal to the current value x, plus some white noise z. If q_1 is selected such that $q(z) = q(-z)$,then the probability of moving from x to y reduces to

$$\alpha(x, y) = \min\left[\frac{\pi(y)}{\pi(x)}, 1\right]. \tag{4.10}$$

This is referred as the 'random walk' chain. Common choices for q_1 include Normal distributions with mean equal to the current value x and variance selected such that we have a favorable acceptance rate of approximately .45 as proposed by (Roberts *et al.* , 1997). The scale parameter of the candidate-generating density determines the acceptance rate and if the parameter space is fairly covered , i.e. our candidates can come from anywhere in the parameter space. If the scale parameter is relatively large, then generated values have high probability of being far from current value resulting in low probability of acceptance. If the scale is too small, the chain will take longer to reach approximate convergence and spaces with low probability will be under-sampled. The scale parameter is hence called the tuning parameter of our chain. A

reduction of scale parameter in the former case, and an increase in scale parameter, will remedy the issues.

One can construct an M-H algorithm, but care is needed when adjusting tuning parameters for each set of data. (Roberts & Rosenthal, 2009) provide examples of how to make the M-H algorithm adapt to different scenarios of data. In one example, they use the aforementioned random walk M-H algorithm and periodically check the acceptance rate for the previous 50 iterations of the chain. If the acceptance rate is below the desired .45, then the adaptive chain will increase the tuning parameter, and will decrease if acceptance rate above .45. This adaptive tuning is performed until the process has traversed the transient state. With this detailed introduction to the M-H algorithm, we will now apply the method in detecting the change-point.

4.4 M-H Algorithm for Posterior Distribution of The Change-Point

As mentioned earlier, the likelihood ratio test will determine if a change-point exists, but is not specifically focused on pinpointing the location of that change. In this section, we use a M-H algorithm to generate values from the posterior distribution of the change-point k. These generated values will give the scientist a posterior probability distribution for the location of the change-point. We now make one subtle but important change from our assumptions of Chapter 3 where we used the GLRT approach. For the change-point k, we now assume $\underset{\sim}{\alpha}_1 \ldots \underset{\sim}{\alpha}_k$ are i.i.d. vectors of observations i.e. we assume $\underset{\sim}{\alpha}_1 \ldots \underset{\sim}{\alpha}_k$ follow the same unimodal distribution while

the remaining $\underset{\sim}{\alpha}_{k+1} \ldots \underset{\sim}{\alpha}_n$ vectors are i.i.d. from correspond to the same bimodal distribution i.e. there are two sets of parameter values one before change and one after the change. This avoids the multiple parameter vectors, one for each step $j \in \{1, \ldots, n\}$, that were allowed in the earlier GLRT set-up of Chapter 3.

Before we present our algorithm and the results we introduce the steps involved.

Let $\boldsymbol{\theta}_u = (p_u, \mu_u, \delta_u, c_u, \kappa_u)$ and $\boldsymbol{\theta}_b = (p_b, \mu_b, \delta_b, c_b, \kappa_b)$ represent the parameter vectors under the unimodal and bimodal parameter spaces. We wish to simulate from the posterior of the change-point distribution k which is given by,

$$\pi(k|\boldsymbol{\theta}_u, \boldsymbol{\theta}_b, \boldsymbol{\alpha}) = \frac{f_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}|\boldsymbol{\theta}_u, \boldsymbol{\theta}_b)\pi(\boldsymbol{\theta}_u)\pi(\boldsymbol{\theta}_b)\pi(k)}{\int\int\int f_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}|\boldsymbol{\theta}_u, \boldsymbol{\theta}_b)\pi(\boldsymbol{\theta}_u)\pi(\boldsymbol{\theta}_b)\pi(k)d\boldsymbol{\theta}_u d\boldsymbol{\theta}_b d\boldsymbol{\alpha}} \tag{4.11}$$

The constant in the denominator of posterior is not needed to sample from the posterior of k when using the M-H algorithm. In this approach, we use a different re-parametrization compared with the GLRT to avoid identifiability issues, since we wish to recover all parameters involved including the change-point k. For example, with $\delta = |\mu_1 - \mu_2|$, suppose the triple (δ, κ, p) gives a unimodal distribution, then any (μ_1, μ_2) with difference δ will give the same unimodal shape but have shifted centers. We can recover the true values of the pair (μ_1, μ_2) if we re-express our mixCND in 1.15 as,

$$f_\alpha(\alpha|\boldsymbol{\theta}) = pCN(\mu + c\delta, \kappa) + (1-p)CN(\mu, \kappa), \tag{4.12}$$

where $0 < p < 1/2$, $\kappa > 0$, $0 \leq \mu < 2\pi$, $0 \leq \delta \leq \pi$, and $c = \{1, -1\}$. This parametrization allows for easy reference in checking modality and for recovering the values of all parameters involved. For a given value of κ, Figure 4.1, shows that for a given value κ, the parameter space

is symmetric around the horizontal line $p = 1/2$. By symmetry we can restrict $0 < p < 1/2$, and thus speed up computations. Recall, area within the wineglass-shaped curve is the bimodal parameter space and outside the wineglass shape is the unimodal parameter space. An increase in the value of κ, will increase the area within the wineglass region.

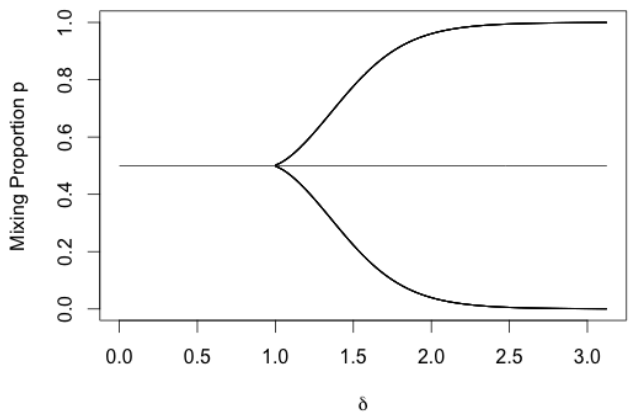

Figure 4.1: Given κ, Unimodal and Bimodal Parameter from Table 1.1 is Symmetric Around $p = 1/2$.

The symmetry in the parameter space allows the restriction on p. Also, $\mu_1 = \mu + c\delta$ and $\mu_2 = \mu$ as in 1.15, gives the location of (μ_1, μ_2). Where δ gives the absolute difference between the mean directions and c gives the direction of the difference. Also in 1.15, $\mu_1 = \mu + c\delta$ and $\mu_2 = \mu$ in 4.12. The parameter c takes value either -1 or 1, depending if $\mu_1 - \mu_2$ is a negative or positive difference.

We now introduce the probability distributions and the priors used.

- $f_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}|\boldsymbol{\theta}_u, \boldsymbol{\theta}_b) = \prod_{j=1}^{k} \text{mixCN}(\underset{\sim}{\alpha}_j|\boldsymbol{\theta}_u) \prod_{j=k+1}^{n} \text{mixCN}(\underset{\sim}{\alpha}_j|\boldsymbol{\theta}_b)$

  where $\text{mixCN}(\underset{\sim}{\alpha}_j|\boldsymbol{\theta}_u) = p_u CN(\mu_u + c\delta_u, \kappa_u) + (1 - p_u)CN(\mu_u, \kappa_u)$,

  and similarly for $\text{mixCN}(\underset{\sim}{\alpha}_j|\boldsymbol{\theta}_b)$

- $\pi(\boldsymbol{\theta}_u)$ Uniform distribution on unimodal space given by Table 1.1
- $\pi(\boldsymbol{\theta}_b)$ Uniform distribution on bimodal space given by Table 1.1
- $\pi(k)$ discrete Uniform distribution on values $k = 1, \ldots, n$.

Our uniform priors are indication that we lack prior information of form of unimodal/bimodal parameter spaces or where the change-point location may be. Now we will provide an outline for the M-H algorithm. (Note that superscript in $\boldsymbol{\theta}^{(c)}$ represents candidate and $c^{(t+1)}, c_u, c_b$ denote the sign of difference between μ_1 and μ_2. The c has two different meanings and we list to avoid any confusion.) Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_u, \boldsymbol{\theta}_b, k)$ and we use symmetric candidate generating functions therefore the probability of process to move is as in 4.10, where $\alpha(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^{(c)})$ is given by,

$$\frac{\prod_{j=1}^{k} \text{mixCN}(\underset{\sim}{\alpha}_j|\boldsymbol{\theta}_u^{(c)}) \prod_{j=k+1}^{n} \text{mixCN}(\underset{\sim}{\alpha}_j|\boldsymbol{\theta}_b^{(c)})|\text{J}^{(c)}|}{\prod_{i=1}^{k} \text{mixCN}(\underset{\sim}{\alpha}_j|\boldsymbol{\theta}_u^{(t)}) \prod_{i=k+1}^{n} \text{mixCN}(\underset{\sim}{\alpha}_j|\boldsymbol{\theta}_b^{(t)})|\text{J}^{(t)}|} \tag{4.13}$$

where $|\text{J}^{(t)}|$ and $|\text{J}^{(c)}|$ are the determents of the Jacobian matrices. $\boldsymbol{\theta}^{(t)}$ and $\boldsymbol{\theta}^{(c)}$ represent the current and candidate value of parameter vector $\boldsymbol{\theta}$. The need for the Jacobian matrices is due to the fact that we generate functions of some parameters that ensure we have a symmetric generating function. The list of generating functions for $t \in \{1, \ldots, N\}$ is given by,

- $\omega^{(c)} \sim \text{N}(\text{logit}(2p^{(t)}), \sigma_\omega^2)$, thus $p^{(c)} = \frac{\exp \omega^{(c)}}{2(1+\exp \omega^{(c)})^2}$
- $\mu^{(c)} \sim \text{CN}(\mu^{(t)}, \tau_\mu)$
- $\psi^{(c)} \sim \text{CN}(2\delta^{(t)}, \tau_\psi)$, thus $\delta^{(c)} = \frac{\psi^{(c)}}{2}$

- $c^{(t+1)} = 1, w.p. \propto \prod_{i=1}^{k} \text{mixCN}(\underset{\sim}{\alpha}_j|\boldsymbol{\theta}^{(t)}(c=1))|\text{J}^{(t)}|$

  $= -1, w.p. \propto \prod_{i=1}^{k} \text{mixCN}(\underset{\sim}{\alpha}_j|\boldsymbol{\theta}^{(t)}(c=-1))|\text{J}^{(t)}|$

- $\eta^{(c)} \sim \text{N}(\log(\kappa^{(t)}), \sigma_\kappa^2)$, thus $\kappa^{(c)} = \exp \eta^{(c)}$

- $k^{(t+1)}$ is sampled from discrete distribution on $\{1, \ldots, n\}$ w.p.

  $\mathbb{P}(K = i) \propto \prod_{j=1}^{i} \text{mixCN}(\underset{\sim}{\alpha}_j|\boldsymbol{\theta}_u^{(t)}) \prod_{j=i+1}^{n} \text{mixCN}(\underset{\sim}{\alpha}_j|\boldsymbol{\theta}_b^{(t)})|\text{J}^{(t)}|$ where $i = 1, \ldots, n$.

The first five generating functions are in general form, but they are used for both the unimodal and bimodal parameter sets. Each candidate value of $\theta^{(c)}$ which are $\omega^{(c)}, \mu^{(c)}, \psi^{(c)}, \eta^{(c)}$, are generated from a distribution that is centered around the current value, $\theta^{(t)}$. The scale parameters $\sigma_\omega^2, \tau_\mu, \tau_\psi$ and σ_κ^2 are the tuning parameters of our algorithm.

- Repeat for $t = 1, \ldots, N$

- Begin with $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\theta}_u^{(0)}, \boldsymbol{\theta}_b^{(0)}, k)$, where parameters are randomly selected within their respective parameter spaces.

- Generate $p^{(c)}$

  - If $p_u^{(c)}$ falls in the unimodal parameter space (given $\delta_u^{(c)}, \kappa_u^{(c)}$) continue to (*), otherwise $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)}$

    * Generate $u \sim U(0,1)$, if $u \leq \alpha(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^{(c)})$

      —set $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(c)}$

    * Else $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)}$

- Generate $\mu^{(c)}$ and $u \sim U(0,1)$

- If $u \leq \alpha(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^{(c)})$

  —set $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(c)}$

- Else $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)}$

- Generate $\delta^{(c)}$

  - If $\delta_u^{(c)}$ falls in the unimodal parameter space (given $p_u^{(c)}, \kappa_u^{(c)}$) continue to (*), otherwise $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)}$

    * Generate $u \sim U(0,1)$, if $u \leq \alpha(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^{(c)})$

      —set $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(c)}$

    * Else $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)}$

- Generate $c^{(t+1)}$ according to aforementioned sampling technique.

- Generate $\kappa^{(c)}$

  - If $\kappa_u^{(c)}$ falls in the unimodal parameter space (given $p_u^{(c)}, \delta_u^{(c)}$) continue to (*), otherwise $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)}$

    * Generate $u \sim U(0,1)$, if $u \leq \alpha(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^{(c)})$

      —set $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(c)}$

    * Else $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)}$

- Repeat for parameters in $\boldsymbol{\theta}_b$, but check if triplet $(p_b, \delta_b, \kappa_b)$ are bimodal.

- $k^{(t+1)}$ is sampled from aforementioned sampling technique.

- Increment t and repeat. Return the values for $t \in \{1, \ldots, N\}$, $\{\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots, \boldsymbol{\theta}^{(N)}\}$

When generating the values for δ, κ and p, we need to check if we are in the unimodal or bimodal parameter space before we continue with the usual M-H acceptance-rejection step. In our first example (Simulation 1) we simulate a scenario with $\underset{\sim}{\alpha}_1, \ldots, \underset{\sim}{\alpha}_8$, i.e. 8 vectors of data each with 20 observations. Here the first 7 vectors of data come from a mixture with $\boldsymbol{\theta}_u = (p_u = .7, \mu_u = \pi/2, \delta_u = \pi/8, \kappa_u = 1)$ and the last with $\boldsymbol{\theta}_b = (p_b = .65, \mu_b = \pi/8, \delta_b = \pi, \kappa_b = 5)$. The density of the unimodal and the bimodal density is given in Figure 4.2, showing that both these unimodal and bimodal densities are far from each other in shape and center of mass.

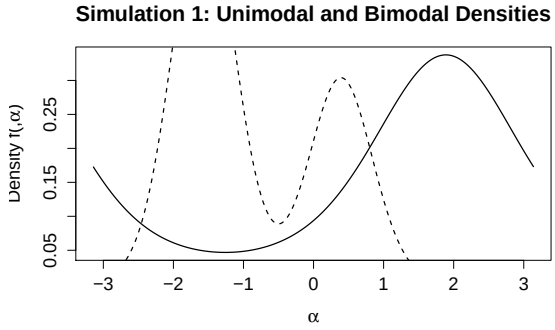

Figure 4.2: Densities of Unimodal and Bimodal Mixtures used in Simulation 1

We ran our M-H algorithm for $N = 100,000$ simulations and the trace plots for most parameters are given in the Figures 4.3, 4.4, and 4.5. For each parameter we appear to have approximately converged to the stationary univariate distribution given by the trace plots in Figure 4.4. For the unimodal trace plots in Figure 4.5 we have satisfactory results. In the trace plot for μ_u it seems there are jumps to larger values but this due to the alternating value from 1 to -1 for parameter c_u, thus $\pi/8$ will become $(2\pi - \pi/8)$, which are equivalent. The unimodal distributions are very similar so this alteration in c_u does not disrupt our process. The trace plot **??** for the change-point is most impressive as the posterior distribution concentrates at $k = 7$ very quickly and remains there for the remainder of the iterations. The value is correct

as the first 7 vectors of data come from a unimodal distribution and the last vector comes from a bimodal distribution.

In our next example (Simulation 2) we simulate from unimodal and bimodal distributions more similar in terms of shape and center of mass. In Figure 4.6 we have the densities of unimodal and bimodal distributions used for simulation. Here $\boldsymbol{\theta}_b = (p_b = .3, \mu_b = \pi/4, \delta_b = \pi/2, \kappa_b = 4)$ and $\boldsymbol{\theta}_u = (p_u = .3, \mu_u = 0, \delta_u = 3\pi/8, \kappa_u = 2)$. The distributions have approximately the same center and the bimodal density is not very pronounced, meaning the bimodal parameter values are near the unimodal boundary. Here we simulate 8 vectors of data again where the first three are simulated under the unimodal density and the last five under the bimodal density ($k = 3$).

In Figure 4.7, the trace plot fluctuates for k between $k = 3$ and $k = 4$, with slightly more emphasis on $k = 3$. This is not the exact value but still a good approximation given that data were simulated from two very similarly shaped distributions. In Figures 4.8 and 4.9 we have the trace plots for the unimodal and bimodal parameters. The bimodal parameters converge quite well, and in the unimodal parameter set have some jumps in μ_u which is caused by parameter c_u.

Overall, this method works very well with the additional but reasonable assumption of identically distributed unimodal and bimodal distributions. If this assumption does not hold, then the GLRT procedure discussed in the earlier Chapter is a more suitable test.

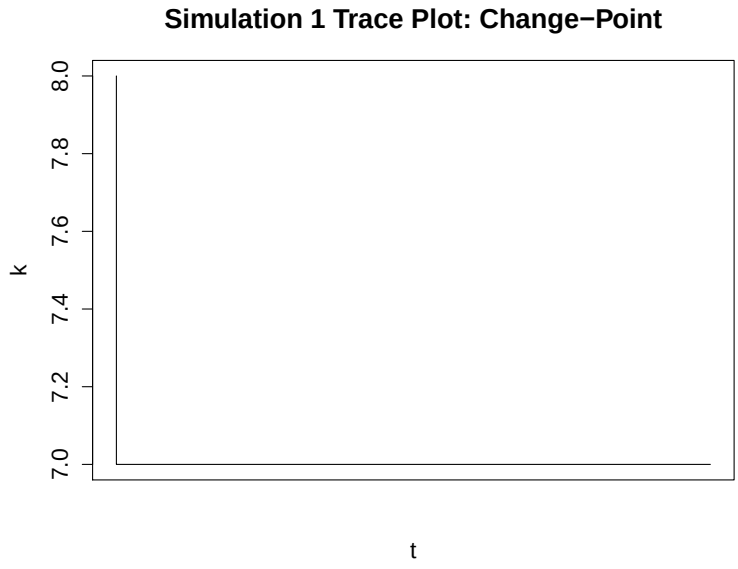

Figure 4.3: Trace Plot for Change-Point k in Simulation 1

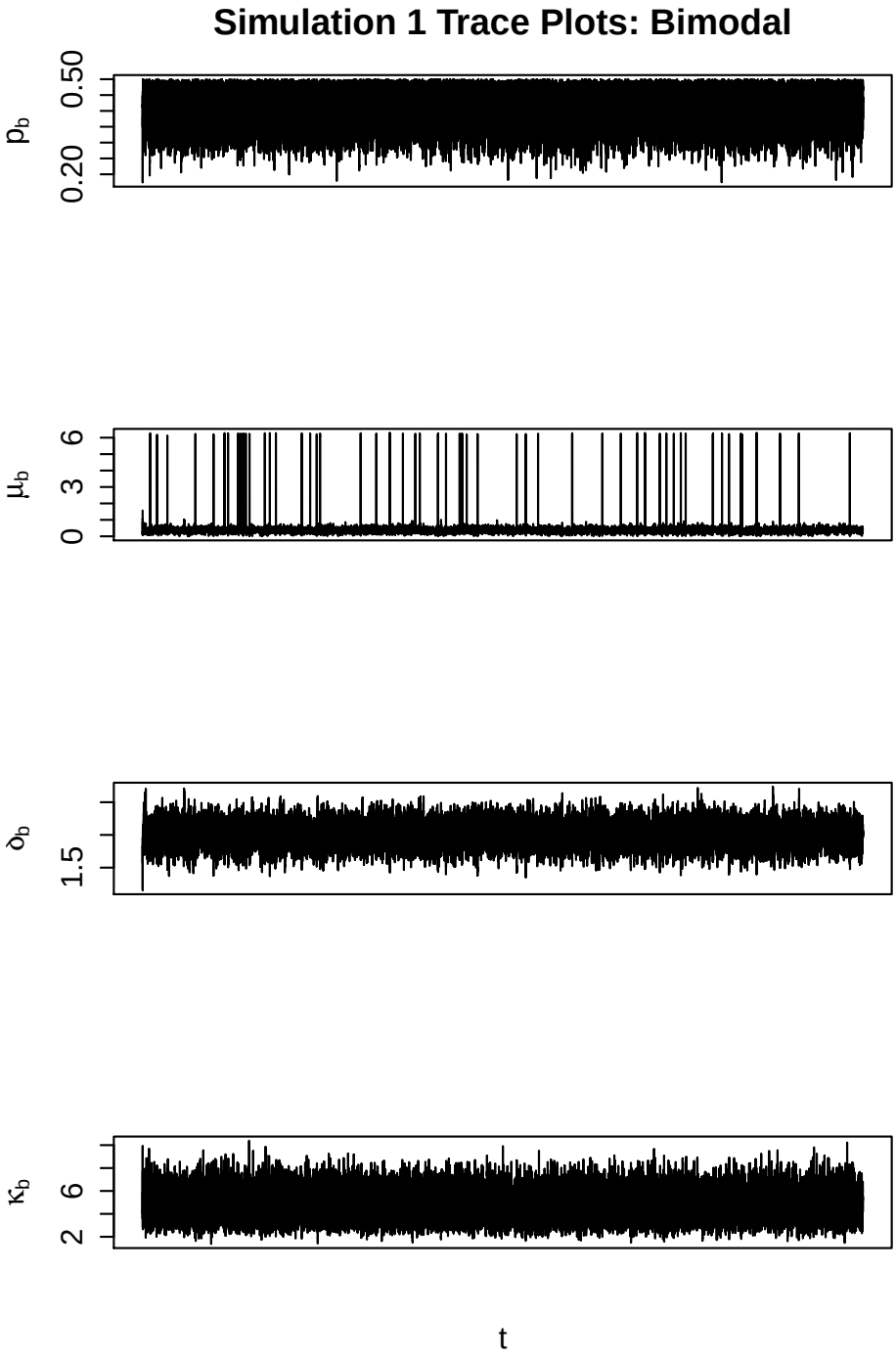

Figure 4.4: Trace Plots for $\boldsymbol{\theta}_b = (p_b = .65, \mu_b = \pi/8, \delta_b = \pi, \kappa_b = 5)$ in Simulation 1

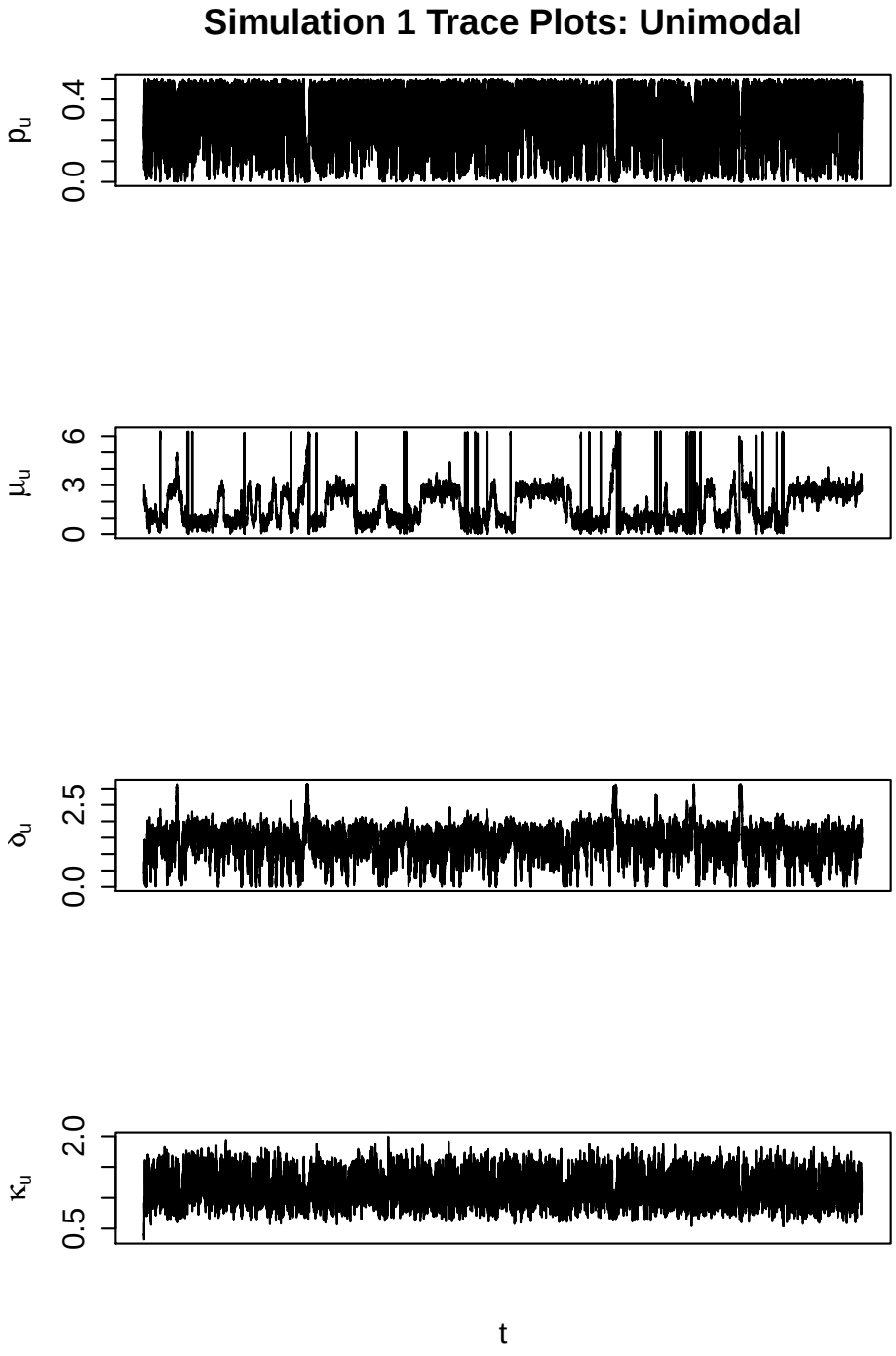

Figure 4.5: Trace Plots for $\boldsymbol{\theta}_u = (p_u = .7, \mu_u = \pi/2, \delta_u = \pi/8, \kappa_u = 1)$ in Simulation 1

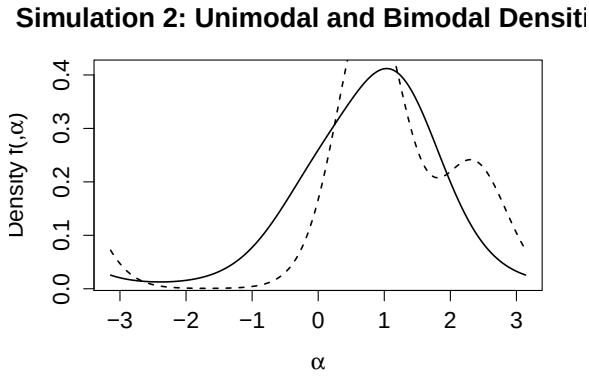

Figure 4.6: Densities of Unimodal and Bimodal Mixtures in Simulation 2

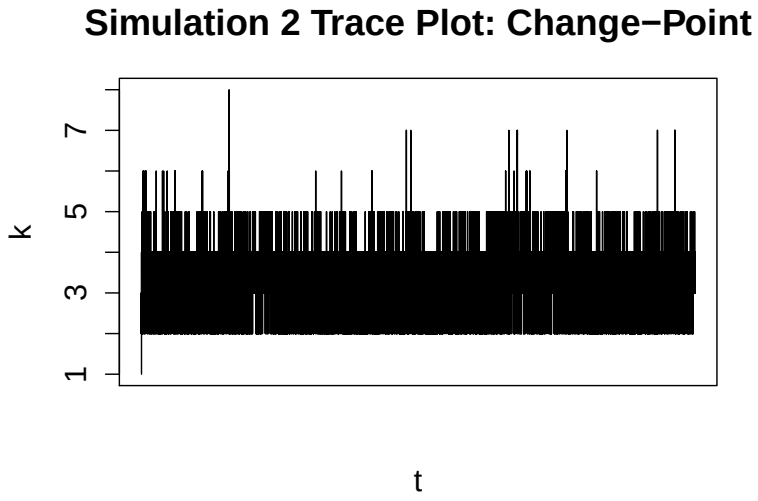

Figure 4.7: Trace Plot for Change-Point k in Simulation 2

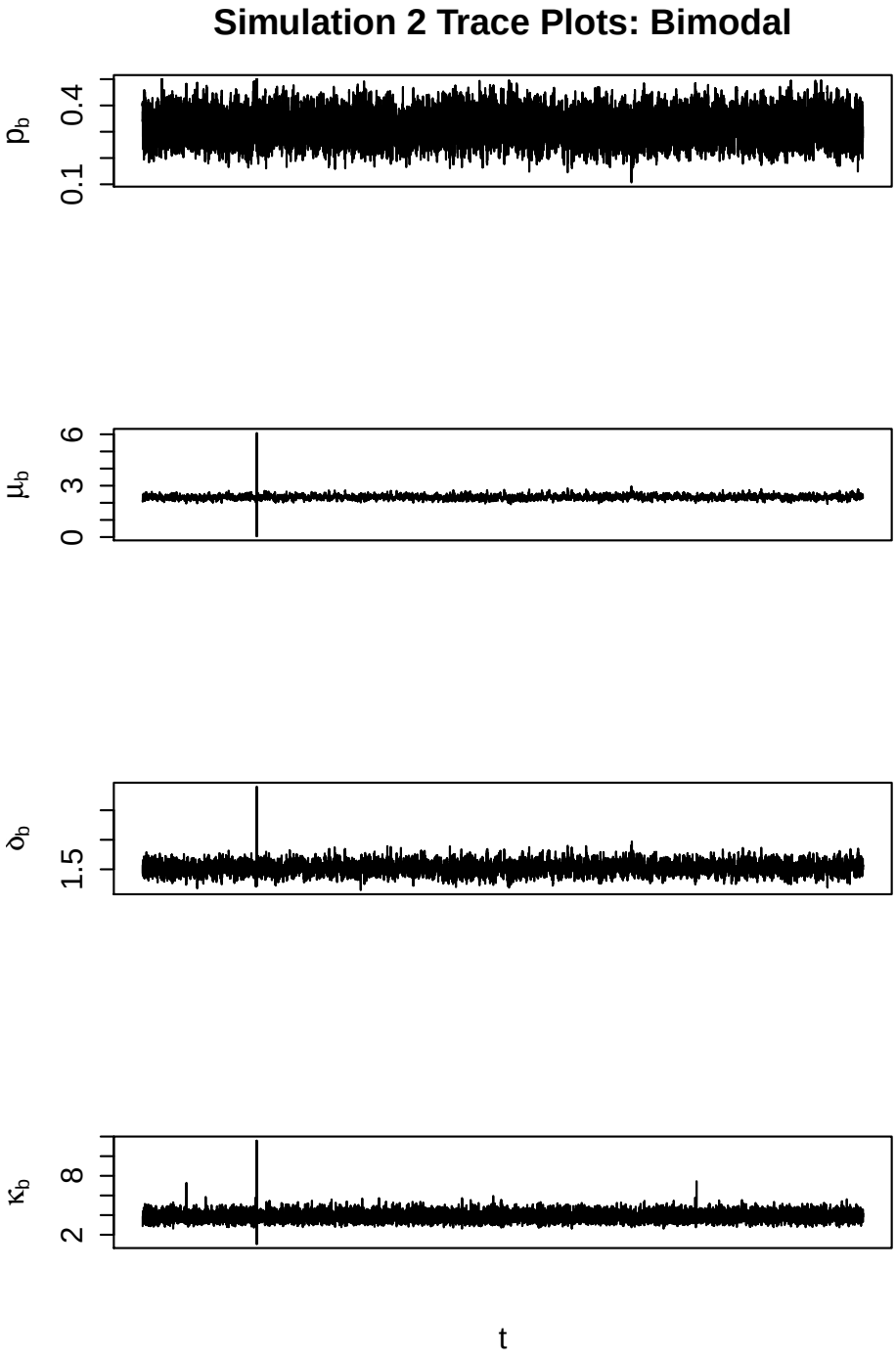

Figure 4.8: Trace Plots for $\boldsymbol{\theta}_b = (p_b = .3, \mu_b = pi/4, \delta_b = \pi/2, \kappa_b = 4)$ in Simulation 2

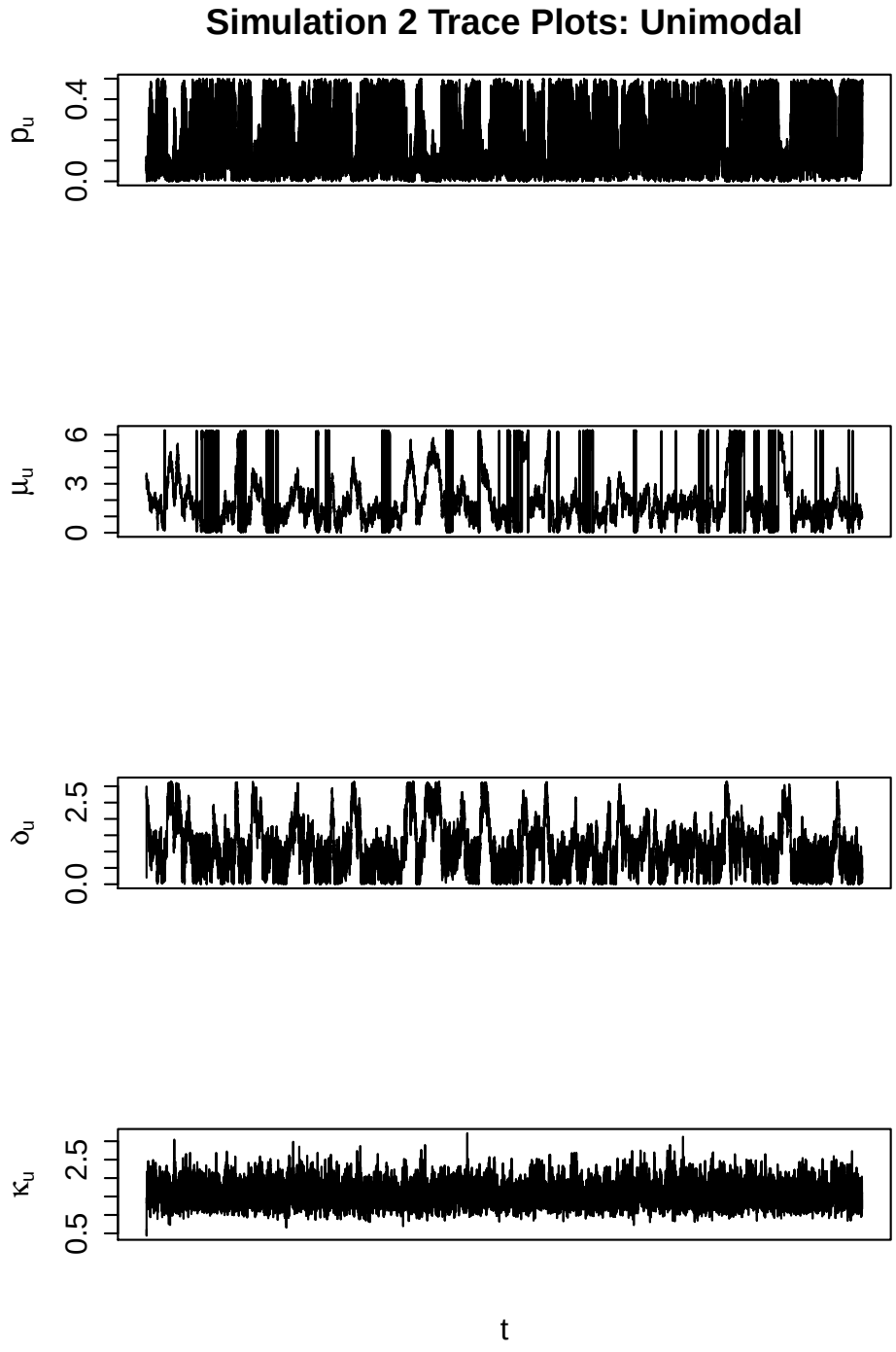

Figure 4.9: Trace Plots for $\boldsymbol{\theta}_u = (p_u = .3, \mu_u = 0, \delta_u = 3\pi/8, \kappa_u = 2)$ in Simulation 2

Chapter 5

Future Work

5.1 Preliminary Test Estimation

In the discussions on the PTE we relied primarily on numerical simulations, but can aim to try to derive analytical forms for the MSE and MRE of the PTEs. We also will explore if the non-centrality parameter has a connection to those obtained in linear models setting. Though through our simulations, MRE in the circular context behaves similarly to the linear versions.

5.2 Change-point Problems

In our Bayesian approach for detecting a change in the number of modes, for each scenario, we needed a manual adjustment of tuning parameters. One could follow the work of Roberts & Rosenthal (2009), and implement an adaptive procedure. As stated there, an increase in value of tuning parameter will decrease the acceptance rate, while a decrease will increase the acceptance rate. An adaptive process can periodically monitor the acceptance rate, then

increase the tuning parameter if acceptance rate is above 44% or decrease paramter if the acceptance rate falls below 44%.

If we suspect that the unimodal/bimodal distributions are identical then the M-H algorithm can be employed. If they are not identically distributed then our M-H algorithm will not suffice, whereas use of the GLRT will be useful. In future work, we may be able to develop an M-H algorithm for the non-identically distributed case, but it will require many more parameters in the M-H simualtion.

A wrapped circular distribution comes from wrapping a linear distribution around the unit circle. One example is wrapping a mixture of Normal distributions. Many interesting linear properties like symmetry and bimodality of such a mixture are retained in the circular case (see Jammalamadaka & Kozubowski (2015)). For example, if a mixture of Normals is bimodal, then the wrapped version will also be bimodal. One can bring to bear the unimodal and multimodal parameter subspaces of mixtures of Normals to our change-point problem, and we expect the resulting analysis to be simpler than dealing with the mixCND that we analyzed.

Bibliography

Aghaeepour, Nima, Finak, Greg, Hoos, Holger, Mosmann, Tim R, Brinkman, Ryan, Gottardo, Raphael, Scheuermann, Richard H, Consortium, FlowCAP, Consortium, DREAM, *et al.* . 2013. Critical assessment of automated flow cytometry data analysis techniques. *Nature methods*, **10**(3), 228–238.

Bancroft, TA. 1964. Analysis and inference for incompletely specified models involving the use of preliminary test (s) of significance. *Biometrics*, 427–442.

Bancroft, TA. 1965. Inference for incompletely specified models in physical sciences. *Bulletin of the International Statistical Institue*, **41**(1), 497–515.

Bancroft, Theodore Alfonso. 1944. On biases in estimation due to the use of preliminary tests of significance. *The Annals of Mathematical Statistics*, **15**(2), 190–204.

Basu, Sanjib, & Jammalamadaka, SR. 2000. Unimodality in circular data: a Bayes test. *Advances on Methodological and Applied Aspects of Probability and Statistics*, **1**, 4–153.

Chernoff, Herman. 1954. On the distribution of the likelihood ratio. *The Annals of Mathematical Statistics*, 573–578.

Chib, Siddhartha, & Greenberg, Edward. 1995. Understanding the metropolis-hastings algorithm. *The american statistician*, **49**(4), 327–335.

Darzynkiewicz, Zbigniew, Crissman, Harry, & Jacobberger, James W. 2004. Cytometry of the cell cycle: cycling through history. *Cytometry Part A*, **58**(1), 21–32.

(Eds.)Krishan, Awtar, Krishnamurthy, H, & Totey, Satish. 2011. *Applications of Flow Cytometry in Stem Cell Research and Tissue Regeneration*. John Wiley and Sons.

Gay, David M. 1990. Usage summary for selected optimization routines. *Computing science technical report*, **153**, 1–21.

Ghosh, Kaushik, Jammalamadaka, S Rao, & Vasudaven, Mangalam. 1999. Change-point problems for the von Mises distribution. *Journal of Applied Statistics*, **26**(4), 423–434.

Han, Chien-Pai, & Bancroft, TA. 1968. On pooling means when variance is unknown. *Journal of the American Statistical Association*, **63**(324), 1333–1342.

Ho, Hsiu J, Lin, Tsung I, Chang, Hannah H, Haase, Steven B, Huang, Sui, & Pyne, Saumyadipta. 2012. Parametric modeling of cellular state transitions as measured with flow cytometry. *BMC bioinformatics*, **13**(Suppl 5), S5.

Holzmann, Hajo, & Vollmer, Sebastian. 2008. A likelihood ratio test for bimodality in two-component mixtures with application to regional income distribution in the EU. *AStA Advances in Statistical Analysis*, **92**(1), 57–69.

James, William, & Stein, Charles. 1961. Estimation with quadratic loss. *Pages 361–379 of: Proceedings of the fourth Berkeley symposium on mathematical statistics and probability.* A, vol. 1, no. 1961.

Jammalamadaka, S Rao, & Kozubowski, Tomasz J. 2015. Wrapped circular distributions from Mixtures. *In preparation.*

Jammalamadaka, S Rao, & SenGupta, Ashis. 2001. *Topics in circular statistics.* Vol. 5. World Scientific.

Jammalamadaka, Sreenivasa R, & Lund, Ulric J. 2006. The effect of wind direction on ozone levels: a case study. *Environmental and Ecological Statistics*, **13**(3), 287–298.

Mardia, Kanti V, & Jupp, Peter E. 1999. *Directional statistics.* Vol. 494. John Wiley & Sons.

Mardia, KV, & Sutton, TW. 1975. On the modes of a mixture of two von Mises distributions. *Biometrika*, 699–701.

Metropolis, Nicholas, Rosenbluth, Arianna W, Rosenbluth, Marshall N, Teller, Augusta H, & Teller, Edward. 1953. Equation of state calculations by fast computing machines. *The journal of chemical physics*, **21**(6), 1087–1092.

Muralidharan, K, & Parikh, Rajiv. 2012. Bayesian Inferences on Mixture of Two Von-Mises Distributions. *Journal of Indian Society for Probability and Statistics Vol*, **14**, 56–69.

Nava, M M, & Jammalamadaka, S Rao. 2008. *Circular Statistics: Applications in Cricket and Ozone Levels.* Special Interest Group of the Mathematical Association of America: Statistics Poster Award, Joint Mathematics Conference, Washington D.C.

Ohtani, Kazuhiro. 1991. Estimation of the variance in a normal population after the one-sided pre-test for the mean. *Communications in statistics-theory and methods*, **20**(1), 219–234.

Pyne, Saumyadipta, Hu, Xinli, Wang, Kui, Rossin, Elizabeth, Lin, Tsung-I, Maier, Lisa M, Baecher-Allan, Clare, McLachlan, Geoffrey J, Tamayo, Pablo, Hafler, David A, *et al.* . 2009. Automated high-dimensional flow cytometric data analysis. *Proceedings of the National Academy of Sciences*, **106**(21), 8519–8524.

Roberts, Gareth O, & Rosenthal, Jeffrey S. 2009. Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics*, **18**(2), 349–367.

Roberts, Gareth O, Gelman, Andrew, Gilks, Walter R, *et al.* . 1997. Weak convergence and optimal scaling of random walk Metropolis algorithms. *The annals of applied probability*, **7**(1), 110–120.

Saleh, AK Md Ehsanes. 2006. *Theory of preliminary test and Stein-type estimation with applications*. Vol. 517. John Wiley & Sons.

Self, Steven G, & Liang, Kung-Yee. 1987. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, **82**(398), 605–610.

Snedecor, G. W. 1938. *Statistical Methods*. Collegiate Press, Iowa.

Stein, Charles, *et al.* . 1955. A necessary and sufficient condition for admissibility. *The Annals of Mathematical Statistics*, **26**(3), 518–522.

Stein, Charles, *et al.* . 1956. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Pages 197–206 of: Proceedings of the Third Berkeley symposium on mathematical statistics and probability.* A, vol. 1, no. 399.

Stephens, MA. 1962. Exact and approximate tests for directions. I. *Biometrika*, 463–477.

Upton, Graham JG. 1986. Approximate confidence intervals for the mean direction of a von Mises distribution. *Biometrika*, **73**(2), 525–527.

von Mises, R. 1918. Über die Ganzzahligkeit der Atomgewichte und verwandte Fragen. *Phys. z*, **19**, 490–500.