

AN ALTERNATIVE ESTIMATE OF PREFERRED DIRECTION FOR CIRCULAR DATA

Bennett Sango Otieno

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Statistics

APPROVED:

Dr. Christine M. Anderson-Cook, **CHAIR**

Dr. Samantha C. Bates

Dr. Eric P. Smith

Dr. George R. Terrell

Dr. Keying Ye

July 25, 2002

Blacksburg, Virginia

Keywords: Circular Data, Preferred Direction, Hodges-Lehmann Estimator.

Copyright 2002. Bennett Sango Otieno

An Alternative Estimate of Preferred Direction for Circular Data

Bennett Sango Otieno

ABSTRACT

Circular or angular data occur in many fields of applied statistics. A common problem of interest in circular data is estimating a preferred direction and its corresponding distribution. This problem is complicated by the so-called wrap-around effect, which exists because there is no minimum or maximum on the circle. The usual statistics employed for linear data are inappropriate for directional data, as they do not account for the circular nature of directional data. Common choices for summarizing the preferred direction are the sample circular mean, and sample circular median. A newly proposed circular analog of the Hodges-Lehmann estimator is proposed, as an alternative estimate of preferred direction. The new measure of preferred direction is a robust compromise between circular mean and circular median. Theoretical results show that the new measure of preferred direction is asymptotically more efficient than the circular median and that its asymptotic efficiency relative to the circular mean is quite comparable. Descriptions of how to use the methods for constructing confidence intervals and testing hypotheses are provided. Simulation results demonstrate the relative strengths and weaknesses of the new approach for a variety of distributions.

In Memory Of My Late Papa

and

To My Loving Mama.

Acknowledgments

First and foremost, I thank The Lord Jesus Christ for giving me the courage to face the challenges of each day during the entire period in school.

I sincerely thank Dr. Christine M. Anderson-Cook (my Committee Chair) for encouraging me to pursue Circular Statistics. I am very grateful for her commitment throughout the research work. Her close involvement and keen interest positively stirred my progress in studying Circular Statistics. Her availability to read the earlier manuscripts of this work plus the many suggestions thereafter significantly expanded my understanding of Circular Statistics. It was indeed a rare privilege to know and work under her supervision. Thank you Dr. A-C.

Thank you: Dr. Samantha Bates, Dr. Eric Smith, Dr. George Terrell and Dr. Keying Ye for agreeing to serve on my committee and making sure there was a one-one correspondence between what I wrote and what I meant. I thank the Department of Statistics for providing the funds to enable me to pursue my studies at Virginia Tech for three years. Thank you: Michele Marini, Mike Box and Brent Brownfield for making my computing experience during my research work less painful. To Dr. Megan Waterman, thanks for the encouragement when the simulations took forever to end. To Linda Breeding, thank you for reminding me that my long drives on I-77 to and from Ohio have got to stop.

To my late Papa and to my loving mama, I owe this to you. To my wife Karen and Son Tumaini which means HOPE, thank you for believing with me that this work will come to a respectable end. And especially to Karen, thank you for your friendship and for taking care of Tumaini while I was away. You have endured tough times that wont be erased in both my mind and heart. To my parents-in-law, sisters-in-law and brother -in-law, I say thanks for all the encouraging letters, emails, phone calls and prayers. To my brothers and sisters, thanks for being patient even when I took forever to reply your letters.

Finally to ALL friends both in Kenya and the U.S.A., who prayed and provided many resources to Karen, Tumaini and I, in ways we can never repay - THANKS, and may God Bless You ALL.

Contents

1	Introduction	1
1.1	Circular Data	2
1.2	The Need for Appropriate Measures and Analysis	11
1.3	Fundamental Characteristics of Directional Data	13
1.3.1	The Mean Direction and the Resultant Length	14
1.3.2	The Median Direction	17
1.3.3	The Measures of Dispersion	21
2	Literature Review	23
2.1	Directional Data	23
2.1.1	Robust Estimation of Preferred Direction	24
2.1.2	Outliers in Circular Data	26
2.2	Hodges-Lehmann Estimator for Linear Data	27
2.3	Influence Functions for Estimates of Location for Linear Data	30
2.4	Bootstrap Methods	33
2.4.1	Introduction	33

2.4.2	Confidence Intervals Involving Measures of Preferred Direction	34
3	New Measures of Preferred Direction	38
3.1	Introduction	38
3.2	Unique Solution to Circular Median	39
3.3	Comparison of Mardia Median & New Median	42
3.3.1	Discussions and Conclusions on Relative Performance of Mardia Median and New Median	46
3.4	New Measures of Preferred Direction for Circular Data	46
3.5	Comparison of HL1, HL2 & HL3	49
3.5.1	Simulation to compare HL1, HL2, and HL3	49
3.5.2	Discussions and Conclusions on Relative Performance of HL1, HL2 and HL3	53
3.6	Theoretical Results for a Limited Range of Concentrated Distributions . . .	53
3.6.1	Introduction	53
3.6.2	Approximations of von Mises distribution $VM(\mu, \kappa)$	54
3.6.3	Probability of Wrapping	57
3.6.4	Circular and Linear point estimates for $\kappa \geq 2$	59
3.7	Influence Functions for Measures of Preferred Direction	66
3.7.1	Position of an Outlier on the Influence curve of the mean, median and HL2	70
3.8	Asymptotic Relative Efficiency	71

4	Simulation Study	76
4.1	Overview	76
4.2	Uncontaminated Data	79
4.2.1	Effect of Sample Size on the Mean, the Median and HL	79
4.2.2	Effect of Concentration Parameter (κ) on the Mean, the Median and HL	81
4.3	Comparing Theoretical and Simulated Results for HL	83
4.4	Contamination in Spread	85
4.4.1	Effect of Sample Size on the Mean, the Median and HL	85
4.4.2	Effect of Concentration Parameter (κ) on the Mean, the Median and HL	87
4.4.3	Effect of increasing contamination in spread on the Mean, the Median and HL	89
4.5	Contamination in Location	91
4.5.1	Effect of Sample Size on the Mean, the Median and HL	91
4.5.2	Effect of Concentration Parameter (κ) on the Mean, the Median and HL	93
4.5.3	Effect of contamination level on Mean, Median and HL, n=20	95
4.5.4	Effect of increasing amount of shift of mean on Mean, Median and HL, n=20	97
4.5.5	Discussions and Conclusions on Relative Performance of Mean, Median and HL	99

5	Bootstrap: Confidence Interval and Test of Hypothesis	100
5.1	Introduction	100
5.2	Bootstrap method	102
5.2.1	Nonparametric Bootstrap	103
5.2.2	Parametric Bootstrap	103
5.3	Algorithm for Bootstrapping HL	104
5.4	Examples	107
5.4.1	Example 1: Frog Data	107
5.4.2	Example 2: Cross-bed Azimuths of Palaeocurrents	109
5.5	Simulation Study	111
6	Conclusions and Future Study	116
6.1	Conclusions	116
6.2	Future Research	117
	References	119
	Appendix	127
	Appendix A	127
	Appendix B	131
	Appendix C	134
	Appendix D	138
	Appendix E	171

List of Figures

1.1	Representation of Circular Data	3
1.2	Relationship between rectangular and polar co-ordinates	7
1.3	Common distributions	9
1.4	Homing Directions of 14 Northern cricket frogs clockwise from North	12
1.5	Calculation of mean direction by forming a vector polygon	15
1.6	Circular median for even and odd number of observations	18
1.7	P^* : Linear Median of 75^0 , P : Circular Median of 52^0 for Ackermann 1997 Data.	19
1.8	Circular mean, \mathbf{m} , median, \mathbf{P} , and antimedian, \mathbf{Q} , of a sample.	20
1.9	Circular distance is the arc length LMN.	22
3.1	Non Uniqueness of Circular median	39
3.2	Original Observation \mathbf{o} , Potential Medians \mathbf{p}	41
3.3	Special features	42
3.4	Mardia Median & New Median for VM(0, 2)	44
3.5	Mardia median & New median for VM (0, κ), $n = 20$	45
3.6	Hodges-Lehmann estimate for linear data	48

3.7	Hodges-Lehmann estimate for circular data	49
3.8	HL1, HL2 & HL3 for VM(0, 2)	51
3.9	HL1, HL2 & HL3 for VM(0, κ), $n = 20$	52
3.10	Plot of $\hat{\sigma}^2 = [-2\log A[\kappa]]$, and $\hat{\sigma}^2 = \frac{1}{\kappa}$ vs. Concentration Parameter (κ) for a single observation	55
3.11	Plot of $\hat{\sigma}^2 = [-2\log A[\frac{3\kappa}{\pi}]]$, and $\hat{\sigma}^2 = \frac{\pi}{3\kappa}$ vs. Concentration Parameter (κ) for a pairwise circular mean.	56
3.12	Split circle: $b = \frac{\pi}{2} + \pi$, $-b = \frac{\pi}{2} - \pi$	58
3.13	Plot of Probability of Wrapping vs. Sample size	59
3.14	Adjacent angles	60
3.15	Calculation of median direction	63
3.16	Distribution of Circular Hodges-Lehmann Estimator for $\kappa = 2$	65
3.17	Influence Functions for $1 \leq \kappa \leq 8$	69
3.18	Data with a Single Outlier	70
4.1	Uncontaminated von Mises, VM(μ, κ)	77
4.2	Contaminated Data	78
4.3	Mean, Median and HL for VM(0,2)	80
4.4	Effect of κ on Mean, Median and HL for VM (0, κ), $n = 20$	82
4.5	Mean, Median and HL for 70%VM(0, 2) & 30% Uniform, $\epsilon = 0.3$	86
4.6	Mean, Median and HL for 70%VM (0, κ) & 30%Uniform, $\epsilon = 0.3$ & $n = 20$	88
4.7	Effect of increasing the spread on Mean, Median and HL, for $\kappa = 2$, & $n = 20$	90

4.8	Mean, Median and HL for VM (0, 2) & VM($\frac{\pi}{4}$, 2), $\epsilon = 0.3$	92
4.9	Mean, Median and HL for 70%VM(0, κ) & 30%VM($\frac{\pi}{4}$, κ), $\epsilon = 0.3$, $n = 20$	94
4.10	Effect of contamination level on Mean, Median and HL, $n=20$	96
4.11	Effect of Shifting the mean direction from 0 to $\frac{\pi}{2}$ on Mean, Median and HL, $\kappa = 2$ & $n = 20$	98
5.1	Symmetric Arc Confidence Interval	105
5.2	Equal-Tailed Confidence Interval	105
5.3	Likelihood-Based Arc Confidence Interval	106
5.4	95% Bootstrap C.I for Circular Mean, $n = 14$, $B = 1000$	108
5.5	95% Bootstrap C.I for Circular Median, $n = 14$, $B = 1000$	108
5.6	95%Bootstrap C.I for HL, $n = 14$, $B = 1000$	109
5.7	95% Bootstrap C.I for Circular Mean, $n = 30$, $B = 1000$	110
5.8	95% Bootstrap C.I for Circular Median, $n = 30$, $B = 1000$	110
5.9	95% Bootstrap C.I for HL, $n = 30$, $B = 1000$	111

List of Tables

4.1	Theoretical and Simulated Mean Resultant Lengths and Concentration Parameters for distribution of HL for various concentration parameters.	84
5.1	Bootstrap Confidence Intervals for $n = 10, \kappa = 2, M = 1000, B = 500$. . .	113
5.2	Bootstrap Confidence Intervals for $n = 10, \kappa = 10, M = 1000, B = 500$. . .	115

Chapter 1

Introduction

A common one-dimensional statistical problem is the estimation of a location parameter, and its corresponding distribution. For directional data (random sample of measured directions from a reference point), this problem involves identification of a preferred direction for a random process generating circular or angular data, and is complicated by the wrap-around nature of this type of data, with no maximum and minimum.

This dissertation develops an alternative estimate of preferred direction for circular data. The motivation behind the new estimate is to balance robustness with using information from all observations. This dissertation presents a method for estimating preferred direction from pairwise circular means, that is analogous to the Hodges-Lehmann estimate of center for linear data. The algorithm uses the definition of circular median as stated in Mardia (1972, p. 28). Properties of this estimate are explored and compared to some of the existing measures of preferred direction. S-Plus functions are provided in *Appendix E* to compute this estimate, since solutions to many directional data problems are computationally intensive and often not obtainable in simple closed analytical forms (Jammalamadaka & SenGupta, 2001).

In the following sections of Chapter 1, a brief overview of circular data including the graphical representation and fundamental quantities are presented. Some common distributions on the circle are also discussed. Chapter 2 presents existing estimates of center, a discussion of outliers, and some existing tests for the preferred direction for directional data. Also included in this section are a brief description of the Hodges-Lehmann estimate of location for linear data, and an outline of bootstrap methods. In Chapter 3, an alternative solution to guarantee uniqueness of circular median is proposed, the new measure of preferred direction is described for unimodal circular data, and influence functions (IF) of all measures of preferred direction considered are given. The Asymptotic Relative Efficiencies (ARE) for the measures of preferred direction are also obtained. Simulation studies to compare relative performance of the new measure to the existing ones are presented in Chapter 4. Bootstrap confidence interval and hypothesis testing using the new measure are discussed and compared to existing methods in Chapter 5. Finally, future research topics are outlined in Chapter 6.

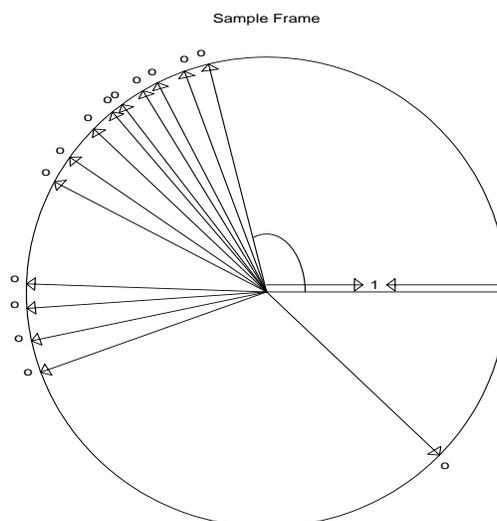
1.1 Circular Data

Statistical problems where the data are in the form of angular measurements giving orientations or angles in the plane (circular data) or in space (spherical data) arise in diverse scientific fields. Circular data is the simplest case of this category of data called directional data, where the single response is not scalar, but angular or directional. The basic statistical assumption is that the data are randomly sampled from a population of directions. Observations arise either from direct measurement of angles or they may arise from the measurement of times reduced modulo some period and converted into angles according to the periodicity of time, such as days or years. They are commonly summarized as locations on a unit circle or as angles over a 360° or 2π radians range, with the endpoints of each range corresponding to the same location on the circle.

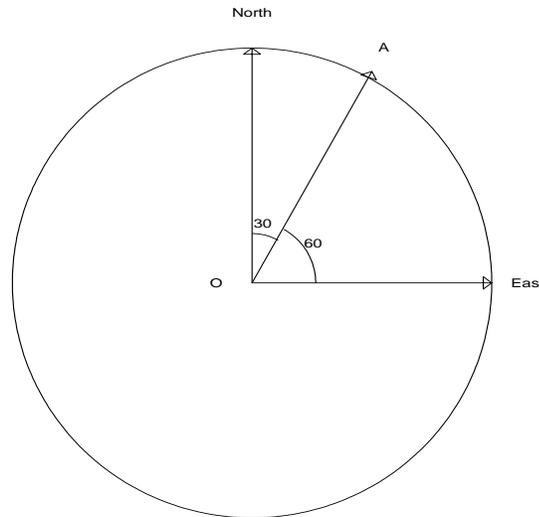
A convenient sample frame for circular data is the circumference of a unit circle centered at the origin with each point on the circumference representing a direction, or, a unit vector since magnitude has no relevance. Figure 1.1a illustrates how unit lengths starting at the origin and pointing in the direction of their angle are used to represent individual observations. It is because of this representation that angular observations in two-dimensions are called circular data. However, numerical representation as an angle is not necessarily unique since the angular value depends on the choice of what is labeled as the zero-direction (true East or true North (“azimuth”)) and the sense of rotation (counter-clockwise or clockwise). In Figure 1.1b, the angle \mathbf{A} is 30° if the zero direction is true North and the sense of rotation is clockwise, however, if zero is taken to be true East and the sense of rotation is counter-clockwise, then angle \mathbf{A} is 60° . Clearly $30^\circ \neq 60^\circ$, hence it is important that any method of estimation or hypothesis testing handles this consistently.

Figure 1.1: Representation of Circular Data

1.1a. Point on the circumference on a unit circle



1.1b. Value depends on choice of origin and sense of rotation

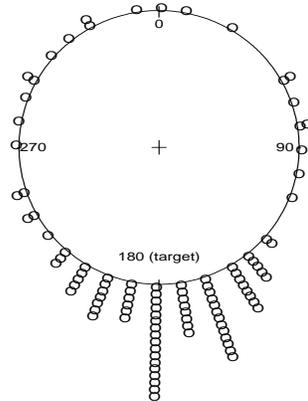


In addition there is no natural ordering or ranking of observations, since whether one direction is “larger” than the other depends on whether clockwise or counter-clockwise is treated as being the positive direction as well as where the “zero” angle is located. This renders rank-based methods used for linear data essentially inapplicable for circular data (Mardia, 1972 and Jammalamadaka & SenGupta, 2001). Therefore it is important to make sure that our conclusions (i.e. data summaries, inferences etc) are a function of the given observations and not dependent on the arbitrary values by which we refer to them.

The two main sources that give rise to circular data correspond to the two principal circular measuring instruments, the compass and the clock. Examples of circular data measured by the compass include the flight directions of birds, animal migration, paleomagnetic directions, or the directions of wind.

Example 1: (Fisher 1993, p. 243)

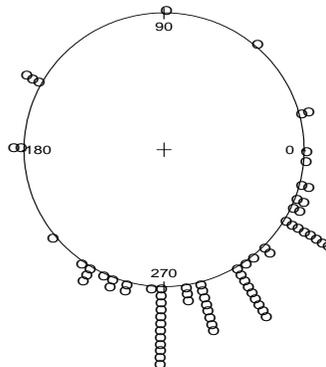
Directions chosen by 100 ants in response to
an evenly illuminated black target



Most ants tend to find the target, however, several ants miss the target giving some observations around the entire possible range.

Example 2: (Fisher 1993, p. 254)

Wind directions
at a Mountain site in ACT, Australia



The most common wind direction is between South and East.

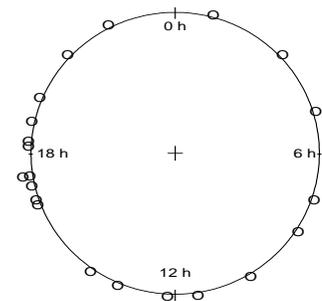
Typical observations measured by the clock include arrival times (on a 24-hour clock) of patients at a causality unit in a hospital. Similar type of data arise as times of day (or times of year) of appropriate events, for example the times of day at which thunderstorms occur, the times of the year at which heavy rain occurs and time of the day a major traffic accident occurs, see Example 3. Notice that in each case, each subject has only one measurement but there is a cyclic pattern in time. That is, the quantity of interest is a function of where in the cycle the observation occurs.

Example 3: (Batschelet 1981, p.13)

Distribution of traffic accidents in a city

Major traffic accidents in a city recorded during several days

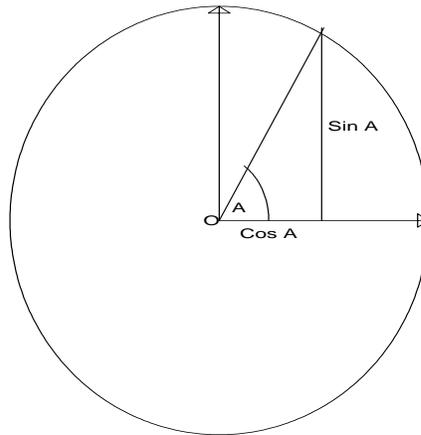
00 : 56h	03 : 08	04 : 52
07 : 16	08 : 08	10 : 00
11 : 24	12 : 08	13 : 28
14 : 16	16 : 20	16 : 44
17 : 04	17 : 20	17 : 24
18 : 08	18 : 16	18 : 56
19 : 32	20 : 52	22 : 08



Two major approaches to directional statistics, namely, the intrinsic approach (directions are considered as points on the circle itself) and the embedding approach (directions are considered as special points in the plane) are commonly used (Mardia & Jupp, 2000, p. 14). The embedding approach of regarding each point θ on the circle as the unit vector

$X = (\cos \theta, \sin \theta)^T$ in the plane enables us to take expectations and thereby define unbiasedness.

Figure 1.2: Relationship between rectangular and polar co-ordinates



A single observation $\mathbf{A} = \theta^0$ ($0 < \theta^0 \leq 360^0$) represents the angle made by the vector with the positive x-axis (the point $(1, 0)$ on the unit circle) in the counter-clockwise direction. The Cartesian co-ordinates of the vector are $(x, y) = (\cos\theta^0, \sin\theta^0)$ while the polar co-ordinates are $(r, A) = (1, \theta^0)$. See Figure 1.2.

The key characteristic that differentiates circular data from data measured on a linear scale is its wrap-around nature with no maximum or minimum. That is, the “beginning” coincides with the “end”, i.e., $0 = 2\pi$ and in general the measurement is periodic with θ being the

same as $\theta + 2p\pi$ for any integer p . Differences between the theories of statistics on the line and on the circle can be attributed to the fact that the circle is a closed curve while the line is not. Thus, distribution functions, characteristic functions and moments on the circle have to be defined by taking into account the natural periodicity of the circle.

A circular distribution (CD) is one whose total probability is concentrated on the circumference of a unit circle. A set of identically independent random variables from such a distribution is referred to as a random sample from the CD. See Jammalamadaka & Sen-Gupta (2001, p. 25-63) for a detailed discussion of circular probability distributions. Two frequently used families of distributions for circular data include the von Mises and the Uniform distribution.

The importance of the von Mises distribution is similar to the Normal distribution on the line (Mardia, 1972). It was introduced by von Mises (1918) to study the deviations of measured atomic weights from integral values. It is a symmetric unimodal distribution characterized by a mean direction μ , and concentration parameter κ , with probability density function

$$f(\theta) = [2\pi I_0(\kappa)]^{-1} \exp [\kappa \cos(\theta - \mu)] \quad 0 \leq \theta, \mu < 2\pi, \quad 0 \leq \kappa < \infty, \quad (1.1)$$

where

$$I_0(\kappa) = (2\pi)^{-1} \int_0^{2\pi} \exp [\kappa \cos(\phi)] d\phi = \sum_{j=0}^{\infty} -\frac{1}{(j)^2} \left(\frac{\kappa^2}{4}\right)^j$$

is the modified Bessel function of order zero. See Fisher(1993, p.50) for a series expansion and methods for evaluating $I_0(\kappa)$. κ is a concentration parameter, which quantifies the dispersion. As κ increase from zero, $f(\theta)$ peaks higher about μ . See Figure 1.3a. Note, we say that the circular random variable θ is symmetric about a given direction μ if its distribution has the property $f(\mu + \theta) = f(\mu - \theta)$, for all θ , where addition or subtraction is modulo (2π) .

If κ is zero, then $f(\theta) = \frac{1}{2\pi}$ and the distribution is uniform with no preferred direction.

The circular uniform distribution has the following density

$$f(\theta) = \frac{1}{2\pi}, \quad 0 \leq \theta \leq 2\pi . \tag{1.2}$$

All directions are equally likely, hence this is also known as the Isotropic distribution. This distribution represents the state of no “preferred direction”, since the total probability is spread out uniformly on the circumference of a circle. See Figure 1.3b. The uniform distribution on the circle has the property that the sample mean direction and the sample length of the resultant vector are independent. Similar property is held by the normal distribution for linear data (Kent et al, 1979).

Figure 1.3: Common distributions

1.3a. Probability density functions of von Mises Distribution for different values of the concentration parameter (κ) and $\mu = \frac{\pi}{2}$.

1.3b. Uniform distribution

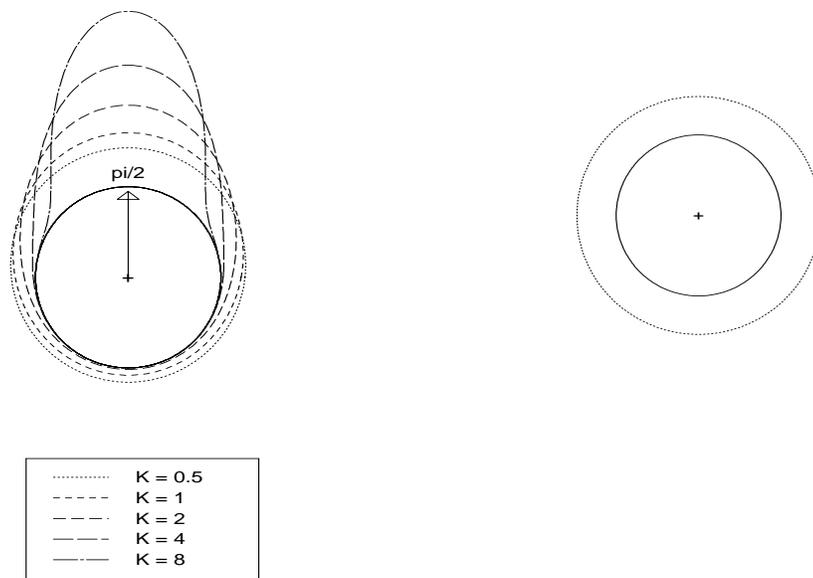


Figure 1.3 shows the plot of Uniform distribution and von Mises distributions with $\mu = \frac{\pi}{2}$ and different values of κ . Note that as $\kappa \rightarrow 0$, the tail of the distribution becomes heavier, and as $\kappa \rightarrow \infty$, the distribution tends to a single spike at $\mu = \frac{\pi}{2}$. Note that if a distribution is symmetric about the direction $\theta = \mu$, it is also symmetric about $\theta = \mu + \pi$. When $\kappa \geq 2$, the von Mises distribution $VM(\mu, \kappa)$, can be approximated by the wrapped normal distribution $WN(\mu, \rho)$, which is a symmetric unimodal distribution obtained by wrapping a normal $N(\mu, \sigma^2)$ distribution around the circle.

A circular r.v θ is said to have a WN distribution if its pdf is

$$f_W(\theta) = (2\pi)^{-1} + \pi^{-1} \sum_{p=1}^{\infty} \rho^{p^2} \cos[p(\theta - \mu)], \quad 0 \leq \mu, \theta \leq 2\pi, \quad 0 \leq \rho \leq 1, \quad (1.3)$$

where μ and $\rho = \exp(-\frac{1}{2}\sigma^2)$ are the mean direction and mean resultant length (see Section 1.2 for the definition) respectively. The value of $\rho = 0$ corresponds to the circular uniform distribution, and as ρ increases to 1, the distribution concentrates increasingly around μ . The wrapped normal was introduced by Zernike (1928) and later studied by Wintner (1933), Levy (1939), and Gumbel et al. (1953). Stephens (1963) matched the first trigonometric moments of the von Mises and wrapped normal distributions, with $\rho = e^{-\frac{1}{2}\sigma^2} = A(\kappa) = \frac{I_1(\kappa)}{I_0(\kappa)}$ establishing that the two have a close relationship. The similarity of the two distributions has also been noted and to some extent explained by Kendall (1974a, 1974b), Lewis (1974) and Kent (1976). Based on the difficulty in distinguishing the two distributions, Collett and Lewis (1981) conclude that decision on whether to use a von Mises model or a wrapped normal model, depends on which of the two is most convenient.

Note, if Y is the random variable on the line, the random variable $X_w = Y(\text{mod } 2\pi)$ has the wrapped distribution whose distribution function is given by

$$F_w(\theta) = \sum_{k=-\infty}^{\infty} [F(\theta + 2\pi k) - F(2\pi k)], \quad 0 < \theta \leq 2\pi. \quad (1.4)$$

The operation $X_w = Y(\text{mod } 2\pi)$, corresponds to taking the real line and wrapping it around the circle of unit radius, accumulating probability over all the overlapping points $x = \theta, \theta \pm 2\pi, \theta \pm 4\pi, \dots$. This is clearly a many-to-one mapping so that if $g(\theta)$ represents the circular density and $f(x)$ the density of the real-valued random variable, we have

$$g(\theta) = \sum_{m=-\infty}^{\infty} f(\theta + 2m\pi), \quad 0 \leq \theta < 2\pi.$$

Other models used to analyze symmetric unimodal circular data are the Cardiod distribution, $C(\mu, \rho)$ and the Wrapped Cauchy distribution, $WC(\mu, \rho)$ (Mardia, 1972, p. 51, 56).

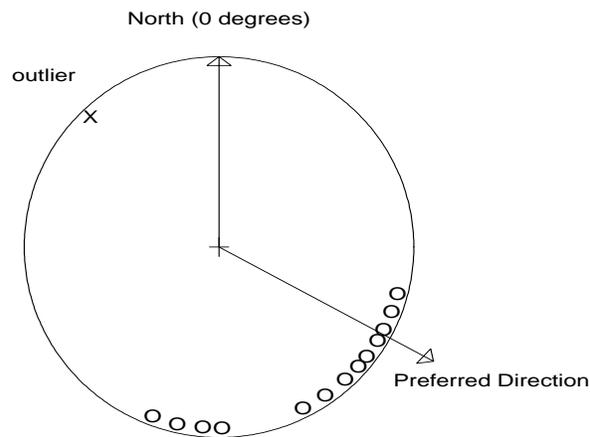
On the circle there is a limit to how far an observation may fall from the others, so one may expect fewer problems due to outlier. However, circumstances which give rise to outliers on the line, such as contaminated observations, misrecordings and values from a distribution other than that of the main sample, also occur in the circle. Consider for example data relating to the homing ability of Northern cricket frog, *Acris crepitans*, given by Ferguson et. al. (1967). The dot plot of the data in Figure 1.4 indicates that one observation $X = 316^0$ may be considered as an outlier relative to a basic von Mises model. S-Plus functions for obtaining circular plots such as Figure 1.4 are given in *Appendix E*.

1.2 The Need for Appropriate Measures and Analysis

A large part of parametric statistical inference for circular data is derived based on one or two models and there has not been enough discussion on model-robustness, i.e., to justify their validity and use when the data is actually from another model (Jammalamadaka & SenGupta, 2001). As a result, modeling asymmetric data sets, which frequently occur in practice provides some challenges because of the lack of appropriate models.

Many applied scientists (biologists, geologists, social and behavioral scientists) dealing with

Figure 1.4: Homing Directions of 14 Northern cricket frogs clockwise from North



circular data have fallen into the trap of using the more common, but inappropriate linear methods like ranking, for example Zar (1999). In addition one has to distinguish cases when time is considered in the usual time-series analysis which is a linear variable, compared to situations where one is considering timing only within a cycle, which is most usefully treated as a circular variable. The key distinction between these two approaches is whether the change across time or within cycle is of primary interest.

1.3 Fundamental Characteristics of Directional Data

The inherent periodicity of circular data brings with it a peculiar nature that does not occur elsewhere in statistics. Consider two angles which are 2 degrees apart. If the interval $[-180^0, 180^0]$ is chosen, the two angles would be -1^0 and 1^0 . On the interval $[0^0, 360^0]$, the angles will be 1^0 and 359^0 . If viewed graphically on the circle, no problem is apparent, but numerically potential problems exist. For instance, consider estimating the mean direction of the latter pair of angles. Clearly, these observations are centered about 0^0 . However, using naive linear methods, the sample mean and standard deviation of these two observations would be 180^0 and 253^0 respectively. Had the pair of angles been 1^0 & -1^0 , and we used the naive linear methods, we would get more sensible values of 0 and $\sqrt{2}$ as the sample mean and standard deviation, respectively. This illustrates the need for different measures of location and scale when dealing with circular data. However, since the choice of a zero-direction and the sense of rotation is arbitrary, one needs decision procedures, which are invariant under such choices. A point estimate $\tilde{\theta}$, is said to be location (translation) invariant if

$$\tilde{\theta}[(\theta_1 + \eta), \dots, (\theta_n + \eta)] = \eta + \tilde{\theta}[\theta_1, \dots, \theta_n], \quad (1.5)$$

for every η and $(\theta_1, \dots, \theta_n)$. That is, if the data is shifted by a certain amount η , the value of the point estimate also changes by the same amount.

Three common choices for summarizing the preferred direction are the mean direction, the median direction and the modal direction. (Fisher, 1993, pp.1). The sample mean direction is usually preferred for moderately large samples, because when combined with a measure of sample dispersion, it acts as a summary of the data suitable for comparison and amalgamation with other such information. The sample median can be thought of as balancing the number of observations on two halves of the circle and will be discussed in section 1.3.2. The sample modal direction is the direction corresponding to the maximum concentration of the data and is less useful because of difficulties in its calculation, in drawing inferences, and in ascertaining its sampling error.

All three measures of preferred direction are undefined if the sample data are equally spaced around the circle. This is sensible because if the data are symmetric around the circle then there is no preferred direction. In case of bimodal data, there are two preferred directions, consequently the three measures are also not meaningful. We expect any new measures to have these two properties as well. In this dissertation, we shall emphasize estimating the preferred direction for unimodal circular data i.e., point or concentrate towards a single direction.

1.3.1 The Mean Direction and the Resultant Length

An appropriate and meaningful measure of the mean direction for a set of directions which are unimodal is obtained by treating the data as unit vectors and using the direction of their resultant vector. For a set of angular measurements $\theta_1, \dots, \theta_n$, we convert each observation to its rectangular form $(\cos \theta_i, \sin \theta_i)$, $i = 1, \dots, n$. We obtain the resultant vector of these (e_1, \dots, e_n) n unit vectors from the origin by summing them component-wise, to get

$$R = \left(\sum_{i=1}^n \cos \theta_i, \sum_{i=1}^n \sin \theta_i \right) = (C_n, S_n), \text{ say.} \quad (1.6)$$

Let $|R| = \sqrt{(C_n^2 + S_n^2)} > 0$ represent the length of the resultant vector \mathbf{R} . The direction of this resultant vector \mathbf{R} is known as the circular mean direction, and is denoted by $\bar{\theta}$. A “quadrant-specific” inverse of the tangent definition of the circular mean direction is,

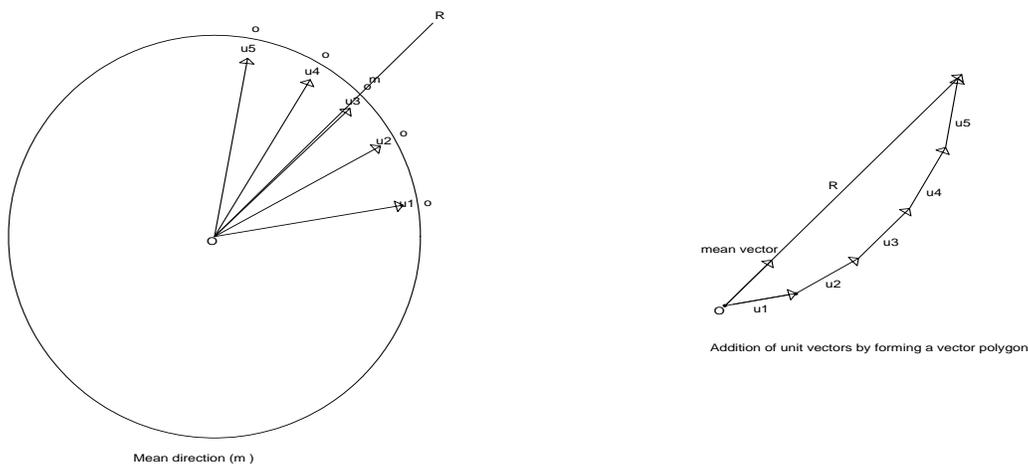
$$\bar{\theta} = \begin{cases} \arctan (S_n/C_n) , & \text{if } C_n > 0 \\ \frac{\pi}{2} , & \text{if } C_n = 0 \text{ and } S_n > 0 \\ \frac{-\pi}{2} , & \text{if } C_n = 0 \text{ and } S_n < 0 \\ \pi + \arctan (S_n/C_n), & \text{otherwise.} \end{cases} \quad (1.7)$$

Note, the inverse tangent function, \tan^{-1} (or \arctan), takes values in $[-\frac{\pi}{2}, \frac{\pi}{2}]$. This can be computed using the program *ave.ang()* in *Appendix E*. The above definition, provides us the

correct unique inverse on $[0, 2\pi)$, which takes into account the signs of C_n and S_n (Mardia, 1972, p.22). Note that in the context of circular statistics, $\bar{\theta}$ does not denote the standard linear average $(\theta_1 + \dots + \theta_n)/n$.

Geometrically, the mean direction is equivalently obtained by forming the vector polygon as shown in Figure 1.5.

Figure 1.5: Calculation of mean direction by forming a vector polygon



The resultant vector is the sum of the unit vectors, $R = \sum_{i=1}^n u_i$. This is a vector with length between 0 and n , and pointing in the mean direction of the sample $\bar{\theta}$. The sample mean resultant length (standardized length) is given by $\bar{R} = \frac{|E|}{n}$, with $\bar{R} \in [0, 1]$. If the data are closely clustered around the mean, then \bar{R} is close to 1. However, if the data are evenly spread around the circle, \bar{R} will be near zero. Hence \bar{R} is a natural measure of spread.

Thus, the resultant vector can be decomposed into two nonparametric components; the mean direction $\bar{\theta}$ and the mean resultant length \bar{R} , which form a useful starting point for any analysis. Jammalamadaka & SenGupta (2001, p. 14), prove that $\bar{\theta}$ is location invariant,

i.e., if the data is shifted by a certain amount, the value of $\bar{\theta}$ also changes by the same amount.

Proposition 1.1: $\bar{\theta}$ is invariant with respect to changes in the sense of rotation, i.e., when we switch from clockwise to counter-clockwise so that θ 's become $(2\pi - \theta)$'s, then $\bar{\theta}$ becomes $(2\pi - \bar{\theta})$.

Proof: Let $(\theta_1, \theta_2, \dots, \theta_n)$ have mean direction $\bar{\theta}$. We will show that $(2\pi - \theta_1, 2\pi - \theta_2, \dots, 2\pi - \theta_n)$ have mean direction $(2\pi - \bar{\theta})$.

Suppose \mathbf{R}^* is the resultant vector of the new set of observations i.e., after the shift. Then we have

$$\mathbf{R}^* = \left(\sum_{i=1}^n \cos(2\pi - \theta_i), \sum_{i=1}^n \sin(2\pi - \theta_i) \right) = (C_n^*, S_n^*) \text{ (say).}$$

$$\begin{aligned} \text{Then, } C_n^* &= \sum_{i=1}^n \cos(2\pi - \theta_i) \\ &= \sum_{i=1}^n (\cos \theta_i \cos(2\pi) + \sin \theta_i \sin(2\pi)) \\ &= C_n \cos(2\pi) + S_n \sin(2\pi) \\ &= C_n, \quad \text{since } \cos(2\pi) = 1 \text{ and } \sin(2\pi) = 0 \\ &= R \cos \bar{\theta} \\ &= R \cos(2\pi) \cos \bar{\theta} + R \sin(2\pi) \sin \bar{\theta} \\ &= R \cos(2\pi - \bar{\theta}). \end{aligned}$$

Similarly $S_n^* = R \sin(2\pi - \bar{\theta})$. Now, $R^* = \|\mathbf{R}^*\| = \sqrt{(C_n^{*2} + S_n^{*2})} = R = \sqrt{(C_n^2 + S_n^2)}$. Hence, $\frac{C_n^*}{R^*} = \cos(2\pi - \bar{\theta})$, $\frac{S_n^*}{R^*} = \sin(2\pi - \bar{\theta})$.

Therefore, the point estimate does not depend on what direction is taken to be the positive direction. Hence, any practitioner using this estimate need not be wary of the zero-direction.

Any measure of preferred direction we develop should have such a property as well. See Mardia (1972, p.45) for more properties of the mean direction.

1.3.2 The Median Direction

For the purposes of robust estimation, it is desirable to have a version of the sample median for circular data. As a nonparametric and robust estimate for the preferred direction of a distribution, it has a different character from the sample mean as illustrated by different breakdown properties. The circular median was defined more formally by Fisher and Powell (1989) as the angle about which the sum of absolute angular deviations is minimized. We will return to this in Section 1.3.3.

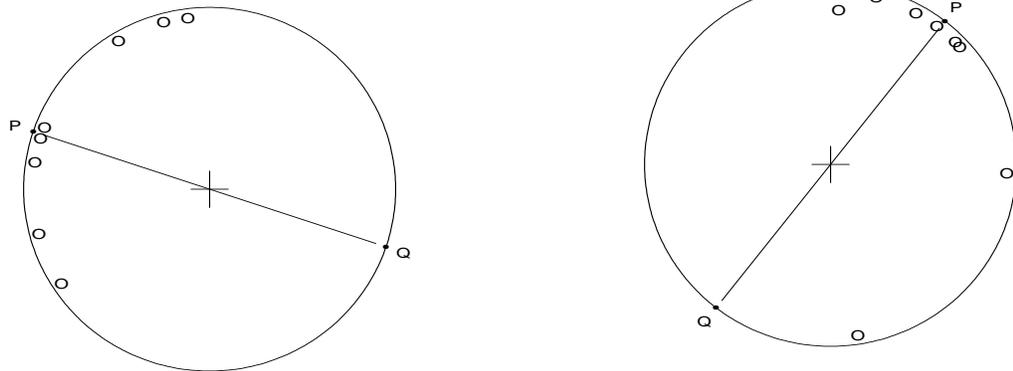
According to Mardia (1972, p. 28, 31), the sample median direction $\tilde{\theta}$ of angles $\theta_1, \dots, \theta_n$ is the point \mathbf{P} on the circumference of the circle that satisfies the following two properties: (a) The diameter PQ through \mathbf{P} divides the circle into semi-circles, each with an equal number of observed data points and, (b) the majority of the observed data is closer to \mathbf{P} than to the anti-median \mathbf{Q} . Mardia (1972, p.46-47) proved that the circular median of a unimodal distribution is unique. It is also rotationally invariant as shown by Ackermann (1997). Wehrly and Shine (1981) obtained the influence function (IF), also called influence curve (IC), for the circular median as well as the circular mean direction. They observe that the circular median is sensitive to rounding or grouping of data. Mardia (1972, p.29) recommends adopting the convention used for the linear median to obtain the circular median for grouped data.

Like in the case of the median for linear data, circular median is defined separately for odd and even number of observations.

Figure 1.6: Circular median for even and odd number of observations

The median(\mathbf{P}) is the midpoint of two observations.

The median(\mathbf{P}) is one of the observations.

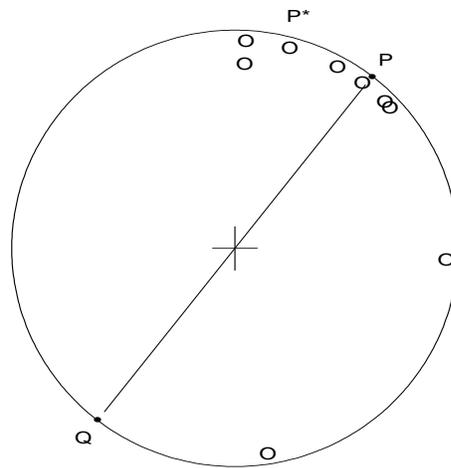


When n is odd, the sample median is one of the data points. When n is even, the sample median is taken to be the midpoint of two appropriate adjacent data points. Figure 1.6 depicts the circular median for even and odd sample sizes. Note the balance between the number of points in both half circles. In both cases, the majority of sample points are closer to \mathbf{P} (median), than to \mathbf{Q} (antimedian).

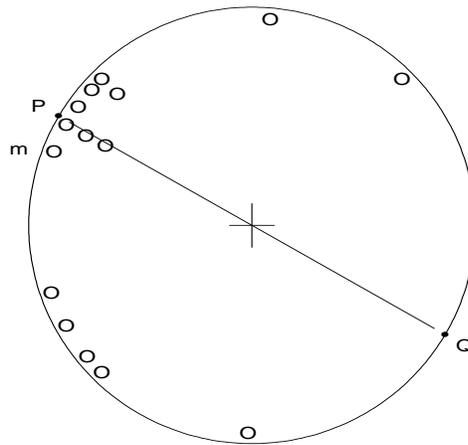
However, procedures that are based on ranking data for computing the median for linear cannot be applied to circular data. For example, consider the following data set (in degrees) 43, 45, 52, 61, 75, 88, 88, 279, 357 (Ackermann, 1997), shown in Figure 1.7. If we treat these data as linear measures, then the median is 75° . However when considered as circular variables, the median is 52° . Clearly $52^{\circ} \neq 75^{\circ}$, and in addition, a line through 75° will

not lead to equal number of observations on each semi-circle.

Figure 1.7: P^* : Linear Median of 75° , P : Circular Median of 52° for Ackermann 1997 Data.



Also, the mean and median directions typically yield different estimates of preferred direction. Figure 1.8 shows an example in which the circular median (denoted by \mathbf{P}) is one of the sample values, while the circular mean (denoted by \mathbf{m}) is not necessarily one of the sample values.

Figure 1.8: Circular mean, \mathbf{m} , median, \mathbf{P} , and antimedian, \mathbf{Q} , of a sample.

The two can coincide if the underlying distribution is symmetric about the reference direction. Ease of computation and availability of relevant statistical theory (e.g., for calculating confidence regions or pooling independent estimates of the same quantity) makes the mean direction the most commonly used measure of preferred direction, particularly for moderate to large samples (Fisher, 1993, p. 72). However, the median direction is preferred for small samples which are clustered around a single value (Shepherd and Fisher, 1982). Robust estimation of the preferred direction for the von Mises distribution has been based mainly on the median direction approach.

1.3.3 The Measures of Dispersion

The measures of spread associated with the circular mean and the circular median directions are the circular variance and the circular mean deviation respectively (Mardia, 1972).

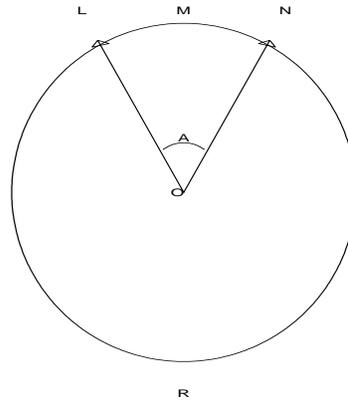
The circular variance \mathbf{S}_0 , is a common dispersion statistic defined in terms of the length of the standardized resultant vector using

$$S_0 = 1 - \bar{R}, \quad (1.8)$$

where $0 \leq S_0 \leq 1$ since $0 \leq \bar{R} \leq 1$. Minimum variation occurs when $S_0 = 0$ ($\bar{R} = 1$), and corresponds to all of the observations in a given sample occurring at precisely the same location. A natural upper limit to the possible variation occurs for data uniformly distributed around the circle, and corresponds to $S_0 = 1$ ($\bar{R} = 0$). Calculation of \bar{R} , and hence \mathbf{S}_0 is straightforward, and the interpretation of results does not depend on assumptions about the original data (Anderson-Cook, 1996). See Mardia (1972, p.45) for additional properties of the circular variance. Some authors including Batschelet (1981), refer to the quantity $2(1 - \bar{R})$ as the circular variance. We shall use the definition given in equation (1.8).

The distance between \mathbf{L} and \mathbf{N} in Figure 1.9 can be the length of the arc \mathbf{LMN} or the arc length \mathbf{LRN} . But since the arc \mathbf{LMN} is shorter, the *circular distance* between \mathbf{L} and \mathbf{N} is defined to be the arc length \mathbf{LMN} . Note the angular distance $|L, N| = |L - N|$ is not periodic and is therefore linear, not a circular variable. Consequently, only linear statistical methods should be applied to angular distances, Batschelet (1981, p.4).

Figure 1.9: Circular distance is the arc length LMN.



The circular mean deviation is a measure of spread associated with any measure of the preferred direction, $\tilde{\theta}$. It is defined about $\tilde{\theta}$ using

$$d(\tilde{\theta}) = \pi - \frac{1}{n} \sum_1^n | \pi - | \theta_i - \tilde{\theta} ||, \quad (1.9)$$

that is, the mean distance between the preferred direction and the data points. Mardia (1972, p. 30-31) showed that it has a minimum when we use the sample median as the measure of the preferred direction.

We define the circular median absolute deviation from $\tilde{\theta}$ to be

$$median(| \theta_1 - \tilde{\theta} |, \dots, | \theta_n - \tilde{\theta} |), \quad (1.10)$$

that is, the median absolute distance between the preferred direction and the data points, where n is the sample size.

To evaluate our new measure of preferred direction, we shall use the three measures; circular variance, circular mean deviation and circular median absolute deviation.

Chapter 2

Literature Review

2.1 Directional Data

Introduction

The standard texts on directional data are Mardia (1972) and Fisher (1993). Batschelet (1981) gives a less mathematical account of applications of circular data to the analysis of biological data. Fisher, et al. (1987) gives an account of methods for the analysis of spherical data. Upton and Fingleton (1989), and Mardia & Jupp (2000) discuss both two and three dimensional data. Most recently Jammalamadaka & SenGupta (2001) discuss circular data.

The commonly used parametric model, the von Mises distribution, for analyzing directional data assumes unimodality and axial symmetry of a given data set. Since this is not always the case, the search for robust methods leads naturally to techniques which are nonparametric or are distribution free. In linear inference one can justify an assumption of normality [for example, when one deals with averages of large samples], but there is no corresponding rationale for invoking the von Mises distribution in directional inference. As a consequence,

the need for distribution-free methods is highly desirable in directional data analysis (Rao, 1984).

Inappropriate applications of linear methods to circular data are in the book by Zar (1999, p. 607, 624-625), who does not consider the fact that, the zero or positive direction in the circle is arbitrary. Many of his proposed methods are not rotationally invariant.

2.1.1 Robust Estimation of Preferred Direction

The finiteness of the circle creates new challenges as readily as it solves others. While linear approximations may solve ad hoc data analysis problems, they are not suitable for routine data processing (Fisher, 1993). For example, the investigation of Wehrly and Shine (1981) of the robustness properties of both the circular mean and median using influence curves, revealed that the circular mean is quite robust, in contrast to the sample mean on the real line. Watson (1983a, 1986) computed the influence functions of some classical estimators of location and scale for circular distributions. Lenth (1981) defined a class of M-estimators for the location of circular distributions. While Mardia (1972), Fisher (1985) and Durcharme and Milasevic (1987a) generalized the median to the sphere. Durcharme and Milasevic (1987b), derived the asymptotic distribution and the asymptotic relative efficiency for the circular median for symmetric distributions on the circle. They show that in the presence of outliers, the circular median is more efficient than both the mean direction and an estimator proposed by Watson (1983a). A survey was provided by Jupp & Mardia (1989).

Other notable results involving the circular median are for example by Liu & Singh (1992) and Purkayastha (1991, 1995a, 1995b). Liu & Singh (1992), call the Mardia's (1972) median for circular data the arc distance median for circular data. They obtain it by minimizing the sum of geodesic distances of an arbitrary point on the circle given a set of observations on the circle and by simultaneously searching for the point where this minimum is attained.

Purkayastha(1991) introduced a rotationally symmetric model of the form

$$f(\theta, \underline{\mu}, \kappa) = C(\kappa) \exp[\kappa \cos^{-1}(\theta^T \underline{\mu})], \quad (2.1)$$

where $C(\kappa)$ is some function of κ , for which the sample circular median is the maximum likelihood estimate of μ . Purkayastha (1995a), derives the asymptotic linear representation Mardia's (1972) median for circular data.

Batschelet (1981, p. 18-19) commented on the possibilities of obtaining several disjoint ranges for the median given small samples and disperse data. He concluded that the median has limited practical use unless data are clustered about a single mean. Anderson (1993, p. 106) observed that the non-uniqueness of the circular median is more pronounced for small data sets with small concentration. The breakdown point and the M- and L-estimators of the mean direction for the von Mises distribution have been studied by He and Simpson (1992). They conclude that the directional median is the most standardized bias (SB)-robust estimator for the circular mean. This result is analogous to the linear data case in which where the median is also the most SB-robust estimator of the mean (its Kullback-Liebler standardized breakdown slope is bounded away from 0 uniformly over the set of positive real numbers). The notion of standardized bias robustness was introduced by Ko and Guttorp (1988). They standardized the influence function with respect to the dispersion of the data and showed the non-robustness of the classical (maximum likelihood) estimators for spherical distributions. Some aspects of Bayesian inference for the von Mises-Fisher distribution are provided by Mardia & El-Atoum (1976).

2.1.2 Outliers in Circular Data

The outlier problem in directional data is somewhat different from that in the linear case. An effort to understand outliers in directional data, has led many to question if indeed there is an outlier problem for directional data, Small (1990). A detailed discussion of the outlier problem in directional data can be found in Ko and Guttorp (1988).

How far an observation is from the “center” in the directional data setup should be judged by using an appropriate “circular distance”. Jammalamadaka and SenGupta (2001), state that unlike in the linear case, outliers in circular data need not be too large or too small, but could be away from the “central” part of the data. Current research on dealing with the handling of unusual observations (or outliers) in the circular/directional data can be grouped into two major areas. The first area is robust statistical methods (whereby outlying observations are automatically given less weight in estimation) and the second is for the detection of outliers techniques (whereby outlying observations are deleted from the sample using objective criteria). In this dissertation, we shall restrict ourselves to the area of robust statistical methods.

When the data come from a disperse distribution on the circle, a small amount of contamination by outliers would not be noticed and would have little effect on estimates of location or spread, (Lenth, 1981). We focus our attention on cases where the majority of the data come from a fairly narrow distribution, but there is a possibility that a few spurious observations may be present. For example consider data from a study of the flight patterns of birds, in which most of the birds are migrating but a few are distracted. We shall use the word “outlier” to refer to a surprising observation, which is suspiciously far from the main data mass.

Misrecording of data, unwittingly sampling from a second population or vagaries of sampling

resulting in the occasional isolated values are some of the ways in which outliers can occur in circular data. Such an observation will have the largest angular deviation from the average direction. Barnett and Lewis (1994) and Beckman and Cook (1983), include the results of Collett (1980), who analyzed the performance of four statistics for the detection of a single outlier. A Bayesian approach to outlier problem has been studied by Bagchi and Guttman (1990).

2.2 Hodges-Lehmann Estimator for Linear Data

For linear data, the sample mean and median are estimates of the population mean and median, respectively. The sample mean is highly sensitive to outliers and therefore non-robust. However, Zielinski (1987) showed that under restrictions on outliers, the sample mean may be more robust than the median. On the circle, observations have a limited range (distributions on the circle are bounded) so the findings of Wehrly and Shine (1981), who concluded that the circular mean is quite robust, are therefore not surprising.

The classical median and the Hodges-Lehmann estimate (defined as the median of pairwise averages by Hodges and Lehmann (1963)) are well-known estimates of location in one-sample univariate linear problems. The Hodges-Lehmann estimator is derived from the Wilcoxon signed-rank test statistic. Sirvanci (1982) noted that the Hodges-Lehmann estimator is a nonlinear estimator which is asymptotically best linear unbiased for the logistic distribution, $F(x) = [1 + \exp(-x)]^{-1}, x \geq 0$.

In this section, we shall give a brief review of the Hodges-Lehmann estimate. The motivation for the Hodges-Lehmann estimator (due to Hodges and Lehmann (1963)) is as follows. Given a random sample X_1, \dots, X_n , let $T(X_n) = T(X_1, \dots, X_n)$ be a test statistic for some hypothesis concerning a location parameter θ , say. Suppose that $T(X_n - \mathbf{a}\mathbf{1}_n) =$

$T(X_1 - \mathbf{a}, \dots, X_n - \mathbf{a})$ is a monotone (decreasing) as a function of \mathbf{a} , and that values of T near $T = 0$ are in accordance with the null hypothesis. If T is a rank statistic, it is likely to be a step-function (Fisher, 1982). Choose as the estimator of θ , the value of \mathbf{a} for which T is closest to 0.

For a random sample of size n denoted by X_1, \dots, X_n , the procedure for calculating the Hodges-Lehmann estimator of the population center and a $(1 - \alpha)100\%$ confidence interval for the population median is as follows:

- 1) Form the $M = n(n + 1)/2$ Walsh averages, $W_{ij} = \frac{X_i + X_j}{2}$, $1 \leq i \leq j \leq n$.
- 2) Arrange the W_{ij} 's in order of magnitude from smallest to largest. Let $W^{(1)} \leq \dots \leq W^{(M)}$, denote these ordered values.
- 3) The median of the W_{ij} 's, given by

$$\tilde{\theta} = \begin{cases} W^{(k+1)}, & \text{if } M = 2k + 1 \\ \frac{W^{(k)} + W^{(k+1)}}{2}, & \text{if } M = 2k \end{cases}, \quad (2.2)$$

is the Hodges-Lehmann estimator of the population center.

Further, define

$$T^+ = \# \left(\frac{X_i + X_j}{2} > 0 \right), \quad i \leq j. \quad (2.3)$$

If $P(T^+ \leq n^*) = \alpha/2 = P(T^+ \geq (M - n^*))$, then $[W^{(n^*+1)}, W^{(M-n^*)}]$ is the $(1 - \alpha)100\%$ confidence interval for the population median based on T^+ , where

$$n^* = n(n + 1)/4 - 0.5 - Z_{\alpha/2} \sqrt{\left(\frac{(n+1)(2n+1)}{24} \right)} \quad (\text{Hettmansperger, 1984, p. 39}).$$

Some properties of the Hodges-Lehmann estimator for location are stated without proof

below:

Property 1: $\tilde{\theta}$ is translation invariant (Fisher, 1982).

Property 2: $\tilde{\theta}$ is unbiased for θ (Manoukian 1986, p. 190-191).

Property 3: $\tilde{\theta}$ is median unbiased for θ , that is, $P_{\theta}(\tilde{\theta} \leq \theta) = 1/2$, and θ is a median for the distribution of $\tilde{\theta}$. See Randles and Wolfe (1979, p.216). This concept is useful in situations where a Hodges-Lehmann estimator is not symmetrically distributed (and thus not necessarily median unbiased for θ). Thus the assumption of symmetry for the underlying distribution can be dropped and still have median unbiasedness for certain Hodges-Lehmann estimators under some regularity conditions.

Property 4: The distribution of $(\tilde{\theta} - \theta)$ is independent of θ (Lehmann and D'Abrera, 1998, p. 177).

Property 5: If the distribution of (X_1, \dots, X_n) is symmetric about θ , the same is true for the distribution of $\tilde{\theta}$ (Lehmann and D'Abrera, 1998, p. 177).

Property 6: Suppose that the distribution L is continuous, then the distribution of $\tilde{\theta}$ is also continuous, so that $P(\tilde{\theta} = d) = 0$ for any given d . This property distinguishes $\tilde{\theta}$ from the discrete distribution of the associated Wilcoxon Statistics (Lehmann and D'Abrera, 1998, p. 178).

The main advantage of the Hodges-Lehmann estimator is that it is robust against outliers in the one-sample problem through Hampel's robustness measures, namely the breakdown

point and the influence curve (more generally known as the influence function). The H-L estimator has a breakdown point of 0.29, i.e. 29% is the least portion of data contamination needed to drive the estimate beyond all bounds, (Hettmansperger and McKean, 1998). The asymptotic standard deviation (which depends on the underlying distribution) and the asymptotic normality theory were first obtained by Hodges and Lehmann (1963). Sheather (1987) described a method of estimating the asymptotic standard error of the Hodges-Lehmann estimator based on generalized least squares. The efficiency of this estimator is discussed by among others, Hodges & Lehmann (1963), Bickel (1965), Hoyland (1965), and Gastwirth & Rubin (1969).

Note, that the total number of Walsh averages is n^2 , but computation of the Hodges-Lehmann estimate requires only $n(n+1)/2$. Huber (1981, p. 9, 63) observes that the Hodges-Lehmann estimate is the median of all n^2 pairwise averages, however, the more customary versions use only $n(n-1)/2$ or $n(n+1)/2$ pairwise averages, since all three are asymptotically equivalent. We shall develop new measures of preferred direction for circular data which are analogous to the Hodges-Lehmann estimate of center for linear data described above. For one measure, we use $n(n-1)/2$ pairwise circular means excluding the data paired with itself. The other new measure uses the $n(n+1)/2$ pairwise circular means including the data paired with itself. We shall also use all the n^2 pairwise circular means to obtain a third measure. We anticipate that the three new measures will be asymptotically equivalent and will in circular data study their properties by simulation.

2.3 Influence Functions for Estimates of Location for Linear Data

The influence function is an important concept in describing the robustness of estimators. This concept is best understood by thinking of estimators and their corresponding parameters

as functionals (“functions of functions”). In our case, functionals will be functions of the cumulative density function (CDF) \mathbf{F} or the empirical cumulative density function (ECDF) \mathbf{F}_n , where $F_n = \frac{\#(x^{(i)} \leq x)}{n} = \frac{i}{n}$ at $x = x^{(i)}$. Functions of the CDF are the parameters of interest and the functions of the ECDF are the estimators of these parameters. Note, since the ECDF converges uniformly to the CDF as n goes to infinity (by the “Glivenko Cantelli Lemma”), this implies that the estimators based on the empirical substitution principle converge in probability, to the parameter, i.e., $T(F_n) \xrightarrow{Prob.} T(F)$. Hence, studying certain properties of estimator $T(F_n)$, when based on large samples, can be used as approximations of those properties of the parameter $T(F)$. Common examples of functionals include mean, variance and median functionals. These are respectively,

$$T_1(F) = \int x dF(x) = \begin{cases} \int x f(x) d(x), & \text{x continuous} \\ \sum x_i f(x_i), & \text{if x is discrete,} \end{cases} \quad (2.4)$$

$$T_2(F) = \int (x - T_1(F))^2 dF(x) = \int x^2 dF(x) - [\int x dF(x)]^2, \quad (2.5)$$

$$\text{and} \quad T_{\frac{1}{2}}(F) = F^{-1}\left(\frac{1}{2}\right) \quad (2.6)$$

The corresponding functional estimators are

$$T_1(F_n) = \int x dF_n(x) = \frac{\sum_{i=1}^n x_i}{n} = \bar{x}, \quad (2.7)$$

$$T_2(F_n) = \int (x - T_1(F_n))^2 dF_n(x) = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = s^2, \quad (2.8)$$

$$\text{and} \quad T_{\frac{1}{2}}(F_n) = \begin{cases} F_n^{-1}\left(\frac{n+1}{2n}\right), & \text{n odd} \\ \frac{1}{2}F_n^{-1}\left(\frac{1}{2}\right) + \frac{1}{2}F_n^{-1}\left(\frac{1}{2} + \frac{1}{n}\right), & \text{n even} \end{cases} \quad (2.9)$$

(Birch, 2002, p. 167). For theoretical convenience,

$$T_{\frac{1}{2}}(F_n) = F_n^{-1}\left(\frac{1}{2}\right), \text{ for large } n. \quad (2.10)$$

The behavior of functionals can be studied by examining the behavior of its derivative, just like in the case of functions where examining its derivative enables one to learn about the

function. The derivative of a functional is called the Influence Function. The influence function is interpreted as the asymptotic bias of $\mathbf{T}(\mathbf{F})$ arising from a small contamination of the distribution $\mathbf{F}(\mathbf{x})$ at x_0 . Simply, the influence function measures the “influence” of x_0 on the parameter $\mathbf{T}(\mathbf{F})$. That is, it measures the change in $\mathbf{F}(\mathbf{x})$, in the limit, as the point x_0 is given special emphasis, through the weight ϵ . Therefore, a desirable property for an estimator is a bounded influence function.

The influence function (IF) (Hampel, 1974) of an estimator $\mathbf{T}(\mathbf{F})$ at a distribution \mathbf{F} is the directional derivative from above

$$IF(x_0; T, F) = \lim_{\epsilon \rightarrow 0^+} \frac{[T((1 - \epsilon)F(x) + \epsilon\delta_{x_0}(x)) - T(F(x))]}{\epsilon}, \quad 0 \leq \epsilon \leq 1, \quad (2.11)$$

where the estimator is regarded as a function of the empirical distribution \mathbf{F}_n ,

$T_n(x_1, \dots, x_n) = T(F_n)$ and $\delta_{x_0}(x)$ denotes the distribution that puts mass 1 at the point x_0 , i.e,

$$\delta_{x_0} = \begin{cases} 0, & x < x_0 \\ 1, & x \geq x_0 \end{cases}, \quad (2.12)$$

and $(1 - \epsilon)F(x) + \epsilon\delta_{x_0}(x)$ is an altered CDF which places special emphasis on the point x_0 .

The influence function can also be used to evaluate the asymptotic efficiency of an estimator, since the asymptotic variance can be obtained by integrating the square of the influence function. See for example Birch (2002, p. 175-179). Note for any asymptotically normal estimator, i.e.

$$\sqrt{n}(T_n - T(F)) \rightarrow N(0, V(T, F)), \quad (2.13)$$

the asymptotic variance $V(T, F)$ is given by $V(T, F) = \int IF^2(x; T, F)dF(x)$.

We shall use the influence function to study the robustness properties as well as to obtain the asymptotic distribution of our new measures of preferred direction.

2.4 Bootstrap Methods

2.4.1 Introduction

Bootstrap methods have found use in statistics in situations where distributional assumptions are kept to a minimum or when distributional results for the quantity of interest do not exist. The distribution of a statistic can be assessed by obtaining B resamples of the data by sampling from a surrogate for the population distribution namely the data. The statistic of interest is evaluated for each of the bootstrap samples and the variability of these B values is taken as an estimate of the variability of the statistic over the population. The fundamental assumption of bootstrapping is that the observed data are representative of the underlying population. By resampling observations from the observed data, the process of sampling observations from the population is mimicked. A general reference for bootstrap methods is Efron & Tibshirani (1993).

Attention to the use of bootstrap methods for directional data has been made in part by Watson (1983b) and Fisher et. al.(1987). Mardia & Jupp (2000, p. 277) advocate the use of bootstrap methods for directional data since the distributions of the statistics commonly used for inference are frequently intractable. Fisher and Hall (1992) applied the notions of pivoting, percentile-t confidence region, the iterated bootstrap and the parametric bootstrap to the problems of calculating a confidence region for a mean direction, finding an interval estimate for the concentration parameter of the von Mises distribution and testing the *closeness* of several mean directions. In addition, bootstrap confidence regions based on pivotal statistics have been noted to have smaller coverage error, or boundaries closer to those of exact confidence regions than regions derived from nonpivotal statistics (Fisher and Hall, 1989).

Two gaps in bootstrap theory for directional data deal with non-i.i.d observations common

in geology where observations are hardly ever randomly selected (Rao, 1975) and the lack of a commercial computer package for the analysis of directional data (Jupp and Mardia, 1989). In the case of linear data, Liu (1988) and Liu and Singh (1995), have shown that use of classical i.i.d bootstrap on data that is not i.i.d is frequently appropriate since it captures the first order limit and also retains second order asymptotic properties in the case of the sample mean.

In this dissertation, our new measures of center are derived from the pairwise circular means. Mardia (1972, p.98) asserts that the marginal probability density function of the circular mean cannot be simplified. The distribution of the circular mean direction is intractable, hence bootstrap methods provide the only alternative for coming up with such a distribution, particularly for small samples.

2.4.2 Confidence Intervals Involving Measures of Preferred Direction

Confidence intervals for the mean direction can be obtained either by obtaining the parametric standard error, Fisher (1993, p.88-89, p. 206) or by nonparametric bootstrap, Fisher and Powell (1989) and Fisher (1993, p. 206). For $n > 25$, the estimate of the circular standard error of $\bar{\theta}$ is given by $\hat{\sigma}_{VM} = \frac{1}{\sqrt{(n\bar{R}\hat{\kappa})}}$, where $\bar{\theta}$ and $\hat{\kappa}$ are the estimates of μ (given by equation (1.6)). \bar{R} is the mean resultant length (see Section 1.3.1), and $\hat{\kappa}$ (given below).

$$\hat{\kappa} = \begin{cases} 2\bar{R} + \bar{R}^3 + \frac{5\bar{R}^5}{6}, & \bar{R} < 0.53 \\ -0.4 + 1.39\bar{R} + \frac{0.43}{(1-\bar{R})}, & 0.53 \leq \bar{R} < 0.85 \\ \frac{1}{(\bar{R}^3 - 4\bar{R}^2 + 3\bar{R})}, & \bar{R} \geq 0.85 \end{cases} .$$

A $100(1 - \alpha)\%$ confidence interval for μ is given by $\bar{\theta} \pm \sin^{-1}\left(z_{\frac{1}{2}\alpha}\hat{\sigma}_{VM}\right)$, where $z_{\frac{1}{2}\alpha}$ is the upper $100(\frac{1}{2}\alpha)$ percentage point of the Normal $N(0, 1)$ distribution. See Fisher (1993, p. 88) and Jammalamadaka & SenGupta (2001, p. 96). See also Fisher and Lewis (1983) and Watson (1983a, Chapter 4).

A general class of bootstrap confidence arcs exists, Fisher and Hall (1989, 1992). These are the Symmetric arc, Equal-Tailed arc and the Likelihood-Based arc. The symmetric intervals are often shorter than equal-tailed intervals and have better accuracy even in highly asymmetric circumstances, Hall (1988a). Fisher and Hall (1989) show that the equal-tailed and likelihood based have coverage error $O(n^{-1})$ and the symmetric version has coverage error $O(n^{-2})$. These results are similar to those of Beran (1987) and Hall (1988a,b) obtained for linear data.

The Equal-Tailed arc method uses the point estimate of the preferred direction as the middle observation, then defines the endpoints of the confidence interval as the location, where $\frac{(1-\alpha)}{2}$ of the bootstrap values $\tilde{\theta}^*$ lie between the edge and the preferred direction. Fisher (1993, p. 206), refers to this procedure as the Basic Method. A $(1 - \alpha)100\%$ confidence interval for the population preferred direction is constructed as follows. Compute the difference between the preferred direction of the original sample and that of the b^{th} bootstrap sample

$$\gamma_b = \tilde{\theta}_b^* - \bar{\theta}, \quad (-\pi \leq \gamma_b < \pi), \quad b = 1, \dots, B,$$

where $\tilde{\theta}_b^*$ is the sample preferred direction for the b^{th} bootstrap sample. Take this value to be the middle observation. Define the endpoints of the confidence interval as the observation where l is the largest integer less than or equal to $(\frac{1}{2}B\alpha + \frac{1}{2})$ position and m^{th} (where $B-l$) position. Thus a $(1 - \alpha)100\%$ confidence interval for the population preferred direction is (point estimate + observation at the l^{th} position, point estimate + observation at the m^{th} position). Note, the Equal-Tailed arc procedure, is the analog of equal-tailed confidence intervals, commonly used type of bootstrap confidence interval in linear data (Efron, 1981,1982, 1985).

The Symmetric-arc method selects the angle, Δ , such that $(1 - \alpha)B$ of the $\tilde{\theta}^*$ values lie within the interval. Note that, Δ above and below the point estimate is the same. Fisher

(1993, p.206), refers to this procedure as the Symmetric Distribution Method. A $(1 - \alpha)100\%$ confidence interval for the population preferred direction is constructed as follows. Calculate the absolute difference between the preferred direction of the original sample and that of the b^{th} bootstrap sample,

$$\psi_b = |\tilde{\theta}_b^* - \bar{\theta}|, \quad b = 1, \dots, B,$$

where $\tilde{\theta}_b^*$ is the sample preferred direction for the b^{th} bootstrap sample. Note, these B values can be treated as linear variables (Batschelet, 1981, p.4). Sort the B values into increasing order to obtain $\psi_{(1)} \dots \psi_{(B)}$ say. Let $l = \text{integer part of } B\alpha + \frac{1}{2}$, $m = B - l$. A $(1 - \alpha)100\%$ confidence interval for μ is given by $\bar{\theta} \pm \Delta$, where $\Delta = \psi_{(m)}$.

A third distinct procedure for constructing bootstrap confidence intervals is the Likelihood-Based arc. This method, enables one to find the narrowest interval that satisfies the requirement of the $(1 - \alpha)100\%$ confidence interval. The likelihood-based arc is chosen as the shortest arc containing $(1 - \alpha)B$ of the θ^* values.

An alternative to the bootstrap for the circular median, with data that are not too dispersed (that is, data concentrated on an arc substantially less than the whole circumference) can be found by analogy with methods for linear data, Fisher (1993, pp. 72-73). The procedure is as follows. For any integer m greater than zero, count off m θ -values to the left and to the right of the sample median (obtained as in Section 1.4.2) (not counting the point estimate itself) to get lower and upper data values $\theta_{(L_m)}$ and $\theta_{(U_m)}$ respectively. Exact α -levels for the confidence interval $(\theta_{(L_m)}, \theta_{(U_m)})$ are given in Appendix A6, Fisher (1993), for $n < 16$. For $n > 16$, an approximate $100(1 - \alpha)\%$ confidence interval can be obtained by setting $m = 1 + \text{integer part of } \left(\frac{1}{2}\sqrt{n}\right) z_{\frac{1}{2}\alpha}$, where $z_{\frac{1}{2}\alpha}$ is the upper $100(\frac{1}{2}\alpha)\%$ of the $N(0, 1)$ distribution.

To determine relative performance of the mean, median and Hodges-Lehmann estimator,

bootstrap confidence intervals (Symmetric-Arc, Equal-Tail and Likelihood-Based) for each of the measures of preferred direction will be compared to the alternative confidence interval for the median in Chapter 5.

In order to make inferences about a particular sample of circular random variables, we shall use the fact that confidence intervals and tests of hypotheses are related in the following way. Suppose we wish to test the null hypothesis that $\mu = \mu_0$ versus the alternative that μ differs from μ_0 at the significance level α . If we have a $(1 - \alpha)100\%$ confidence interval for θ , we can construct an α -level hypothesis test by simply accepting the null hypothesis that $\mu = \mu_0$ if μ_0 is contained in the $(1 - \alpha)100\%$ confidence interval for μ and rejecting the null hypothesis if μ is outside the interval. This is the format that we shall take in this dissertation. In circular data, only two-sided alternative hypothesis are feasible since one-sided tests only make sense if the data can be sensibly ordered, which is not the case for circular data.

Chapter 3

New Measures of Preferred Direction

3.1 Introduction

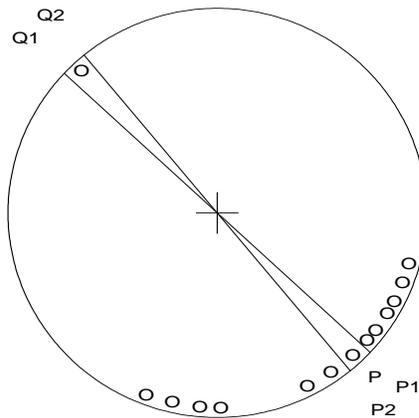
This dissertation provides an alternative to the circular mean and median, for estimating the preferred direction. An appropriate measure should be a compromise between the advantages and disadvantages of the circular median and the circular mean. Methods that offer an alternative to the lack of robustness of the sample mean are desired because outliers can have a considerable effect on estimates. For example, Fisher & Lewis (1983), observe that palaeocurrent data can sometimes have quite skewed distributions. Hence, the need for an estimate of center that will be robust under general conditions. In Section 3.2, an alternative unique solution to circular median is obtained. The new measure of preferred direction, the circular analog to the Hodges-Lehmann estimate is presented in Section 3.4.

3.2 Unique Solution to Circular Median

The procedure for computing the circular median proposed by Mardia (1972, p.28,31) occasionally leads to a non-unique estimate of the circular median since there can sometimes be two or more diameters that divide the data equally. Many authors, for example He and Simpson (1992), advocate the use of circular median as an estimate of preferred direction especially in situations where the data are not von Mises distributed. Consequently, a strategy is needed to deal with non-unique circular median estimates for small samples.

Figure 3.1: Non Uniqueness of Circular median

Both P_1 & P_2 satisfy the definition of 7 observations per semicircle



As an illustration of the non-uniqueness of the circular median, consider the Frog data

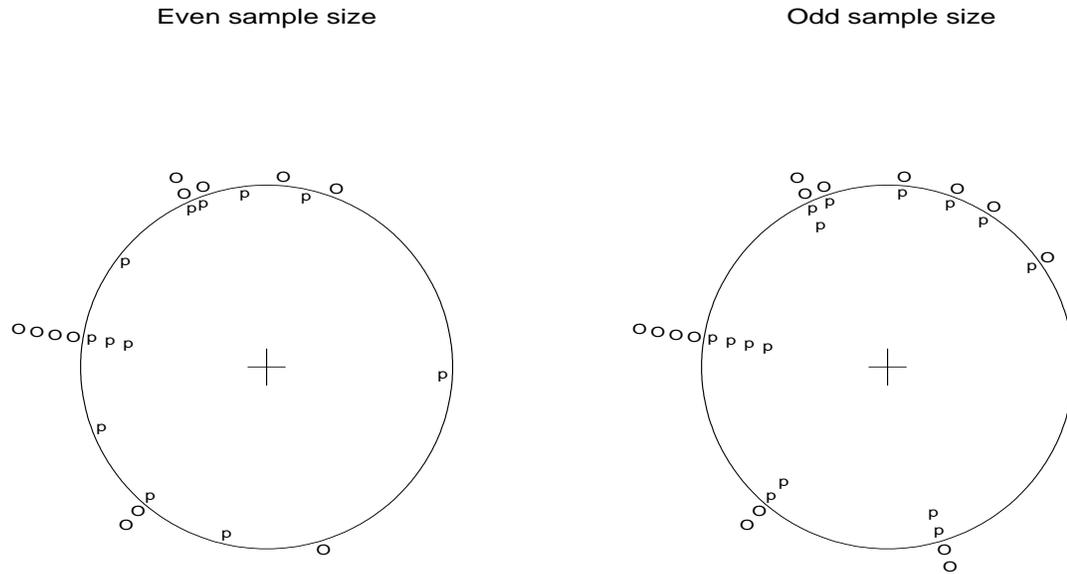
(Ferguson, et. al. (1967)) in Figure 3.1. Notice that the data points labeled P_1 & P_2 both satisfy the definition of a circular median as defined by Mardia (1972, p. 28, 31) since the diameter through each one of them leads to half the observations being on each half-circle. The points Q_1 & Q_2 are the antimedian of P_1 & P_2 respectively.

In this section, we adapt the existing definition of circular median as defined by Mardia (1972, p. 28, 31) and propose that the estimate of the population circular median be an average of two or more values that would be considered as sample medians. The point \mathbf{P} in Figure 3.1 is the circular mean of the two sample medians (P_1 & P_2). We conjecture that \mathbf{P} will be more robust to rounding and will be a unique estimate since it involves local averaging, Cabrera et.al. (1994).

We propose the following algorithm for computing the circular median. See Appendix E for an S-Plus function. Suppose $\theta_1, \dots, \theta_n$ is a random sample of circular data from a unimodal distribution.

Step 1: Find all values (denoted by \mathbf{p} in Figure 3.2) satisfying the definition of median. For even samples, the candidate values are the midpoints of all neighboring observations. For odd samples, the candidate values are the observations themselves.

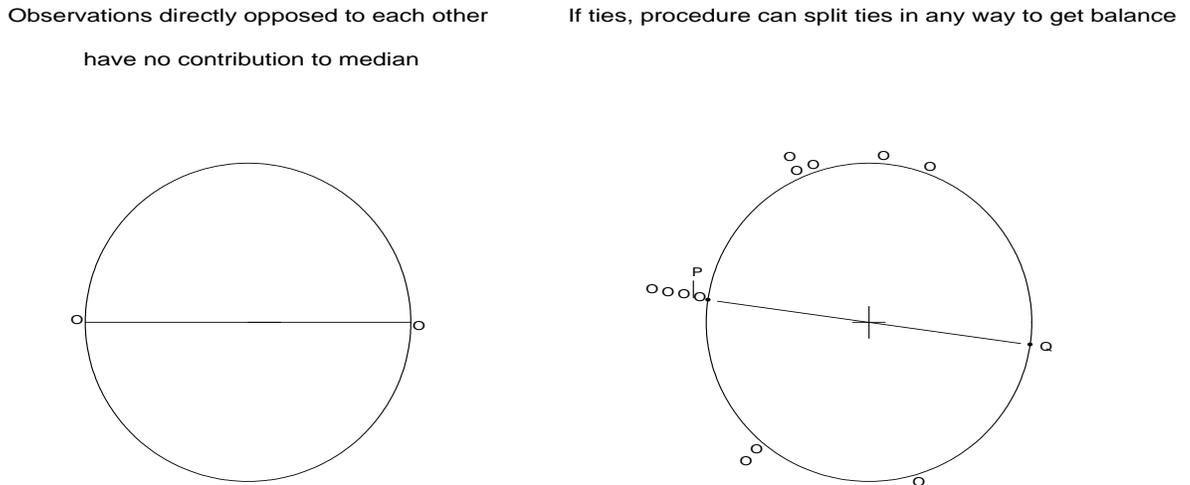
Figure 3.2: Original Observation \mathbf{o} , Potential Medians \mathbf{p}



Step 2a: For the “Mardia Median”, compute the circular mean deviation for each angle satisfying the median definition in Step 1. Select value(s) satisfying the definition of median that has (have) the smallest circular mean deviation to be the estimate of the median (Fisher, 1993, pp. 35-36). In case of ties, take the circular mean of these angles as the unique estimate of the median.

Step 2b: For the “New Median”, compute the circular mean for all angles satisfying the definition of median obtained in Step 1. This gives a unique estimate of the median.

Figure 3.3: Special features



Both of the above procedures incorporate the two features shown in Figure 3.3. Observations directly opposite each other do not contribute to the preferred direction, since in such a case the observations balance each other for all possible choices of medians. Another important feature is that of breaking ties, as these procedures have the flexibility of finding a balancing point no matter how many observations are tied, by mimicking the midranking idea for linear data.

3.3 Comparison of Mardia Median & New Median

To determine the relative performance of Mardia Median and the New Median, data was simulated from a von Mises distribution. Without loss of generality, the center of all the distributions considered was $\mu = 0$. 10000 samples each of sizes between 5 & 20 from the distributions with 4 dispersion values around the circle were obtained. For each sample, the sample circular means, and circular medians (both Mardia Median and New Median) were

computed. The results were summarized using the following measures: Circular mean of all 10000 samples given using equation (1.7), 95% Empirical Confidence Interval, that is, the central 95% of the 10000 values, Circular Variance (CV) given by equation (1.8), Circular Mean Deviation (CMD) given by equation (1.9), and Circular Median Absolute Deviation (CMAD) given by equation (1.10).

Figure 3.4 illustrates the effect of sample size on the three measures for $\kappa = 2$. All measures appear unbiased, and the confidence bands for the mean is narrowest compared to the two medians (which are identical). The confidence bands become narrower as sample size increases for all the three measures, see Figure 3.4(a). The mean has the smallest circular variance, while the two medians have identical circular variances over the whole range of sample sizes considered, see Figure 3.4(b). In terms of circular mean deviation (CMD), the two medians have the smallest compared to the mean as expected, see Figure 3.4(c). Similarly, the two medians have the smallest circular median absolute deviation (CMAD) compared to the mean, see Figure 3.4(d). These results were similar for other concentration parameters studied as well.

The effect of concentration parameter on the three measures of preferred direction is illustrated in Figure 3.5 for $n = 20$. All measures appear unbiased, and the confidence bands for the mean is narrowest compared to the two medians (which are identical). The confidence bands become narrower as the concentration parameter increases for all the three measures, see Figure 3.5(a). The mean has the smallest circular variance, while the two medians have identical circular variances over the whole range of concentration parameter considered, see Figure 3.5(b). The circular mean deviation (CMD) for the three measures are nearly identical, with the two medians have the smallest compared to the mean, see Figure 3.5(c). Similarly, the two medians have the smallest circular median absolute deviation (CMAD) compared to the mean, see Figure 3.5(d). These results were similar for other sample sizes studied as well.

Figure 3.4: Mardia Median & New Median for VM(0, 2)

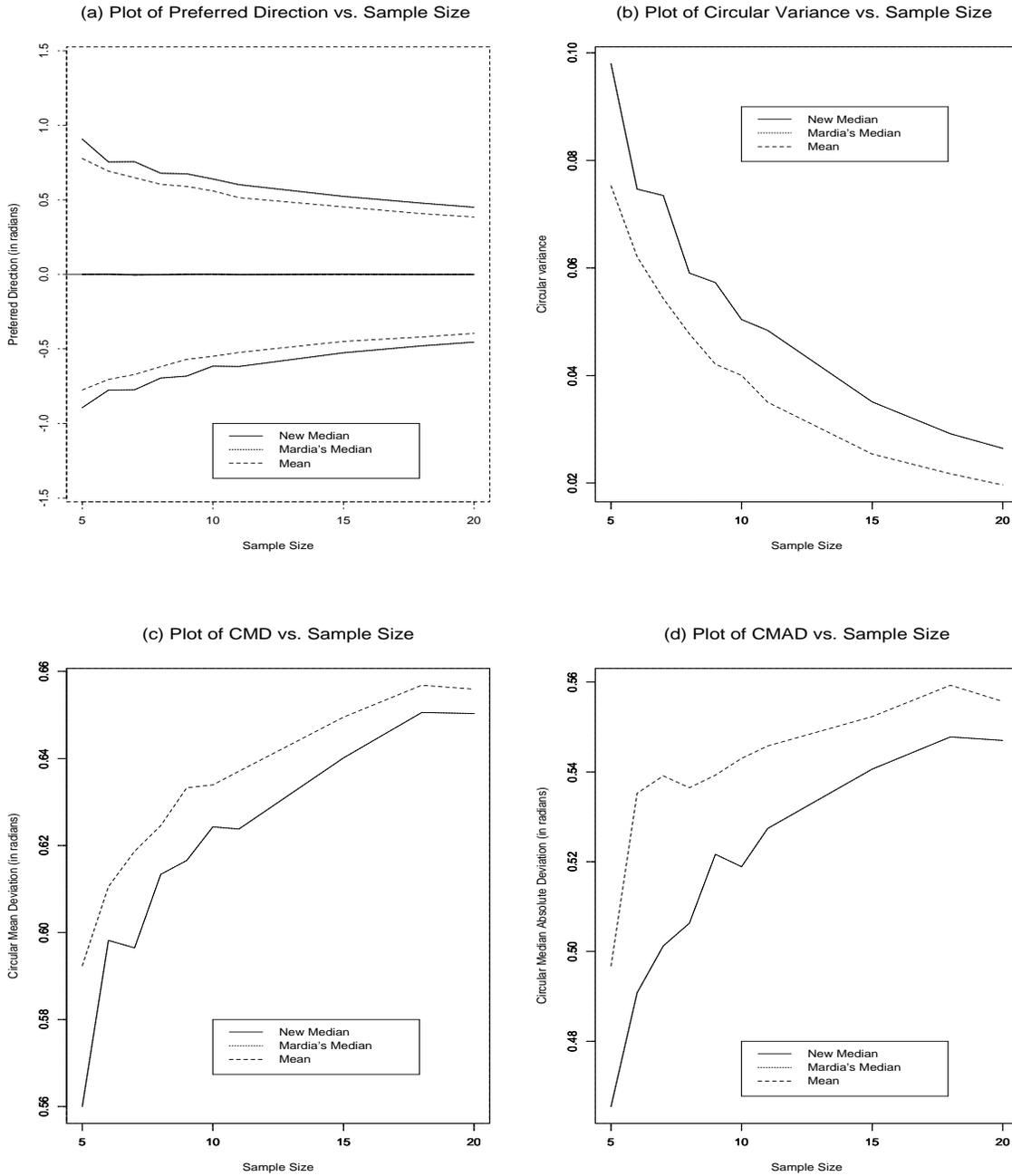
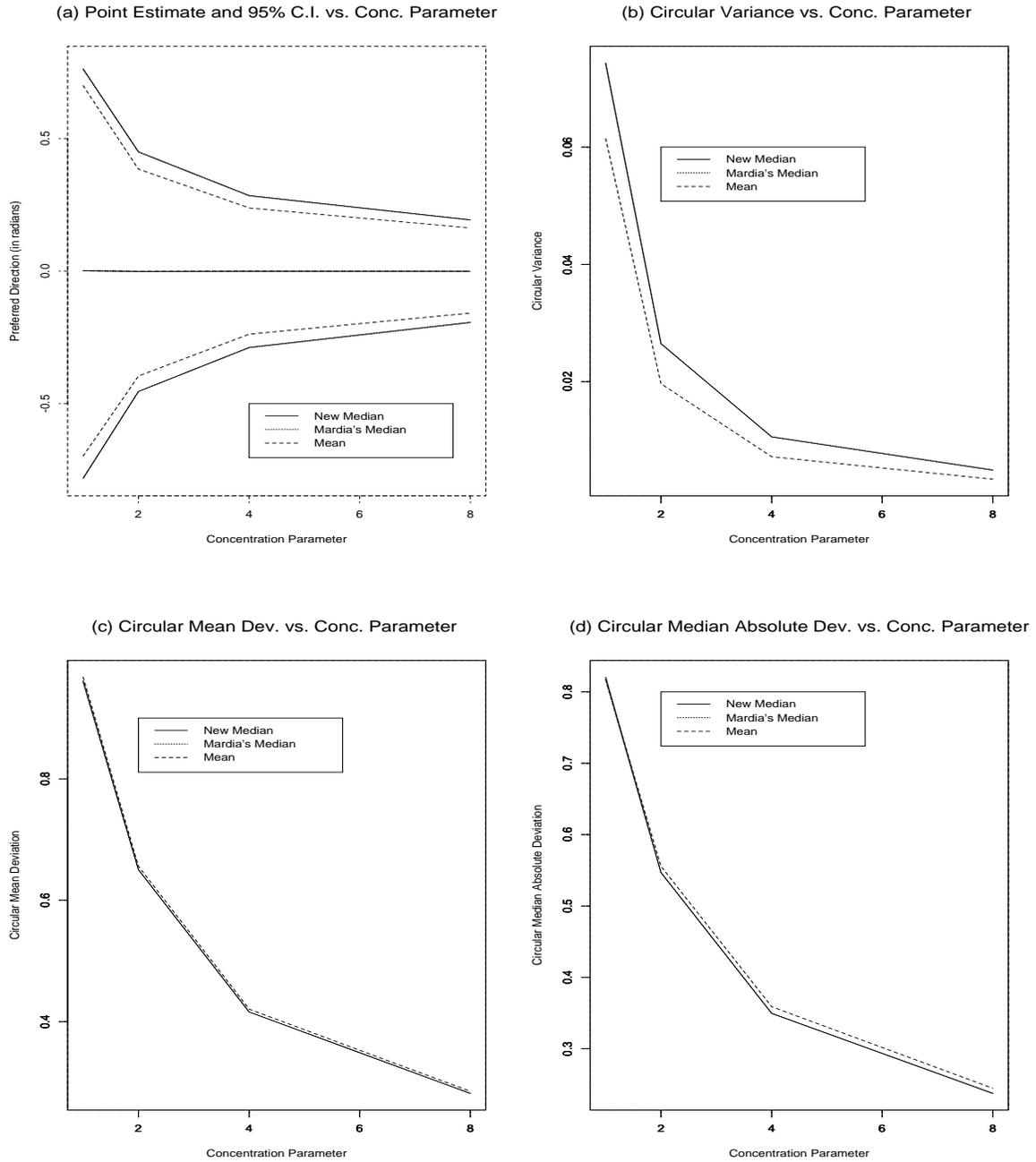


Figure 3.5: Mardia median & New median for VM $(0, \kappa)$, $n = 20$



3.3.1 Discussions and Conclusions on Relative Performance of Mardia Median and New Median

Different concentrations give similar results for von Mises data, see *Appendix A*. For a fixed sample size, the Mardia Median and New Median are identical for all combinations of sample sizes and concentrations studied, see *Appendix A* for $n = 10$. Most strikingly, the two estimators, Mardia Median and New Median are approximately identical, this implies that either of the two can be used as an estimate of preferred direction. We shall use Mardia Median in this dissertation, since it is the better established of the two measures. However, computationally, the new measure is easier and faster to work with. The two measures are both meaningless for uniform and bimodal data since “preferred direction” is not sensible here. Both Mardia Median and New Median are robust alternatives to the mean.

3.4 New Measures of Preferred Direction for Circular Data

The motivation behind the new measures of preferred direction is two-fold. First we want a compromise between circular mean (occasionally non-robust) and the robust circular median. Second, we want an estimate of preferred direction that down weights outliers sparingly and is more robust to rounding and grouping. The circular median down weights outliers significantly but is sensitive to rounding and grouping (Wehrly and Shine, 1981).

As discussed in section 2.2, the Hodges-Lehmann estimate (Hodges-Lehmann, 1963) for location for linear data is the “median” of all n^2 pairwise “averages” (Walsh averages). More customary versions use only $n(n - 1)/2$ (using all distinct pairs of observations) or $n(n + 1)/2$ (using the observations plus all distinct pairs of observations) pairwise means, since all three are asymptotically equivalent (Huber, 1981, p.63). In Figure 3.6, we consider

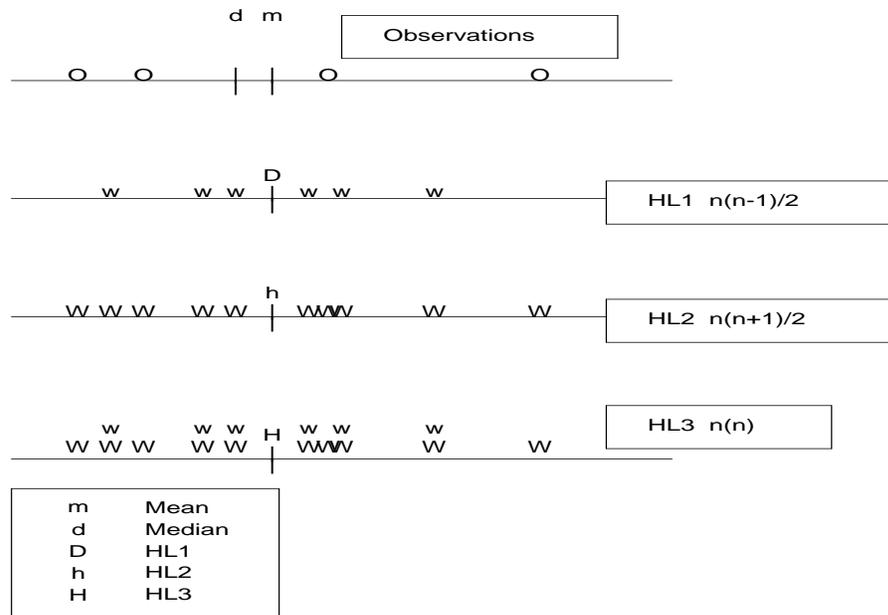
a simple example with 4 linear observations denoted by circles on the top line. The mean and median are denoted by \mathbf{m} and \mathbf{d} , respectively, on the top line. The HL1 estimate denoted by \mathbf{D} is the median of pairwise means excluding observation with itself, denoted by \mathbf{w} on top of the second line. The HL2 estimate denoted by \mathbf{h} is the median of the pairwise means used to obtain HL1 plus individual observations. While HL3 denoted by \mathbf{H} is the median of all possible pairwise means, that is observations used to obtain HL1 plus those used to obtain HL2. The Hodges-Lehmann is most commonly taken to be HL2. It is an established competitor to mean and median for the center of the distribution with excellent robustness and efficiency properties (Hollander and Wolfe, 1999, pp. 54, 74). Notice there can be different locations for some of the five measures of center for this data set, which is typical for many samples.

We propose an adaptation of this well known robust estimate for preferred direction as an alternative estimate of preferred direction for circular data. Our new measure eliminates some of the small sample problems of the circular median as noted by Anderson (1993). It is the “circular median” of all pairwise “circular means”.

Like in the linear case, we have three cases, namely, HL1, HL2 and HL3, shown in Figure 3.7. HL1 is the circular median obtained by using circular means (denoted by \mathbf{w}) of all pairs of observations, while HL2 is the circular median obtained by using pairwise circular means including individual observations, i.e \mathbf{w} 's and \mathbf{z} 's. HL3 is the circular median of all possible pairwise circular means. Notice how the circular Hodges-Lehmann estimates divide the respective circular means evenly on the two semicircles. In this example, the three measures give slightly different estimates for preferred direction.

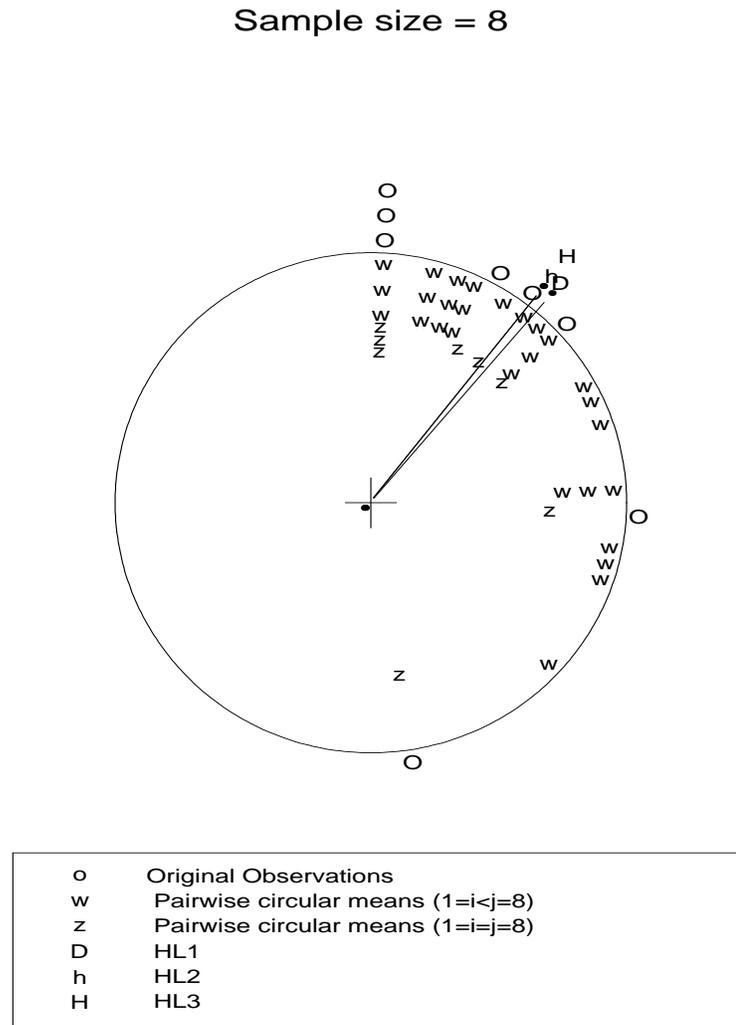
The estimates of preferred direction obtained using these procedures, are location invariant, since they satisfy the definition of circular median, which is location invariant. The approach used is feasible regardless of sample size or the presence of ties. The S-Plus function for com-

Figure 3.6: Hodges-Lehmann estimate for linear data



puting these new measures of preferred direction is given in the *Appendix E*. The circular Hodges-Lehmann estimates offer a competitive measures of preferred direction to the commonly used mean direction. Moreover, no ranking is used in computing our estimates, since on the circle there is no minimum/maximum. This distinguishes our procedure from the one suggested by Zar (1999), which is dependent on the chosen range.

Figure 3.7: Hodges-Lehmann estimate for circular data



3.5 Comparison of HL1, HL2 & HL3

3.5.1 Simulation to compare HL1, HL2, and HL3

To determine the relative performance of HL1, HL2 and HL3 as estimates of preferred direction for circular data, data was generated from a von Mises distribution. Without loss

of generality, $\mu = 0$ was used as the center of all the distributions considered. 10000 samples of sizes between 5 & 20 from the distributions with 4 dispersions around the circle were obtained. For each sample, the sample HL1, HL2 and HL3 were computed. The results were summarized using the following measures: Circular mean of all 10000 samples given by equation (1.7), 95% Empirical Confidence Interval, that is, the central 95% of the 10000 values, Circular Variance (CV) given by equation (1.8), Circular Mean Deviation (CMD) given by equation (1.9), and Circular Median Absolute Deviation (CMAD) given by equation (1.10).

Figure 3.8 illustrates the effect of sample size on the three measures for $\kappa = 2$. All measures appear unbiased, and the confidence bands are nearly identical. The confidence bands become narrower as sample size increases for all the three measures, see Figure 3.8(a). The three measures have nearly identical circular variances over the whole range of sample sizes considered, see Figure 3.8(b). HL2 and HL3 have smaller circular mean deviation (CMD), compared to HL1 over most of the range of sample sizes considered, see Figure 3.8(c). Similarly, HL2 and HL3 have smaller circular median absolute deviation (CMAD) compared to HL1, see Figure 3.8(d). These results were similar for other concentration parameters studied as well.

The effect of the concentration parameter on the three measures of preferred direction is illustrated in Figure 3.9 for $n = 20$. Again all three measures perform nearly identically. All measures appear unbiased with identical confidence bands. The confidence bands become narrower as the concentration parameter increases for all the three measures, see Figure 3.9(a). The circular variances, the circular mean deviation (CMD) and the circular median absolute deviation (CMAD) for the three measures are identical over the whole range of concentration parameters considered. As the concentration parameter increases, the circular variance, CMD and CMAD decreases for all the three measures, see Figures 3.9(b), 3.9(c) and 3.9(d) respectively. These results were similar for other sample sizes studied as well.

Figure 3.8: HL1, HL2 & HL3 for VM(0, 2)

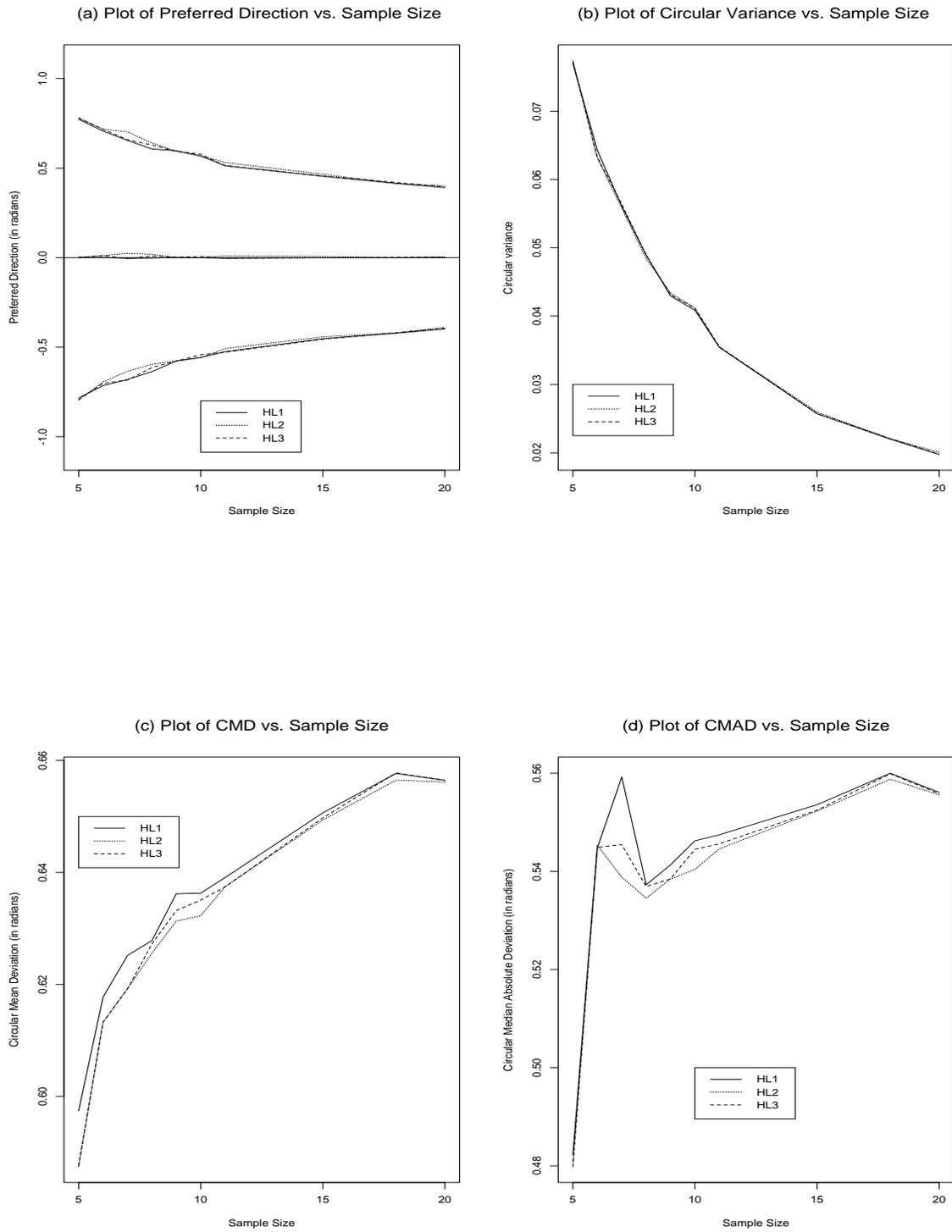
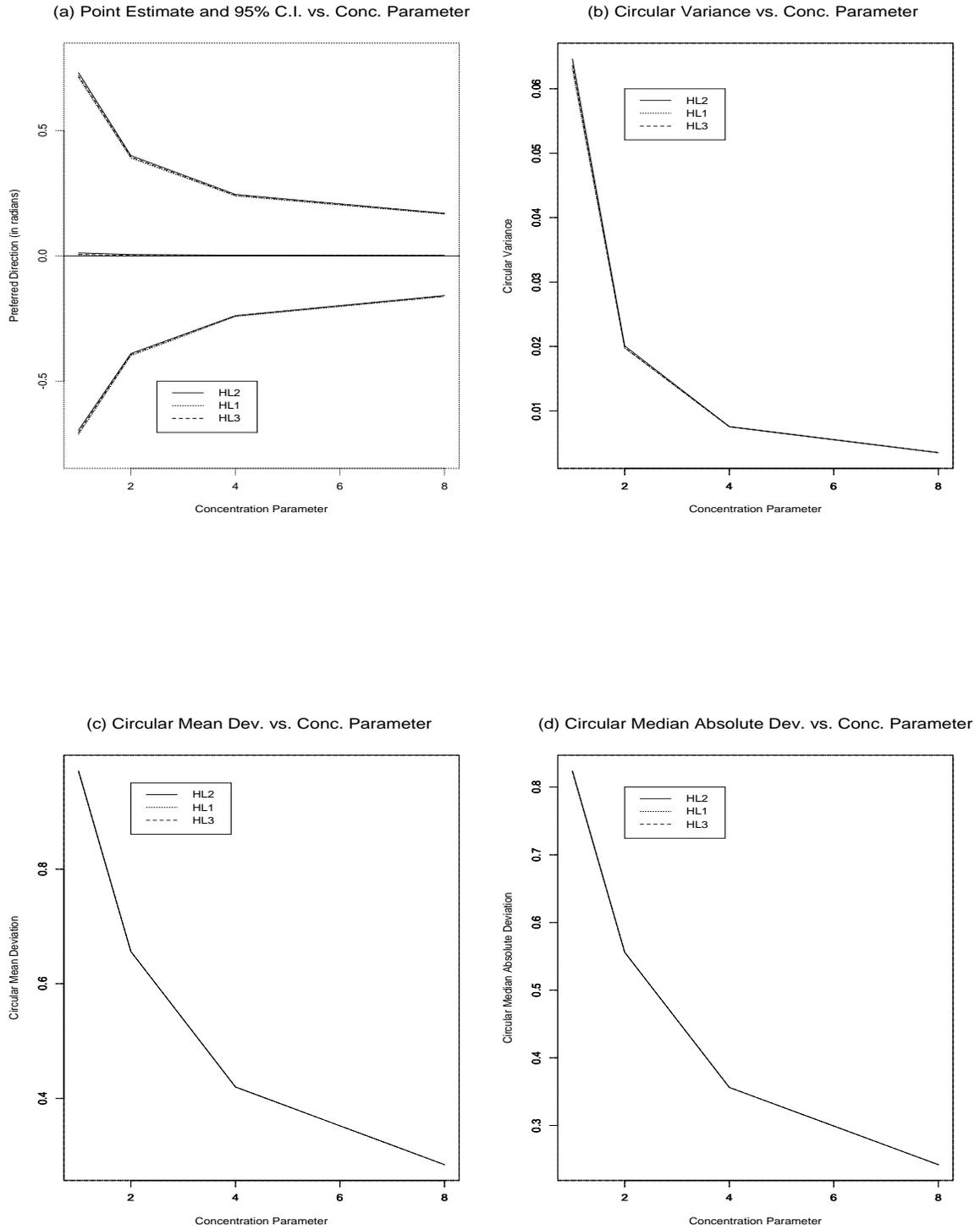


Figure 3.9: HL1, HL2 & HL3 for $VM(0, \kappa)$, $n = 20$



3.5.2 Discussions and Conclusions on Relative Performance of HL1, HL2 and HL3

Different concentrations give similar results for von Mises data, see *Appendix C*, for $k = 1, 4$ & 8 . For a fixed sample size, the three measures are virtually identical, see *Appendix C* for $n = 10$. More variability between measures is observed for smaller sample sizes. Overall, the three measures HL1, HL2 and HL3 are approximately identical, as is the case for linear data (Huber, 1981). Thus any of the three measures can be used as an estimate of preferred direction. In the remainder of this dissertation, we shall use HL2, matching the most common choice for linear data.

3.6 Theoretical Results for a Limited Range of Concentrated Distributions

3.6.1 Introduction

The new measure of preferred direction, the circular Hodges-Lehmann estimate is the circular median of the pairwise circular means. Problems of getting the distribution of circular mean direction have been noted by Mardia (1972, p. 98) and Mardia & Jupp (2000, p.69). They conclude that the marginal probability density function of the circular mean cannot be simplified. The fact that the process of wrapping is not reversible (Mardia, 1972, p.54) is also a major problem, since one cannot expect to uniquely unwrap a distribution from the circle onto the line.

Also, ignoring the circular nature of the data and treating it as linear is in general not well defined, hence one cannot sensibly obtain the Hodges-Lehmann estimator from pairwise averages without restrictions on the range of the data. A further complication is the fact that

convolutions under the most common model (the von Mises distribution) are not feasible as shown by Mardia & Jupp (2000) and Jammalamadaka & SenGupta (2001). Under such limitations, a general theory for the new measure is not easy, thus we shall restrict our discussion to data which comes from a distribution whose range is quite concentrated around a single mode. That is, data from concentrated distributions whose probability of a single observation wrapping is essentially zero.

In general, we shall obtain the distributional properties of our new measures using parametric and nonparametric bootstrap methods. See Chapter 5 for more details.

3.6.2 Approximations of von Mises distribution $VM(\mu, \kappa)$

The Wrapped Normal distribution $WN(\mu, \rho)$ is obtained by wrapping the $N(\mu, \sigma^2)$ distribution onto the circle, where we select $\sigma^2 = -2 \log \rho$, which implies that, $\rho = \exp[-\frac{\sigma^2}{2}]$. But for large κ , in particular $\kappa \geq 2$ (Fisher, 1987),

$$VM(\mu, \kappa) \quad \sim \quad N(\mu, \frac{1}{\kappa}) \quad \sim \quad WN(\mu, \exp[-\frac{1}{2\kappa}]) \quad (3.1)$$

This approximation is very accurate for $\kappa > 10$, Mardia & Jupp (2000, p. 41).

However, more generally, any von Mises distribution can be approximated by a Wrapped Normal distribution. Kent (1978) showed that the following approximation holds to a higher order in κ , that is,

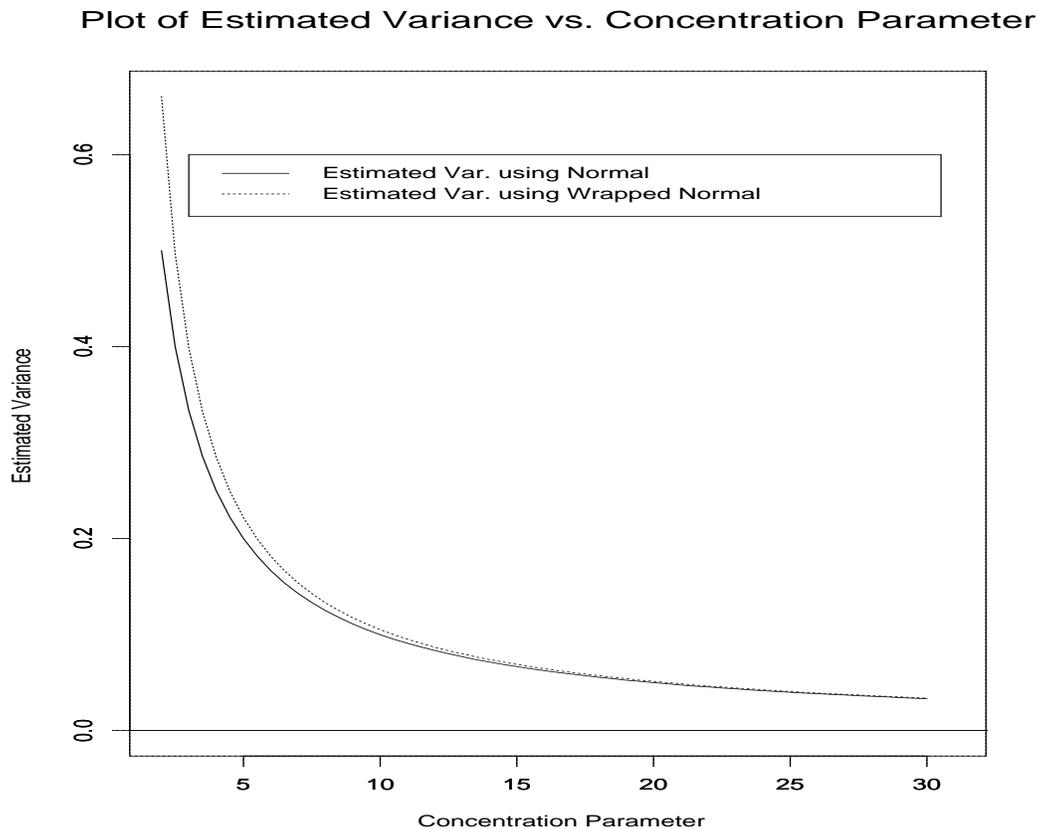
$$f_{VM}(\theta; \mu, \kappa) - f_{WN}(\theta; \mu, A(\kappa)) = O(\kappa^{-\frac{1}{2}}) \quad \kappa \longrightarrow \infty. \quad (3.2)$$

where $f_{VM}(\theta; \mu, \kappa)$ and $f_{WN}(\theta; \mu, A(\kappa))$ denote the densities of the von Mises distribution $VM(\mu, \kappa)$ and the approximating Wrapped Normal distribution $WN(\mu, A(\kappa))$, respectively. Note $A(\kappa) = \frac{I_1(\kappa)}{I_0(\kappa)}$, where $I_0(\kappa)$ & $I_1(\kappa)$ are Modified Bessel Functions of the first kind of order zero and one respectively (Abramowitz and Stegun, 1965). Mardia & Jupp (2000,

pp.37-38) derive this approximation as a first-order approximation for large κ by equating their first trigonometric moments. Numerical studies by Stephens (1963) verified that the approximation is quite satisfactory for intermediate values of κ .

Figure 3.10 shows the plots of the estimated variance against the concentration parameter, κ ; when a von Mises distribution is estimated by a Normal and a Wrapped Normal distribution for a single observation. Note $\hat{\sigma}^2 = -2\log A(\kappa)$ and $\hat{\sigma}^2 = \frac{1}{\kappa}$ are the estimates of σ^2 when VM (μ, κ) is approximated by WN $(\mu, \rho = A(\kappa))$ and N $(\mu, \frac{1}{\kappa})$ respectively. The approximations are quite close for $\kappa \geq 10$. Hence, either choice suffices.

Figure 3.10: Plot of $\hat{\sigma}^2 = [-2\log A[\kappa]]$, and $\hat{\sigma}^2 = \frac{1}{\kappa}$ vs. Concentration Parameter (κ) for a single observation

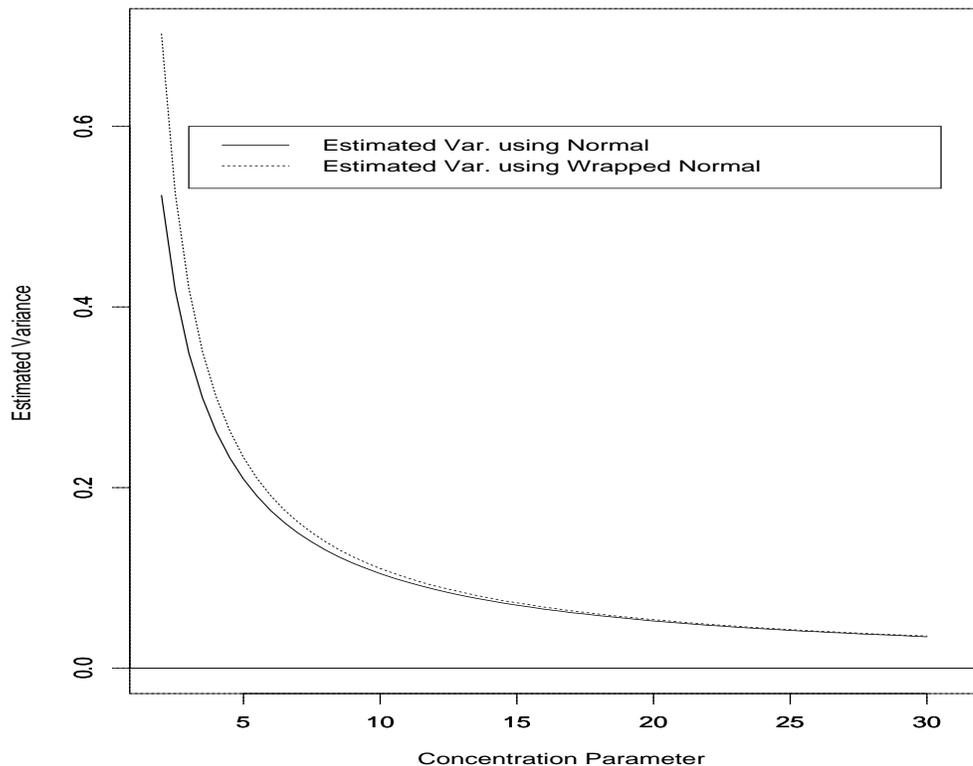


Theorem 3.6.1

For a sample from a von Mises distribution with $\kappa \geq 2$, the circular Hodges-Lehmann estimator $\hat{\theta}_{HL}^c = \text{circular median}(\bar{\theta}_{1,1}^c, \bar{\theta}_{1,2}^c, \dots, \bar{\theta}_{n-1,n}^c, \bar{\theta}_{n,n}^c)$, where $\bar{\theta}_{ij}^c$ is the pairwise circular mean of observations θ_i and θ_j defined as $\bar{\theta}_{ij}^c = \left[\tan^{-1} \left(\frac{\sin \theta_i + \sin \theta_j}{\cos \theta_i + \cos \theta_j} \right) \right], i \leq j \leq n$, is approximately distributed as $\text{VM} \left(\theta_{HL}^c, \frac{3n\kappa}{\pi} \right)$. Figure 3.11 shows how the WN and N approximations are related for various VM distributions.

Figure 3.11: Plot of $\hat{\sigma}^2 = \left[-2\log A \left[\frac{3\kappa}{\pi} \right] \right]$, and $\hat{\sigma}^2 = \frac{\pi}{3\kappa}$ vs. Concentration Parameter (κ) for a pairwise circular mean.

Plot of Estimated Variance vs. Concentration Parameter



To prove this theorem requires the three lemmas in Section 3.6.4. To motivate the lemmas, suppose that $\theta_1, \dots, \theta_n$ is an independent random sample from the von Mises distribution

$VM(\mu, \kappa)$, with density given by equation (1.1). For $\kappa \geq 2$, the $VM(\mu, \kappa)$ distribution is effectively a $N(\mu, \sigma^2)$ distribution with $\sigma^2 = \frac{1}{\kappa}$ (Fisher, 1987). See Mardia (1972, Section 3.9) for more details.

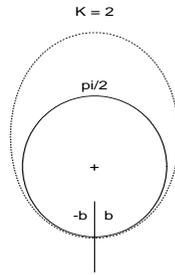
3.6.3 Probability of Wrapping

In order to avoid problems with the process of unwrapping not being reversible, consider data from the von Mises where the equivalent WN distribution has virtually no probability of containing values that will need to be wrapped from outside the original range. Suppose that $\theta_1, \dots, \theta_n$ is an independent random sample from the von Mises distribution $VM(\mu, \kappa)$, the probability of a single observation being wrapped from outside the original range is given by

$$\begin{aligned}
 & \text{Prob}[\text{One Observation Being Wrapped}] \\
 &= \text{Prob}[\theta < \mu - \pi] + \text{Prob}[\theta > \mu + \pi] \\
 &= 1 - [2 * (F(\pi; \mu, \sigma) - F(0; \mu, \sigma))]. \tag{3.3}
 \end{aligned}$$

Figure 3.12 below, is the density function of von Mises with mean $\frac{\pi}{2}$, and concentration parameter $\kappa = 2$. Notice that the area under the curve beyond the two endpoints $\frac{\pi}{2} - \pi$ and $\frac{\pi}{2} + \pi$, is close to zero. We shall limit our discussion in these sections to such distributions.

Figure 3.12: Split circle: $b = \frac{\pi}{2} + \pi$, $-b = \frac{\pi}{2} - \pi$

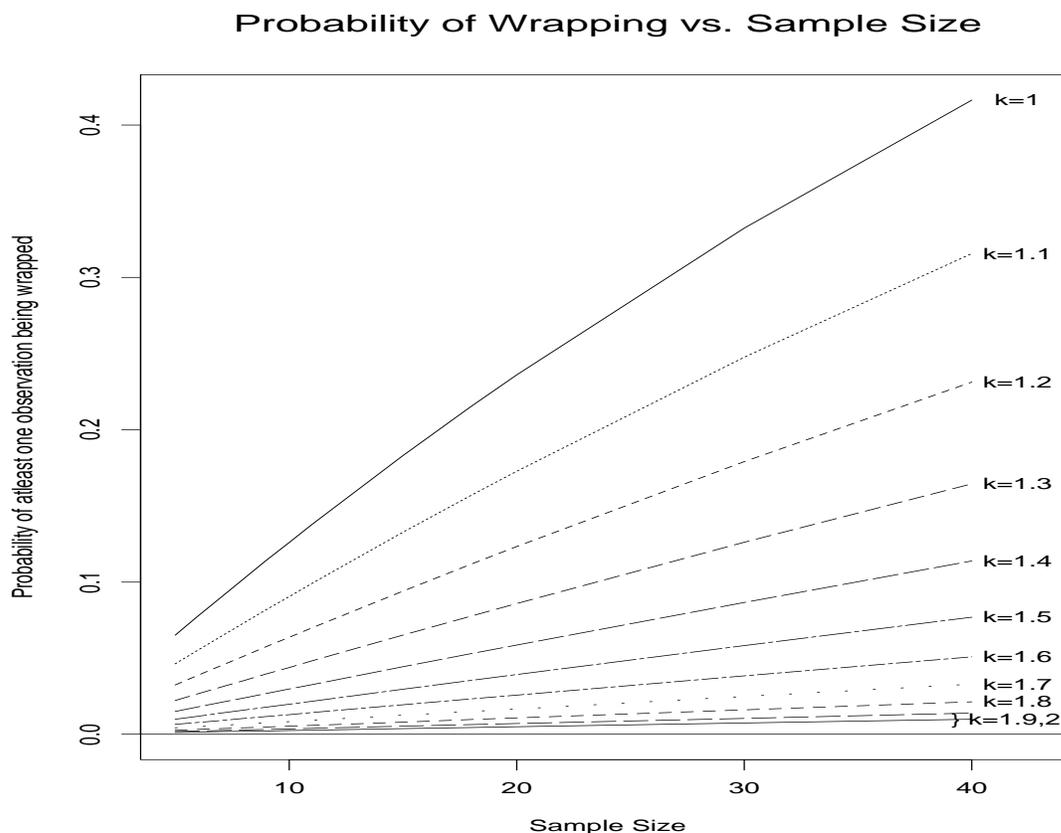


From Figure 3.12, we see that for symmetric unimodal distributions with high concentrations about the mode, the probability of wrapping is very close to zero. Note the probability of n independent observations being wrapped from outside the original range is

$$\begin{aligned} & \text{Prob}[\text{At least One Observation Being Wrapped}] \\ &= 1 - \text{Prob}[\text{None Wrapped}] \\ &= 1 - [2 * (F(\pi; \mu, \sigma) - F(0; \mu, \sigma))]^n, \text{ where } n \text{ is the sample size.} \end{aligned} \tag{3.4}$$

Figure 3.13 is obtained by evaluating this probability for different values of n and various dispersions. Note that probability of at least one observation being wrapped tends to zero as κ increases, with probability being close to zero for $\kappa \geq 2$. Thus for data from the von Mises distribution, with observations limited to a smaller portion of the circle, we claim that the probability of wrapping, as σ decreases or κ increases, will be sufficiently small allowing some theoretical work to be done on the new measure.

Figure 3.13: Plot of Probability of Wrapping vs. Sample size



3.6.4 Circular and Linear point estimates for $\kappa \geq 2$

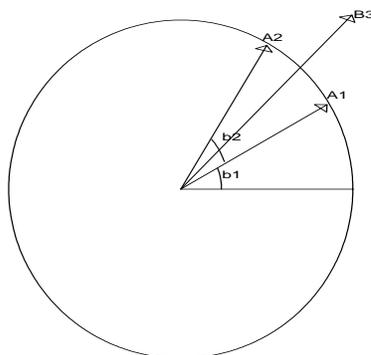
Lemma 3.6.1

Consider a sample of size 2. Let θ_1 and θ_2 be the ordered observations. Without loss of generality, we can describe θ_1 and θ_2 as θ and $\theta + \phi$ respectively, where $\phi < \pi$, see Figure 3.14. Then for an appropriately chosen range (where the endpoints do not lie in the arc defined by ϕ), $\bar{\theta}_{\text{linear}} = \bar{\theta}_{\text{circular}}$, where

$$\bar{\theta}_{\text{linear}} = \theta + \frac{1}{2}\phi \quad \text{and} \quad \bar{\theta}_{\text{circular}} = \tan^{-1} \left[\frac{\sin\theta + \sin(\theta + \phi)}{\cos\theta + \cos(\theta + \phi)} \right].$$

Figure 3.14: Adjacent angles

Two Adjacent Angles



where $b1 = \theta$, $b2 = \phi$, $A1 = (\cos(\theta), \sin(\theta))$ & $A2 = (\cos(\theta + \phi), \sin(\theta + \phi))$, in Cartesian coordinates.

Proof of Lemma 3.6.1

Recall the following trigonometric identities:

$$\cos(\theta + \frac{1}{2}\phi) = \cos(\theta)\cos(\frac{1}{2}\phi) - \sin(\theta)\sin(\frac{1}{2}\phi)$$

$$\sin(\theta + \frac{1}{2}\phi) = \cos(\theta)\sin(\frac{1}{2}\phi) + \sin(\theta)\cos(\frac{1}{2}\phi)$$

$$2\cos^2(\frac{1}{2}\phi) = (1 + \cos(\phi))$$

$$\sin(\phi) = 2\sin(\frac{1}{2}\phi)\cos(\frac{1}{2}\phi)$$

$$\begin{aligned} \text{Now } \sin(\theta + \phi) + \sin(\theta) &= (\sin(\theta)\cos(\phi) + \cos(\theta)\sin(\phi)) + \sin(\theta) \\ &= \sin(\theta)(1 + \cos(\phi)) + \cos(\theta)\sin(\phi) \\ &= 2\sin(\theta)\cos^2(\frac{1}{2}\phi) + 2\cos(\theta)\sin(\frac{1}{2}\phi)\cos(\frac{1}{2}\phi) \end{aligned}$$

$$\begin{aligned}
&= 2\cos\left(\frac{1}{2}\phi\right) \left[\sin(\theta)\cos\left(\frac{1}{2}\phi\right) + \cos(\theta)\sin\left(\frac{1}{2}\phi\right) \right] \\
&= 2\cos\left(\frac{1}{2}\phi\right) \left[\sin\left(\theta + \frac{1}{2}\phi\right) \right]
\end{aligned}$$

Next

$$\begin{aligned}
&\cos(\theta + \phi) + \cos(\theta) \\
&= (\cos(\theta)\cos(\phi) - \sin(\theta)\sin(\phi)) + \cos(\theta) \\
&= \cos(\theta)(1 + \cos(\phi)) - \sin(\theta)\sin(\phi) \\
&= 2\cos(\theta)\cos^2\left(\frac{1}{2}\phi\right) - 2\sin(\theta)\sin\left(\frac{1}{2}\phi\right)\cos\left(\frac{1}{2}\phi\right) \\
&= 2\cos\left(\frac{1}{2}\phi\right) \left[\cos(\theta)\cos\left(\frac{1}{2}\phi\right) - \sin(\theta)\sin\left(\frac{1}{2}\phi\right) \right] \\
&= 2\cos\left(\frac{1}{2}\phi\right) \left[\cos\left(\theta + \frac{1}{2}\phi\right) \right]
\end{aligned}$$

Thus

$$\begin{aligned}
&\frac{\sin(\theta + \phi) + \sin(\theta)}{\cos(\theta + \phi) + \cos(\theta)} \\
&= \frac{2\cos\left(\frac{1}{2}\phi\right) \left[\sin\left(\theta + \frac{1}{2}\phi\right) \right]}{2\cos\left(\frac{1}{2}\phi\right) \left[\cos\left(\theta + \frac{1}{2}\phi\right) \right]} \\
&= \frac{\left[\sin\left(\theta + \frac{1}{2}\phi\right) \right]}{\left[\cos\left(\theta + \frac{1}{2}\phi\right) \right]} \\
&= \tan\left(\theta + \frac{1}{2}\phi\right)
\end{aligned}$$

Therefore, $\bar{\theta}_{circular} = \tan^{-1} \left[\tan\left(\theta + \frac{1}{2}\phi\right) \right] = \left(\theta + \frac{1}{2}\phi\right) = \bar{\theta}_{linear}$.

Lemma 3.6.2

Suppose that $\theta_1, \dots, \theta_n$ is a set of concentrated observations. The circular Mardia median matches the linear median for an appropriately selected range of angles, with the endpoints of the range chosen to be at the antimedians.

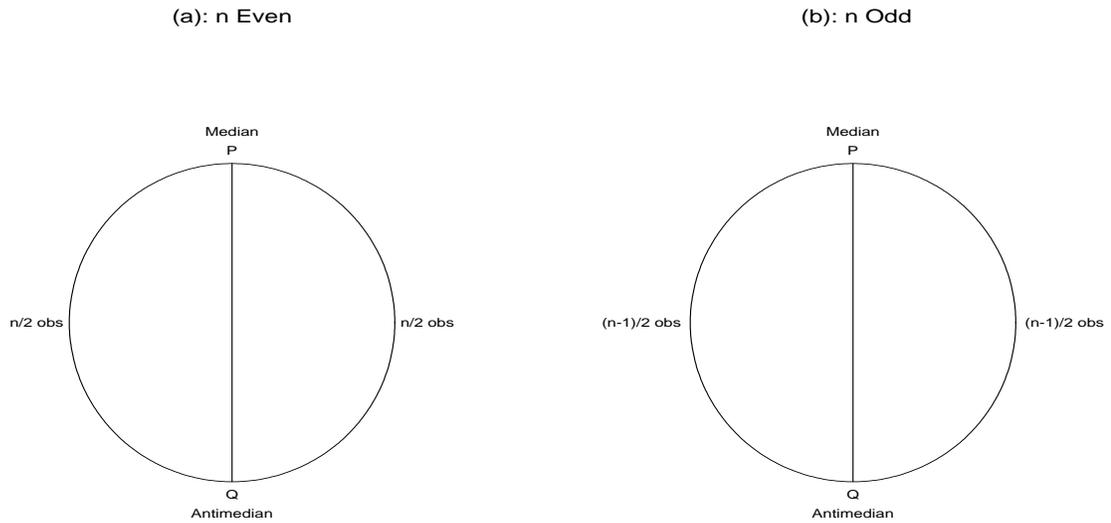
Proof of Lemma 3.6.2

If n is even, then by definition, the circular median (\mathbf{P}) is the circular mean (which is equivalent to the linear average by Lemma 3.6.1) of the two central observations such that half of the observations ($\frac{n}{2}$) lie on either side of the diameter (\mathbf{PQ}) and more observations are closer to \mathbf{P} than to \mathbf{Q} . See Figure 3.15a. If we select the range of our observations so that the antimedial (\mathbf{Q}) is the dividing point, we can assign ranks to these observations starting from either side of \mathbf{Q} . The linear median is the average of the observations whose corresponding ranks are $\frac{n}{2}$ and $\frac{n}{2} + 1$. This is equivalent to the circular median (\mathbf{P}), which will have $\frac{n}{2}$ observations in each half circle.

Similarly, if n is odd, the circular median (\mathbf{P}) is the central observation such that half of the observations ($\frac{(n-1)}{2}$) lie on either side of the diameter (\mathbf{PQ}) and more observations are closer to \mathbf{P} than to \mathbf{Q} . See Figure 3.15b. Again if we cut the circle at the antimedial (\mathbf{Q}), we can assign ranks to these observations starting from either side of \mathbf{Q} . The linear median is the observation whose corresponding rank is $\frac{(n+1)}{2}$. This is equivalent to the circular Mardia median (\mathbf{P}), which has equal numbers of observations in each half circle.

Note this result implies that the circular Mardia median is a linear median for the data from an appropriately chosen interval. It does not imply that any linear median will satisfy the definition of the circular median.

Figure 3.15: Calculation of median direction



Lemma 3.6.3

For a suitably chosen range of the observations, the circular median of the pairwise circular means ($\hat{\theta}_{HL}^c$) is equivalent to the linear Hodges-Lehmann estimate ($\hat{\theta}_{HL}^l$), where

$$\hat{\theta}_{HL}^c = \text{circular median} \left(\bar{\theta}_{1,1}^c, \bar{\theta}_{1,2}^c, \dots, \bar{\theta}_{n-1,n}^c, \bar{\theta}_{n,n}^c \right),$$

$$\text{with } \bar{\theta}_{ij}^c = \left[\tan^{-1} \left(\frac{\sin \theta_i + \sin \theta_j}{\cos \theta_i + \cos \theta_j} \right) \right]$$

and $\hat{\theta}_{HL}^l = \text{median} \left(\bar{\theta}_{1,1}^l, \bar{\theta}_{1,2}^l, \dots, \bar{\theta}_{n-1,n}^l, \bar{\theta}_{n,n}^l \right)$, with $\bar{\theta}_{ij}^l = \left(\frac{\theta_i + \theta_j}{2} \right)$.

Proof of Lemma 3.6.3

Obtained by combining the results of Lemma 3.6.1 and Lemma 3.6.2.

Proof of theorem 3.6.1

Suppose $\theta_1, \dots, \theta_n$ are i.i.d $VM(\mu, \kappa)$. $\hat{\theta}_{HL}^c$ is obtained as the circular median of the pairwise circular means. By Lemma 3.6.3, $\hat{\theta}_{HL}^c = \hat{\theta}_{HL}^l$. Using Randles & Wolfe (1979, p. 224-226), it can be shown that $(\hat{\theta}_{HL}^l - \theta_{HL}^l)$ is distributed approximately as $N\left(0, \frac{1}{12n[\int f^2(x)dx]^2}\right)$, where $12n[\int f^2(\theta, \mu, \kappa)d\theta]^2 = \frac{3n\kappa}{\pi}$. To show this, recall (Section 3.6.2) that for large κ , a von Mises distribution with mean μ and concentration parameter κ can be approximated by a $N(\mu, \frac{1}{\kappa})$ distribution, Mardia & Jupp(2000, p.41) The density function of $N(\mu, \frac{1}{\kappa})$ is given by

$$f(\theta, \mu, \kappa) = \frac{1}{\sqrt{2\pi\frac{1}{\kappa}}}\exp\left[-\frac{(\theta - \mu)^2}{\frac{2}{\kappa}}\right].$$

$$\begin{aligned} \text{Thus} \quad & \int f^2(\theta) d\theta \\ &= \int \frac{1}{2\pi\frac{1}{\kappa}}\exp\left[-\frac{(\theta - \mu)^2}{\frac{1}{\kappa}}\right]d\theta \\ &= \frac{1}{\frac{\sqrt{2\pi}}{\kappa}} \int \frac{\left(\frac{1}{2\kappa}\right)^{\frac{1}{2}}}{\sqrt{2\pi}\left(\frac{1}{2\kappa}\right)^{\frac{1}{2}}}\exp\left[-\frac{(\theta - \mu)^2}{\frac{1}{\kappa}}\right]d\theta \\ &= \frac{\left(\frac{1}{2\kappa}\right)^{\frac{1}{2}}}{\frac{\sqrt{2\pi}}{\kappa}} \\ &= \frac{1}{2} \left[\frac{\kappa}{\pi}\right]^{\frac{1}{2}} \end{aligned}$$

$$\text{This implies that} \quad 12n \left[\int f^2(\theta, \mu, \kappa)d\theta \right]^2 = 12n \frac{1}{4} \frac{\kappa}{\pi} = \frac{3n\kappa}{\pi}.$$

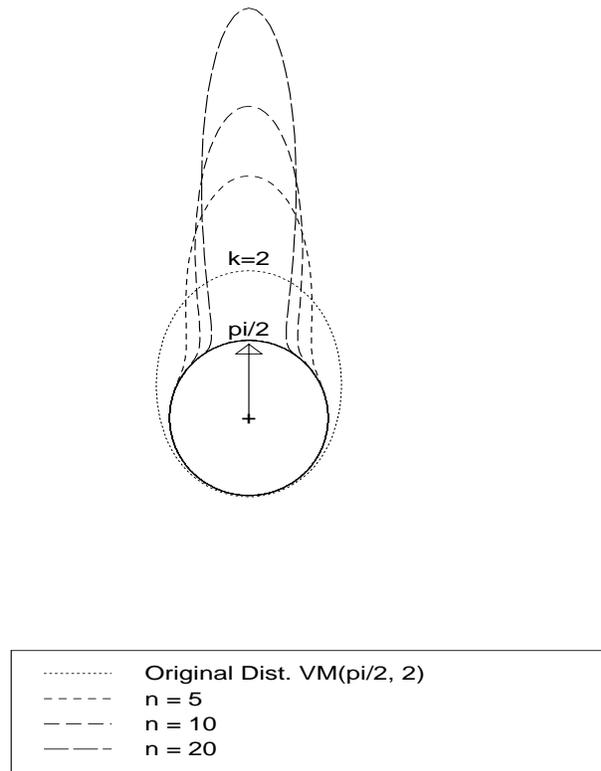
Next, recall if X is a random variable on the line, the corresponding random variable of the wrapped distribution is given by $X_w = X(\text{mod}2\pi)$. Thus a Wrapped Normal variate with mean μ and mean resultant length $\rho = \exp\left[\frac{-\sigma^2}{2}\right]$, is obtained by wrapping a Normal variate with mean μ and variance σ^2 . Consequently, $\theta^* = (\hat{\theta}_{HL}^l - \theta_{HL}^l) \text{ mod } 2\pi$, is approximately distributed as Wrapped Normal with mean 0 and mean resultant length is $\rho = \exp\left[\frac{-\pi}{6n\kappa}\right]$,

since $\sigma = \frac{3n\kappa}{\pi}$, which can be approximated by von Mises with mean 0 and concentration parameter

$$\mathbf{A}^{-1} \left[\exp \left(\frac{-\pi}{6n\kappa} \right) \right] \simeq \frac{3n\kappa}{\pi}. \quad \text{Thus} \quad \hat{\theta}_{HL}^c \sim VM \left(\theta_{HL}^c, A^{-1} \left[\exp \left[\frac{-\pi}{6n\kappa} \right] \right] \right).$$

$$\text{That is} \quad \hat{\theta}_{HL}^c \sim VM \left(\theta_{HL}^c, \frac{3n\kappa}{\pi} \right).$$

Figure 3.16: Distribution of Circular Hodges-Lehmann Estimator for $\kappa = 2$



Note from Figure 3.16, that as the sample size increases, the distribution of the circular Hodges-Lehmann estimator becomes more concentrated at the single mode.

3.7 Influence Functions for Measures of Preferred Direction

The study of the stability of parameter estimates for directional data dates back at least to Watson (1967). Wehrly & Shine (1981) and Watson (1986) evaluated the robustness of the circular mean via an influence function introduced by Hampel (1968, 1974) and concluded that the estimator is somewhat robust to fixed amounts of contamination and to local shifts, since its influence function is bounded. On the other hand, the influence curve for the circular median has a jump at the antimode. This implies that the circular median is sensitive to rounding or grouping of data (Wehrly & Shine, 1981). Ko and Guttorp (1988) introduced the notion of scale-standardized-bias (or SB) robustness to adjust for the concentration of the data on the sphere. They showed that the circular mean is not SB-robust with their scaling. He and Simpson (1992) studied various measures of stability of estimates on the sphere.

Consider a circular distribution \mathbf{F} which is unimodal and symmetric about the unknown direction μ_0 . The influence functions of the circular mean and circular median are given below in Theorems 3.7.1 and 3.7.2, respectively. These results are stated without proof in Wehrly and Shine (1981), hence we provide their proofs in *Appendix B*.

Theorem 3.7.1:

The influence function (IF) for the circular mean direction is given by

$$IF(\theta) = \frac{\sin(\theta - \mu_0)}{\rho}, \quad (3.5)$$

where ρ is the mean resultant length. For any given value of ρ , this influence function and

its derivative are bounded by $\pm\rho^{-1}$.

Theorem 3.7.2:

Without loss of generality for notational simplicity, assume that $\mu \in [0, \pi]$. The influence function for the circular median direction is given by

$$IF(\theta) = \frac{\frac{1}{2} \text{sgn}(\theta - \mu_0)}{[f(\mu_0) - f(\mu_0 + \pi)]}, \quad (\mu_0 - \pi < \theta < \mu_0 + \pi), \quad (3.6)$$

where $\text{sgn}(x) = 1, 0, \text{ or } -1$ as $x > 0, x = 0, \text{ or } x < 0$, respectively.

Remark: Both the influence functions of the circular mean and circular median are bounded, this implies that observations with large circular distance from the “center” cannot have an arbitrarily large effect on the estimates. However, unlike the influence function of the circular mean which is continuous, the influence function of the circular median has jumps at the antimode. The new measure, the circular Hodges-Lehmann estimate has a bounded influence function, see Theorem 3.7.3.

Theorem 3.7.3

For a sample from a von Mises distribution with a limited range of concentrated parameter values, $\kappa \geq 2$, the influence function of the circular Hodges-Lehmann estimator ($\hat{\theta}_{HL}^c$) is given by

$$IF(\theta) = \frac{F(\theta) - \frac{1}{2}}{\left(\frac{\kappa}{4\pi}\right)^{\frac{1}{2}}}, \quad (3.7)$$

where $\mathbf{F}(\cdot)$ is the distribution of θ_i , $i = 1, 2, \dots, n$. Note that this influence function is a centered and scaled cdf and is therefore bounded. Note that, it is also discontinuous at the antimode, like the influence function of the circular median (see Theorem 3.7.2).

Proof of Theorem 3.7.3

Assume that θ_i and θ_j are iid, with distribution function $F(\theta)$. Let $\Phi = \frac{(\theta_i + \theta_j)}{2}$, $i \leq j$.

By Lemma 1, Φ is equivalent to the pairwise circular mean of θ_i and θ_j . The functional of $\hat{\theta}_{HL}^c$ is the Pseudo-Median Locational functional $F = F^{*-1}(\frac{1}{2})$, where

$$F^*(\phi) = P(\Phi \leq \phi) = \int F(2\phi - \theta)h(\theta)d\theta,$$

Hettmansperger & McKean (1998, p.3,10-11). Since by Lemma 3.5.3, $\hat{\theta}_{HL}^c = \hat{\theta}_{HL}^l$, it can be shown that

$$IF(\theta) = \frac{F(\theta) - \frac{1}{2}}{\int f^2(\theta)d\theta},$$

(Hettmansperger & McKean,1998, p. 40-41). Thus, the influence function of the circular Hodges-Lehmann estimator is obtained as

$$IF(\theta) = \frac{F(\theta) - \frac{1}{2}}{\int f^2(\theta)d\theta} = \frac{F(\theta) - \frac{1}{2}}{\left(\frac{\kappa}{4\pi}\right)^{\frac{1}{2}}},$$

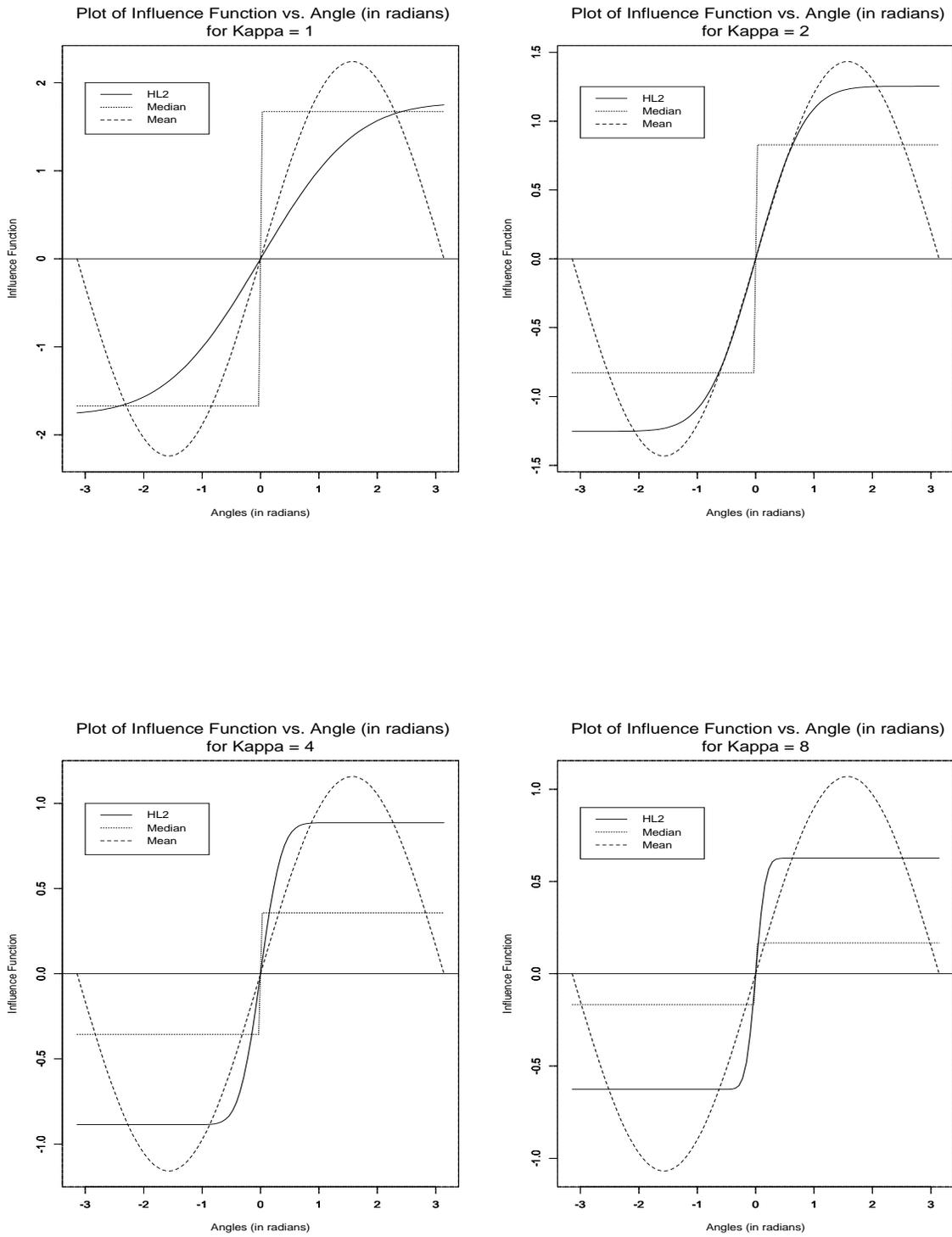
and for large κ ,

$$f(\alpha) = \sqrt{\left(\frac{1}{\frac{2\pi}{\kappa}}\right) \exp\left(\frac{-(\alpha - \mu)^2}{\frac{2}{\kappa}}\right)} \quad \text{so}$$

$$\begin{aligned} & \int f^2(\alpha) d\alpha \\ &= \int \frac{1}{\frac{2\pi}{\kappa}} \exp\left[\frac{-(\alpha - \mu)^2}{\frac{1}{\kappa}}\right] d\alpha \\ &= \frac{1}{2} \left[\frac{\kappa}{\pi}\right]^{\frac{1}{2}}. \end{aligned}$$

Figure 3.17 are plots of the influence functions of the circular mean, circular median and the circular Hodges-Lehmann estimators for preferred direction for various concentration parameters. Notice that all the estimators have curves which are bounded. Also, as the concentration parameter (κ) increases, the influence function of the circular median stays relatively unchanged followed by the circular Hodges-Lehmann estimator. This is similar to the linear case.

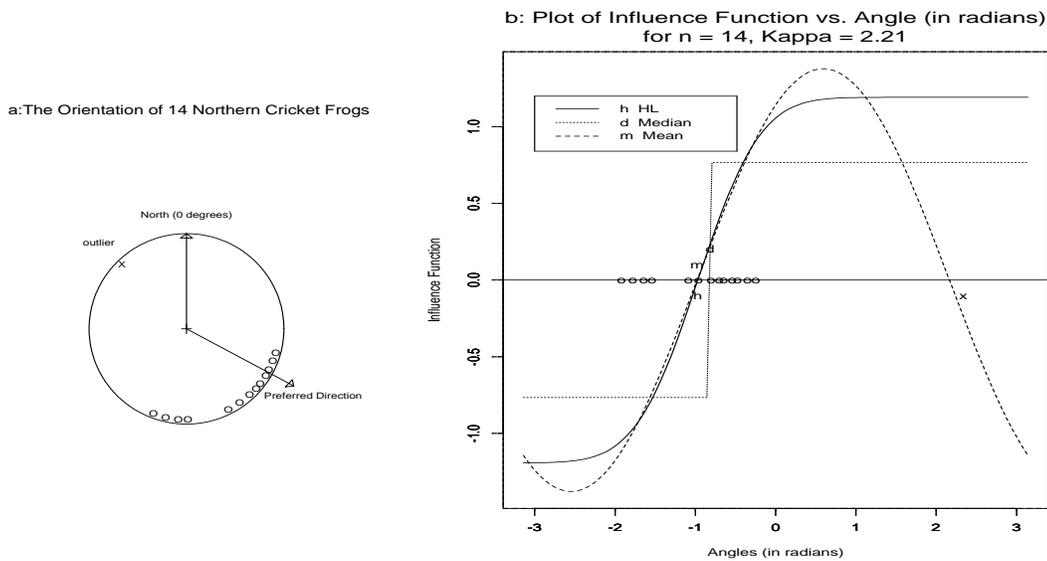
Figure 3.17: Influence Functions for $1 \leq \kappa \leq 8$



Also, as κ increases, the influence function for all the three measures decreases, however, overall the influence function for the mean is largest for angles closest to $\frac{\pi}{2}$ radians from the preferred direction. The maximum influence for the mean occurs at $\frac{\pi}{2}$ or $-\frac{\pi}{2}$ from the mode for all κ , while for both the median and HL, the maximum occurs uniformly for a range away from the preferred direction. Overall, HL seems like a compromise between the mean and the median.

3.7.1 Position of an Outlier on the Influence curve of the mean, median and HL2

Figure 3.18: Data with a Single Outlier



Consider the following example of Frog migration taken from Collett (1980), shown in Figure 3.18a. For this data the circular mean is $-0.977(124^0)$, \bar{R} is 0.725, thus $\hat{\kappa} = 2.21$ for the best fitting von Mises. (Table A.3, Fisher, 1993, p. 224). The circular median is $-0.816(133.25^0)$ and circular Hodges-Lehmann is $-0.969(124.5^0)$. Using $\hat{\kappa} = 2.21$, Figure

3.18b gives the influence curves of the mean, median and HL2. Note that the measure least influenced by observation x is the circular mean, since x is nearer to the antimode. However, the circular median is downweights by the observations nearer to the center of the data followed by HL2. The influence of an outlier on the sample circular median is bounded at either a constant positive or a constant negative value, regardless of how far the outlier is from the center of the data. On the other hand, the HL2 estimator is influenced less by observations near the center, and accounts for outlier. The influence curve for the circular mean is similar to that of the redescending Φ function, See Andrews et. al. (1972) for details regarding redescending functions.

3.8 Asymptotic Relative Efficiency

For a fixed underlying distribution, let $\tilde{\theta}$ and $\hat{\theta}$ be two consistent estimators of the preferred direction, then the relative efficiency (RE) $\tilde{\theta}$ relative to $\hat{\theta}$ is

$$\text{RE} (\tilde{\theta}; \hat{\theta}) = \frac{\text{Variance of } \hat{\theta}}{\text{Variance of } \tilde{\theta}} .$$

This is interpreted as, $\hat{\theta}$ needs RE times as many observations as $\tilde{\theta}$ for approximately the same precision. The asymptotic relative efficiency (ARE) is the limit of the RE as the sample size $n \rightarrow \infty$. More generally, the ARE of an estimator $\tilde{\theta}$ relative to another estimator $\hat{\theta}$ is defined by

$$\text{ARE} (\tilde{\theta}; \hat{\theta}) = \frac{\text{Asymptotic Variance of } \hat{\theta}}{\text{Asymptotic Variance of } \tilde{\theta}} .$$

This notion of efficiency is also known as Pitman efficiency, (Fisher, 1982).

We shall use the general definition of ARE to obtain the ARE for the measures of preferred direction: circular mean, circular median and circular Hodges-Lehmann estimator.

Theorem 3.8.1

The ARE of the circular median relative to the circular mean is $\frac{2}{\rho\pi} [1 - e^{-2\kappa}]^2$ for the von Mises distribution.

Proof of Theorem 3.8.1

Recall for circular mean that ,

$$IF(\theta) = \frac{1}{\rho} \sin(\theta - \mu_0). \quad \text{This implies} \quad IF^2(\theta) = \frac{1}{\rho^2} \sin^2(\theta - \mu_0).$$

Therefore the Asymptotic Variance for the circular mean

$$\begin{aligned} E [IF^2(\theta)] &= \frac{1}{\rho^2} \int \sin^2(\theta - \mu_0) dF(\theta - \mu_0) \\ &= \frac{1}{\rho^2} \frac{A(\kappa)}{\kappa} = \frac{1}{\rho\kappa}, \end{aligned}$$

since $E [\sin(\theta - \mu_0)] = 0$ & $\text{Var} [\sin(\theta - \mu_0)] \simeq \frac{A(\kappa)}{\kappa}$, (Mardia & Jupp, 2000, p.84, 87) and $\rho = A(\kappa)$.

Next, for the circular median

$$\begin{aligned} IF(\theta) &= \frac{\frac{1}{2} \text{sgn}(\theta - \mu_0)}{[f(\mu_0) - f(\mu_0 + \pi)]} \\ &= \begin{cases} \frac{-\frac{1}{2}}{[f(\mu_0) - f(\mu_0 + \pi)]} & \text{if } (\theta - \mu_0) < 0 \\ \frac{\frac{1}{2}}{[f(\mu_0) - f(\mu_0 + \pi)]} & \text{if } (\theta - \mu_0) > 0 \end{cases} . \end{aligned}$$

Therefore, the Asymptotic variance of the circular median is

$$\begin{aligned} E [IF^2(\theta)] &= \frac{1}{2} \left[\frac{-1}{2[f(\mu_0) - f(\mu_0 + \pi)]} \right]^2 + \frac{1}{2} \left[\frac{1}{2[f(\mu_0) - f(\mu_0 + \pi)]} \right]^2 \\ &= \frac{1}{4[f(\mu_0) - f(\mu_0 + \pi)]^2}. \end{aligned}$$

But $f(\mu_0) = \frac{e^\kappa}{2\pi I_0(\kappa)}$, and $f(\mu_0 + \pi) = \frac{e^{-\kappa}}{2\pi I_0(\kappa)}$, from Jammalamadaka & SenGupta (2001, p.36). Therefore $f(\mu_0) - f(\mu_0 + \pi) = \frac{1}{2\pi I_0(\kappa)} (e^\kappa - e^{-\kappa})$. This implies

$$4[f(\mu_0) - f(\mu_0 + \pi)]^2$$

$$\begin{aligned}
&= \frac{4}{4\pi^2 I_0^2(\kappa)} (e^\kappa - e^{-\kappa})^2 \\
&= \frac{1}{\pi^2 \left[\frac{e^\kappa}{\sqrt{(2\pi\kappa)}} \right]^2} (e^\kappa - e^{-\kappa})^2, \quad \text{since } I_0^2(\kappa) = \left[\frac{e^\kappa}{\sqrt{(2\pi\kappa)}} \right]^2 \\
&= \frac{1}{\pi^2 \left[\frac{e^{2\kappa}}{2\pi\kappa} \right]} (e^\kappa - e^{-\kappa})^2 \\
&= \frac{2\kappa (e^\kappa - e^{-\kappa})^2}{\pi e^{2\kappa}} \\
&= \frac{2\kappa}{\pi} (1 - e^{-2\kappa})^2.
\end{aligned}$$

Thus the Asymptotic variance of circular median is $\frac{\pi}{2\kappa} (1 - e^{-2\kappa})^{-2}$. Therefore

$$\begin{aligned}
&\text{ARE}(\tilde{\theta}_{med}^c; \bar{\theta}_{mean}^c) \\
&= \frac{\text{var}(\bar{\theta}_{mean}^c)}{\text{var}(\tilde{\theta}_{med}^c)} \\
&= \frac{1}{\rho\kappa} \frac{2\kappa}{\pi} (1 - e^{-2\kappa})^2 \\
&= \frac{2}{\rho\pi} (1 - e^{-2\kappa})^2.
\end{aligned}$$

Note that as $\kappa \rightarrow \infty$, $\text{ARE}(\tilde{\theta}_{med}^c; \bar{\theta}_{mean}^c) \rightarrow \frac{2}{\pi}$, since $\rho \rightarrow 1$ and $e^{-2\kappa} \rightarrow 0$.

This is equivalent to the linear case, where the Asymptotic Relative efficiency of the median relative to the mean at the normal distribution is 0.637, Hollander & Wolfe (1999, p.105).

Theorem 3.8.2

The ARE of the circular Hodges-Lehmann estimator relative to the circular mean is $\frac{3}{\rho\pi}$ for the von Mises distribution.

Proof of Theorem 3.8.2

From Theorem 3.8.1, the Asymptotic variance for the circular mean is $\frac{1}{\kappa\rho}$.

To obtain the Asymptotic variance for circular Hodges-Lehmann estimator we use

$$IF(\theta) = \frac{F(\theta) - \frac{1}{2}}{\left(\frac{\kappa}{4\pi}\right)^{\frac{1}{2}}},$$

to obtain

$$IF^2(\theta) = \frac{\left[F(\theta) - \frac{1}{2}\right]^2}{\left(\frac{\kappa}{4\pi}\right)}.$$

Hence

Asymptotic Variance of HL2

$$\begin{aligned} &= E\left[IF^2(\theta)\right] \\ &= \frac{4\pi}{\kappa} \int \left(F(\theta) - \frac{1}{2}\right)^2 dF(\theta) \\ &= \frac{4\pi}{\kappa} \frac{1}{12} \\ &= \frac{\pi}{3\kappa}, \end{aligned}$$

using $\int \left(F(\theta) - \frac{1}{2}\right)^2 dF(\theta) = \frac{1}{12}$, (Hettmansperger, 1984, p. 46). Therefore

$$\begin{aligned} &\text{ARE}\left(\tilde{\theta}_{HL}^c; \bar{\theta}_{mean}^c\right) \\ &= \frac{\text{var}(\bar{\theta}_{mean}^c)}{\text{var}(\tilde{\theta}_{HL}^c)} \\ &= \frac{1}{\rho\kappa} \frac{3\kappa}{\pi} \\ &= \frac{3}{\rho\pi}. \end{aligned}$$

Note that as $\kappa \rightarrow \infty$, $\text{ARE}\left(\tilde{\theta}_{HL}^c; \bar{\theta}_{mean}^c\right) \rightarrow \frac{3}{\pi}$, since $\rho \rightarrow 1$. This is equivalent to the linear case, where the Asymptotic Relative efficiency of the Hodges-Lehmann estimator relative to the mean at the normal distribution is 0.955, Hollander & Wolfe (1999, p.104).

Theorem 3.8.3

The ARE of the circular median relative to the circular Hodges-Lehmann estimator is $\frac{2}{3} (1 - e^{-2\kappa})^2$ for the von Mises distribution.

Proof of Theorem 3.8.3

From Theorems 3.8.1 and 3.8.2, the asymptotic variance of the circular median and the circular Hodges-Lehmann estimator are $\frac{\pi}{2\kappa} (1 - e^{-2\kappa})^{-2}$ and $\frac{\pi}{3\kappa}$ respectively. Therefore

$$\begin{aligned} \text{ARE}(\tilde{\theta}_{med}^c; \tilde{\theta}_{HL}^c) &= \frac{\text{var}(\tilde{\theta}_{HL}^c)}{\text{var}(\tilde{\theta}_{med}^c)} \\ &= \frac{2\kappa}{\pi} (1 - e^{-2\kappa})^2 \left[\frac{\pi}{3\kappa} \right] \\ &= \frac{2}{3} (1 - e^{-2\kappa})^2. \end{aligned}$$

Note that as $\kappa \rightarrow \infty$, $\text{ARE}(\tilde{\theta}_{HL}^c; \tilde{\theta}_{med}^c) \rightarrow \frac{2}{3}$, since $e^{-2\kappa} \rightarrow 0$. This is equivalent to the linear case, where the ARE for the median relative to HL at the normal distribution is 0.667, Manoukian (1986, p. 194-195).

Chapter 4

Simulation Study

4.1 Overview

The simulation study was designed to compare the three measures of preferred direction for circular data; circular mean, circular median, and the circular Hodges-Lehmann estimator (using $\frac{n(n+1)}{2}$ pairwise means), referred to from now on as the mean, median and HL respectively.

10000 samples of sizes 5, 6, 7, 8, 9, 10, 11, 15, 18, 20 from the distributions with 4 dispersions around the circle were obtained. For each sample, the three measures were computed. The results were summarized using the following measures: Circular mean of the 10000 samples given by equation (1.7), 95% Empirical Confidence Interval (the central 95% of the 10000 values), Circular Variance (CV) given by equation (1.8), which is sample mean resultant length, Circular Mean Deviation (CMD) given by equation (1.9), and Circular Median Absolute Deviation (CMAD) given by equation (1.10). Three types of data sets were considered and are illustrated by Figures 4.1 and 4.2.

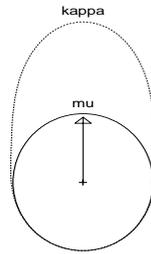
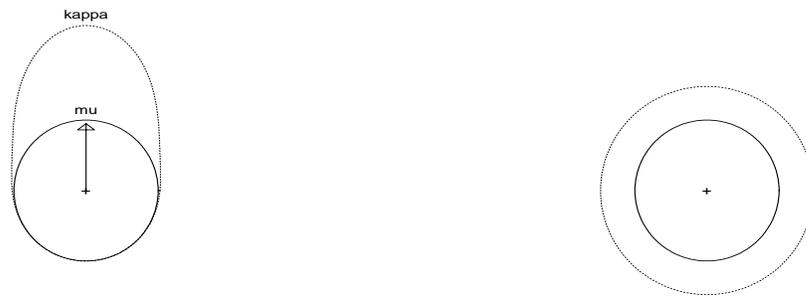
Figure 4.1: Uncontaminated von Mises, $VM(\mu, \kappa)$ 

Figure 4.1 is a plot of a von Mises density with mean μ and concentration parameter κ . Data generated from such a distribution alone is called uncontaminated data in our study. The value of μ was set equal to zero without loss of generality, while the values of κ in the study were 1, 2, 4 & 8.

Two types of contaminated data, see Figure 4.2, were considered in the study. These result from contamination in spread and contamination in location. For contamination in spread, data were generated with probability $(1 - \epsilon)$ from a von Mises distribution and with probability ϵ from Uniform $(-\pi, \pi)$ distribution. In case of contamination in location, data were generated with probability $(1 - \epsilon)$ from a von Mises distribution and with probability ϵ from a von Mises distribution with mean μ^* , where μ^* takes values $\frac{\pi}{8}$, $\frac{\pi}{4}$ & $\frac{\pi}{2}$, and constant concentration. The fractions of contamination, ϵ , considered were 0, 0.1, 0.2 & 0.3.

Figure 4.2: Contaminated Data

(a): Contamination in spread: $(1 - \epsilon)VM$ & ϵ Uniform



(b): Contamination in location: $(1 - \epsilon)VM(\mu_1, \kappa)$ & $\epsilon VM(\mu_2, \kappa)$



4.2 Uncontaminated Data

We now present the results for the uncontaminated data for which data was generated from a von Mises distribution with mean zero and various concentration parameters.

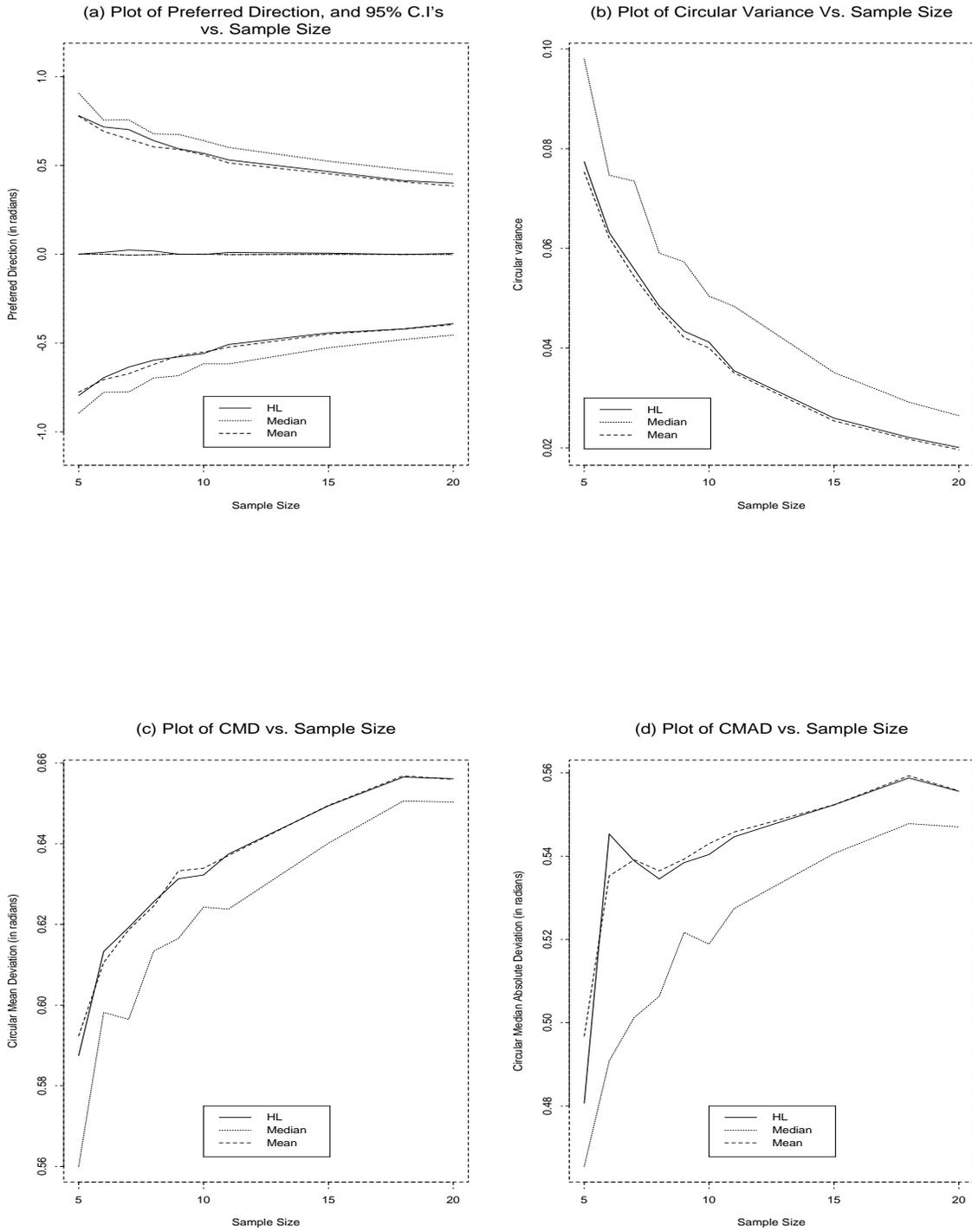
4.2.1 Effect of Sample Size on the Mean, the Median and HL

The effect of sample size on the three measures of preferred direction: the mean, the median and HL for data from an uncontaminated von Mises distribution is illustrated in Figure 4.3 for $\kappa = 2$. All measures appear unbiased. The mean has the narrowest confidence band, followed closely by HL. The confidence bands for the median are widest over the whole range of sample sizes considered as seen Figure 4.3(a). The confidence band for HL is sandwiched between that of the mean and the median. In general, the confidence bands become narrower as sample size increases for all the three measures.

In terms of circular variance, see Figure 4.3(b), the mean and HL compete favorably, on the other hand, the median has the largest circular variance over the whole range of sample sizes considered. Note that the circular variance of HL is sandwiched between the circular variances of the mean and the median, but closer to that of the mean than to the median. Circular variance decreases as the sample size increases for all measures.

Figure 4.3(c) shows that the median has the smallest circular mean deviation (CMD), this is not surprising since CMD is designed to be smallest for the median. Note, however, the CMD for the mean and HL are almost identical over the whole range of sample sizes considered. Similarly, from Figure 4.3(d), we observe that the circular median absolute deviation (CMAD) is smallest for the median and largest for the mean and HL over the whole range of sample sizes considered. Similar results to Figure 4.3 were obtained for other concentration parameters, See *Appendix D*, for the results of $\kappa = 1, 4$ & 8 .

Figure 4.3: Mean, Median and HL for VM(0,2)



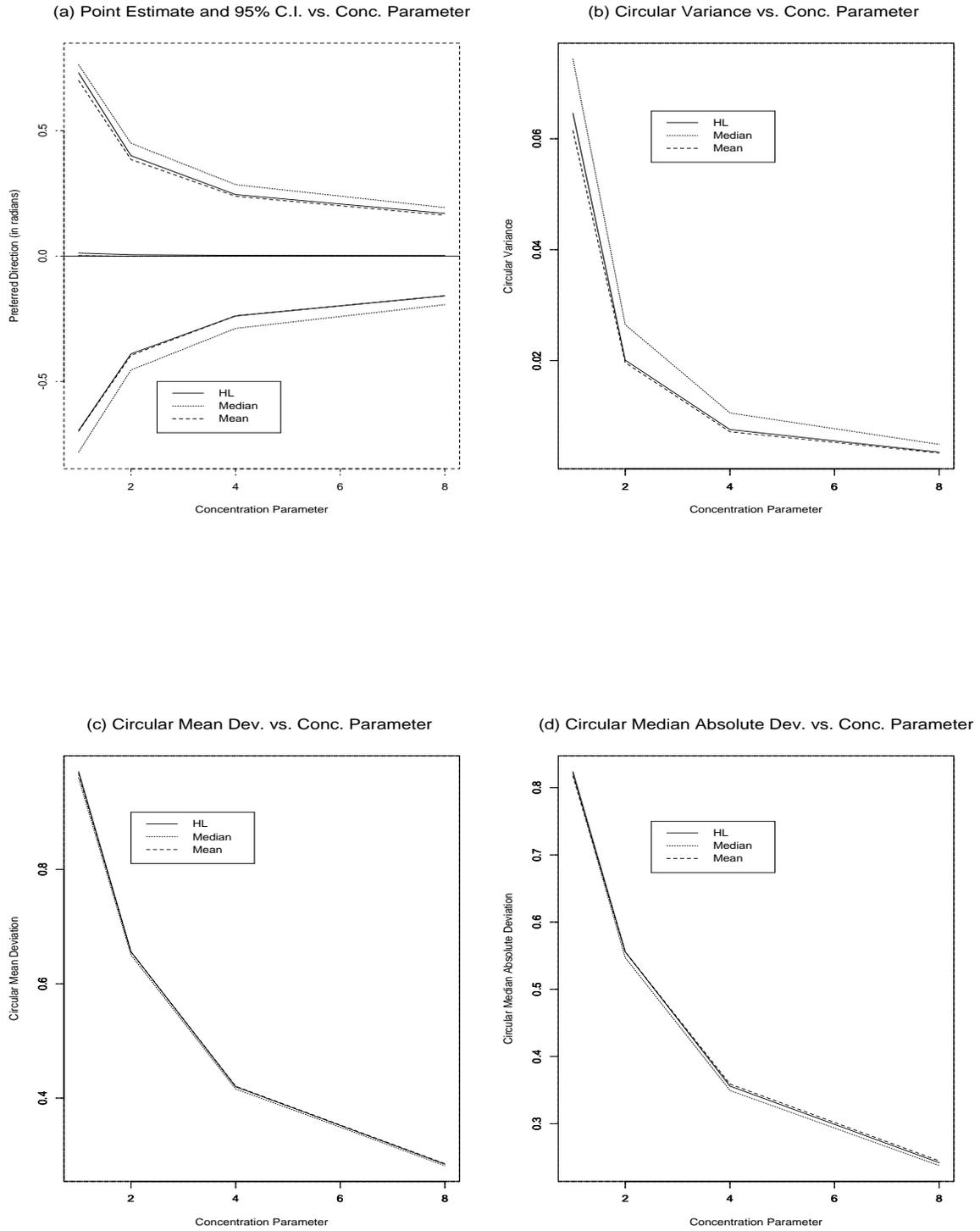
4.2.2 Effect of Concentration Parameter (κ) on the Mean, the Median and HL

Effect of concentration parameter (κ) on the three measures of preferred direction: the mean, the median and HL for data from uncontaminated von Mises is illustrated in Figure 4.4 for $n = 20$. All measures appear unbiased. The mean has the narrowest confidence band, followed closely by HL. The confidence bands for the median are widest over the whole range of κ s considered, see Figure 4.4(a). The confidence band for HL is sandwiched between that of the mean and the median. In general, the confidence bands become narrower as κ increases for all the three measures.

In terms of circular variance, see Figure 4.4(b), the mean and HL compete favorably, on the other hand, the median has the largest circular variance over the whole range of concentration parameters considered. Note that the circular variance of HL is sandwiched between the circular variances of the mean and the median, but closer to that of the mean than to the median. Circular variance decreases as the concentration parameter increases for all measures.

Figure 4.4(c) the three measures have circular mean deviation (CMD) that are almost identical over the whole range of concentration parameters considered. In the case of circular median absolute deviation (CMAD), the median has the smallest CMAD followed closely by the mean and HL (which are almost identical over the whole range of concentration parameters considered), see Figure 4.4 (d). Similar results to Figure 4.4 were obtained for other sample sizes, See *Appendix D*, for the results of $n = 10$.

Figure 4.4: Effect of κ on Mean, Median and HL for VM $(0, \kappa)$, $n = 20$



4.3 Comparing Theoretical and Simulated Results for HL

A sample from a von Mises distribution with mean μ and concentration parameter κ has a mean resultant length $\rho = A(\kappa) = \left[\frac{I_1(\kappa)}{I_0(\kappa)} \right]$, Fisher (1993, p. 225). In Section 3.6, we derived the distribution of HL, see Theorem 3.6.1.

$$\hat{\theta}_{HL}^c \sim \text{VM} \left(\theta_{HL}^c, \frac{3n\kappa}{\pi} \right)$$

The concentration parameter is $\frac{3n\kappa}{\pi}$ and the mean resultant length is $\rho = \left[\exp \left[\frac{-\pi}{6n\kappa} \right] \right]$, where n is the sample size.

Data was generated from von Mises distribution with mean zero and various concentration parameters, for various sample sizes. The concentration parameters considered were $\kappa = 2, 4$ & 8 while the sample sizes were $n = 5, 10$ & 20 . For each data set, we estimated the mean resultant length from which we obtain an estimate of the concentration parameter using Appendix A3 in Fisher (1993, p. 224).

Table 4.1 are the estimates (theoretical & observed from simulations) of the mean resultant length and concentration parameters for the Hodges-Lehmann estimator. For example, the entries in the first two rows of Table 4.1, are obtained as follows: theory mean resultant length is equal to $\exp \left[\frac{-\pi}{6(5)(2)} \right] = 0.949$ and theory concentration parameter is equal to $A^{-1} \left(\frac{3(5)(2)}{\pi} \right) = 10.3$. On the other hand observed mean resultant length is 0.923 , then using Appendix A3 in Fisher (1993, pp. 224), we obtain observed concentration parameter as 6.95 . For all the concentration parameters considered both the mean resultant length and the concentration parameter for theory and observed match closely as the sample size increases. In general, the approximations are better for large κ and large sample sizes. Overall, the simulated results appear consistent with the theoretical results.

Table 4.1: Theoretical and Simulated Mean Resultant Lengths and Concentration Parameters for distribution of HL for various concentration parameters.

Sample Size	Kappa (κ)		Mean Resultant Length	Concentration Parameter
5	2	Theory*	0.949	10.3
		Observed**	0.923	6.95
	4	Theory	0.979	20.3
		Observed	0.969	16.9
	8	Theory	0.987	33.6
		Observed	0.986	33.6
10	2	Theory	0.974	20.3
		Observed	0.959	12.8
	4	Theory	0.987	33.6
		Observed	0.985	33.6
	8	Theory	0.993	71.7
		Observed	0.993	71.7
20	2	Theory	0.987	33.6
		Observed	0.980	25.3
	4	Theory	0.993	71.7
		Observed	0.992	62.7
	8	Theory	0.997	167
		Observed	0.997	167

* Theory from Theorem 3.6.1

** Observed based on 10000 simulations

4.4 Contamination in Spread

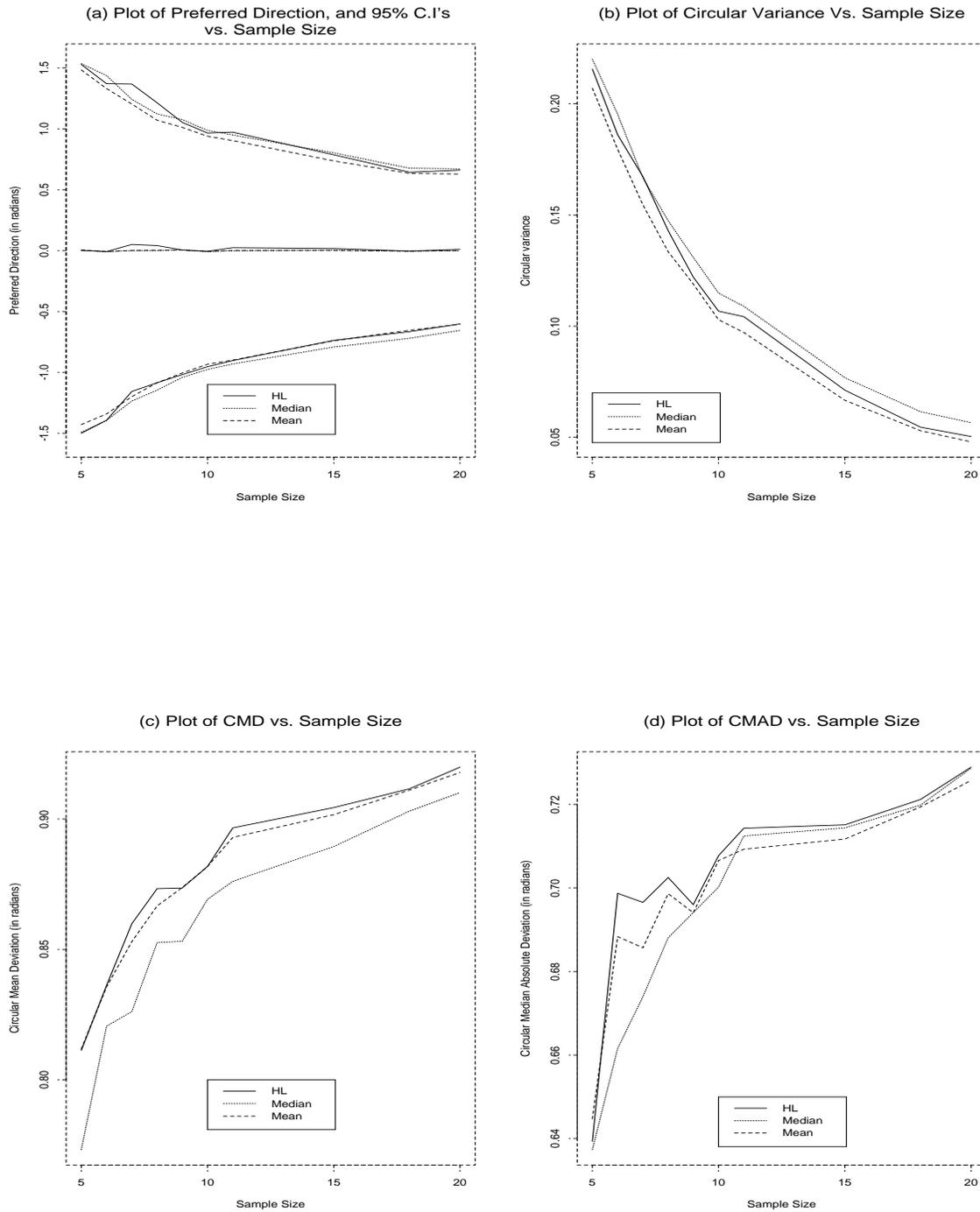
In the second type of data considered, data was generated with probability ϵ from the Circular Uniform and with probability $(1 - \epsilon)$ from von Mises distribution with mean zero and various concentration parameters. The proportion of contamination (ϵ) considered were 0, 0.1, 0.2 & 0.3.

4.4.1 Effect of Sample Size on the Mean, the Median and HL

The effect of sample size on the three measures of preferred direction: the mean, the median and HL for data from contaminated von Mises is illustrated in Figure 4.5 for $\kappa = 2$, & $\epsilon = 0.3$. All measures appear unbiased. The mean has the narrowest confidence band, followed closely by HL. The confidence bands for the median are widest over the whole range of sample sizes considered, see Figure 4.5(a). The confidence band for HL is generally sandwiched between that of the mean and the median. The confidence bands become narrower as sample size increases for all the three measures, but are in general wider than in the no contamination case in Figure 4.3(a).

In terms of circular variance, in Figure 4.5(b), the mean is the smallest followed by HL with the median having the largest circular variance over the whole range of sample sizes considered. Note that the circular variance of HL is sandwiched between the circular variances of the mean and the median. Circular variance decreases as the sample size increases for all measures, but CV is much higher than in no the contamination case in Figure 4.3(b). Figure 4.5(c) shows that the median has the smallest circular mean deviation (CMD), followed by the mean. HL has the largest CMD over most of the range of sample sizes considered. Similarly, from Figure 4.5(d), we observe that the circular median absolute deviation (CMAD) is smallest for the median and largest for the mean and HL. Similar results to Figure 4.5 were obtained for other concentration parameters, See *Appendix D*, for the results of $\kappa = 1, 4$ & 8 .

Figure 4.5: Mean, Median and HL for 70%VM(0, 2) & 30% Uniform, $\epsilon = 0.3$



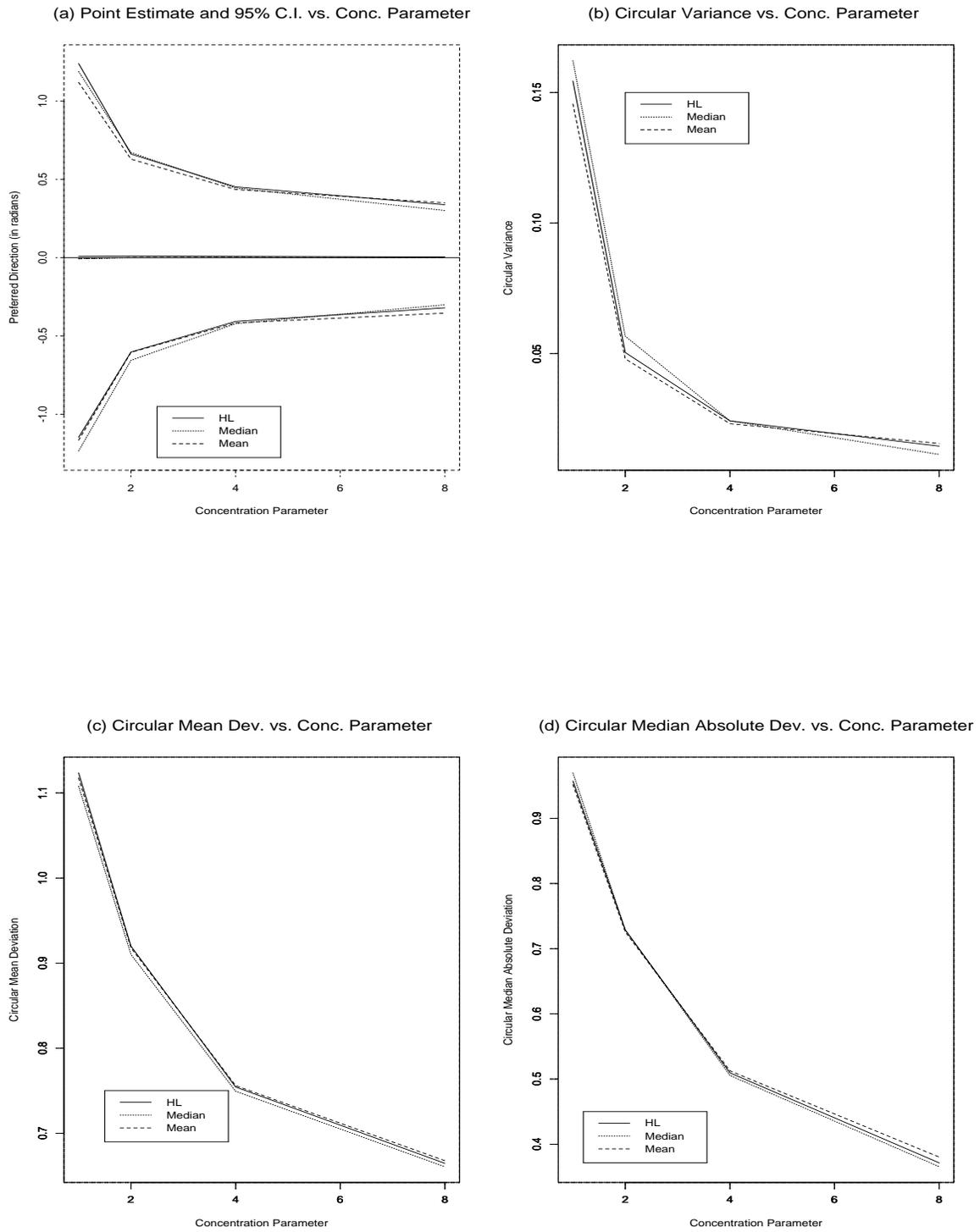
4.4.2 Effect of Concentration Parameter (κ) on the Mean, the Median and HL

The effect of the concentration parameter (κ) on the three measures of preferred direction: the mean, the median and HL for data from contaminated von Mises is illustrated in Figure 4.6 for $\epsilon = 0.3$ & $n = 20$. All measures appear unbiased. The mean has the narrowest confidence band, followed closely by HL. The confidence bands for the median are widest over the whole range of κ s considered, see Figure 4.6(a). The confidence band for HL is sandwiched between that of the mean and the median. In general, the confidence bands become narrower as κ increases for all the three measures.

In terms of circular variance, the mean and HL compete favorably (Figure 4.6(b)), while the median has the largest circular variance over most of the range of concentration parameters considered. Note that the circular variance of HL is sandwiched between the circular variances of the mean and the median. Circular variance decreases as the concentration parameter increases for all measures.

Figure 4.6(c) shows the three measures have circular mean deviation (CMD) that are almost identical over most of the range of concentration parameters considered. In the case of circular median absolute deviation (CMAD), the median has the smallest CMAD followed closely by the mean and HL (which are almost identical over the whole range of concentration parameters considered), see Figure 4.6(d). Note that both the median and HL have smaller CMD and CMAD for $\kappa \geq 6$ compared to the mean. Similar results to Figure 4.6 were obtained for other sample sizes, See *Appendix D*, for the results of $n = 10$.

Figure 4.6: Mean, Median and HL for 70%VM $(0, \kappa)$ & 30%Uniform, $\epsilon = 0.3$ & $n = 20$



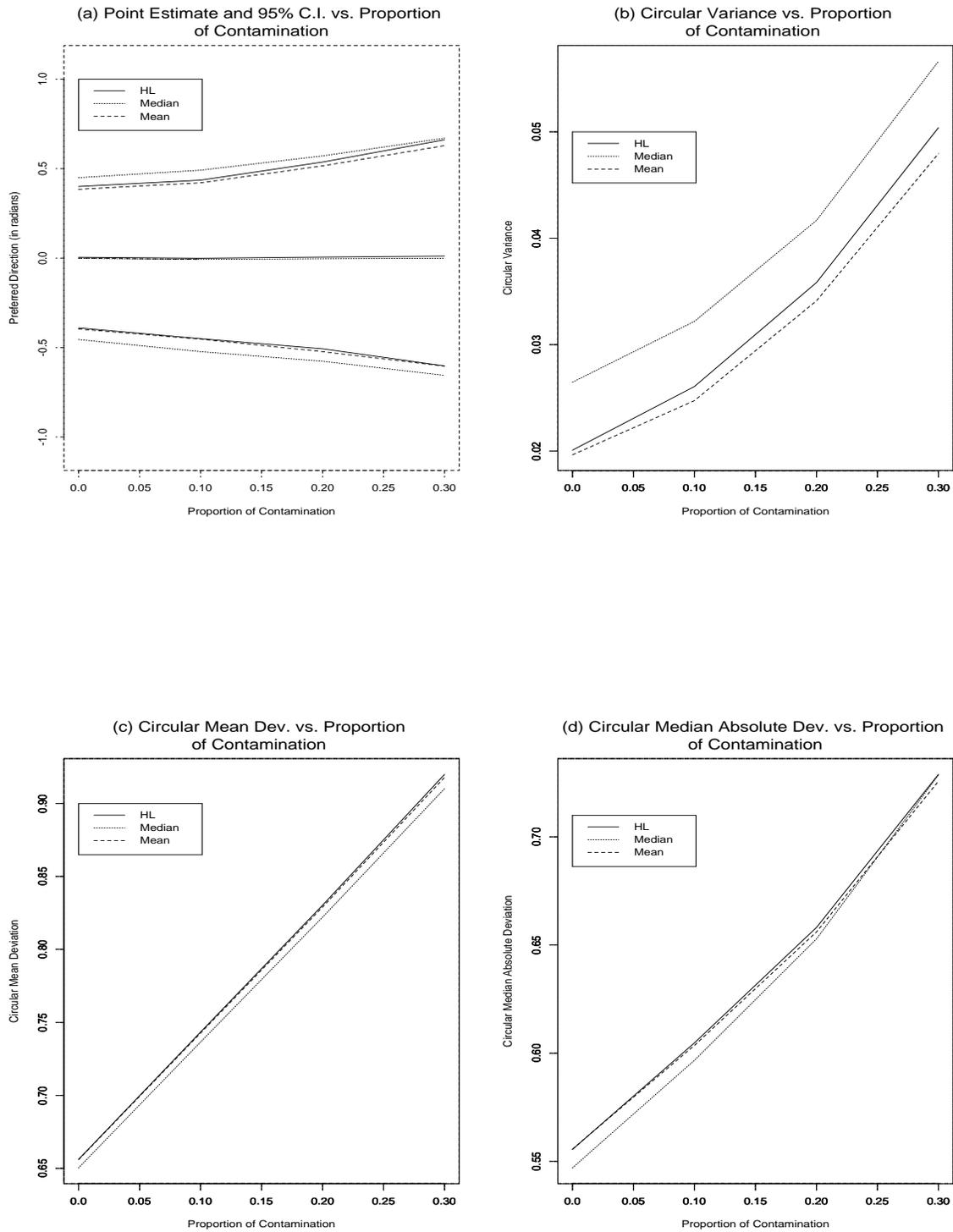
4.4.3 Effect of increasing contamination in spread on the Mean, the Median and HL

The effect of increasing contamination in spread on the three measures of preferred direction: the mean, the median and HL for data from contaminated von Mises is illustrated in Figure 4.7 for $\kappa = 2$, & $n = 20$. All measures appear unbiased. The mean has the narrowest confidence band, followed closely by HL. The confidence bands for the median are widest over the whole range of sample sizes considered, see Figure 4.7(a). The confidence band for HL is sandwiched between that of the mean and the median. In general, the confidence bands become wider as the proportion of contamination increases for all the three measures.

In terms of circular variance, in Figure 4.7(b), the mean has the smallest followed by HL with the median having the largest circular variance over the whole range of proportion of contamination considered. Note that the circular variance of HL is sandwiched between the circular variances of the mean and the median, but closer to that of the mean than to the median. Circular variance increases as the amount of contamination increases for all measures.

Figure 4.7(c) shows that the median has the smallest circular mean deviation (CMD), followed by the mean and HL (which are almost identical) the whole range of amount of contamination considered. Similarly, from Figure 4.7(d), we observe that the circular median absolute deviation (CMAD) is smallest for the median and largest for the mean and HL over most of the range of amount of contamination considered. However, the CMAD for the three measures is almost identical when the proportion of contamination greater or equal 0.25. Note that in general, both CMD and CMAD increase as the proportion of contamination increases for all the three measures. Similar results to Figure 4.7 were obtained for other concentration parameters, See *Appendix D*, for the results of $n = 10$.

Figure 4.7: Effect of increasing the spread on Mean, Median and HL, for $\kappa = 2$, & $n = 20$



4.5 Contamination in Location

The final type of data was generated with probability $(1 - \epsilon)$ from von Mises distribution with mean zero and with probability ϵ from von Mises distribution with mean μ^* , where μ^* take values $\frac{\pi}{8}$, $\frac{\pi}{4}$ & $\frac{\pi}{2}$, and various concentration parameters respectively. The proportions of contamination considered were 0, 0.1, 0.2 & 0.3.

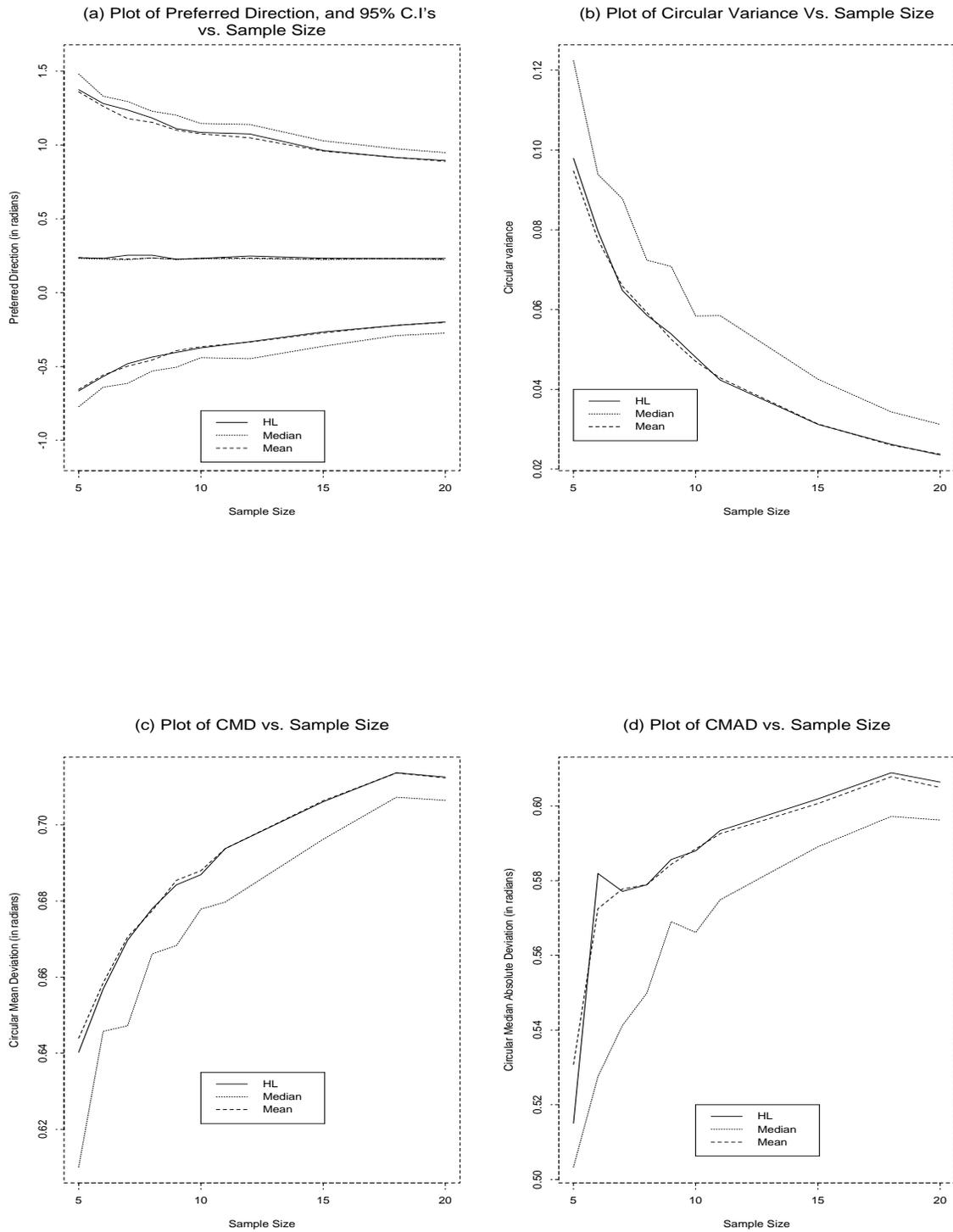
4.5.1 Effect of Sample Size on the Mean, the Median and HL

The effect of sample size on the three measures of preferred direction: the mean, the median and HL for data from contaminated von Mises is illustrated in Figure 4.8 for $\kappa = 2$. Notice that now since the data does not have mean 0, the estimates of preferred direction have shifted to reflect this. The confidence bands becomes narrower as sample size increases for all the three measures, with the mean having the narrowest followed closely by HL. The median has the widest confidence band over the whole range of sample sizes considered, see Figure 4.8(a).

In terms of circular variance, in Figure 4.8(b), the mean and HL have the smallest while, the median has the largest circular variance over the whole range of sample sizes considered. The mean and HL are virtually identical now for all samples. Circular variance decreases as the sample size increases for all measures.

Figure 4.8(c) shows that the median has the smallest circular mean deviation (CMD), followed by the mean and HL (which are almost identical) over most of the range of sample sizes considered. Similarly, from Figure 4.8(d), we observe that the circular median absolute deviation (CMAD) is smallest for the median and largest for the mean and HL over the whole range of sample sizes considered. Similar results to Figure 4.8 were obtained for other concentration parameters, See *Appendix D* for the results of $\kappa = 1, 4$ & 8 .

Figure 4.8: Mean, Median and HL for VM (0, 2) & VM($\frac{\pi}{4}$, 2), $\epsilon = 0.3$



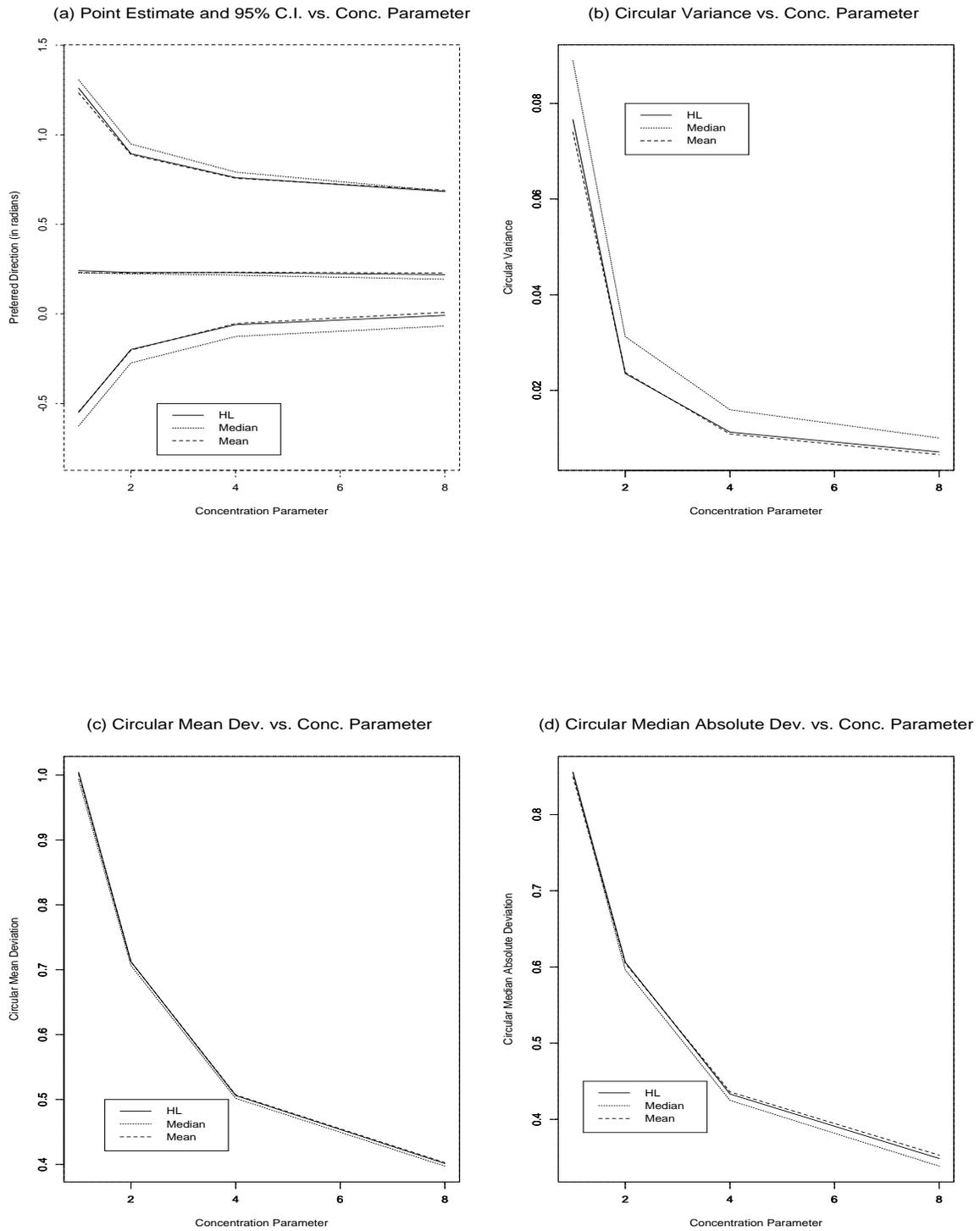
4.5.2 Effect of Concentration Parameter (κ) on the Mean, the Median and HL

The effect of concentration parameter (κ) on the three measures of preferred direction: the mean, the median and HL for data from contaminated von Mises is illustrated in Figure 4.9 for $\epsilon = 0.3$, & $n = 20$. All measures reflect the shift in preferred direction through contamination. The mean has the narrowest confidence band, followed closely by HL. The confidence bands for the median are widest over the whole range of κ s considered, see Figure 4.9(a). The confidence band for HL is sandwiched between that of the mean and the median, but very close to the mean. In general, the confidence bands become narrower as κ increases for all the three measures.

In terms of circular variance, see Figure 4.9(b), the mean and HL compete favorably, while the median has the largest circular variance over most of the range of concentration parameters considered. Note that the circular variance of HL is sandwiched between the circular variances of the mean and the median, but closer to that of the mean than to the median. Circular variance decreases as the concentration parameter increases for all measures.

Figure 4.9(c) the three measures have circular mean deviation (CMD) that are almost identical over most of the range of concentration parameters considered. In the case of circular median absolute deviation (CMAD), the median has the smallest CMAD followed closely by the mean and HL (which are almost identical over the whole range of concentration parameters considered), see Figure 4.9(d). Note that both the median and HL have smaller CMD and CMAD for $\kappa \geq 4$ compared to the mean. Similar results to Figure 4.9 were obtained for other sample sizes, See *Appendix D*, for the results of $n = 10$.

Figure 4.9: Mean, Median and HL for 70%VM(0, κ) & 30%VM($\frac{\pi}{4}$, κ), $\epsilon = 0.3$, $n = 20$



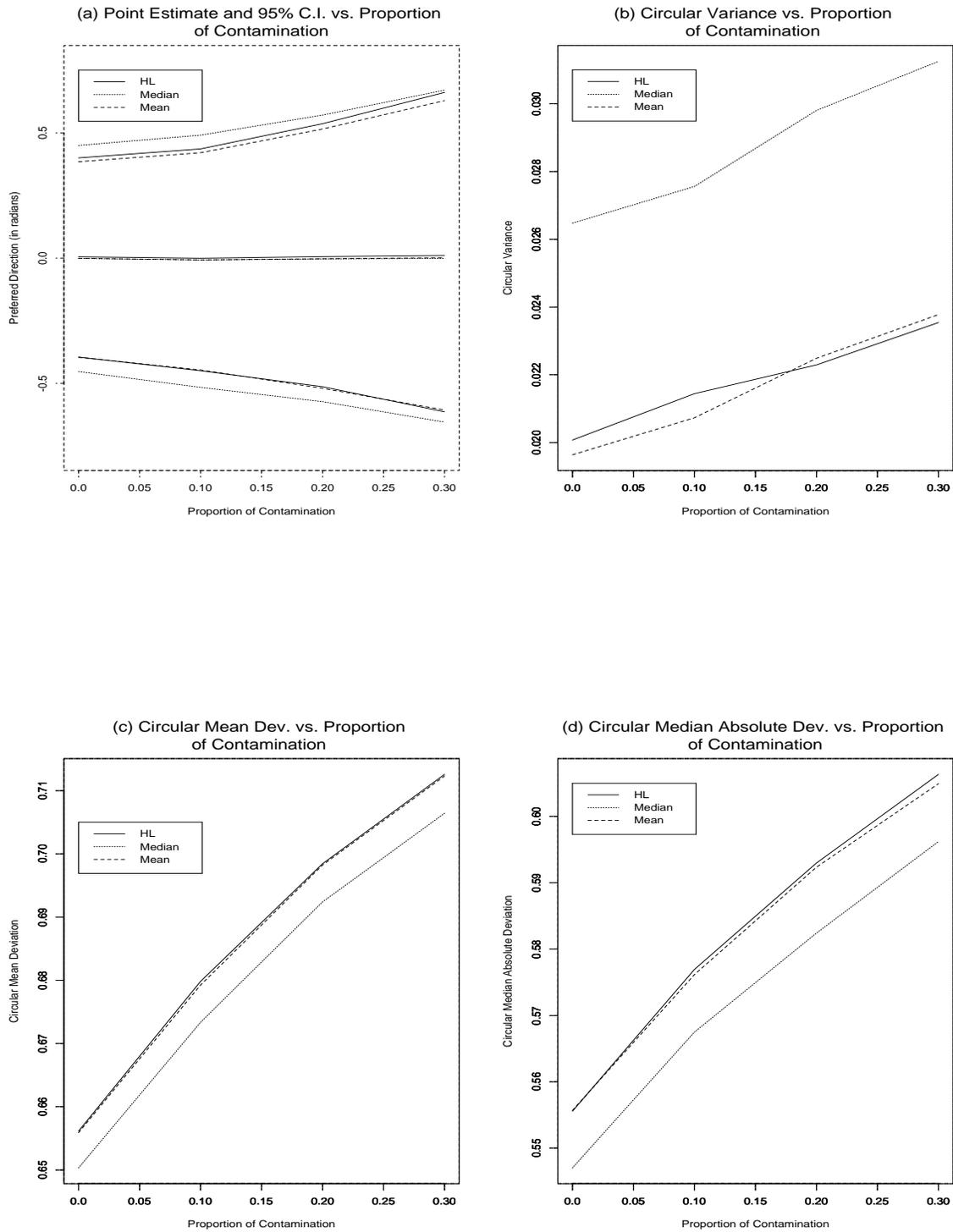
4.5.3 Effect of contamination level on Mean, Median and HL, $n=20$

The effect of increasing the contamination level on the three measures of preferred direction: the mean, the median and HL for data from contaminated von Mises is illustrated in Figure 4.10 for $\kappa = 2$, & $n = 20$. Notice the shift in preferred direction as the proportion of contamination increases. The mean has the narrowest confidence band, followed closely by HL. The confidence bands for the median are widest over the whole range of sample sizes considered, see Figure 4.10(a). The confidence band for HL is sandwiched between that of the mean and the median. In general, the confidence bands become wider as the proportion of contamination increases for all the three measures.

In terms of circular variance, in Figure 4.10(b), the mean and HL compete favorably, while the median has the largest circular variance over the whole range of proportion of contamination considered. The mean and HL seem comparable depending on the percentage of contamination. Circular variance increases as the amount of contamination increases for all measures.

Figure 4.10(c) shows that the median has the smallest circular mean deviation (CMD), followed by the mean and HL (which are almost identical) the whole range of amount of contamination considered. Similarly, from Figure 4.10(d), we observe that the circular median absolute deviation (CMAD) is smallest for the median and largest for the mean and HL over most of the range of amount of contamination considered. Similar results to Figure 4.10 were obtained for other concentration parameters and sample sizes, See Appendix D, for the results of $\kappa = 1, 4$ & 8 .

Figure 4.10: Effect of contamination level on Mean, Median and HL, $n=20$



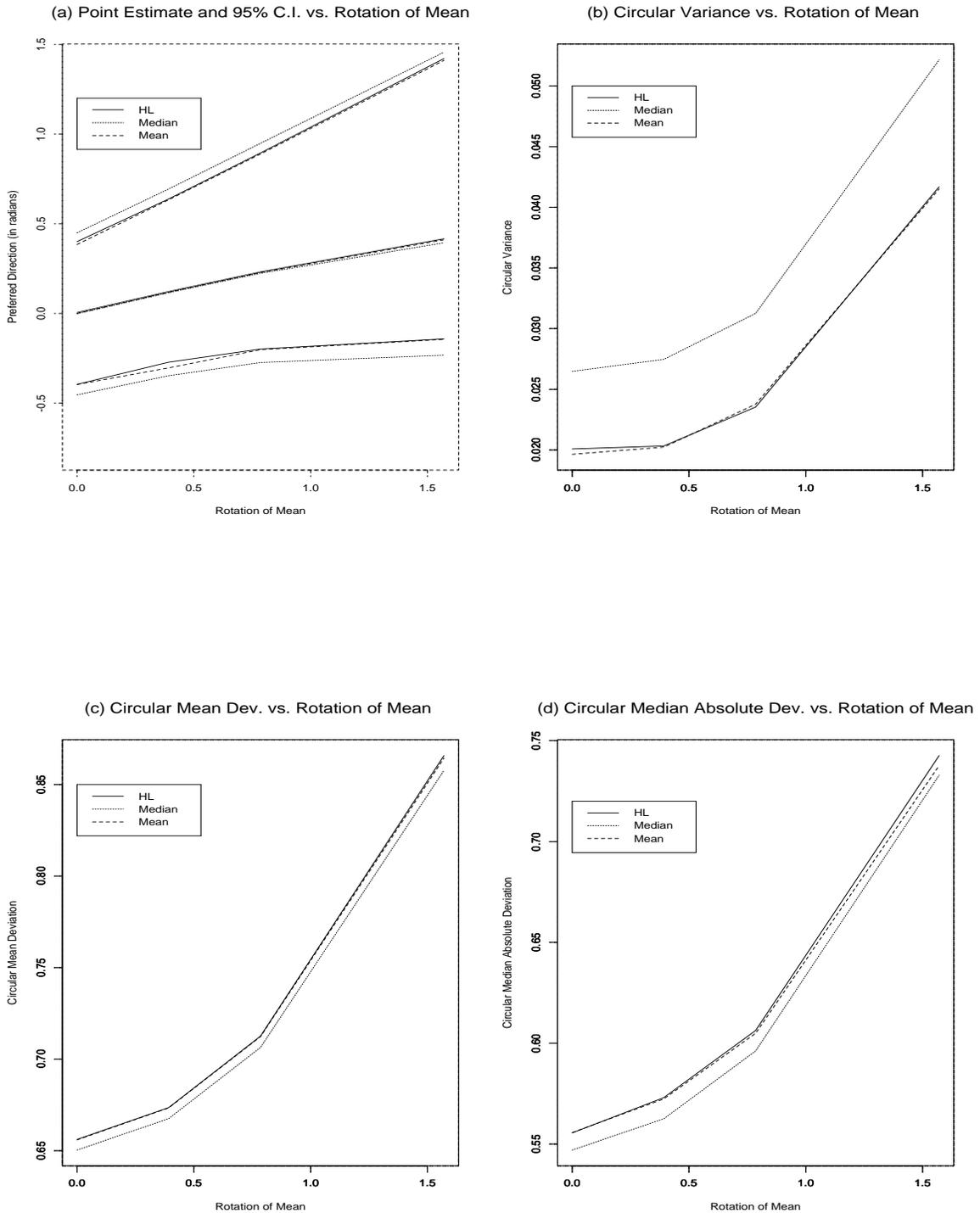
4.5.4 Effect of increasing amount of shift of mean on Mean, Median and HL, $n=20$

The effect of increasing amount of rotation on the three measures of preferred direction: the mean, the median and HL for data from contaminated von Mises is illustrated in Figure 4.11 for $\kappa = 2$ & $n = 20$. Notice the shift in preferred direction as the proportion of contamination increases. The mean has the narrowest confidence band, followed closely by HL. The confidence bands for the median are widest over the whole range of sample sizes considered, see Figure 4.11(a). The confidence band for HL is sandwiched between that of the mean and the median. In general, the confidence bands become wider as the proportion of rotation increases for all the three measures.

In terms of circular variance, in Figure 4.11(b), the mean and HL are the smallest, the median has the largest circular variance over the whole range of rotation of mean considered. Circular variance increases as the amount of rotation of the mean increases for all measures.

Figure 4.11(c) shows that the median has the smallest circular mean deviation (CMD), followed by the mean and HL (which are almost identical) the whole range of amount of rotation of the mean considered. Similarly, from Figure 4.11(d), we observe that the circular median absolute deviation (CMAD) is smallest for the median and largest for the mean and HL over most of the range of amount of rotation of the mean considered. Note that in general, both CMD and CMAD increase as the proportion of contamination increases for all the three measures. Similar results to Figure 4.11 were obtained for other concentration parameters and sample sizes, See Appendix D, for the results of $\kappa = 1, 4$ & 8 .

Figure 4.11: Effect of Shifting the mean direction from 0 to $\frac{\pi}{2}$ on Mean, Median and HL, $\kappa = 2$ & $n = 20$



4.5.5 Discussions and Conclusions on Relative Performance of Mean, Median and HL

Different concentrations give similar results for both uncontaminated and contaminated data. The mean, unlike in the linear case, does not break down in case of contaminated data, and most results stay relatively consistent even as amount of contamination increases. Under contamination of spread, contamination in direction and larger sample sizes, the HL estimate is consistently ranked second. However, for different measures, the mean or median fluctuate between best and worst. This result shows the robust behavior of the HL measure over a variety of measures of dispersion.

Overall, the new measure (HL) is a good compromise between circular mean and circular median. Relative performance of the mean and HL estimates are consistent for both uncontaminated and contaminated data. HL is less robust to outliers compared to the median, however it is an efficient alternative, since it has a smaller circular variance. HL also provides a robust alternative to the mean especially in situations where the model of choice of circular data (the von Mises distribution) is in doubt.

Chapter 5

Bootstrap: Confidence Interval and Test of Hypothesis

5.1 Introduction

The statistic used to estimate the population preferred direction is the sample preferred direction. Because of sampling error, one cannot expect the sample preferred direction to be exactly equal to the population preferred direction. Thus, it is important to provide information about the accuracy of the estimate in addition to just a single point estimate. This leads to the discussion of confidence intervals, the main topic of this Chapter.

Recall that a point estimate for a parameter is the value of a statistic that is used to estimate the parameter. On the other hand, a confidence interval estimate for a parameter consists of an interval of numbers together with a percentage that specifies how confident we are that the parameter lies in the interval. The confidence percentage is called the confidence level. For a fixed sample size, the greater the confidence level, the greater the width of the confidence interval.

Given a sample of circular data, a common reason for obtaining the sampling distribution is to test some hypothesis. A hypothesis of interest could be for example, to test whether a sample could reasonably have arisen from a population in which the preferred direction was $\mu = \mu_0$, some particular value.

If the primary concern is deciding whether a population preferred direction (μ) is different from a specified value (μ_0), then the alternative hypothesis should be $\mu \neq \mu_0$. Such a hypothesis is called a two-tailed test. One-tailed tests would occur if the primary interest is deciding whether a population preferred direction is less than or greater than a specified value. However, such one-tailed tests are not sensible in circular data, since there is no natural ordering of observations, given the wrap-around nature of the data. Thus we shall only discuss two-tailed tests in this Chapter.

Two-tailed hypothesis tests and confidence intervals are closely related. In a two-tailed hypothesis test for a population preferred direction at the significance level α , the null hypothesis is rejected if and only if the value of the preferred direction in the null hypothesis lies outside the $(1 - \alpha)$ -level confidence interval for the population preferred direction. Conversely, the null hypothesis is not rejected if and only if the value of the population preferred direction in the null hypothesis lies within the $(1 - \alpha)$ -level confidence interval for the population preferred direction.

Bootstrap methods (useful especially in situations where distributional assumptions are kept to a minimum or when the distributional results for the quantity of interest do not exist) have found enormous use in directional data. For example, a general approach to calculating bootstrap confidence arcs (symmetric, equal-tailed and likelihood-based arcs, see Section 2.4) based on pivots is proposed by Fisher and Hall (1989), improving on the method proposed

by Durchame et al (1985), which allows only symmetric confidence arcs. In this Chapter, we will use Fisher and Hall's approach to obtain the bootstrap confidence intervals for our new measure of preferred direction; the circular Hodges-Lehmann estimator.

In Chapter 3, we proposed an adaptation of the Hodges-Lehmann estimator for linear data as an alternative estimate of preferred direction for circular data. However, since no simple form of the distribution of the pairwise circular means exists for all κ values for both von Mises and non von Mises data, (Mardia, 1972, p. 97-98, 126 and Mardia & Jupp, 2000, p. 69), we shall present both nonparametric and parametric bootstrap methods to estimate the bootstrap confidence intervals for HL and compare them to those for the circular mean and circular median. We hope to identify that confidence interval method which is best.

5.2 Bootstrap method

The bootstrap is a computer intensive resampling technique (introduced by Efron, 1979), for estimating the variance, confidence intervals and bias of an estimator and sampling distribution of a given statistic, with little or no assumptions about the distribution of the statistic. The bootstrap method involves empirically estimating the entire sampling distribution of $\hat{\theta}$, by examining the variation of the statistic within the sample. Note the sampling distribution of $\hat{\theta}$ can be thought of as the distribution of values of the statistic calculated from an infinite number of random samples of size n from a given population (Mooney & Duval, 1993).

Two commonly used bootstrap approaches are nonparametric bootstrap and parametric bootstrap. In the following sections, we shall explain how both of these approaches are used and when they are most appropriate.

5.2.1 Nonparametric Bootstrap

In the nonparametric bootstrap approach, we do not assume that we know the population probability distribution. Thus from the observed data, we not only obtain an estimate of the parameter of interest, $\tilde{\theta}$, but we also obtain an estimate of the entire distribution from which it came. Application of the method requires generation of bootstrap samples: samples of size n drawn at random with replacement from the empirical distribution function F_n . F_n is the distribution which assigns probability $\frac{1}{n}$ to each observation θ_i for $i = 1, 2, \dots, n$. For each bootstrap sample, an estimator $\tilde{\theta}^*$ is computed. Given B bootstrap samples, the empirical distribution of the B values of $\tilde{\theta}^*$'s can be used to estimate the characteristics of $\tilde{\theta}$ including the variability, the mean, and quantiles for $\tilde{\theta}$. Typically, B should be between 50 – 200 to estimate the standard error of $\tilde{\theta}$, and at least 500 to estimate the confidence intervals around $\tilde{\theta}$ (Efron & Tibshirani, 1993, Sec. 9).

5.2.2 Parametric Bootstrap

The idea for the parametric bootstrap was derived from the situation which we have a good idea about the underlying distribution of the population from which the original data is drawn, but do not know how to translate this knowledge into distributional results for the estimate of the parameter of interest. The parametric bootstrap works as follows for our situation. Assume that data (Θ) comes from a von Mises with mean μ and concentration parameter κ . From the data, calculate the parameter estimates $\hat{\mu}$ and $\hat{\kappa}$. Sample Θ^*, n observations from von Mises with mean $\hat{\mu}$ and concentration parameter $\hat{\kappa}$. Compute the estimator, $\tilde{\theta}^*$. Repeat this process B times. Use the B estimators to get information about the distribution of $\tilde{\theta}^*$. This approach allows for a wide variety of values for the new sample observations. It uses and incorporates all known information about the observations into the estimation of the distribution of the estimated parameter of interest. In addition, the parametric bootstrap method tends to give smoother estimates of the distribution for small sample sizes and for parameters of interest that use the numerical values of only a small

number of data values (e.g. median, min, max).

Note that, if we define Monte Carlo methods to include sampling from estimated distributions, parametric bootstrap may be viewed as a form of Monte Carlo. If the bootstrap estimate is the limit of this Monte Carlo estimate as the number of bootstrap samples tends to infinity, then the bootstrap samples give a Monte Carlo approximation to the bootstrap estimate.

5.3 Algorithm for Bootstrapping HL

For the nonparametric bootstrap, the following four stages are used.

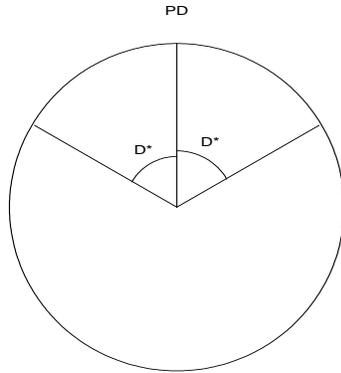
Stage 1. Resampling. Draw a random sample of n values with replacement, from original sample denoted by \mathcal{S} . Denote the bootstrap resample by \mathcal{S}^* . Note that some of the original sample values may appear more than once, and others not at all.

Stage 2. Calculation of bootstrap estimate. Compute $\frac{n(n+1)}{2}$ pairwise circular means. Obtain $\tilde{\theta}^*$, the circular Hodges-Lehmann estimate from all $\frac{n(n+1)}{2}$ values, as described in Chapter 3, Section 3.4. Note that to estimate the distribution of the circular mean or circular median, these statistics would be calculated from the bootstrap sample.

Stage 3. Repetition. Repeat Stages 1 and 2 to obtain a total of B bootstrap estimates $\tilde{\theta}_1^*, \tilde{\theta}_2^*, \dots, \tilde{\theta}_B^*$. An alternate point estimate of the circular Hodges-Lehmann estimator is the circular mean of these B $\tilde{\theta}^*$'s, and the corresponding resultant length is the length of the mean vector of these $\tilde{\theta}^*$'s.

Stage 4 (a): Confidence Intervals. Confidence Intervals can be calculated for the B bootstrap estimates in a number of different ways. Three approaches proposed by Fisher & Hall (1989) are presented below for a $(1 - \alpha)100\%$ Confidence Interval.

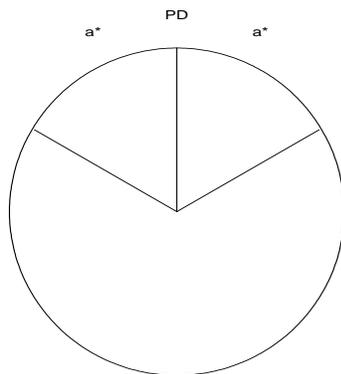
Figure 5.1: Symmetric Arc Confidence Interval



Technique 1: Symmetric Arc

This method uses the point estimate, PD, as the middle of the interval and selects the angle D^* such that $(1 - \alpha)B$ of the $\tilde{\theta}^*$ values lie within the interval. This approach is designed for measures with an assumed symmetric distribution.

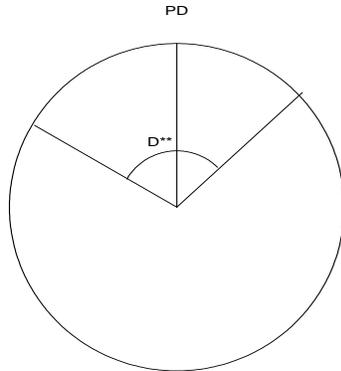
Figure 5.2: Equal-Tailed Confidence Interval



Technique 2: Equal-Tailed Arc

This method uses the point estimate of preferred direction, PD, as the middle observation. It then defines the endpoints of the confidence interval as the location where $\frac{(1 - \alpha)}{2}$ of the bootstrap values $\tilde{\theta}^*$ lie between the edge of the interval and PD. This approach is flexible enough to deal with skewed distributions.

Figure 5.3: Likelihood-Based Arc Confidence Interval



Technique 3: Likelihood-Based Arc

This method chooses the shortest arc containing $(1 - \alpha)B$ of the θ^* values, where D^{**} is the width of the confidence interval. This is the most flexible method.

Stage 4 (b): Hypothesis Testing. This approach relies on the connection between confidence intervals and tests of hypotheses. If $\mu \in (1 - \alpha)100\%$ confidence interval, then fail to reject the null hypothesis that $\mu = \mu_0$, else reject the null hypothesis in favor of the alternative hypothesis that $\mu \neq \mu_0$, at level α .

The parametric bootstrap replaces Step 1 with sampling from the assumed distribution to obtain a sample of size n.

An S-Plus program to implement these procedures (both nonparametric and parametric) for our new measures is given in *Appendix E*.

Preliminary study of Bias-Corrected intervals based on the Symmetric Arc, Equal-Tailed Arc and Likelihood-Based Arc, showed no improvements from the uncorrected intervals. A

detailed simulation study needs to be carried out for a variety of data sets to establish if these methods (uncorrected and bias-corrected) lead to similar intervals for circular data.

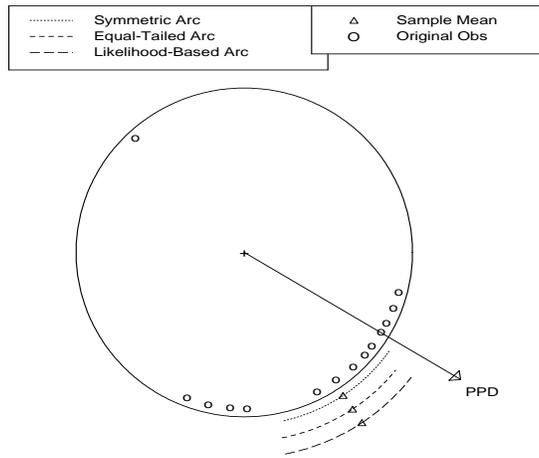
5.4 Examples

In this Section, we compare the results of three bootstrap procedures (Symmetric-Arc, Equal-Tailed Arc and Likelihood-Based Arc) and the Median Theory procedure on two data sets, the frog data of Ferguson et. al. (1967) and cross-bed azimuths of palaeocurrents of Fisher and Powell (1989).

5.4.1 Example 1: Frog Data

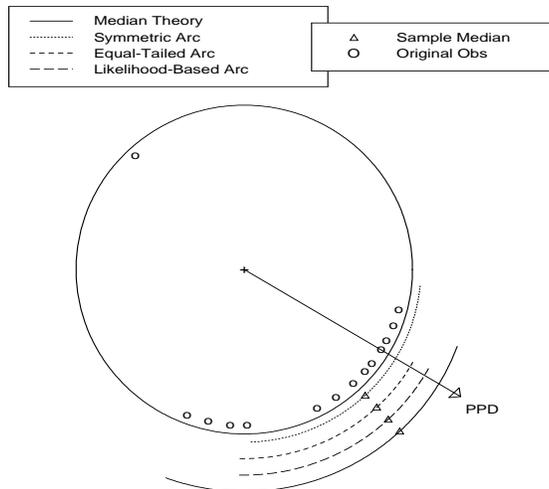
The data in this example relates the homing ability of the Northern cricket frog, *Acris crepitans*, as studied by Ferguson, et. al. (1967). For this data set, it is thought that the preferred direction for the population is 121^0 (where 0^0 is taken to be true North, and angles are measured in a clockwise direction), (Collett, 1980). The sample appears to consist of a single modal group, with one observation which can be classified as an outlier. We wish to obtain the point estimate as well as the corresponding confidence interval using three measures of preferred direction, namely: the mean, the median and the new measure, HL.

Figure 5.4: 95% Bootstrap C.I for Circular Mean, $n = 14$, $B = 1000$



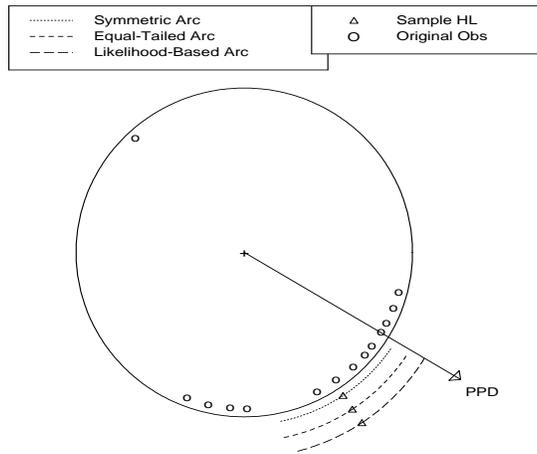
Note that the three Confidence Intervals have similar widths., with the Symmetric-Arc being slightly shifted. The hypothesized population preferred direction is denoted by “PPD”. Using the connection between Confidence Intervals and Hypothesis Testing, we can say that we reject the hypothesis that the population mean is 121^0 , since for all intervals the “PPD” is outside the confidence arc.

Figure 5.5: 95% Bootstrap C.I for Circular Median, $n = 14$, $B = 1000$



The confidence interval based on the Median Theory method of Fisher and Powell (1989) has the widest width, followed by the Symmetric-Arc. The Equal-Tailed and Likelihood-Based approaches yield identical results. All the four procedures, lead us to fail to reject the null hypothesis that the population median is 121^0 .

Figure 5.6: 95% Bootstrap C.I for HL, $n = 14$, $B = 1000$



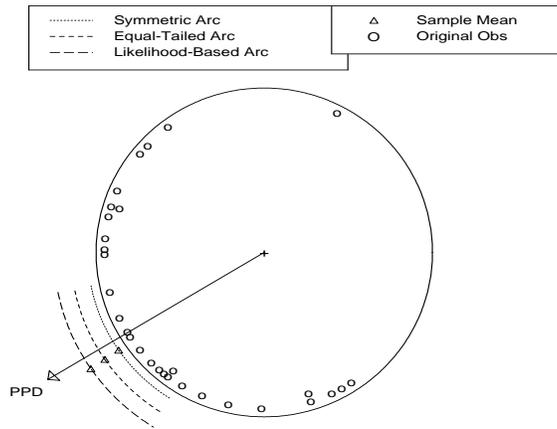
For the HL estimate of preferred direction, the Symmetric-Arc, Equal-Tailed and Likelihood-Based confidence intervals all appear to have similar widths although they are shifted. The null hypothesis that the population preferred direction based on the Hodges Lehmann measure is 121° , is rejected using all the three procedures.

Note that for this data set, the confidence intervals using the mean and HL as estimates of preferred direction, lead to the rejection of the null hypothesis. Using the median as the estimate of preferred direction leads us to fail to reject the null hypothesis.

5.4.2 Example 2: Cross-bed Azimuths of Palaeocurrents

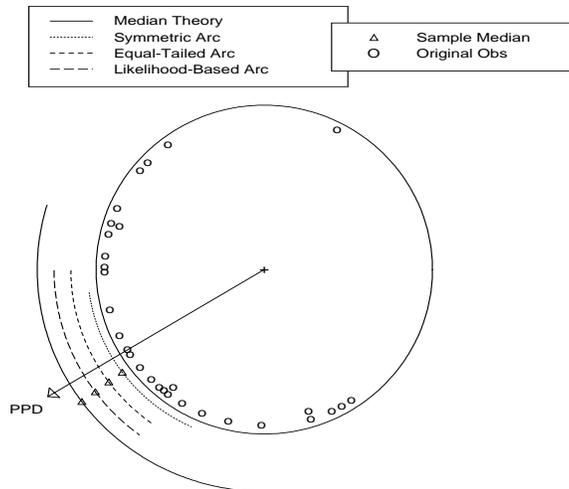
The data in this example are 30 measurements of cross-bed azimuths of paleocurrents measured in the Belford Anticline in New South Wales (Fisher, 1993, p. 59). The sample appear to consist of a single modal group, with one observation which can be classified as an outlier. It is desired to summarise the data with the point estimate and corresponding confidence interval, which can then be compared and possibly combined with corresponding information from other samples in the region. It is thought from previous studies that the population preferred direction (PPD) is 239° , where 0° is taken to be true North and angles are measured in a clockwise direction.

Figure 5.7: 95% Bootstrap C.I for Circular Mean, $n = 30$, $B = 1000$



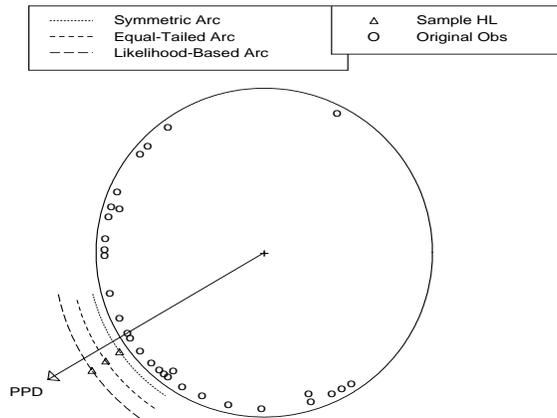
For the mean, all three confidence intervals are identical in terms of confidence width. The population preferred direction, “PPD” lies within all the intervals, hence we conclude that the data are consistent with a population with this “PPD”.

Figure 5.8: 95% Bootstrap C.I for Circular Median, $n = 30$, $B = 1000$



The confidence interval based on the Median Theory method has the widest length, followed by the Symmetric-Arc. The Equal-Tailed and Likelihood-Based methods are identical. All intervals include “PPD”, hence we conclude that the the data was drawn from a population with this “PPD”.

Figure 5.9: 95% Bootstrap C.I for HL, $n = 30$, $B = 1000$



For the HL estimate of preferred direction, all the three confidence intervals have similar widths. Here, the population preferred direction, lies inside all the three intervals, implying that we should accept the null hypothesis that this sample came from a population with this “PPD”.

Note that for this data set, the confidence intervals using the three measures as estimates of preferred direction, all lead us to fail to reject the null hypothesis.

5.5 Simulation Study

We conducted a study to compare the relative performance of the different nonparametric confidence intervals for the mean, median and HL based on symmetric arc, equal-tailed arc and likelihood-based arc. The parametric confidence intervals using the von Mises assumption were not considered due to limitations of time. The data considered were values from von Mises distribution with concentration parameters 2 and 10. Symmetric Arc (SYMA), Equal-Tailed Arc (ETA), and Likelihood-Based Arc (LBA) were constructed as described in Section 5.3. The theoretical confidence interval for the median (MT), given in Section 2.4.2, was also constructed.

The number of bootstrap samples in the study was 500, repeated 1000 times at $\alpha = 0.1$ and $\alpha = 0.05$ as in Fisher and Hall (1989) and Durchame et al (1985). For each confidence interval a number of measures were obtained:

- the Nominal Coverage, which is the $(1 - \alpha)$ assumed level.
- the Simulated Coverage, which calculates the number of observed confidence intervals actually containing the true preferred direction.
- Interval Width, which is the average Width of the Interval from the 1000 simulations.
- Resultant Length, which is a circular measure of variability of the interval widths of the 1000 simulation results.

Here, we report results of a small-sample simulation study on estimating preferred direction for the von Mises distribution. Median Theory confidence intervals (see Section 2.4.2), and the Symmetric-Arc (SYMA), Equal-Tailed Arc (ETA), and Likelihood-Based Arc (LBA) bootstrap confidence intervals for each of the mean, median and HL, were constructed. Results for a sample of size 10, for $\kappa = 2$ & $\kappa = 10$ are summarized in Tables 5.1 and 5.2, respectively.

Table 5.1: Bootstrap Confidence Intervals for $n = 10, \kappa = 2, M = 1000, B = 500$.

Nom.Cov.	Measure	Type	Sim. Cov.	Int. Width	Res.Length
0.950	Mean	SYMA	0.902	1.038	0.944
		ETA	0.908	1.032	0.947
		LBA	0.902	1.007	0.949
0.950	Median	MT	0.977*	1.688	0.885
		SYMA	0.922	1.310	0.879
		ETA	0.879	1.211	0.919
		LBA	0.879	1.093	0.928
0.950	HL	SYMA	0.924	1.114	0.938
		ETA	0.920	1.069	0.946
		LBA	0.895	1.013	0.951
0.900	Mean	SYMA	0.850	0.871	0.960
		ETA	0.855	0.873	0.962
		LBA	0.847	0.851	0.964
0.900	Median	MT	0.884**	1.102	0.924
		SYMA	0.863	1.111	0.901
		ETA	0.833	0.983	0.931
		LBA	0.831	0.907	0.939
0.900	HL	SYMA	0.874	0.935	0.953
		ETA	0.871	0.910	0.960
		LBA	0.833	0.852	0.965

Median Theory (MT), Symmetric-Arc (SYMA), Equal-Tailed Arc (ETA) and Likelihood-Based Arc (LBA)

* True Nominal Coverage is 0.978, ** True Nominal Coverage is 0.893, Fisher (1993, p. 226).

From Table 5.1 with the results of $\kappa = 2$, indicate that within each measure, all bootstrap intervals have true coverage below nominal value and both the 95% & 90% confidence intervals seem to miss nominal coverage by similar amounts. In terms of interval width and coverage, the Likelihood-Based has the shortest, with Equal-Tailed, generally next shortest followed by the Symmetric-Arc, for all the three measures. A similar ordering is observed for the variability of the widths as well. That is, the Likelihood-Based has the largest Resultant length (thus, the smallest circular variance), followed by the Equal-Tailed, and then the Symmetric-Arc. The measures have the following ordering in terms of shortest interval width and smallest variability of widths: the mean, then HL, the median. This ordering holds for the Symmetric-Arc, Equal-Tailed and Likelihood-Based bootstrap confidence intervals. These results match those obtained from the simulations in Chapter 4.

In general, the Median Theory, leads to intervals that have larger widths and larger variability compared to the three bootstrap confidence intervals. However, this method is the only one with observed coverage probability consistent with the Median Theory in Table A.6, in Fisher (1993, p. 226).

The results for $\kappa = 10$ in Table 5.2, give similar results to those of $\kappa = 2$. However, the interval widths are much shorter and the variability of the widths are much smaller compared to those in Table 5.1. Note that, across all measures, the confidence intervals converge as κ increases. Also, within each measure, the different types of confidence intervals become more similar as κ increases.

Based on these results, the median theory is a good choice because of its appropriate size given a specified Confidence level. Of the bootstrap methods, the Likelihood-Based seems best, despite the fact that it does not use the point estimate. The Equal-Tailed confidence interval also performs well. Comparing the mean, median and new estimate HL, the mean

generally leads to the smallest intervals with similar true coverage probabilities. The HL measure also performs quite well.

Table 5.2: Bootstrap Confidence Intervals for $n = 10, \kappa = 10, M = 1000, B = 500$.

Nom. Cov.	Measure	Type	True Cov.	Int. Width	Res.Length
0.950	Mean	SYMA	0.907	0.372	0.996
		ETA	0.905	0.371	0.996
		LBA	0.890	0.364	0.997
0.950	Median	MT	0.977*	0.657	0.984
		SYMA	0.929	0.518	0.985
		ETA	0.886	0.467	0.980
		LBA	0.886	0.425	0.992
0.950	HL	SYMA	0.922	0.419	0.994
		ETA	0.923	0.402	0.995
		LBA	0.892	0.381	0.996
0.900	Mean	SYMA	0.855	0.315	0.997
		ETA	0.863	0.315	0.997
		LBA	0.843	0.309	0.997
0.900	Median	MT	0.880**	0.425	0.988
		SYMA	0.881	0.419	0.988
		ETA	0.845	0.378	0.991
		LBA	0.845	0.350	0.993
0.900	HL	SYMA	0.877	0.353	0.995
		ETA	0.869	0.343	0.996
		LBA	0.844	0.317	0.997

Median Theory (MT), Symmetric-Arc (SYMA), Equal-Tailed Arc (ETA) and Likelihood-Based Arc (LBA)

True Nominal Coverage is 0.978, ** True Nominal Coverage is 0.893, Fisher (1993, p. 226).

Chapter 6

Conclusions and Future Study

6.1 Conclusions

In conclusion, a new measure, the circular Hodges-Lehmann estimate has been proposed. Three variations of this measure were given that all perform similarly, as with their counterparts in the linear case. Theoretical results for the new measure were obtained for concentrated von Mises data for the distribution, influence function and asymptotic relative efficiency relative to the circular mean and median. As the data become extremely concentrated the results become consistent with the linear case. The circular Hodges-Lehmann estimator has comparable efficiency to mean and is superior to median. Most strikingly, it consistently ranks second and is never worst for all considered measures of dispersion, for both uncontaminated and contaminated data. A method for using the new measure for estimating confidence intervals and hypothesis testing for all concentrations is obtained using extensions to the bootstrap confidence methodology. The new measure again leads to comparable results with the circular mean and shorter bootstrap confidence intervals compared to the circular median. Overall, the circular Hodges-Lehmann estimate is a solid alternative to the established circular mean and circular median with some of the desirable features of each.

6.2 Future Research

[1]. Durchame and Milasevic (1987b), obtain the Asymptotic Relative Efficiency for median to mean for the contaminated von Mises distribution. They assume that F has density $(1 - \epsilon)f_{\kappa}(\theta) + \epsilon f_{\lambda\kappa}(\theta)$, where $0 \leq \epsilon \leq 1$, where $f_{\kappa}(\theta)$ is the density of the von Mises distribution. Their simulation study reveals that the circular median is more efficient than the mean. It remains of interest to determine what happens to the efficiency of HL relative to both the mean and the median under such contaminations.

[2]. For linear data, HL for the two sample problem is discussed in several texts for example Hettmansperger (1984, p.139). We will extend the procedure developed in this dissertation to the two sample problem in circular data (paired and non-paired) to compare preferred directions.

[3]. HL for linear data is an optimal estimator for location for the logistic distribution (Lehmann, 1953). We will determine if a similar distribution on the circle has the circular HL as an optimal estimator of preferred direction.

[4]. Fisher (1993, p. 205-206) suggests fitting a nonparametric density to bootstrap estimates of the population mean μ if for example the sample is believed to come from a von Mises distribution with mean μ and concentration parameter κ . We will determine how a nonparametric density fit from bootstrap estimates of the population HL compares with that based on the mean.

[5]. The bias-corrected (BC) method for constructing bootstrap confidence intervals, assumes that $\tilde{\theta}^*$, $\tilde{\theta}$ and θ are distributed around a constant, $z_0\sigma$, where σ is the standard deviation of the respective distribution, and z_0 is a biasing constant. Details of how to

implement the BC method can be found for example in Mooney and Duval (1993, p. 37-43). Preliminary applications of the BC method to Symmetric Arc, Equal-Tailed Arc and Likelihood-Based Arc had little effect on the confidence intervals constructed in Chapter 5. An extensive simulation study to compare relative performance for the mean, the median and HL based on the BC method needs to be carried out in order to arrive at general conclusions.

Bibliography

- [1] Abramowitz, M. and Stegun, I.A. (1965). Handbook of Mathematical Functions. Dover, NY.
- [2] Ackermann, H. (1997). A note on circular nonparametrical classification. *Biometrical Journal*, 5, pp. 577-587.
- [3] Anderson, C.M. (1993). Location and Dispersion Analysis of Factorial Experiments with Directional Data. PhD Thesis, Dept. of Statistics and Actuarial Science, University of Waterloo, Ontario, Canada.
- [4] Anderson-Cook, C.M. (1996). Analysis of location and dispersion effects from factorial experiments with a directional response. S. Ghosh and C.R. Rao eds. *Handbook of Statistics*, vol. 13, pp. 241-259.
- [5] Andrews, D.R., Bickel, P.J., Hampel, F.R., Huber, P.J., Rogers, W.H., and Tukey, J.W. (1972). Robust Estimates of Location: Survey and Advances. Princeton Univ. Press. New Jersey.
- [6] Bagchi, P., and Guttman, I. (1990). Spuriousity and outliers in directional data. *Journal of Applied Statistics*, 17, pp. 341-350.
- [7] Barnett, V. and Lewis, T. (1994). Outliers in Statistical data, 3rd, ed. Wiley, Chichester.
- [8] Batschelet, E. (1981). Circular Statistics in Biology. Academic Press. London.

- [9] Beran, R. (1987). Prepivoting to Reduce Level Error of Confidence Sets. *Biometrika*, 74, pp. 457-468.
- [10] Beckman, R.J., and Cook, R.D. (1983). Outliers..... *Technometrics*, 25, 2, pp. 119-163.
- [11] Bickel, P. J., and Hodges, J. L. Jr. (1967). The asymptotic theory of Galton's test and a related simple estimate of location. *Ann. Math. Statist.*, 38, pp. 73-89.
- [12] Birch, J.B. (2002). Contemporary Applied Statistics: An Exploratory and Robust Data Analysis Approach. Lecture Notes. Virginia Polytechnic and State University.
- [13] Cabrera, J., Maguluri, G. and Singh, K. (1994). An odd property of the sample median *Statistics & Probability Letters* 19, pp. 349-354.
- [14] Collett, D. (1980). Outliers in circular data. *Appli. Statist.*, 29, pp. 50-57.
- [15] Collett, D., and Lewis, T. (1981). Discriminating between the von Mises and wrapped normal distributions. *Australian Journal of Statistics*, 23, 73-79.
- [16] Durcharme, G.R., Jhun, M., Romano, J., and Truong, K.N. (1985). Bootstrap Confidence Cones for Directional Data. *Biometrika*, 72, pp. 637-645.
- [17] Durcharme, G.R., and Milasevic, P. (1987a). Spatial median and directional data. *Biometrika*, 74, pp. 212-215.
- [18] Durcharme, G.R., and Milasevic, P. (1987b). Some asymptotic properties of the circular median. *Communications in Statist. Theory and Methods*, 16, pp. 163-169.
- [19] Efron, B. (1979). Bootstrap Methods: Another look at the Jackknife. *Ann. Statist.*, 7 pp. 1-26.
- [20] Efron, B. (1981). Nonparametric Standard Errors and Confidence Intervals. (with discussion). *The Canadian Journal of Statistics*, 9, pp. 139-172.

- [21] Efron, B. (1982). The Jackknife, the Bootstrap and Other Resampling Plans. Society for Industrial and Applied Mathematics. Philadelphia.
- [22] Efron, B. (1985). Bootstrap Confidence Intervals for a Class of Parametric Problems. *Biometrika*, 72, pp. 45-58.
- [23] Efron, B., and Tibshirani, R.J. (1993). An Introduction to the Bootstrap. Chapman Hall. New York.
- [24] Ferguson, D.E., Landreth, H.F., and McKeown, J.P. (1967). Sun compass orientation of northern cricket frog, *Acris crepitans*. *Anim. Behav.*, 15, pp. 45-53.
- [25] Fisher, N.I. (1982). Non-Parametric Statistics. *Math. Scientist*, 7, pp. 25-47.
- [26] Fisher, N.I. (1985). Spherical medians. *Journal of the Royal Statistical Society*, B47, 342-348.
- [27] Fisher, N.I. (1987). Problems with the current definitions of the standard deviations of wind direction. *J. Clim. Appl. Meteorol.* 26, pp. 1522-1529.
- [28] Fisher, N.I. (1993). Statistical analysis of circular data. Cambridge University Press. Cambridge.
- [29] Fisher, N.I., and Lewis, T. (1983). Estimating the common mean direction of several circular or spherical distributions with differing dispersions. *Biometrika* , 70, pp. 333-341.
- [30] Fisher, N.I., Lewis, T., and Embleton, B.J.J. (1987). Statistical Analysis of Spherical data. Cambridge University Press. Cambridge.
- [31] Fisher, N.I., and Powell, C. McA. (1989). Statistical analysis of two-dimensional palaeocurrent data :methods and examples *Australian Journal of Earth Science*, 36, pp. 91-107.

- [32] Fisher, N.I., and Hall, P. (1989). Bootstrap Confidence Regions for directional data. *Journal of the American Statistical Association*, 84, 408, pp. 996-1002.
- [33] Fisher, N.I., and Hall, P. (1990). New Statistical methods for directional data-I. Bootstrap comparison of mean directions and the fold test in palaeomagnetism *Geophys. J. Int.*, 101, pp. 305-313.
- [34] Fisher, N.I., and Hall, P. (1992). Bootstrap methods for directional data. *The Art of statistical science*, edited by K.V. Mardia. Wiley. NY.
- [35] Gastwirth, J.L., and Rubin, H. (1969). The behavior of robust estimators on dependent data *Annals of Statistics*, 3, pp. 1070-1100.
- [36] Gumbel, E.J., Greenwood, J.A., and Durand, D. (1953). The circular normal distribution. Theory and Tables. *Journal of the American Statistical Association.*, 48, pp. 131-152.
- [37] Hampel, F.R. (1968). Contributions to the theory of robust estimation. Ph.D. Thesis, Univ. California, Berkeley.
- [38] Hall, P. (1988a). On the Bootstrap and Symmetric Confidence Intervals. *Journal of the Royal Statistical Association. Ser. B*, 50, pp. 35-45.
- [39] Hall, P. (1988b). Theoretical Comparison of Bootstrap Confidence Intervals. *The Annals of Statistics*, 16, pp. 927-985.
- [40] Hampel, F.R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association.*, 69, pp. 383-393.
- [41] He, X., and Simpson, D.G. (1992). Robust direction estimation. *Annals of Statistics*, 20, 1, pp. 351-369.
- [42] Hettmansperger, T.P. (1984). *Statistical Inference Based on Ranks*. Wiley, NY.

- [43] Hettmansperger, T.P., and McKean, J.W. (1998). Robust Nonparametric Statistical methods. Wiley, NY.
- [44] Hodges, J.L.Jr., and Lehmann, E.L. (1963). Estimates of location based on ranks tests. *Annals of Mathematical Statistics*, 34, 2, pp. 598-611.
- [45] Hollander, M., and Wolfe, D.A. (1999). Nonparametric Statistical Models. Wiley. NY.
- [46] Hoyland, A. (1965). Robustness of the Hodges-Lehmann estimates for shift. *Annals of Mathematical Statistics*, 36, pp. 174-197.
- [47] Huber, P.J. (1981). Robust Statistics. Wiley & Sons, NY.
- [48] Jammalamadaka, S. R., and SenGupta, A. (2001). Topics in Circular Statistics. World Scientific. New Jersey.
- [49] Jupp, P.E., and Mardia, K.V. (1989) . A unified view of the theory of directional statistics 1975-1988. *Internat. Stat. Rev.*, 57, pp. 261-294.
- [50] Kendall, D.G. (1974a). Hunting quanta. *Trans. R. Soc. Lond.*, A 276, pp. 231-266.
- [51] Kendall, D.G. (1974b). Pole-seeking Brownian motion and bird navigation. *J. R. Stat. Soc. B*, 36, pp. 365-417.
- [52] Kent, J.T. (1976). Distributions, Processes and Statistics on Spheres. Ph.D Thesis. University of Cambridge. Cambridge.
- [53] Kent, J.T., Mardia, K.V., and Rao, J.S. (1979). A characterization of Uniform distribution on the circle. *Ann. Statist.* , 7, pp. 197-209.
- [54] Ko, D., and Guttorp, P. (1988). Robustness of estimators for directional data. *Annals of Statistics*, 16, pp. 609-618.
- [55] Lehmann, E.L. (1953). The power of rank tests. *Ann. Math. Statist.*, 24, pp. 23-43.

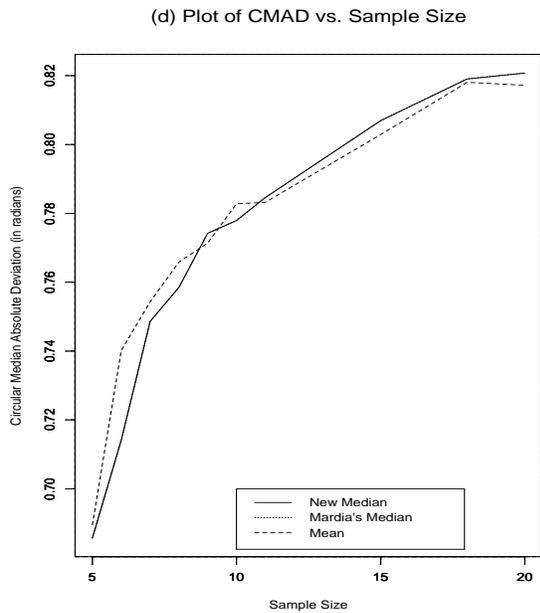
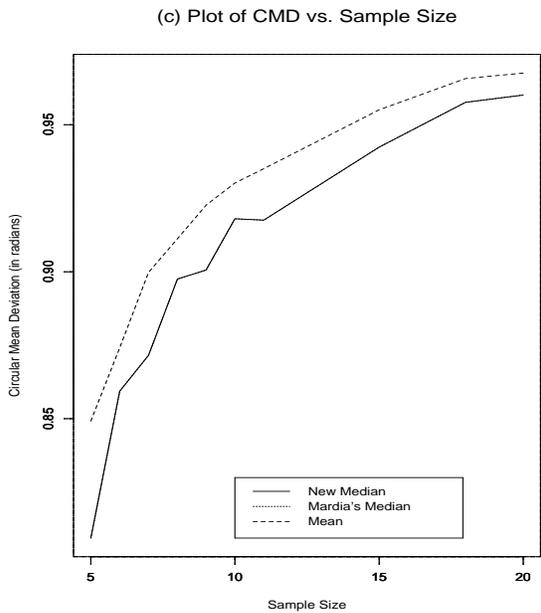
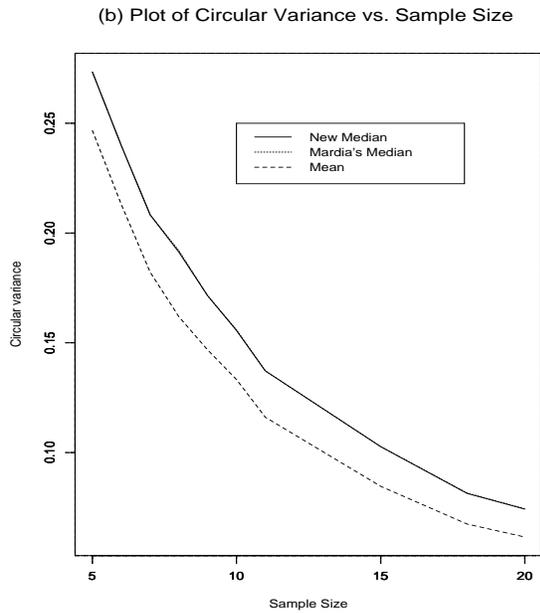
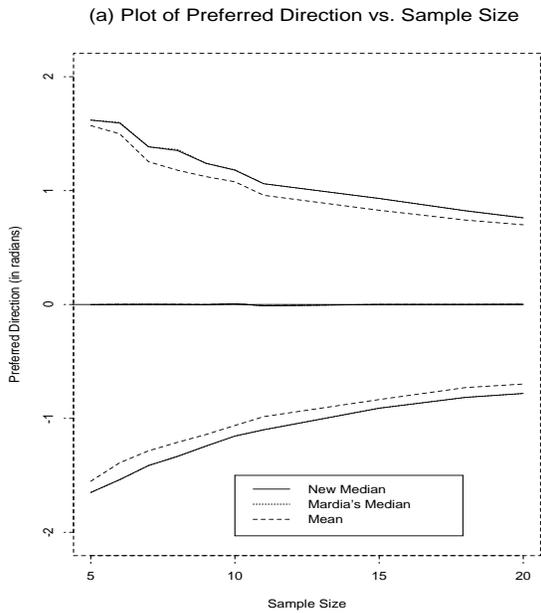
- [56] Lehmann, E.L. and D'Abrera (1998). Nonparametrics: Statistical Methods Based on Ranks. Prentice Hall, Upper Saddle River, New Jersey.
- [57] Lenth, R.V. (1981). Robust measures of location for directional data. *Technometrics*, 23, pp. 77-81.
- [58] Levy, P. (1939). L'addition des variable aleatoires defines sur une circonference. Bull. Soc. Math. France, 67, pp. 1-41.
- [59] Lewis, T. (1974). Discussion on "Pole Seeking Brownian motion and bird navigation" by Kendall, D. G. *Journal of the Royal Statistical Society B* 36, pp. 365-402.
- [60] Liu, R. (1988) Bootstrap procedures under some non-i.i.d models. *Annals of Statistics*, 16, pp. 1696-1708.
- [61] Liu, R., and Singh, K. (1992). Ordering directional data: concepts of data depth on circles and spheres. *Annals of Statistics*, 20, pp. 1468-1484.
- [62] Liu, R., and Singh, K. (1995). Using i.i.d bootstrap inference for general non - i.i.d models. . *Journal of Statistical Planning and Inference*, 43, pp. 67-75.
- [63] Manoukian, E.B. (1986). Mathematical Nonparametric Statistics. Gordon and Breach Science Publishers, NY.
- [64] Mardia, K.V. (1972). Statistics of Directional Data. London: Academic Press.
- [65] Mardia, K.V. (1975). Statistics of directional data. (With Discussion.). *J.R. Statist. Soc. B*, 37, pp. 349-393.
- [66] Mardia, K.V., and El-Atoum (1976). Bayesian inference for the von Mises-Fisher distribution. *Biometrika* 63, 1, pp. 203-206.
- [67] Mardia, K.V., and Jupp, P.E. (2000). Directional Statistics. Wiley. Chichester.

- [68] Mooney, C.Z., and Duval, R.D. (1993). Bootstrapping: A Nonparametric Approach to Statistical Inference. Sage Publications, Newbury Park.
- [69] Purkayastha, S. (1991). A rotationally symmetric directional distribution: obtained through maximum likelihood characterization. *Sankhya A*, 53, pp. 70-83.
- [70] Purkayastha, S. (1995a). An almost sure representation of sample circular median. *Journal of Statistical Planning and Inference*, 46, pp. 77-91.
- [71] Purkayastha, S. (1995b). On the asymptotic efficiency of the sample circular median. *Statistics and Decisions*, 13, pp. 243-252.
- [72] Rao, C.R. (1975). Some problems of sample surveys. Suppl. *Appl. Prob.*, 7, pp. 50-61.
- [73] Rao, J.S. (1984). Nonparametric methods in directional data. Chapter 31 in Handbook of Statistics, Vol. 4. Nonparametric statistics, editors P.R. Krishnaiah and P.K. Sen. pp. 755-770. Elsevier Science Publishers. Amsterdam.
- [74] Randles, R.H., and Wolfe, D.A. (1979). Introduction to the theory of Nonparametric Statistics, Wiley, NY.
- [75] Sirvanci, M. (1982). An estimator of location for dependent data. *Biometrika*, 69, 2. pp. 473-476.
- [76] Sheather, S.J. (1987). A new method of estimating the asymptotic standard error of the Hodges-Lehmann estimator based on generalized least squares. *Australian Journal of Statistics*, 29, pp. 66-83.
- [77] Shephard, J., and Fisher, N.I. (1982). Rapid method of mapping fracture trends in collieries. *Trans. Soc. Min. Eng. AIME* 270, pp. 1931-1932.
- [78] Small, C. G. (1990). A Survey of Multidimensional Medians. *International Statistical Review*, 58, pp. 263-277.
- [79] Stephens, M.A. (1963). Random walk on a circle. *Biometrika*, Vol. 50, pp. 385-390.

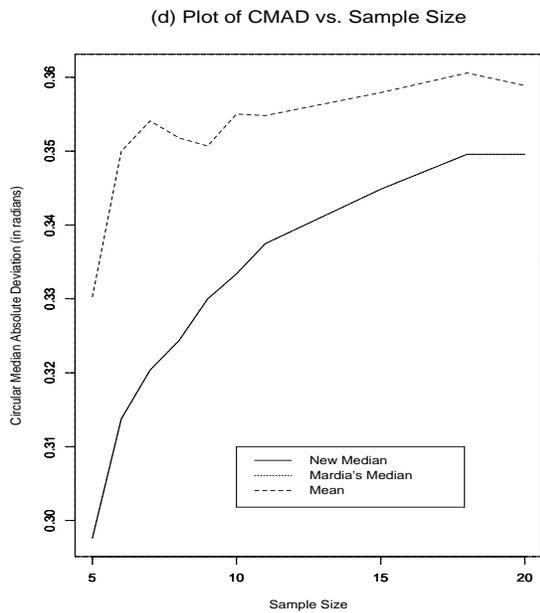
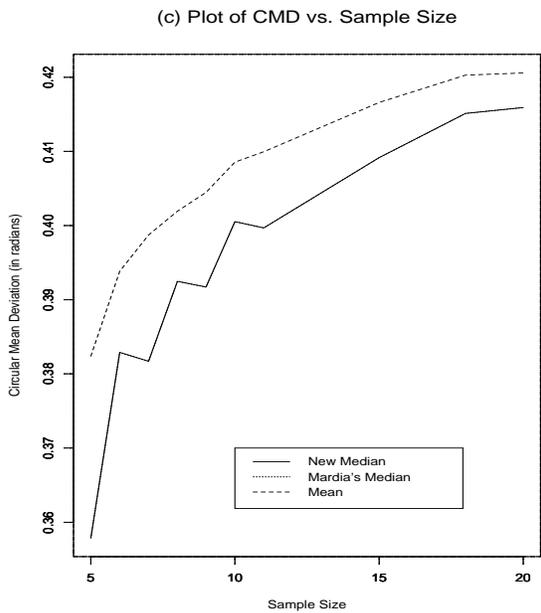
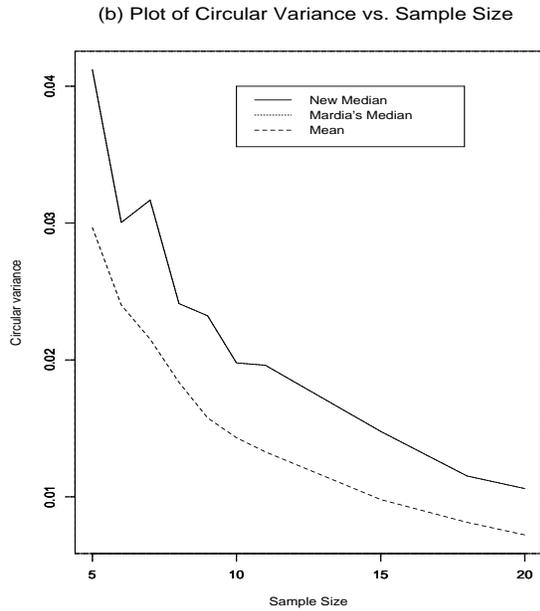
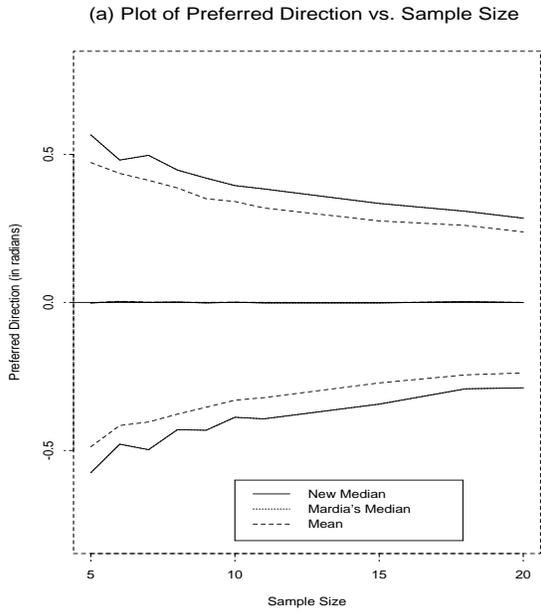
- [80] Upton, G.J.G., and Fingleton, B. (1989). Spatial Data Analysis by Example. Volume 2. Categorical and Directional Data. Wiley. NY.
- [81] Von Mises, R. (1918). Uberdie Ganzzahligkeit der Atom gewicht und verwandte Frogren. *Physikal. z.* 19, pp. 490-500.
- [82] Watson, G.S. (1967). Some problems in the statistics of directions. *Bull. 36th Session, ISI*. Sidney, Australia.
- [83] Watson, G.S. (1983a). Statistics on spheres. Wiley. NY.
- [84] Watson, G.S. (1983b). The computer simulation treatment of directional data. In Proceedings of the Geological conference, Kharagpur, India, Dec. 1983. *Indian journal of Earth Science*, pp. 19-23.
- [85] Watson, G.S. (1986). Some estimation theory on the sphere. *Ann. Inst. Statist. Math.*, 38, pp. 263-275.
- [86] Wehrly, T., and Shine, E.P. (1981). Influence curves for directional data. *Biometrika*, 68, pp. 334-335.
- [87] Wintner, A. (1933). On the stable distribution law. *American Journal of Mathematics*, 55, pp. 335-339.
- [88] Zar, J.H. (1999). Biostatistical Analysis. 4th Edition. Prentice-Hall, Inc.
- [89] Zernike, F. (1928). Wahrscheinlickkeitsrechnung und Mathematische Statistik. In Handbuch der Physik, 3, pp. 419-492. Springer Berlin.
- [90] Zielinski, R. (1987). Robustness of sample mean and sample median under restrictions on outliers. *Zastosowania Matematyki Applicationes Mathematicae*, XIX, 2, pp. 239-240.

Appendix A: Mean, Mardia Median and New Median

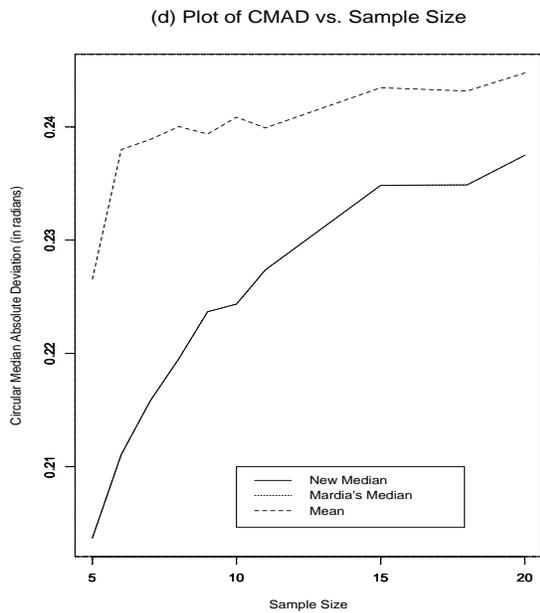
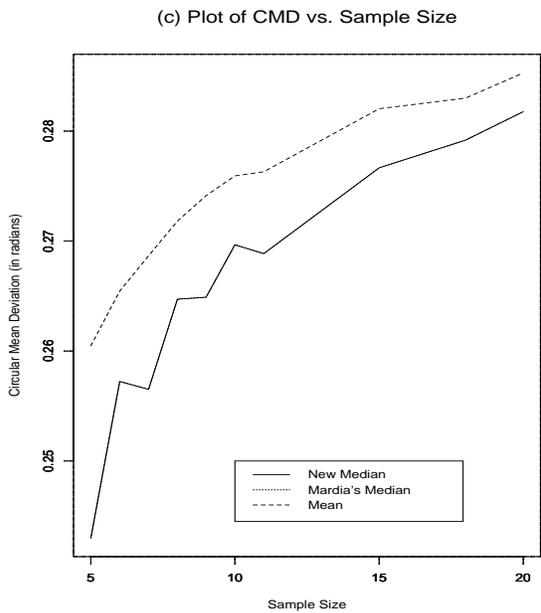
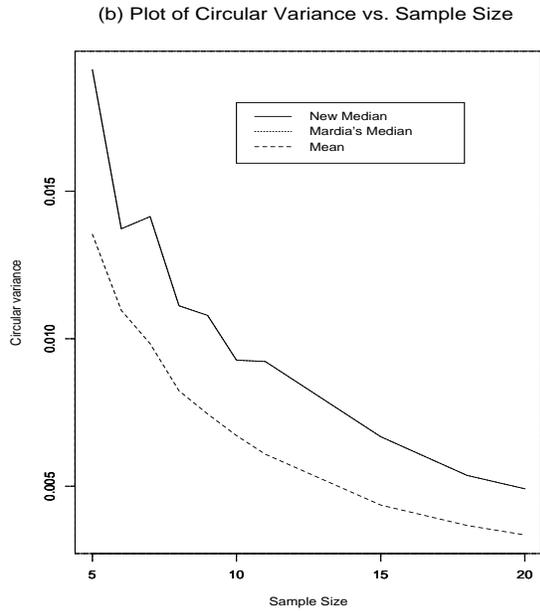
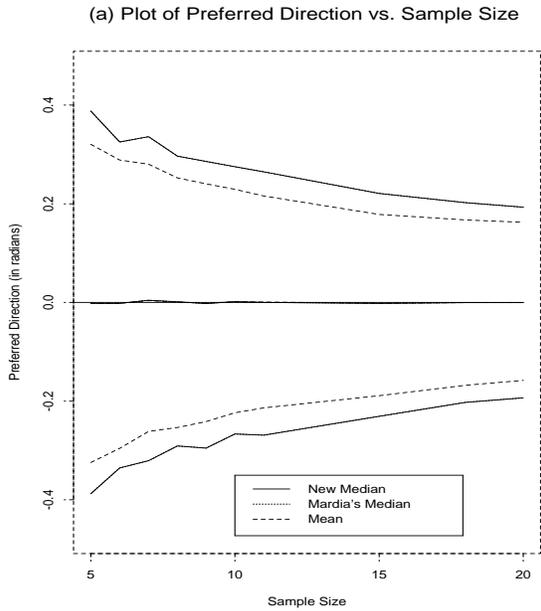
Mardia Median & New Median for VM(0,1)



Mardia Median & New Median for VM(0, 4)

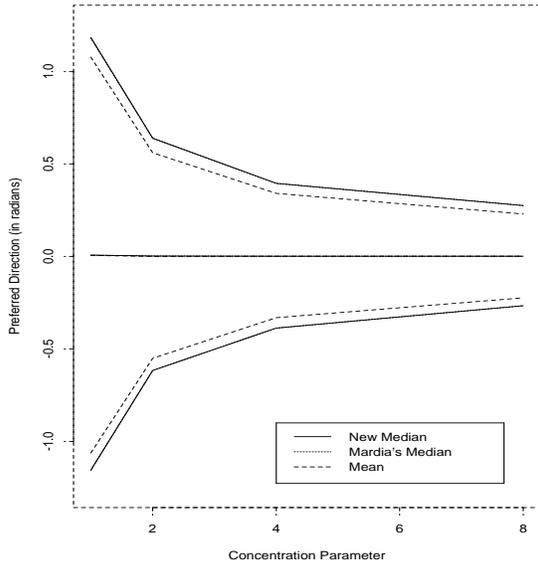


Mardia Median & New Median for VM(0, 8)

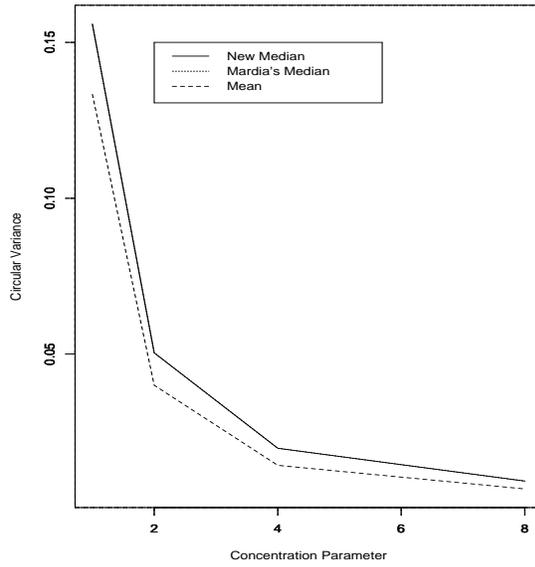


Mardia median & New median for VM $(0, \kappa)$, $n = 10$

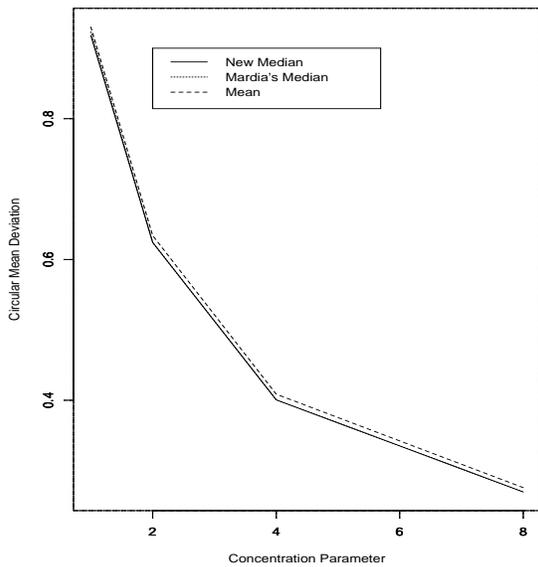
(a) Point Estimate and 95% C.I. vs. Conc. Parameter



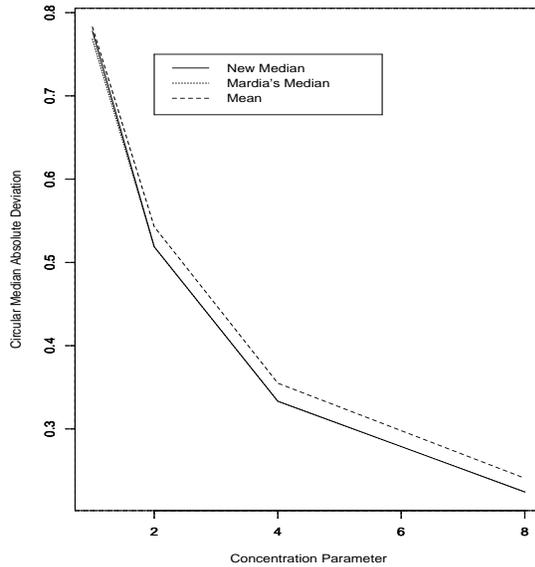
(b) Circular Variance vs. Conc. Parameter



(c) Circular Mean Dev. vs. Conc. Parameter



(d) Circular Median Absolute Dev. vs. Conc. Parameter



Appendix B: Proofs of Theorems 3.7.1 and 3.7.2

B.1 Theorem 3.7.1:

The influence function (IF) for the circular mean direction is given by

$$IF(\theta) = \frac{\sin(\theta - \mu_0)}{\rho}, \quad (1)$$

where ρ is the mean resultant length. For any given value of ρ , this influence function and its derivative are bounded by $\pm\rho^{-1}$.

Proof of Theorem 3.7.1

$$\text{Recall} \quad \int \frac{du}{u^2 + a^2} = \frac{1}{a} \tan^{-1} \frac{u}{a} + C$$

The functional for the circular population mean is $T(F) = \tan^{-1} \left[\frac{\int_0^{2\pi} \sin \theta dF(\theta)}{\int_0^{2\pi} \cos \theta dF(\theta)} \right]$

Let $\tilde{F}(\theta) = (1 - \epsilon)F(\theta) + \epsilon\delta_{\theta_0}(\theta)$, then WLOG, the functional for the circular population mean is

$$T(\tilde{F}) = \tan^{-1} \left[\frac{\int_0^{2\pi} \sin(\theta - \mu_0) dF(\theta - \mu_0)}{\int_0^{2\pi} \cos(\theta - \mu_0) dF(\theta - \mu_0)} \right]$$

$$\begin{aligned} \text{Now} \quad & \int_0^{2\pi} \sin(\theta - \mu_0) d\tilde{F}(\theta - \mu_0) \\ &= \int_0^{2\pi} \sin(\theta - \mu_0) d[(1 - \epsilon)\tilde{F}(\theta - \mu_0) + \epsilon\delta_{\theta_0}(\theta - \mu_0)] \\ &= (1 - \epsilon) \int_0^{2\pi} \sin(\theta - \mu_0) d\tilde{F}(\theta - \mu_0) + \epsilon \int_0^{2\pi} \sin(\theta - \mu_0) d\delta_{\theta_0}(\theta - \mu_0) \\ &= (1 - \epsilon)E[\sin(\theta - \mu_0)] + \epsilon \sin(\theta - \mu_0) \\ &= \epsilon \sin(\theta - \mu_0), \quad \text{since } E[\sin(\theta - \mu_0)] = 0, \text{ Mardia(1972, p. 45)}. \end{aligned}$$

$$\text{Next} \quad \int_0^{2\pi} \cos(\theta - \mu_0) d\tilde{F}(\theta - \mu_0)$$

$$\begin{aligned}
&= \int_0^{2\pi} \cos(\theta - \mu_0) d[(1 - \epsilon)\tilde{F}(\theta - \mu_0) + \epsilon\delta_{\theta_0}(\theta - \mu_0)] \\
&= (1 - \epsilon) \int_0^{2\pi} \cos(\theta - \mu_0) d\tilde{F}(\theta - \mu_0) + \epsilon \int_0^{2\pi} \cos(\theta - \mu_0) d\delta_{\theta_0}(\theta - \mu_0) \\
&= (1 - \epsilon)E[\cos(\theta - \mu_0)] + \epsilon \cos(\theta - \mu_0) \\
&= (1 - \epsilon)\rho + \epsilon \cos(\theta - \mu_0), \quad \text{since } E[\cos(\theta - \mu_0)] = \rho, \text{ Mardia(1972, p. 45)}.
\end{aligned}$$

$$\text{Hence } T(\tilde{F}) = \tan^{-1} \left[\frac{\epsilon \sin(\theta - \mu_0)}{(1 - \epsilon)\rho + \epsilon \cos(\theta - \mu_0)} \right].$$

$$\text{Thus } \left. \frac{T(\tilde{F})}{d\epsilon} \right|_{\epsilon=0} = \frac{-\rho \sin(\theta - \mu_0)}{\left[\frac{\epsilon \sin(\theta - \mu_0)}{(1 - \epsilon)\rho + \epsilon \cos(\theta - \mu_0)} \right]^2 + [(1 - \epsilon)\rho + \epsilon \cos(\theta - \mu_0)]^2}$$

$$\text{Therefore } \left. \frac{T(\tilde{F})}{d\epsilon} \right|_{\epsilon=0} = \left[\frac{\rho \sin(\theta - \mu_0)}{\rho^2} \right] = \frac{\sin(\theta - \mu_0)}{\rho}.$$

B.2 Theorem 3.7.2:

Without loss of generality for notational simplicity, assume that $\mu \in [0, \pi]$. The influence function for the circular median direction is given by

$$IF(\theta) = \frac{\frac{1}{2} \text{sgn}(\theta - \mu_0)}{[f(\mu_0) - f(\mu_0 + \pi)]}, \quad (\mu_0 - \pi < \theta < \mu_0 + \pi), \quad (2)$$

where $\text{sgn}(x) = 1, 0, \text{ or } -1$ as $x > 0, x = 0, \text{ or } x < 0$, respectively.

Proof of Theorem 3.7.2

Since $T_{\frac{1}{2}}(F) = F^{-1}(\frac{1}{2}) = \theta_{\frac{1}{2}}$, it follows that $F(T_{\frac{1}{2}}(F)) = F(F^{-1}(\frac{1}{2})) = F(\theta_{\frac{1}{2}}) = \frac{1}{2}$.

Expressing the median of the altered cdf \tilde{F} , we have

$$\begin{aligned}
L &= \tilde{F}\left(\tilde{F}^{-1}\left(\frac{1}{2}\right)\right) = \tilde{F}\left(\theta_{\frac{1}{2}}\right) \\
&= (1 - \epsilon)F\left(\tilde{\theta}_{\frac{1}{2}}\right) + \epsilon\delta_{\theta_0}\left(\tilde{\theta}_{\frac{1}{2}}\right) = \frac{1}{2}
\end{aligned}$$

Differentiating both sides of L above with respect to ϵ , we obtain

$$\frac{dL}{d\epsilon} = (-1)F(\tilde{\theta}_{\frac{1}{2}}) + (1 - \epsilon)\frac{d\tilde{\theta}_{\frac{1}{2}}}{d\epsilon}f(\tilde{\theta}_{\frac{1}{2}}) + \delta_{\theta_0}(\tilde{\theta}_{\frac{1}{2}}) + \epsilon\frac{d\delta_{\theta_0}(\tilde{\theta}_{\frac{1}{2}})}{d\epsilon} = 0$$

Setting $\epsilon = 0$, we obtain

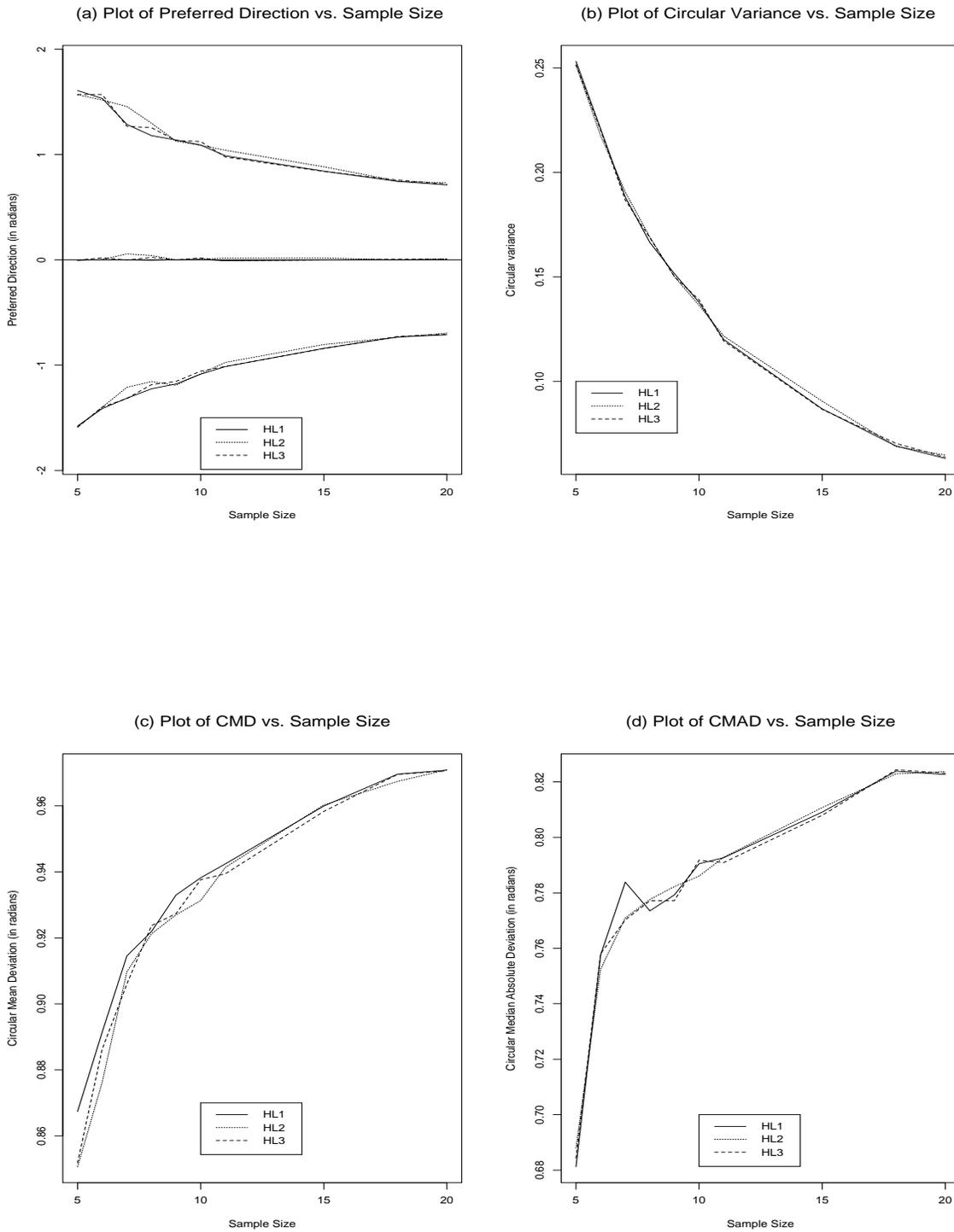
$$\frac{d(\tilde{\theta}_{\frac{1}{2}})}{d\epsilon} = \frac{\frac{1}{2} - \delta_{\theta_0}(\tilde{\theta}_{\frac{1}{2}})}{f(\tilde{\theta}_{\frac{1}{2}})}$$

$$\begin{aligned} \text{That is } \quad \text{IF}(\theta_0, \theta_{\frac{1}{2}}) &= \frac{\frac{1}{2} - \delta_{\theta_0}(\tilde{\theta}_{\frac{1}{2}})}{f(\tilde{\theta}_{\frac{1}{2}})} \\ &= \begin{cases} \frac{-1}{2f(\theta_{\frac{1}{2}})}, & \text{for } \theta_0 \leq \theta_{\frac{1}{2}} \\ \frac{1}{2f(\theta_{\frac{1}{2}})}, & \text{for } \theta_0 \geq \theta_{\frac{1}{2}} \end{cases} \end{aligned}$$

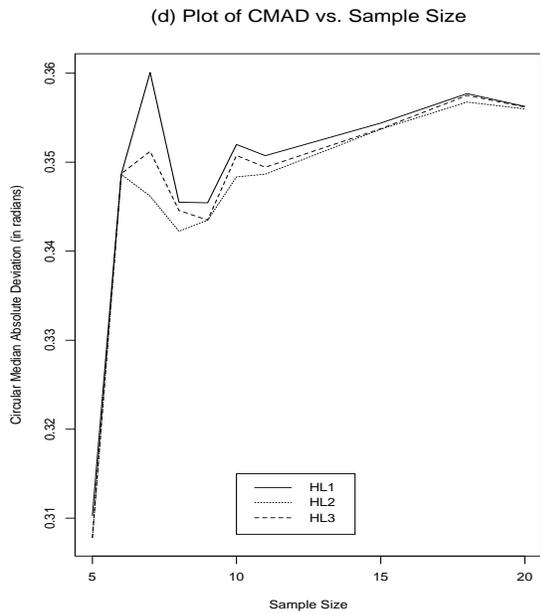
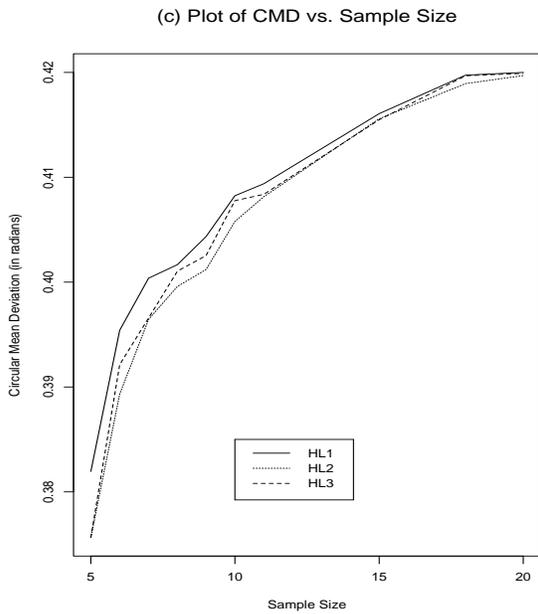
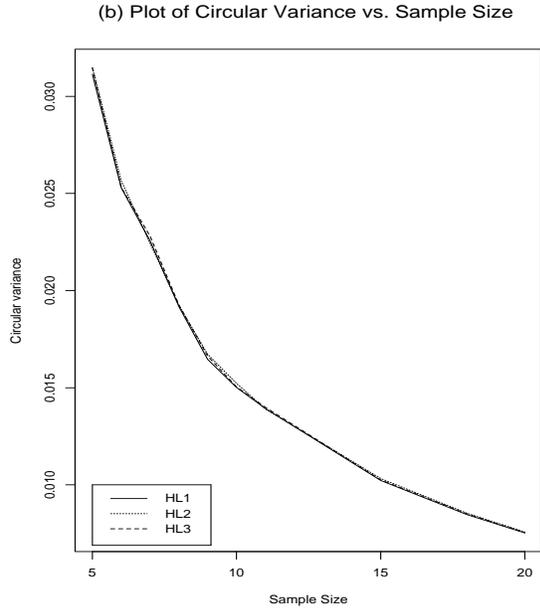
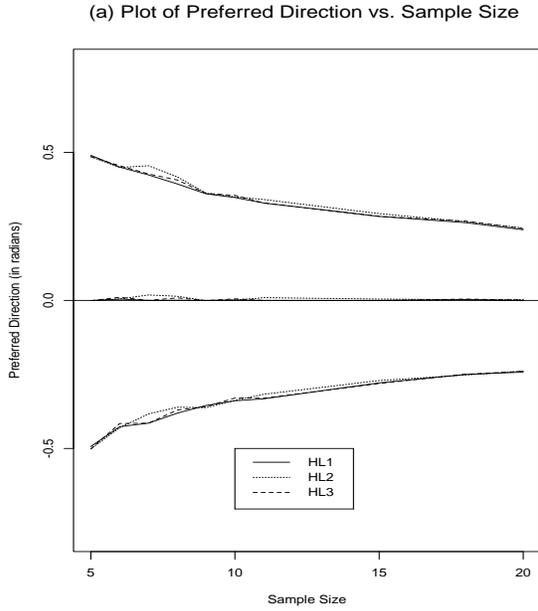
$$\text{since } \quad \delta_{\theta_0}(\theta_{\frac{1}{2}}) = \begin{cases} 0, & \theta_{\frac{1}{2}} < \theta_0 \\ 1, & \theta_{\frac{1}{2}} \geq \theta_0 \end{cases}.$$

Appendix C : HL1, HL2 and HL3

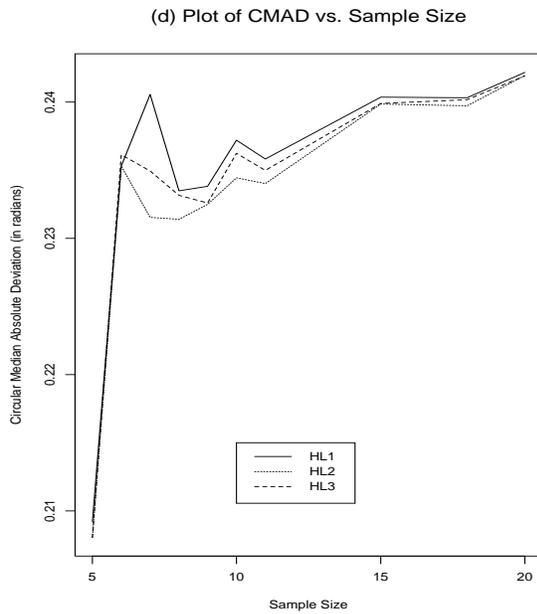
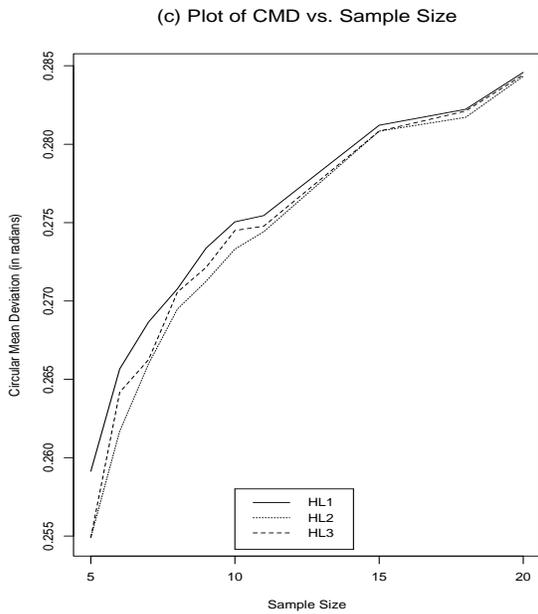
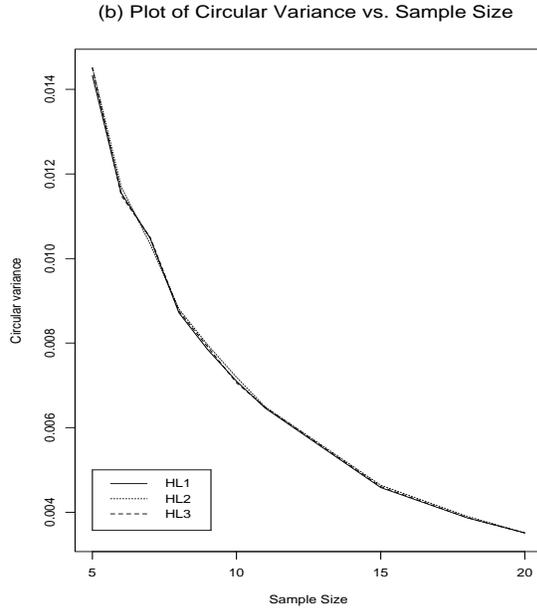
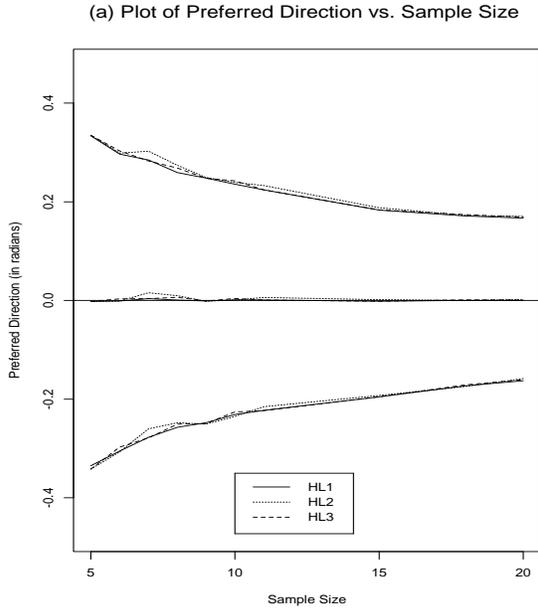
HL1, HL2 & HL3 for VM(0, 1)



HL1, HL2 & HL3 for VM(0, 4)

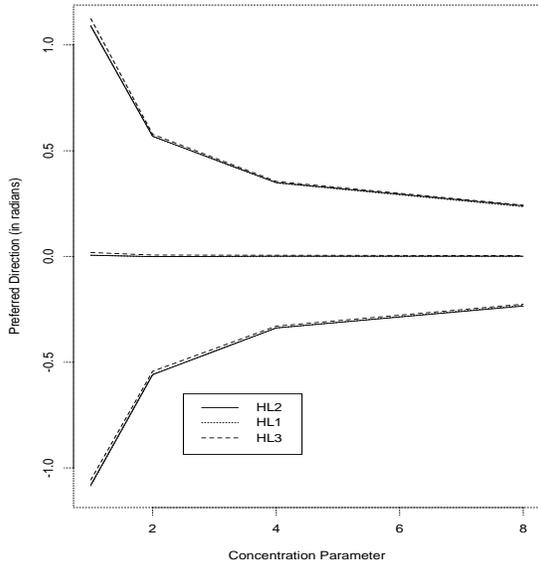


HL1, HL2 & HL3 for VM(0, 8)

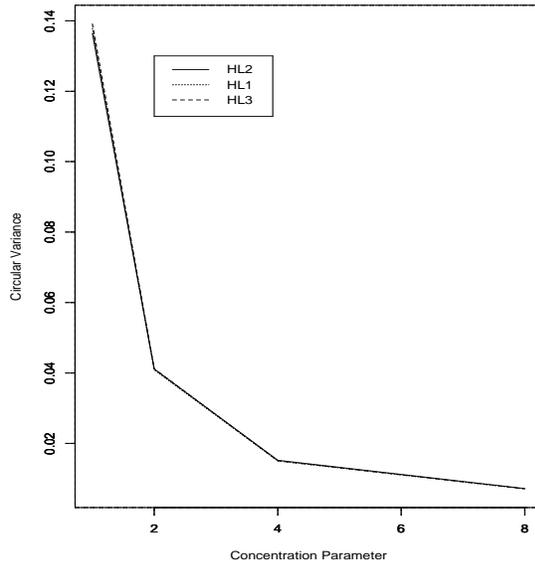


HL1, HL2 & HL3 for $VM(0, \kappa)$, $n = 10$

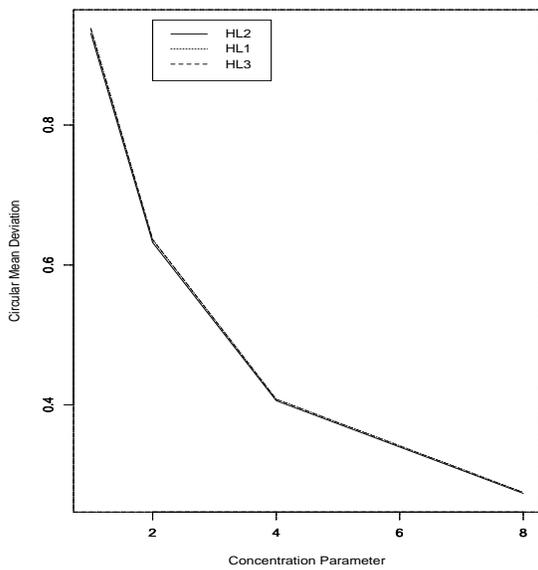
(a) Point Estimate and 95% C.I. vs. Conc. Parameter



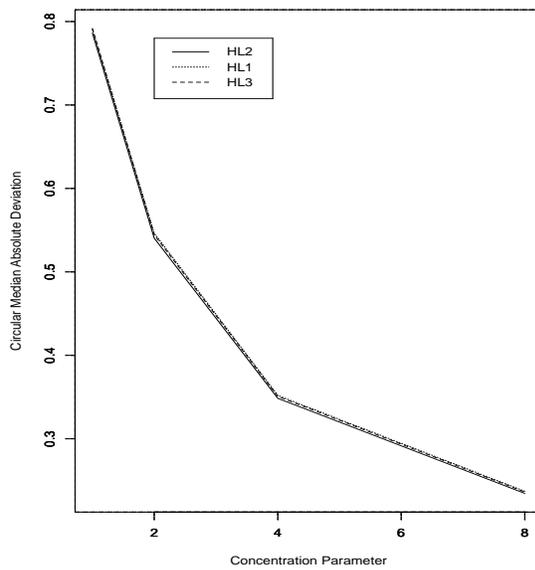
(b) Circular Variance vs. Conc. Parameter



(c) Circular Mean Dev. vs. Conc. Parameter



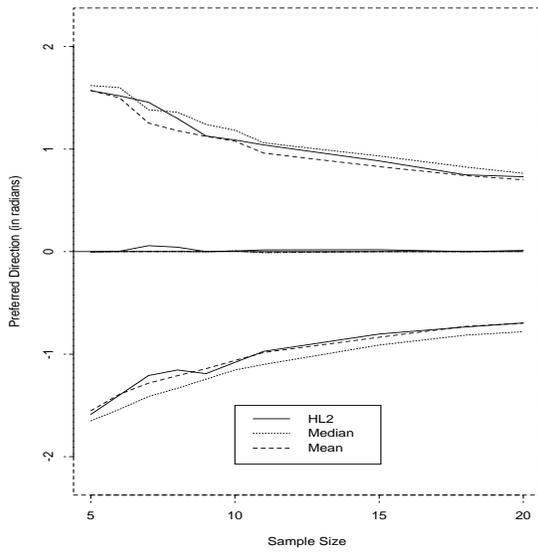
(d) Circular Median Absolute Dev. vs. Conc. Parameter



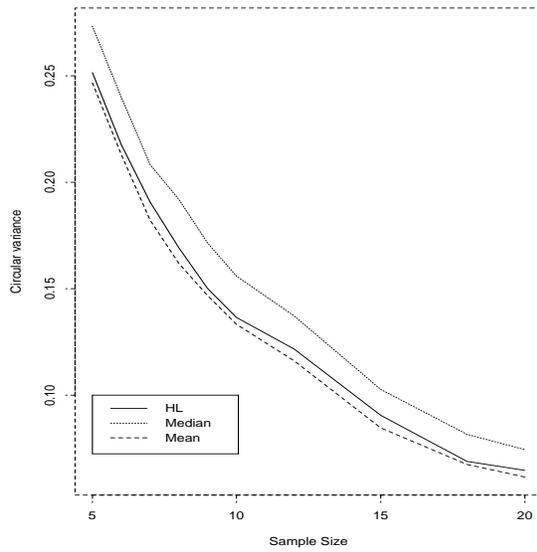
Appendix D : Mean, Median and HL

Mean, Median and HL for VM(0,1)

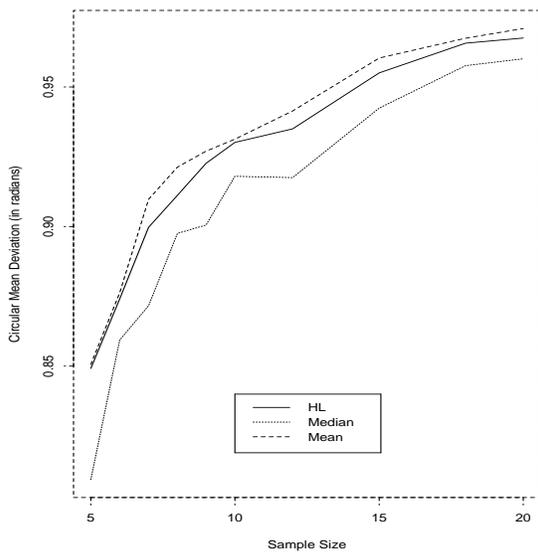
Plot of Preferred Direction vs. Sample Size



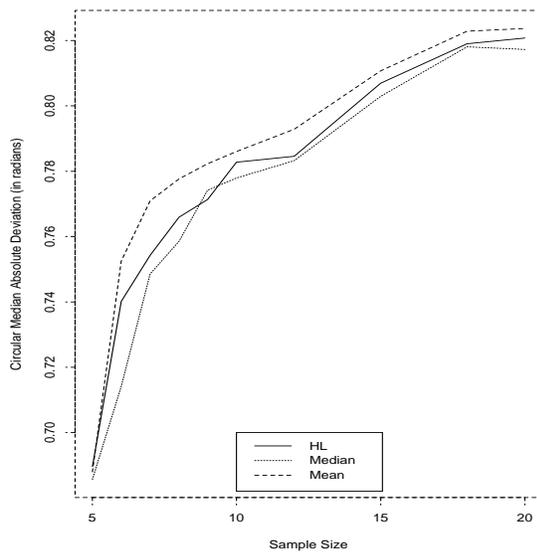
(b) Plot of Circular Variance Vs. Sample Size



(c) Plot of CMD vs. Sample Size

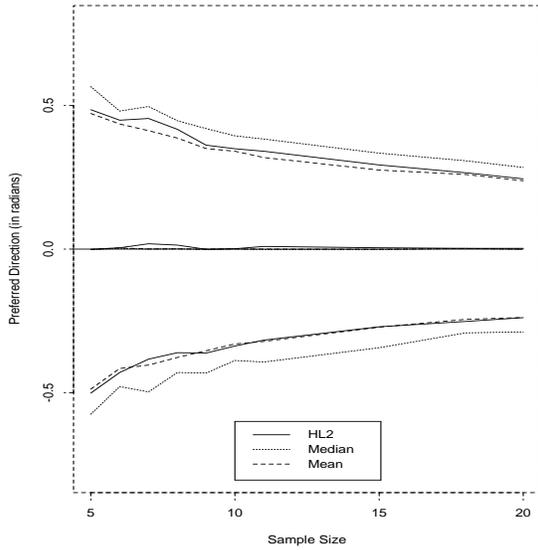


(d) Plot of CMAD vs. Sample Size

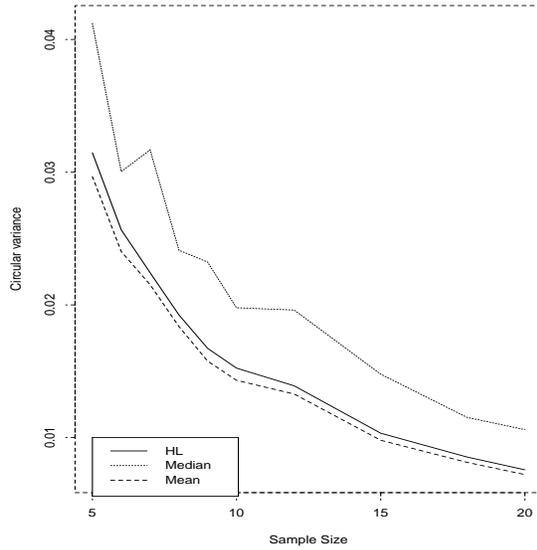


Mean, Median and HL for VM(0,4)

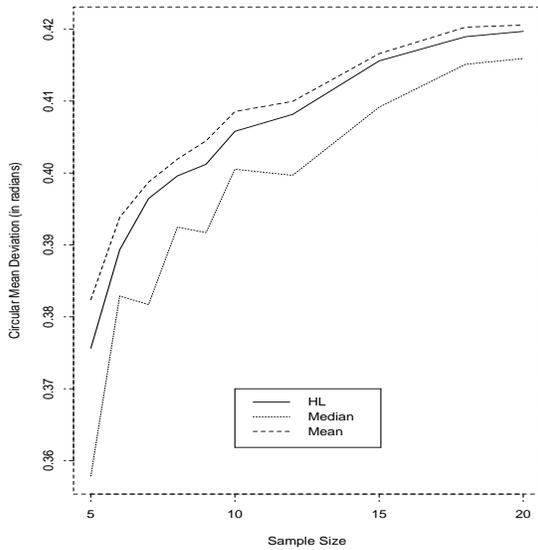
Plot of Preferred Direction vs. Sample Size



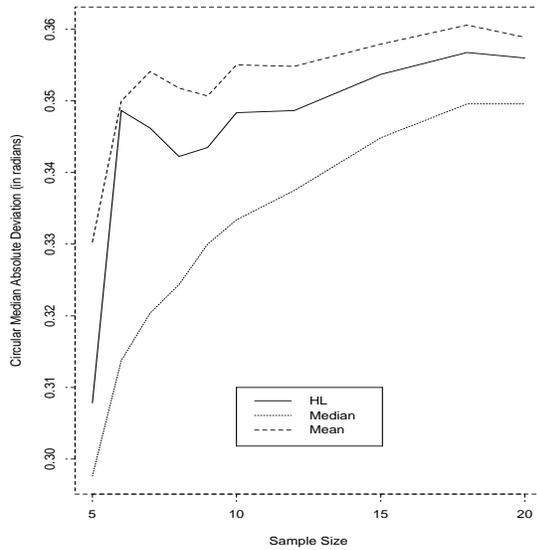
(b) Plot of Circular Variance Vs. Sample Size



(c) Plot of CMD vs. Sample Size

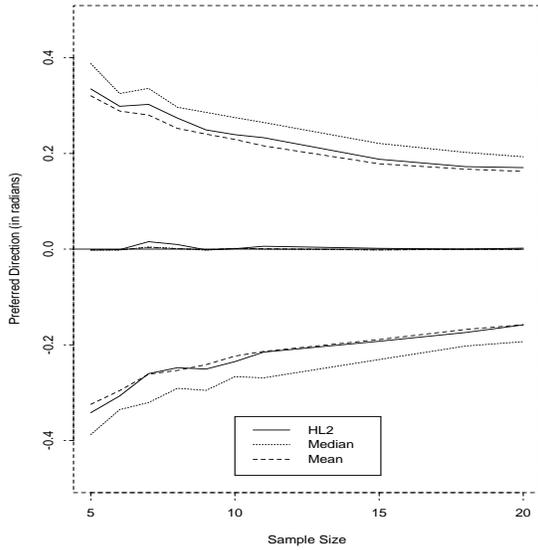


(d) Plot of CMAD vs. Sample Size

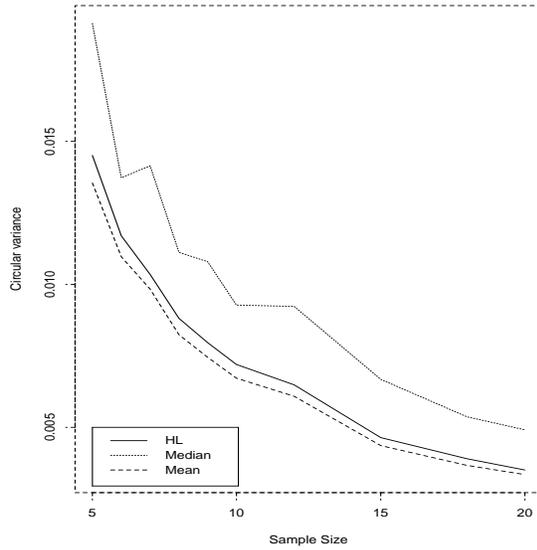


Mean, Median and HL for VM(0,8)

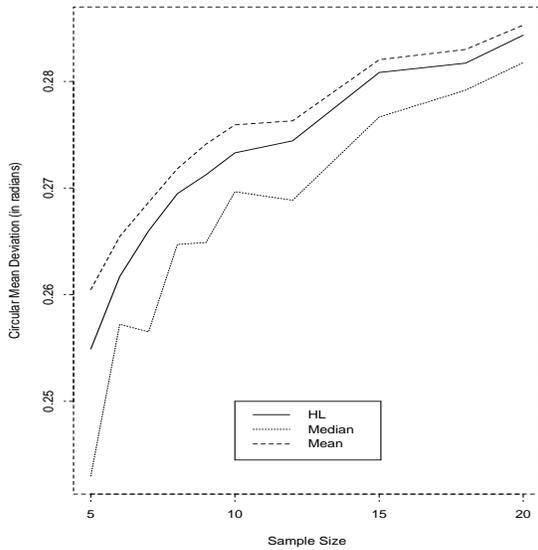
Plot of Preferred Direction vs. Sample Size



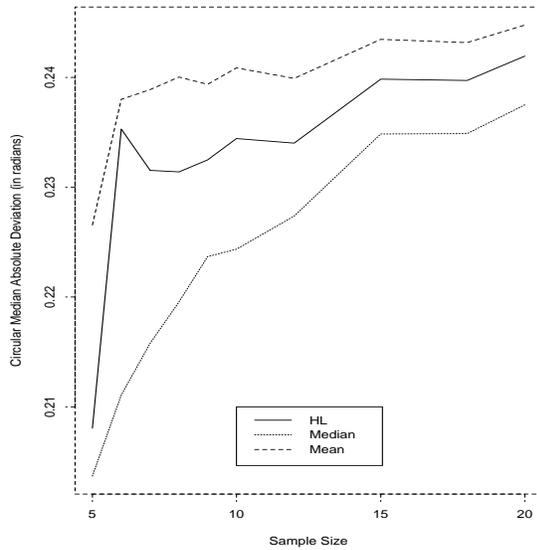
(b) Plot of Circular Variance Vs. Sample Size



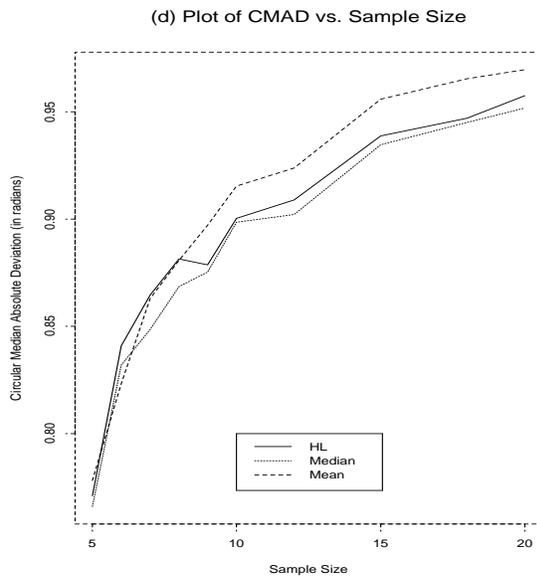
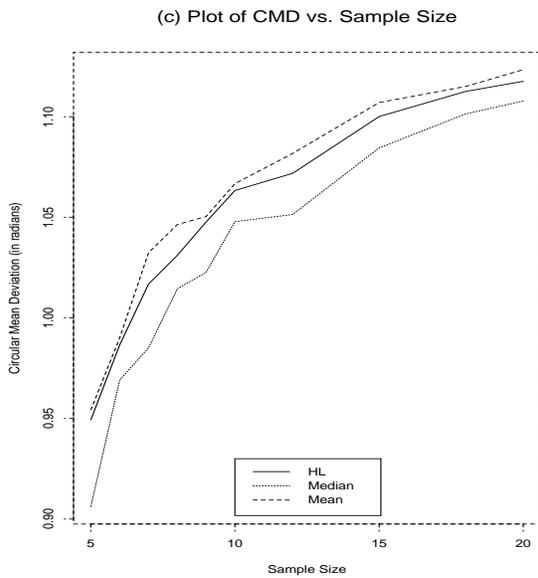
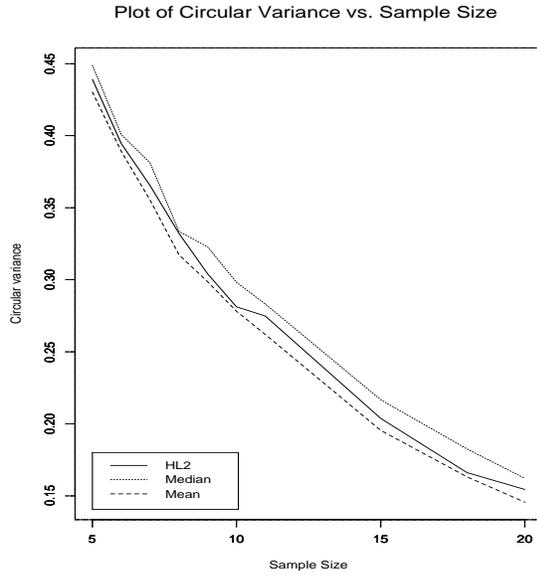
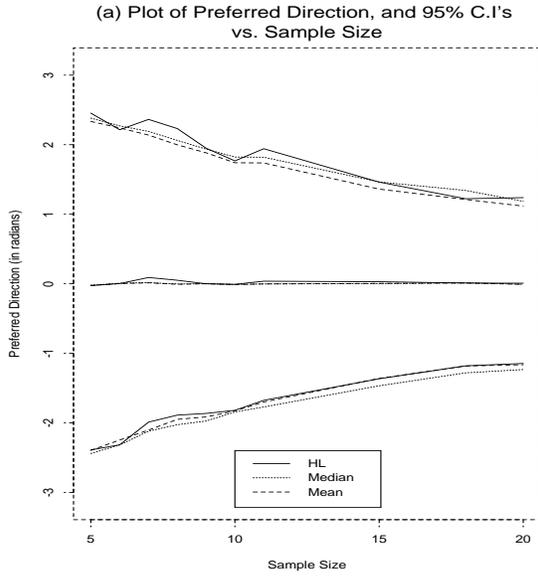
(c) Plot of CMD vs. Sample Size



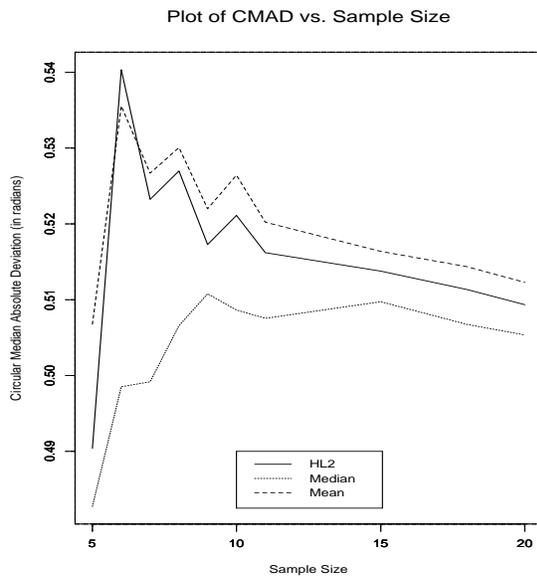
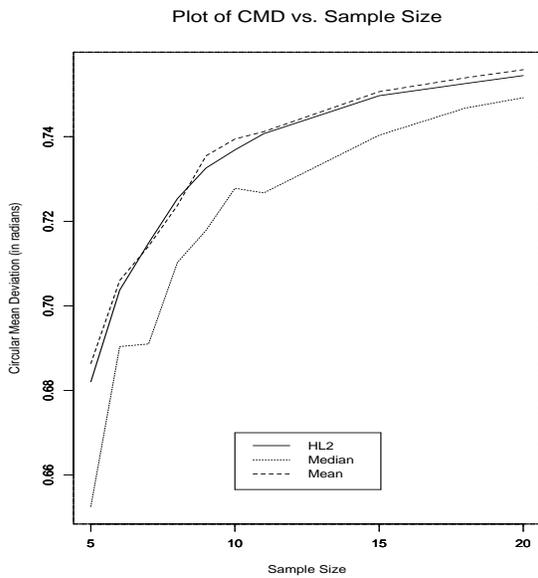
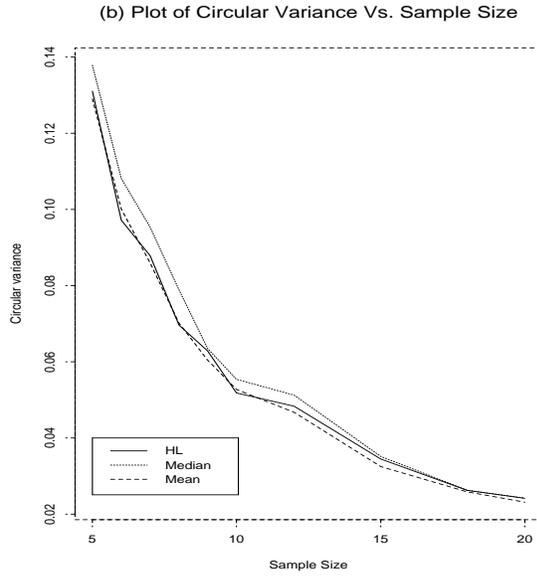
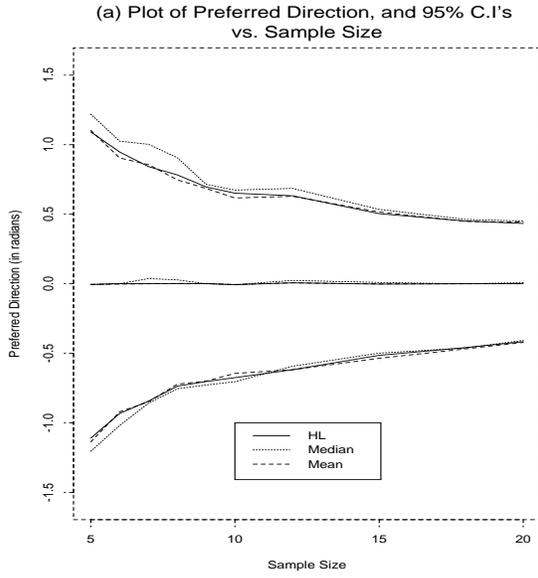
(d) Plot of CMAD vs. Sample Size



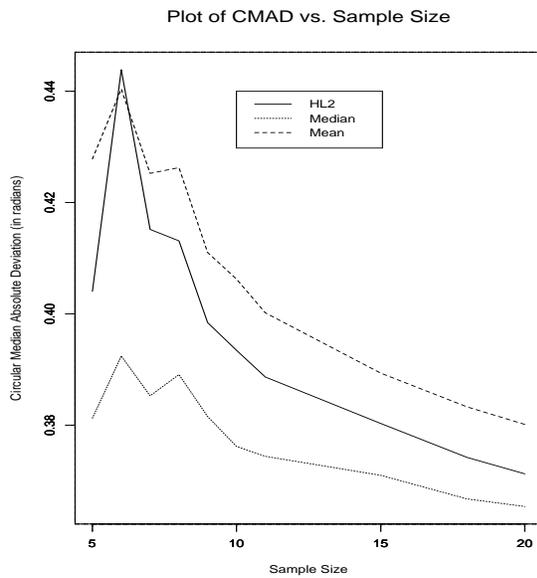
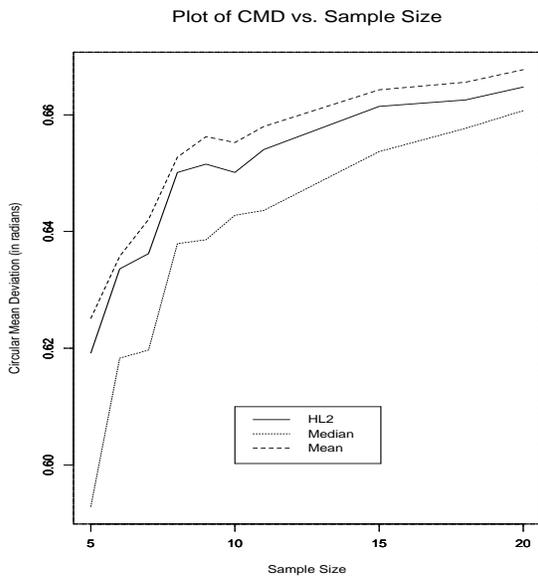
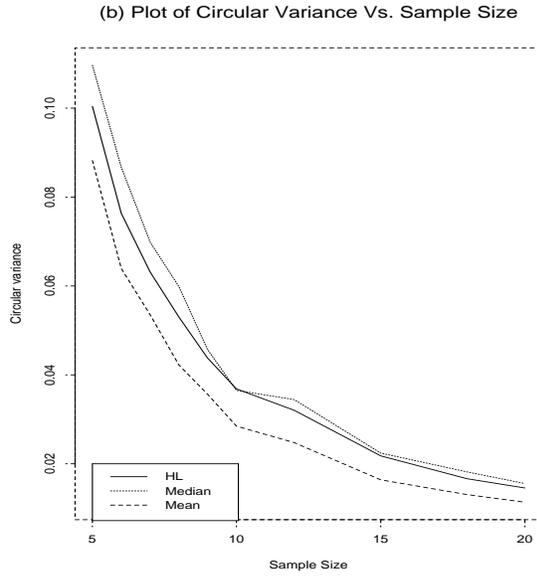
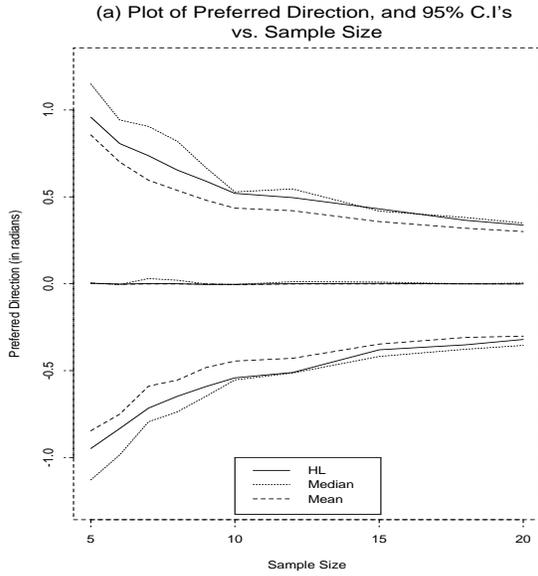
Mean, Median and HL for 70%VM(0,1) & 30% Uniform



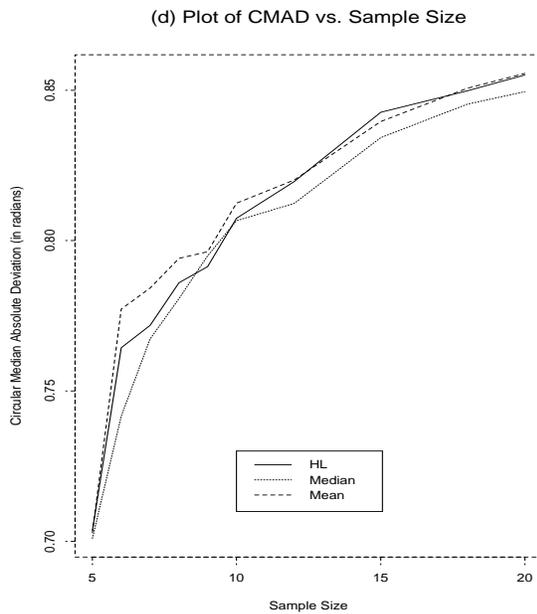
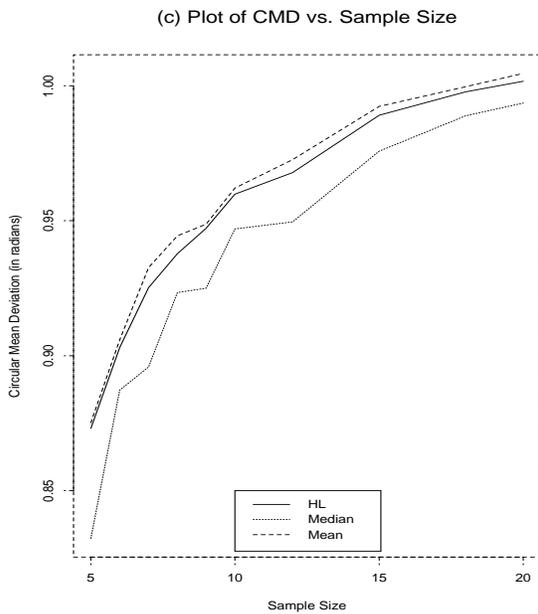
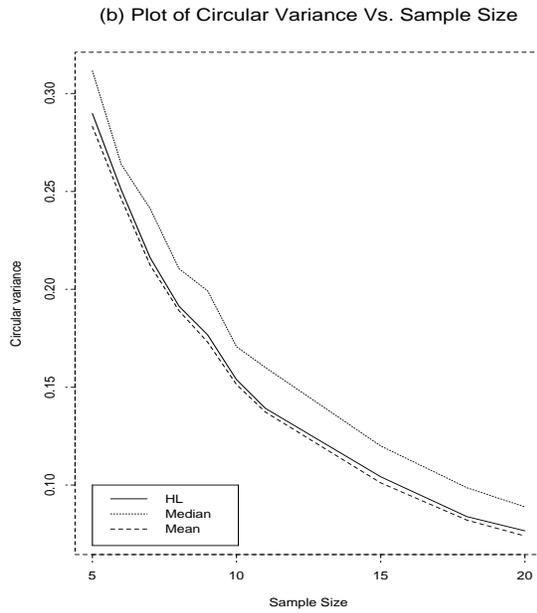
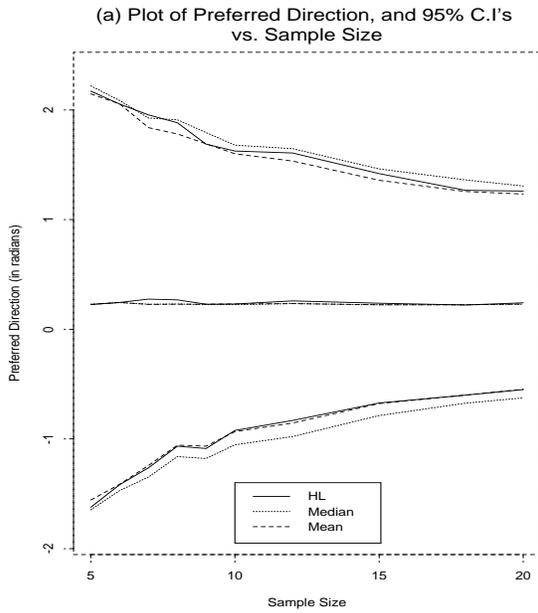
Mean, Median and HL for 70%VM(0,4) & 30% Uniform



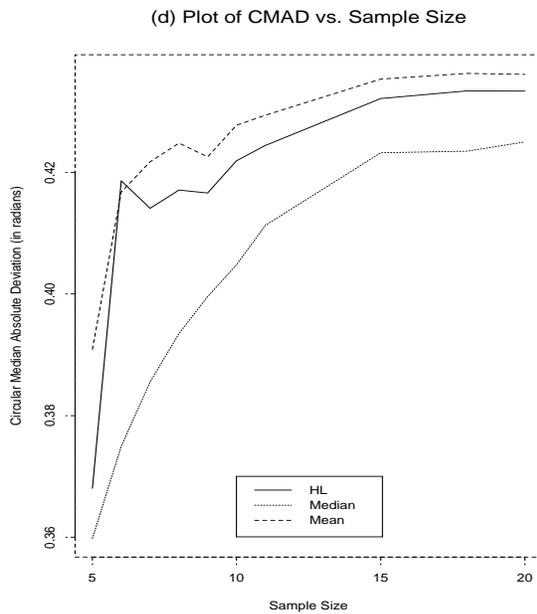
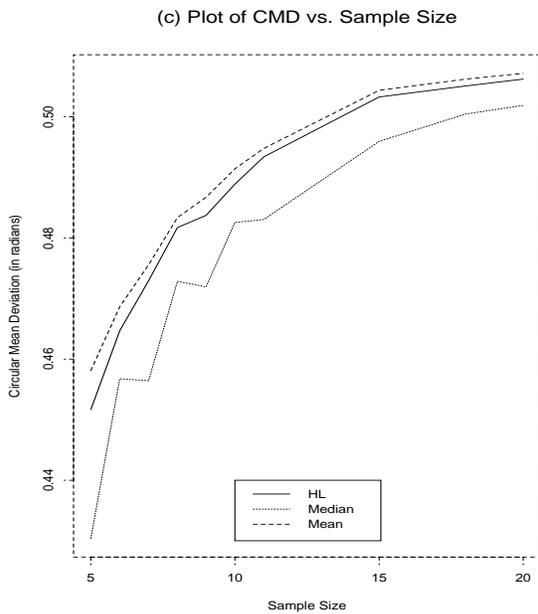
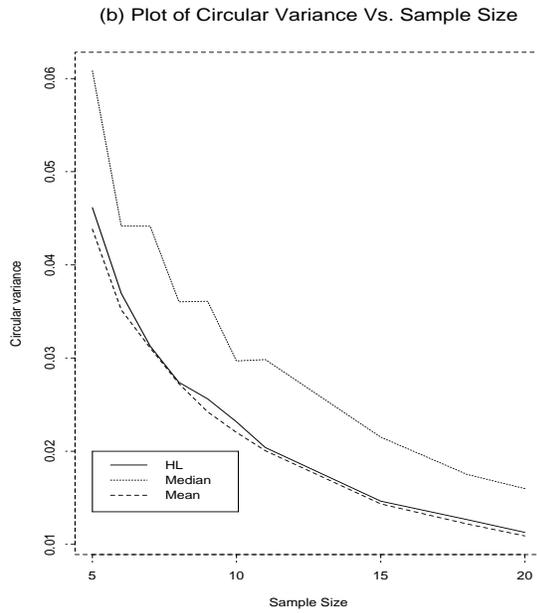
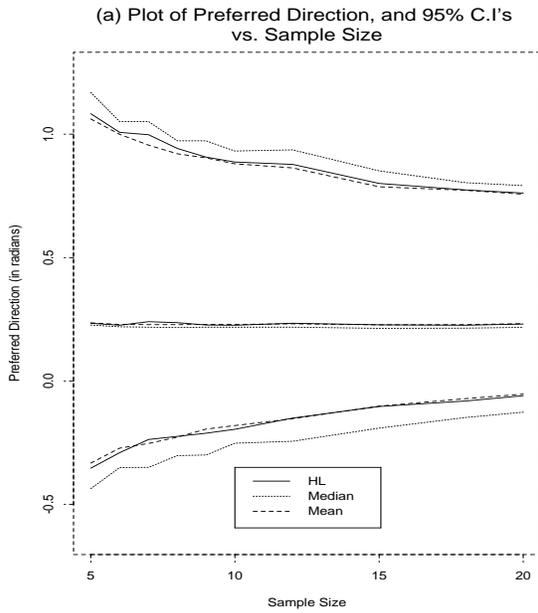
Mean, Median and HL for 70%VM(0, 8) & 30% Uniform



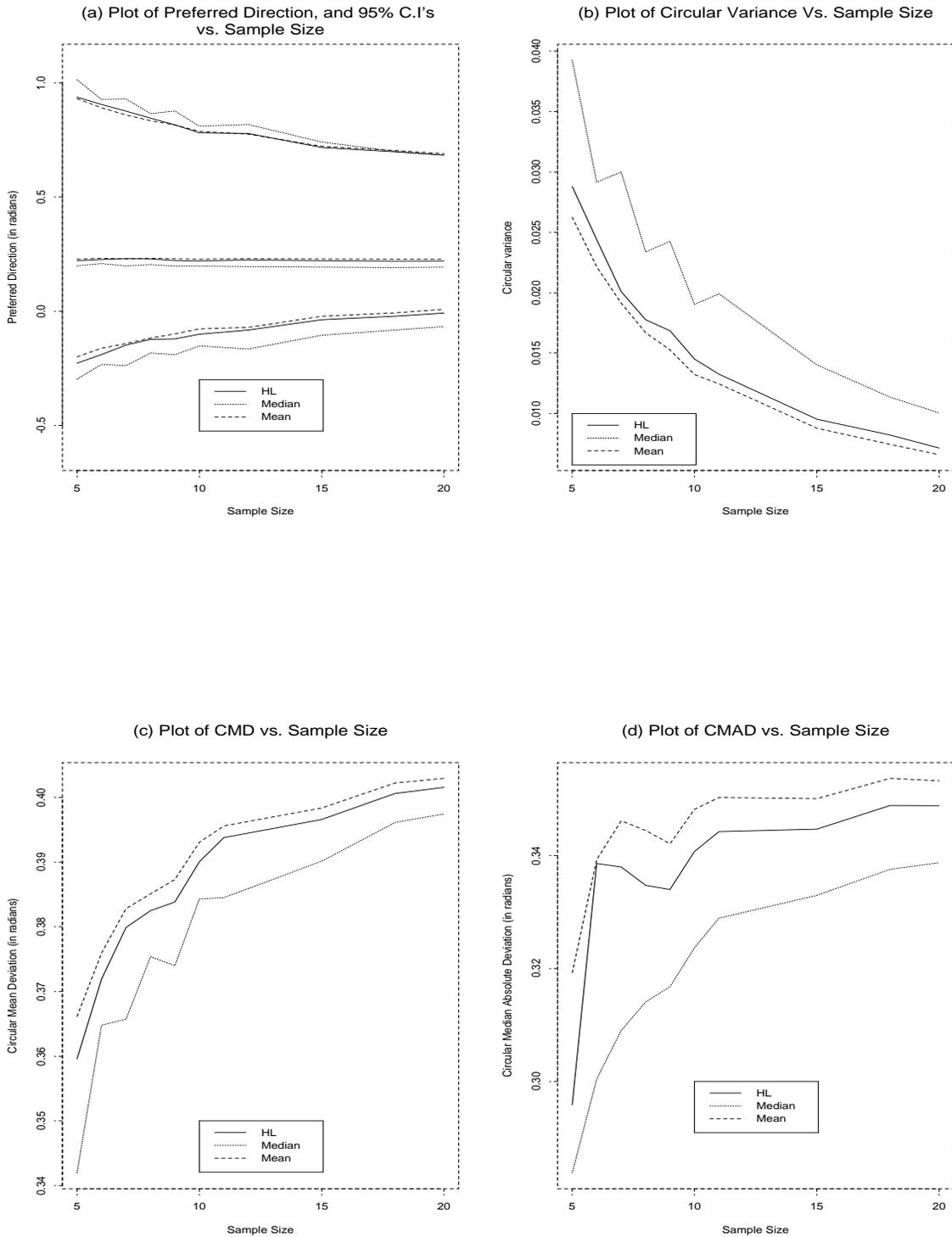
Mean, Median and HL for VM (0, 1), for $\epsilon = 0.3$, $\mu_1 = 0$, $\mu_2 = \frac{\pi}{4}$



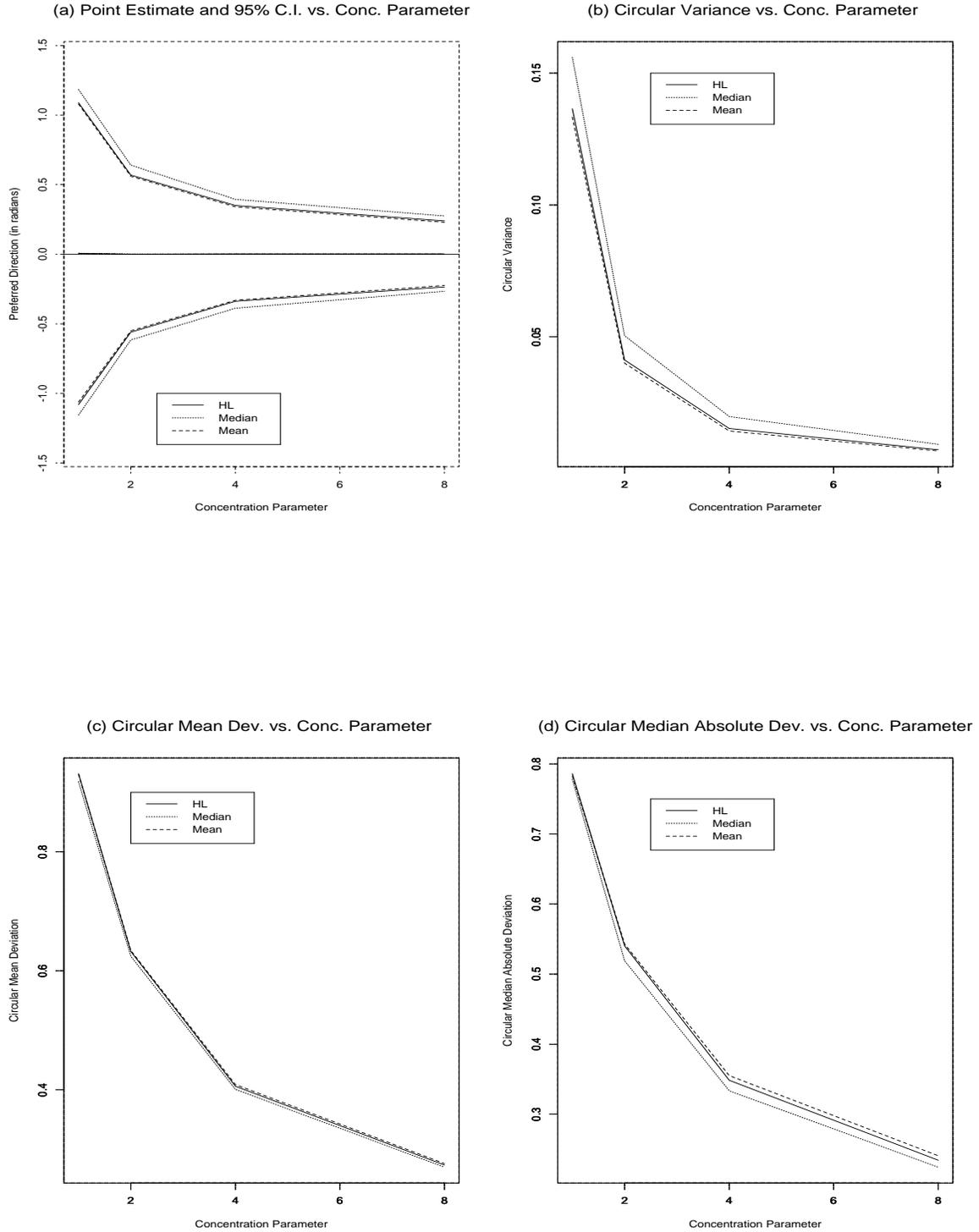
Mean, Median and HL for VM (0, 4), for $\epsilon = 0.3$, $\mu_1 = 0$, $\mu_2 = \frac{\pi}{4}$



Mean, Median and HL for VM (0, 8), for $\epsilon = 0.3$, $\mu_1 = 0$, $\mu_2 = \frac{\pi}{4}$

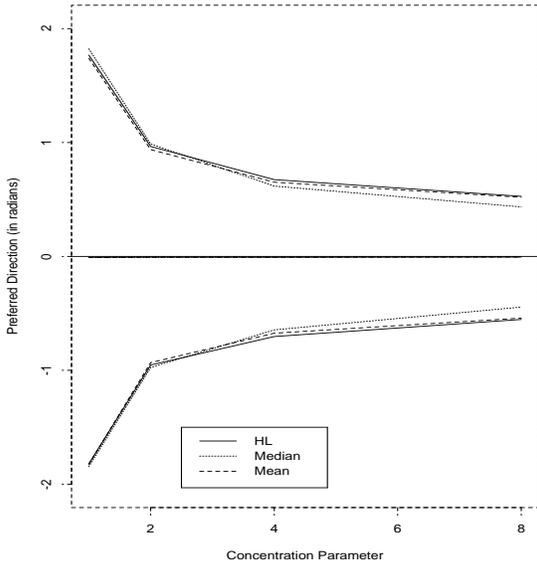


Effect of κ on Mean, Median and HL for VM $(0, \kappa)$, $n = 10$

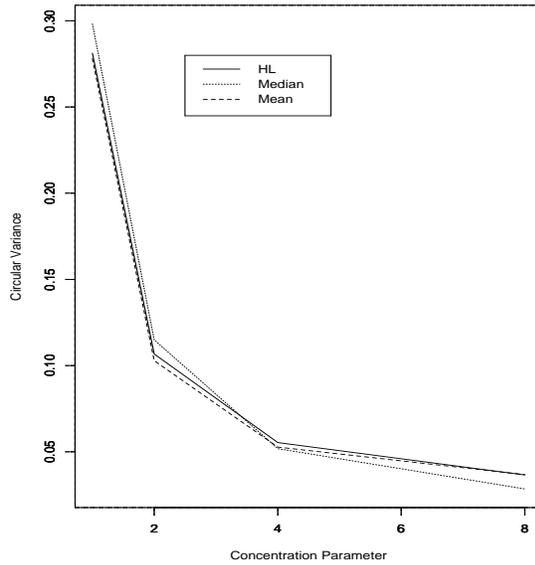


Mean, Median and HL for 70%VM $(0, \kappa)$ & 30%Uniform, $\epsilon = 0.3, n = 10$

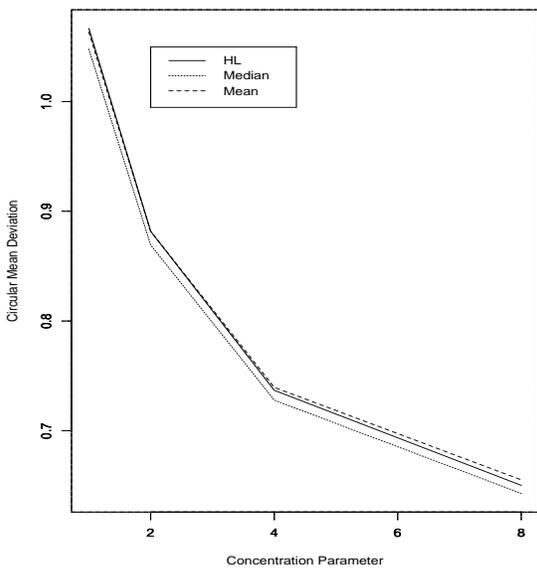
(a) Point Estimate and 95% C.I. vs. Conc. Parameter



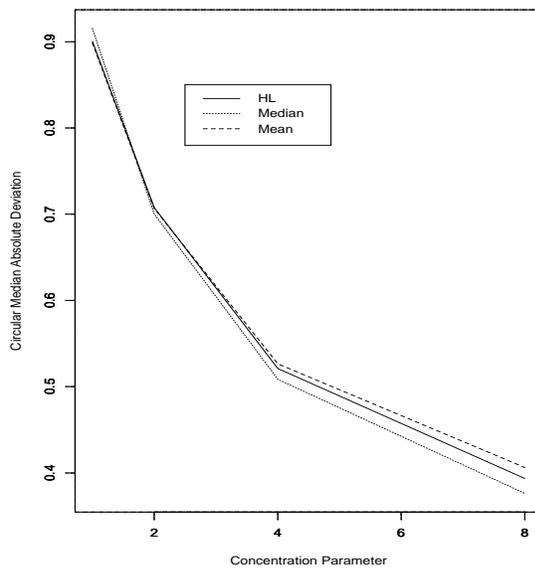
(b) Circular Variance vs. Conc. Parameter



(c) Circular Mean Dev. vs. Conc. Parameter

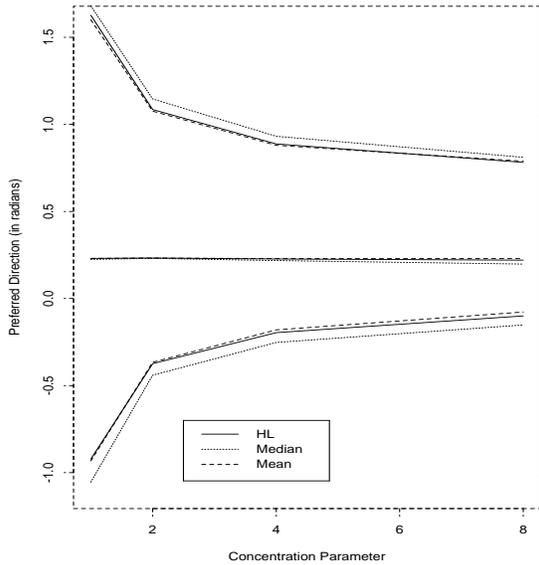


(d) Circular Median Absolute Dev. vs. Conc. Parameter

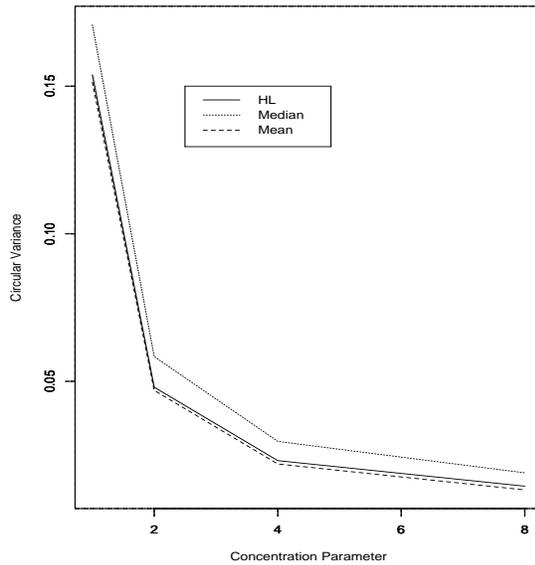


Mean, Median and HL for 70%VM(0, κ) & 30%VM($\frac{\pi}{4}$, κ), $n = 10$

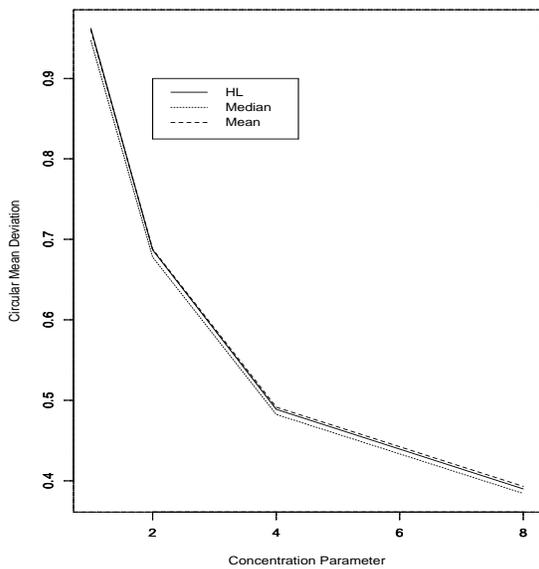
(a) Point Estimate and 95% C.I. vs. Conc. Parameter



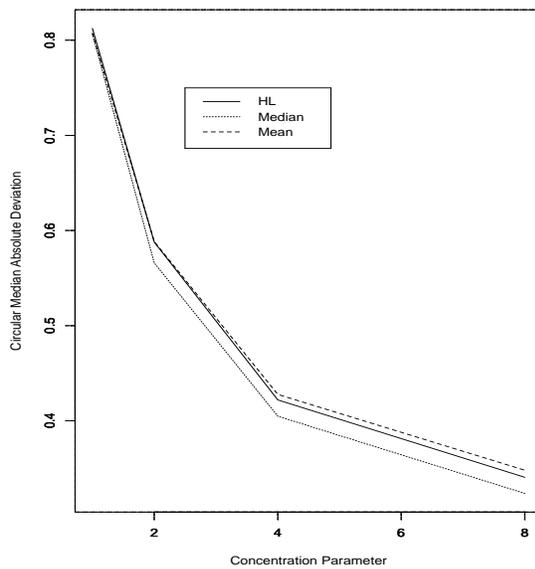
(b) Circular Variance vs. Conc. Parameter



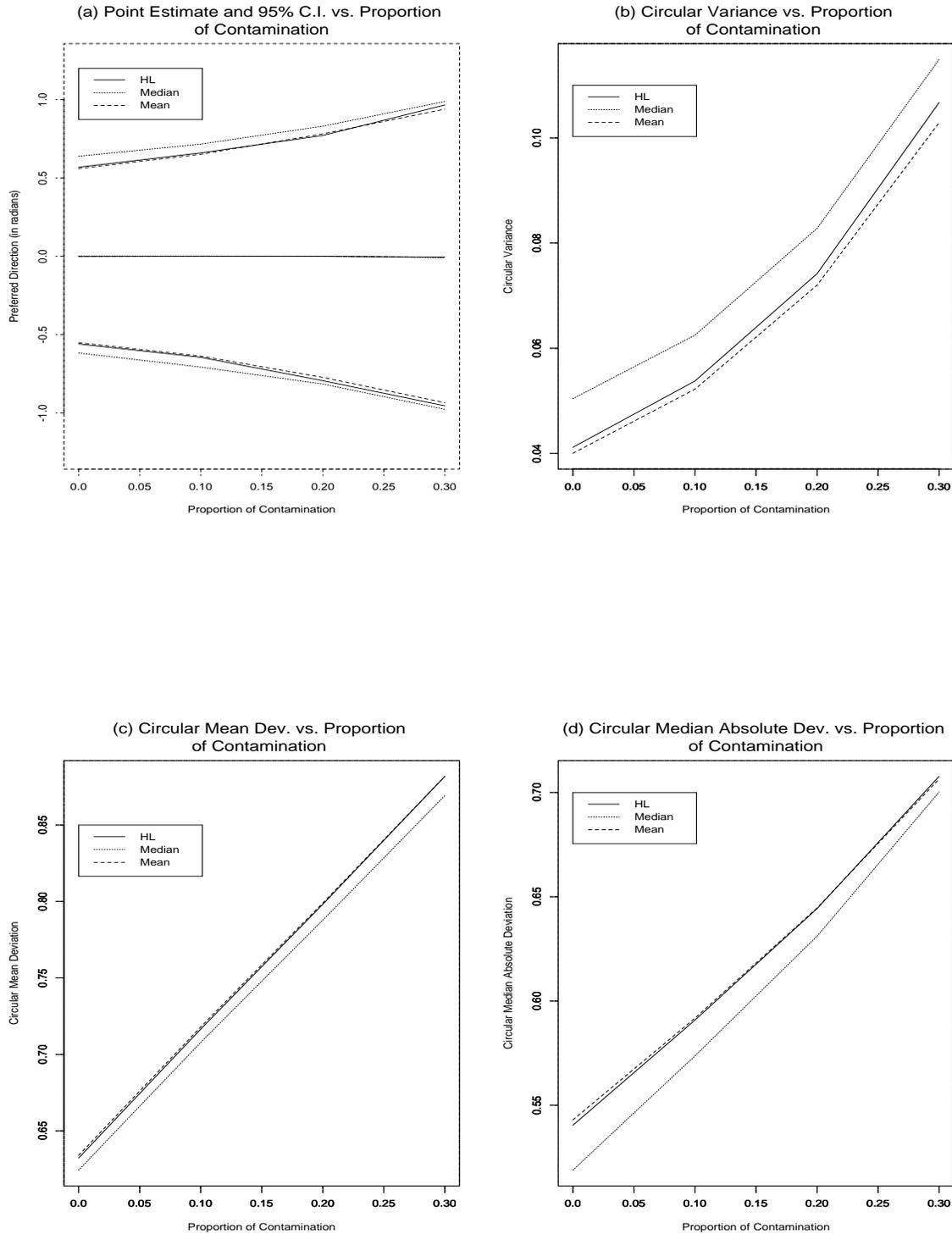
(c) Circular Mean Dev. vs. Conc. Parameter



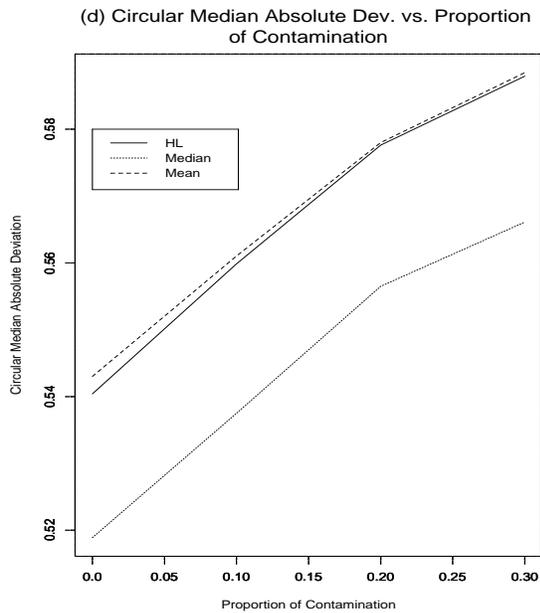
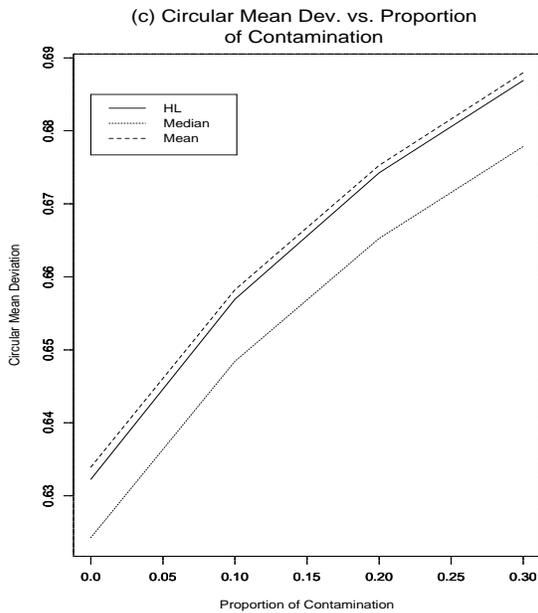
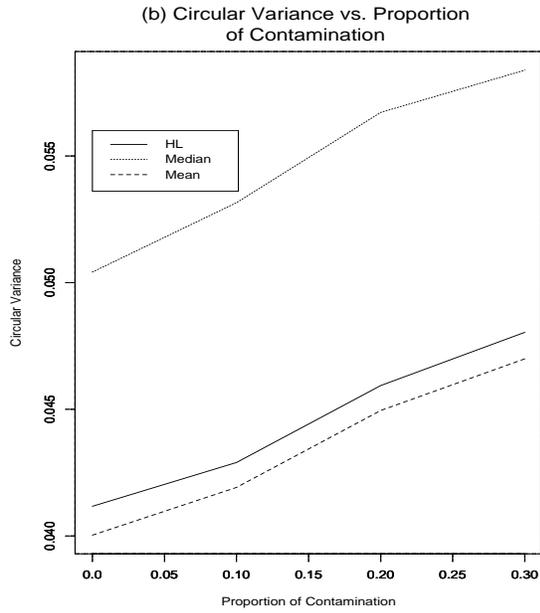
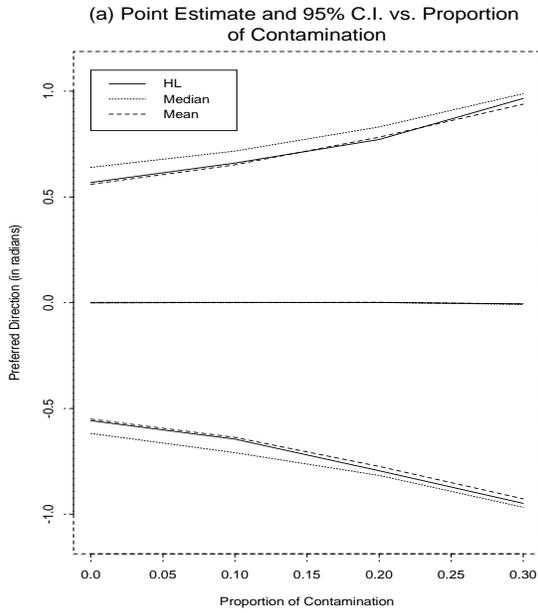
(d) Circular Median Absolute Dev. vs. Conc. Parameter



Effect of increasing the spread on Mean, Median and HL, for $\kappa = 2, n = 10$

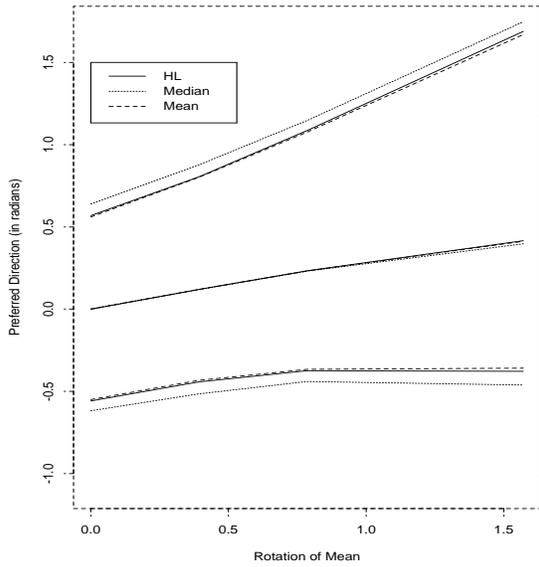


Effect of contamination level on Mean, Median and HL, n=10

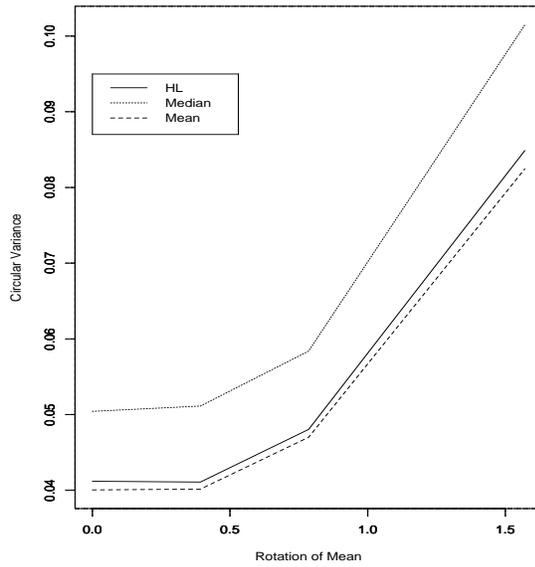


Effect of Rotating the mean direction from 0 to $\frac{\pi}{2}$ on Mean, Median and HL, n=10

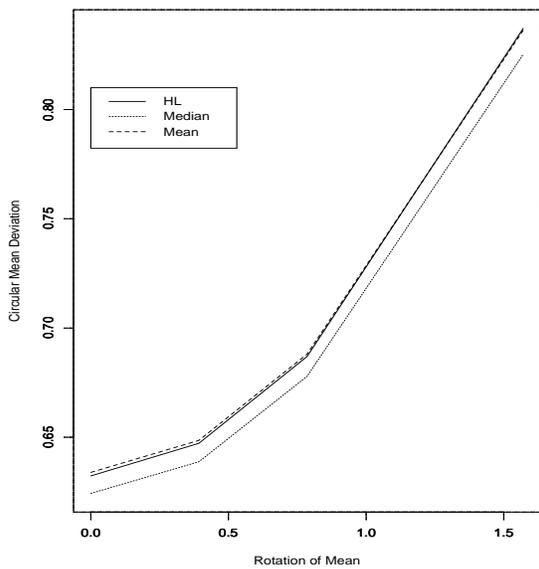
(a) Point Estimate and 95% C.I. vs. Rotation of Mean



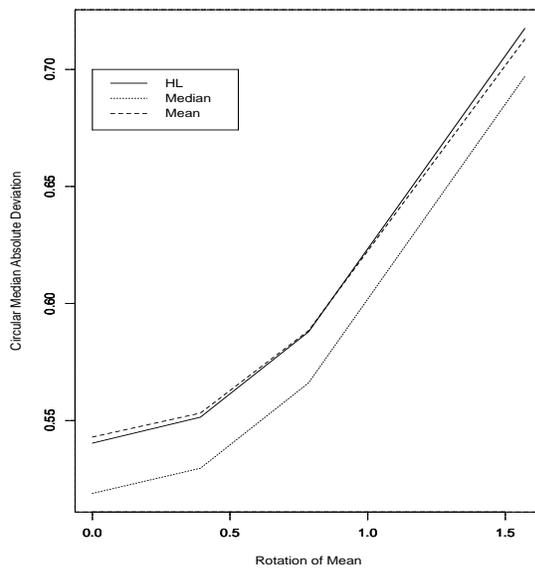
(b) Circular Variance vs. Rotation of Mean



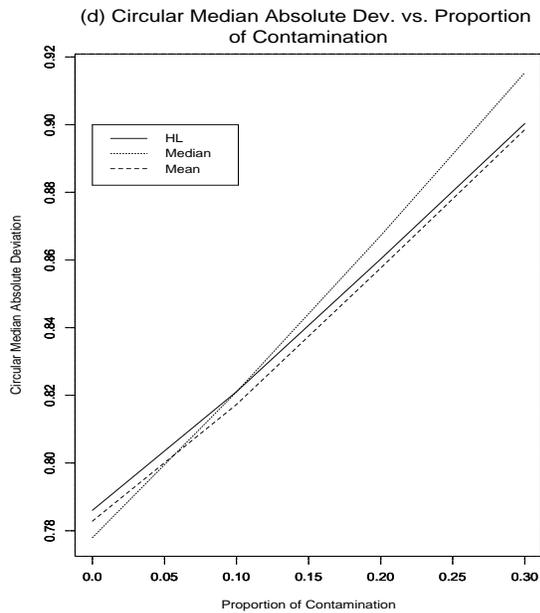
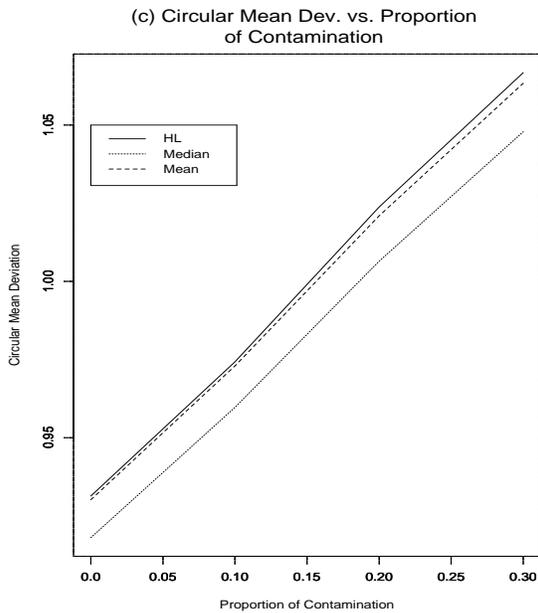
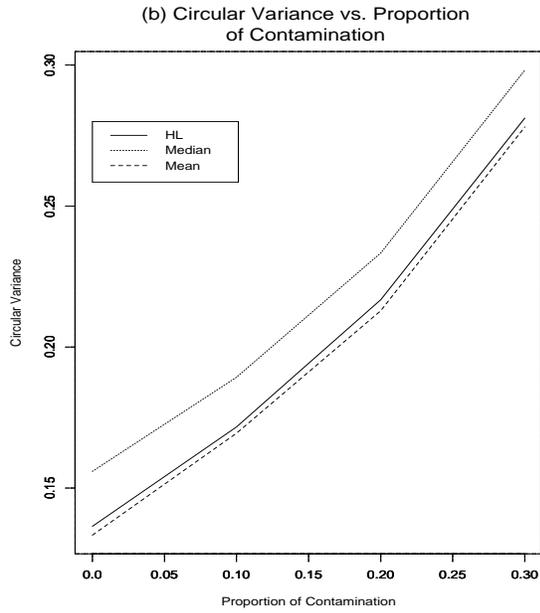
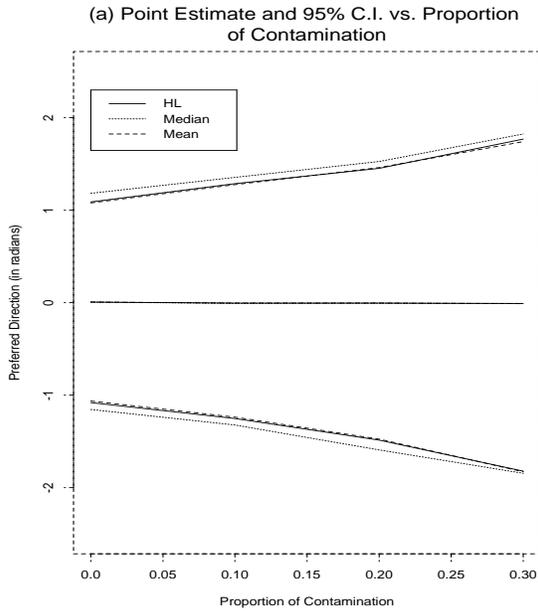
(c) Circular Mean Dev. vs. Rotation of Mean



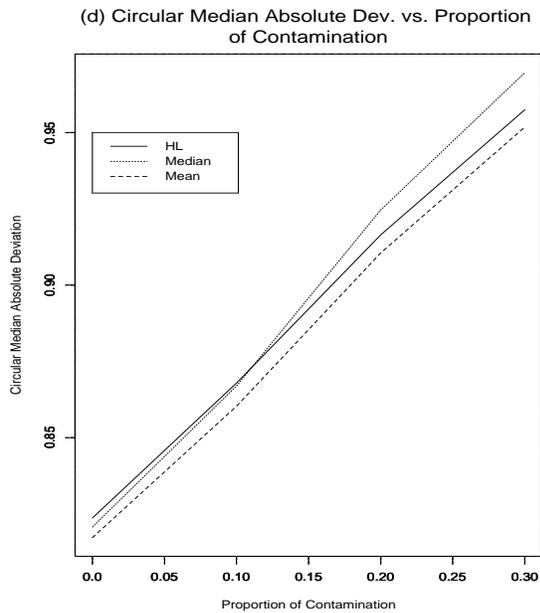
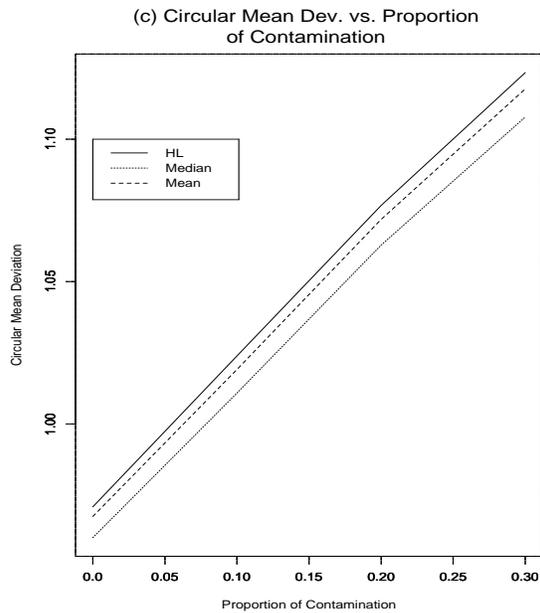
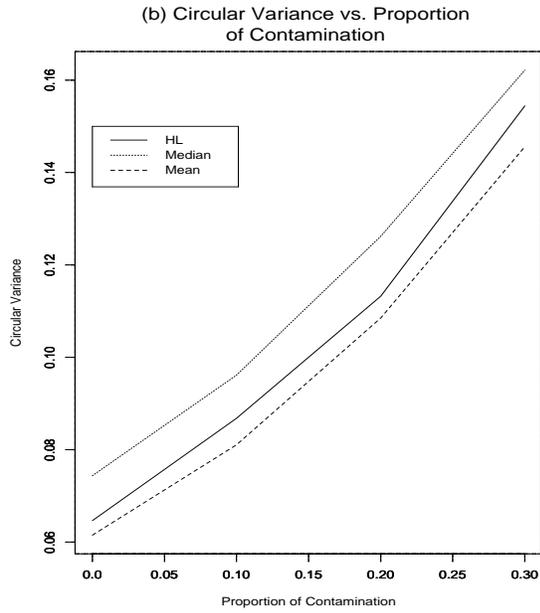
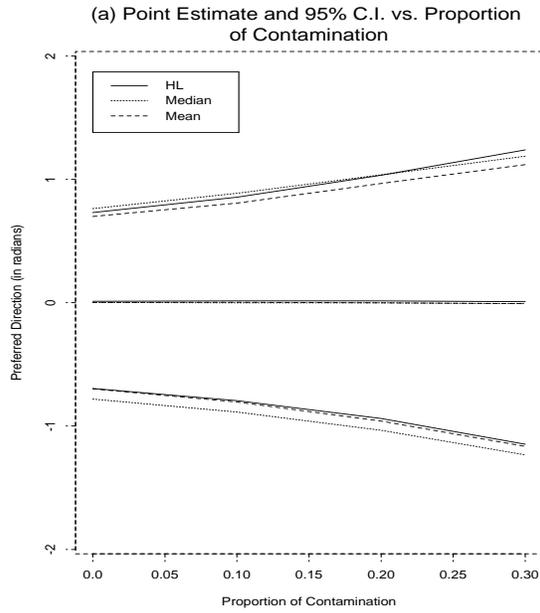
(d) Circular Median Absolute Dev. vs. Rotation of Mean



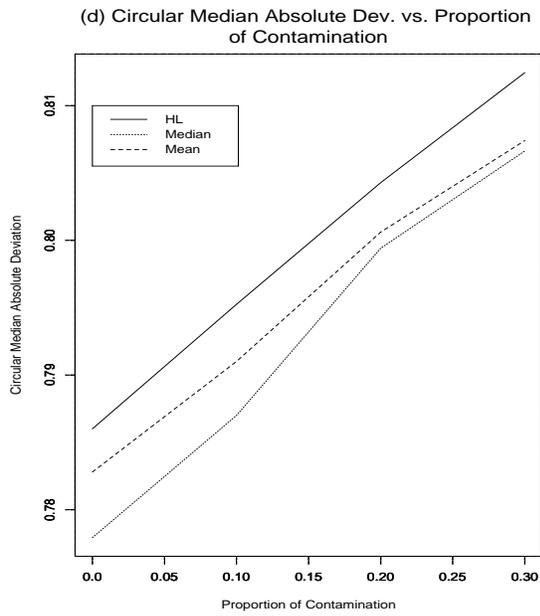
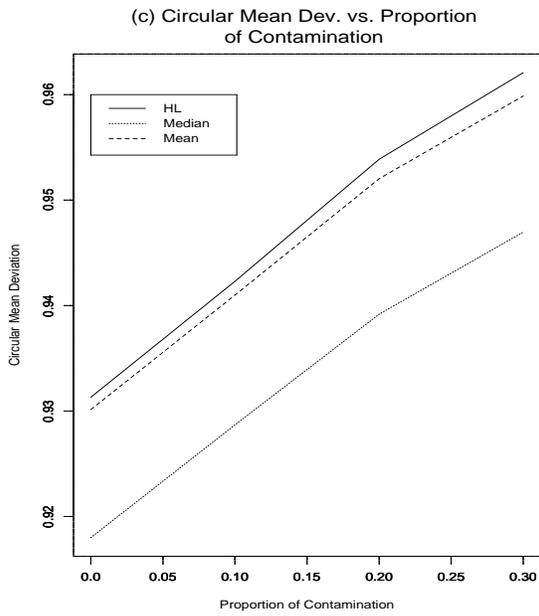
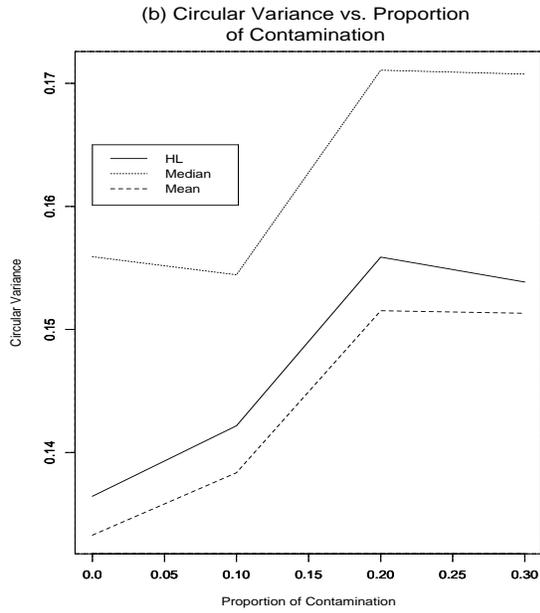
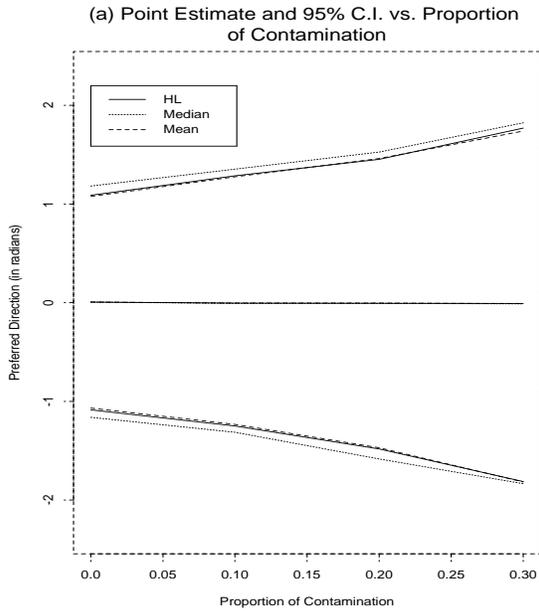
Effect of increasing the spread on Mean, Median and HL, for $\kappa = 1, n = 10$



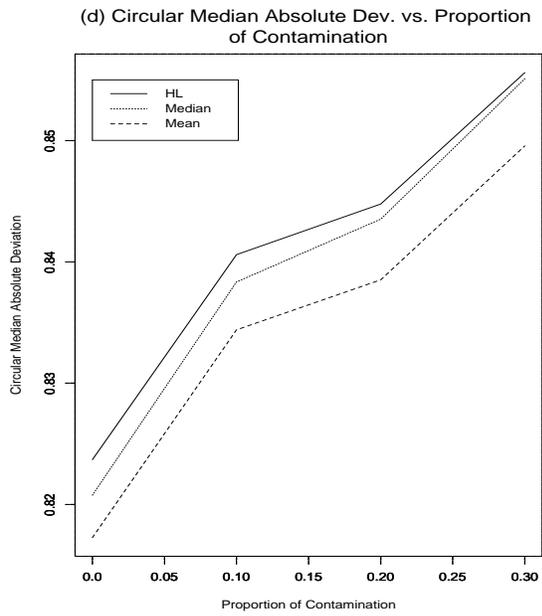
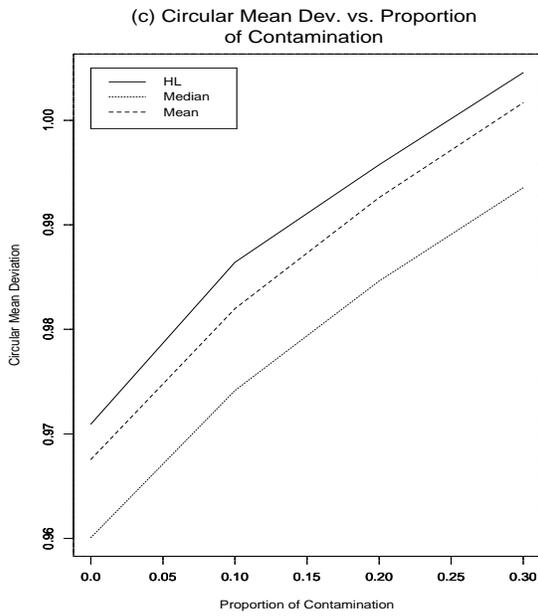
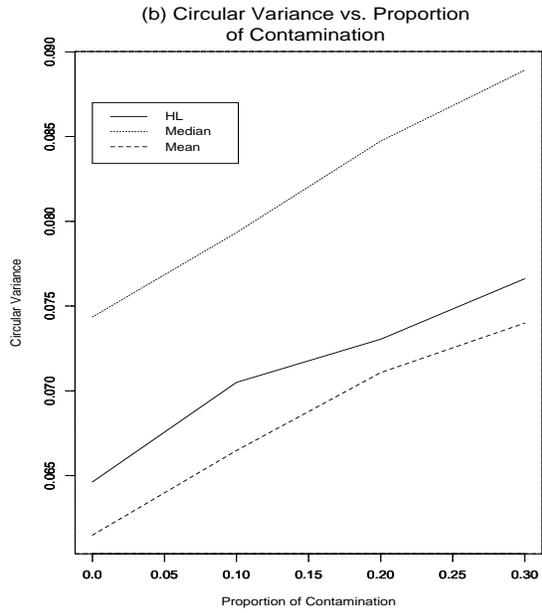
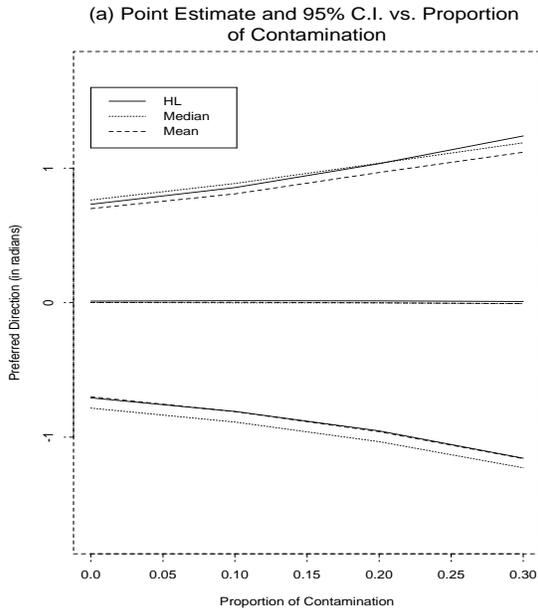
Effect of increasing the spread on Mean, Median and HL, for $\kappa = 1, n = 20$



Effect of contamination level on Mean, Median and HL, $\kappa = 1, n = 10$



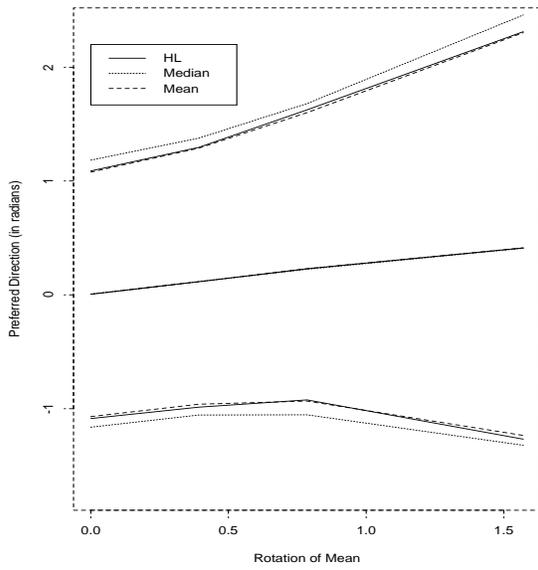
Effect of contamination level on Mean, Median and HL, $\kappa = 1, n = 20$



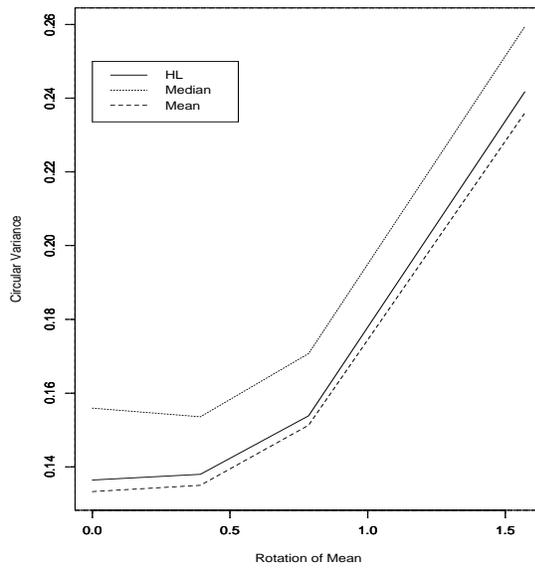
Effect of Rotating the mean direction from 0 to $\frac{\pi}{2}$ on Mean, Median and HL

$$\kappa = 1, \epsilon = 0.3, n = 10$$

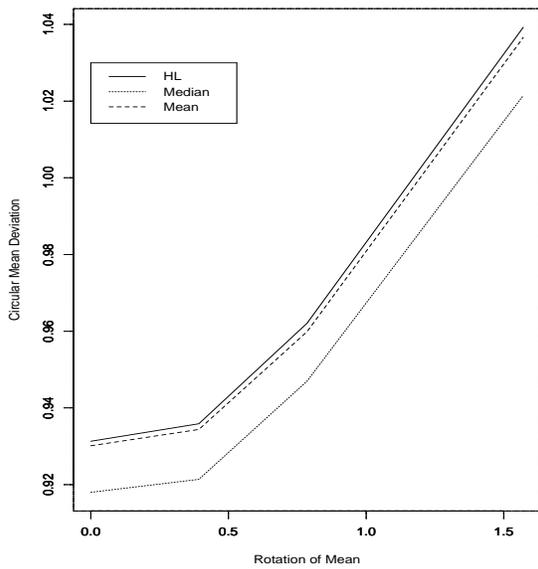
(a) Point Estimate and 95% C.I. vs. Rotation of Mean



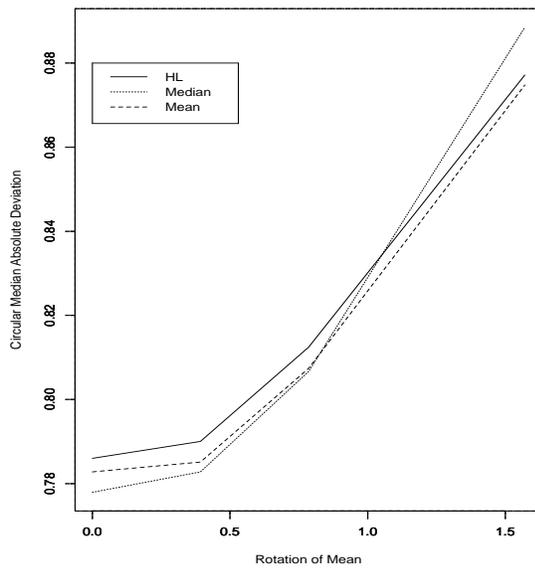
(b) Circular Variance vs. Rotation of Mean



(c) Circular Mean Dev. vs. Rotation of Mean

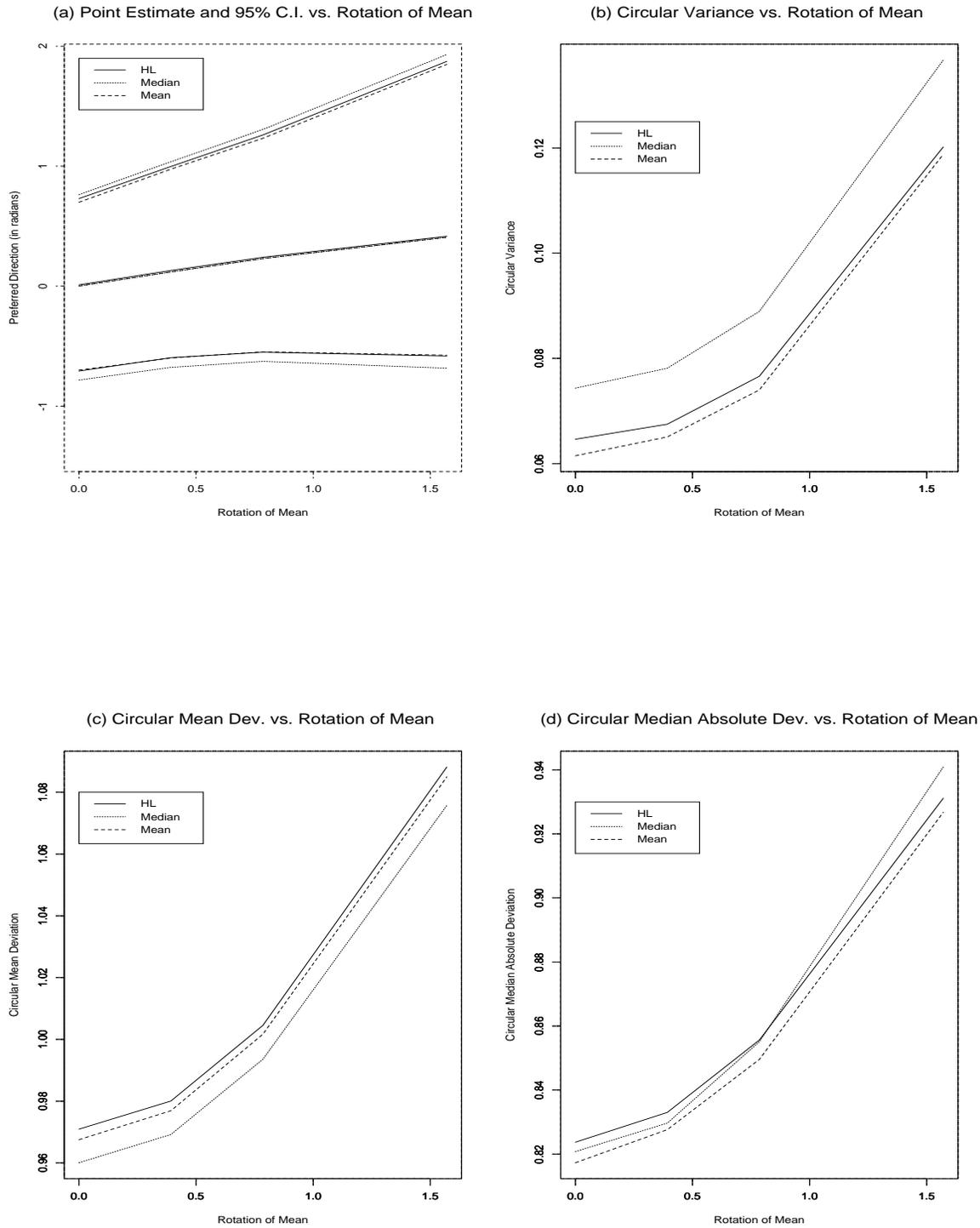


(d) Circular Median Absolute Dev. vs. Rotation of Mean

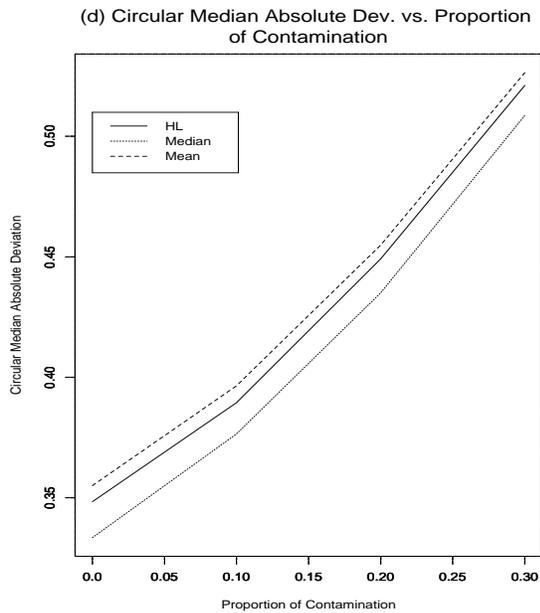
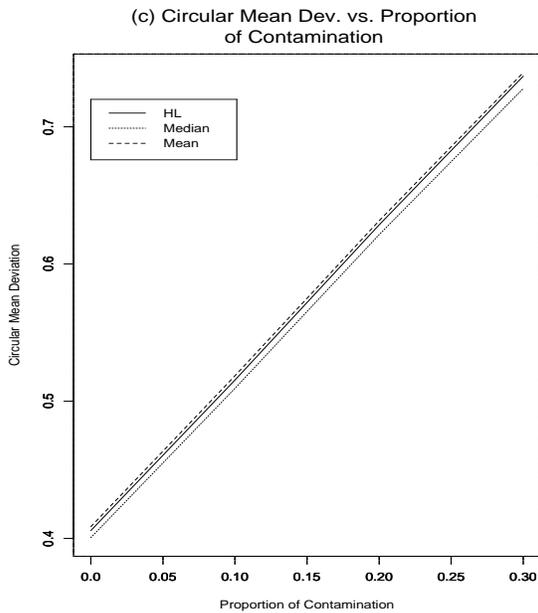
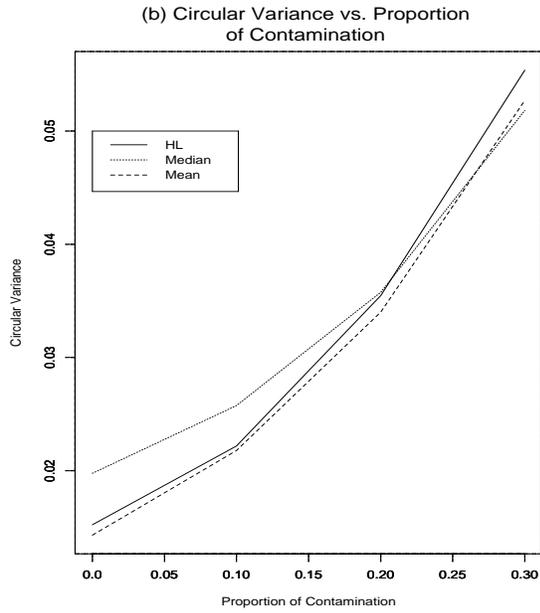
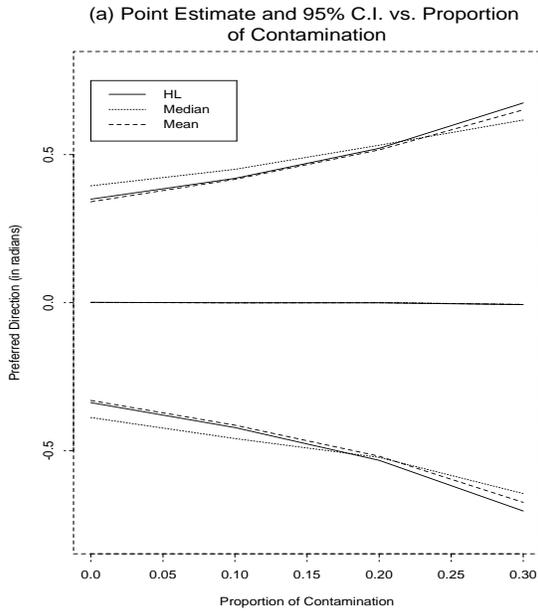


Effect of Shifting the mean direction from 0 to $\frac{\pi}{2}$ on Mean, Median and HL

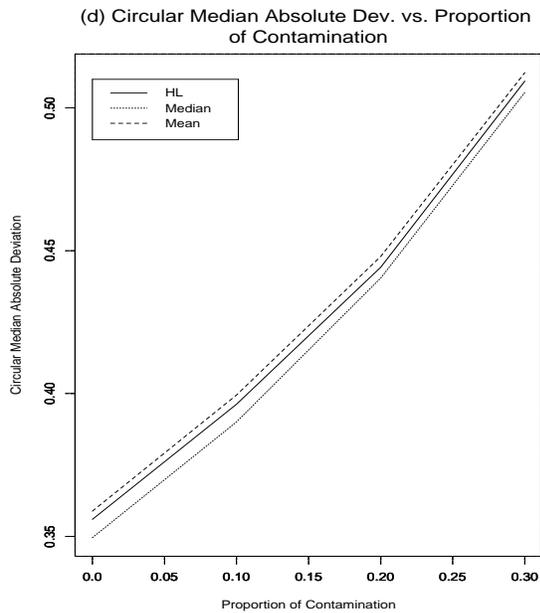
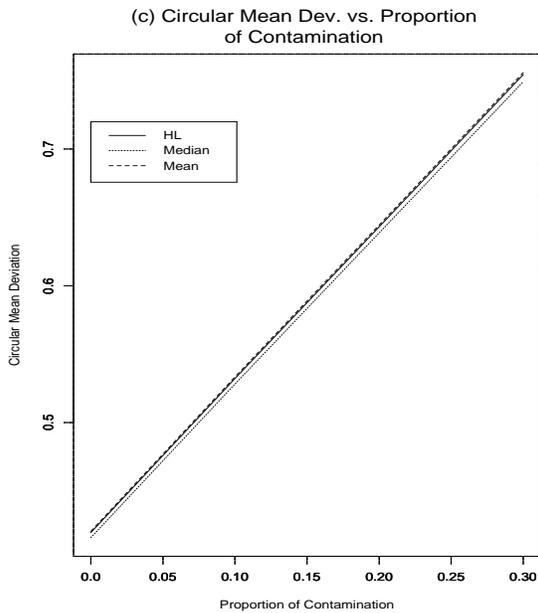
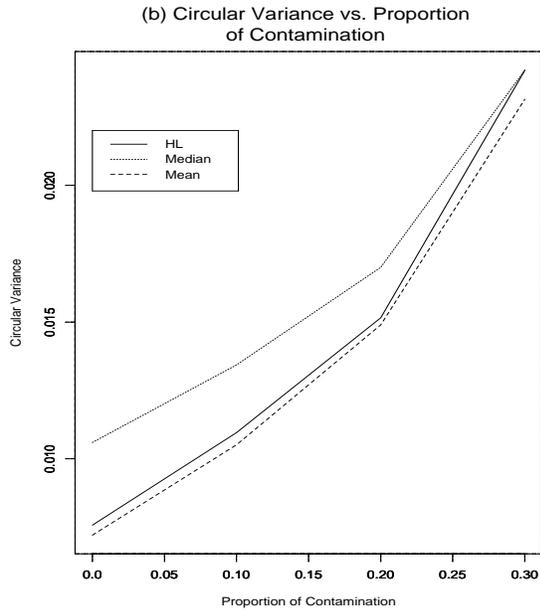
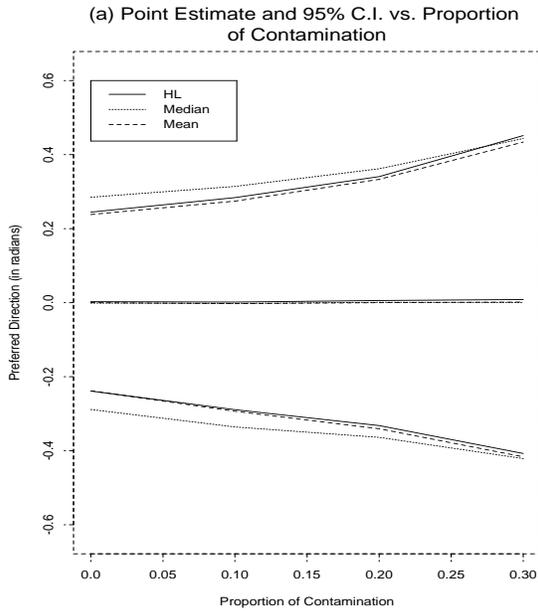
$$\kappa = 1, \epsilon = 0.3, n = 20$$



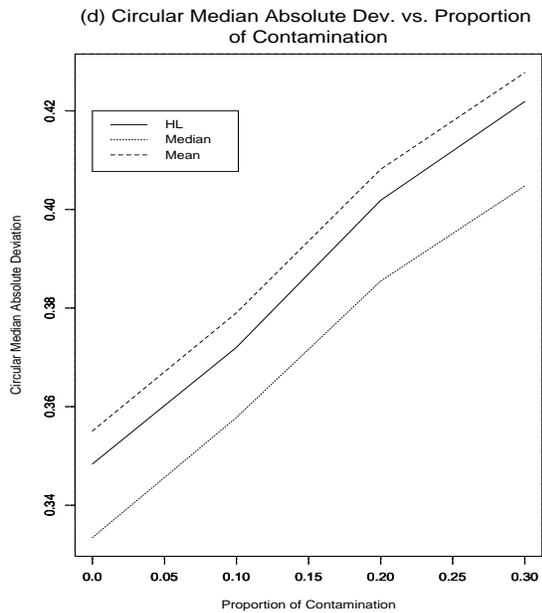
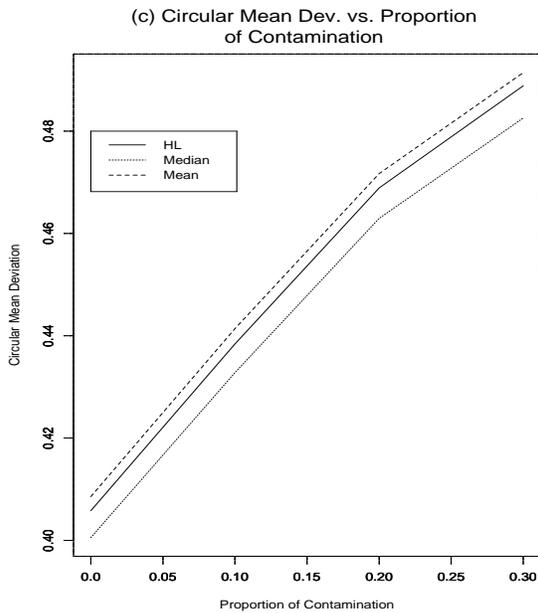
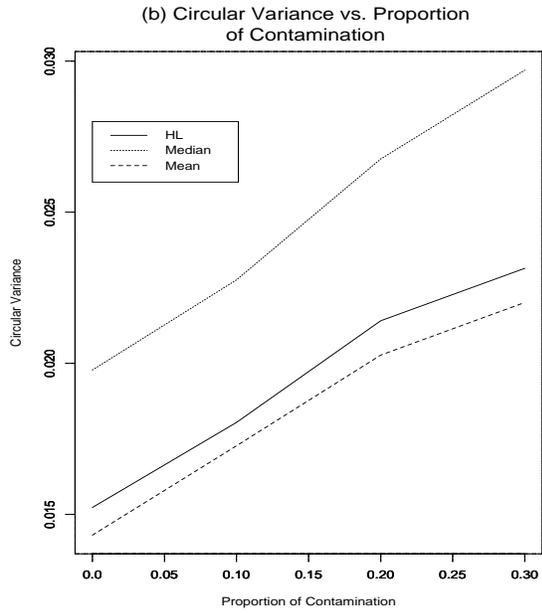
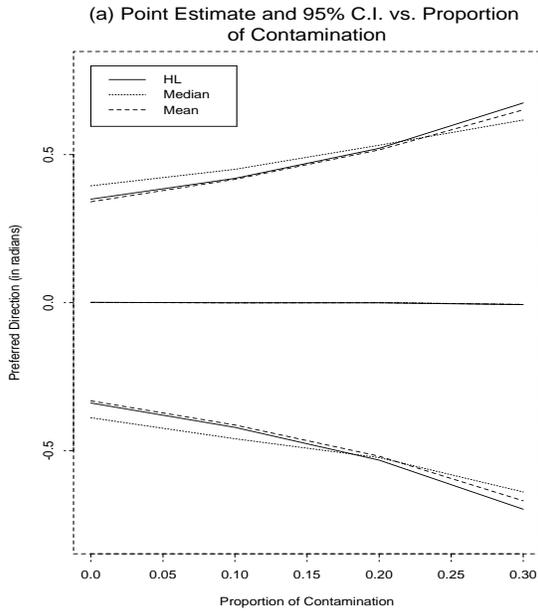
Effect of increasing the spread on Mean, Median and HL, for $\kappa = 4, n = 10$



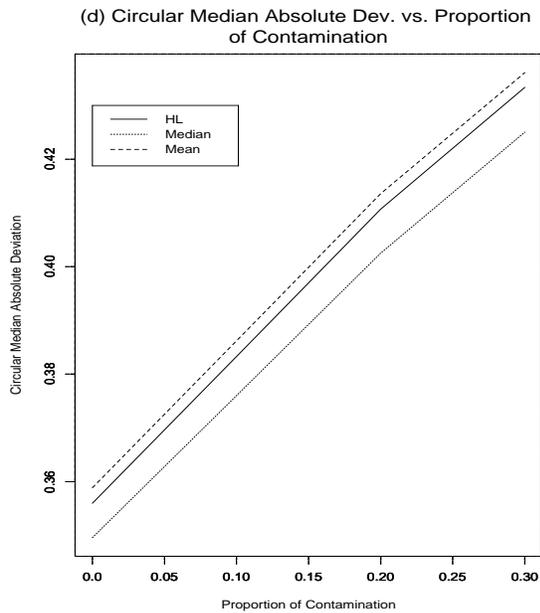
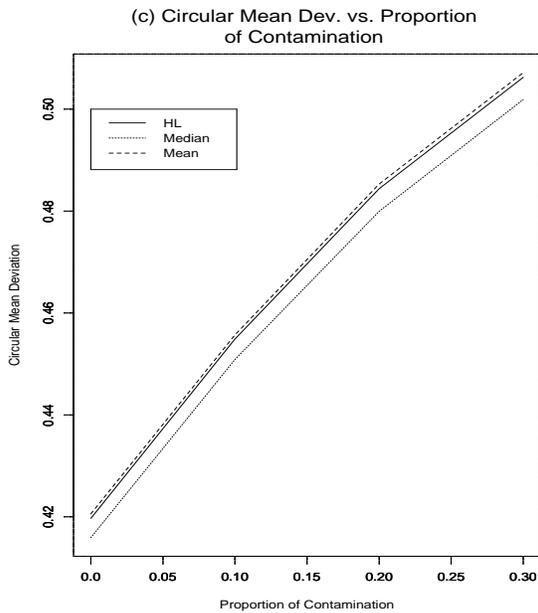
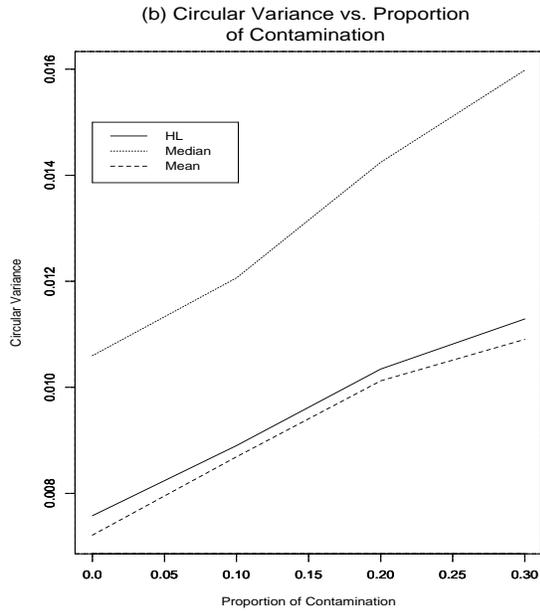
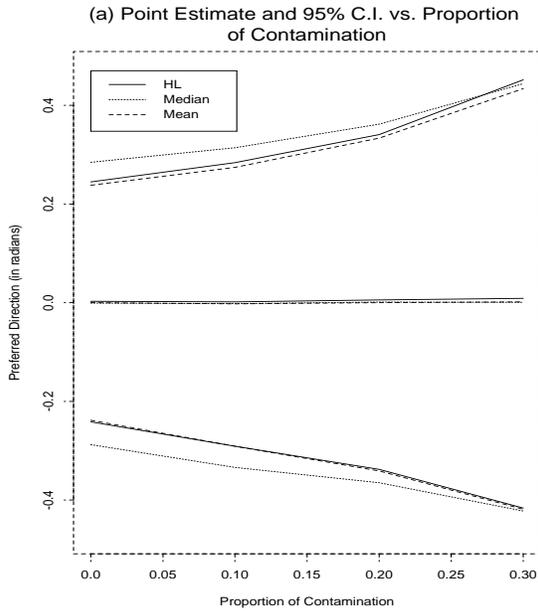
Effect of increasing the spread on Mean, Median and HL, for $\kappa = 4, n = 20$



Effect of contamination level on Mean, Median and HL, $\kappa = 4, n = 10$



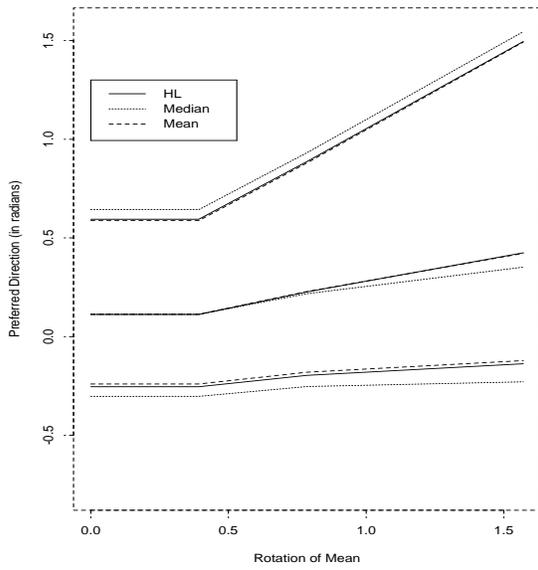
Effect of contamination level on Mean, Median and HL, $\kappa = 4, n = 20$



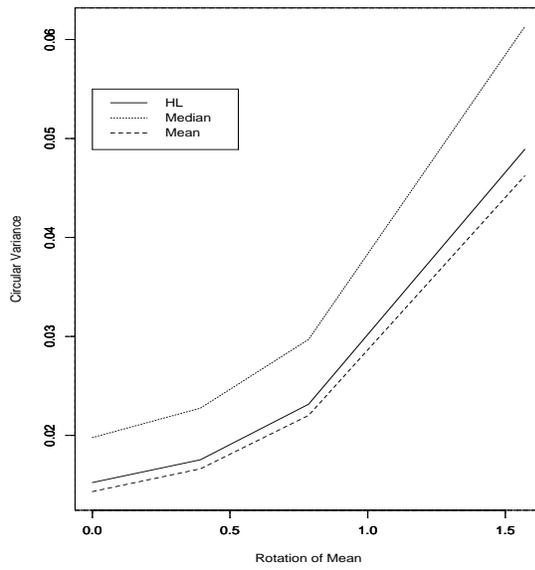
Effect of Shifting the mean direction from 0 to $\frac{\pi}{2}$ on Mean, Median and HL

$$\kappa = 4, \epsilon = 0.3, n = 10$$

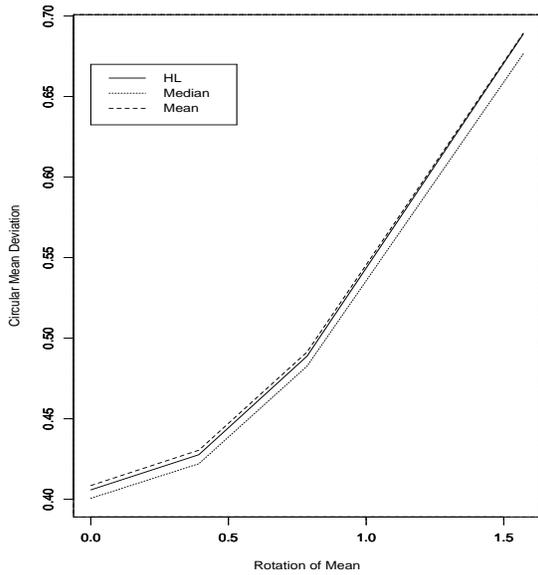
(a) Point Estimate and 95% C.I. vs. Rotation of Mean



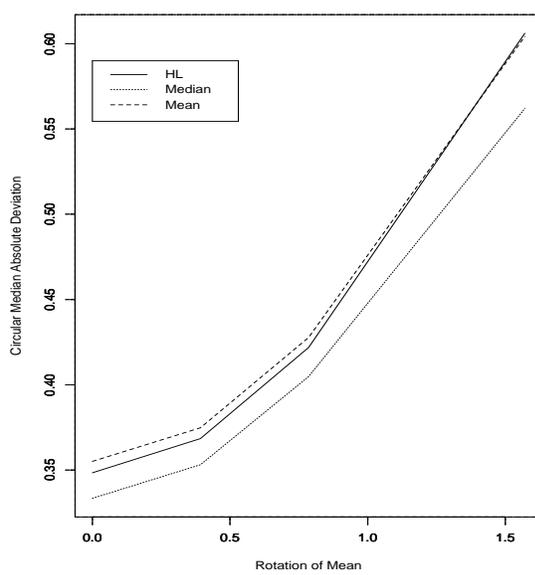
(b) Circular Variance vs. Rotation of Mean



(c) Circular Mean Dev. vs. Rotation of Mean



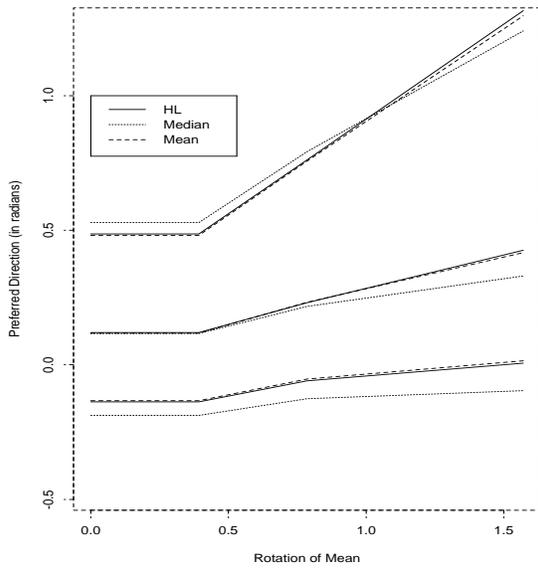
(d) Circular Median Absolute Dev. vs. Rotation of Mean



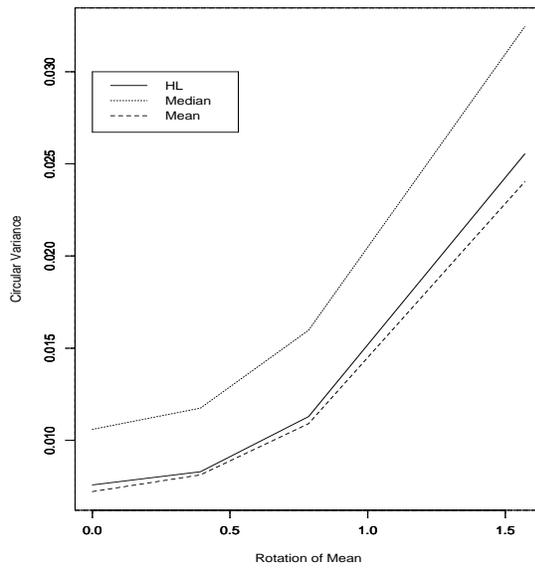
Effect of Shifting the mean direction from 0 to $\frac{\pi}{2}$ on Mean, Median and HL

$$\kappa = 4, \epsilon = 0.3, n = 20$$

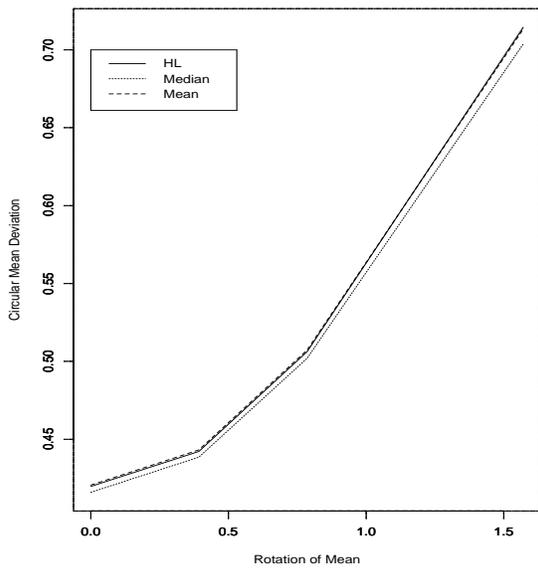
(a) Point Estimate and 95% C.I. vs. Rotation of Mean



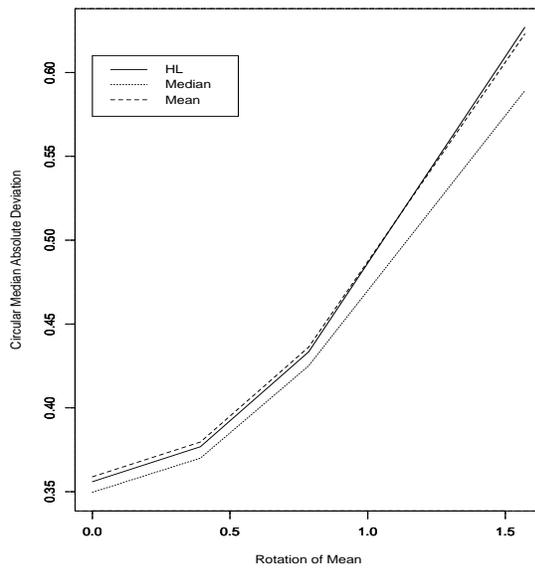
(b) Circular Variance vs. Rotation of Mean



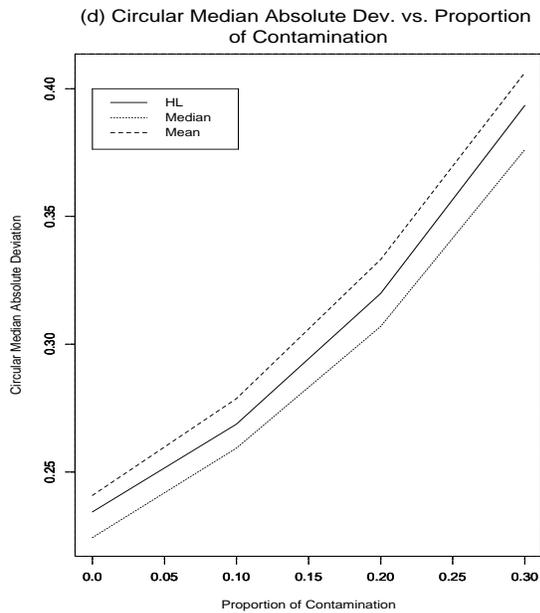
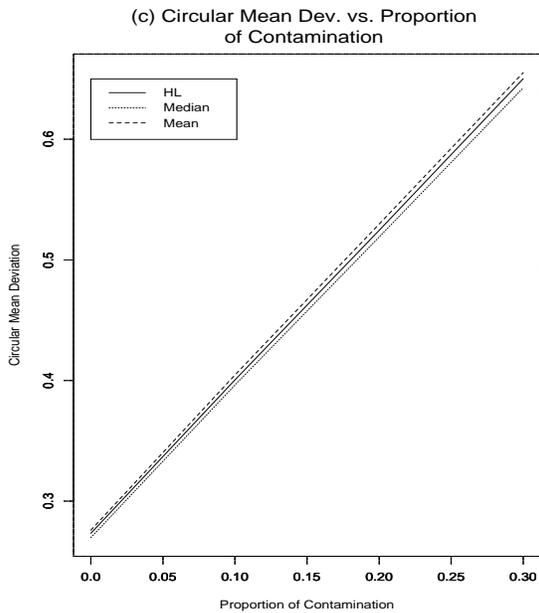
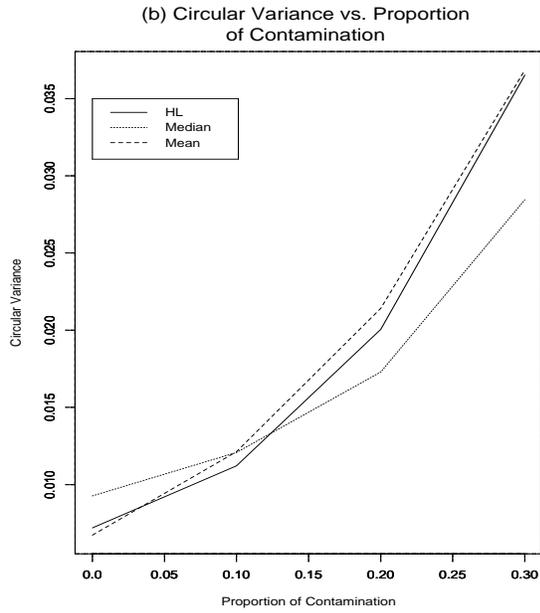
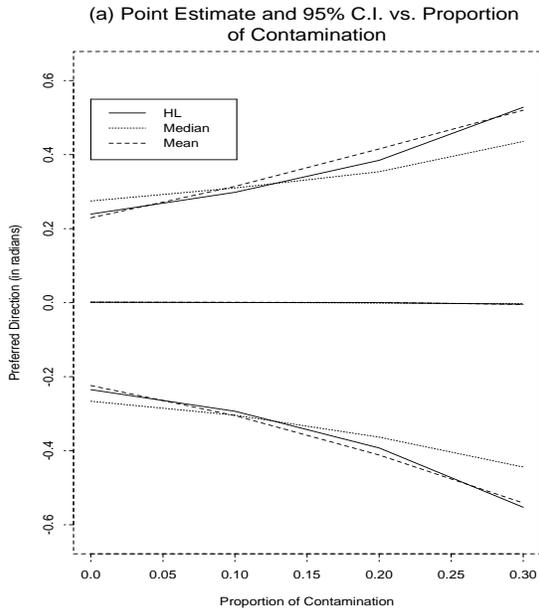
(c) Circular Mean Dev. vs. Rotation of Mean



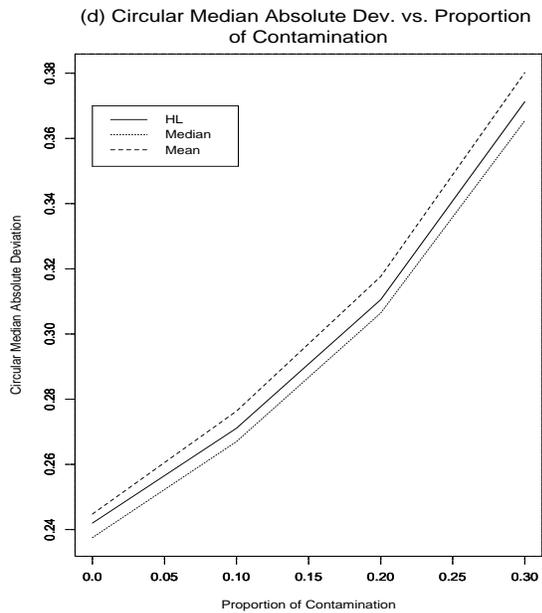
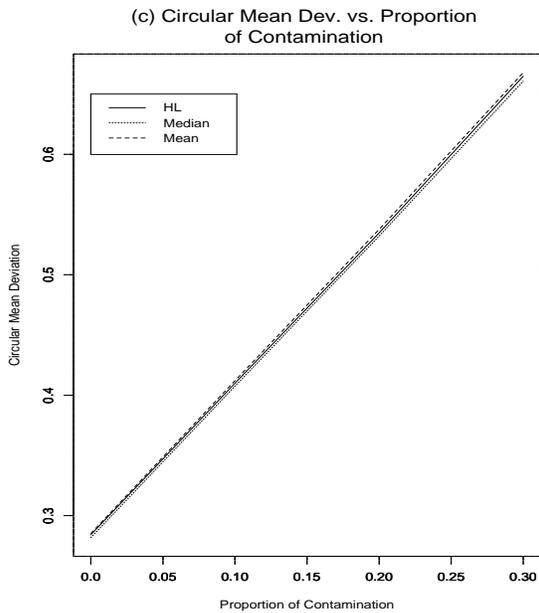
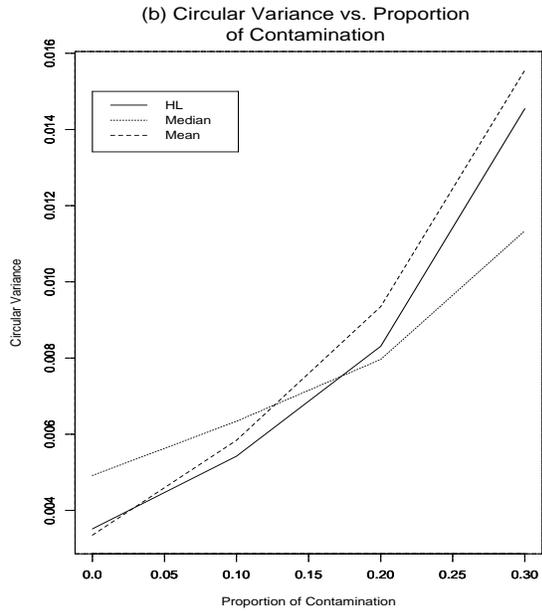
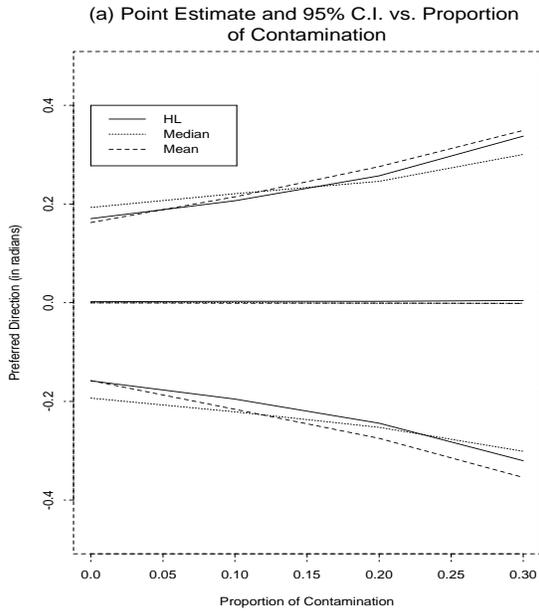
(d) Circular Median Absolute Dev. vs. Rotation of Mean



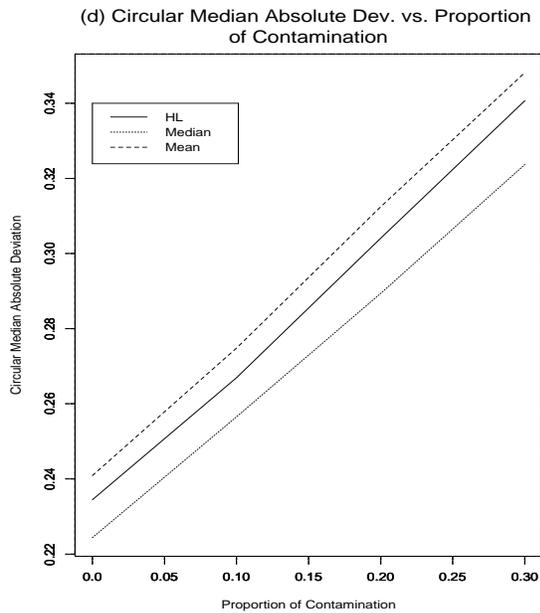
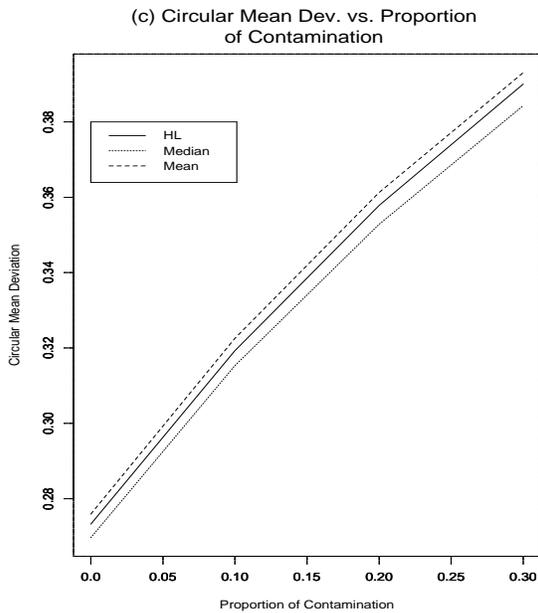
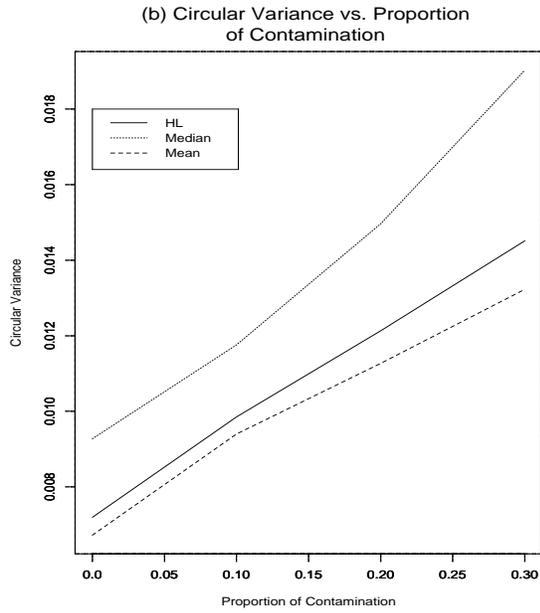
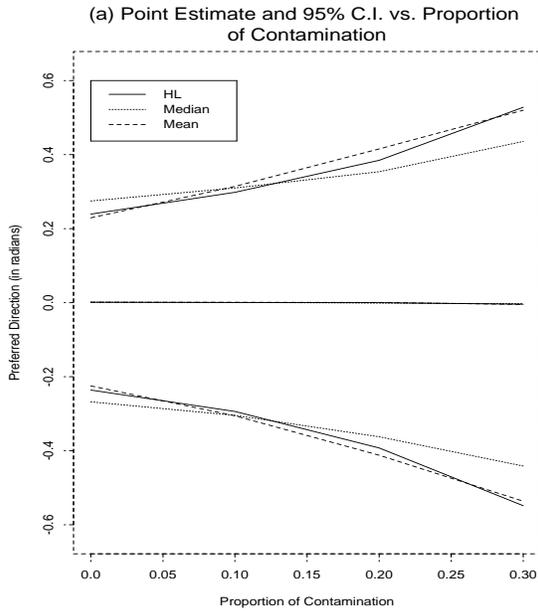
Effect of increasing the spread on Mean, Median and HL, for $\kappa = 8, n = 10$



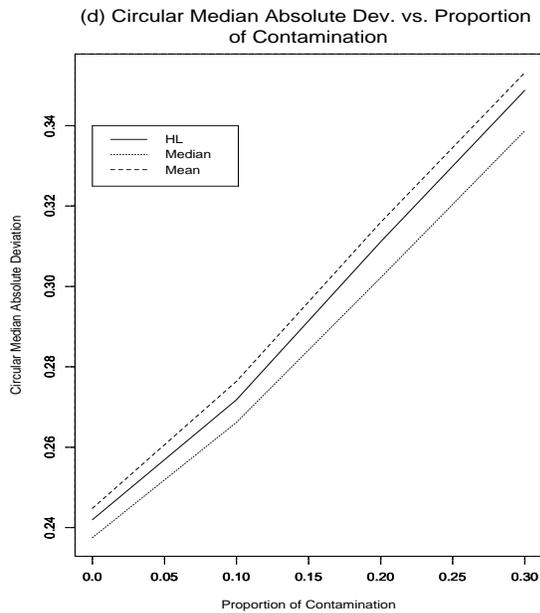
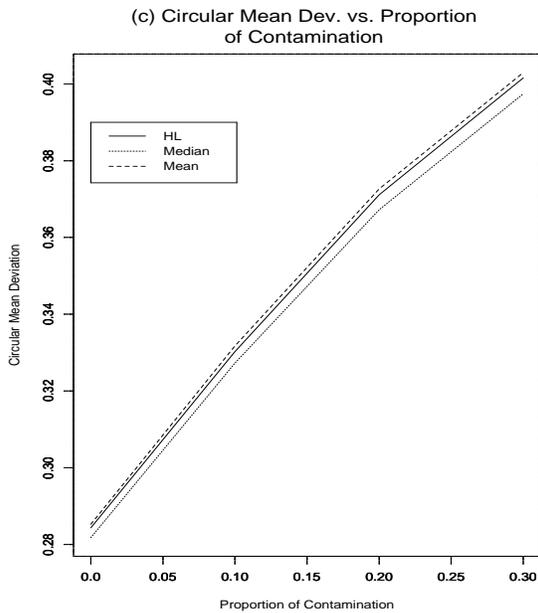
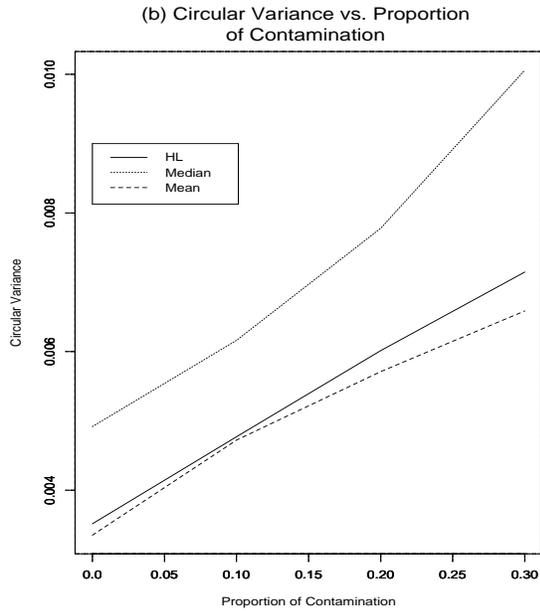
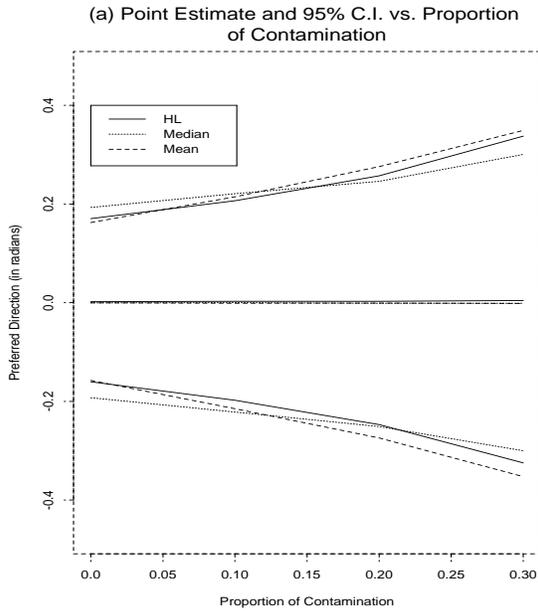
Effect of increasing the spread on Mean, Median and HL, for $\kappa = 8, n = 20$



Effect of contamination level on Mean, Median and HL, $\kappa = 8, n = 10$



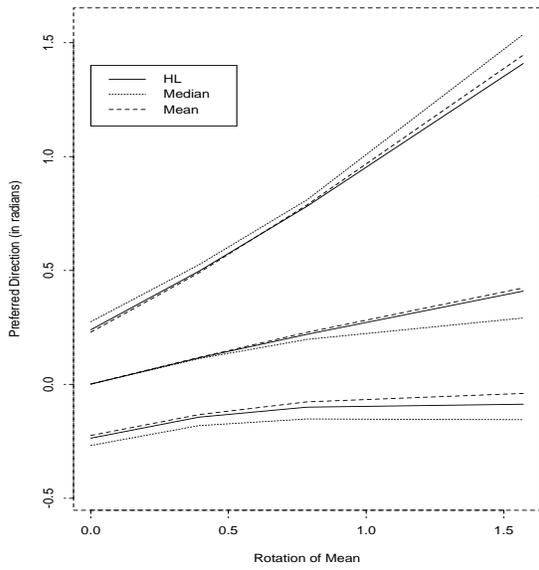
Effect of contamination level on Mean, Median and HL, $\kappa = 8, n = 20$



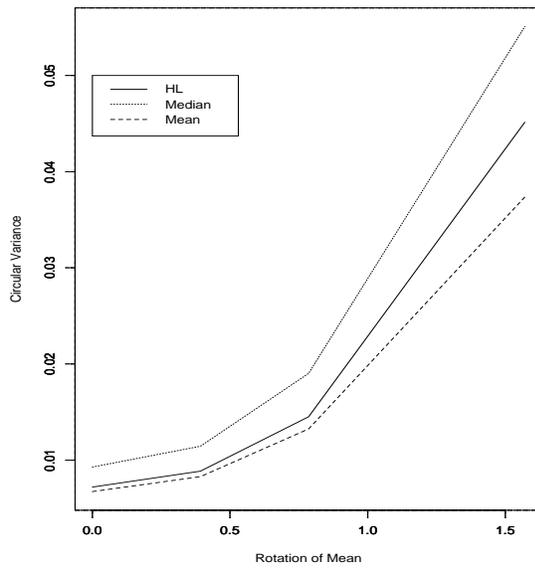
Effect of Shifting the mean direction from 0 to $\frac{\pi}{2}$ on Mean, Median and HL

$$\kappa = 8, \epsilon = 0.3, n = 10$$

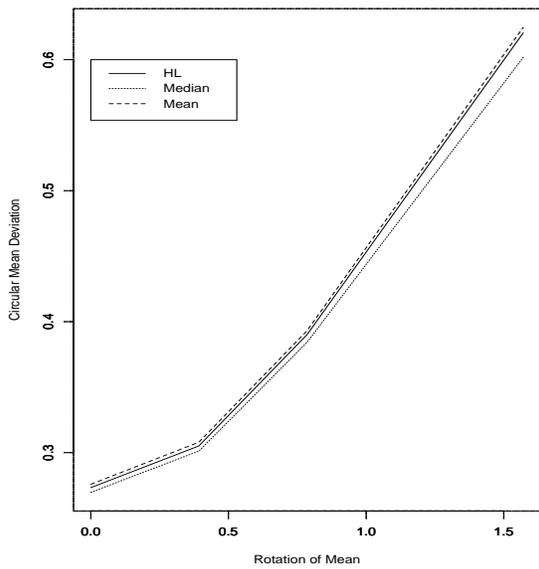
(a) Point Estimate and 95% C.I. vs. Rotation of Mean



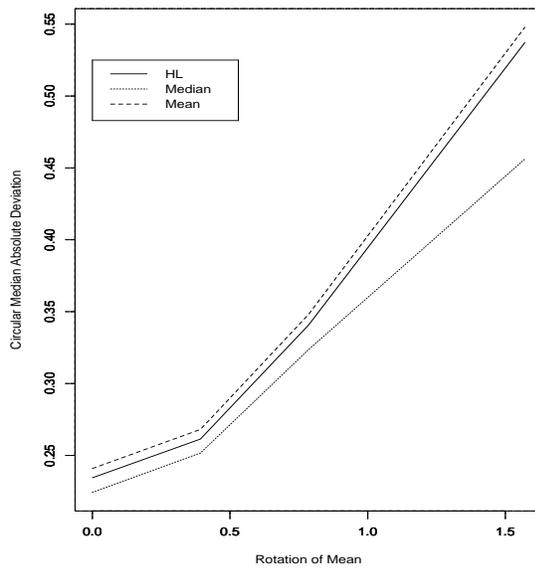
(b) Circular Variance vs. Rotation of Mean



(c) Circular Mean Dev. vs. Rotation of Mean



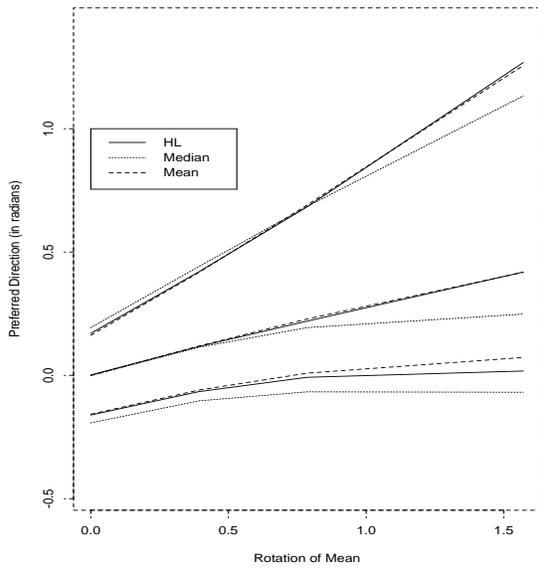
(d) Circular Median Absolute Dev. vs. Rotation of Mean



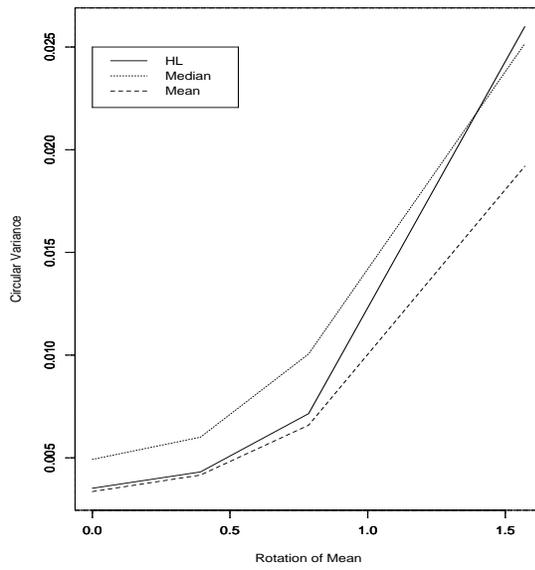
Effect of Shifting the mean direction from 0 to $\frac{\pi}{2}$ on Mean, Median and HL

$$\kappa = 8, \epsilon = 0.3, n = 20$$

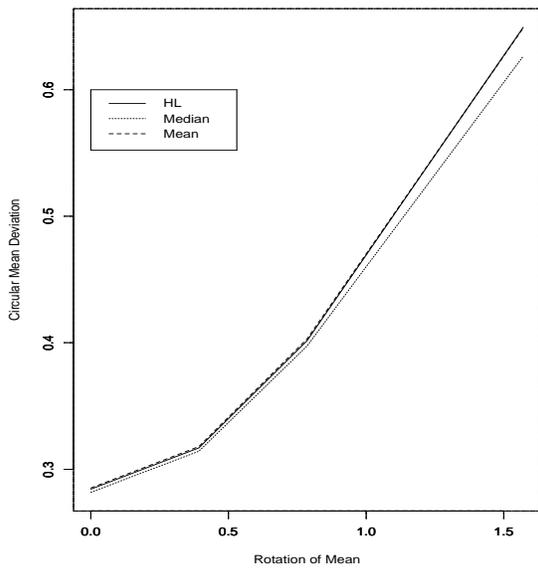
(a) Point Estimate and 95% C.I. vs. Rotation of Mean



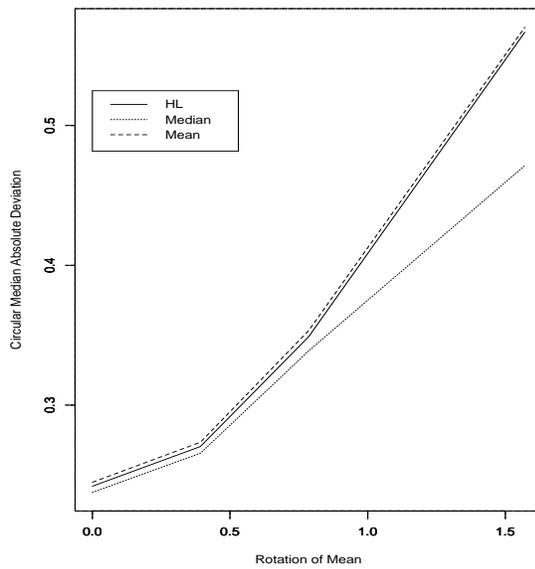
(b) Circular Variance vs. Rotation of Mean



(c) Circular Mean Dev. vs. Rotation of Mean



(d) Circular Median Absolute Dev. vs. Rotation of Mean



Appendix E : S-Plus Programs

There are no standard software for analyzing circular data. Jammalamadaka & SenGupta (2001), provide a 3.5M Floppy disk containing S-Plus functions for both Windows and Unix with their text. However, these functions are only useful for the procedures that they discuss. For this reason, ALL the code provided in this appendix was written specifically for the analyses/methods discussed in this dissertation.

The programming language chosen was S-Plus. The reason for choosing S-Plus is because it is very customizable. Its object-oriented design allows for creating one function to handle many different data types. This makes code design more versatile in a variety of applications.

The single drawback for using S-Plus over a lower level programming language such as C or Fortran is its inefficiency during large simulations. For this reason, many of the simulations provided in this dissertation took many hours to complete. Future simulation research in this area would greatly benefit from code that is written in a more efficient language.

Below are the necessary S-Plus functions to perform all the analyses in this dissertation.

- E.1 bofunc() & bofig()
- E.2 vmplot()
- E.3 ave.ang()
- E.4 posmedf()
- E.5 checkeven()
- E.6 checkodd()
- E.7 cmed()
- E.8 meandev()

E.9	<code>cmedM()</code>
E.10	<code>sipfunc()</code>
E.11	<code>ppcircmed()</code>
E.12	<code>sipfunc2()</code>
E.13	<code>ppcircmed2()</code>
E.14	<code>sipfunc3()</code>
E.15	<code>ppcircmeda()</code>
E.16	<code>simumat()</code>
E.17	<code>simumat1()</code>
E.18	<code>simumat2()</code>
E.19	<code>simumat3()</code>
E.20	<code>rangeang()</code>
E.21	<code>meddev()</code>
E.22	<code>simuNWNVM1()</code>
E.23	<code>simuNWNVM2()</code>
E.24	<code>simuNWNVM3()</code>
E.25	<code>simuNWNVMa()</code>
E.26a	<code>funcCI()</code>
E.26b	<code>funcmn()</code>
E.27	<code>funcmed()</code>
E.28	<code>bootSACI()</code>
E.29	<code>bootETCI()</code>
E.30	<code>bootLBCI()</code>
E.31	<code>simul.bootF()</code>
E.32	<code>bootfunc2()</code>

E.1 bofunc() and bofig()

These functions draws a circular scatter diagram

```

bofunc ← function(x) {
lenx ← length(x)
resang ← rep(0, lenx)
reslen ← rep(0, lenx)
sinx ← sum(sin(x))
cosx ← sum(cos(x))
for(i in 1:lenx) {
# Computes circular mean and resultant length of two observations
resang[i] ← atan(sin(x[i]), cos(x[i]))
reslen[i] ← sqrt((sin(x[i]))2 + (cos(x[i]))2) }
# Plots circle
plot(cos(0:360/180 * pi), sin(0:360/180 * pi), xlim = c(-1.05,1.05), ylim = c(-1.05,1.05), xlab
= " ", ylab = " ", type = "l", axes= F)
new.ang ← ((2 * ceiling((90 * x)/pi) - 1) * pi)/180
s.ang ← sort(new.ang)
# Puts observation # 1 on circumference of circle
points(1.05 * cos(s.ang[1]), 1.05 * sin(s.ang[1]), pch = "O")
dist ← 1.05
for(i in 2:lenx) {
if(s.ang[i] == s.ang[i - 1]) {
dist ← dist + 0.05 }
else { dist ← 1.05 }
# Puts other observations on the circumference of circle
points(dist * cos(s.ang[i]), dist * sin(s.ang[i]), pch= "O") }
aang ← ave.ang(x)

```

```

for(i in 1:length(resang)) {
lines(c(0, reslen[i] * cos(resang[i])), c(0, reslen[i] * sin(resang[i]))) }
aang2 ← ave.ang(resang)
return(list(resang, reslen, aang, aang2)) }

```

```

bofig ← function(x){
lenx ←length(x)
sx ← sort(sipfunc2(x))
# Plots circle
plot(cos(0:360/180 * pi), sin(0:360/180 * pi), xlim = c(-1.55,1.55),
ylim = c(-1.55,1.55), xlab = " ", ylab = " ", type = "l", axes= F)
new.ang ← ((2 * ceiling((90 * x)/pi) - 1) * pi)/180
s.ang ←sort(new.ang)
# Puts observation # 1 on circumference of the circle
points(1.05 * cos(s.ang[1]), 1.05 * sin(s.ang[1]), pch = "O")
dist ← 1.05
for(i in 2:lenx) {
# Plots identical observations stacked on each other
if(s.ang[i] == s.ang[i - 1]) {
dist ← dist + 0.10 }
else {
dist ← 1.05}
points(dist * cos(s.ang[i]), dist * sin(s.ang[i]), pch = "O")}
newsx ← ((2*ceiling((90*sx)/pi)-1)*pi)/180
nsx ← sort(newsx)
points(.95 * cos(nsx[1]), .95 * sin(nsx[1]), pch = "X")
for(i in 2:length(sx)) {
if(nsx[i]== nsx[i - 1]) { dist ← dist - 0.10}

```

```

else {
dist ← .95 }
#Puts point estimates near the circumference of the circle
points(dist * cos(nsx[i]), dist * sin(nsx[i]), pch = "X")
points(1.15*cos(ppcircmed(x)),1.15*sin(ppcircmed(x)),pch="D")
points(1.15*cos(ppcircmed2(x)),1.15*sin(ppcircmed2(x)),pch="h")
}

```

E.2 vmpplot()

This function plots von Mises Circular Density

```

vmpplot ← function(kappa,mu,b,add= F){
angs ← 0:360/180 *pi
# Computes exponent part of the VM density
temp← exp(kappa*cos(angs-mu))
tempsum ← sum(temp)
# Scales exponent term in order for area under the curve (density) to add up to one
tempnew ← temp/tempsum *100
if (add==F){
# Draws circle
plot (cos(angs),sin(angs),xlim=c(-2.8,2.8),
ylim= c(-2.8,2.8),xlab ="" , ylab ="" ,type = "l" ,lty= 1, axes= F)}
# Plots density
lines((1+tempnew)*cos(angs),(1+tempnew)*sin(angs), type= "l" ,lty= b+1)
return(NULL)
}

```

E.3 ave.ang()

This function calculates circular mean direction

```
ave.ang ← function(a) {
y ← sum(sin(a))
x ← sum(cos(a))
ifelse(round(x, 10) == 0 and round(y, 10) == 0, 9999, atan(y, x)) }
# If both x and y are zero, then no circular mean exists, so assign it a large number (9999).
```

E.4 posmedf()

This function calculates all potential medians for even samples

```
posmedf ← function(x) {
lenx ← length(x)
sx ← sort(x)
sx2 ← sx[c(2:lenx,1)]
# Determines closest neighbors of a fixed observation
posmed ← c()
for(i in 1:lenx) {
posmed[i] ← ave.ang(c(sx[i],sx2[i])) }
# Computes circular mean of two adjacent observations
posmed ← posmed[posmed ≠ 9999]
posmed }
```

E.5 checkeven()

This function checks if the number of possible medians is even

```
checkeven ← function(x){
```

```

lenx ← length(x)
sx ← sort(x)
check ← c()
# Computes possible medians
posmed ← posmedf(x)
for(i in 1:length(posmed)){
#Takes posmed[i] as the center, i.e. draws diameter at posmed[i] and counts observations
on either side of the diameter
newx ← sx-posmed[i]
check[i] ← ifelse(sum(round(cos(newx),10) > 0) > lenx/2, 9999, posmed[i])}
nposmed ← check[check ≠ 9999]
nposmed}

```

E.6 checkodd()

This function checks if the number of possible medians is odd

```

checkodd ← function(x) {
lenx ← length(x)
sx ← sort(x)
check ← c()
posmed ← sx
# Each observation is a possible median
for (i in 1:length(posmed)) {
newx ← sx-posmed[i]
#Takes posmed[i] as the center, i.e. draws diameter at posmed[i] and counts observations
on either side of the diameter
check[i] ← ifelse(sum(cos(newx) > 0) > (lenx-1)/2, 9999, posmed[i]) }
nposmed ← check[check ≠ 9999]

```

```
nposmed }
```

E.7 cmed()

This function calculates circular median “New Median”

```
cmed←function(x){
lenx ←length(x)
sx ← sort(x)
difsin ←c()
numties ←c()
if(lenx/2 == round(lenx/2)) {
# Checks if sample size is odd or even
# Computes median if sample size is even
posmed←checkeven(x)
for(i in 1:length(posmed)) {
newx ← sx - posmed[i]
difsin[i] ←sum(round(sin(newx),10)> 0) - sum(round(sin(newx),10) < 0)
numties[i] ← sum(round(newx, 10) == 0)}
}
else {
# Computes median if sample size is odd
posmed ← checkodd(x)
for(i in 1:length(posmed)) {
newx ← sx - posmed[i]
difsin[i] ← sum(round(sin(newx),10) > 0) - sum(round(sin(newx),10) < 0)
numties[i] ← sum(round(newx, 10) == 0)}
}
# Checks for ties
```

```

cm ← c(posmed[round(difsin, 10) == 0 | abs(difsin) > numties])
circmed ← ave.ang(cm) }
#Takes into account if possible circmed are equidistant from mean direction circmed

```

E.8 meandev()

This function calculates circular mean deviation

```

meandev ← function(x, teta) {
ifelse(teta == 9999, 9999, (pi - mean(round(abs(pi -
(abs(rangeang( x - teta))), 10))))))}

```

E.9 cmedM()

This function calculates circular median as defined in Mardia (1972, p. 28, 31)

```

cmedM ← function(x) {
lenx ← length(x)
sx ← sort(x)
sx2 ← c(sx[2:lenx], sx[1])
# Determines closest neighbors of a fixed observation
posmed ← rep(0, lenx)
difsin ← rep(0, lenx)
numties ← rep(0, lenx)
med ← c()
if(lenx/2 == round(lenx/2)) {
# Checks if sample is odd or even
posmed ← posmedf(x)
# Computes median if sample size is even

```

```

for(i in 1:length(posmed)) {
newx ← sx - posmed[i]
difsin[i] ← sum(round(sin(newx),10) > 0) - sum(round(sin(newx),10) < 0)
numties[i] ← sum(round(newx, 10) == 0) } } else {
# Computes median if sample size is even
posmed ← checkodd(x)
for(i in 1:length(posmed)) { newx ← sx - posmed[i]
difsin[i] ← sum(round(sin(newx),10)> 0) - sum(round(sin(newx),10) < 0)
numties[i] ← sum(round(newx, 10) == 0) } }
# Checks for ties
cm ← c(posmed[round(difsin, 10) == 0 | round(abs(difsin),10) < numties])
for (i in 1:length(cm)) {
# Computes the circular mean deviation for candidate medians
med[i] ← meandev(x,cm[i]) }
circmed ← ave.ang(cm[round(med,10) == round(min(med),10)]) }
# Chooses the candidate medians with smallest circular mean deviations and takes circular
mean of them if more that one.

```

E.10 sipfunc()

This function calculates pairwise circular means for HL1

```

sipfunc ← function(x) {
lenx ← length(x)
pang ← matrix(9999, nrow = lenx, ncol = lenx)
for(i in 1:(lenx - 1)) { for(j in (i + 1):lenx) {
# Computes pairwise circular means excluding observation with itself
if((round(sin(x[i]) + sin(x[j]), 10) ≠ 0) and (round(cos(x[i]) + cos(x[j]), 10) ≠ 0)) {
pang[i, j] ← ave.ang1(c(x[i], x[j])) } } }

```

```
$ Keeps only values that are not equal to 9999
ang ← pang[pang ≠ 9999]
ang }
```

E.11 ppcircmed()

This function calculates HL1

```
ppcircmed ← function(x) {
# Obtains Mardia median of the pairwise circular means
cmedM(sipfunc(x))
}
```

E.12 sipfunc2()

This function calculates pairwise circular means for HL2

```
sipfunc2 ← function(x) {
lenx ← length(x)
pang ← matrix(9999, nrow = lenx, ncol = lenx)
for(i in 1:lenx) {
for(j in i:lenx) {
if(i == j) { # Computes pairwise circular means including observation with itself
pang[i, j] ← x[i] }
else if((round(sin(x[i]) + sin(x[j]), 10) ≠ 0) and (round(cos(x[i]) + cos(x[j]), 10) ≠ 0)) {
pang[i, j] ← ave.ang(c(x[i], x[j])) }}}
ang ← pang[pang ≠ 9999]
ang }
```

E.13 ppcircmed2()

This function calculates HL2

```
ppcircmed2 ← function(x) {
# Obtains Mardia median of the pairwise circular means
cmedM(sipfunc2(x))
}
```

E.14 sipfunc3()

This function calculates pairwise averages for HL3 for use in simuNWNVMa function

```
sipfunc3←function(x){
# Computes ALL pairwise circular means
c(sipfunc(x),sipfunc2(x))}
```

E.15 ppcircmeda()

This function calculates HL3

```
ppcircmeda←function(x){
#Computes Mardia median of the pairwise circular means
cmedM(sipfunc3(x)) }
```

E.16 simumat()

This function simulates data from $(N(\mu, \sigma)) \bmod(2*\pi) == WN(\mu, A(\kappa)) == VM(\mu, \kappa)$

```
simumat ←function(n, mu, s, M) {
```

```

res ← matrix(nrow = n, ncol = M)
res1 ← c()
for(i in 1:M) {
#Converts normal variates to wrapped normal variates
res1 ← rnorm(n, mu, s) %% (2 * pi)
res[, i] ← res1 }
res }

```

E.17 simumat1()

This function simulates data from contaminated $[(N(\mu, \sigma_1)) \bmod(2\pi) == WN(\mu, A(\kappa)) == VM(\mu, \kappa)]$ with propability (1-p) and $[(N(\mu, \sigma_2)) \bmod(2\pi) == WN(\mu_2, A(\kappa)) == VM(\mu_2, \kappa)]$ with probability p

```

simumat1←function(n,mu,s1,s2,p,M){
res←matrix(nrow=n,ncol=M)
res1←c()
for(i in 1:M){
#Determines number of contaminated observations using binomial distribution
numcont←rbinom(1,n,p)
#Converts normal variates to wrapped normal variates
rVM←rnorm((n-numcont),mu,s1)%%(2*pi)
rVm←rnorm(numcont,mu,s2)%%(2*pi)
# Sample contains some outliers
res1←c(rVM, rVm)%%(2*pi)
res[,i]←sort(res1)}
res }

```

E.18 simumat2()

This function simulates data from uncontaminated VM with probability (1-p) and uniform (-pi, pi) with probability p

```

simumat2←function(n,mu,s1,p,M){
res←matrix(nrow=n,ncol=M)
res1←c()
for(i in 1:M){
#Determines number of contaminated observations using binomial distribution
numcont←rbinom(1,n,p)
#Converts normal variates to wrapped normal variates
rVM←rnorm((n-numcont),mu,s1)%%(2*pi)
rUnif←runif(numcont,-pi,pi)# Sample contains some outliers
res1j-c(rVM, rUnif)%%(2*pi)
res[,i]←sort(res1)}
res}

```

E.19 simumat2()

This function simulates data from contaminated $[(N(\mu_1, \sigma)) \bmod(2*\pi) == WN(\mu_1, A(\kappa)) == VM(\mu_1, \kappa)]$ with propability (1-p) and $[(N(\mu_2, \sigma)) \bmod(2 * \pi) == WN(\mu_2, A(\kappa)) == VM(\mu_2, \kappa)]$ with probability p

```

simumat3←function(n, mu1,mu2, s1, p, M){
res ←matrix(nrow = n, ncol = M)
res1 ← c()
for(i in 1:M) {

```

```

#Determines number of contaminated observations using binomial distribution
numcont←rbinom(1,n,p)
#Converts normal variates to wrapped normal variates
rVM←rnorm((n-numcont),mu1,s1)%%(2*pi)
rVm←rnorm(numcont,mu2,s1)%%(2*pi)
# Sample contains some outliers
res1←c(rVM, rVm)%%(2*pi)
res[,i]←sort(res1)}
res}

```

E.20 rangeang()

This function guarantees that angles are in the right range for computation of confidence intervals

```

rangeang ← function(x) {
# Converts observations to  $(-\pi, \pi)$  range
ang ← ifelse(x < - pi, x + 2 * pi, x)
ang2 ← ifelse(ang > pi, ang - 2 * pi, ang)
return(ang2) }

```

E.21 meddev()

This function calculates circular median absolute deviation

```

#Computes median deviation Hampel(1974) meddev← function(x,teta){
ifelse(teta==9999, 9999, pi-median(abs(pi-abs(rangeang(x-teta)))))) }

```

E.22 simuNWNVM1()

This function calculates circular mean direction, circular median, and their corresponding resultant lengths, 95% empirical confidence intervals, and circular mean deviations

```

simuNWNVM1←function(M){
meanD ← c()
meanD1←c()
medD ← c()
medD1 ← c()
mmedD←c()
mmedD1←c()
res2 ← rep(0, M)
res3 ← rep(0, M)
res4 ← rep (0,M)
for(i in 1:M) {
res1 ← res[, i]
#Computes point estimates
res2[i]←ave.ang(res1)
res3[i] ←cmed(res1)
res4[i] ←cmedM1(res1)
#Computes circular mean dev. and circular median absolute dev.
meanD[i]←meandev1(res1,res2[i])
meanD1[i]←meddev(res1,res2[i])
medD[i]←meandev1(res1,res3[i])
medD1[i]←meddev(res1,res3[i])
mmedD[i]←meandev1(res1,res4[i])
mmedD1[i]←meddev(res1,res4[i])}
#Keeps all values except 9999

```

```

nres2 ← res2[res2 ≠ 9999]
nres3 ← res3[res3 ≠ 9999]
nres4 ← res4[res4 ≠ 9999]
nres2 ← sort(nres2)
scmean ← ave.ang(nres2)
# Converts observations to  $(-\pi, \pi)$  range
newscmn ← rangeang(nres2)
# Computes 95% Empirical C.I.s, mean resultant lengths and point estimates
for the mean, Mardia median and New median
cicmean ← quantile(newscmn, c(0.025, 0.975)) + scmean
srcmean ← sqrt((((sum(round(sin(nres2), 10)))/length(nres2))^2) +
(((sum(round(cos(nres2), 10)))/length(nres2))^2))
nres3 ← sort(nres3)
scmedian ← ave.ang(nres3)
nscmedian ← rangeang(nres3)
cicmedian ← quantile(nscmedian, c(0.025, 0.975)) + scmedian
srcmedian ← sqrt((((sum(round(sin(nres3), 10)))/length(nres3))^2) +
(((sum(round(cos(nres3), 10)))/length(nres3))^2))
nres4 ← sort(nres4)
scmed ← ave.ang(nres4)
nscmed ← rangeang(nres4)
cicmed ← quantile(nscmed, c(0.025, 0.975)) + scmed
srcmed ← sqrt((((sum(round(sin(nres4), 10)))/length(nres4))^2) +
(((sum(round(cos(nres4), 10)))/length(nres4))^2))
# Computes mean dev. and median absolute dev. for the three measures
meanD ← meanD[meanD ≠ 9999]
meanD ← sort(meanD)
meand ← ave.ang(meanD)

```

```

meanD1 ← meanD1[meanD1 ≠ 9999]
meanD1 ← sort(meanD1)
meand1 ← ave.ang(meanD1)
medD ← medD[medD ≠ 9999]
medD ← sort(medD)
medd ← ave.ang(medD)
medD1 ← medD1[medD1 ≠ 9999]
medD1 ← sort(medD1)
medd1 ← ave.ang(medD1)
mmedD ← mmedD[mmedD ≠ 9999]
mmedD ← sort(mmedD)
mmedd ← ave.ang(mmedD)
mmedD1 ← medD1[mmedD1 ≠ 9999]
mmedD1 ← sort(mmedD1)
mmedd1 ← ave.ang(mmedD1)
list(WNVM = res, cmean = res2, scmean = scmean, cicmean=cicmean,srcmean=srcmean,
meand = meand, meand1=meand1,cmedian= res3, scmedian = scmedian, cicmedian = ci-
cmedian, srcmedian = srcmedian,medd = medd, medd1=medd1,scmed= scmed, cicmed =
cicmed, srcmed = srcmed,mmedd = mmedd, mmedd1=mmedd1)}

```

E.23 `simuNWNVM2()`

This function calculates HL1 and corresponding resultant length, 95% empirical confidence interval, and circular mean deviation

```

simuNWNVM2 ← function(M) {
res2 ← rep(0, M)
for(i in 1:M) {
res1 ← res[, i]

```

```

# Computes point estimate
res2[i] ← ppcircmed(res1) }
nres2 ← res2[res2 ≠ 9999]
sMHL ← ave.ang(nres2)
# Converts observations to  $(-\pi, \pi)$  range
newsMHL ← rangeang(nres2 - sMHL[sMHL ≠ 9999])
#Computes 95% Empirical Confidence Interval
ciMHL ← quantile(newsMHL, c(0.025, 0.975)) + sMHL[sMHL ≠ 9999]
# Computes mean resultant length
srMHL ← sqrt((((sum(sin(nres2)))/M)2 + (((sum(cos(nres2)))/M)2))
#Computes circular mean deviation for HL1
MHLd ← meandev(nres2,sMHL)
list(WNVNVM=res, MHL = res2, sMHL = sMHL, ciMHL= ciMHL, srMHL =srMHL, MHLd=
MHLd)}

```

E.24 simuNWNVM3()

This function calculates HL2 and corresponding resultant length, 95% empirical confidence interval, and circular mean deviation

```

simuNWNVM3 ← function(M) {
res2 ← rep(0, M)
for(i in 1:M) {
res1 ← res[, i]
# Computes point estimate
res2[i] ← ppcircmed2(res1) }
nres2 ← res2[res2 ≠ 9999]
sHL ← ave.ang(nres2)
# Converts observations to  $(-\pi, \pi)$  range

```

```

newsHL ← rangeang(nres2 - sHL[sHL ≠ 9999])
#Computes 95% Empirical Confidence Interval
ciHL ← quantile(newsHL, c(0.025, 0.975)) + sHL[sHL ≠ 9999]
# Computes mean resultant length
srHL ← sqrt((((sum(sin(nres2)))/M)2 + (((sum(cos(nres2)))/M)2))
#Computes circular mean deviation
HLd ← meandev(nres2,sHL)
list(WNVm = res, HL = res2, sHL = sHL, ciHL =ciHL, srHL = srHL, HLd= HLd)}

```

E.25 simuNWNVMa()

This function calculates HL3 and corresponding resultant length, 95% empirical confidence interval, and circular mean deviation

```

simuNWNVMa←function(M){
mmhld←c()
res2 ← rep(0, M)
for(i in 1:M) {
res1 ← res[, i]
# Computes point estimate
res2[i] ← ppcircmeda(res1)
#Computes circular mean deviation
mmhld[i]←meandev(res1,res2[i])}
nres2 ← res2[res2 ≠ 9999]
sMMHL ← ave.ang(nres2)
# Converts observations to  $(-\pi, \pi)$  range
newsMMHL ← rangeang(nres2 - sMMHL)
newsMMHL←newsMMHL[newsMMHL > -9990]
#Computes 95% Empirical Confidence Intervals for HL3

```

```

ciMMHL ← quantile(newsMMHL, c(0.025, 0.975)) + sMMHL
# Computes mean resultant length
srMMHL← sqrt((((sum(round(sin(nres2),10)))/length(nres2))^2) + (((sum(round(cos(nres2),10)))/length(nres2))^2))
mmhld←mmhld[mmhld≠ 9999]
MMHLd←ave.ang(mmhld)
list(WNVM = res, MMHL = res2, sMMHL = sMMHL, ciMMHL = ciMMHL, srMMHL = srMMHL, MMHLd= MMHLd)}

```

E.26a funCI()

This function calculates 95% Confidence Interval for the circular mean (Fisher 1993, p.88 and Jamalamadaka & SenGupta (2001,p.96)

```

funcCI←function(x,func1){
n←length(x)
# Computes point estimate
pe1←func1(x)
#Computes mean resultant length, S.E and lower and upper confidence limits
rbar←sqrt((((sum(round(sin(x),10)))/n)^2) + (((sum(round(cos(x),10)))/n)^2))
kaph←akapinv(rbar)
serr←1/(sqrt(n*rbar*kaph))
l1←pe1-asin(1.96*serr)
u1←pe1+asin(1.96*serr)
return(pe1,rbar,kaph,serr,l1,u1)
}

```

where

```
# MLE of  $\kappa$ 
```

```

akappainv←function(x){
ifelse(x<0.85,(-0.4 + 1.39 * x + 0.43/(1 - x)),(x3 - 4 * x2 + 3 * x)-1) }
akapinv←function(x){
ifelse(x<0.53,(2 * x + x3 + (5 * x5)/6), akappainv(x))}

```

Since akapinv is biased (Fisher, 1993, p.88), use

```

kaphat←function(x){
n←length(x)
x1←akapinv(x)
return(x2)}

```

E.26b funcmn()

This function calculates Bootstrap confidence intervals for mean direction Fisher & Mc Powell (1989)

```

funcmn←function(data,func, B, alfa){
n ←length(data)
#Computes point estimate
pe ← func(data)
#Obtains bootstrap samples
bootx ←bootsample(data,B)
peb ← c()
for(i in 1:B) {
#Computes point estimate from bootstrap sample
peb[i] ← func(bootx[, i])}
speb ← sort(peb)
#Position of lower limit

```

```

L ← trunc(0.5 + (0.5 * B * alfa))
#position of upper limit
m ← B - L
nbootx1 ← sort(rangeang(speb - pe)) #Basic Method
l1 ← pe + nbootx1[L + 1] #Lower C.I.
u1 ← pe + nbootx1[m] #Upper C.I.
w1 ← abs(u1 - l1)
nbootx2 ←abs(sort(speb - pe)) #symmetric distribution
l2 ←pe - nbootx2[m] #Lower C.I.
u2 ← pe + nbootx2[m] #Upper C.I.
w2 ← abs(u2 - l2)
return(pe, l1,u1,w1,l2,u2,w2,L,m)}

```

where

#Generating Bootstrap Samples

```

bootsample←function(x,B){
n ←length(x)
bootsam←matrix(nrow= n,ncol= B)
bootx←c()
for(i in 1:B) {
#Sampling With Replacement
bootx ← sample(x, n, replace = T)
bootsam[, i]←bootx }
bootsam}

```

E.27 funcmed()

This function calculates 95% Confidence Interval for the circular median for $n \geq 15$ (Fisher 1993, p.72-73)

```
funcmed←function(x,func){
n←length(x)
pe←func(x)
#Computes C.I.
m←(ifelse(n > 15, (1+trunc(0.5 * n0.5 * (1.96))), (1+trunc(0.5 * n0.5 * (1.96)) - 2)))
newx←sort(rangeang(x-pe))
l1←newx[m]+pe      # lower C.L.
u1←newx[(n-(m-1))+pe] # upper C.L.
w1←abs(u1-l1)     # interval width
return(pe,l1,u1,w1,m)}
```

E. 28 bootSACI()

Symmetric Arc Bootstrap Confidence Interval

```
bootSACI←function(data,func1,func2,func3,B,alfa){
n←length(data)
datab1←c(1:B)
datab2←c(1:B)
datab3←c(1:B)
#Draws B samples and computes B bootstrap estimates
for (i in 1:B){
ndatab←sample(data,n, replace= T)
datab1[i]←func1(ndatab)
datab2[i]←func2(ndatab)
```

```

datab3[i]←func3(ndatab)}
#mean resultant length, point estimate, lower and upper C.L. for mean
rL1 ←sqrt((((sum(round(sin(sort(datab1)),10)))/length(sort(datab1)))2) +
(((sum(round(cos(sort(datab1)),10)))/length(sort(datab1)))2))
pe1←func1(data)
sdata1←sort(abs(rangeang(datab1-pe1)))
bound←quantile(sdata1, (1-alfa))
l1←pe1-bound
u1←pe1+bound
#mean resultant length, point estimate, lower and upper C.L. for median
rL2 ←sqrt((((sum(round(sin(sort(datab2)),10)))/length(sort(datab2)))2) +
(((sum(round(cos(sort(datab2)),10)))/length(sort(datab2)))2))
pe2←func2(data)
sdata2←sort(abs(rangeang(datab2-pe2)))
bound←quantile(sdata2, (1-alfa))
l2←pe2-bound
u2←pe2+bound
#mean resultant length, point estimate, lower and upper C.L. for HL
pe3←func3(data)
rL3 ← sqrt((((sum(round(sin(sort(datab3)),10)))/length(sort(datab3)))2) +
(((sum(round(cos(sort(datab3)),10)))/length(sort(datab3)))2))
sdata3←sort(abs(rangeang(datab3-pe3)))
bound←quantile(sdata3, (1-alfa))
l3←pe3-bound
u3←pe3+bound
return(rL1,pe1,l1,u1,rL2,pe2,l2,u2,rL3,pe3,l3,u3)}

```

E.29 bootETCI()

This function calculates Equal-Tailed Bootstrap Confidence Interval

```

bootETCI←function(data,func1,func2,func3,B,alfa){
n←length(data)
datab1←c(1:B)
datab2←c(1:B)
datab3←c(1:B)
#Draws B samples and computes B bootstrap estimates
for (i in 1:B){
ndatab←sample(data,n, replace= T)
datab1[i]←func1(ndatab)
datab2[i]←func2(ndatab)
datab3[i]←func3(ndatab)}
#mean resultant length, point estimate, lower and upper C.L. for mean
pe1←ave.ang(data)
rL1← sqrt((((sum(round(sin(sort(datab1)),10)))/length(sort(datab1))))^2) +
(((sum(round(cos(sort(datab1)),10)))/length(sort(datab1))))^2))
sdatab1←sort((rangeang(datab1-pe1)))
l1←pe1+quantile(sdatab1, alfa/2)
u1←pe1+quantile(sdatab1,(1-alfa/2))
#mean resultant length, point estimate, lower and upper C.L. for median
pe2←func2(data)
rL2←sqrt((((sum(round(sin(sort(datab2)),10)))/length(sort(datab2))))^2) +
(((sum(round(cos(sort(datab2)),10)))/length(sort(datab2))))^2))
sdatab2←sort((rangeang(datab2-pe2)))
l2←pe2+quantile(sdatab2, alfa/2)
u2←pe2+quantile(sdatab2,(1-alfa/2))

```

```

#mean resultant length, point estimate, lower and upper C.L. for HL
pe3←func3(data)
rL3 ← sqrt((((sum(round(sin(sort(datab3)),10)))/length(sort(datab3)))2 +
(((sum(round(cos(sort(datab3)),10)))/length(sort(datab3)))2))
sdatab3←sort((rangeang(datab3-pe3)))
l3←pe3+quantile(sdatab3, alfa/2)
u3←pe3+quantile(sdatab3,(1-alfa/2))
return(rL1,pe1,l1,u1,rL2,pe2,l2,u2,rL3,pe3,l3,u3)}

```

E. 30 bootLBCI()

This function calculates Likelihood-Based Bootstrap Confidence Interval

```

bootLBCI←function(data,func1,func2,func3,B,alfa){
n←length(data)
datab1←c(1:B)
datab2←c(1:B)
datab3←c(1:B)
#Draws B samples and computes B bootstrap estimates
for (i in 1:B){
ndatab←sample(data,n, replace=T)
datab1[i]←func1(ndatab)
datab2[i]←func2(ndatab)
datab3[i]←func3(ndatab)}
#mean resultant length, point estimate, lower and upper C.L. for mean
pe1←func1(data)
rL1←sqrt((((sum(round(sin(sort(datab1)),10)))/length(sort(datab1)))2 +
(((sum(round(cos(sort(datab1)),10)))/length(sort(datab1)))2))

```

```

sdatab1←sort(rangeang(datab1))
lenB←length(sdatab1)
minang←rep(0,lenB)
for (i in 1:lenB){
minang[i]←quantile((sdatab1-sdatab1[i])%%(2*pi),1-alfa)}
poslower←sdatab1[round(minang-min(minang),10)==0]
l1←pe1-sort(abs(rangeang(poslower-pe1)))[1]
u1←l1+min(minang)
#mean resultant length, point estimate, lower and upper C.L. for median
pe2←func2(data)
rL2←sqrt((((sum(round(sin(sort(datab2)),10)))/length(sort(datab2)))2 +
(((sum(round(cos(sort(datab2)),10)))/length(sort(datab2)))2))
sdatab2←sort(rangeang(datab2))
lenB←length(sdatab2)
minang←rep(0,lenB)
for (i in 1:lenB){
minang[i]←quantile((sdatab2-sdatab2[i])%%(2*pi),1-alfa)}
poslower←sdatab2[round(minang-min(minang),10)==0]
l2←pe2-sort(abs(rangeang(poslower-pe2)))[1]
u2←l2+min(minang)
#mean resultant length, point estimate, lower and upper C.L. for HL
pe3←unc3(data)
rL3 ←sqrt((((sum(round(sin(sort(datab3)),10)))/length(sort(datab3)))2 +
(((sum(round(cos(sort(datab3)),10)))/length(sort(datab3)))2))
sdatab3←sort(rangeang(datab3))
lenB←length(sdatab3)
minang←rep(0,lenB)
for (i in 1:lenB){

```

```

minang[i]←quantile((sdat3-sdat3[i])%%(2*pi),1-alfa)}
poslower←sdat3[round((minang-min(minang),10)==0)]
l3←pe3-sort(abs(rangeang(poslower-pe3)))[1]
u3←l3+min(minang)
return(rL1,pe1,l1,u1,rL2,pe2,l2,u2,rL3,pe3,l3,u3)}

```

E.31 simul.bootF()

This function simulates the confidence intervals

```

simul.bootCI←function(n,mu,s,M,func1,func2,func3,B,alfa){
fish1←c()
wfish1←c()
fish2←c()
wfish2←c()
fish3←c()
wfish3←c()
SA1←c()
SA2←c()
SA3←c()
wSA1←c()
wSA2←c()
wSA3←c()
ET1←c()
ET2←c()
ET3←c()
wET1←c()
wET2←c()
wET3←c()
}

```

```

LB1←c()
LB2←c()
LB3←c()
wLB1←c()
wLB2←c()
wLB3←c()
ciL ← matrix(nrow = M, ncol = 24)
#Generates M samples of size n from WN distribution
#Computes C.I.s based on the Basic Method, Symmetric Distr.
#Median Theory, Symmetric-Arc, Equal-Tailed and Likelihood-Based
for(i in 1:M) {
data←rnorm(n,mu,s)%%(2*pi)
bresmn←funcmn(data,func1,B,alfa)
bresmed←funcmed(data,func2)
bresSA←bootSA(data,func1,func2,func3,B,alfa)
bresET←bootET(data,func1,func2,func3,B,alfa)
bresLB←bootLB(data,func1,func2,func3,B,alfa)
#Basic Method
ciL[i,1]←bresmnl1
ciL[i,2]←bresmnu1
fish1[i]←ifelse((bresmnl1 < μ&μ < bresmnu1),1,0)
wfish1[i]←abs(bresmnu1-bresmnl1)
#Symmetric Distribution method
ciL[i,3]←bresmnl2
ciL[i,4]←bresmnu2
fish2[i]←ifelse((bresmnl2 < μ&μ < bresmnu2),1,0)
wfish2[i]←abs(bresmnu2-bresmnl2)
#Median Theory

```

```

ciL[i,5]←bresmedl1
ciL[i,6]←bresmedu1
fish3[i]←ifelse((bresmedl1 <  $\mu$ & $\mu$  < bresmedu1),1,0)
wfish3[i]←abs(bresmedu1-bresmedl1)
#Symmetric-Arc
ciL[i, 7]← bresSA1
ciL[i, 8]← bresSAu1
SA1[i]←ifelse((bresSA1 <  $\mu$ & $\mu$  < bresSAu1),1,0)
wSA1[i]←abs(bresSAu1 -bresSA1)
ciL[i, 9]← bresSA2
ciL[i, 10]← bresSAu2
SA2[i]←ifelse((bresSA2 <  $\mu$ & $\mu$  < bresSAu2),1,0)
wSA2[i]←abs(bresSAu2 -bresSA2)
ciL[i, 11]← bresSA3
ciL[i, 12]← bresSAu3
SA3[i]←ifelse((bresSA3 <  $\mu$ & $\mu$  < bresSAu3),1,0)
wSA3[i]←abs(bresSAu3 -bresSA3)
#Equal-Tailed Arc
ciL[i, 13]← bresET1
ciL[i, 14]← bresETu1
ET1[i]←ifelse((bresET1 <  $\mu$ & $\mu$  < bresETu1),1,0)
wET1[i]←abs(bresETu1 -bresET1)
ciL[i, 15]← bresET2
ciL[i, 16]← bresETu2
ET2[i]←ifelse((bresET2 <  $\mu$ & $\mu$  < bresETu2),1,0)
wET2[i]←abs(bresETu2 -bresET2)
ciL[i, 17]← bresET3
ciL[i, 18]← bresETu3

```

```

ET3[i]←ifelse((bresETl3 < μ&μ < bresETu3),1,0)
wET3[i]←abs(bresETu3 -bresETl3)
#Likelihood-Based Arc
ciL[i, 19]← bresLB1
ciL[i, 20]←bresLBu1
LB1[i]←ifelse((bresLB1 < μ&μ < bresLBu1),1,0)
wLB1[i]←abs(bresLBu1 -bresLB1)
ciL[i, 21]← bresLB2
ciL[i, 22]← bresLBu2
LB2[i]←ifelse((bresLB2 < μ&μ < bresLBu2),1,0)
wLB2[i]←abs(bresLBu2 -bresLB2)
ciL[i, 23]← bresLB3
ciL[i, 24]← bresLBu3
LB3[i]←ifelse((bresLB3 < μ&μ < bresLBu3),1,0)
wLB3[i]←abs(bresLBu3 -bresLB3)}
ciL
#Basic Method
fish1←sum(fish1)
rLfish1←sqrt((((sum(round(sin(sort(wfish1)), 10)))/length(sort(wfish1))))2 +
( ((sum(round(cos(sort(wfish1)), 10)))/length(sort(wfish1))))2))
wfish1←func1(sort(wfish1))
#Symmetric Distribution
fish2←sum(fish2)
rLfish2←sqrt((((sum(round(sin(sort(wfish2)), 10)))/length(sort(wfish2))))2 +
( ((sum(round(cos(sort(wfish2)), 10)))/length(sort(wfish2))))2))
wfish2←func1(sort(wfish2))
# Median Theory
fish3←sum(fish3)

```

```

rLfish3←sqrt((((sum(round(sin(sort(wfish3)), 10)))/length(sort(wfish3)))2 +
( ((sum(round(cos(sort(wfish3)), 10)))/length(sort(wfish3)))2))
wfish3←func1(sort(wfish3))
#Symmetric-Arc
SA1←sum(SA1)
rLSA1←sqrt((((sum(round(sin(sort(wSA1)), 10)))/length(sort(wSA1)))2 +
( ((sum(round(cos(sort(wSA1)), 10)))/length(sort(wSA1)))2))
wSA1←func1(sort(wSA1))
SA2←sum(sort(SA2))
rLSA2←sqrt((((sum(round(sin(sort(wSA2)), 10)))/length(sort(wSA2)))2 +
( ((sum(round(cos(sort(wSA2)), 10)))/length(sort(wSA2)))2))
wSA2←func1(sort(wSA2))
SA3←sum(sort(SA3))
rLSA3←sqrt((((sum(round(sin(sort(wSA3)), 10)))/length(sort(wSA3)))2 +
( ((sum(round(cos(sort(wSA3)), 10)))/length(sort(wSA3)))2))
wSA3←func1(sort(wSA3))
# Equal-Tailed Arc
ET1←sum(ET1)
rLET1←sqrt((((sum(round(sin(sort(wET1)), 10)))/length(sort(wET1)))2 +
( ((sum(round(cos(sort(wET1)), 10)))/length(sort(wET1)))2))
wET1←func1(sort(wET1))
ET2←sum(sort(LB2))
rLET2←sqrt((((sum(round(sin(sort(wET2)), 10)))/length(sort(wET2)))2 +
( ((sum(round(cos(sort(wET2)), 10)))/length(sort(wET2)))2))
wET2←func1(sort(wET2))
ET3←sum(sort(ET3))
rLET3←sqrt((((sum(round(sin(sort(wET3)), 10)))/length(sort(wET3)))2 +
( ((sum(round(cos(sort(wET3)), 10)))/length(sort(wET3)))2))

```

```

wET3←func1(sort(wET3))
#Likelihood-Based Arc
LB1←sum(LB1)
rLLB1←sqrt((((sum(round(sin(sort(wLB1)), 10)))/length(sort(wLB1)))2 +
( ((sum(round(cos(sort(wLB1)), 10)))/length(sort(wLB1)))2))
wLB1←func1(sort(wLB1))
LB2←sum(sort(LB2))
rLLB2←sqrt((((sum(round(sin(sort(wLB2)), 10)))/length(sort(wLB2)))2 +
( ((sum(round(cos(sort(wLB2)), 10)))/length(sort(wLB2)))2))
wLB2←func1(sort(wLB2))
LB3←sum(sort(LB3))
rLLB3←sqrt((((sum(round(sin(sort(wLB3)), 10)))/length(sort(wLB3)))2 +
( ((sum(round(cos(sort(wLB3)), 10)))/length(sort(wLB3)))2))
wLB3←func1(sort(wLB3))
return(fish1,wfish1,rLfish1,fish2,wfish2,rLfish2,fish3,wfish3,rLfish3,
SA1,wSA1,rLSA1,SA2,wSA2,rLSA2,SA3, wSA3,rLSA3,
ET1,wET1,rLET1,ET2,wET2,rLET2,ET3,wET3,rLET3,
LB1,wLB1,rLLB1,LB2,wLB2,rLLB2,LB3,wLB3,rLLB3)}

```

E.32 bootfunc2()

This function plots the confidence intervals

```

bootfunc2←function(x){
lenx ← length(x)
#Draws circle
plot(cos(0:360/180 * pi), sin(0:360/180 * pi), xlim = c(-1.5,1.5), ylim = c(-1.5,1.5), xlab= "
", ylab = " ", type = "l", axes= F)
new.ang ← ((2 * ceiling((90 * x)/pi) - 1) * pi)/180

```

```
s.ang ←sort(new.ang)
l ←length(x)
points(0.95 * cos(s.ang[1]), 0.95 * sin(s.ang[1]), pch = "o") dist ←0.95
for(i in 2:l) { if(s.ang[i] == s.ang[i - 1]) { dist←dist - 0.05}
else { dist ←0.95 }
points(dist * cos(s.ang[i]), dist * sin(s.ang[i]), pch = "o")}}

# Used with bootfunc2() to plot confidence bands

normalsize points(1.25 * cos(pe), 1.25 * sin(pe), pch = 2)
lowline ←seq(round(lower, 2), round(pe, 2), 0.01)
uppline ← seq(round(pe, 2), round(upper, 2), 0.01)
lines(1.25 * cos(lowline), 1.25 * sin(lowline), type = "l", lty = 1)
lines(1.25 * cos(uppline), 1.25 * sin(uppline), type = "l", lty = 1)
```

Vita

BENNETT SANGO OTIENO

Bennett Sango Otieno was born on April 22, 1967, Busia, Kenya. His academic achievements include a B.S.(1990) in Mathematics, an M.S.(1994) in Statistics , both from Egerton University, Njoro, Kenya, and Ph.D.(2002) in Statistics from Virginia Tech. Upon graduating from Virginia Tech, Sango Otieno will continue his academic career as a Visiting Assistant Professor of Mathematics at The College of Wooster in Wooster, Ohio, starting August, 2002.