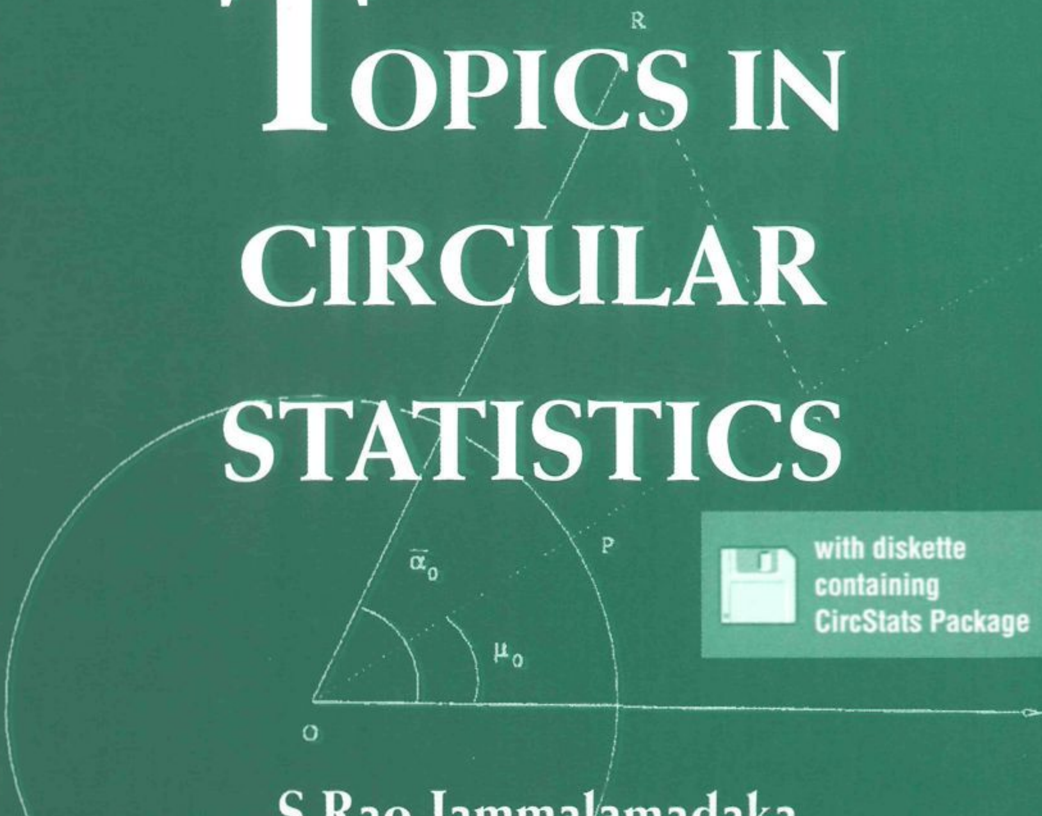
Series on

Multivariate Analysis

• Vol. 5 •

TOPICS IN CIRCULAR STATISTICS



with diskette containing CircStats Package

S Rao Jammalamadaka

A SenGupta

World Scientific

TOPICS IN CIRCULAR STATISTICS

SERIES ON MULTIVARIATE ANALYSIS

Editor: M M Rao

Published

Vol. 1: Martingales and Stochastic Analysis
J. Yeh

Vol. 2: Multidimensional Second Order Stochastic Processes
Y. Kakihara

Vol. 3: Mathematical Methods in Sample Surveys
H. G. Tucker

Vol. 4: Abstract Methods in Information Theory
Y. Kakihara

Vol. 5: Topics in Circular Statistics
S. R. Jammalamadaka and A. SenGupta

Forthcoming

Convolution Structures and Stochastic Processes
R. Lasser

Series on Multivariate Analysis • Vol. 5 •

TOPICS IN CIRCULAR STATISTICS

S Rao Jammalamadaka

Department of Statistics and Applied Probability
University of California, Santa Barbara
USA

A SenGupta

Applied Statistics Unit
Indian Statistical Institute
India

World Scientific
Singapore • New Jersey • London • Hong Kong

Published by

World Scientific Publishing Co. Pte. Ltd.

P O Box 128, Farrer Road, Singapore 912805

USA office: Suite 1B, 1060 Main Street, River Edge, NJ 07661

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

Library of Congress Cataloging-in-Publication Data

Jammalamadaka, S. Rao.

Topics in circular statistics / S. Rao Jammalamadaka, A. SenGupta.

p. cm. -- (Series on multivariate analysis ; vol. 5)

Includes bibliographical references and indexes.

ISBN 9810237782

1. Circular data. 2. Mathematical statistics. I. SenGupta, Ambar, 1963– II. Title. III. Series.

QA276 .J36 2001

519.5--dc21 2001024071

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

Printed in Singapore by Uto-Print

Preface

This research monograph on circular data analysis covers recent theoretical advances in the field after providing an introduction to this novel area. No attempt has been made to cover spherical models or their analysis although there are many essential common features with circular data analysis. The first seven chapters provide an up-to-date, if not exhaustive, coverage of circular modeling and analysis. The focus then shifts to more recent research in this field in such areas as change-point problems, predictive distributions, circular correlation and regression, classification and discriminant analysis, etc., highlighting mainly the authors' own research interests and that of their co-workers. The statistical methods in this area, as in many others, evolved as a series of practical and sometimes ad-hoc tools. Whenever possible, we attempt to analyze these methods in the framework of optimal statistical inference, as for example, when we examine the circular normal distribution as a member of the curved exponential family. Besides learning about the novelty of circular data analysis and the essential techniques, we hope a careful reader can pick up possible topics for future research on topics such as optimal properties of procedures, modeling, and robustness. As stated before, it is not our intent to provide comprehensive coverage of all the tools and techniques in the field but to give an introduction to the area and describe some current research.

Along with such theoretical discussions, we also provide a package of computational subroutines in `SPlus`, called `CircStats`, which allows users to analyze their own data sets. This important feature should make this monograph very useful to practitioners of circular data analysis.

Chapter 1 starts with an introduction to the field while Chapter 2 provides a discussion of circular models and methods for generating them. Sampling distributions are covered in Chapter 3 and parametric estimation and testing in Chapters 4, 5, and 6. Chapter 7 covers nonparametric methods and

Chapter 8 discusses circular correlation and regression. Recent research on predictive densities, outliers, and change point problems are covered in the next three chapters, while the final chapter reviews some important miscellaneous topics.

We have had considerable help in completing this work and we would like to thank Dr. Kaushik Ghosh, A. Laha, C. Pal, Debashis Paul, Huaxin You, and Anna Valeva for helping with LaTeX and Dr. Saralees Nadarajah for reading the manuscript. The package of computer routines, `CircStats` based on `SPlus`, is the work of Dr. Ulric Lund whom we would like to thank for letting us incorporate it as part of this monograph.

Contents

This page is intentionally left blank

Chapter 1

Introduction

1.1 Introduction

In many diverse scientific fields, the measurements are directions. For instance, a biologist may be measuring the direction of flight of a bird or the orientation of an animal, while a geologist may be interested in the direction of the earth's magnetic pole. Such directions may be in two dimensions as in the first two examples or in three dimensions like the last one. A set of such observations on directions is referred to as *directional data.*

Two-dimensional directions can be represented as angles measured with respect to some suitably chosen "zero direction", i.e., the starting point and a "sense of rotation", i.e., whether clockwise or anti-clockwise, is taken as the positive direction. Since a direction has no magnitude, these can be conveniently represented as points on the circumference of a unit circle centered at the origin or as unit vectors connecting the origin to these points. Because of this circular representation, observations on such two-dimensional directions are also called *circular data.* Similarly, directions in three dimensions may be represented by two angles (akin to the representation of points on the earth's surface by their longitude and latitude), as unit vectors in three dimensions, or as points on the surface of a unit sphere. Because of this, directional data in three dimensions are also referred to as *spherical data.*

Directional data have many unique and novel features both in terms of modeling and in their statistical treatment. For instance, the numerical representation of a two-dimensional direction as an angle or a unit vector is not necessarily unique since the angular value depends on the choice of what is labeled as the zero-direction and the sense of rotation. What is considered 60^0 by a mathematician who takes true East as the zero-direction and anti-clockwise as the positive direction comes out to be 30^0 to a Geologist who takes true North as the zero and clockwise as the positive direction, called the "azimuth".

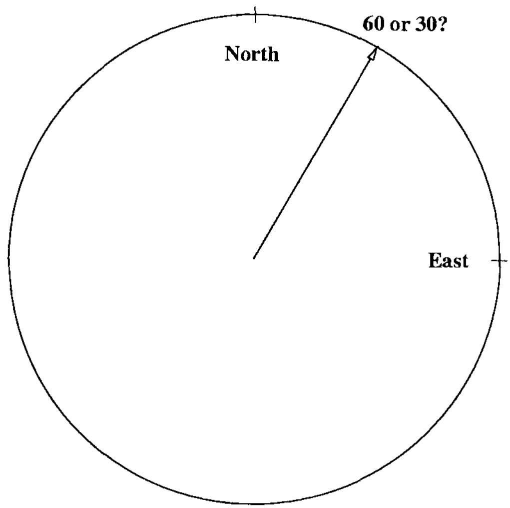


Figure 1.1: Value depends on choice of origin and sense of rotation.

The same is true of the specific values assigned to any spherical data set. It is therefore 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. That is, we should aim at conclusions that do not depend on the arbitrary choice of origin and sense of rotation. Again, because of this arbitrariness, there is also no natural ordering or ranking of the observations since whether one direction is "larger" than the other depends on whether clockwise or anti-clockwise is treated as being the positive direction as well as where the "zero" is. This makes rank-based methods essentially inapplicable (see Chapter 7). Finally, since the "beginning" coincides with the "end" i.e., $0 = 2\pi$ and in general the measurement is periodic with θ being the same as $\theta + p.2\pi$ for any integer p, methods for dealing with directional data should take careful note of how to measure the distance between any two points.

Such distinctive features make directional analysis substantially different from the standard "linear" statistical analysis of univariate or multivariate data that one finds in most statistics books. The need for "invariance" of statistical methods and measures with respect to the choice of this arbitrary zero-direction and sense of rotation, makes many of the usual linear techniques and measures often misleading, if not entirely meaningless. Commonly used summary measures on the real line, such as the sample mean and variance turn out to be inappropriate as do all the moments and cumulants. See Examples 1.1 and 1.2 later. Analytical tools such as the moment generating function and other generating functions are equally useless. Many notions such as correlation and regression as well as their statistical measures, need to be re-invented for directional data (see Chapter 8). Similarly, such ideas of statistical inference as unbiasedness, loss functions, variance bounds, monotonicity of power functions of tests, etc., need to be redefined with caution.

Some commonly used parametric models for directional data are described in this book. However, much like for the linear case, a large part of parametric statistical inference for directional data, is derived based on just 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. Even modeling asymmetric data sets, which frequently occur in practice, provides some challenges because of the paucity of appropriate models. All in all, this area of directional data provides an inquisitive reader with many open research problems and is a fertile area for developing new statistical methods and inferential tools. There is also an opportunity to develop new and novel "applications" to problems arising in Natural, Physical, Medical as well as Social Sciences.

In this monograph, our goal is to make the readers aware of the limitations of the standard linear statistical methods, to present some basic ideas about this new area of statistics and to discuss some of the newer methodologies. We hope that the applied scientists will gain sufficient knowledge to enable them to correctly model, analyze and make inferences on the directional data problems that they come across. Since solutions to many directional data problems are non-trivial and often not obtainable in simple closed analytical forms, related computer software is essential for practitioners to be able to use these methods. The `CircStats` package included with this book has been used to analyze data sets and illustrative examples and sample outputs have been provided throughout the book. Built using the popular software `SPlus`, this package will make the statistical procedures and tools described here readily accessible to such users. Finally, we believe that there is enough material and discussion here to challenge the more inquisitive and theoretically oriented reader to investigate the many open problems that exist in this field.

Other significant books in this field include those by Mardia (1972), Batschelet (1981), Fisher (1993), and Mardia and Jupp (2000). Readers interested in directions in 3 dimensions or spherical data should consult the books by Fisher et al. (1987) and Watson (1983b)

1.2 Applications and Background

1.2.1 Some Examples

Directional data in two and/or three dimensions, arise quite frequently in many natural and physical sciences like Biology, Medicine, Ecology, Geology etc. For instance, biologists studying bird-migrations record the flight directions of just-released birds as they disappear over the horizon, while some others record the directional movements or orientations of certain biological organisms. In an experiment on homing pigeons, Schmidt-Koenig (1963) obtained the following data set on the vanishing angles of 15 birds released singly (arranged in increasing order):

$$85, 135, 135, 140, 145, 150, 150, 150, 160, 185, 200, 210, 220, 225, 270.$$

The data are represented as points on a circle in Figure 1.2. Table 1.1 gives data on orientations of turtles after laying eggs. See Batschelet (1981) for a very nice account of the applications of circular statistics to Biology.

Jammalamadaka et al. (1986) discuss an interesting medical application where the angle of knee flexion was measured to assess the recovery of orthopaedic patients. Ecologists consider the prevailing wind direction as an important factor in many studies including those which involve pollutant transport. Geologists study paleocurrents to infer about the direction of flow of rivers in the past (see Sengupta and Rao (1967)) and analyze paleomagnetic directions of the earth's magnetic pole to investigate the phenomenon of pole-reversal as well as in support of the hypothesis of continental drift. See e.g., Fuller et al. (1996).

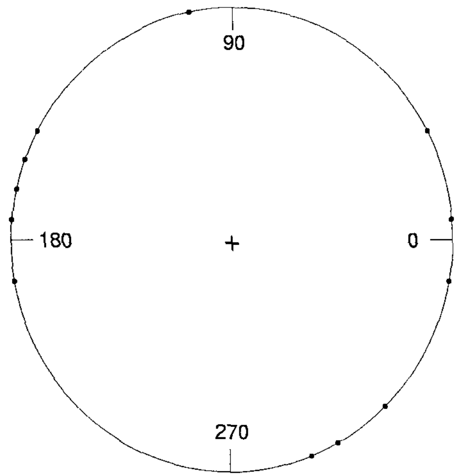


Figure 1.2: Vanishing angles of pigeons released in Schmidt-Koenig experiment.

Table 1.1: Orientations of 76 turtles after laying eggs (Gould's data cited by Stephens, 1969).

Direction (in degrees) clockwise from north									
8	9	13	13	14	18	22	27	30	34
38	38	40	44	45	47	48	48	48	48
50	53	56	57	58	58	61	63	64	64
64	65	65	68	70	73	78	78	78	83
83	88	88	88	90	92	92	93	95	96
98	100	103	106	113	118	138	153	153	155
204	215	223	226	237	238	243	244	250	251
257	268	285	319	343	350				

Also, any periodic phenomenon with a known period say a day, a month or a year, can be represented on the circle where the circumference corresponds to this period, by aggregating as necessary, data over several individuals or periods. Examples include arrival times of patients say with heart attacks at a hospital over the day, or the timing of breast cancer surgery within the menstrual cycle. As another example, the circle may be taken to represent the 365 days in the year and one could plot the occurrence of airplane accidents to see if they are uniformly distributed over the different seasons of the year.

In von Mises (1918), the hypothesis of interest was whether the atomic weights of elements known till then, were indeed integers, subject to measurement errors. The fractional parts of the atomic weights of the 24 lightest elements, converted to angles, are shown in Table 1.2, giving a distribution on the circle.

Table 1.2: Fractional parts of the atomic weights (as known in 1918) of the 24 lightest elements

Fractional part (in degrees)	Frequency
0	12
3.6	1
36	6
72	1
108	2
169.2	1
324	1

In studying circadian or other rhythms, the circle may be used to represent one cycle, and the interest may lie in the *timing* of an event within this cycle, say for instance when the body temperature or the blood pressure peaks within the day. Thus certain aspects in the study of biological rhythms, provide an important example from biology which can be put into this circular data framework. Because biological rhythms control characteristics such as sleep-wake cycles, hormonal pulsatility, body temperature, mental alertness, reproductive cycles and so on, there has been a renewed interest among the medical professionals in such topics as chronobiology, chronotherapy and the study of biological clock. See the books "Chronobiology and Chronomedicine" by Morgan (1990) and "Circadian Cancer Therapy" by

Table 1.3: Frequency Distributions of Cross-bedding Azimuths in the 3 Units of Kamthi River (Sengupta and Rao, 1966)

Azimuth (in degrees)	Lower Kamthi	Middle Kamthi	Upper Kamthi
0—19	14	50	75
20—39	14	62	75
40—59	11	33	15
60—79	13	9	25
80—99	9	1	7
100—119	16	3	3
120—139	0	0	3
140—159	4	0	0
160—179	0	0	0
180—199	3	0	0
200—219	4	2	21
220—239	0	8	8
240—259	0	0	24
260—279	0	11	16
280—299	6	5	36
300—319	7	20	75
320—339	1	53	90
340—359	21	41	107

Hrushesky (1994). Improved modeling and analysis in this important area is crucial not only for inference and prediction but also in formulating important health policies.

1.2.2 The Need for Appropriate Analysis

As we know, statistical analysis – modeling and inference, forms the basis for objective generalizations from observed data. As the above examples illustrate, directional data occur in many scientific studies and calls for the use of specialized statistical tools and techniques, which are not yet widely known and appreciated. Because of this unfamiliarity, applied scientists including biologists, geologists, social and behavioral scientists dealing with such data have sometimes fallen into the trap of using the more common but inappro-

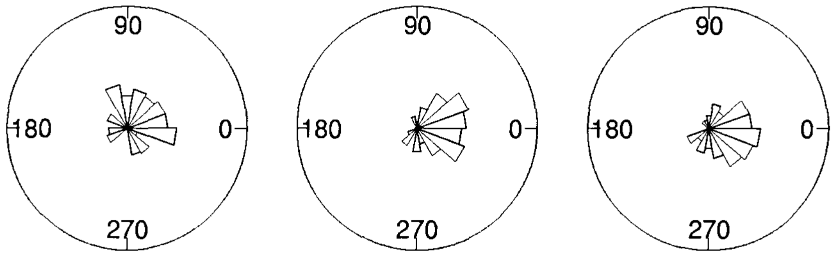


Figure 1.3: Rose diagrams for the three units (lower, middle and upper) of Kamthi river.

priate linear methods like computing the sample mean, the sample variance, and applying the t-test or doing a simple analysis of variance.

In an emerging area called *chronotherapeutics*, proper use of circular statistical methods could lead to improved understanding of the rhythms as well as factors that influence these rhythms. However many scientific studies do not undertake a full statistical analysis, depending instead, on simple graphical or other descriptive tools.

One has to distinguish time as it occurs in the usual time-series analysis which is a linear variable, as distinguished from situations where one is considering timing within a cycle, which is to be treated as a circular variable. This is a significant distinction, since, in the latter case, it is inappropriate to use sample means and variances as one would do in a standard regression or time-series analysis. In a recent study, 'days from menstrual period' was used as an explanatory variable. However, one should note that in this context, 2 days (near the beginning of the cycle) is closer to 25 days (near the end of the cycle) than it is to say 9 days, in the circular sense and hence, a straightforward analysis using packages will give unreliable and misleading results.

If the response variable is circular, like the 'time of day when a drug reaches its peak effectiveness' or when maximum blood concentration is attained, it is clear that one has to make judicious use of circular statistics. For example, if one wants to model the time of peak effectiveness as a function of time of administering the drug and possibly various other linear variables including dosage etc., we have a circular response variable and a circular explanatory variable and one should use what is known as *circular regression* as discussed in Chapter 8. As a simple control problem, it would then be possible to use such a model and the knowledge of the rhythmic pattern to devise an optimal treatment for controlling an individual patient's blood pressure. For example, this can be done by controlling the time of administering the drug and/or dosage so as to maximize the chance that the time at which the peak effectiveness of the drug (response variable) occurs, matches the time when the patient has peak blood pressure.

1.3 Descriptive Statistics

As pointed out earlier, circular data which is the main focus of this monograph, can be represented as angles or as points on the circumference of a unit circle. The directional position may be uniquely determined by two co-ordinates. For this purpose, we can use the rectangular co-ordinate system with origin O and two (or three) perpendicular axes X, Y through O. Any point P on the plane can be represented as (X, Y) in terms of its rectangular co-ordinates or as (r, α) in terms of its polar co-ordinates where r is the distance to the origin and α - its direction. For the exceptional point, namely the origin O, $r = 0$ and no direction is indicated i.e., α is not defined.

It is easy to convert the polar co-ordinates into rectangular co-ordinates and vice versa. This is done by means of the trigonometric functions *sine* and *cosine*. Take a point P with polar co-ordinates (r, α).

From Figure 1.4, the rectangular co-ordinates of point P are given by

$$x = r\cos\alpha, \quad y = r\sin\alpha. \tag{1.3.1}$$

As we said before, in directional analysis we are interested in the direction and not in the magnitude of the vector and therefore we take these vectors to be of unit length (i.e., $r = 1$) for convenience. Each direction thus corresponds to a point P on the circumference of the unit circle. Alternatively, this point on the circumference of a unit circle can be specified just by the angle. Now,

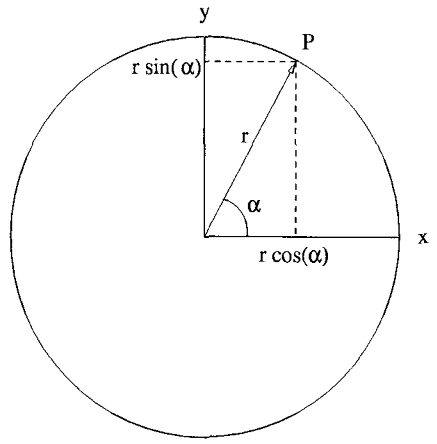


Figure 1.4: Relation between rectangular and polar co-ordinates.

if the point P lies on the circumference of the unit circle, the conversion between polar and rectangular coordinates is simply

$$(1,\ \alpha) \Leftrightarrow (x = \cos\alpha,\ y = \sin\alpha). \tag{1.3.2}$$

Remark 1.1 *Although we write $\{\mathbf{u}_i\} = \{(x_i,\ y_i)\}$, such data cannot be considered as bivariate data on the plane, since x_i's and y_i's are related through the equation $x_i^2 + y_i^2 = 1\ \forall i$. These points are thus all restricted to be on the circumference of a circle, which is of Lebesgue measure (area) zero on the plane.*

Remark 1.2 *Three-dimensional directions can also be similarly represented in several equivalent ways : as points on the surface of a unit sphere centered at the origin, as a vector $(X,\ Y,\ Z)$ of length one, or in terms of two angles $(\alpha_1,\ \alpha_2)$ in the range $[0,\pi)\times[0,2\pi)$. In geographical terms, these 2 angles are called the latitude and longitude. Again, one can go back and forth between the rectangular co-ordinates $(X,\ Y,\ Z)$ and the polar co-ordinates $(1,\ \alpha_1,\ \alpha_2)$ by using the relationship*

$$(X,\ Y,\ Z) = (\ \cos\alpha_1,\ \sin\alpha_1\cos\alpha_2,\ \sin\alpha_1\sin\alpha_2\ ).$$

1.3.1 Measure of Center

In order to define a mean direction – sometimes also called the "preferred direction", for a given unimodal circular data set, one might be tempted to calculate the arithmetic mean of the angles. As mentioned earlier, the sample mean as well as standard deviation and other higher moments, suffer from their strong dependence on the choice of zero direction and the sense of rotation. The next 2 examples illustrate why they are inappropriate as descriptive measures of the center, for circular data.

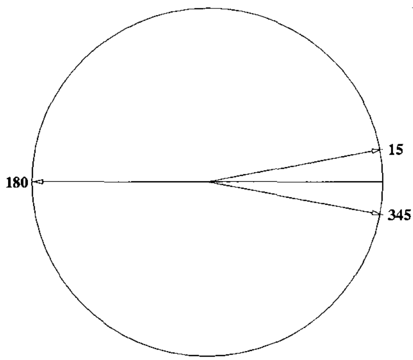


Figure 1.5: The arithmetic mean points the wrong way!

Example 1.1 *Measured in the standard way, (with East as the zero direction and anti-clockwise as the positive sense of rotation), suppose two birds flew at* 15^0 *and* 345^0*. Their arithmetic mean, which is* 180^0*, points due West whereas the two observations essentially point towards East. See Figure 1.5.*

Example 1.2 *Assume that an angular data set consists of the four directions* 50^0, 160^0, 210^0 *and* 300^0 *measured in the usual way (East is zero with the counter-clockwise rotation being positive in Figure 1.6). Their arithmetic mean is* 180^0*, pointing that their mean direction is due West.*

On the other hand, if the same four directions were measured as angles in the range -180^0 *to* $+180^0$ *(with the same zero direction and counter-clockwise as positive rotation), their values become* -50^0, -160^0, 150^0 *and* 60^0*. Their arithmetic mean now becomes* 0^0 *i.e., due East, which is the exact opposite to the previous mean which pointed due West. Assuming next that North*

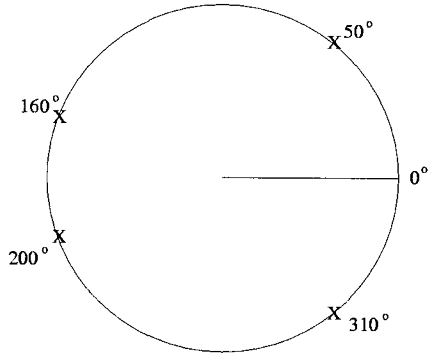


Figure 1.6: The value of the sample mean depends on the choice of origin.

is zero with the counter-clockwise sense of rotation being positive, the same four directions take values 40^0, 150^0, 240^0 *and* 290^0. *Their arithmetic mean now is* 180^0 *i.e., due South. Finally, measuring the same four directions using angles ranging from* -180^0 *to* $+180^0$ *again with North as zero and the counter-clockwise rotation as positive, gives due North as their arithmetic mean. Thus the usual arithmetic mean of the sample can point to E, W, N or S (and things in between) for the same data, depending on the choice of origin and sense of rotation.*

From these examples, it is clear that the arithmetic mean, which is commonly used for linear data, is not so much a measure of the 'center' for a given set of observed directions, as it is a function of the choice of the zero-direction and sense of rotation. Hence, it should be avoided as a measure of center for directions. The sample variance s^2, which depends on the sample mean, also suffers from the same problem and thus one needs alternate measures of center and dispersion when dealing with circular data.

An appropriate and meaningful measure of the mean direction for a set of directions which are unimodal i.e., point or concentrate towards a single direction, is obtained by treating the data as unit vectors and using the *direction of their resultant vector.* Computation of the direction of the resultant vector for a given angular data set, is done as follows :

Let $\alpha_1, \alpha_2, \ldots, \alpha_n$ be a set of circular observations given in terms of angles. Consider the polar to rectangular transformation for each observation (see 1.3.2), i.e.

$$(\cos\alpha_i,\ \sin\alpha_i),\ i = 1,\ \ldots,\ n.$$

We obtain the resultant vector of these n unit vectors by summing them component-wise, to get

$$\mathbf{R} = \left(\sum_{i=1}^{n}\cos\alpha_i,\ \sum_{i=1}^{n}\sin\alpha_i\right) = (C,\ S),\ \text{say.} \tag{1.3.3}$$

Let

$$R = \|\mathbf{R}\| = \sqrt{C^2 + S^2}$$

represent the length of the resultant vector $\mathbf{R}$. The direction of this resultant vector $\mathbf{R}$, which is proposed as the circular mean direction, is denoted by $\overline{\alpha}_0$, and is defined as

$$\overline{\alpha}_0 = \arg\{\sum_{j=1}^{n}\cos\alpha_j + i\sum_{j=1}^{n}\sin\alpha_j\}$$

or by the equations

$$\cos\overline{\alpha}_0 = \frac{C}{R},\quad \sin\overline{\alpha}_0 = \frac{S}{R}. \tag{1.3.4}$$

More explicitly, it is given by the "quadrant-specific" inverse of the tangent that we shall refer to from now on as,

$$\overline{\alpha}_0 = arctan^*(S/C),$$

where

$$\overline{\alpha}_0 = arctan^*(S/C) = \begin{cases} arctan(S/C), & \text{if } C > 0,\ S \geq 0, \\ \pi/2, & \text{if } C = 0,\ S > 0, \\ arctan(S/C) + \pi, & \text{if } C < 0, \\ arctan(S/C) + 2\pi & \text{if } C \geq 0,\ S < 0, \\ \text{undefined}, & \text{if } C = 0,\ S = 0. \end{cases} \tag{1.3.5}$$

The somewhat complicated looking definition of a quadrant-specific inverse of the tangent is necessitated by the fact that $tan(\theta) = tan\ (\theta + \pi)$, so that there are 2 inverses for any given θ. Since *arctan* is usually defined so as to take values in $(-\frac{\pi}{2}, +\frac{\pi}{2})$, Definition 1.3.5, provides us the correct unique inverse on $[0, 2\pi)$, which takes into account the signs of C and S. The following proposition shows that such an $\overline{\alpha}_0$ reflects the center for the data set and does not depend on the choice of origin or the sense of rotation.

Proposition 1.1 *$\overline{\alpha}_0$ is rotationally equivariant, i.e., if the data is shifted by a certain amount, the value of $\overline{\alpha}_0$ also changes by the same amount.*

Proof: Let $(\alpha_1,\ \alpha_2,\ \ldots,\ \alpha_n)$ have mean direction $\overline{\alpha}_0$. We will show that $(\alpha_1 + c,\ \alpha_2 + c,\ \ldots,\ \alpha_n + c)$ have mean direction $\overline{\alpha}_0 + c$.

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

$$\mathbf{R}' = \left(\sum_{i=1}^{n} \cos(\alpha_i + c),\ \sum_{i=1}^{n} \sin(\alpha_i + c)\right) = (C',\ S') \quad \text{(say)}.$$

Then,

$$\begin{aligned}
C' &= \sum_{i=1}^{n} \cos(\alpha_i + c) \\
&= \sum_{i=1}^{n} (\cos\alpha_i \cos c - \sin\alpha_i \sin c) \\
&= C\cos c - S\sin c \quad \text{(from Equation (1.3.3))} \\
&= R\cos\overline{\alpha}_0 \cos c - R\sin\overline{\alpha}_0 \sin c \quad \text{(from Equation (1.3.4))} \\
&= R\cos(\overline{\alpha}_0 + c).
\end{aligned}$$

Similarly, $S' = R\sin(\overline{\alpha}_0 + c)$. Now,

$$R' = \|\mathbf{R}'\| = \sqrt{C'^2 + S'^2} = R = \sqrt{C^2 + S^2}.$$

Hence,

$$\frac{C'}{R'} = \cos(\overline{\alpha}_0 + c),\ \frac{S'}{R'} = \sin(\overline{\alpha}_0 + c),$$

proving the result. □

Similarly one can check that $\overline{\alpha}_0$ is equivariant with respect to changes in the sense of rotation, i.e., when we switch from clockwise to anticlockwise so that α's become $(2\pi - \alpha)$'s, then $\overline{\alpha}_0$ becomes $(2\pi - \overline{\alpha}_0)$.

To summarize, for a given set of angles, first we transform them into rectangular co-ordinates i.e., find the *cosine* and *sine* of each angle. These are added component-wise to find the resultant vector, using Equation (1.3.3). Then we obtain the direction of this resultant vector, namely $\overline{\alpha}_0$, using Equation (1.3.5). This represents the circular mean direction for the given set of angles. This can be computed using the program `circ.mean` in the package `CircStats`. Clearly, when $\mathbf{R} = \mathbf{0}$, i.e., when both $C = 0$ and $S = 0$, a circular mean cannot be defined, so that formally

Definition 1.1 *If the resultant vector is of positive length, its direction* $\overline{\alpha}_0$, *is taken as the circular mean direction. We say that no mean direction exists if the resultant vector is of length* 0.

The case when the resultant vector has zero length corresponds to the situation where the data is spread evenly or uniformly over the circle, with no concentration towards any direction. In this case, it is clear that the data do not indicate any preferred or mean direction.

1.3.2 Circular Distance and Measure of Dispersion

Not only does the direction $\overline{\alpha}_0$ of the vector resultant $\mathbf{R}$ provide a mean direction as we just observed, but its length

$$R = \|\mathbf{R}\|$$

is a useful measure for unimodal data, of how *concentrated* the data is towards this center. This can be seen easily since if all the angles (unit vectors) point in the same direction indicating large concentration, R can be as large as n. Conversely, if the data is evenly spread over the circle indicating no concentration, R can be as small as zero. We now show with an appropriate "distance" measure on the circle, $(n - R)$ is indeed the right analogue of the usual sample variance.

One reasonable measure of "circular distance" between any two points is to take the *smaller of the two arclengths between the points* along the circumference, i.e., for any two angles α and β

$$d_0(\alpha, \beta) = \min(\alpha - \beta, 2\pi - (\alpha - \beta)) = \pi - \mid \pi - \mid \alpha - \beta \mid\mid . \qquad (1.3.6)$$

For example, the distance between A and B can be the length of the arc ANB or the arclength ASB (see Figure 1.7). But since the arc ANB is shorter than the arc ASB, we define the circular distance to be the arclength ANB. Clearly no two points on the circumference of a circle can be farther than π or the circular distance (1.3.6) always lies in $[0, \pi]$.

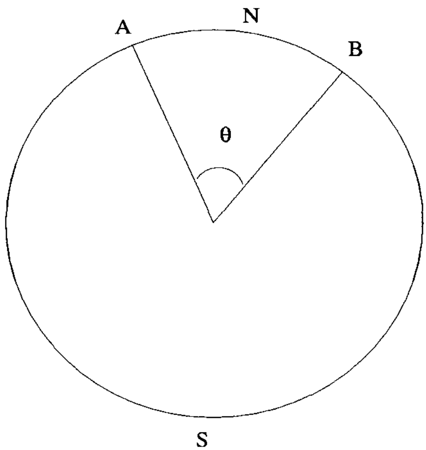


Figure 1.7: Circular 'distance' d_0 is the arclength ANB.

Another closely related definition of "circular distance" between points A, B in Figure 1.7, is given by

$$d(\alpha,\ \beta) = (1 - \cos(\alpha - \beta)), \tag{1.3.7}$$

where α and β represent the angles corresponding to A, B respectively. If θ is the angle between the points A and B, this is clearly a monotone increasing function of θ, taking the value 0 when $\theta = 0$ and increasing to 2 when $\theta = \pi$.

Proposition 1.2 below justifies $(n - R)$ as a measure of dispersion, on grounds similar to those in choosing the sample variance s^2 for the linear case. Representing the observations as unit vectors, $\{\mathbf{u}_i,\ i = 1,\ \ldots,\ n\}$, let $D_{\mathbf{v}}(\mathbf{u}_1,\ \mathbf{u}_2,\ \ldots,\ \mathbf{u}_n)$ denote the sample dispersion with respect to some arbitrary unit vector $\mathbf{v} = (a,\ b)$. If θ_i is the angle between $\mathbf{u}_i$ and this arbitrary unit vector $\mathbf{v}$, (see Figure 1.8) $0 \leq \theta_i \leq \pi$, we have the circular distance $d(\mathbf{v},\ \mathbf{u}_i) = 1 - \cos\theta_i$, using Equation (1.3.7).

Now,

$$\begin{aligned} D_{\mathbf{v}}(\mathbf{u}_1,\ \mathbf{u}_2,\ \ldots,\ \mathbf{u}_n) &= \sum_{i=1}^{n} d(\mathbf{v},\ \mathbf{u}_i) \\ &= n - \sum_{i=1}^{n} \cos\theta_i. \end{aligned} \tag{1.3.8}$$

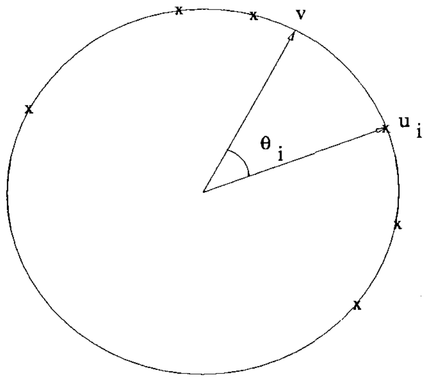


Figure 1.8: 'Circular distance' between **v** and observations **u**ᵢ on a circle.

Analogous to the linear case (where recall that the dispersion around a value c, viz. $\sum(x_i - c)^2$ is a minimum when $c = \overline{x}$ and this minimum value is used to define the sample variance), we can ask if there is a point on the circumference **v** such that $D_{\mathbf{v}}(\mathbf{u}_1,\ \mathbf{u}_2,\ \ldots,\ \mathbf{u}_n)$ is minimized.

Proposition 1.2 $D_{\mathbf{v}}(\mathbf{u}_1,\ \mathbf{u}_2,\ \ldots,\ \mathbf{u}_n)$ *is minimized when* **v** *is the normalized resultant vector*

$$\mathbf{v}^* = \left(\frac{C}{R},\quad \frac{S}{R}\right)$$

and the corresponding sample measure of dispersion is given by $(n - R)$.

Proof: Let

$$\mathbf{u}_i = (\cos\alpha_i,\ \sin\alpha_i) = (x_i,\ y_i),$$

say, and

$$\mathbf{v} = (a,\ b),$$

where $a^2 + b^2 = 1$. Then

$$d(\mathbf{v},\ \mathbf{u}_i) = 1 - \cos\theta_i = 1 - (\mathbf{v}\cdot\mathbf{u}_i) = 1 - (ax_i + by_i),$$

from 1.3.7. Here $(\mathbf{v}\cdot\mathbf{u}_i)$ is the dot product or the inner product of the vectors **v** and $\mathbf{u}_i$. Thus the problem is to minimize

$$D_{\mathbf{v}}(\mathbf{u}_1,\ \ldots,\ \mathbf{u}_n) = n - \sum_{i=1}^{n}\cos\theta_i,$$

with respect to the choice of $\mathbf{v}$. Since $a^2+b^2=1$, we have $2a(da)+2b(db)=0$ and

$$\frac{db}{da} = -\frac{a}{b}. \tag{1.3.9}$$

Differentiating $D_{\mathbf{v}}$ with respect to a and equating to zero, we get using Equation (1.3.9),

$$0 = \frac{dD_{\mathbf{v}}}{da} = -\sum_{i=1}^{n}\left\{x_i - \frac{a}{b}y_i\right\}.$$

Hence,

$$\frac{a}{b} = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} y_i} = \frac{C}{S}, \text{ say.}$$

Since $a^2+b^2=1$, we have,

$$\frac{C^2}{S^2} = \frac{a^2}{1-a^2},$$

which implies

$$a = \frac{C}{R}, \qquad b = \frac{S}{R}.$$

Hence, $D_{\mathbf{v}}$ is minimum when $\mathbf{v} = \mathbf{v}^* = (\frac{C}{R}, \frac{S}{R})$, the normalized resultant vector. And the circular sample dispersion with respect to this is given by

$$\begin{aligned}
D_{\mathbf{v}^*} &= \sum_{i=1}^{n} d(\mathbf{v}^*, \mathbf{u}_i) \\
&= n - \sum_{i=1}^{n}\left\{\frac{C}{R}x_i + \frac{S}{R}y_i\right\} \\
&= n - \left(\frac{C}{R}C + \frac{S}{R}S\right) \\
&= n - R.
\end{aligned}$$

□

A value of R near 0 means the dispersion is large whereas R values close to n implies that the set of observations has small dispersion or more concentration towards the center. Note that $(n-R)$ measures the dispersion of the sample relative to the center, viz., the sample mean direction. No population parameter is involved. However, if we are given the population mean direction μ_0, (also called the *polar direction*), then denote the "polar vector" by $\boldsymbol{P} = (\cos\mu_0, \ \sin\mu_0)$. Then it is natural to measure the dispersion around this known mean direction so that we get,

$$D_{\boldsymbol{P}} = n - \sum_{i=1}^{n} \cos(\alpha_i - \mu_0) = n - V_0,$$

where

$$\begin{aligned} V_0 &= \sum_{i=1}^{n} \{\cos\alpha_i \cos\mu_0 + \sin\alpha_i \sin\mu_0\} \\ &= X\cos\mu_0 + Y\sin\mu_0 \\ &= R\cos\overline{\alpha}_0\cos\mu_0 + R\sin\overline{\alpha}_0\sin\mu_0 \\ &= R\cos(\overline{\alpha}_0 - \mu_0) \\ &= R\cdot c \qquad (\text{where } c = \cos(\overline{\alpha}_0 - \mu_0)). \end{aligned} \tag{1.3.10}$$

This V_0, corresponding to the length OC in Figure 1.9, represents the length of the projection of the sample resultant vector towards the polar direction $\boldsymbol{P}$. Since $c = \cos(\overline{\alpha}_0 - \mu_0)$ is bounded by one, $V_0 \leq R$ and is *equal* if and only if $\overline{\alpha}_0 = \mu_0$. This result is also restated in the second half of the following Theorem.

The following result is worth noting:

Theorem 1.1 *If* $\overline{\alpha}_0$ *is the direction of the vector resultant of* $(\alpha_1, \alpha_2, ..., \alpha_n)$*, then*

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

and

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

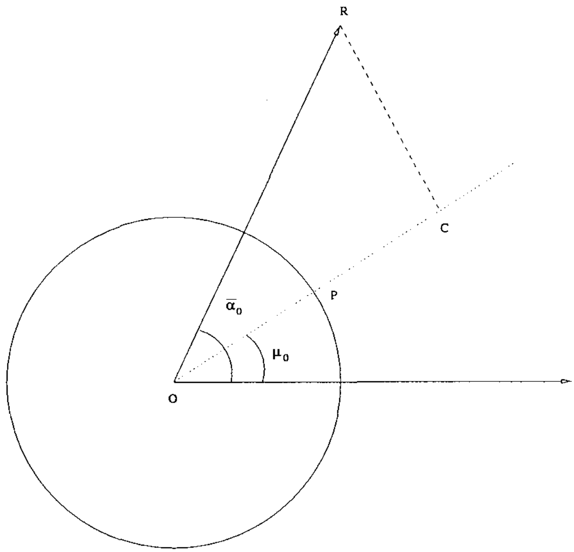


Figure 1.9: Projection of the resultant on the polar vector.

Proof: It follows that

$$\begin{aligned}\sum_{i=1}^{n}\sin(\alpha_i-\overline{\alpha}_0) &= \sum_{i=1}^{n}\{\sin\alpha_i\cos\overline{\alpha}_0-\cos\alpha_i\sin\overline{\alpha}_0\}\\ &= Y\cos\overline{\alpha}_0 - X\sin\overline{\alpha}_0\\ &= R\cos\overline{\alpha}_0\sin\overline{\alpha}_0 - R\cos\overline{\alpha}_0\sin\overline{\alpha}_0\\ &= 0.\end{aligned}$$

Similarly,

$$\begin{aligned}\sum_{i=1}^{n}\cos(\alpha_i-\overline{\alpha}_0) &= \sum_{i=1}^{n}\{\cos\alpha_i\cos\overline{\alpha}_0+\sin\alpha_i\sin\overline{\alpha}_0\}\\ &= X\cos\overline{\alpha}_0 + Y\sin\overline{\alpha}_0\\ &= R\cos^2\overline{\alpha}_0 + R\sin^2\overline{\alpha}_0\\ &= R.\end{aligned}$$

Remark 1.3 *These results can be put in context by drawing again some analogies to linear data, say* $(x_1, ..., x_n)$ *with sample mean* $\overline{x}$ *and population mean* μ*. Note that* $D_{\mathbf{v}^*} = (n - R)$ *corresponds to* $\sum(x_i - \overline{x})^2$ *while* $D_{\mathbf{P}} = (n - V)$ *corresponds to* $\sum(x_i - \mu)^2$*, the dispersion around the true mean. Theorem 1.1 reminds one of the fact that* $\sum(x_i - \overline{x}) = 0$ *whereas* $\sum(x_i - \overline{x})^2$ *relates to the sample variance.*

These results appear more intuitive and seem less mysterious if one notes that

$$\sin\theta \approx \theta \quad \text{and} \quad \cos\theta \approx (1 - \frac{\theta^2}{2}) \quad \text{for small } \theta. \tag{1.3.11}$$

Using second part of Theorem 1.1 in the first step and the Taylor expansion for $\cos\theta$ in the second step, we get

$$\begin{aligned} 2(1 - \overline{R}) &= \frac{2}{n}\sum_{i=1}^{n}(1 - \cos(\alpha_i - \overline{\alpha}_0)) \\ &\simeq \frac{1}{n}\sum_{i=1}^{n}(\alpha_i - \overline{\alpha}_0)^2 \\ &= s^2. \end{aligned} \tag{1.3.12}$$

This led some to define $2(1 - \overline{R})$ as a "circular variance." However it is preferable to stay with the original concentration measure R instead of trying to find its approximate linear equivalents.

1.3.3 Higher Moments

Recall our definition of C and S as the sums of the *cosine* and *sine* components. Instead, we could as well have considered their averages giving us

$$\overline{C}_n = \frac{1}{n}\sum_{j=1}^{n}\cos\alpha_j \quad \text{and} \quad \overline{S}_n = \frac{1}{n}\sum_{j=1}^{n}\sin\alpha_j.$$

Since $e^{i\alpha} = (\cos\alpha + i\sin\alpha)$, we note that

$$\overline{C}_n + i\overline{S}_n = \frac{1}{n}\sum_{j=1}^{n}e^{i\alpha_j}.$$

Similarly we can define the higher order sample moments by taking higher powers of $\{e^{i\alpha_j}\}$ and averaging these. However, since

$$e^{ip\alpha} = \cos p\alpha + i \sin p\alpha,$$

we have

$$\begin{aligned}
\frac{1}{n}\sum_{j=1}^{n}(e^{i\alpha_j})^p &= \frac{1}{n}\sum_{j=1}^{n} e^{ip\alpha_j} \\
&= \frac{1}{n}\sum_{j=1}^{n}\cos p\alpha_j + i\frac{1}{n}\sum_{j=1}^{n}\sin p\alpha_j \\
&= \overline{C}_n(p) + i\overline{S}_n(p), \text{ say, } p = 0,\ 1,\ 2,\ \ldots, \qquad (1.3.13)
\end{aligned}$$

where $\left(\overline{C}_n(p),\ \overline{S}_n(p)\right)$ are called the p^{th} order trigonometric moments based on the sample. They are indeed the empirical or sample analogs of the coefficients in the Fourier series expansion of the circular density, which are discussed in Equation (2.1.3).

Remark 1.4 Grouped data: *Occasionally data comes in a grouped form as in the example of Kamthi data (see Table 1.3). In this case, we make the usual assumption that all the observations in an interval, are at the mid-point of that interval. If for instance, the n original observations are grouped into k classes with i^{th} class having a mid-point of α_i and frequency f_i, $i = 1, \ldots, k$; $(\sum_{i=1}^{k} f_i = n)$, then we compute*

$$\overline{C}_n = \frac{1}{n}\sum_{i=1}^{k} f_i(\cos\alpha_i), \qquad \overline{S}_n = \frac{1}{n}\sum_{i=1}^{k} f_i(\sin\alpha_i)$$

and

$$\overline{R} = \sqrt{\overline{C}^2 + \overline{S}^2}.$$

Similarly higher trigonometric moments are calculated as

$$\overline{C}_n(p) = \frac{1}{n}\sum_{i=1}^{k} f_i(\cos p\alpha_i), \qquad \overline{S}_n(p) = \frac{1}{n}\sum_{i=1}^{k} f_i(\sin p\alpha_i).$$

For most practical purposes, no correction for grouping is required unless the grouping is very coarse, say with interval exceeding 45^0 in length.

Remark 1.5 Axial data: *Sometimes the angular data refers to the "axis", as for instance, the long axis of particles in sediments or the optical axis of a crystal, rather than a direction. Such observations of axes are referred to as* <u>*axial data*</u>. *In this case, any direction* θ *is identified with its opposite direction* $(\theta + \pi)$. *Such axial data are handled by the device of "doubling the angles", i.e., transforming* θ *to* (2θ) *which removes the directional ambiguity. For instance, the "mean axis" corresponding to the axial data* $\theta_1, \theta_2, \ldots, \theta_n$ *is obtained as follows: Take the mean direction* $\overline{\alpha}_0$ *of the doubled angles* $(\alpha_i = 2\theta_i, i = 1, 2, \ldots, n)$ *(as in 1.3.5), then take half of this, namely* $\frac{1}{2}\overline{\alpha}_0$ *or equivalently* $(\frac{1}{2}\overline{\alpha}_0 + 180^0)$. *See Equation (2.2.19) with* $\ell = 2$ *for a circular model appropriate for axial data.*

The following example illustrates how one can use the accompanying `CircStats` package to compute the sample mean, dispersion and the trigonometric moments.

Example 1.3 *Suppose the data consisted of* 51^0, 67^0, 40^0, 109^0, 31^0, 358^0 *and we wished to compute their circular mean and measures of dispersion as well as the first four sample trigonometric moments. We use the programs circ.mean, circ.disp, and trig.moment in the package* ***CircStats*** *with the following results:*

```
> x_c(51,67,40,109,31,358}
> circ.mean(rad(x))
[1] 0.8487044
> circ.summary(rad(x))
  n       rho  mean.dir
1 6 0.8364365 0.8487044
> trig.moment(rad(x),1)
      mu.p          rho.p     cos.p     sin.p
1 0.8487044 0.8364365 0.5528477 0.6276826
> trig.moment(rad(x),2)
      mu.p          rho.p        cos.p     sin.p
1 1.588121 0.4800428 -0.008316171 0.4799708
> trig.moment(rad(x),3)
     mu.p          rho.p        cos.p     sin.p
1 1.963919 0.2365641 -0.09062174 0.2185183
> trig.moment(rad(x),4)
     mu.p          rho.p      cos.p      sin.p
```

```
1 2.685556 0.2255761 -0.2025234 0.09934215
```

Thus,

$$\begin{aligned} \overline{C}(1) &= 0.55, & \overline{S}(1) &= 0.63, \\ \overline{C}(2) &= -0.01, & \overline{S}(2) &= 0.48, \\ \overline{C}(3) &= -0.09, & \overline{S}(3) &= 0.22, \\ \overline{C}(4) &= -0.20, & \overline{S}(4) &= 0.10. \end{aligned}$$

Chapter 2

Circular Probability Distributions

2.1 Introduction

A circular distribution is a probability distribution whose total probability is concentrated on the circumference of a unit circle. Since each point on the circumference represents a direction, such a distribution is a way of assigning probabilities to different directions or defining a directional distribution. The range of a circular random variable (rv) θ, measured in radians, may be taken to be $[0,\ 2\pi)$ or $[-\pi,\ \pi)$. Circular distributions are essentially of two types: they may be discrete — assigning probability masses only to a countable number of directions, or may be absolutely continuous (with respect to the Lebesgue measure on the circumference). In the latter case, a probability density function (pdf) $f(\theta)$ exists and has the following basic properties:

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

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

(iii) $f(\theta) = f(\theta + k \cdot 2\pi)$ for any integer k (i.e., f is periodic).

Figures 2.1 and 2.2 show the circular and linear representations of a continuous circular distribution, respectively.

As on the real line, such a distribution can also be described via its "characteristic function". However, since θ is a periodic r.v. having the

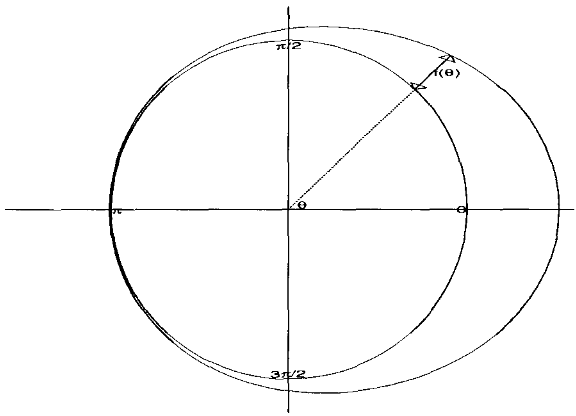


Figure 2.1: Circular representation of a continuous circular distribution.

same distribution as $(\theta + 2\pi)$, the characteristic function of such a r.v. has the property:

$$\varphi_\theta(t) \stackrel{def}{=} E(e^{it\theta}) = E(e^{it(\theta+2\pi)}) = e^{it2\pi} \cdot \varphi_\theta(t).$$

Hence, either $\varphi_\theta(t) = 0$ or $e^{it2\pi} = 1$ i.e., t must be an integer. Thus, for circular r.v.s, the characteristic function needs to be defined only at integer values. The value of the characteristic function at an integer p is also called the p^{th} *trigonometric moment* of θ. If $F(\theta)$ denotes the cdf of the r.v., we may write

$$\varphi_\theta(p) = E(e^{ip\theta}) = \int_0^{2\pi} e^{ip\theta} dF(\theta) = \rho_p e^{i\mu_p}, \qquad p = 0,\ \pm 1,\ \pm 2,\ \ldots. \tag{2.1.1}$$

Clearly $\varphi_0 = 1$, and

$$|\varphi_p| = \rho_p = \|E(e^{ip\theta})\| \leq E(\|e^{ip\theta}\|) = 1. \tag{2.1.2}$$

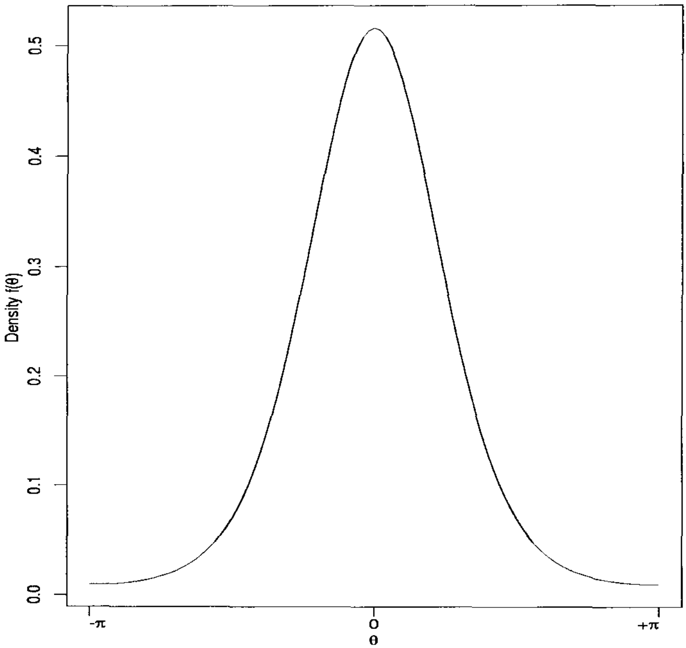


Figure 2.2: Linear representation of a continuous circular distribution.

Also $\overline{\varphi}_p$, the complex conjugate of φ_p, is φ_{-p}. One can also view these trigonometric moments in terms of

$$\alpha_p = E(\cos p\theta), \qquad \beta_p = E(\sin p\theta), \qquad p = 0,\ \pm1,\ \pm2,\ \ldots, \tag{2.1.3}$$

and these are related by the equation

$$\rho_p = \sqrt{\alpha_p^2 + \beta_p^2}, \qquad \text{and} \qquad \mu_p = arctan^* \left(\frac{\beta_p}{\alpha_p}\right).$$

In particular, consider the first trigonometric moment namely

$$\varphi_1 = \alpha_1 + i\beta_1 = \rho_1 \cdot e^{i\mu_1}, \tag{2.1.4}$$

which plays the most prominent role. We shall denote this length ρ_1, which lies between 0 and 1 (see Equation (2.1.2)), simply as ρ and the mean direction μ_1 as simply μ i.e.,

$$\rho_1 = \rho \quad \text{and} \quad \mu_1 = \mu.$$

The length ρ and when it is non-zero, the direction μ, of this first trigonometric moment φ_1 are used to provide "theoretical" or "population" measures of the concentration and the mean direction of θ, respectively. It can be seen that the larger this ρ, i.e., the closer it is to 1, the more the concentration towards the mean direction μ. At the two extremes, $\rho = 0$ corresponds to an isotropic or uniform distribution while $\rho = 1$ corresponds to a degenerate distribution which concentrates at the single point μ. Again the sample or empirical analogs of μ and ρ are $\overline{\alpha}_0$ and $\bar{R} = R/n$ respectively, as defined in Section 1.3 while the sample analogs of the trigonometric moments are discussed in Section 1.3.3.

Remark 2.1 *Unlike in the linear case, the circular mean (as defined above) of the sum of 2 angles is not, in general, the sum of their circular means, even when interpreted modulo 2π. Such additivity of means can be seen to hold if the 2 random angles are independent and neither of them uniform.*

Indeed if we consider $Z = e^{i\theta}$ as the random variable, these are the so-called raw moments of Z since

$$\varphi_\theta(p) = E[Z^p] = E[\cos p\theta + i \sin p\theta] = \alpha_p + i\beta_p.$$

On the other hand, by applying the binomial expansion to $Z = \cos\theta + i\sin\theta$, we have

$$\alpha_p + i\beta_p = \sum_{j=0}^{p} \binom{p}{j} i^{p-j} E[\cos^j \theta \sin^{(p-j)} \theta],$$

providing a relation between the "trigonometric moments" of θ with the usual moments obtained by taking powers of $(\cos\theta)$ and $(\sin\theta)$. From this relationship, one can get all the joint moments of the *sine* and *cosine* components from the trigonometric moments and vice-versa.

Indeed, one can recover the pdf from either of these sequences. For convenience, from now on, we shall write $\varphi_\theta(p)$ as simply φ_p. These trigonometric moments of θ are the same as the Fourier coefficients in the Fourier series expansion of the pdf $f(\theta)$ (see Carslaw (1930), Chapter VII). If the circular pdf is square integrable on $[0, 2\pi)$ (in fact a much weaker set of conditions called "Dirichlet Conditions" will suffice for the convergence of Fourier Series, see pp.225-226 of Carslaw (1930)), one can recover this circular density from the Fourier coefficients, through the Fourier expansion

$$f(\theta) = \frac{1}{2\pi} \sum_{p=-\infty}^{\infty} \varphi_p e^{-ip\theta} = \frac{1}{2\pi} \left[1 + 2 \sum_{p=1}^{\infty} (\alpha_p \cos p\theta + \beta_p \sin p\theta) \right]. \quad (2.1.5)$$

This is analogous to the "inversion formula" for characteristic functions of real-valued r.v.s.

The pth *central* trigonometric moment can also be defined as:

$$E(e^{ip(\theta-\mu)}) = \alpha_p^* + i\beta_p^*,$$

μ being the mean direction.

We say that the circular r.v. θ is symmetric about a given direction μ if its distribution has the property

$$f(\mu + \theta) = f(\mu - \theta), \qquad \forall\, \theta.$$

Here, as elsewhere, we interpret addition or subtraction of angles modulo (2π). In particular symmetry around the zero direction corresponds to invariance of the density under the transformation $\theta \mapsto -\theta(mod\ 2\pi)$. Note that if a distribution is symmetric about the direction $\theta = \mu$, it is also symmetric about $\theta = \mu + \pi$. For a symmetric distribution, it can be checked that

$$\beta_p^* = 0, \qquad p = 0, \pm 1, \pm 2, \dots \quad (2.1.6)$$

so that the central trigonometric moments are all real. This is analogous to the fact that symmetric r.v.'s on the real line have real characteristic functions. In this case, we have:

$$f(\theta) = \frac{1}{2\pi} \left\{ 1 + 2 \sum_{p=1}^{\infty} \alpha_p \cos(p\theta) \right\}. \quad (2.1.7)$$

2.1.1 Some Methods of Obtaining Circular Distributions

We noted that a circular r.v. may be represented either in terms of the angle θ, $(0 \leq \theta < 2\pi)$ or as the two-dimensional unit vector $(X = \cos\theta,\ Y = \sin\theta)'$.

Many useful and interesting circular models may be generated from known probability distributions on the real line or on the plane, by a variety of mechanisms. We describe a few such general methods below:

(i) By *wrapping* a linear distribution around the unit circle;

(ii) Through *characterizing* properties such as maximum entropy, etc;

(iii) By transforming a bivariate linear r.v. to just its directional component, the so-called *offset distributions*;

(iv) One may start with a distribution on the real line $\mathbb{R}$, and apply a *stereographic projection* that identifies points x on $\mathbb{R}$ with those on the circumference of the circle, say θ. This correspondence is one-to-one except for the fact that the mass if any, at both $+\infty$ and $-\infty$, are identified with π. Such a correspondence is shown in Figure 2.3.

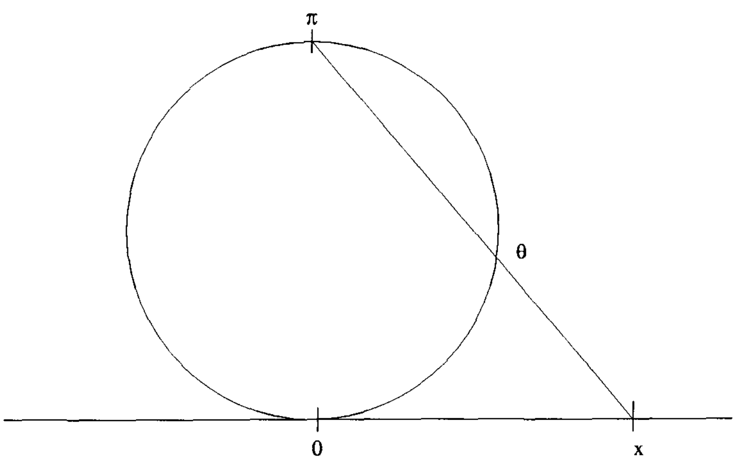


Figure 2.3: Stereographic projection.

We briefly explain each of these methods and present some examples in the following discussion.

Wrapped Distributions

Any linear r.v. X on the real line may be transformed to a circular r.v. by reducing its modulo 2π i.e., we define

$$\theta = X(mod\ 2\pi).$$

This operation corresponds to taking the real line and wrapping it around the circle of unit radius, accumulating probability over all the overlapping points $x = \theta,\ \theta \pm 2\pi,\ \theta \pm 4\pi,\ \ldots$. 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 r.v. , we have

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

By this technique, both discrete and continuous wrapped distributions may be constructed. An alternate procedure for wrapping, which might be more convenient sometimes (see Craig (1988)), but resulting in a different circular density, is to take $\theta = 2\pi(X - [X])$, where $[X]$ is the integer part of X. This amounts to wrapping around a circle of *unit* circumference first and then rescaling $[0,\ 1)$ to $[0,\ 2\pi)$.

Among the continuous wrapped distributions, special mention should to be made of the following two: Wrapped Normal (WN) and the Wrapped Cauchy (WC) distributions. While the pdf of WN does not particularly simplify, the pdf of WC admits a simplification by virtue of the geometric series occurring there. We discuss these in some detail in the following sections and elsewhere in the book. Note that both these are members of a more general family of wrapped distributions, called the "Wrapped Stable (WS) family", which is discussed in Section 2.2.8. The following result provides a way to represent the pdf of a wrapped circular distribution from the knowledge of the characteristic function say $\phi_X(t)$ of the linear r.v. X.

Proposition 2.1 *The trigonometric moment of order p for a wrapped circular distribution corresponds to the value of the characteristic function of the unwrapped r.v. X, say $\phi_X(t)$ at the integer value p, i.e., $\varphi_p = \phi_X(p)$.*

Proof: Indeed, using the cdfs G and F of θ and X respectively, we have

$$\begin{aligned} \varphi_p &= \int_0^{2\pi} e^{ip\theta} dG_\Theta(\theta) & (2.1.8) \\ &= \sum_{k=-\infty}^{\infty} \int_{2\pi k}^{2\pi(k+1)} e^{ip\theta} dF_X(\theta) \\ &= \int_{-\infty}^{\infty} e^{ipx} dF_X(x) = \phi_X(p). & (2.1.9) \end{aligned}$$

□

Characterization Properties

It is often instructive to ask if there are distributions on the circle which enjoy certain desirable properties. For instance, one may ask which distribution has the maximum *entropy* subject to having non-zero first trigonometric moment. Without such a condition on the existence of the mean, it is easy to see that the uniform distribution has the maximum entropy. As we will demonstrate in Section 2.2.4, the Circular Normal distribution has the maximum entropy property. This distribution is also the one whose mean direction is estimated with maximum likelihood by the direction of the resultant vector, $\overline{\alpha}_0$. This characterization, which we shall prove later, goes back to von Mises (1918) which is why circular normal distribution is also called the von Mises distribution.

If we ask which distribution on the circle has the property that the sample mean direction and the length of the resultant vector (which as we saw, represents the concentration) are *independent*, then the uniform or isotropic distribution is the answer (see Kent et al. (1979)). This characterization of the uniform distribution is similar to and as important as that of the normal distribution on the line as the one for which the sample mean and sample variance are independent.

Offset Distributions

Offset distributions are obtained from bivariate distributions on the plane by seeking just the distribution of the directional component. This is done by accumulating probabilities over all different lengths for a given direction. We transform the bivariate random vector (X, Y) into polar co-ordinates (R, θ) and integrate over R for a given θ. If $f(x, y)$ denotes the joint distribution of a bivariate distribution on the plane, then the resulting circular offset distribution, say $g(\theta)$, is given by

$$g(\theta) = \int_0^\infty f(r\cos\theta, r\sin\theta) r\,dr.$$

We discuss an example of this in Section 2.2.5.

2.2 Circular Distributions

2.2.1 Uniform Distribution

If the total probability is spread out uniformly on the circumference of a circle, we get the Circular Uniform (CU) distribution with the constant density

$$f(\theta) = \frac{1}{2\pi}, \qquad 0 \leq \theta < 2\pi. \tag{2.2.1}$$

All directions are equally likely and hence this is also known as the *isotropic* or random distribution. Its trigonometric moments α_p and β_p of all orders p are zero, except for α_0, which is unity. Since the length of the first trigonometric moment ρ is 0 in this case, this distribution has no well-defined mean or preferred direction. As mentioned already, the uniform distribution has the maximum entropy when one does not ask for the existence of a mean direction. Also the length R and direction $\overline{\alpha}_0$ of the sample resultant vector are independent, if and only if the sample is from a uniform distribution (see Kent et al. (1979)). Uniform distributions play a pivotal role in circular analysis because they represent the state of *no* "mean direction" or "preferred direction".

When a distribution is not uniform, we may think of a certain concentration towards one or more preferred directions. More commonly, there is just one preferred direction or mode and the distribution is said to be *unimodal.* We now discuss some important unimodal distributions, many of which are also symmetric around this preferred direction.

2.2.2 Cardioid Distribution

This distribution, derived from the cardioid curve (cf. Jeffreys (1961), p. 328) has the probability density function:

$$\begin{aligned} f(\theta;\ \mu,\ \rho) &= \frac{1}{2\pi}\{1+2\rho\cos(\theta-\mu)\}, \\ & \quad 0 \le \mu < 2\pi,\ -\frac{1}{2} < \rho < \frac{1}{2}. \end{aligned} \tag{2.2.2}$$

This distribution is unimodal and symmetric around μ. Its trigonometric moments are given by

$$\varphi_1 = |\rho| \cdot e^{i\mu} \ \text{ and } \ \varphi_p = 0 \ \text{ for } \ p \ge 2,$$

so that the parameter ρ denotes the concentration and μ denotes the mean direction. A related distribution, which is often used to model the directional spectra of ocean waves was considered by Cartwright (1963) and has density

$$f(\theta) = c(p)\cos^{2p}[(\theta-\mu)/2], \qquad p \ge 1, \tag{2.2.3}$$

where

$$c(p) = \frac{2^{(}2p-1)\cdot\Gamma^2(p+1)}{\pi\cdot\Gamma(2p+1)}.$$

For $p = 1$, this reduces to the Cardioid distribution with $\rho = 1/2$.

2.2.3 A Triangular Distribution

This distribution, with its pdf in the shape of a triangle, symmetric about $\theta = 0$, has the probability density function:

$$f(\theta;\ \rho) = \frac{1}{8\pi}\{4-\pi^2\rho+2\pi\rho|\pi-\theta|\}, \tag{2.2.4}$$

$$0 \le \theta < 2\pi,\ 0 \le \rho \le \frac{4}{\pi^2}. \tag{2.2.5}$$

Its trigonometric moments are given by

$$\varphi_1 = \rho \quad \text{and} \quad \varphi_{2p-1} = \rho/(2p-1)^2$$

for any integer p, with all the even moments being zero. Thus the distribution has zero as the mean direction and ρ measures the concentration.

2.2.4 Circular Normal (CN) Distribution

A circular r.v. θ is said to have a von Mises or a Circular Normal (CN) distribution if it has the density function:

$$f(\theta;\ \mu,\ \kappa) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(\theta-\mu)}, \qquad 0 \le \theta < 2\pi, \tag{2.2.6}$$

where $0 \le \mu < 2\pi$ and $\kappa \ge 0$ are parameters. Here $I_0(\kappa)$ in the normalizing constant is the modified Bessel function of the first kind and order zero (see Appendix A for some relevant and useful properties of Bessel functions), and is given by

$$I_0(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} \exp(\kappa \cos\theta)\, d\theta = \sum_{r=0}^{\infty} \left(\frac{\kappa}{2}\right)^{2r} \left(\frac{1}{r!}\right)^2 . \tag{2.2.7}$$

This distribution was introduced as a statistical model by von Mises (1918) and was discussed earlier by Langevin (1905), in the context of physics. Gumbel et al. (1953), who provide a nice discussion of this distribution and some of its properties, call it a "*Circular Normal distribution*" to emphasize its importance and similarities to the Normal distribution on the real line. The CN distribution has been extensively studied and inference techniques well developed. Therefore this is the model of choice for circular data in most applied problems. All through this monograph, we will use both these names – the von Mises distribution $vM(\mu, \kappa)$ and the Circular Normal (CN) distribution $CN(\mu, \kappa)$, interchangeably. Although not as useful, the cumulative distribution function of the CN distribution is obtained by integrating the series version of the pdf given in (2.2.10) and we get

$$F(\theta) = \frac{1}{2\pi I_0(\kappa)} \left\{ \theta I_0(\kappa) + 2 \sum_{p=1}^{\infty} \frac{I_p(\kappa) \sin p(\theta-\mu)}{p} \right\}, \qquad 0 \le \theta < 2\pi,$$

where the $I_p(\kappa)$ is the modified Bessel function of the first kind of order p (see Appendix A).

It is easy to verify the following properties of the von Mises density:

A <u>Symmetry:</u> By the symmetry of the *cosine* function, this density is seen to be symmetric about the direction μ (as well as $\mu + \pi$).

B Mode at μ: Since the *cosine* function has a maximum value at zero, the Circular Normal density is maximum at $\theta = \mu$, i.e., μ is the *modal direction* with the maximum value

$$f(\mu) = \frac{e^{\kappa}}{2\pi I_0(\kappa)}. \tag{2.2.8}$$

C Antimode at $(\mu \pm \pi)$: Again since $\cos\pi = -1$ is the minimum value, at the opposite end, when $\theta = \mu \pm \pi$, we have the minimum density, namely,

$$f(\mu \pm \pi) = \frac{e^{-\kappa}}{2\pi I_0(\kappa)}. \tag{2.2.9}$$

Hence, $\mu \pm \pi$ is the *anti-modal direction.*

D Role of κ: From Equations (2.2.8) and (2.2.9), we see that:

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

Hence, the larger the value of κ, the larger will be the ratio of $f(\mu)$ to $f(\mu \pm \pi)$ indicating higher concentration towards the population mean direction μ. Thus, κ is a parameter which measures the concentration towards the mean direction μ. See Figure (2.4).

E Trigonometric Moments:

Using the relations

$$\frac{1}{2\pi}\int_0^{2\pi} \cos(p\theta)\exp(\kappa\cos\theta)\ d\theta = I_p(\kappa)$$

and

$$\frac{1}{2\pi}\int_0^{2\pi} \sin(n\theta)\exp(\kappa\cos\theta)\ d\theta = 0,$$

one can see that the trigonometric moments for this density are given

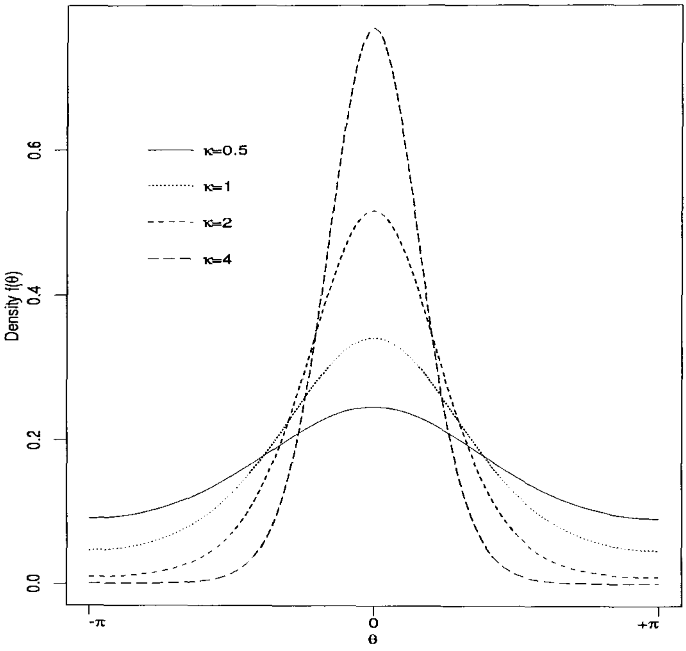


Figure 2.4: Circular Normal (von Mises) densities with $\mu = 0^0$ and $\kappa = 0.5,\ 1,\ 2,$ and 4.

by

$$\begin{aligned}
\varphi_p &= \frac{1}{2\pi I_0(\kappa)} \int_0^{2\pi} e^{ip\alpha} \exp(\kappa\cos(\alpha-\mu))\ d\alpha \\
&= \frac{1}{2\pi I_0(\kappa)} \int_0^{2\pi} e^{ip(\theta+\mu)} \exp(\kappa\cos\theta) d\theta \\
&= \frac{e^{ip\mu}}{2\pi I_0(\kappa)} \int_0^{2\pi} [\cos(p\theta) + i\sin(p\theta)] \exp(\kappa\cos\theta) d\theta \\
&= \frac{I_p(\kappa)}{I_0(\kappa)} \cdot e^{ip\mu}, \qquad p = 0,\ \pm1,\ \pm2,\ \ldots.
\end{aligned}$$

Thus, the length ρ of the polar vector or the first trigonometric moment, is seen to be

$$I_1(\kappa)/I_0(\kappa) = A(\kappa), \text{ say.}$$

The population mean direction, given by the direction of this polar vector, is μ, as one would expect from symmetry considerations. The central trigonometric moments can be seen to be

$$\alpha_p^* = I_p(\kappa)/I_0(\kappa) \cdot \cos(p\mu)$$

whereas $\beta_p^* = 0$ by virtue of the symmetry of the CN density. From the Fourier series expansion in (2.1.5), we can write the pdf in the series form

$$\frac{1}{2\pi I_0(\kappa)}\left\{I_0(\kappa) + 2\sum_{p=1}^{\infty} I_p(\kappa)\cos p(\theta - \mu)\right\}, \qquad 0 \leq \theta < 2\pi. \quad (2.2.10)$$

F <u>A Characterization:</u> We saw in Chapter 1 how the direction of the sample resultant vector $\overline{\alpha}_0$ provides a reasonable mean direction for a given sample. von Mises (1918) asked whether there is a circular model for which such an $\overline{\alpha}_0$ provides a maximum likelihood estimator of its location parameter i.e., characterize the circular distribution for which the population mean direction γ is estimated by the direction of the sample resultant with maximum probability. As we see below (cf. von Mises (1918)), it turns out that the density in (2.2.6) has this characterizing property. It might be recalled that the normal distribution has a similar property i.e., it is the one on the real line whose location parameter is estimated with maximum likelihood by the sample mean.

Let $\alpha_1, \alpha_2, \ldots, \alpha_n$ be observations from the density $f(\alpha - \gamma)$. Since the likelihood is given by

$$L = \prod_{i=1}^{n} f(\alpha_i - \gamma)$$

we obtain the likelihood equation

$$\frac{\partial \log L}{\partial \gamma} = \text{const} \cdot \sum_{i=1}^{n} \frac{f'(\alpha_i - \gamma)}{f(\alpha_i - \gamma)} = 0. \qquad (2.2.11)$$

On the other hand, if γ were to be estimated by $\overline{\alpha}_0$, we also have

$$\sum_{i=1}^{n} \sin(\alpha_i - \gamma) = 0. \tag{2.2.12}$$

Since Equations (2.2.11) and (2.2.12) hold for all arbitrary α_i and all n, the equality must hold term by term and thus,

$$\frac{f'(\alpha - \gamma)}{f(\alpha - \gamma)} = \kappa \sin(\alpha - \gamma)$$

for some constant κ. Hence,

$$f(\alpha - \gamma) = c \cdot e^{\kappa \cos(\alpha - \gamma)},$$

giving us the density (2.2.6).

G Maximum Entropy Property:

It is easy to see from Theorem 13.2.1 of Kagan et al. (1973), that the circular distribution which maximizes the *entropy*

$$-\int_0^{2\pi} f(\alpha) \log f(\alpha) d\alpha$$

subject to

$$E(\cos\alpha) = a \quad \text{and} \quad E(\sin\alpha) = b$$

is of the form

$$e^{\lambda_1 \cos\alpha + \lambda_2 \sin\alpha}.$$

This yields the $CN(\mu, \kappa)$ distribution as the maximum entropy distribution subject to having a given

$$a = \frac{I_1(\kappa)}{I_0(\kappa)} \cos\mu$$

and

$$b = \frac{I_1(\kappa)}{I_0(\kappa)} \sin\mu.$$

H As a Conditional Offset Distribution:

Let $X \sim N(\cos\mu, 1/\kappa)$ and $Y \sim N(\sin\mu, 1/\kappa)$ be independent. Transforming (X, Y) into polar coordinates $X = R\cos\theta, Y = R\sin\theta$, the joint density of (R, θ) is

$$\frac{\kappa r}{2\pi}\, e^{-\frac{\kappa}{2}(r^2-2r\cos(\theta-\mu)+1)}, \qquad 0 < r < \infty, \qquad 0 \le \theta < 2\pi.$$

The conditional density of $\theta | R = 1$ is thus $CN(\mu, \kappa)$. (See also Remark 2.2 later on.)

I A Member of the Curved Exponential Family:

Observe that the CN distribution is a member of the two-parameter Regular Exponential Family (REF) whereas for the particular case of known κ, it becomes a member of the (2, 1) Curved Exponential Family (see Section 2.5). Because of the somewhat technical nature of the topic, we discuss these aspects in the Appendix at the end of this chapter. As an exponential family member, certain trigonometric moments may be easily found e.g., by repeated differentiation under the integral sign. Also in certain cases, optimal tests are readily available.

J Convolution of Two von Mises Distributions:

Suppose θ_1 and θ_2 are $CN(\mu_1, \kappa_1)$ and $CN(\mu_2, \kappa_2)$ respectively and are independent. Then the convolution has the density

$$\begin{aligned} g(\theta) &= \frac{1}{4\pi^2 I_0(\kappa_1) I_0(\kappa_2)} \\ &\quad \times \int_0^{2\pi} \exp\{\kappa_1\cos(\alpha-\mu_1) + \kappa_2\cos(\theta-\alpha-\mu_2)\} d\alpha \\ &= \frac{1}{4\pi^2 I_0(\kappa_1) I_0(\kappa_2)} \int_0^{2\pi} \exp\{[\kappa_1\cos\mu_1 + \kappa_2\cos(\theta-\mu_2)]\cos\alpha \\ &\quad + [\kappa_1\sin\mu_1 + \kappa_2\sin(\theta-\mu_2)]\sin\alpha\} d\alpha \\ &= \frac{1}{2\pi I_0(\kappa_1) I_0(\kappa_2)} I_0(\sqrt{\kappa_1^2+\kappa_2^2+2\kappa_1\kappa_2\cos(\theta-(\mu_1+\mu_2))}), \end{aligned} \tag{2.2.13}$$

using Result (A.0.1) from Appendix A. However Wrapped Normal distribution provides a close approximation in practice, to a von Mises distribution, and it is closed under convolution. This could provide an approximation to the convolution of two Circular Normal distributions although, if convolutions are indeed needed, one is better off starting with the Wrapped Normal as the model of choice.

The following proposition says that for sufficiently large κ, the CN distribution can be *approximated* by a linear normal distribution.

Proposition 2.2 *As* $\kappa \longrightarrow \infty$ *,*

$$\beta = \sqrt{\kappa}(\alpha - \gamma) \stackrel{d}{\Longrightarrow} N(0,1).$$

Proof: Recall the CN density

$$f(\alpha) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(\alpha-\gamma)}, \quad 0 \le \alpha < 2\pi.$$

Let $\beta = \sqrt{\kappa}(\alpha - \gamma)$. Then, for large κ and hence small $\frac{\beta}{\sqrt{\kappa}}$,

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

From the Taylor series expansion, $\cos\theta \simeq 1 - \frac{\theta^2}{2}$. Using the change-of-variable formula and the fact that for large κ (see Appendix A), $I_0(\kappa) \simeq \exp(\kappa)/\sqrt{2\pi\kappa}$, we have

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

□

Remark 2.2 *This distribution, as we remarked earlier, may be traced back to Langevin (1905) who proposed the following model in p dimensions. On the surface of a p-dimensional sphere with unit radius, if μ of unit length denotes the mean direction, then towards any direction represented by a vector of unit length* $\mathbf{x}$, *the density is given by*

$$C_p(\kappa)e^{\kappa \mathbf{x}'\mu}, \qquad ||\mathbf{x}|| = 1, \quad ||\mu|| = 1.$$

It can be checked out that the normalizing constant $C_p(\kappa)$ is given by

$$C_p(\kappa) = \frac{\kappa^{\frac{p}{2}-1}}{(2\pi)^{\frac{p}{2}} I_{(\frac{p}{2}-1)}(\kappa)}.$$

In particular for the circular case (p=2), this constant reduces to

$$C_2(\kappa) = \frac{1}{2\pi I_0(\kappa)},$$

giving the von Mises or the CN distribution and for the spherical case (p=3), this constant reduces to

$$C_3(\kappa) = \frac{\kappa}{4\pi \sinh(\kappa)},$$

yielding the distribution whose sampling properties were discussed by R.A. Fisher in a fundamental paper (Fisher (1953)). Also one obtains the Langevin distribution by conditioning a p−dimensional r.v. $\mathbf{X}$ *with the $N_p(\boldsymbol{\mu}, \sigma^2 I)$ distribution to be of length one i.e.,*

$$\mathbf{X} \sim N_p(\boldsymbol{\mu}, \sigma^2 I) \quad \text{given} \quad ||\mathbf{X}|| = 1.$$

It is easy to see why this is so, since the multivariate normal has density

$$\begin{aligned} & c\ e^{-\frac{1}{2\sigma^2}(\mathbf{x}-\boldsymbol{\mu})'(\mathbf{x}-\boldsymbol{\mu})} \\ = \ & c\ e^{-\frac{1}{2\sigma^2}(\mathbf{x}'\mathbf{x}+\boldsymbol{\mu}'\boldsymbol{\mu}-2\mathbf{x}'\boldsymbol{\mu})} \\ \sim \ & c'\ e^{\kappa \mathbf{x}'\boldsymbol{\mu}}, \qquad \textit{with } \ \kappa = \frac{1}{\sigma^2} \end{aligned}$$

since $\mathbf{x}'\mathbf{x} = 1 = \boldsymbol{\mu}'\boldsymbol{\mu}$. *This generalizes what has been stated earlier under Property H.*

2.2.5 Offset Normal Distribution

The offset normal (ON) distribution is derived from the bivariate normal distribution $\phi(x,\ y;\ \boldsymbol{\mu},\ \boldsymbol{\Sigma})$ with mean $\boldsymbol{\mu} = (\mu,\ \nu)'$ and covariance matrix $\boldsymbol{\Sigma}$. If ρ denotes the correlation between the variables and $\sigma_1^2,\ \sigma_2^2$ their variances, the probability density function of the offset normal distribution is given by

$$f(\theta) = \frac{1}{C(\theta)}\left\{\phi(\mu,\ \nu;\ 0,\ \boldsymbol{\Sigma}) + aD(\theta)\Phi[D(\theta)]\phi\left[\frac{a(\mu\sin\theta - \nu\cos\theta)}{\sqrt{C(\theta)}}\right]\right\},$$

where

$$\begin{aligned} a &= \frac{1}{\sigma_1\sigma_2\sqrt{1-\rho^2}}, \\ C(\theta) &= a^2(\sigma_2^2\cos^2\theta - \rho\sigma_1\sigma_2\sin 2\theta + \sigma_1^2\sin^2\theta), \\ D(\theta) &= \frac{a^2}{\sqrt{C(\theta)}}[\mu\sigma_2(\sigma_2\cos\theta - \rho\sigma_1\sin\theta) + \nu\sigma_1(\sigma_1\sin\theta - \rho\sigma_2\cos\theta)] \end{aligned}$$

and $\phi(\cdot),\ \Phi(\cdot)$ are the pdf and cdf of $N(0,\ 1)$ respectively.
The particular case when $\boldsymbol{\mu} = \mathbf{0}$ and $\rho = 0$ leads to the density function

$$f(\theta) = \frac{\sqrt{1-b^2}}{2\pi(1 - b\cos 2\theta)},$$

where

$$b = \frac{\sigma_1^2 - \sigma_2^2}{\sigma_1^2 + \sigma_2^2}.$$

This has been widely used by meteorologists for wind direction under the assumption that the X and Y components of the wind vector are independently distributed as $N(0,\ \sigma_1^2)$ and $N(0,\ \sigma_2^2)$ respectively.

Another particular case of this offset normal distribution with zero means and unit variances and correlation ρ leads to

$$f(\theta) = \frac{\sqrt{1-\rho^2}}{2\pi(1 - \rho\sin 2\theta)}.$$

If further $\rho = 0$, this results in a uniform distribution because of the circular symmetry of such a bivariate distribution. Because of the complicated and unwieldy nature, the general pdf is not commonly used for either modeling or inference.

2.2.6 Wrapped Normal (WN) Distribution

A wrapped normal (WN) distribution is obtained by wrapping a $N(\mu, \sigma^2)$ distribution around the circle. Its pdf is given by

$$\begin{aligned} g(\theta) &= \sum_{m=-\infty}^{\infty} f(\alpha + 2m\pi) \\ &= \frac{1}{\sigma\sqrt{2\pi}} \sum_{m=-\infty}^{\infty} \exp\left[\frac{-(\theta - \mu - 2\pi m)^2}{2\sigma^2}\right]. \end{aligned} \tag{2.2.14}$$

By making use of the theory of theta functions (cf. Whittaker and Watson (1944), p.124), an alternate and more useful representation of this density can be shown to be

$$g(\theta) = \frac{1}{2\pi}\left(1 + 2\sum_{p=1}^{\infty} \rho^{p^2} \cos p(\theta - \mu)\right), \tag{2.2.15}$$

where $\rho = e^{\frac{-\sigma^2}{2}}$. The representation (2.2.15) can also be obtained by using the characteristic function of the $N(\mu,\ \sigma^2)$ and Proposition 2.1. It is clear that the density can be adequately approximated by just the first few terms of the infinite series, depending on the value of σ^2. It is unimodal and symmetric about the value $\theta = \mu$. It appears in the central limit theorem on the circle and in connection with Brownian Motion on the circle (see Stephens (1963)). The wrapped normal distribution possesses the additive property i.e., the convolution of two WN variables is also WN, unlike the von Mises distributions. Specifically if $\theta_1 \sim WN(\mu_1, \rho_1), \theta_2 \sim WN(\mu_2, \rho_2)$ are independent, then $\theta_1 + \theta_2 \sim WN(\mu_1 + \mu_2, \rho_1\rho_2)$. This follows from the fact that

$$\theta_i = X_i(mod 2\pi),$$

where

$$X_i \sim N(\mu_i, \sigma_i^2),\ i = 1, 2,$$

are independent. Now

$$\theta_1 + \theta_2 = X_1(mod 2\pi) + X_2(mod 2\pi) = (X_1 + X_2)(mod 2\pi).$$

However because of independence,

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

with corresponding concentration parameter

$$e^{-\frac{\sigma_1^2+\sigma_2^2}{2}} = e^{-\frac{\sigma_1^2}{2}} e^{-\frac{\sigma_2^2}{2}} = \rho_1 \cdot \rho_2.$$

Its shape is quite similar to that of the CND and one can find a matching CND for a given WN by having the same center and equating the concentration $I_1(\kappa)/I_2(\kappa)$ of the CN with the $\exp(-\sigma^2/2)$ of the WN (see Kendall (1974)). In particular, since

$$\frac{I_1(\kappa)}{I_0(\kappa)} \approx 1 - \frac{1}{2\kappa}$$

for small κ (see Appendix A) and

$$\exp\frac{-\sigma^2}{2} \approx 1 - \frac{-\sigma^2}{2},$$

for small σ^2 (< 0.4, say), we get matching CN density (2.2.6) and WN density (2.2.15) when

$$\kappa = \frac{-\sigma^2}{2}.$$

2.2.7 Wrapped Cauchy (WC) Distribution

The wrapped Cauchy(WC) distribution is obtained by wrapping the Cauchy distribution on the real line with density

$$f(x) = \left(\frac{1}{\pi}\right) \frac{\sigma}{\sigma^2 + (x-\mu)^2}, \quad -\infty < x < \infty$$

around the circle. It has the probability density function

$$\begin{aligned} g(\theta) &= \frac{1}{2\pi}\left(1 + 2\sum_{k=1}^{\infty} \rho^k \cos k(\theta-\mu)\right) \\ &= \frac{1}{2\pi}\frac{1-\rho^2}{1+\rho^2-2\rho\cos(\theta-\mu)}, \quad 0 \le \theta < 2\pi, \qquad (2.2.16) \end{aligned}$$

where $\rho = e^{-\sigma}$. The equality of the two expressions above is verified by equating the real parts of the geometric series identity

$$\sum_{k=1}^{\infty} a^k = \frac{a}{1-a}$$

with $a = \rho e^{-i(\theta-\mu)}$. The distribution is unimodal and symmetric. It possesses the additive property and earlier references go back to Wintner (1933) and Lévy (1939). The maximum likelihood estimation for this model is discussed in Section 4.4.

2.2.8 General Wrapped Stable (WS) Distributions

Wrapped α-stable distributions are constructed via Proposition (2.1) by using the characteristic function of the α-stable of the real line, which is given by (see Lukacs (1970))

$$\varphi(t) = \begin{cases} \exp\{-\tau^\alpha |t|^\alpha [1 - i\beta \operatorname{sgn}(t) \tan \frac{\alpha\pi}{2}] + i\mu t\}, & \text{if } \alpha \in (0,1) \cup (1,2], \\ \exp\{-\tau|t| + i\mu t\}, & \text{if } \alpha = 1, \end{cases}$$

with $\tau \geq 0, |\beta| \leq 1, 0 < \alpha \leq 2$, while μ is a real number. As stated in Proposition 2.1, the Fourier coefficients for a wrapped circular model correspond to the characteristic function at integer values for the unwrapped model. Thus, using Equation (2.1.5), the density function of a wrapped α-stable random variable for $\theta \in [0, 2\pi)$, is given by

$$f(\theta) = \frac{1}{2\pi} + \frac{1}{\pi} \sum_{k=1}^{\infty} \exp\{-\tau^\alpha k^\alpha\} \cos\left\{k(\theta - \mu) - \tau^\alpha k^\alpha \beta \tan \frac{\alpha\pi}{2}\right\}, \quad (2.2.17)$$

when $\alpha \in (0,1) \cup (1,2]$, with μ conveniently redefined as $\mu \stackrel{\text{def}}{=} \mu \,(\text{mod } 2\pi)$. Note that although there is generally no closed form expression for the density of an α-stable distribution on the real line, we are able to write such density for the wrapped case, at least as an infinite series. The particular case corresponding to $\beta = 0$ gives us the *Symmetric Wrapped Stable* (SWS) family of circular densities, which we will simply refer to as Wrapped Stable (WS), are given by

$$f(\theta; \rho, \alpha, \mu) = \frac{1}{2\pi} + \frac{1}{\pi} \sum_{k=1}^{\infty} \rho^{k^\alpha} \cos\{k(\theta - \mu)\}, \qquad (2.2.18)$$

where

$$\rho = \exp(-\tau^\alpha).$$

We shall denote such a distribution as $\mathrm{WS}(\alpha, \rho, \mu)$. In particular the wrapped normal density is a special case of Equation (2.2.17) with $\alpha = 2$ and $\beta = 0$, and gives the density in Equation (2.2.15) with $\rho = e^{-\frac{\sigma^2}{2}}$. The wrapped Cauchy density corresponds to the case $\alpha = 1$ and $\beta = 0$, and simplifies to Equation (2.2.16) with $\rho = e^{-\tau}$. Symmetric wrapped stable distributions have the following convolution property.

Lemma 2.1 *Let θ_1 and θ_2 be independent observations from the symmetric wrapped stable distributions WS(α, ρ_1, μ_1) and WS(α, ρ_2, μ_2) respectively. Then the distribution of $(\theta_1 - \theta_2)$ (mod 2π) is WS($\alpha, \rho, \mu_1 - \mu_2$) where $\rho = exp(-\tau_1^\alpha - (-\tau_2)^\alpha)$.*

Proof: For $j = 1, 2$, since θ_j has the WS distribution given above, its characteristic function (or Fourier series) is given by

$$\phi_j(p) = e^{i\mu_j p} e^{-(\tau_j p)^\alpha} = e^{i\mu_j p} (\rho_j)^{p^\alpha}, \; p = 1, \; 2, \; \ldots .$$

Therefore the c.f. of $\theta_1 - \theta_2$ say, $\phi^*(p), \; p = 1, \; 2, \; \ldots$ is

$$\begin{aligned} \phi^*(p) &= E\{e^{ip(\theta_1 - \theta_2)}\} \\ &= E\{e^{ip\theta_1}\} E\{e^{-ip\theta_2}\} \\ &= e^{i\mu_1 p} e^{-(\tau_1 p)^\alpha} e^{-i\mu_2 p} e^{-(-\tau_2 p)^\alpha} \\ &= e^{i(\mu_1 - \mu_2)p} . (\rho)^{p^\alpha}, \end{aligned}$$

proving the result. □

The wrapped α-stable class forms an important subclass of circular models for the following reasons: First, from the fact that the Fourier coefficients for a wrapped circular model correspond to the characteristic function at integer values for the unwrapped model, one can directly deduce that many of the important properties of linear α-stables are also enjoyed by the WS family, as e.g. domain of attraction, closure under convolution etc. Second, both the commonly used CN density and wrapped normal density, do not provide sufficient degree of flexibility, even when we restrict ourselves to symmetric densities. A practical illustration of this fact can be found in the data of Example 10.2 where the best-fitting CN distribution does not give an acceptable fit in contrast to an excellent fit given by an appropriately chosen member of the WS family.

We reproduce in Figure 2.5 some examples of symmetric wrapped α-stable densities for various choices of dispersion parameter τ and shape parameter

α. All densities have $\mu = 0$ and are plotted over the support $[-\pi, \pi)$. These figures are from Gatto and Jammalamadaka (2000) where additional discussion on conditional inference for the WS family can be found. It appears that by considering values of α smaller than 2 leads to densities with heavy tails which cannot be reproduced by wrapped normals even with larger variance. In particular, in Figure 2.5, we plot three wrapped symmetric α-stable densities with $\alpha = 0.4$, 0.8, 1.2 and $\tau = 1$ (solid lines), and three wrapped normal densities ($\alpha = 2$) with $\tau = 0.5$, 1, 1.2 (dashed lines). The various curves are easy to identify, since small values of α correspond to heavy tails, and large values of τ correspond to high dispersion. We can see that a wrapped normal density with the same extreme tail value as another wrapped symmetric α-stable density, would differ substantially in shape. Thus, wrapped α-stable densities allow for arbitrarily heavier tails than can the wrapped normal densities.

2.2.9 Variations of the CN Distribution

CN Distribution on part of the Circumference

One can define a CN distribution restricted to any arc of arbitrary length say $2\pi/\ell$ for an integer ℓ, rather than the full circle, with the resulting density

$$g(\theta) = \frac{\ell}{2\pi I_0(\kappa)} \exp[\kappa \cos \ell(\theta - \mu)], \quad 0 \le \theta,\ \mu < \frac{2\pi}{\ell}. \tag{2.2.19}$$

When ℓ=2, and the circular r.v. is restricted to a semi-circular arc, i.e., $\theta \in [0,\ \pi)$, it is called an "axial distribution". Occasionally, the measurements result in "axial data" as for instance when one observes the axes of pebbles rather than directions, and such distributions are relevant in modeling axial data.

ℓ-modal CN Distribution

A multi-modal CN density, say with ℓ equidistant modes, is obtained through the density

$$f(\theta) = \frac{1}{2\pi I_0(\kappa)} \exp\{\kappa \cos \ell(\theta - \mu)\}, \quad 0 \le \theta,\ \mu < 2\pi. \tag{2.2.20}$$

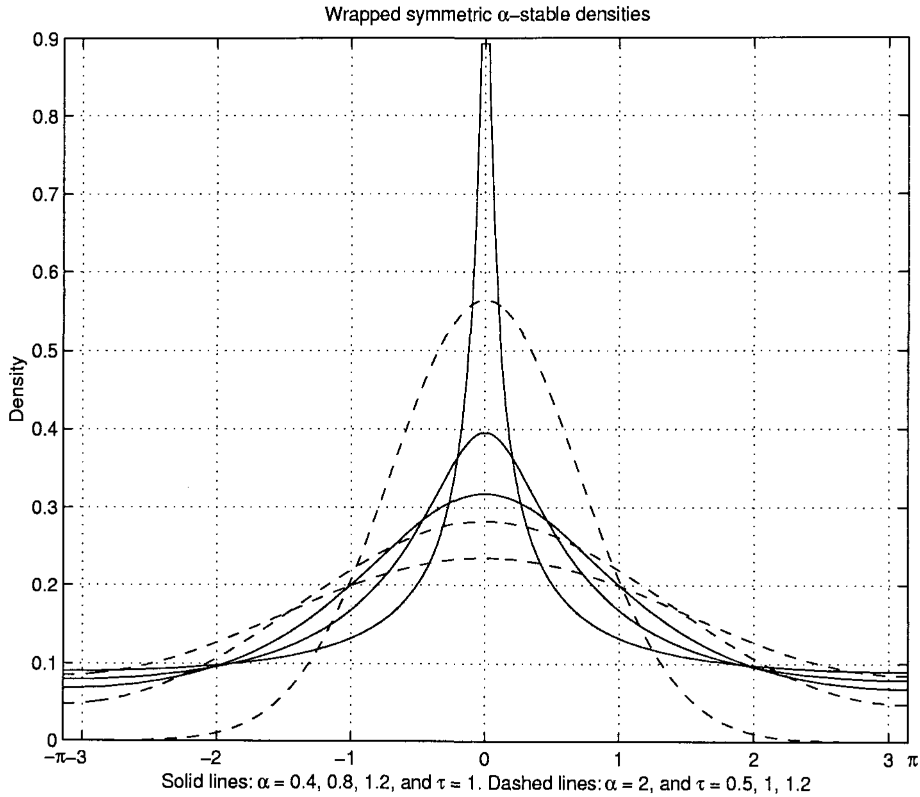


Figure 2.5: Wrapped symmetric α-stable densities with $\alpha = 0.4$, 0.8, 1.2 and $\tau = 1$ (solid lines), and with $\alpha = 2$ and $\tau = 0.5$, 1, 1.2 (dashed lines); all with $\mu = 0$.

Mixtures

Bimodal distributions may also be obtained by considering mixtures of say two CN distributions or other combinations of densities. Although these introduce flexibility and enrichment in modeling, they invite complications for statistical inference due to lack of sufficiency, invariance, etc., as well as the increased number of parameters involved. One should note however that the mixture of 2 unimodal distributions need not always lead to bimodality.

Consider the mixture of two von Mises distributions with probability density

$$\begin{aligned} f(\theta;\ \mu,\ \kappa) &= \frac{1}{2}\{CN(\theta;\ \mu,\ \kappa)+CN(\theta;\ -\mu,\ \kappa)\} \\ &= \frac{1}{4\pi I_0(\kappa)}\left\{e^{\kappa\cos(\alpha-\mu)}+e^{\kappa\cos(\alpha+\mu)}\right\}, \end{aligned} \tag{2.2.21}$$

where $CN(\theta;\ \mu,\ \kappa)$ is the von Mises density with mean direction μ and concentration parameter κ. This mixture (2.2.21) is not necessarily bimodal. For example, for $\mu \leq \arccos[\{-1+(1+4\kappa^2)^{1/2}\}/2\kappa]$, it is platykurtic and unimodal (see Mardia and Sutton (1975)). Bartels (1984), who was interested in modeling the direction of travel of *Dictyostelium slugs* in response to light and heat, used the above mixture (2.2.21). Basu and Jammalamadaka (2000) provide a Bayes procedure for testing if such CN mixtures are unimodal.

Wrapped Stable and Uniform Mixture (WSM)

An important family is obtained by mixing a wrapped stable distribution with a uniform $[0,\ 2\pi)$ distribution giving the density

$$\begin{aligned} f^*(\theta;p,\alpha,\rho,\mu_0) &= (1-p)\frac{1}{2\pi}+p\frac{1}{2\pi}\left\{1+2\sum_{k=1}^{\infty}\rho^{k^\alpha}\cos k(\theta-\mu_0)\right\}, \\ & \qquad p\in[0,1],\ \mu_0\in[0,2\pi),\rho\in[0,1),\alpha\in(0,2]. \end{aligned} \tag{2.2.22}$$

We will refer to this class of densities as the Wrapped Stable Mixture (WSM) family. SenGupta and Pal (2001a) provide applications of and optimal test procedures for testing for no mixtures in this model.

2.2.10 A Circular Beta Model

Saw (1978) suggested a method for generating a family of distributions based on the tangent normal decomposition of a distribution on the n-dimensional sphere, S^{n-1}. Lai (1994) uses this idea to construct a family of symmetric circular distributions. To make the convention that the modal vector is always pointing towards the same direction as the mean vector, we will assume without loss of generality, that $\alpha \geq \beta$. In the circular case (i.e., distribution on the unit circle S^1), the density function of the circular beta model becomes

$$f(\boldsymbol{x}) = \frac{1}{2^{\alpha+\beta}B(\alpha,\ \beta)}(1+\boldsymbol{\mu}'\boldsymbol{x})^{\alpha-\frac{1}{2}}(1-\boldsymbol{\mu}'\boldsymbol{x})^{\beta-\frac{1}{2}},\quad \boldsymbol{x}\in S^1, \tag{2.2.23}$$

where $B(\cdot,\ \cdot)$ is the beta function.

Equivalently, when the polar coordinate system is used, the probability density function can be written in terms of the angle made with the positive x-axis and we get

$$f(\theta) = \frac{1}{2^{\alpha+\beta}B(\alpha,\ \beta)}(1+\cos\theta)^{\alpha-\frac{1}{2}}(1-\cos\theta)^{\beta-\frac{1}{2}},\quad \theta\in[0,\ 2\pi).$$

One gets various shapes depending on the choice of α and β. For instance, it can be seen that when the values of α and β are both less than 0.5, the probability function is unbounded with mode at both 0^0 and 180^0, indicating a bimodal distribution. When the value of α is equal to 0.5 and the value of β is less than 0.5, the mode is at 0^0, indicating a unimodal distribution. The probability function in this case is bounded. When both parameters are equal to 0.5, the probability density function is that of the uniform distribution. When the values of both parameters are greater than 0.5, the distribution changes to equatorial or small circle depending on whether both values are equal or not.

A somewhat related directional distribution was proposed by McCullagh (1989) and has the form

$$h(t) = \frac{(1-t^2)^{\nu-\frac{1}{2}}}{(1-2\theta t+\theta^2)^{\nu}\ B(\nu+\frac{1}{2},\ \frac{1}{2})},\quad t\in(-1,\ 1),$$

where the parameters satisfy $\nu > -\frac{1}{2}$ and $-1\le\theta\le 1$.
Hence, the density defined for S^{n-1} is

$$f(\boldsymbol{x}) = \frac{\Gamma\left(\frac{n-1}{2}\right)}{2\pi^{\frac{n-1}{2}}}\frac{(1-(\boldsymbol{\mu}'\boldsymbol{x})^2)^{\nu+1-\frac{n}{2}}}{(1-2\theta(\boldsymbol{\mu}'\boldsymbol{x})+\theta^2)^{\nu}\ B(\nu+\frac{1}{2},\ \frac{1}{2})}.$$

For the circular case ($n=2$), this reduces to

$$f(\boldsymbol{x}) = \frac{1}{2}\frac{(1-(\boldsymbol{\mu}'\boldsymbol{x})^2)^{\nu}}{(1-2\theta\boldsymbol{\mu}'\boldsymbol{x}+\theta^2)^{\nu}\ B(\nu+\frac{1}{2},\ \frac{1}{2})}.$$

These r.v.s may be generated by using the method suggested in Michael et al. (1976). When $\nu < 0.5$, the distribution is bimodal and when $\nu \ge 0.5$, the distribution is a small circle or equatorial in the spherical case.

2.2.11 Asymmetric Circular Distributions

Although one comes across many examples in practice, there are not many useful models of asymmetric circular distributions. Wrapping a linear random variable, even an asymmetric one, need not result in an asymmetric circular density. Motivated by the multiparameter exponential families, (see Rukhin (1972) and Beran (1979)), one may consider a family of asymmetric circular densities given by

$$\begin{aligned} f(\theta; \mu_1, \mu_2, \kappa_1, \kappa_2) &= C \cdot \exp\{\kappa_1 \cos(\theta - \mu_1) + \kappa_2 \cos 2(\theta - \mu_2)\}, \\ & \quad 0 \le \theta, \mu_1, \mu_2 < 2\pi, \quad \kappa_1, \kappa_2 \geq 0, \end{aligned} \tag{2.2.24}$$

where C is the normalizing constant. This 4-parameter density can be used to represent symmetric or asymmetric, unimodal or bimodal shapes depending on the choice of the parameters.

Another example of an asymmetric axial distribution is provided by the marginal density of the axial component of the Langevin-Fisher distribution on the sphere. Its density is given by,

$$f(\theta; \kappa) = C \cdot \sin\theta \exp\{\kappa \cos(\theta)\}, \qquad 0 \le \theta < \pi, \ \kappa > 0. \tag{2.2.25}$$

2.3 Bivariate Circular Distributions

2.3.1 A Bivariate von Mises Distribution

The general results on maximum entropy characterization (see e.g., Kagan et al. (1973)) led Mardia (1975) to propose the following distribution in an attempt to define a bivariate analogue of the von Mises distribution on the circle. The density given by

$$\begin{aligned} f_{\Theta,\Phi}(\theta, \phi) &= C \ \exp\{\kappa_1 \cos(\theta - \mu)\kappa_2 \cos(\phi - \nu) \\ & \quad + \rho\sqrt{\kappa_1 \kappa_2} \cos(\theta \pm \phi - \psi)\}, \end{aligned} \tag{2.3.1}$$

where $\rho \le 1$, C is a normalizing constant and $\kappa_1, \ \kappa_2 > 0$, has the attractive property of being maximum entropy density under the constraints

$$E[e^{i\theta}] = e^{i\mu}, \ E[e^{i\phi}] = e^{i\nu} \quad \text{and} \quad E[e^{i(\theta \pm \phi)}] = \rho e^{i\psi}.$$

On the other hand, another attractive model which maximizes entropy subject to

$$E[e^{i\theta}] = e^{i\mu}, \ E[e^{i\phi}] = e^{i\nu} \quad \text{and} \quad E[\sin(\theta - \mu)\sin(\phi - \nu)] = c,$$

i.e., given means, concentrations and circular correlation ρ_c as defined in 8.2.2, has the form,

$$f(\theta, \phi) = C. \exp\{\kappa_1 \cos(\theta - \mu) + \kappa_2 \cos(\phi - \nu) + \kappa_3 \sin(\theta - \mu) \sin(\phi - \nu)\}$$

for $(0 \leq \theta, \phi < 2\pi)$.

In general, one can obtain a family of joint densities (see Wehrly and Johnson (1979)) for Θ and Φ by taking

$$f_{\Theta,\,\Phi}(\theta,\ \phi) = g(F_1(\theta) + F_2(\phi))\ f_1(\theta)\ f_2(\phi),$$

where g, f_1 and f_2 are univariate circular densities and F_1 and F_2 are the cumulative distribution functions (cdf's) corresponding to f_1 and f_2. The marginal densities of Θ and Φ are f_1 and f_2 respectively.

2.3.2 Wrapped Bivariate Normal Distribution

This distribution is a direct generalization of the univariate wrapped normal and is discussed in Jammalamadaka and Sarma (1988). See Chapter 8 for more details. Suppose

$$(X,\ Y) \sim N\left(\mathbf{0},\ \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right).$$

Define $\Theta = X (\text{mod } 2\pi)$, $\Phi = Y (\text{mod } 2\pi)$, so that their joint density is given by

$$f_{\Theta,\Phi}(\theta,\ \phi) = \sum_{j,k \in \mathbb{Z}} f_{X,\,Y}(\theta + 2\pi\ j,\ \phi + 2\pi\ k),$$

where $f_{X,Y}(x,\ y)$ is the bivariate normal density. The marginals are wrapped normal with parameters σ_1^2 and σ_2^2 respectively. For Θ, Φ to have identical marginals, obviously, we must have $\sigma_1 = \sigma_2$. One may ask which bivariate extension of the von Mises distribution is most similar to such a bivariate WN distribution, in view of such similarity between the univariate versions of these distributions.

2.3.3 Circular-Linear Distribution

In examples dealing with medicine and *rhythmometry* (see Batschelet (1981)), one needs to model a circular r.v. θ along with a linear r.v. X resulting in a distribution on the surface of a cylinder $[0, 2\pi) \times R^1$. One can for instance take θ to be von Mises and $X|\theta$ to be normal with its mean depending on θ, as in Mardia and Sutton (1978). Other examples of such circular-linear distributions can be found in Johnson and Wehrly (1978).

2.4 Generation of Circular Random Variables

Generation of circular uniform variables is straightforward. There are several efficient algorithms which generate CN variables with given μ and κ. The subroutine in our package which does this is called "`rvm`" for random variable generation from von Mises (or CN). The 3 arguments inside refer to the sample size needed, μ, and κ.

Example 2.1 *Using the* ***CircStats*** *routines, we can generate a random sample of size 20 from the* $CN(60^0,\ 12)$ *distribution in the following way (the result is given in degrees):*

```
> deg(rvm(20,rad(60),12))
 [1]  26.54263    66.77400  76.14425 100.93612  29.73252  27.27067
 [7]  48.63416    94.32601  77.71981  65.63407  56.04723  20.41976
[13]  90.05026    82.48932  70.36470  82.63763  48.99205  65.32162
[18]  53.62310    36.62527
```

Generation of circular random variables from any member of the wrapped stable family is performed by first generating linear random variables from the corresponding stable distribution and then taking mod 2π. The subroutine is called "`wrpstab`".

Example 2.2 *Again using the* ***CircStats*** *routines, we generate 20 observations from the wrapped stable distribution with index* $\alpha = 1.5$ *and skewness parameter* $\beta = .5$.

```
> deg(rwrpstab(20,1.5,.5))
 [1]  41.330141  40.883784 314.124611  48.326819 282.889739
 [6] 355.909996  54.849052  61.900085  20.223693  22.712903
[11] 290.045797 305.590105  16.224927 256.262106   5.362626
[16] 296.286103 312.692494 348.369930  44.970269  82.276158
```

Finally circular random variables from any member of the Wrapped Stable Mixture family are generated by probability allocation using the mixing proportion p. The following SPlus routine produces example from the desired mixture.

```
> rmixedwrpstab <- function(n, index1, skewness1, index2,
                                    skewness2, p)
{
     result <- c(1:n)
     for(i in 1:n) {
         test <- runif(1)
         if (test < p)
              result[i] <- rwrpstab(1, index1, skewness1)
              else result[i] <- rwrpstab(1, index2, skewness2)
     }
     result
}
```

Example 2.3 *Then application of the function with values of the parameters, e.g., sample size* $n = 20$, $\alpha_1 = 1.5$, $\beta_1 = .5$, $\alpha_2 = .7$, $\beta_2 = -1$, *and mixing proportion* $p = .3$, *results in:*

```
> deg(rmixedwrpstab(20,1.5,.5,.7,-1,.3))
 [1] 122.24260  43.38932 332.41061  10.15483  64.52638 187.03352
 [7]  11.53673 294.95383 294.38497 268.25848 218.25336 337.76633
[13]  43.79028 352.73323 216.02179 178.46674  36.96057 330.45315
[18] 213.29043 329.82699
```

Remark 2.3 *Observe that if we can generate samples of a given size from any distribution or mixture, we can compute a Monte Carlo sampling distribution of any useful statistic, say the length of the resultant. This can be done by sampling from the distribution repeatedly, calculating the statistic of interest and plotting the distribution of values of this statistic. In many cases, this may be our last resort for generating sampling distribution of a statistic if the distribution theory gets complicated. Theoretical derivations of the sampling distributions of circular statistics of interest, is the focus of our next chapter.*

2.5 Appendix: Curved Exponential Families and CND

We present here some of the ideas in curved exponential families (CEFs) and how they relate to the CND. See Efron (1975), Amari (1985) for a detailed discussion on the CEFs. Let Y be a r.v. taking on values in a nonempty open subset O of a Euclidean space and let

$$P(\tilde{\Theta}) = \{P_\eta \mid \eta \in \Theta\}$$

be a class of probability measures on $\mathcal{L}$, where the parameter space $\tilde{\Theta}$ is a nonempty open subset of $\mathbb{R}^p$ and for $\eta_1 \neq \eta_2$ in Θ, assume $P_{\eta_1} \neq P_{\eta_2}$. Next let,

$$\Theta = \{\eta \in \tilde{\Theta} \mid \eta = \psi(\theta),\ \theta \in L\}$$

be a 'surface' in $\tilde{\Theta}$ parametrized by θ, where L is a nonempty open subset of $\mathbb{R}^q$ with $q < p$ and $\psi(\cdot)$ is a known Borel bimeasurable bijection from L onto its image $\psi(L)$ in $\tilde{\Theta}$. We call the subfamily $P(\tilde{\Theta}) = \{P_\eta \in \tilde{\Theta}\}$, a *curved* family in $P(\tilde{\Theta})$. Further, let

$$P(\tilde{\Theta}) = \{f(t \mid \theta) = \exp\{< t,\ \eta(\theta) > -h(t) - \tau(\theta)\},\ \theta \in L\}, \qquad (2.5.1)$$

where

$$< t,\ \eta(\theta) >= \sum_{i=1}^{p} t_i \eta_i(\theta),$$

with range of t_i, $i = 1, \ldots, p$ independent of θ. $P(\tilde{\Theta})$ given by Equation (2.5.1) can be looked upon as a "reduced" dimensional exponential family with respect to a σ-finite measure ν where

$$dim(\Theta) = q < p = dim(Q),$$

Q being a minimal sufficient statistic. Consider the family in (2.5.1) with $q = 1$, i.e. the one-parameter curved exponential family. Let $\sum_\theta = cov_\theta(T)$. Denote the component-wise derivatives of $\eta(\theta)$ with respect to θ by,

$$\dot{\eta}(\theta) \equiv (\partial/\partial\theta)\eta(\theta), \qquad \ddot{\eta}(\theta) \equiv (\partial^2/\partial\theta^2)\eta(\theta).$$

Assume that these derivatives exist continuously in a neighborhood of a value of θ where we wish to define the curvature. Let,

$$\gamma_\theta^2 = \mid M_\theta \mid / \nu_{20}^3(\theta),$$

where

$$M_\theta \equiv \begin{bmatrix} \nu_{20}(\theta) & \nu_{11}(\theta) \\ \nu_{11}(\theta) & \nu_{02}(\theta) \end{bmatrix} \equiv \begin{bmatrix} \dot{\eta}(\theta)' \sum_\theta \dot{\eta}(\theta) & \dot{\eta}(\theta)' \sum_\theta \ddot{\eta}(\theta) \\ \ddot{\eta}(\theta)' \sum_\theta \dot{\eta}(\theta) & \ddot{\eta}(\theta)' \sum_\theta \ddot{\eta}(\theta) \end{bmatrix}.$$

Then, γ_θ is called the *statistical curvature* of P at θ.

Geometry of the CN Distribution

Case I: κ is unknown

Let $\alpha_1, \ldots, \alpha_n$ be a random sample from $CN(\mu, \kappa)$ i.e. from the pdf

$$f(\alpha) = [2\pi I_0(\kappa)]^{-1} \exp[\kappa \cos(\alpha - \mu)], \quad 0 \le \alpha, \mu < 2\pi, 0 \le \kappa < \infty. \quad (2.5.2)$$

It is clear that

$$(\sum_{i=1}^{n} \cos \alpha_i, \sum_{i=1}^{n} \sin \alpha_i) = (C, S)$$

is sufficient for (μ, κ) and (2.5.2) is a member of the regular exponential family. Now, we may rewrite

$$f(\alpha) \equiv \exp[\theta^i x_i - \phi(\theta)],$$

where $x_1 = \cos \alpha$ and $x_2 = \sin \alpha$. The natural parameters are

$$\theta^1 = \kappa \cos \mu, \ \theta^2 = \rho \sin \mu.$$

The expectation parameters given by $\nu_i = E(X_i) \equiv \partial_i \phi(\theta)$ are $\nu_1 = \rho \cos \mu$, $\nu_2 = \rho \sin \mu$, where $\rho = I_1(\kappa)/I_0(\kappa) = A(\kappa)$. The mean direction μ and the concentration parameter κ are frequently used as the parameter $\gamma = (\gamma^1, \gamma^2)$, $\gamma^1 = \mu$, $\gamma^2 = \kappa$ to specify the family $\mathcal{S} = \{CN(\mu, \kappa)\}$ of the circular normal distributions. The parameter space is then the infinite open-top rectangle in the first quadrant with its base on 0 to 2π. The natural basis $\{\partial_i\}$ is $\partial_1 = \partial/\partial\mu$, $\partial_2 = \partial/\partial\kappa$. The tangent vector T_θ is spanned by these vectors. We can identify T_θ, the differentiation operator representation of the tangent space with $T_\theta^{(1)}$ as the random variable or 1-representation of the same tangent space.

From the log-likelihood function, $l(\alpha, \gamma)$, the basis $\partial_i l$ of the 1-representation is

$$\partial_1 l = \kappa \sin(\alpha - \mu), \qquad \partial_2 l = A(\kappa) + \cos(\alpha - \mu).$$

The space $T_\theta^{(1)}$ is spanned by these two r.v.s, so that it consists of all the linear trigonometric functions in α defined below whose expectation vanishes i.e.,

$$T_\theta^{(1)} = \{a\kappa \sin(\alpha - \mu) + b\cos(\alpha - \mu) + c\},$$

with $c = -bA(\kappa)$. Let

$$B_i^\nu = \frac{\partial \theta^\nu}{\partial \gamma^i} = \begin{bmatrix} -\kappa \sin \mu & \cos \mu \\ \kappa \cos \mu & \sin \mu \end{bmatrix}. \tag{2.5.3}$$

Let us consider the metric in the manifold of circular normal distributions. The metric tensor $g_{ij}(\gamma)$ or Fisher information matrix in the coordinate system $\gamma = (\mu, \kappa)$ of the circular normal family $CN(\mu, \kappa)$ is easily calculated by using $g_{ij}(\gamma) = -E[\partial_i \partial_j l(\alpha, \gamma)]$. The various expectations involved are computed by using the results:

(i) $E \sin p(\alpha - \mu) = 0, \quad E \cos p(\alpha - \mu) = I_p(\kappa)/I_0(\kappa), \quad p = 1, 2, \ldots;$

(ii) $\sin \alpha$ and $\cos \alpha$ are odd and even functions respectively;

(iii) $f(\alpha)$ is symmetric in μ;

(iv) $\int \partial_2^i \partial_1^j f(\alpha) d\alpha \equiv 0, \qquad i, j = 1, 2, \ldots.$

Note that $g_{12}(\gamma)$ and $g_{21}(\gamma)$ vanish identically, since $g_{12}(\gamma) = E[\sin(\alpha - \mu)] \equiv 0$. So the basis vectors ∂_1 and ∂_2 are always orthogonal. Thus we get,

Proposition 2.3 *The coordinate system γ is an orthogonal system, composed of two families of mutually orthogonal coordinate curves, $\gamma^1 = \mu =$ const and $\gamma^2 = \kappa =$ const. Note however that the length of ∂_i depends on the position γ; i.e.,*

$$|\, \partial_1 \,|^2 = \text{Var } [\kappa \sin(\alpha - \mu)] = \kappa A(\kappa),$$

$$|\, \partial_2 \,|^2 = \text{Var } [A(\kappa) + \cos(\alpha - \mu)] = 1 - A(\kappa)/\kappa - A^2(\kappa).$$

One can now calculate the Riemannian distance between two points, i.e., two CN distributions and the Riemannian geodesic curve connecting two CN distributions.

Case II: κ is known

Let κ be known, say equal to 1 for simplicity. Note that the model M consisting of the $CN(\beta, 1)$ is a submanifold imbedded in the space $\mathcal{L}$ of the circular normal distributions $CN(\beta, \kappa)$ with the coordinate system $\theta = (\beta, \kappa)$. From (2.5.2), it is clear that the CN family is a particular member of the regular exponential family and our $CN(\beta, 1)$ can be rewritten as $q(x, u) = f\{x, \theta(u)\}$, where

$$\theta(u) = (\theta^1(u), \theta^2(u)) \equiv (\cos u, \sin u).$$

Since this gives a smooth imbedding in the space of the exponential family, the family $M = \{q(x, u)\}$ is a (2, 1) Curved Exponential Family. The family M forms a one-dimensional submanifold, i.e., a curve imbedded in the two-dimensional manifold $\mathcal{L}$. It is a unit circle in the θ-plane of the natural coordinates θ, since $\theta(u)$ satisfies $(\theta^1)^2 + (\theta^2)^2 = 1$. If we use the expectation parameter $\eta = (\eta_1, \eta_2)$ the η-coordinates $\eta(u)$ of the distribution specified by u are, $\eta_1 = A(1)\cos u, \eta_2 = A(1)\sin u$. Hence, the family M is represented also by a circle, $\eta_1^2 + \eta_2^2 = A^2(1)$, centered at 0, in the η-plane of the expectation coordinates. In terms of the expectation parameter, the tangent vector B_{ai} of $M(a = 1)$ is

$$B_{ai} = \frac{\partial \eta_i}{\partial u} = A(1)[-\sin u, \cos u]$$

and a metric in $\mathcal{L}$ is defined by the tensor $g^{ij}(u) = E\{\partial_i l \cdot \partial_j l\}$. Now,

$$\begin{aligned} l &= -\ln 2\pi - \ln I_0(\kappa) + \kappa \cos\alpha \cos\mu + \kappa \sin\alpha \sin\mu \\ &= -\ln 2\pi - \ln I_0(A^{-1}(\rho)) + (A^{-1}(\rho)\eta_1 \cos\alpha)/\rho + (A^{-1}(\rho)\sin\alpha)/\rho, \end{aligned}$$

where $\rho^2 = \eta_1^2 + \eta_2^2$. In terms of the natural parameters,

$$l = -\ln 2\pi - \ln I_0(((\theta^1)^2 + (\theta^2)^2)^{1/2}) + \theta^1 \cos\alpha + \theta^2 \sin\alpha,$$

thus

$$\frac{\partial l}{\partial \theta^1} = -A(\kappa)\cos\mu + \cos\alpha \quad \text{and} \quad \frac{\partial l}{\partial \theta^2} = -A(\kappa)\sin\mu + \sin\alpha.$$

Hence,

$$
\begin{aligned}
g_{11} &= E(\frac{\partial l}{\partial \theta^1})^2 \\
&= -A^2(\kappa)\cos^2\mu + \frac{I_0(\kappa) - I_2(\kappa\cos 2\mu)}{2I_0(\kappa)}, \\
g_{22} &= E(\frac{\partial l}{\partial \theta^2})^2 \\
&= -A^2(\kappa) + \frac{I_0(\kappa) - I_2(\kappa)\cos 2\mu}{2I_0(\kappa)} \\
\text{and } g_{12} &= E(\frac{\partial l}{\partial \theta^1}\cdot\frac{\partial l}{\partial \theta^2}) \\
&= \frac{-A^2(\kappa)\sin 2\mu + E(\sin 2\alpha)}{2} \\
&= -\frac{1}{2}A^2(\kappa)\sin 2\mu.
\end{aligned}
$$

This gives the metric g_{ij} defined in $\mathcal{L}$, from which the metric g_{ab} of M is given by $g_{ab} = B^i_a B^j_b g_{ij}$ where $\theta^1 \equiv \theta^1(u) = \cos u$ and $\theta^2 \equiv \theta^2(u) = \sin u$. Then,

$$B^i_a(u) \equiv \partial_a \theta^i(u) = [-\sin u, \cos u], \quad a = 1, i = 1, 2$$

and hence,

$$g_{ab} = g_{11}\sin^2 u + g_{22}\cos^2 u - 2g_{12}\sin u\cos u.$$

A vector orthogonal to M, say $n^i(u)$ is such that, $n^i B^j_b g_{ij} = 0$. Normalize n^i to $B^i_\kappa(u)$ such that

$$B^i_\kappa B^j_b g_{ij} = 0 \text{ and } B^i_\kappa B^j_\lambda g_{ij} = 1.$$

Given n^i satisfying the orthogonality condition, this orthonormalization with respect to the metric g_{ij} is achieved by taking $B^i_\kappa(u) = [n^i n^j g_{ij}]^{1/2} n^i$. Figure 2.6 below is instructive and provides further clarification of the above discussions.

Let us attach a one-dimensional submanifold $D(u)$ of $\mathcal{L}$ to each point u such that $D(u)$ transverses M at the point $\theta(u)$ or $\eta(u)$. Here we attach a straight line $D(u)$ at each $\eta(u)$ in the η-coordinate system. The equation for $D(u)$ is,

$$\eta_1 = A(1)\cos u + [A(1)\cos u]v, \quad \eta_2 = A(1)\sin u + [A(1)\sin u]v,$$

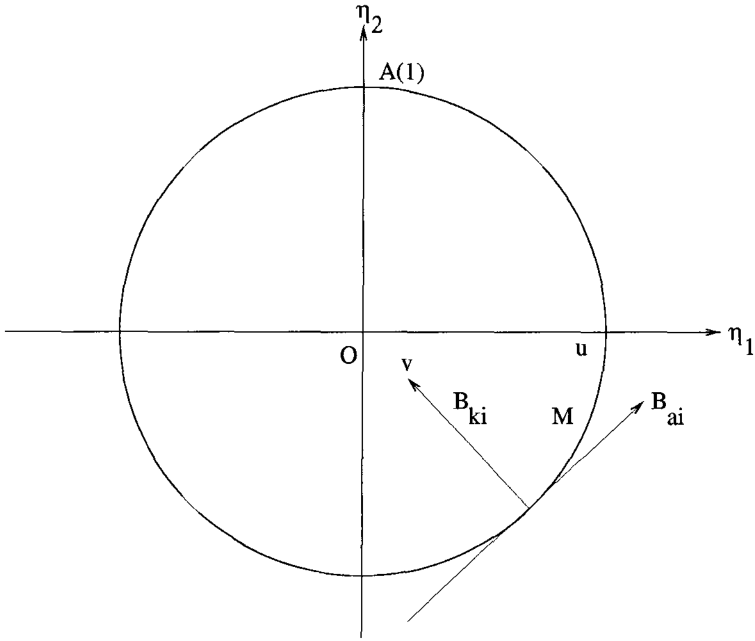


Figure 2.6: The curved exponential family CN(β, 1).

i.e., $D(u) = \{\eta \mid \eta_1, \eta_2\}$ where v is the parameter specifying points on $D(u)$. Here v can be regarded as a coordinate on the line $D(u)$, where the origin $v = 0$ is chosen at the intersection of M and $D(u)$.

The imbedding $\theta = \theta(u)$ is given by $\theta^1(u) = u, \theta^2(u) = 1$. Jacobian matrix $B^i_a(u)$ of the above coordinate transformation is

$$B^i_a(u) = \partial\theta^i/\partial u^a = (1, 0), \qquad a = 1, \quad i = 1, 2.$$

We attach to each point $u \in M$, a rigging ancillary submanifold $R(u)$ as shown in Figure 2.7, i.e. $R(u)$ consists of the CN distributions with fixed mean β and varying κ. Let $v = \kappa = 1$ be the coordinate of a point $CN(\beta, \kappa)$ in $R(u)$. Then (u, v) forms a coordinate system of $\mathcal{L}$ with the coordinate transformation $\theta^1(u, v) = u, \theta^2(u, v) = v + 1$ and the Jacobian matrix,

$$B^i_\alpha = \frac{\partial\theta^i(u, v)}{\partial\xi^\alpha} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

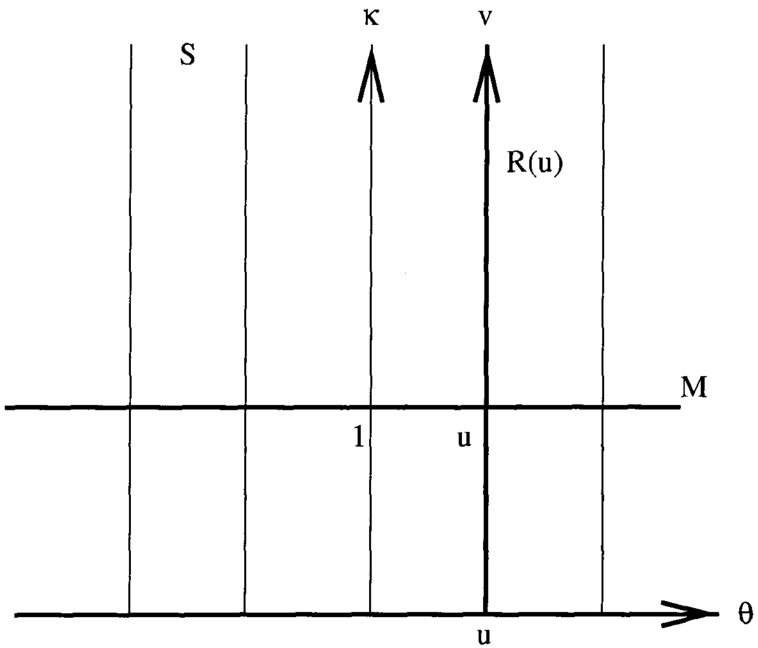


Figure 2.7: The rigging ancillary submanifold $R(u)$.

where $\xi = (u, v)$ and $\delta_\alpha B^i_\alpha = 0$ holds. The metric tensor $g_{\alpha\beta}$ is given by,

$$g_{\alpha\beta} = B^i_\alpha B^j_\beta g_{ij} = \begin{bmatrix} t_1(\kappa) & 0 \\ 0 & t_2(\kappa) \end{bmatrix},$$

where $t_i(\kappa), i = 1, 2$, are given below and $g_{ij} \equiv <\partial_i, \partial_j> = E[\partial_i l \cdot \partial_j l]$. In particular,

$$g_{11} = E[\kappa^2 \sin^2(\alpha - \beta)] = \kappa A(\kappa) \equiv t_1(\kappa),$$

$$\begin{aligned} g_{22} &= E[\cos^2(\alpha - \beta) + A^2(\kappa) - 2A(\kappa)\cos(\alpha - \beta)] \\ &= 1 - A(\kappa)/\kappa - A^2(\kappa) \equiv t_2(\kappa), \end{aligned}$$

and

$$g_{12} = E[-\kappa A(\kappa)\sin(\alpha - \beta) + \kappa \sin(\alpha - \beta)\cos(\alpha - \beta)] = 0.$$

Let a, b, c, standing for subscript 1, be indices for u and κ, α, μ, standing for subscript 2, be indices for v. So, $g_{a\kappa} = 0$ and this implies that the metric (Fisher information) of M is $g_{ab}(u) = t_1(1)$ where we put $v = 0$, i.e. $\kappa = 1$ and the metric of $R(u)$ on M is $g_{\kappa\lambda}(u) = t_2(1)$ and ∂_a and ∂_κ are mutually orthogonal, $g_{a\kappa}(u) = 0$. Thus we have,

Proposition 2.4 *The ancillary family is orthogonal. The α-connection $\Gamma^{(\alpha)}_{\alpha\beta\gamma}(u)$ in the associated coordinate system $\xi = (u, v)$ is,*

$$\begin{aligned}\Gamma^{(\alpha)}_{abc}(u) &\equiv E\left[\left\{\partial_a\partial_b l(x,u) + \frac{1-\alpha}{2}\partial_a l \partial_b l\right\}\partial_c l\right] \\ &= \Gamma^{(1)}_{abc} + \frac{1-\alpha}{2}T_{abc},\end{aligned}$$

where T_{abc} is a third order tensor.

Proof: Since $\partial_a \equiv \partial l/\partial\beta = \kappa\sin(\alpha-\beta)$,

$$\Gamma^{(\alpha)}_{abc}(u) = E[-\kappa^2\cos(\alpha-\beta)\sin(\alpha-\beta) + \frac{1-\alpha}{2}\kappa^3\sin^3(\alpha-\beta)] = 0.$$

Since $\partial_a \equiv \partial l/\partial\beta = \kappa = -A(\kappa) + \cos(\alpha-\beta)$, we get similarly

$$\begin{aligned}\Gamma^{(\alpha)}_{\kappa\lambda\mu}(u) &= E[-A'(\kappa)\{-A(\kappa)+\cos(\alpha-\beta)\} + \frac{1-\alpha}{2}\{\cos(\alpha-\beta) - A(\kappa)\}^3] \\ &= \frac{1-\alpha}{4I_0(\kappa)} \\ &\quad \times[2E\cos^3(\alpha-\beta) - 3A(\kappa)(1+I_2(\kappa) + 3A^2(\kappa)I_1(\kappa) - A^3(\kappa)] \\ &\neq 0 \text{ for any } \alpha, \text{ identically in } \kappa\end{aligned}$$

and

$$\begin{aligned}\Gamma^{(\alpha)}_{ab\kappa}(u) &= E[\{-A(\kappa)+\cos(\alpha-\beta)\}\cdot \\ &\quad \{-\kappa\cos(\alpha-\beta) + \frac{1-\alpha}{2}\kappa^2\sin^2(\alpha-\beta)\}] \\ &= \frac{1}{2I_0(\kappa)}[2\kappa A(\kappa)I_1(\kappa) - \kappa(1+I_2(\kappa)) \\ &\quad -(1-\alpha)\left\{\frac{1}{2}\kappa^2 A(\kappa)(1-I_2(\kappa))\right. \\ &\quad \left. + (I_1(\kappa) - \kappa(1-I_2(\kappa)) + \frac{1}{2}\kappa(1+I_2(\kappa)))\right\}],\end{aligned}$$

where $E\sin^2(\alpha-\beta)\cos(\alpha-\beta)$ may be obtained, for example, by using the identity, $\int \partial_\kappa \partial_\beta^2 f(\alpha)d\alpha \equiv 0$. Thus, $\Gamma^{(\alpha)}_{ab\kappa}(u) \neq 0$ for any α, identically in κ. Hence,

$$H^{(\alpha)}_{ab\kappa}(u) = \Gamma^{(\alpha)}_{ab\kappa}(u),$$

which gives the imbedding curvature of $R(u)$ in $\mathcal{L}$ at $\theta = \theta(u)$ does not vanish for any α, identically in $v = \kappa - 1$. However, the Riemann-Cristoffel curvature vanishes identically, since every one-dimensional manifold must be curvature-free. The above observations give,

Proposition 2.5 *The model M is not an α-flat submanifold in $\mathcal{L}$, is not an α-geodesic and, in particular, is not a 0-geodesic. However M is itself α-flat. The auxiliary submanifolds $R(u)$ are α-flat for any α. The coordinate u of M is α-affine for any α, while the coordinate v of $R(u)$ is not α-affine for any α.*

Chapter 3

Some Sampling Distributions

3.1 Introduction

For purposes of statistical inference about a given model, we need to know the sampling distributions of relevant statistics, based on a sample from that model. The most common models for which such useful sampling distributions are available, are the Circular Uniform (CU) distribution (2.2.1) and the Circular Normal (CN) model (2.2.6). Given a "random sample" or a set of independently and identically distributed (i.i.d.) random variables denoted by $\alpha_1, \alpha_2, \ldots, \alpha_n$, from some model, the basic statistics of interest for inference purposes are the sums of *sines* and *cosines*,

$$C = \sum_{i=1}^{n} \cos \alpha_i, \quad S = \sum_{i=1}^{n} \sin \alpha_i, \quad \text{and} \quad R = \sqrt{C^2 + S^2}, \tag{3.1.1}$$

where R is the length of the resultant vector. Equivalently one can consider the averages,

$$\overline{C} = \frac{1}{n}\sum_{i=1}^{n} \cos \alpha_i, \quad \overline{S} = \frac{1}{n}\sum_{i=1}^{n} \sin \alpha_i, \quad \text{and} \quad \overline{R} = \sqrt{\overline{C}^2 + \overline{S}^2} = \frac{R}{n}.$$

The joint probability distribution of (C, S) readily yield those of $(\overline{\alpha}_0, R)$ because of the one-to-one transformation

$$C = R \cos \overline{\alpha}_0, \qquad S = R \sin \overline{\alpha}_0. \tag{3.1.2}$$

The exact distributions of these statistics, even in the simple case when the data is from a Circular Uniform distribution, are rather complicated.

We start by discussing Pearson's random walk problem which is related to the distribution of R for uniform samples. This result can then be used to obtain other sampling distributions of interest for the uniform as well as for the CN or von Mises distribution for a single sample. Section 3.4 considers some asymptotic and related results while Section 3.5 discusses sampling distributions of interest for two-sample inference. In the final section, some approximate sampling distributions for large κ, are discussed.

3.2 Generalized Pearson's Random Walk Problem

In 1904, Nobel-Prize winning biologist Ronald Ross, who was working on the Malaria vector, was interested in how the density of mosquitoes might drop off as one goes farther from their breeding ground. Karl Pearson, who was asked about this, in turn, posed this as a "random walk" problem the following year (see Pearson (1905)) in the journal "Nature". A two-dimensional version of this general problem is as follows: An individual starts from an (original) point O and walks a length of a_1 in a random direction, then changes his/her direction again randomly to cover a distance a_2 and etc. After traversing n segments of lengths $a_1, a_2, \ldots, a_n$, he/she ends at a terminal point T. Let R be the length of the distance from the original point O to the terminal point T, i.e., the length of the line-segment OT. Then it is clear that the range of R is $0 \le R \le \sum a_i$. However the problem of interest here is to get the probability distribution of this length R within this range.

For an easy special case, consider $n = 2$, $a_1 = 1 = a_2$. Let the initial direction be OA. From A he/she takes the direction AT. Let $\angle OAT = \theta$. Then, from the statement of the problem, we have $\theta \sim U(0,\ 2\pi)$ so that

$$f(\theta) = \frac{1}{2\pi}, \qquad 0 \le \theta < 2\pi.$$

From the law of cosines,

$$\begin{aligned}
(OT)^2 &= (OA)^2 + (AT)^2 - 2(OA)(AT)\cos\theta \\
\text{or,} \qquad R^2 &= 2(1 - \cos\theta) \\
&= 4\sin^2\frac{\theta}{2}. \qquad (3.2.1)
\end{aligned}$$

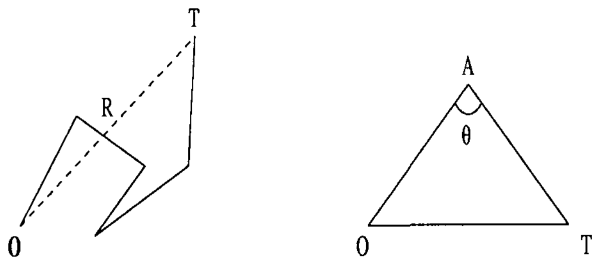


Figure 3.1: Pearson's random walk problem.

Differentiating (3.2.1),

$$\begin{aligned} 2rdr &= 2\sin\theta d\theta \\ \text{or,} \quad \frac{d\theta}{dr} &= \frac{r}{\sin\theta}. \end{aligned}$$

Again, from Equation (3.2.1), $2\cos\theta = 2 - r^2$ which gives

$$\sin\theta = \sqrt{1-(1-\frac{r^2}{2})^2}.$$

Taking into account that this is a two-to-one mapping we get the pdf of R to be,

$$g(r) = 2\frac{1}{2\pi}\frac{r}{\sqrt{1-(1-\frac{r^2}{2})^2}} = \frac{2}{\pi}\frac{1}{\sqrt{4-r^2}}, \qquad 0 \le r \le 2.$$

3.2.1 The General Case

For treating the general case, we need to make extensive use of another related Bessel function called $J_\nu(t)$ which for $\nu > -\frac{1}{2}$ is defined by

$$J_\nu(t) = \frac{(t/2)^\nu}{\Gamma(1/2)\Gamma(\nu+1/2)}\int_0^\pi e^{it\cos\theta}\sin^{2\nu}\theta d\theta.$$

See Appendix A for some useful properties.

Let R_k denote the length of the resultant of the first k lines of lengths $a_1, a_2, \ldots, a_k$. Let θ_k denote the angle between R_k and a_{k+1}. Obviously,

R_k is a function of $\theta_1,\ \theta_2,\ \ldots,\theta_{k-1}$ each of which is a random variable with uniform distribution on $(-\pi,\ \pi)$ under the assumptions. The cumulative probability distribution of $R_n = R$ is given by

$$P(R \leq r|a_1,\ \ldots,\ a_n) = \frac{1}{(2\pi)^{n-1}} \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} d\theta_{n-1} d\theta_{n-2} \cdots d\theta_1,$$

where $\theta_1,\ \ldots,\ \theta_{n-2}$ have range $(-\pi,\ \pi)$ but θ_{n-1} has a range which makes $R_n = R \leq r$. We now use an important property of the Bessel functions, called the Weber's discontinuous factor theorem (see Appendix A), which says

$$r\int_0^\infty J_0(Rt)J_1(rt)dt = \begin{cases} 1 & \text{if} \quad R < r, \\ 0 & \text{if} \quad R > r. \end{cases}$$

Using this as an "indicator function" that restricts R to the correct range, we obtain

$$\begin{aligned} & P(R \leq r|a_1, \ldots,\ a_n) \\ = \ & \frac{1}{(2\pi)^{n-1}} \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} \int_0^\infty r J_0(Rt) J_1(rt) dt d\theta_{n-1} \cdots d\theta_1. \end{aligned} \tag{3.2.2}$$

Now we put $R^2 = R_n^2 = R_{n-1}^2 + a_n^2 - 2a_n R_{n-1} \cos\theta_{n-1}$ in (3.2.2) and use another property of the Bessel functions (which follows from the Naumann addition formula; see Appendix A) namely

$$\int_0^{2\pi} J_0(t\sqrt{a^2+b^2-2ab\cos\theta})d\theta = 2\pi J_0(ta)J_0(tb).$$

Interchanging the integrals in t and θ_{n-1} and integrating over θ_{n-1} first, we get

$$\frac{2\pi}{(2\pi)^{n-1}} \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} \int_0^\infty r J_0(R_{n-1}t) J_0(a_n t) J_1(rt) dt d\theta_{n-2} \cdots d\theta_1.$$

Again since $R_{n-1}^2 = R_{n-2}^2 + a_{n-1}^2 - 2a_{n-1}R_{n-2}\cos\theta_{n-2}$, inserting this for R_{n-1} and proceeding as before, we obtain the distribution function of R in n steps, viz.,

$$F(r) = P(R \leq r|a_1,\ \cdots,\ a_n) = r\int_0^\infty J_1(rt) \prod_{i=1}^n J_0(a_i t)dt.$$

By differentiating this, using the fact,

$$\frac{dtJ_1(t)}{dt} = tJ_0(t),$$

one gets the density of R, for $0 \le r \le \sum a_i$, namely

$$\begin{aligned} f(r|a_1, \cdots, a_n) &= \frac{dF(r)}{dr} \\ &= \frac{d}{dr}\int_0^\infty \frac{rt}{t} J_1(rt) \prod_{i=1}^n J_0(a_i t)dt \\ &= \int_0^\infty \frac{drtJ_1(rt)}{dr}\frac{1}{t}\prod_{i=1}^n J_0(a_i t)dt, \qquad (0 \le r \le \textstyle\sum_{i=1}^n a_i) \\ &= \int_0^\infty rtJ_0(rt)\prod_{i=1}^n J_0(a_i t)dt. \end{aligned} \tag{3.2.3}$$

This is known as Kluyver (1906)'s solution of Pearson's random walk problem.

In particular, if we have n unit vectors from the circular uniform distribution and we are interested in the length of the resultant, this corresponds to the special case $a_i = 1$ for $i = 1, \ldots, n$. From 3.2.3, we obtain

$$f_0(r) = r\int_0^\infty J_0(rt)J_0^n(t)tdt, \qquad 0 \le r \le n \tag{3.2.4}$$

as the density function of R, the length of n unit vectors from the Uniform distribution. It is convenient to use the notation

$$\psi_n(r) = \int_0^\infty J_0(rt)J_0^n(t)tdt, \qquad 0 \le r \le n \tag{3.2.5}$$

for the integral in (3.2.4) so that we will write

$$f_0(r) = r\psi_n(r) \tag{3.2.6}$$

as the density of R for the uniform case. This forms the basis for deriving various sampling distributions for the uniform as well as CN samples.

3.3 Sampling Distributions for CN Distribution

Let $\alpha_1, \alpha_2, \ldots, \alpha_n$ be from CN distribution with parameters μ and κ. Then the joint density of the observations $(\alpha_1, \alpha_2, \ldots, \alpha_n)$ is given by

$$\begin{aligned} f_\kappa(\alpha_1, \ldots, \alpha_n) &= \frac{1}{I_0^n(\kappa)(2\pi)^n} e^{(\kappa\cos\mu)(\sum_{i=1}^n \cos\alpha_i)+(\kappa\sin\mu)(\sum_{i=1}^n \sin\alpha_i)} \\ &= \frac{e^{(\kappa\cos\mu)(\sum_{i=1}^n \cos\alpha_i)+(\kappa\sin\mu)(\sum_{i=1}^n \sin\alpha_i)}}{I_0^n(\kappa)(2\pi)^n} \\ &= \frac{e^{(\kappa\cos\mu)(\sum_{i=1}^n \cos\alpha_i)+(\kappa\sin\mu)(\sum_{i=1}^n \sin\alpha_i)}}{I_0^n(\kappa)} f_0(\alpha_1, \ldots, \alpha_n), \end{aligned} \tag{3.3.1}$$

where $f_0(\alpha_1, \ldots, \alpha_n)$ is the joint density of the data under the uniform model. This connection enables us to move from sampling distributions for uniform distribution with $\kappa = 0$ to the corresponding ones for CN with $\kappa > 0$. From Equation (3.3.1), we see $\overline{C}$ and $\overline{S}$ are jointly sufficient for this model. Let β denote the angle that the resultant vector makes with the polar direction μ and c its *cosine*, i.e.,

$$\cos(\overline{\alpha}_0 - \mu) = \cos\beta = c.$$

When μ is given,

$$V = \sum_{i=1}^n \cos(\alpha_i - \mu) = R\cos(\hat{\mu} - \mu) = R\cos\beta = R \cdot c$$

is sufficient for κ. Denote by $f_0(R, c)$ the joint distribution of R and c when $\kappa = 0$, i.e., when the observations are from the uniform distribution.

For uniformly distributed random variables on $[0, 2\pi]$, β is not only uniformly distributed but is also independent of the magnitude of R. We already remarked a converse to this result, namely that the independence of β and R characterizes the circular uniform distribution, was established by Kent et al. (1979) Thus for uniform samples, β or equivalently $\overline{\alpha}_0$ has a $U(0, 2\pi)$ distribution. Since R has the distribution given in (3.2.4) and $\overline{\alpha}_0$ is uniformly distributed and they are independent, we obtain

$$f_0(r, \overline{\alpha}_0) = r\psi_n(r)\frac{1}{2\pi}. \tag{3.3.2}$$

3.3.1 Distribution of (C, S)

Let C, S be as defined in (3.1.1) and let R denote the length and $\overline{\alpha}_0$, the direction of the vector resultant $\mathbf{R}$ for a sample $(\alpha_1, \ldots, \alpha_n)$. We will use the subscript "κ" and "0" for the sampling distributions when the sample is from a CN with $\kappa > 0$ and a uniform with $\kappa = 0$, respectively. Making the one-to-one transformation (3.1.2) from $(R, \overline{\alpha}_0)$ to (C, S) and recalling $(dcds = rdrd\overline{\alpha}_0)$, we have from the density in (3.2.5)

$$f_0(c, s) = \frac{1}{2\pi}\psi_n(\sqrt{c^2 + s^2}). \tag{3.3.3}$$

The joint distribution of (C, S) for the CN case is obtained from that of the Uniform case by using (3.3.1) and we get

$$\begin{aligned} f_\kappa(c, s) &= \int_{\mathcal{A}} f_\kappa(\alpha_1, \ldots, \alpha_n) d\alpha_1 \ldots d\alpha_n \\ &= \frac{e^{(\kappa\cos\mu)c + (\kappa\sin\mu)s}}{I_0^n(\kappa)} \int_{\mathcal{A}} f_0(\alpha_1, \ldots, \alpha_n) d\alpha_1 \ldots d\alpha_n \\ &= \frac{e^{(\kappa\cos\mu)c + (\kappa\sin\mu)s}}{I_0^n(\kappa)} f_0(c, s), \end{aligned} \tag{3.3.4}$$

where

$$\mathcal{A} = \left\{ (\alpha_1, \ldots, \alpha_n) : \sum_{i=1}^n \cos\alpha_i = c, \quad \sum_{i=1}^n \sin\alpha_i = s \right\}.$$

Substituting (3.3.3) in (3.3.4), we get

$$f_\kappa(c, s) = \frac{e^{(\kappa\cos\mu)c + (\kappa\sin\mu)s}}{2\pi\ I_0^n(\kappa)} \psi_n\left(\sqrt{c^2 + s^2}\right). \tag{3.3.5}$$

3.3.2 Distribution of R

Since $\beta \sim U(0, 2\pi)$, for uniform samples, by a simple transformation, this implies that $c = \cos\beta$ has the pdf

$$f_0(c) = \frac{1}{\pi\sqrt{1 - c^2}}, \ -1 \leq c \leq 1.$$

Using independence of R and β (see also (3.2.5)) we obtain

$$f_0(r, c) = f_0(r) f_0(c) = \frac{r\psi_n(r)}{\pi\sqrt{1 - c^2}}.$$

Now by a simple elegant argument as in (3.3.4), we can move from the uniform distribution to the CN distribution. For samples from the CN distribution, by direct definition,

$$f_\kappa(r,c) = \int \left(\frac{1}{2\pi}\right)^n \frac{e^{\kappa rc}}{I_0^n(\kappa)} d\alpha = \frac{e^{\kappa rc}}{I_0^n(\kappa)} \int \left(\frac{1}{2\pi}\right)^n d\alpha,$$

where the integration is over all samples that have the given values of R and c. But this integral is equal to the $f_0(r,c)$ corresponding to the uniform samples that we already have. Thus

$$f_\kappa(r,c) = \frac{e^{\kappa rc}}{I_0^n(\kappa)} \frac{r\psi_n(r)}{\pi\sqrt{1-c^2}}. \tag{3.3.6}$$

From this joint density, one gets the density of the resultant length for samples from the CND as

$$f_\kappa(r) = \int_{-1}^{1} f_\kappa(r,c)dc = \frac{I_0(\kappa r)}{I_0^n(\kappa)} r\psi_n(r). \tag{3.3.7}$$

From Equation (3.3.6), one can also find the conditional distribution of R given V as follows. Making the transformation $(r,c) \mapsto (v,r)$. The Jacobian of the transformation is given by

$$J = \frac{\partial(r,v)}{\partial(r,c)} = r.$$

Hence, in the range $0 \le r \le n$, $-r \le v \le r$,

$$f_\kappa(r,v) = \frac{e^{\kappa v}}{I_0^n(\kappa)} \frac{r\psi_n(r)}{\pi\sqrt{1-(\frac{v}{r})^2}} \frac{1}{r} = \frac{1}{\pi} \frac{e^{\kappa v}}{I_0^n(\kappa)} \frac{r\psi_n(r)}{\sqrt{r^2-v^2}} \tag{3.3.8}$$

so that

$$f_\kappa(r|v) = \frac{f_\kappa(r,v)}{f_\kappa(v)} = \frac{r\psi_n(r)}{\sqrt{r^2-v^2} f_0(v)\pi}. \tag{3.3.9}$$

Starting with this joint distribution $f_\kappa(r,c)$ in (3.3.6), one can also check that the conditional distribution of

$$\overline{\alpha}_0 \mid R = r \sim CN(\mu, \kappa r). \tag{3.3.10}$$

3.3.3 Distribution of V

Let $f_0(v)$ denote the pdf of

$$V = \sum_{i=1}^{n} \cos(\alpha_i - \mu)$$

when α_i's are from the uniform distribution. Since the characteristic function of $\cos\alpha$ is

$$\int_{-\infty}^{\infty} e^{it\cos\alpha}\frac{1}{2\pi}d\alpha = \int_{0}^{2\pi} e^{it\cos\alpha}\frac{1}{2\pi}d\alpha = J_0(t),$$

and since $\alpha_1, \ldots, \alpha_n$ are independent, the characteristic function of v is their product $J_0^n(t)$. Inverting this, one gets the density of V for uniform samples, as

$$\begin{aligned} f_0(v) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itv}J_0^n(t)dt \\ &= \frac{1}{\pi}\int_{0}^{\infty} \frac{e^{itv}+e^{-itv}}{2}J_0^n(t)dt \\ &= \frac{1}{\pi}\int_{0}^{\infty} J_0^n(t)\cos(vt)dt. \end{aligned} \tag{3.3.11}$$

By arguments similar to those in (3.3.1) and (3.3.4) one can go from Uniform samples to CN samples. Thus $f_\kappa(v)$, the distribution of V for observations from a CN distribution with $\kappa > 0$ is obtained as

$$f_\kappa(v) = \frac{e^{\kappa v}}{I_0^n(\kappa)}f_0(v) = \frac{e^{\kappa v}}{\pi I_0^n(\kappa)}\int_{0}^{\infty} J_0^n(t)\cos(vt)dt. \tag{3.3.12}$$

3.4 Some Large Sample Results

The exact distributions that we have discussed thus far for the CU and CN samples are rather complicated and somewhat intimidating. However for large samples, the results are considerably simpler. Observe that both C and S are sums of i.i.d. random variables with finite variance. Therefore an easy application of the bivariate Central Limit Theorem yields the bivariate normal as their asymptotic joint distribution, whose parameters are obtained directly from the trigonometric moments of the underlying distribution.

3.4.1 Series Approximations for R and V

In Pearson's random walk problem, the large sample distribution of R, due to Rayleigh (1880), can be obtained by the following simple argument:

Given a sample $\alpha_1, \alpha_2, \ldots, \alpha_n$ from a circular uniform distribution i.e., $(\alpha_1, \ \ldots, \ \alpha_n) \sim U(0, \ 2\pi)$, write $c = \cos\alpha$ and $s = \sin\alpha$. Then we have

$$E(c) = \int_0^{2\pi} \cos\alpha \frac{1}{2\pi} d\alpha = 0,$$

$$E(s) = \int_0^{2\pi} \sin\alpha \frac{1}{2\pi} d\alpha = 0$$

and similarly it can be checked that

$$E(c^2) = \frac{1}{2} = E(s^2), \qquad E(c \cdot s) = 0.$$

Now since the second moment matrix is finite, the multivariate central limit theorem holds and we obtain that $(\overline{C}, \ \overline{S})$ converges to $N_2(\boldsymbol{\mu}, \ \boldsymbol{\Sigma})$ with

$$\boldsymbol{\mu} = \mathbf{0} \ \text{ and } \ \boldsymbol{\Sigma} = \begin{pmatrix} \frac{1}{2n} & 0 \\ 0 & \frac{1}{2n} \end{pmatrix}.$$

It is interesting to note that $\overline{C}$ and $\overline{S}$ become asymptotically independent although $c = \cos\alpha$ and $s = \sin\alpha$ are not independent. Since $\overline{C}$ and $\overline{S}$ converge in distribution to a $N(0, \ \frac{1}{2n})$, $(\sqrt{2n}\ \overline{C})^2$ and $(\sqrt{2n}\ \overline{S})^2$ converge to χ_1^2 each and their sum,

$$Z \equiv \frac{2R^2}{n} = 2n(\overline{C}^2 + \overline{S}^2) \sim \chi_2^2 \quad \text{for large } n. \tag{3.4.1}$$

Using this result, one may write an approximate density of R for large samples, as

$$f(r) = g(z)\left|\frac{dz}{dr}\right| = \frac{1}{2}e^{-\frac{r^2}{n}}\frac{4r}{n} = \frac{2r}{n}e^{-\frac{r^2}{n}}, \tag{3.4.2}$$

with the corresponding distribution function

$$F_n(r) = 1 - e^{-\frac{r^2}{n}}, \qquad 0 \le r \le n.$$

This is known as Rayleigh's approximation for the length of the sample resultant and is valid for large n. A much more accurate Laguerre series expansion due to Pearson (1906) (and quoted by Greenwood and Durand (1955)) is given by

$$\begin{aligned} F_n(r) &= 1 - e^{-z}\left[1 + \frac{1}{2n}\left(z - \frac{z^2}{2!}\right) + \frac{1}{12n^2}\left(-z + \frac{11z^2}{2!} - \frac{19z^3}{3!} + \frac{9z^4}{4!}\right)\right. \\ &\left. + \frac{1}{24n^3}\left(-2z - \frac{4z^2}{2!} + \frac{69z^3}{3!} - \frac{163z^4}{4!} + \frac{145z^5}{5!} - \frac{45z^6}{6!}\right)\right], \end{aligned} \tag{3.4.3}$$

where $z = r^2/n$.

On the other hand if the dispersion is calculated around a specified or hypothesized mean direction μ_0 using

$$V = \sum_{i=1}^{n} \cos(\alpha_i - \mu_0) = R\cos(\overline{\alpha}_0 - \mu_0),$$

its exact distribution is given in (3.3.11). If we assume, without loss of generality, that $\mu_0 = 0$, then clearly $V = C$. Its asymptotic distribution distribution under uniformity, being the sum of i.i.d. components, is normal with mean zero and variance $n/2$. A useful Edgeworth expansion based on Hermite polynomials is given by

$$\begin{aligned} F_n(v) &= P(V \leq v) \\ &= \Phi(z) + \frac{(z^2 - 3z)\phi(z)}{16n} - \frac{(15z + 305z^3 - 125z^5 + 9z^7)\phi(z)}{4608n^2}, \end{aligned} \tag{3.4.4}$$

where $z = \sqrt{\frac{2}{n}}.v$ and $\phi(.)$ and $\Phi(.)$ are the standard normal pdf and cdf, respectively.

3.4.2 Central Limit Type Results

One might ask what corresponds to the Central Limit Theorem for directional variables. Suppose $(\alpha_1, \alpha_2, \ldots, \alpha_n)$ is an i.i.d. sample from a common (non-lattice) circular distribution with cdf $G(\theta)$. Then the distribution of the sum

$$S_n = (\alpha_1 + \cdots + \alpha_n)(\text{mod } 2\pi)$$

converges to the uniform distribution as $n \to \infty$. If $\{\phi_p\}$ corresponds to the characteristic function (or Fourier coefficients) corresponding to $G(\theta)$, we have seen

$$\phi(0) = 1$$

and

$$|\phi(p)| < 1 \quad \text{for } p \neq 0.$$

Therefore for the sum S_n,

$$\phi_{S_n}(0) = 1$$

while

$$\phi_{S_n}(p) = \phi^n(p) \to 0 \quad \text{for } p \neq 0.$$

This, as we have seen is the characteristic function of the uniform distribution on $(0, 2\pi)$, proving the assertion. On the other hand, if one were to normalize the sum differently, then the limiting distribution can be seen to be a WN distribution. For convenience, let the i.i.d. α_i's be on the interval $(-\pi, \pi)$ with $E(\alpha_i) = 0$ and $E(\alpha_i^2) = \sigma^2$, say. Then $n^{-\frac{1}{2}} \sum \alpha_i$ converges to a $N(0, \sigma^2)$ distribution by the Central Limit Theorem and hence

$$S_n^* = \frac{\sum_{i=1}^n \alpha_i}{\sqrt{n}} (\text{mod } 2\pi)$$

will converge to a $WN(0, \sigma^2)$ distribution.

3.4.3 Large Sample Results for Statistics Based on Moments

Given a sample $\alpha_1, \ldots, \alpha_n$, many of the statistics we deal with can be expressed in terms of (C, S) or $(\bar{C}, \bar{S})$. Because all the moments are bounded, the Central Limit Theorem applies to the joint distribution of $(\bar{C}, \bar{S})$. Specifically, suppose $\alpha_1, \ldots, \alpha_n$ are i.i.d. random angles in $[0, 2\pi)$, with common first two trigonometric moments given by

$$Ee^{i\alpha} = \rho e^{i\mu} = (\alpha_1 + i\beta_1), \qquad Ee^{i2\alpha} = \rho_2 e^{i\mu_2} = (\alpha_2 + i\beta_2). \tag{3.4.5}$$

The multivariate central limit theorem applied to the i.i.d. unit vectors $(\cos\alpha_j, \sin\alpha_j)$, $j = 1, \ldots, n$, gives us the result that

$$\frac{1}{\sqrt{n}} \begin{pmatrix} \sum_{j=1}^n (\cos\alpha_j - E\cos\alpha_1) \\ \sum_{j=1}^n (\sin\alpha_j - E\sin\alpha_1) \end{pmatrix} \sim N(0, \Sigma),$$

where

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{CC} & \sigma_{CS} \\ \sigma_{SC} & \sigma_{SS} \end{pmatrix}$$

is the variance-covariance matrix. The elements of Σ can be identified via the first two trigonometric moments of the distribution, given by (3.4.5). Recall that from the definition of the first trigonometric moment,

$$(\alpha_1, \beta_1) = (E\cos\alpha_1, E\sin\alpha_1) = (\rho\cos\mu, \rho\sin\mu). \tag{3.4.6}$$

Using standard trigonometric identities

$$\cos 2\alpha = 2\cos^2\alpha - 1 = 1 - 2\sin^2\alpha,$$

and

$$\sin 2\alpha = 2\sin\alpha\cos\alpha,$$

one can check that

$$\sigma_{CC} = \frac{1}{2} + \frac{1}{2}\rho_2\cos\mu_2 - \rho^2\cos^2\mu = \frac{1}{2}\left(1 + \alpha_2 - 2\alpha_1^2\right),$$

$$\sigma_{CS} = \sigma_{SC} = \frac{1}{2}\rho_2\sin\mu_2 - \rho^2\cos\mu\sin\mu = \frac{1}{2}\left(\beta_2 - 2\alpha_1\beta_1\right), \tag{3.4.7}$$

and

$$\sigma_{SS} = \frac{1}{2} - \frac{1}{2}\rho_2\cos\mu_2 - \rho^2\sin^2\mu = \frac{1}{2}\left(1 - \alpha_2 - 2\beta_1^2\right).$$

For finding the asymptotic distribution of statistics which can be expressed in terms of $(\bar{C}, \bar{S})$, we can appeal to the so-called δ-method which is stated in the following lemma. See for instance Rao (1973), pp. 387–388.

Lemma 3.1 *Let* $(\sqrt{n}(T_{1n}-\theta_1), \ldots, \sqrt{n}(T_{kn}-\theta_k))$ *have asymptotic k-variate normal distribution with mean zero and covariance matrix* $\Sigma = ((\sigma_{ij}))$ *with* $\sigma_{ij} = Cov(T_i, T_j)$, $i = 1, \ldots, k$ *and* $j = 1, \ldots, k$. *Let g be a function of k variables which is totally differentiable. Then,*

$$\sqrt{n}[g(T_{1n}, \ldots, T_{kn}) - g(\theta_1, \ldots, \theta_k)]$$

has the asymptotic normal distribution with mean zero and variance

$$\sum_{i=1}^{k}\sum_{j=1}^{k}\sigma_{ij}\frac{dg}{d\theta_i}\frac{dg}{d\theta_j}.$$

In particular, the following result gives the asymptotic distribution of the sample mean direction, $\overline{\alpha}_0$ defined in (1.3.5).

Proposition 3.1 *Suppose $\alpha_1, \ldots, \alpha_n$ are i.i.d. random angles in $[0, 2\pi)$, with common first two trigonometric moments given by*

$$Ee^{i\alpha} = \rho e^{i\mu}, \qquad Ee^{i2\alpha} = \rho_2 e^{i\mu_2}. \tag{3.4.8}$$

Let $\overline{\alpha}_0$ be the circular mean direction of the unit vectors associated with these random angles, as defined in (1.3.5). Then the limiting distribution of the tangent of this mean direction is given by

$$\sqrt{n}(\tan \overline{\alpha}_0 - \tan \mu) \sim N(0, \sigma^2), \tag{3.4.9}$$

where

$$\sigma^2 = \frac{1 - \rho_2(\cos \mu_2 \cos 2\mu + \sin \mu_2 \sin 2\mu)}{2\rho^2 \cos^4 \mu}$$

is the asymptotic variance.

Proof: Using Lemma 3.1 on the ratio

$$\sum_{j=1}^{n} \sin \alpha_j / \sum_{j=1}^{n} \cos \alpha_j,$$

gives us the limiting variance

$$\zeta^2 = \frac{1}{\alpha_1^2}\left\{\sigma_{CC}\left(\frac{\beta_1}{\alpha_1}\right)^2 + \sigma_{SS} - 2\sigma_{CS}\frac{\beta_1}{\alpha_1}\right\}. \tag{3.4.10}$$

The final expression for σ^2 is obtained after inserting the values of (α_1, β_1) from (3.4.6) and Σ from (3.4.7) into (3.4.10) and with a few additional simplifications. □

3.4.4 First Significant Digit Phenomenon

Benford (1938) observed, through several empirical studies that the first significant (non-zero) digits in a large collection of numbers like the census figures or income numbers etc. does not follow a uniform distribution on numbers $\{1, 2, \ldots, 9\}$ which one might expect. The first significant digit (f.s.d.) has the distribution,

$$P(\text{f.s.d. is} \quad i) = \log_{10}\left(\frac{i+1}{i}\right), \quad i = 1, \ldots, 9.$$

For instance, the odds of the first significant digit being less than or equal to 4 is almost 0.7. This unusual distribution of significant digits can be related to wrapped circular distributions and is sometimes called the "Benford's Law."

Poincaré (1912) observed that the stopping position of a needle which is free to rotate about the center of a disc, is uniformly distributed over different points on the circumference. Suppose that X is the total distance the needle travels after a push. If we assume for convenience, that the wheel is of circumference 1, then the stopping position corresponds to the wrapped r.v.

$$\theta = X(\text{mod } 1).$$

Poincaré's result states that this θ is uniformly distributed if X has a large spread. Such a result is clearly seen to be true for WS families by since the second term in the expression (2.2.18) converges to zero when τ tends to ∞. Somewhat more generally, it can be shown that if X has a continuous distribution and $Y = cX$, then as $c \to \infty$, (i.e., the spread of Y gets large),

$$\phi_Y(p) = \phi_{cX}(p) = \phi_X(cp)$$

which tends to zero for any $p \neq 0$, by the Lebesgue-Reimann theorem.

A consequence of this provides us an explanation for the first significant digit phenomenon. Suppose X denotes the random variable corresponding to the data. Then

$$\theta = \log_{10} X(\text{mod } 1)$$

tends to a uniform distribution by Poincaré's result since $\log_{10} X$ has large spread. Since the first significant digit of a number X is i ($i = 1, 2, \ldots, 9$) if and only if $i \cdot 10^k \leq X < (i+1) \cdot 10^k$ for some integer k,

$$\begin{aligned}
P(\text{f.s.d. of } X = i) &= P(i \cdot 10^k \leq X < (i+1) \cdot 10^k \text{ for some integer } k) \\
&= P(\log_{10} i + k \leq \log_{10} X < \log_{10}(i+1) + k) \\
&= P(\log_{10} i \leq \theta < \log_{10}(i+1)) \\
&= \log_{10}(i+1) - \log_{10} i \\
&= \log_{10}\left(\frac{i+1}{i}\right), \quad i = 1, \ldots, 9.
\end{aligned}$$

For more recent discussions see Raimi (1969) and Hill (1995).

3.5 Two or More Samples

Suppose there are q $(q \geq 2)$ Circular Normal populations each with the same concentration parameter κ but with different polar directions $\mu_1, \ldots, \mu_q$. Let

$$\{\alpha_{i,j} : \ j = 1, \ \ldots, \ n_i, \ i = 1, \ \ldots, \ q\}$$

denote the observations from these populations. Let R_i denote the length of the resultant for the i^{th} population based on a sample of size n_i and R denote the length of the overall resultant based on all the $N = \sum n_i$ observations. The main result of this section is to show that the conditional distribution $f_\kappa(R_1, \ldots, \ R_q|R)$ of the individual sample resultants gives the length of the overall resultant, is independent of κ for CN samples. This conditional distribution, being independent of the nuisance parameter κ, is suitable for inferences on the mean directions of the q populations when κ is unknown.

Consider again the uniform model (with subscript 0 corresponding to $\kappa = 0$) and note that in this case, the overall resultant vector is composed of individual sample resultants with lengths $R_1, \ \ldots, \ R_q$ with random angles between them. Thus, we can apply the result of the generalized random walk (see Equation (3.2.3)) to conclude, *conditionally*

$$f_0(R|R_1, \ \ldots, \ R_q) = R \int_0^\infty J_0(Rt) \left[\prod_{i=1}^{q} J_0(R_i t) \right] t dt.$$

Now suppose there is a common mean direction μ for all the samples (in testing problems, this might be the null hypothesis of interest – so that this provides the null distribution), and let c denote as before, the *cosine* of the difference in the overall sample resultant from the polar direction. Combining known results on the lengths R_i and c viz.,

$$\begin{aligned} f_0(c) &= \frac{1}{\pi\sqrt{1-c^2}}, \qquad -1 \leq c \leq 1, \\ f_0(R_i) &= R_i \int_0^\infty J_0(R_i t) J_0^{n_i}(t) t dt \\ &= R_i \psi_{n_i}(R_i), \end{aligned}$$

and their independence under uniformity, we get the joint density

$$\begin{aligned}
& f_0(R_1, \ldots, R_q, R, c) \\
= \; & f_0(R|R_1, \ldots, R_q) \prod_{i=1}^{q} f_0(R_i) f_0(c) \\
= \; & \frac{1}{\pi\sqrt{1-c^2}} R \int_0^\infty J_0(Rt) \prod_{i=1}^{q} J_0(R_i t) t dt \prod_{i=1}^{q} R_i \psi_{n_i}(R_i). \qquad (3.5.1)
\end{aligned}$$

Appealing to Equation (3.3.1), we have

$$\begin{aligned}
f_\kappa(R_1, \ldots, R_q, R) &= \int_{-1}^{1} f_\kappa(R_1, \ldots, R_q, R, c) dc \\
&= \int_{-1}^{1} \frac{e^{\kappa R c}}{I_0^N(\kappa)} f_0(R_1, \ldots, R_q, R, c) dc \\
&= \frac{I_0(\kappa R)}{I_0^N(\kappa)} R \int_0^\infty J_0(Rt) \prod_{i=1}^{q} J_0(R_i t) t dt. \qquad (3.5.2)
\end{aligned}$$

The last equality follows from the fact that

$$\frac{1}{\pi} \int_{-1}^{1} \frac{e^{\kappa R c}}{\sqrt{1-c^2}} dc = \frac{1}{2\pi} \int_0^{2\pi} e^{\kappa R \cos\theta} d\theta = I_0(\kappa R).$$

From Equation (3.3.7), we have

$$f_\kappa(R) = \frac{I_0(\kappa R)}{I_0^N(R)} R \psi_N(R)$$

so that the conditional distribution

$$\begin{aligned}
& f_\kappa(R_1, \ldots, R_q|R) \\
= \; & \frac{f_\kappa(R_1, \ldots, R_q, R)}{f_\kappa(R)} \\
= \; & \frac{1}{\psi_N(R)} \int_0^\infty J_0(Rt) \prod_{i=1}^{q} J_0(R_i t) t dt \prod_{i=1}^{q} R_i \psi_{n_i}(R_i) \qquad (3.5.3)
\end{aligned}$$

is independent of κ. Therefore this conditional density may be used for constructing exact tests of significance for mean directions of CN distributions.

The special case for q=2 was obtained by Watson and Williams (1956) and can be derived from Equation (3.5.3) as follows. When q=2, (3.5.3) gives

$$\begin{aligned} & f_\kappa(R_1,\ R_2|R) \\ = & \frac{1}{\psi_N(R)}\int_0^\infty J_0(Rt)J_0(R_1t)J_0(R_2t)tdtR_1\psi_{n_1}(R_1)R_2\psi_{n_2}(R_2). \end{aligned} \quad (3.5.4)$$

Now, from properties of Bessel functions, it is known that

$$\int_0^\infty J_0(at)J_0(bt)J_0(ct)dt = \begin{cases} \frac{1}{2\pi A} & \text{if } a,\ b,\ c \text{ make a triangle with area } A, \\ 0 & \text{otherwise.} \end{cases}$$

In this case, R_1, R_2, and R do make a triangle with area

$$\begin{aligned} A &= \sqrt{s(s-R_1)(s-R_2)(s-R)} \\ &= \frac{1}{4}\sqrt{\{(R_1+R_2)^2-R^2\}\{R^2-(R_1-R_2)^2\}}, \end{aligned}$$

so that

$$\begin{aligned} f_\kappa(R_1,\ R_2|R) &= \frac{1}{2\pi A}\frac{R_1\psi_{n_1}(R_1)R_2\psi_{n_2}(R_2)}{\psi_N(R)} \\ &= \frac{2R_1\psi_{n_1}(R_1)R_2\psi_{n_2}(R_2)}{\pi\psi_N(R)\sqrt{((R_1+R_2)^2-R^2)(R^2-(R_1-R_2)^2)}}. \end{aligned}$$

3.6 Approximate Distributions for Large κ

Recall from Proposition 2.2 that for large κ, a CND can be approximated by a normal distribution. Specifically, if $\alpha \sim CN(\mu,\ \kappa)$, then

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

or $\alpha \sim N(\mu, 1/\kappa)$ for large κ. This allows one to use the normal distribution results for suitably large κ.

If $\alpha_1,\ \ldots,\ \alpha_n$ are observations from $CN(\mu,\ \kappa)$, then

$$\theta_i = (\alpha_i-\mu) \sim N(0,\ 1/\kappa),$$

for sufficiently large κ. Since

$$\cos\theta_i \simeq (1 - \theta_i^2/2), \quad \text{or} \quad \theta_i^2 \simeq 2(1 - \cos\theta_i)$$

for small θ_i, from the normal distribution theory

$$(\sqrt{\kappa}\theta_i)^2 = \kappa\theta_i^2 \simeq 2\kappa(1 - \cos\theta_i) \sim \chi_1^2.$$

Summing over all i, we get

$$2\kappa(n - V) \simeq \chi_n^2, \tag{3.6.1}$$

where $V = \sum\cos\theta_i = \sum\cos(\alpha_i - \mu)$. Recall that $(n - V)$ is the sample dispersion about the population mean direction. Also since, $\theta_i \simeq \sin\theta_i$ are independent,

$$\sum_{i=1}^{n}\sin\theta_i \sim N(0, n/\kappa)$$

for sufficiently large κ. From the normal theory again, it follows that

$$\frac{\kappa}{n}\left(\sum_{i=1}^{n}\sin\theta_i\right)^2 \sim \chi_1^2.$$

Since R is invariant over location shifts, i.e., it has the same value based on the $\{\alpha_i\}$s as it has for the shifted $\{\theta_i\}$s, so that

$$R^2 = \left(\sum_{i=1}^{n}\sin\theta_i\right)^2 + \left(\sum_{i=1}^{n}\cos\theta_i\right)^2.$$

Thus we have

$$\frac{\kappa}{n}\left(\sum_{i=1}^{n}\sin\theta_i\right)^2 = \frac{\kappa}{n}\left(R^2 - (V)^2\right) \simeq 2\kappa(R - V) \sim \chi_1^2.$$

The last approximation comes from the fact that for large κ, $R \simeq V \simeq n$ with high probability, so that

$$R^2 - V^2 = (R - V)(R + V) \simeq 2n(R - V).$$

We thus have

$$2\kappa(R - V) \sim \chi_1^2. \tag{3.6.2}$$

Equations (3.6.1) and (3.6.2) lead to the useful χ^2 decomposition

$$2\kappa(n-V) = 2\kappa(n-R) + 2\kappa(R-V) \tag{3.6.3}$$

with the n degrees of freedom on the left hand side split into two terms on the right hand side, one with $(n-1)$ df and the other with 1 df, respectively.

Remark 3.1 *A more direct way to justify these approximations is that when κ is sufficiently large, all observations lie in an arc of sufficiently small length near the polar direction. In such a case, Taylor series approximations for* $\sin\theta$ *and* $\cos\theta$ *may be used with $\theta_i = (\alpha_i - \mu)$. The usual sample variance $s^2 = \sum(\alpha_i - \overline{\alpha})^2/n$ is related to the resultant length as in Equation (1.3.12). Thus the standard result on the distribution of the sample variance for normal samples (with σ^2 replaced by $1/\kappa$) may be invoked to give*

$$\frac{ns^2}{\sigma^2} \simeq 2\kappa(n-R) \sim \chi^2_{n-1}.$$

If on the other hand, the mean direction μ is known, we would use $s_0^2 = \sum(\alpha_i - \mu)^2/n \simeq 2(n-V)/n$ and the corresponding result for the normal samples gives,

$$\frac{ns_0^2}{\sigma^2} \simeq 2\kappa(n-V) \sim \chi^2_n.$$

These approximate results will be found quite useful and are clearly much easier to apply than the exact distribution theory discussed in previous sections. They will be used to construct some useful approximate tests in Chapter 5.

Remark 3.2 *Since these approximations (3.6.1), (3.6.2) and (3.6.3) are not likely to be good for moderately large κ, one may improve on them by replacing κ in these equations by γ, a function of κ, so that the expected values on both sides match i.e.,*

$$E\left[2\gamma(n-v)\right] = n.$$

Such improved approximations have been empirically validated by Stephens (1969b) even for the values of κ as small as 2, when γ is chosen to be

$$\frac{1}{\gamma} = \frac{1}{\kappa} + \frac{3}{8\kappa^2}. \tag{3.6.4}$$

Chapter 4

Estimation of Parameters

4.1 Introduction

In this chapter we consider the problem of estimating the parameters for some circular distributions. The CN distribution is considered in some detail, in which maximum likelihood and the moment estimators of the parameters μ and κ are seen to be the same. The properties of best equivariance and admissibility of these estimators are established. Mixtures of von Mises distributions are briefly discussed for which the EM algorithm provides the MLEs. The ML method however leads to complicated iterative solution for the circular Beta distribution and specific members of the wrapped stable (WS) family and becomes quite unwieldy for the three-parameter WS family of distributions. The method of moments is advocated for such cases. Derivation of "efficient" estimators here remains an open problem. One may also use bootstrap methods and for a discussion of bootstrap methods in connection with directional analysis, see Fisher and Hall (1989).

4.2 CN Distribution

4.2.1 Estimating the Parameters of a CN Distribution

Let $\alpha_1, \alpha_2, \ldots, \alpha_n$ be a set of observations from a CN distribution with parameters μ and κ, i.e., $\alpha_1, \alpha_2, \ldots, \alpha_n$ are i.i.d. with pdf

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

The likelihood function is given by

$$L(\mu,\ \kappa|\alpha_1,\ \alpha_2,\ \ldots,\ \alpha_n) = \frac{1}{[2\pi I_0(\kappa)]^n} e^{\sum_{i=1}^n \kappa\cos(\alpha_i-\mu)},$$

and its logarithm,

$$l = \log L = -n\log(2\pi I_0(\kappa)) + \kappa\sum_{i=1}^n \cos(\alpha_i - \mu).$$

Differentiating this with respect to μ and κ, respectively, and equating to zero, we obtain the likelihood equations

$$\frac{\partial l}{\partial \mu} = \sum_{i=1}^n \sin(\alpha_i - \mu) = 0, \tag{4.2.1}$$

and

$$\frac{\partial l}{\partial \kappa} = -n\frac{I_1(\kappa)}{I_0(\kappa)} + \sum_{i=1}^n \cos(\alpha_i - \mu) = 0, \tag{4.2.2}$$

using the fact that

$$\frac{dI_0(\kappa)}{d\kappa} = I_1(\kappa),$$

the modified Bessel function of order 1. From Equation (4.2.1), we have

$$\begin{aligned}\sum_{i=1}^n (\sin\alpha_i\cos\mu - \cos\alpha_i\sin\mu) &= 0\\ \text{or,}\quad S\cos\mu - C\sin\mu &= 0,\end{aligned}$$

so that,

$$\hat{\mu} = \arctan^*\left(\frac{S}{C}\right) = \overline{\alpha}_0, \tag{4.2.3}$$

where arctan* is as defined in Equation (1.3.5). Plugging this $\hat{\mu}$ into the other Equation (4.2.2) and noting that $\sum_{i=1}^n \cos(\alpha_i - \overline{\alpha}_0) = R$ from Theorem 1.1, we get

$$\begin{aligned}-n\frac{I_1(\kappa)}{I_0(\kappa)} + R &= 0\\ \text{or,}\quad \frac{I_1(\kappa)}{I_0(\kappa)} &= \frac{R}{n},\end{aligned} \tag{4.2.4}$$

so that $\hat{\kappa}$ is obtained as the solution of Equation (4.2.4). Let $A(\kappa)$ denote the ratio of the two Bessel functions,

$$A(\kappa) = \frac{I_1(\kappa)}{I_0(\kappa)}. \tag{4.2.5}$$

It has the following properties (see Appendix A):

- $0 \leq A(\kappa) \leq 1$;
- $A(\kappa) \longrightarrow 0$ as $\kappa \longrightarrow 0$ and $A(\kappa) \longrightarrow 1$ as $\kappa \longrightarrow \infty$;
- $A'(\kappa) \equiv \partial A(\kappa)/\partial\kappa = (1 - \frac{A(\kappa)}{\kappa} - A^2(\kappa)) \geq 0$,

i.e., $A(\kappa)$ is a strictly monotone increasing function of κ so that $\hat{\kappa}$, may be obtained as the unique solution to Equation (4.2.4). For different values of κ the values of $I_1(\kappa)/I_0(\kappa)$ are obtained from program `A1` in our `SPlus` library. It can be checked that the Hessian matrix, i.e., the matrix of second derivatives of l, evaluated at $(\hat{\mu}, \hat{\kappa})$, is the diagonal matrix,

$$\mathrm{Diag}(-\hat{\kappa}R, \ -nA'(\hat{\kappa}))$$

which is negative definite. Thus $\hat{\mu}$ and $\hat{\kappa}$ are indeed the MLEs of μ and κ, respectively. Using the fact that $A'(\kappa) = 1 - A(\kappa)/\kappa - A^2(\kappa)$, the Fisher information matrix in this case, is the diagonal matrix

$$I = \begin{bmatrix} \frac{A(\kappa)}{\kappa} & 0 \\ 0 & 1 - \frac{A(\kappa)}{\kappa} - A^2(\kappa) \end{bmatrix}. \tag{4.2.6}$$

Thus the asymptotic variance-covariance matrix for the MLE's evaluated at $(\hat{\mu}, \hat{\kappa})$ is given by the diagonal matrix

$$\begin{bmatrix} \frac{1}{R\hat{\kappa}} & 0 \\ 0 & \frac{1}{n(1-\frac{\overline{R}}{\hat{\kappa}}-\overline{R}^2)} \end{bmatrix}. \tag{4.2.7}$$

Remark 4.1 *Method of moments estimation involves equating the theoretical trigonometric moments as defined in Equation (2.1.1) with the sample trigonometric moments defined in Equation (1.3.13), and solving for the unknown parameters after setting up as many equations as needed. In this case, equating*

$$E(\cos\alpha) = A(\kappa)\cos\mu,$$

$$E(\sin\alpha) = A(\kappa)\sin\mu$$

to the sample moments $\overline{C}$ and $\overline{S}$, respectively, results again in Equations (4.2.3) and (4.2.4), so that the trigonometric moment estimators coincide with the maximum likelihood estimators for this model.

Remark 4.2 *The MLE of μ remains the same whether or not κ is known. On the other hand, the ML estimate of κ is different when μ is known. The case μ is known results in the likelihood equation*

$$A(\kappa) = V/n \qquad \textit{where } V = \sum_{i=1}^{n} \cos(\alpha_i - \mu).$$

Note that V may be negative in which case this equation is not meaningful. However, by virtue of the properties of $A(\kappa)$ listed above, L is a monotonically decreasing function of κ. Thus, in this case we get the MLE of κ to be,

$$\hat{\kappa} = \begin{cases} A^{-1}(V/n), & \textit{if } V > 0, \\ 0, & \textit{if } V \leq 0. \end{cases} \tag{4.2.8}$$

4.2.2 Optimal Properties of the MLEs

We now explore the performance of the above estimators. Note that though the inverse of the variance is a reasonable measure for exploring the efficiency of $\hat{\kappa}$ which takes values in $\mathbb{R}^+$, the same is not true for $\hat{\mu}$ which is circular. Hence, a circular measure of efficiency needs to be used. Recall from (1.3.7) that $[1 - \cos(\hat{\mu} - \mu)]$ is a measure of the deviation of $\hat{\mu}$ from μ. Thus a reasonable loss function to use in this context would be

$$L(\mu, a) = 1 - \cos(a - \mu),\ 0 \leq a, \mu < 2\pi.$$

We will refer to this as the *"Circular Loss"*.

Using this measure, SenGupta and Maitra (1998) study the best equivariance and admissibility of the MLE of the mean direction for a single CN distribution and in simultaneous estimation with several independent circular normal distributions, all having the same concentration parameter.

We have seen that for κ known, CN(μ, κ) is a member of a (1, 2) curved exponential family. Following the arguments in Kariya (1989), MLE for μ can be shown to be the Best Equivariant Estimator (BEE) under this circular loss. This result also holds for the case when κ is unknown. These results are generalized in the following theorem which establishes the Best Equivariant nature of the MLE of $\underset{\sim}{\mu} = (\mu_1, \cdots, \mu_p)'$ in simultaneous estimation with several independent $\text{CN}(\mu_i, \kappa)$, $i = 1, \ldots, p$ populations for both κ known and unknown cases. The admissibility of the MLE for μ in $CN(\mu, \kappa)$ for a single population follows from Bagchi (1987), Zhong (1991) or as a special case of the results in Watson (1983b) and is a consequence of the Bayes character of the MLE with respect to a uniform prior. Admissibility of the simultaneous MLE $\underset{\sim}{\widehat{\mu}}$ for $\underset{\sim}{\mu} = (\mu_1, \cdots, \mu_p)'$ in p independent $CN(\mu_i, \kappa)$, $i = 1, \ldots, p$ is again established in the following result. The proof of Theorem 4.1 is somewhat involved and is deferred to the end of this chapter.

Theorem 4.1 *(SenGupta and Maitra (1998)) The MLE $\underset{\sim}{\widehat{\mu}}$ of $\underset{\sim}{\mu}$ in the simultaneous estimation problem with p independent $CN(\mu_i, \kappa)$, $i = 1, \ldots, p$, κ known or unknown, is (i) the best equivariant and (ii) an admissible estimator.*

Remark 4.3 *In particular, for the case $p = 1$, the estimator $\bar{\alpha}_0$ is the best equivariant, admissible and minimax estimator of mean direction μ in $CN(\mu, \kappa)$ population whether κ is known or unknown.*

Remark 4.4 *Recall that as $\kappa \longrightarrow \infty$, $\sqrt{\kappa}(\theta - \mu) \xrightarrow{\mathcal{L}} N(0, 1)$ for $\theta \sim CN$ (μ, κ). In view of James-Stein type estimators one would expect the simultaneous MLE to be inadmissible for $p \geq 3$, at least for large κ. Admissibility of the simultaneous MLE as stated in Theorem 4.1 is thus rather counter-intuitive and marks a departure from the normal theory.*

Remark 4.5 *For the sake of completeness, we make a few observations on estimation of the concentration parameter κ. Note that for μ known, say $\mu = 0$, $CN(0, \kappa)$ is a member of the one-parameter regular exponential family (REF) with κ as its canonical parameter. This yields the canonical statistic, $\cos\theta$, as an admissible estimator for κ. However this is not a sensible estimator of κ, since $0 \leq \kappa < \infty$. Other optimality properties of the MLE for the concentration parameter in CN as well as the Langevin distributions, need further investigation. An excellent reference in this context is Brown (1986), in particular pages 136-137, 150 and 204-205.*

4.3 CN Mixtures

Finite mixtures of $CN(\mu_i, \kappa_i)$ with mixing parameters p_i, $i = 1, \ldots, k$ provide a much richer class of circular models with density

$$g(\alpha) = \sum_{i=1}^{k} p_i \frac{e^{\kappa_i \cos(\alpha - \mu_i)}}{2\pi I_0(\kappa_i)}, \ 0 \leq \alpha < 2\pi.$$

Since finite mixtures are known to be identifiable (see e.g. Frazer, Hsu and Walker (1981)), the MLEs can be found by the standard EM algorithm. See Section 4.3.2 of Titterington et al. (1985).

However, the EM algorithm may encounter difficulties in case the concentration parameters κ_i, $i = 1, \ldots, k$ are not well separated - see, e.g. Spurr (1981). In such a situation, the method of trigonometric moments may be used. Spurr (1981) considered the above mixture model for axial data, i.e. $0 \leq \alpha < \pi$ with two components ($k = 2$) and equal κ_i's and noted that a procedure based on the method of moments gave a much simpler yet comparable results to the maximum likelihood method.

The special case $k = 2$, $p_1 = p_2 = \frac{1}{2}$, and $\mu_1 = -\mu_2 = \mu$, referred to in (2.2.21), was considered by Bartels (1984) who discusses the maximum likelihood estimation. From the density (2.2.21), the loglikelihood is given by

$$L(\mu,\ \kappa) = C - n \log \{I_0(\kappa)\} + \sum_{i=1}^{n} \log \left[e^{\kappa \cos(\alpha_i - \mu)} + e^{\kappa \cos(\alpha_i + \mu)} \right]. \qquad (4.3.1)$$

The function is symmetric about $\mu = 0$ and hence, to identify the parameters uniquely, we restrict μ to the interval $[0,\ \pi]$.

An important feature of this loglikelihood function is that $(\partial L / \partial \mu) = 0$ when $\mu = 0$, regardless of the sample values. This means that when $\mu = 0$, $E\{(\partial L/\partial \mu)^2\}$ and $E\{(\partial L/\partial \mu)(\partial L/\partial \kappa)\}$ are both 0, so that the matrix is singular. The conditions for the consistency of $\hat{\mu}$ are still satisfied (see Kendall and Stuart (1979)) while those for asymptotic efficiency or asymptotic normality are not. This does not mean that finite-sample maximum likelihood estimators (MLEs) are necessarily bad, however we cannot rely on the asymptotic theory of MLEs to give us approximate standard errors or confidence intervals for μ when the true μ is near zero. This nonstandard character of the likelihood applies equally to any member of the class of mirror-image distributions.

The fact that $(\partial L/\partial \mu) = 0$ at $\mu = 0$ means that L will always have a local maximum or minimum at $\mu = 0$. Consequently, in the numerical optimization of L, care should be taken to avoid mistaking the local turning point at $\mu = 0$ for the global maximum. There is a positive probability that L, in fact, has its global maximum at $\mu = 0$, and this probability increases as the true value of μ approaches 0. The sampling distribution of the MLE $\hat{\mu}$ will, as a result, have a lump of probability at $\hat{\mu} = 0$, with the remaining probability distributed over $\hat{\mu} > 0$.

Some of the problems arising from the lack of regularity of L can be overcome by a reparametrization of (2.2.21) to obtain the density

$$h(x;\ A,\ B) = C \exp(A \cos x) \cosh(\sqrt{B} \sin x), \tag{4.3.2}$$

where $-\infty < A < \infty$ and $B \geq -\pi^2/4$. If we write $A = \kappa \cos \mu$ and $B = (\kappa \sin \mu)^2$ then, for $B \geq 0$, (4.3.2) corresponds to this bimodal distribution. For $-\pi^2/4 \leq B < 0$, we can use $\cosh iy = \cos y$ to show that (4.3.2) still defines a density, which is platykurtic and unimodal. Unfortunately, the parameters A and B are not as easily interpreted as μ and κ. However, it is easy to verify that $h(x;\ A,\ B)$ is regular, and that $B > 0$ or $B = 0$ as $\mu > 0$ or $\mu = 0$, respectively, (provided we assume $\kappa \neq 0$). Note that if we restrict the MLE of B to be nonnegative, in order to ensure correspondence of (4.3.2) with this density, then when $B = 0$ the asymptotic distribution of $\hat{B}$ consists of a discrete probability of $1/2$ at $\hat{B} = 0$, with the remainder distributed as a half-normal over $\hat{B} > 0$ (Chernoff (1951)). Consequently, when $\mu = B = 0$ we have $P(\hat{\mu} = 0) = P(\hat{B} = 0) = 1/2$ asymptotically.

4.4 ML Estimation for the WC Distribution

Recall the WC density given in (2.2.16) with the location parameter $0 \leq \mu < 2\pi$ and scale parameter $0 \leq \rho < 1$. The ML estimates of μ and ρ can be obtained by a recursive algorithm given by Kent and Tyler (1988). We first reparametrize this WC density by putting

$$\mu_1 = \frac{2\rho \cos \mu}{(1 + \rho^2)}, \qquad \mu_2 = \frac{2\rho \sin \mu}{(1 + \rho^2)},$$

obtaining

$$f(\theta; \mu_1, \mu_2) = \frac{1}{[2\pi \cdot c \cdot (1 - \mu_1 \cos \theta - \mu_2 \sin \theta)]},$$

where

$$c = c(\mu_1, \mu_2) = \frac{1}{\sqrt{1-\mu_1^2-\mu_2^2}}.$$

Consider now a random sample $\theta_1, \ldots, \theta_n$ from a WC distribution. To obtain the likelihood equations it is simpler to introduce another parametrization $\eta_1 = c \cdot \mu_1$ and $\eta_2 = c \cdot \mu_2$. Differentiating the log likelihood function with respect to η_1 and η_2 and noting that $c = \sqrt{1+\eta_1^2+\eta_2^2}$ leads to the likelihood equations

$$\begin{aligned} \frac{1}{c}\sum_{i=1}^{n} w_i \left[\cos\theta_i - \mu_1\right] &= 0, \\ \frac{1}{c}\sum_{i=1}^{n} w_i \left[\sin\theta_i - \mu_2\right] &= 0, \end{aligned} \tag{4.4.1}$$

where $w_i = 1/(1 - \mu_1 \cos\theta_i - \mu_2 \sin\theta_i)$ for $i = 1, \ldots, n$. These equations can be written to express μ_1 and μ_2 as adaptively weighted averages $\cos\theta_i$ and $\sin\theta_i$, respectively; namely,

$$\begin{aligned} \mu_1 &= \left\{\sum_{i=1}^{n} w_i \cos\theta_i\right\} \Big/ \left\{\sum_{i=1}^{n} w_i\right\}, \\ \mu_2 &= \left\{\sum_{i=1}^{n} w_i \sin\theta_i\right\} \Big/ \left\{\sum_{i=1}^{n} w_i\right\}. \end{aligned} \tag{4.4.2}$$

This representation suggests the following iterative re-weighting algorithm for computing the maximum likelihood estimates $\hat{\mu}_1$ and $\hat{\mu}_2$:

(1) Start with arbitrary initial values $\mu_{1,0}$ and $\mu_{2,0}$ with $\mu_{1,0}^2 + \mu_{2,0}^2 < 1$;

(2) Given $\mu_{1,\nu}$ and $\mu_{2,\nu}$ at iteration ν, update the values via

$$\begin{aligned} \mu_{1,\nu+1} &= \left\{\sum_{i=1}^{n} w_{i,\nu} \cos\theta_i\right\} \Big/ \left\{\sum_{i=1}^{n} w_{i,\nu}\right\}, \\ \mu_{2,\nu+1} &= \left\{\sum_{i=1}^{n} w_{i,\nu} \sin\theta_i\right\} \Big/ \left\{\sum_{i=1}^{n} w_{i,\nu}\right\}, \end{aligned} \tag{4.4.3}$$

where $w_{i,\nu} = 1/(1 - \mu_{1,\nu} \cos\theta_i - \mu_{2,\nu} \sin\theta_i)$ for $i = 1, \ldots, n$;

(3) Repeat (2) until the algorithm converges, giving $\hat{\mu}_1$ and $\hat{\mu}_2$.

The existence and uniqueness of the maximum likelihood estimates and the convergence of this algorithm are assured by noting that the WC distribution are related to the angular central Gaussian distribution on the circle (i.e., the distribution of $\mathbf{X}/\| \mathbf{X} \|$ where $\mathbf{X}$ is a bivariate normal), and by using known results for the angular central Gaussian. Simulation studies indicate that very small values of ρ tend to make the estimates rather unreliable.

The routine `wrcauchy.ml` in `CircStats` can find the ML estimates for WC distribution as illustated in the following example.

Example 4.1 *This data set on the direction of flight of homing pigeons, provided by Schmidt-Koenig (1963) , has already been mentioned in Section 1.1. The vanishing angles of 15 birds released singly are given (in increasing order).*

```
> pigeon_c(85,135,135,140,145,150,150,150,160,185,
+ 200,210,220,225,270)
> wrpcauchy.ml(rad(pigeon),4,.5) # mu0 = 4, rho0 = .5
        mu       rho
1 2.633061 0.6921546
> wrpcauchy.ml(rad(pigeon),4,.2) # mu0 = 4, rho0 = .2
        mu       rho
1 2.633061 0.6921546
> wrpcauchy.ml(rad(pigeon),1,.2) # mu0 = 1, rho0 = .2
        mu       rho
1 2.633061 0.6921546
> wrpcauchy.ml(rad(pigeon),0,.1) # mu0 = 0, rho0 = .1
        mu       rho
1 2.633061 0.6921546
> deg(2.633061)
[1] 150.8633
```

Here μ_0 and ρ_0 are initial estimates of μ and ρ, respectively.

4.5 Circular Beta Distribution

Parameter estimation for the circular beta distribution defined by the pdf (2.2.23), has been considered in Lai (1994). One may use the method of moments or the ML approach. In either approach, it is first assumed that the modal vector is known and when it is unknown, it is replaced by the normalized sample mean vector.

Moment Estimation

(i) When the modal vector $\boldsymbol{\mu}$ is *known*. By equating the sample moments to the population moments for $\{x_i, i = 1, \ldots, n\}$, we have

$$m_1 = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\mu}' x_i = \frac{\tilde{\alpha} - \tilde{\beta}}{\tilde{\alpha} + \tilde{\beta}}$$

and

$$m_2 = \frac{1}{n}\sum_{i=1}^{n} (\boldsymbol{\mu}' x_i)^2 = \frac{(\tilde{\alpha} + \tilde{\beta}) + (\tilde{\alpha} - \tilde{\beta})^2}{(\tilde{\alpha} + \tilde{\beta})(\tilde{\alpha} + \tilde{\beta} + 1)},$$

where the superscript '$\sim$' above a parameter denotes its moment estimate. These two equations yield the moment estimates for α and β as

$$\tilde{\alpha} = \frac{1}{2}\frac{(1 + m_1)(1 - m_2)}{m_2 - m_1^2}$$

and

$$\tilde{\beta} = \frac{1}{2}\frac{(1 - m_1)(1 - m_2)}{m_2 - m_1^2}.$$

(ii) When the mean vector is *unknown*, we plug in the estimate

$$\tilde{\boldsymbol{\mu}} = \frac{\sum_{i=1}^{N} x_i}{\|\sum_{i=1}^{N} x_i\|}$$

in place of μ and use the same equations above.

Maximum Likelihood Method

If $\{x_1, x_2, \ldots, x_n\}$ is a random sample of size n from a directional beta population, then the log likelihood function is

$$\begin{aligned}\log L &= n \log \Gamma(\alpha + \beta) - n(\alpha + \beta) \log 2 - \log \Gamma(\alpha) - \log \Gamma(\beta) \\ &\quad + \left(\alpha - \frac{1}{2}\right) \sum_{i=1}^{n} \log (1 + \boldsymbol{\mu}' x_i) + \left(\beta - \frac{1}{2}\right) \sum_{i=1}^{n} \log (1 - \boldsymbol{\mu}' x_i).\end{aligned}$$

Putting $\cos\theta_i = \mu' x_i$ and differentiating with respect to α and β, we have

$$\phi(\alpha) - \phi(\alpha+\beta) = \frac{1}{n}\sum_{i=1}^{n} \log\left(\frac{1+\cos\theta_i}{2}\right)$$

and

$$\phi(\beta) - \phi(\alpha+\beta) = \frac{1}{n}\sum_{i=1}^{n} \log\left(\frac{1-\cos\theta_i}{2}\right),$$

where $\phi(x)$ is the di-gamma function defined by

$$\phi(x) = d\log\Gamma(x)/dx.$$

Efficient iterative techniques like the simplex minimization algorithm (see Nelder and Mead (1965)), using the moment estimates as our initial guesses, may be used to solve the above estimating equations.

4.6 WS Family

For the more general WS family with "α" also unknown, the method of moments gives

$$\hat{\mu} = \overline{\alpha}_0, \tag{4.6.1}$$

$$\hat{\rho} = \overline{R}_1, \tag{4.6.2}$$

$$\hat{\alpha} = \frac{1}{\log 2}\log\left(\frac{\log\overline{R}_2}{\log\overline{R}_1}\right), \tag{4.6.3}$$

where

$$\overline{R}_j = \frac{1}{n}\sum_{i=1}^{n}\cos j(\alpha_i - \overline{\alpha}_0), \quad j = 1,\ 2, \quad \overline{R}_1 \equiv \overline{R}.$$

Lemma 4.1 *$\overline{\alpha}_0$ and $\overline{R}$ provide robust moment estimators for μ_0 and ρ for the Cardioid, WN, WC and in general, for the entire family of WS distributions. Further, $\hat{\rho}$ and $\hat{\alpha}$ are consistent, asymptotically normal (CAN) estimators.*

Proof: The first part follows directly since the corresponding first two defining trigonometric moment equations for μ and ρ do not involve the stable index α. The second part follows by applications of the central limit theorem and the δ-method. □

For the moment estimators given in (4.6.1)–(4.6.3), their standard errors can be derived by using the central limit theorem and the δ-method. Gatto and Jammalamadaka (2000) discuss asymptotic inference for $\hat{\mu}$ given in (4.6.1) conditional on R, which makes it independent of the nuisance parameter. Saddlepoint methods are used to set accurate confidence intervals for μ, even when the sample sizes are moderate.

Remark 4.6 *Note that $\hat{\alpha}$ needs to be properly truncated to be a meaningful estimator of α. Derivation of maximum likelihood and other efficient estimators here, is an open problem.*

Remark 4.7 *Since the WC distribution admits a simple representation, not involving an infinite sum, it was possible to find an iterative scheme for obtaining the MLEs of its parameters. The estimation of the parameters for an arbitrary member of the WS family by the method of maximum likelihood does not result in simple closed form expressions. Recall however that an observation θ from a WS distribution can be looked upon as an observation x from the corresponding unwrapped distribution being wrapped, say k times. This k however is an unobservable integer, giving rise to a missing data framework. Hence, an EM algorithm may be used as was done by Breckling (1989), Fisher and Lee (1994) for dealing with estimation in wrapped distributions.*

4.7 Confidence Intervals

Consider first the CN distribution and the construction of a confidence interval (CI) for the mean direction μ. One method is to consider the circular 'standard error' of the MLE $\hat{\mu}$ for the CN distribution for large samples which is given by $\hat{\sigma}_{\hat{\mu}} = 1/\sqrt{n\bar{R}\hat{\kappa}}$. Using this, a $(1-\alpha)100\%$ CI for μ is given by

$$\left[\hat{\mu} - \arcsin\left(\tau_{\alpha/2}\hat{\sigma}_{\hat{\mu}}\right),\ \hat{\mu} + \arcsin\left(\tau_{\alpha/2}\hat{\sigma}_{\hat{\mu}}\right)\right].$$

A large-sample CI for κ, the concentration parameter, can be obtained by similar methods. For example, one can use the MLE $\hat{\kappa}$ and its asymptotic variance obtained from the corresponding component of the information matrix, given in (4.2.7).

4.8 Appendix: Proofs

Proof of Theorem 4.1 :
Let $\theta_1, \theta_2, \ldots, \theta_p$ be independent with $\theta_i \sim CN(\mu_i, \kappa)$, κ known. Then the pdf of $\underset{\sim}{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)'$ is given by:

$$f(\underset{\sim}{\theta} \mid \underset{\sim}{\mu}, \kappa) = [2\pi I_0(\kappa)]^{-p} \exp\left\{\kappa \sum_{i=1}^{p} \cos(\theta_i - \mu_i)\right\}$$
$$0 < \theta_i, \mu_i < 2\pi, \qquad 1 \le i \le p. \tag{4.8.1}$$

Let $\underset{\sim}{\theta}^{(1)}, \underset{\sim}{\theta}^{(2)}, \ldots, \underset{\sim}{\theta}^{(n)}$ be a random sample from $f(\underset{\sim}{\theta} \mid \underset{\sim}{\mu}, \kappa)$. Then the joint density of $\underset{\sim}{\theta}^{(1)}, \underset{\sim}{\theta}^{(2)}, \ldots, \underset{\sim}{\theta}^{(n)}$ is given by

$$\begin{aligned} f(\underset{\sim}{\theta}^{(1)}, \underset{\sim}{\theta}^{(2)}, \ldots, \underset{\sim}{\theta}^{(n)} \mid \underset{\sim}{\mu}, \kappa) &= [2\pi I_0(\kappa)]^{-np} \exp\left\{\kappa \sum_{j=1}^{n} \sum_{i=1}^{p} \cos(\theta_i^{(j)} - \mu_i)\right\} \\ &= [2\pi I_0(\kappa)]^{-np} \exp\left\{n\kappa \sum_{i=1}^{p} \left(\cos\mu_i \bar{C}_i + \sin\mu_i \bar{S}_i\right)\right\}, \end{aligned} \tag{4.8.2}$$

where

$$\bar{C}_i = \frac{1}{n}\sum_{j=1}^{n} \cos\theta_i^{(j)}, \qquad \bar{S}_i = \frac{1}{n}\sum_{j=1}^{n} \sin\theta_i^{(j)}, \qquad 1 \le i \le p.$$

The MLE of $\underset{\sim}{\mu}$ is given by $\underset{\sim}{\widehat{\mu}}$, where

$$\hat{\mu}_i = \arctan^*\left(\frac{\bar{S}_i}{\bar{C}_i}\right), \qquad 1 \le i \le p. \tag{4.8.3}$$

We now have two cases corresponding to κ being known or unknown. In the first case recall that $CN(\mu, \kappa)$ becomes a member of the curved exponential family. One can appeal to Theorem 2.1 of Kariya (1989) to explore the best equivariance property of the MLE. When κ is unknown, we are in the framework of the regular exponential family and we approach the problem directly. To establish admissibility with κ known or unknown, our approach is to present a suitable prior for each case, so that the corresponding Bayes estimator coincides with the MLE given in Equation (4.8.3).

(i) To explore the best equivariance of the MLE given above, let us first consider the case when $p = 1$, i.e. a single CN population. Let $\theta_1, \theta_2, \ldots, \theta_n$ be a random sample from the $CN(\mu, \kappa)$. Then,

$$f(\theta_1, \theta_2, \ldots, \theta_n) = [2\pi I_0(\kappa)]^{-n} \exp\left\{n\kappa\bar{C}\cos\mu + n\kappa\bar{S}\sin\mu\right\},$$

where

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

Case 1: κ known.

Clearly, $(\bar{C}, \bar{S})$ is sufficient for μ. Let, $\eta_1 = \cos\mu$ and $\eta_2 = \sin\mu$. Then, by (3.3.5), the joint distribution of $(\bar{C}, \bar{S})$ is given by

$$g(\bar{C}, \bar{S}) = [2\pi I_0^n(\kappa)]^{-1} n^2 \exp\left\{n\kappa(\bar{C}\eta_1 + \bar{S}\eta_2)\right\} \psi_n\left(n^2(\bar{C}^2 + \bar{S}^2)\right). \quad (4.8.4)$$

This distribution belongs to the curved exponential family with $\Theta = \{\underset{\sim}{\theta} \in \tilde{\Theta} : \underset{\sim}{\theta} = \phi(\mu), \mu \in \Upsilon\}$ where $\tilde{\Theta} = \mathcal{R}^2$, $\Upsilon = [0, 2\pi]$ and $\phi : \Upsilon \longrightarrow \tilde{\Theta}$ defined by $\phi(\mu) = (\cos\mu, \sin\mu)$ is clearly a bimeasurable bijection onto $\phi(\gamma) = \Theta \subset \tilde{\Theta}$.

Consider the group $\mathcal{G}$ acting on $\mathcal{Z} = \{\underline{y} \in \mathcal{R}^2 : 0 \leq \| \underline{y} \| \leq 1\}$ given by

$$\begin{aligned}\mathcal{G} &= \left\{g_A : \underset{\sim}{X} \to A\underset{\sim}{X} \mid A : n \times n \text{ orthogonal with } \mid A \mid = 1\right\} \\ &= \left\{g_\tau : \underset{\sim}{X} \to A\underset{\sim}{X}, A = \begin{pmatrix} \cos\tau & -\sin\tau \\ \sin\tau & \cos\tau \end{pmatrix}, \quad 0 \leq \tau < 2\pi\right\}.\end{aligned}$$

Then, $\mathcal{G}$ is a topological group and the group action on $\mathcal{Z}$ is measurable (being continuous). Further, the joint distribution of $(y_1, y_2) \equiv g_\tau(\bar{C}, \bar{S})$ is given by

$$\begin{aligned}f(y_1, y_2) &= K(\kappa, n)\exp\left[n\kappa\left\{y_1\left(\eta_1\cos\tau - \eta_2\sin\tau\right)\right.\right. \\ &\quad \left.\left. + \, y_2\left(\eta_1\sin\tau + \eta_2\cos\tau\right)\right\}\right] \psi_n\left(n^2(y_1^2 + y_2^2)\right),\end{aligned}$$

$K(\cdot)$ being a constant. So, $g\mathcal{P}(\tilde{\Theta}) = \mathcal{P}(\tilde{\Theta})$ with $gP_\theta = P_\theta \cdot g^{-1}\ \forall g \in \mathcal{G}$ i.e. $\mathcal{P}(\tilde{\Theta})$ is invariant under $\mathcal{G}$. Further, $\bar{g}_\tau = g_\tau$ (i.e., $\bar{g}_A = g_A$) so that $\bar{\mathcal{G}} = \mathcal{G}$. Also, $\mathcal{G}$ acts homeomorphically on $\mathcal{Z}$ by

$$\begin{pmatrix} \bar{C} \\ \bar{S} \end{pmatrix} \to g_\tau \begin{pmatrix} \bar{C} \\ \bar{S} \end{pmatrix}.$$

Defining $\tilde{g}_\tau = \phi^{-1}\bar{g}_\tau\phi$, we have,

$$\begin{aligned}
\tilde{g}_\tau(\mu) &= \phi^{-1}\bar{g}_\tau\phi(\mu) \\
&= \phi^{-1}\bar{g}_\tau\begin{pmatrix}\cos\mu \\ \sin\mu\end{pmatrix} \\
&= \phi^{-1}\begin{bmatrix}\cos\mu\cos\tau - \sin\mu\sin\tau \\ \cos\mu\sin\tau + \sin\mu\cos\tau\end{bmatrix} \\
&= (\mu+\tau)(\text{mod } 2\pi).
\end{aligned}$$

Defining $\tilde{\mathcal{G}}$ acting on Υ by $\tilde{\mathcal{G}} = \{\tilde{g}_\tau(\theta) = (\theta+\tau)(\text{mod } 2\pi), 0 \le \tau < 2\pi\}$, we have: $\tilde{\mathcal{G}}$ is a homeomorphic image of $\bar{\mathcal{G}}$ (and hence of $\mathcal{G}$) and the subfamily $\mathcal{P}(\Theta) = \{P_{\phi(\eta)} \mid \eta \in \Upsilon\}$ is $\tilde{\mathcal{G}}$-invariant.

It is clear that the orbit of $\bar{\mathcal{G}}$ is Θ so that the action of $\bar{\mathcal{G}}$ on Θ is transitive. It follows easily then that the action of $\tilde{\mathcal{G}}$ on Υ is transitive.

Lemma 4.2 *$R = u(\bar{C},\bar{S}) = \sqrt{\bar{C}^2+\bar{S}^2}$ is maximal invariant statistic under $\mathcal{G}$.*

Proof: We first show that $u(\bar{C},\bar{S})$ is $\mathcal{G}$-invariant.

$$\begin{aligned}
u(g_\tau(\bar{C},\bar{S})) &= \left[(g_\tau(\bar{C},\bar{S}))'(g_\tau(\bar{C},\bar{S}))\right]^{1/2} \\
&= \left[(\bar{C},\bar{S})'\ A'A\ (\bar{C},\bar{S})\right]^{1/2} \\
&= \left[(\bar{C},\bar{S})'\ (\bar{C},\bar{S})\right]^{1/2} \\
&= u\left(\bar{C},\bar{S}\right).
\end{aligned}$$

Next, suppose that $u(\bar{C}_1,\bar{S}_1) = u(\bar{C}_2,\bar{S}_2)$. Then

$$R_1^2 \ = \ \bar{C}_1^2+\bar{S}_1^2 \ = \ \bar{C}_2^2+\bar{S}_2^2 \ = \ R_2^2.$$

If $R_1 = R_2 = 0$, then $\bar{C}_1 = \bar{C}_2 = \bar{S}_1 = \bar{S}_2 = 0$ so that $g_o(\bar{C}_1,\bar{S}_1) = (\bar{C}_2,\bar{S}_2)$. So, now suppose that $R_1 = R_2 > 0$. Let

$$P_i = \frac{1}{R_i}\begin{bmatrix}\bar{C}_i & \bar{S}_i \\ -\bar{S}_i & \bar{C}_i\end{bmatrix}, i = 1,2.$$

Then P_1, P_2 are orthogonal matrices with determinant 1 and

$$P_1\begin{pmatrix}\bar{C}_1 \\ \bar{S}_1\end{pmatrix} = \left(u(\bar{C}_1,\bar{S}_1)\right) = \left(u(\bar{C}_2,\bar{S}_2)\right) = P_2\begin{pmatrix}\bar{C}_2 \\ \bar{S}_2\end{pmatrix}.$$

So,

$$\begin{aligned}\begin{pmatrix}\bar{C}_1\\ \bar{S}_1\end{pmatrix} &= P_1'P_2\begin{pmatrix}\bar{C}_2\\ \bar{S}_2\end{pmatrix},\\ P_1'P_2 &= \frac{1}{R_1R_2}\begin{bmatrix}\bar{C}_1\bar{C}_2+\bar{S}_2\bar{S}_1 & \bar{C}_1\bar{S}_2-\bar{S}_1\bar{C}_2\\ \bar{S}_1\bar{C}_2-\bar{C}_1\bar{S}_2 & \bar{C}_1\bar{C}_2+\bar{S}_2\bar{S}_1\end{bmatrix}.\end{aligned}$$

Get τ such that

$$\begin{aligned}\cos\tau &= \frac{\bar{C}_1\bar{C}_2+\bar{S}_2\bar{S}_1}{R_1R_2},\\ \sin\tau &= \frac{-\bar{C}_1\bar{S}_2+\bar{S}_1\bar{C}_2}{R_1R_2}.\end{aligned}$$

Then

$$g_\tau\begin{pmatrix}\bar{C}_1\\ \bar{S}_1\end{pmatrix}=\begin{pmatrix}\bar{C}_2\\ \bar{S}_2\end{pmatrix},$$

so that $u(\bar{C},\bar{S})=R$ is a maximal invariant under $\mathcal{G}$. □

We now verify the conditions needed for Theorem 2.1 of Kariya (1989) to hold.

Lemma 4.3 *Assumptions 2.1 and 2.2 of Kariya (1989) hold in our above set-up with the $CN(\mu,\kappa),\kappa$ known, model.*

Proof: Note that $\lambda(\eta)=\|\eta\|$ is a maximal invariant parameter under $\mathcal{G}$. So Θ in Equation (4.8.4) may be expressed as $\Theta=\{\theta\in\tilde{\Theta}\mid\lambda(\theta)=1\}$. Further, the map $g_\tau\to\bar{g}_\tau\equiv g_\tau$ is measurable. So, Assumption 2.1 of Kariya (1989) is satisfied.

We next verify Assumption 2.2 of Kariya (1989). The MLE of μ is given by,

$$\hat{\mu}(\bar{C},\bar{S})=\arctan^*\left(\frac{\bar{S}}{\bar{C}}\right),\tag{4.8.5}$$

where arctan* is defined in (1.3.5). It may be noted that excluding the set $\{\{0\}\times[-1,1]\}\bigcup\{[-1,1]\times\{0\}\}$ of measure zero on $\mathcal{Z}$, $\widehat{\mu}(\bar{C},\bar{S})$ defines a bijection from $\mathcal{Z}$ onto $[0,2\pi)$. Define,

$$h(\bar{C},\bar{S})=\frac{1}{\|(\bar{C},\bar{S})\|}\begin{bmatrix}\bar{C} & -\bar{S}\\ \bar{S} & \bar{C}\end{bmatrix}.$$

Then,

$$\begin{aligned} h(g_\tau(\bar{C},\bar{S})) &= \frac{1}{\|(\bar{C},\bar{S})\|}\begin{bmatrix} \bar{C}\cos\tau - \bar{S}\sin\tau & -\bar{C}\cos\tau - \bar{C}\sin\tau \\ \bar{C}\sin\tau + \bar{S}\cos\tau & \bar{C}\cos\tau - \bar{S}\sin\tau \end{bmatrix} \\ &= g_\tau(h(\bar{C},\bar{S})). \end{aligned}$$

Define, $\pi(\bar{C},\bar{S}) = (h(\bar{C},\bar{S}), u(\bar{C},\bar{S}))$. Then π is a continuous map defined from $\mathcal{Z}$ onto $\mathcal{G}\times\mathcal{U}$ where $\mathcal{U} = [0,1]$ is a measurable space.

Let $\pi(\bar{C}_1,\bar{S}_1) = \pi(\bar{C}_2,\bar{S}_2)$. Then,

$$\begin{aligned} \frac{1}{\|(\bar{C}_1,\bar{S}_1)\|}\begin{bmatrix} \bar{C}_1 & -\bar{S}_1 \\ \bar{S}_1 & \bar{C}_1 \end{bmatrix} &= \frac{1}{\|(\bar{C}_2,\bar{S}_2)\|}\begin{bmatrix} \bar{C}_2 & -\bar{S}_2 \\ \bar{S}_2 & \bar{C}_2 \end{bmatrix}, \\ \|(\bar{C}_1,\bar{S}_1)\| &= \|(\bar{C}_2,\bar{S}_2)\|, \end{aligned}$$

and it can be shown that π is injective, surjective and continuous. Further, π^{-1} is well-defined and also continuous.

So, there exists a bijective, bimeasurable map from $\mathcal{Z}$ onto $\mathcal{G}\times\mathcal{U}$ such that if $\pi(z) = (h(z), u(z))$, then $\pi(gz) = (gh(z), u(z))$, where $z = (\bar{C},\bar{S}) \in \mathcal{Z}$ and $\mathcal{U}$ is a measurable space.

Hence, Assumption (2.2) of Kariya (1989) is satisfied. □

Now let the loss function be given by,

$$L(\mu, a) = 1 - \cos(a-\mu), \qquad 0 \le a, \mu < 2\pi.$$

Then by virtue of Lemma 4.3 above and invoking Theorem 2.1 of Kariya (1989) it follows that a best equivariant estimator, when it exists, is

$$\widehat{\mu}(h(\bar{C},\bar{S}), u(\bar{C},\bar{S})) = \tilde{h}(\bar{C},\bar{S})\widehat{\mu}_1(e, u(\bar{C},\bar{S})) = \widehat{\mu}(\bar{C},\bar{S}) + \mu_1^*(u(\bar{C},\bar{S})),$$

where $\widehat{\mu}(\bar{C},\bar{S})$ is as defined in Equation (4.8.5) and μ_1^* minimizes the conditional expectation,

$$E_\mu[1-\cos(\widehat{\mu} + \mu_1(u(\bar{C},\bar{S})) - \mu) \mid u(\bar{C},\bar{S})] = E_0[1-\cos(\widehat{\mu}+\mu_1) \mid u(\bar{C},\bar{S})] \tag{4.8.6}$$

by the transitivity of $\mathcal{G}$. To minimize the above w.r.t. μ_1 observe that μ_1^* satisfies the equation,

$$E_0[\sin(\widehat{\mu}+\mu_1^*) \mid u(\bar{C},\bar{S})] = 0.$$

Again using Equation (3.3.10),

$$(\sin \mu_1^*) E_0[\cos \widehat{\mu} \mid u(\bar{C}, \bar{S})] = (\sin \mu_1^*) A(\kappa R) = 0,$$

yielding $\mu_1^* = 0$ or π.

Now, $E_0[\cos(\widehat{\mu} + \pi) \mid u(\bar{C}, \bar{S})] < 0$, while $E_0[\cos \widehat{\mu} \mid u(\bar{C}, \bar{S})] > 0$, so that Equation (4.8.6) is minimized for $\mu_1^* = 0$. Consequently, the MLE $\widehat{\mu}$ in Equation (4.8.5) is the best equivariant estimator.

Further, $\mathcal{G}$ being compact, the MLE $\widehat{\mu}$ is minimax in the class $\mathcal{D}$ of all estimators (see Ferguson (1967)). It is also admissible in the class $\mathcal{D}$ of all estimators for μ.

Case 2: κ unknown.

We already note that the distribution of $(\bar{C}, \bar{S})$ given in (4.8.4) belongs to the regular exponential family (REF) with $\Theta = \{(\mu, \kappa) : \mu \in [0, 2\pi), \kappa \geq 0\}$. Then, as before, consider the group $\mathcal{G}$ acting on $\mathcal{Z} = \{\underset{\sim}{y} \in \mathcal{R}^2, 0 \leq \| \underset{\sim}{y} \| \leq 1\}$ given by

$$\begin{aligned}
\mathcal{G} &= \left\{ g_A : \underset{\sim}{X} \to A \underset{\sim}{X} \mid A : n \times n \text{orthogonal with} \mid A \mid = 1 \right\} \\
&= \left\{ g_\tau : \underset{\sim}{X} \to A \underset{\sim}{X}, A = \begin{pmatrix} \cos\tau & -\sin\tau \\ \sin\tau & \cos\tau \end{pmatrix}, \quad 0 \leq \tau < 2\pi \right\}.
\end{aligned}$$

Proceeding as before, we have that the joint distribution of $(y_1, y_2) \equiv g_\tau(\bar{C}, \bar{S})$ is given by,

$$\begin{aligned}
f(y_1, y_2) &= \text{const.}(\kappa, n) \exp\left[n\kappa \left\{y_1 \left(\eta_1 \cos\tau - \eta_2 \sin\tau\right)\right.\right. \\
&\quad \left.\left. + y_2 \left(\eta_1 \sin\tau + \eta_2 \cos\tau\right)\right\}\right] \psi_n \left(n^2(y_1^2 + y_2^2)\right).
\end{aligned}$$

So, $g\mathcal{P}(\Theta) = \mathcal{P}(\Theta)$ with $gP_\theta = P_\theta . g^{-1} \forall g \in \mathcal{G}$ i.e. $\mathcal{P}(\Theta)$ is invariant under $\mathcal{G}$. Also, the induced group action on the parameter space Θ is given by $g_\tau(\mu, \kappa) = ((\mu + \tau)(\text{mod } 2\pi), \kappa)$. This shows that the induced group of transformations $\overline{\mathcal{G}}$ acting on the parameter space is not transitive. The same arguments as before provide us with $u(\overline{C}, \overline{S}) = R$ as the maximal invariant statistic. Also, it can be shown, following the arguments similar to the ones used for the case κ known, that the MLE of μ, there also given by $\widehat{\mu}(\overline{C}, \overline{S})$ (in Equation (4.8.5)), is equivariant under the group $\mathcal{G}$. To find the best equivariant estimator under the natural loss, we are to find μ_1, as a measurable function of the maximal invariant, $u(\overline{C}, \overline{S})$ such that the risk,

$$E_{\mu,\kappa}[1 - \cos(\widehat{\mu} + \mu_1(u(\overline{C}, \overline{S})) - \mu) \mid u(\overline{C}, \overline{S})]$$

is minimized uniformly for all $(\mu, \kappa) \in \Theta$. To minimize the above risk w.r.t. μ_1, observe that μ_1^* satisfies the equation

$$E_{\mu,\kappa}\left[\sin\left(\widehat{\mu} + \mu_1^* - \mu\right) \mid u\left(\overline{C}, \overline{S}\right)\right] = 0.$$

Thus, $\sin \mu_1^* E_{\mu,\kappa}[\sin(\widehat{\mu} - \mu) \mid u(\overline{C}, \overline{S})] = 0$, since from (3.3.10),

$$\widehat{\mu} \mid R \sim CN\left(\mu, \kappa R\right),$$

and for $\theta \sim CN(\mu, \kappa)$, $E(\sin(\theta - \mu)) = 0$. Hence, $(\sin \mu_1^*)A(\kappa R) = 0$, yielding $\mu_1^* = 0$ or π. Now, $E_{\mu,\kappa}[\cos(\widehat{\mu} - \mu + \pi) \mid u(\overline{C}, \overline{S})] < 0$, while $E_{\mu,\kappa}[cos(\widehat{\mu} - \mu) \mid u(\overline{C}, \overline{S})] > 0$, so that the risk is uniformly minimized for $\mu_1^* = 0$. This leads us to conclude that the MLE is the Best Equivariant Estimator of the mean direction, under the given natural loss, even when the concentration parameter is unknown.

To summarize, when $p = 1$, we have shown that MLE $\widehat{\mu}$ in Equation (4.8.5) for μ in $CN(\mu, \kappa)$ population is the Best Equivariant, Admissible and Minimax estimator for μ whether κ is known or unknown. $\square$

Let us now consider the simultaneous estimation problem of estimating $\underset{\sim}{\mu} = (\mu_1, \mu_2, \ldots, \mu_p)'$. Then, consider the group $\mathcal{G}$ acting on

$$\mathcal{Z} = \prod_{i=1}^{p} \left\{ \underset{\sim}{y_i} : \parallel \underset{\sim}{y} - \underset{\sim}{y_i} \parallel \leq 1 \right\}$$

given by

$$\begin{aligned}
\mathcal{G} &= \left\{ g_{A_i} : \underset{\sim}{X_i} \to A_i \underset{\sim}{X_i} \mid A_i : 2 \times 2 \text{ orthogonal};\, i = 1, 2, \ldots, p \right\} \\
&= \left\{ g_{\tau} : \underset{\sim}{X_i} \to A_i \underset{\sim}{X_i} | A_i = \begin{pmatrix} \cos\tau_i & -\sin\tau_i \\ \sin\tau_i & \cos\tau_i \end{pmatrix},\ \tau_i \in [0, 2\pi); 1 \leq i < p \right\}.
\end{aligned}$$

Following the same arguments as for the case $p = 1$ above, the induced action on the parameter space is given by $\bar{g}\, \underset{\sim}{\mu} = \{(\mu_i + \tau_i)(\text{mod } 2\pi),\ i = 1, 2, \ldots, p\}$. Further, the action is transitive with the parameter space as the orbit. Recall that the MLE for $\underset{\sim}{\mu}$ is given by $\underset{\sim}{\widehat{\mu}}$ defined in Equation (4.8.3) and it is easy to show that it is an equivariant estimator. Imitating the steps for the case $p = 1$, we get that the equivariant estimators for $\underset{\sim}{\mu} = (\mu_1, \mu_2, \ldots, \mu_p)'$ are of the form

$$\delta(\underset{\sim}{\theta}) = ((\theta_1 + c_1)(\text{mod } 2\pi), (\theta_2 + c_2)(\text{mod } 2\pi), \ldots, (\theta_p + c_p)(\text{mod } 2\pi))'.$$

To find the best equivariant estimator under the loss function,

$$L(\underset{\sim}{a}, \underset{\sim}{\mu}) = p - \sum_{i=1}^{p} \cos(a_i - \mu_i)$$

we look for the equivariant estimator having minimum risk. To minimize the risk, we are to find the values of $c_1, c_2, \ldots, c_p$ such that the risk,

$$\left[E[p - \sum_{i=1}^{p} \cos\left((\widehat{\mu}_i + c_i)(\text{mod } 2\pi) - \mu_i\right)]\right] = E\left[p - \sum_{i=1}^{p} \cos\left(\widehat{\mu}_i + c_i - \mu_i\right)\right]$$

is minimized. This gives us, following steps similar to those used above for $p = 1$, that the minimum risk is achieved when $c_i = 0$; $i = 1, 2, \ldots, p$.

(ii) Next we consider admissibility of the MLE, with again two cases.

Case 1: κ known.

Let $\underset{\sim}{\mu} = (\mu_1, \ldots, \mu_p)'$ have the prior density,

$$\pi(\underset{\sim}{\mu}) = [2\pi I_0(\kappa_0)]^{-p} \exp\left\{\kappa_0 \sum_{i=1}^{p} (\cos\mu_i \cos\xi + \sin\mu_i \sin\xi)\right\}.$$

Then using the joint pdf given in Equation (4.8.2), the posterior density of $\underset{\sim}{\mu}$ is given by,

$$\begin{aligned} &\pi\left(\underset{\sim}{\mu} \mid \underset{\sim}{\theta}^{(1)}, \underset{\sim}{\theta}^{(2)}, \ldots, \underset{\sim}{\theta}^{(n)}; \xi, \kappa_0\right) \\ \propto\ &\exp\left\{\sum_{i=1}^{p} \cos\mu_i(\kappa_0\cos\xi + n\kappa\bar{C}_i) + \sin\mu_i(\kappa_0 \sin\xi + n\kappa\bar{S}_i))\right\}. \end{aligned}$$

Let $\kappa_i^* = [\kappa_0^2 + \kappa^2 R_i^2 + 2\kappa\kappa_0 n(\cos\xi\bar{C}_i + \sin\xi\bar{S}_i)]^{1/2}$, $1 \le i \le p$. Let $\underset{\sim}{\eta} = (\eta_1, \eta_2, \ldots, \eta_p)'$ be the solution to:

$$\begin{aligned} \left(n\kappa\bar{C}_i + \kappa_0\cos\xi\right)/\kappa_i^* &= \cos\eta_i, \\ \left(n\kappa\bar{S}_i + \kappa_0\sin\xi\right)/\kappa_i^* &= \sin\eta_i, \qquad 0 \le \eta_i < 2\pi, 1 \le i \le p. \end{aligned}$$

Then the posterior density of $\underset{\sim}{\mu}$ can be written as

$$\pi\left(\underset{\sim}{\mu} \mid \underset{\sim}{\theta}^{(1)}, \underset{\sim}{\theta}^{(2)}, \ldots, \underset{\sim}{\theta}^{(n)}\right) = \prod_{i=1}^{p} \left[2\pi I_0(\kappa_i^*)\right]^{-1} \exp\left(\kappa_i^* \cos(\mu_i - \eta_i)\right).$$

Consider the estimation of $\underset{\sim}{\mu}$ under the loss function,

$$L(\underset{\sim}{\mu}, \underset{\sim}{a}) = p - \sum_{i=1}^{p} \cos\left(a_i - \mu_i\right).$$

For any estimator $\underset{\sim}{\delta}$, the Bayes risk is given by,

$$r\left(\pi, \underset{\sim}{\delta}\right) = EE\left[p - \sum_{i=1}^{p} \cos\left(\delta_i - \mu_i\right) \mid \underset{\sim}{\theta}\right].$$

To minimize the Bayes risk, we minimize,

$$E\left[p - \sum_{i=1}^{p} \cos\left(\delta_i(\underset{\sim}{\theta}) - \mu_i\right) \mid \underset{\sim}{\theta}\right]$$

and this is done when for each i, $1 \leq i \leq p$, $E[\sin(\delta_i(\underset{\sim}{\theta}) - \mu_i) \mid \underset{\sim}{\theta}] = 0$, or $E[\sin(\eta_i - \mu_i)\cos(\delta_i(\underset{\sim}{\theta}) - \mu_i) \mid \underset{\sim}{\theta}] + E[\cos(\eta_i - \mu_i)\sin(\delta_i(\underset{\sim}{\theta}) - \eta_i) \mid \underset{\sim}{\theta}] = 0$, or, $\delta_i(\underset{\sim}{\theta}) = \eta_i$ or $\eta_i + \pi$. If $\underset{\sim}{\delta}(\underset{\sim}{\theta}) = \underset{\sim}{\eta} + \pi \underset{\sim}{1}$, then $r(\pi, \delta) > p$, so that the Bayes estimator is given by, $\underset{\sim}{\delta}(\underset{\sim}{\theta}) = \underset{\sim}{\eta}$.

Specializing to the case when $\kappa_0 = 0$, we obtain that the Bayes estimator w.r.t. the prior of p independent identical circular uniform populations, is given by $\delta(\underset{\sim}{\theta}) = \underset{\sim}{\eta}$, $\underset{\sim}{\eta}$ being given by

$$\eta_i = \arctan^*\left(\frac{\bar{S}_i}{\bar{C}_i}\right), \quad 1 \leq i \leq p.$$

Thus the MLE is admissible, being the unique Bayes MLE w.r.t. a proper prior.

Case 2: κ unknown.

Observe that even when κ is unknown, the MLE $\underset{\sim}{\widehat{\mu}}$ is given by (4.8.3) as before. Let $\pi(\underset{\sim}{\mu}, \kappa)$ be the prior given by,

$$\begin{aligned}\pi(\underset{\sim}{\mu}, \kappa) \quad \propto \quad & \prod_{i=1}^{p} I_0^{-1}\left\{n\kappa_0\sqrt{\bar{C}_i^2 + \bar{S}_i^2}\right\} I\left(\kappa = \kappa_0\right) \\ & \times \exp\left\{n\kappa_0\left(\bar{C}_i \cos\mu_i + \bar{S}_i \sin\mu_i\right)\right\}, \\ & 0 < \mu_i < 2\pi, \quad 1 \leq i \leq p. \end{aligned} \tag{4.8.7}$$

Let the loss function be as in Case 1 above. Then, to minimize the posterior Bayes risk it is enough to minimize

$$\begin{aligned} & E\left[p - \sum_{i=1}^{p} \cos\left(\delta_i - \mu_i\right) \mid \underset{\sim}{\theta}^{(1)}, \underset{\sim}{\theta}^{(2)}, \ldots, \underset{\sim}{\theta}^{(n)}\right] \\ = \quad & p - \sum_{i=1}^{p} EE^{\kappa}\left[\cos\left(\delta_i - \mu_i\right) \mid \underset{\sim}{\theta}^{(1)}, \ldots, \underset{\sim}{\theta}^{(n)}, \kappa\right] \\ = \quad & p - \sum_{i=1}^{p} E\left[\cos\left(\delta_i - \mu_i\right) \mid \underset{\sim}{\theta}^{(1)}, \ldots, \underset{\sim}{\theta}^{(n)}, \kappa_0\right]. \end{aligned}$$

So this reduces to the case of minimizing the loss for κ known and equal to κ_0. Thus the Bayes estimator for $\underset{\sim}{\mu}$ with respect to the prior (4.8.7) is given by,

$$\underset{\sim}{\delta}\left(\underset{\sim}{\theta}^{(1)}, \underset{\sim}{\theta}^{(2)}, \ldots, \underset{\sim}{\theta}^{(n)}\right) = \underset{\sim}{\widehat{\mu}}, \text{ the MLE.}$$

□

Chapter 5

Tests for Mean Direction and Concentration

5.1 Introduction

In this chapter, the focus is on testing problems for the CN distribution and their optimal properties. Not much is known in terms of optimal tests for the parameters of other circular distributions. See Chang (1991), Coles (1998), Gatto and Jammalamadaka (2000) and SenGupta and Pal (2001a) for a discussion on WS distributions. Tests for the case of a single sample are first studied, starting with the mean direction μ of the CND, against one- and two-sided alternatives - both exact and asymptotic. The curved exponential family (CEF) nature of the CN distribution when κ is known, provides a framework to develop optimal conditional tests. When κ is unknown, the CN distribution becomes a member of the regular exponential family. See Appendix at the end of Chapter 2 for a discussion on curved and regular exponential families.

However, since κ is not a scale parameter, unconditional similar tests or invariant tests are not available and one is constrained to look at conditional tests. We review these conditional tests first. Using the notion of statistical curvature, we consider unconditional optimal tests for these problems, starting first with Locally Most Powerful (LMP) tests. In the case of a known κ, exact LMP tests are developed and some cut-off points tabulated. For the case when κ is unknown, we consider first a conditional test and then derive a C_α - test or the asymptotic LMP test of Neyman (1959).

Next, tests for $H_0 : \kappa = \kappa_0$ are studied. Again κ being not a scale parameter, critical values for such a test will depend on the hypothesized value κ_0. Of particular note is the case of testing $\kappa = 0$ which as we said before, is of prime importance in circular statistical inference. The REF nature of the CN distribution, readily yields the UMP test in case μ is known. Further, since μ may be treated as a location parameter, it is possible to derive the (unconditional) Uniformly Most Powerful Unbiased Invariant (UMPUI) test. For the case of two samples, exact conditional tests are considered, while for more than two populations approximate ANOVA tests are discussed. See also Section 12.4.

5.2 Single Population

5.2.1 Tests for Mean Direction

For testing

$$H_0 : \mu = \mu_0$$

versus the one-sided alternative $H_1 : \mu > \mu_0$, or the two-sided alternative $H_2 : \mu \neq \mu_0$, the LRT is based on the statistic

$$V_0 = \sum_{i=1}^{n} \cos(\alpha_i - \mu_0) = R\cos(\overline{\alpha_0} - \mu_0).$$

As noted before, (see Figure 1.9), V_0 represents the length of the projection of the resultant vector towards the hypothesized mean direction μ_0 and is always smaller than the resultant length R. Indeed the ratio

$$\frac{V_0}{R} = \cos(\overline{\alpha_0} - \mu_0)$$

represents how close the sample mean direction is to the hypothesized mean direction. Small values of V_0 or the ratio V_0/R lead to rejection of the null. In the general case, the distribution of V_0 as well as that of the ratio V_0/R depends on the unknown nuisance parameter κ. However an exact *conditional* test for the mean direction of the CN distribution can be obtained by using the conditional distribution of R given V (see Equation 3.3.9), which is independent of κ. Such a test rejects H_0 if V_0 is too small for a given R or equivalently, if R is too large for a given V_0. Tables of such critical values of

R for given V_0 can be constructed using the distribution in 3.3.9 i.e., find a value r_0 such that

$$P(R \geq r_0|v) = \int_{r_0}^{n} f(r|v_0)dr = \alpha,$$

where

$$f_\kappa(r|v) = \frac{r\psi_n(r)}{\sqrt{r^2 - (v)^2 f_0(v)\pi}}. \tag{5.2.1}$$

Charts which incorporate such critical values of R for given n, v_0 and chosen significance level, may be found in Stephens (1962) and are reproduced in Table 57, Volume II of Biometrika Tables.

We now consider various alternative approaches.

Case 1. κ known.

Without loss of generality, one may take the given hypothesized value, $\mu_0 = 0$ and consider testing $H_0 : \mu = 0$ say against the one-sided alternative $H_1 : \mu > 0$. Clearly when $\mu_0 = 0$, $V_0 = C$, the sum of the cosine components. If κ is known then one can obtain the distribution and critical values of V_0 from (3.3.12) or of the ratio V_0/R from (3.3.4). However, note that since κ is not a scale parameter, tables of such critical values need to be computed for different values of the parameter κ to make the tests useful in practice.

A. A Conditional Test.

Consider testing H_0 against a simple alternative $H_1 : \mu = \mu' > 0$. Let R and $\overline{\alpha}_0$ denote the length and the direction of the vector resultant i.e., $V = R\cos\overline{\alpha}_0$, $S = R\sin\overline{\alpha}_0$. Then given $R = r$ the most powerful test for $H_0 : \mu = 0$ against $H_1 : \mu = \mu'$ is to reject when

$$\sin(\overline{\alpha}_0 - \mu'/2) > k,$$

where $\overline{\alpha}_0 \sim CN(\mu, r)$ and k is a constant chosen to satisfy the level condition. See Mardia (1972) pp. 138-139, who suggests this based on Fisher's principle of ancillarity. To use such a test, one needs to get critical values not only for the different values of κ, but also for various values of the conditioning variable R. Also the power performance of such a conditional test is rather unimpressive.

B. Locally Most Powerful Test.

In practice one might be interested in testing against alternatives close to the null, say $0 < \beta < \pi/2$. Here we consider such one-sided composite alternatives. Since the given distribution is a member of the (2, 1) CEF, a UMP test does not exist and a LMP test is then a reasonable candidate. However, as discussed earlier it would be prudent to look at the statistical curvature associated with this testing problem. This is done in the next subsection. The LMP test, by definition, rejects when

$$\sum_i \partial \log f(\alpha_i)/\partial\beta|_{\beta=0} > k, \qquad \text{or} \qquad S = R\sin\overline{\alpha}_0 > k, \tag{5.2.2}$$

where k is a constant chosen to satisfy the level condition. The exact cut-off points are given in Table 5.1 below.

Table 5.1: Cut-off points, k of LMP test ($\alpha = .05$).

n	5	6	7	8	9	10	11	12	13	14	15
k	2.46	2.70	2.92	3.12	3.30	3.48	3.65	3.81	3.97	4.12	4.26

Proposition 5.1 *The LMP test is an unconditional test, is unbiased, has monotone power function, is consistent and is admissible - all globally for* $0 < \mu < \pi/2$.

Proof: The first property is obvious from Equation (5.2.2). Since μ is a location parameter, monotonicity and hence unbiasedness follow e.g., by stochastic ordering. Consistency is easy to establish and admissibility is a consequence of the uniqueness of the non-randomized LMP critical region. □

C. Likelihood Ratio Test.

Since κ is known and taken to be 1, the density of R for this special case is given by (see (3.3.7))

$$f_{\kappa=1}(r) = \frac{I_0(r)}{I_0^n(1)} r\psi_n(r), \tag{5.2.3}$$

where $\psi_n(r)$ is as in (3.2.5). The LRT reduces to,

$$R - C > k \text{ when } \hat{\mu} > \mu_0 = 0$$

and its "size" is given by,

$$\int_0^n t_0(r) f_{\kappa=1}(r) dr,$$

where

$$t_0(r) = (1/2\pi) \int_a^b \exp(r \cos\theta) d\theta,$$

and

$$a = 2\pi - \arccos(k/r + 1) \text{ and } b = \arccos(k/r + 1) \pmod{2\pi}.$$

Power at μ is obtained by replacing a and b above by,

$$a' = 2\pi - \cos^{-1}(k/r + 1) - \mu \text{ and } b' = \cos^{-1}(k/r + 1) - \mu \pmod{2\pi},$$

respectively. Comparison with LMP shows that LMP is superior in the more reasonable range of 0 to $\pi/2$, almost up to $\pi/2$. Further, with $t = 1.2$ Amari (1985) shows that (see p. 182) for $\mu \cong t\sqrt{ng}$, where $g = A(1)$, in actual computation, it requires quite a large n (more than 30) for the approximation to work, and for the LRT to be superior to the LMP. With $t > 2$ however, the LMP test still dominates in a wide range. For example, $t = 1.2$, $n = 7$ give $\beta = .6788$ while $t = 1.2$ gives $\beta = 1.1314$ while for $n = 30$, the values are $\beta = .3279$ and $\beta = .5465$. Table 5.2 exhibits such power comparisons.

Geometry of the Test.

It is interesting and instructive to look at the geometry of the LMP test. Recall from Section 2.5 that the family M forms a circle centered at $(0, 0)$, with the ancillary family $A(u)$ is given by $\lambda(\eta) = c$, i.e. $\eta_2 = c$. The straight line $A(u)$ intersects M at two points u_- and u_+ for $c < 1$ and at one point, $u^* = 1$ for $c = 1$. See Figure 5.1, a parametric representation independent of c seems infeasible. However, the curvature of $CN(\beta,\ 1)$ needs to be obtained. Now, at $\beta = 0$,

$$\begin{aligned} Var(\sin\alpha) &= A(\kappa)/\kappa, \\ Var(\cos\alpha) &= A'(\kappa), \\ \text{and} \quad Cov(\sin\alpha,\ \cos\alpha) &= 0. \end{aligned}$$

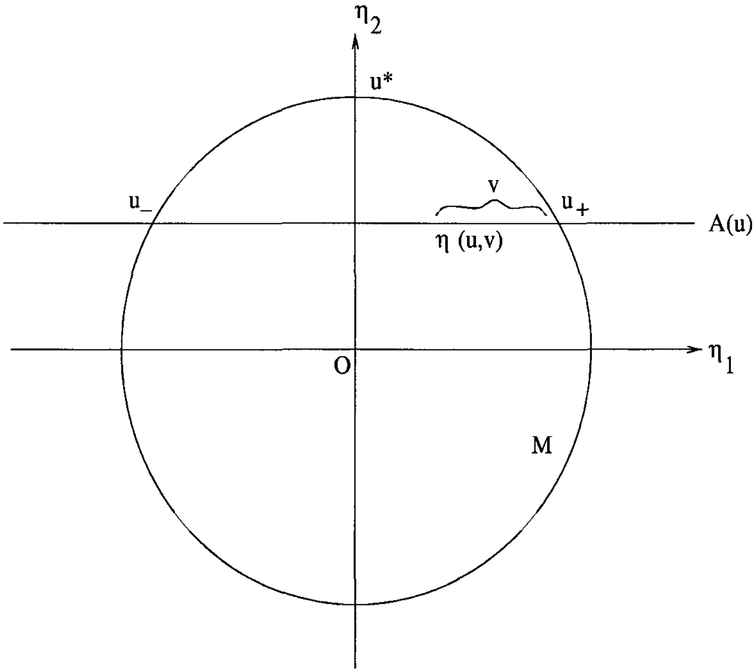


Figure 5.1: Ancillary family and (u, v) - coordinates of the locally most powerful test.

Then, at $\beta = 0$, for any κ, $\gamma_0^2(\kappa) = A(\kappa)/\kappa(A'(\kappa))^2$. So, $\gamma^2(\kappa)$ is $\uparrow$ ($\downarrow$) in κ if $\kappa(A')^3 - A(A')^2 - 2\kappa AA'A'' > (<)0$, where $A \equiv A(\kappa)$. In particular, when $\kappa = 1$, we need $n > 14.266$, i.e. $n = 15$ for $n\gamma_0^2 < 1/8$. Thus, according to Efron's rule, a sample size of 15 will suffice to reduce the curvature below the critical value and to expect the LMP test to work "well" – an easy requirement in practice. Next we obtain the exact power of the unconditional LMP test and compare its performance against the corresponding conditional test discussed earlier. An overall comparison may be obtained by using average power, averaged over the conditioning variable or looking at the power envelope. A general comparison may also be made by evaluating the deviations of the unconditional power from the maximum and minimum powers corresponding to the values of the conditioning variable. From this comparison in Table 5.2, the performance of the LMP test looks quite encouraging. To obtain the cut-off points, note that,

$$\begin{aligned} P(R\sin\overline{\alpha}_0 \geq k \mid H_0) &= \int_0^n P(R\sin\overline{\alpha}_0 \geq k \mid R = r, H_0)p_R(r)dr \\ &= \int_0^n P(\overline{\alpha}_0 \in \omega_r \mid H_0)f_{\kappa=1}(r)dr, \end{aligned} \tag{5.2.4}$$

where $\omega_r = \text{arc}[\{(\pi/2 - \delta_r), (\pi/2 + \delta_r)\}(\text{mod } 2\pi)]$, $\delta_r = \cos^{-1}(k/r)$. But, $\overline{\alpha}_0 \mid (R = r) \sim CN(0, r)$ under H_0 and $f_{\kappa=1}(r)$ is given in (5.2.3). For $\mu = \mu' > 0$, since $(\theta - \mu') \mid (R = r) \sim CN(0, r)$,

$$\text{Power } (\mu') = P(R\sin\overline{\alpha}_0 > k \mid \mu') = \int_0^n P(\overline{\alpha}_0 \in \omega_r' \mid \mu = 0)p_R(r)dr,$$

where $\omega_r' = \text{arc}[\{(\pi/2 - \delta_r - \mu'), (\pi/2 + \delta_r - \mu)\}(\text{mod } 2\pi)]$. Given k, power can be obtained through numerical integration. Some cut-off points are given in Table 5.1 while certain power values are tabulated in Table 5.2. Table 5.1 gives the powers for the best test for H_0 against a single value for the alternative. Comparisons of the performances of the LMP test and this restrictive test can now be made as was indicated earlier. For example, for the best test, at $H_1 : \mu = 10$, maximum power $= .1130$ (for $r = 7$) and minimum power $= .0594$ (for $r = 1$), while the LMP test has power .09.

Next we consider tests for the two-sided alternatives, i.e., test

$$H_0 : \mu = 0 \quad \text{against} \quad H_1 : \mu \neq 0, \tag{5.2.5}$$

κ is specified. We consider the following tests : a test based on the MLE, the LRT and the LMPU tests. The LMPU merits special mention here, since not only does it possess an exact optimality property, but the test statistic is also quite elegant. This results from the symmetry of the CN distribution. We study the geometry of the first two tests, noting that no small-sample optimality for them is known. However based on this study, together with the two-sided LMPU test, we present in Section 5.2.2 a large-sample higher-order power comparison of these three tests.

A. Test Based on the MLE.

As we have seen before, the maximum likelihood estimator of μ is given by, $\hat{\mu} = \arctan^*(S/C)$. For testing $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$, consider the

Table 5.2: Comparison of powers, ζ, of MP and LMP tests.

β'	MP test at $H_1 : \beta = \beta'$		LMP test
	$\min \zeta$ r	$\max \zeta$ r	
2	.0518	.0595	.0566
4	.0536	.0704	.0639
6	.0555	.0829	.0718
8	.0574	.0971	.0805
10	.0594	.1130	.0900
12	.0615	.1308	.1002
14	.0637	.1505	.1112
16	.0652	.1722	.1223
18	.0682	.1959	.1354
20	.0705	.2216	.1487

test which rejects when $\hat{\mu} > k_1$ or $< k_2$. The geometry of this test may be described as follows: In terms of the coordinates $\eta(u) = (\cos u, \sin u)$, M forms a unit circle, $\eta_1^2 + \eta_2^2 = 1$, centered at $(0,0)$ in S. From the definition of μ, it then follows that the ancillary family associated with the test T is given by the family of straight lines which pass through the center (actually origin) of the above circle. For a given significance level, the critical region is bounded by the pair of these lines. We introduce in each ancillary subspace (line in this case) $A(u)$, a local coordinate system v, which is defined as the distance from the intersecting point of M and $A(u)$. Then a point $\eta = (\eta_1, \eta_2)$ in a neighborhood of M can be expressed in terms of the local coordinate system (u, v) as $\eta_1 = (1 - v)\cos u$, $\eta_2 = (1 - v)\sin u$.

B. Likelihood Ratio Test.

The likelihood ratio test statistic, $\lambda(\alpha_1, \ldots, \alpha_n)$ is given by,

$$\lambda(\alpha_1, \ldots, \alpha_n) = \prod_{i=1}^{n} f(\alpha_i, \hat{\mu}) = \exp\left[\sum_{i=1}^{n} \{\cos\alpha_i - \cos(\alpha_i - \hat{\mu})\}\right]$$

or

$$-2\ln\lambda = 2\left[(\cos\hat{\mu} - 1)C + (\sin\hat{\mu})S\right] \equiv U(C, S),$$

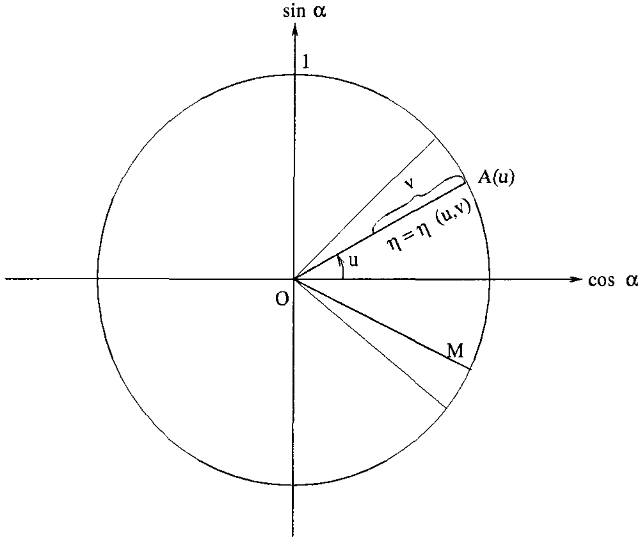


Figure 5.2: Ancillary family and (u, v) - coordinates of the test based on the maximum likelihood estimator.

say. The geometry of this test may be discussed as follows: The ancillary family is given by, $\{U(\eta_1, \eta_2) = k\}$ and therefore,

$$U(\eta_1, \eta_2) = k \Rightarrow \left(\frac{\eta_1}{\sqrt{\eta_1^2 + \eta_2^2}} - 1 \right) \eta_1 + \left(\frac{\eta_1}{\sqrt{\eta_1^2 + \eta_2^2}} - 1 \right) \eta_1 = k/2,$$

i.e.,

$$\eta_2^2 = k(\eta_1 + k/4). \tag{5.2.6}$$

Thus, the ancillary subspaces are defined by the parabolas as given in (5.2.6). Representing these parabolas in a parametric form, let us consider a parabola $A(u)[\equiv \eta_2^2 = k(\eta_1 + k/4)]$ which intersects the unit circle M (our family of circular normal distribution) at the point Q. So, coordinate of Q is $(\cos u, \sin u)$. Let P be any point on $A(u)$. Let the coordinate of P be (η_1, η_2). Let $PN \perp OX$ and $OP = x$. Let ρ correspond to the value of MLE for $\eta = (C, S)$, i.e. $\angle PON = \rho$. The coordinates of P can be expressed in a parametric form in terms of ρ. Clearly, $\eta_1 = x \cos \rho, \eta_2 = x \sin \rho$, and by the property of a parabola,

$$x = k/2 + \eta_1 = k/2 + x\cos\rho \text{ i.e., } x(1-\cos\rho) = k/2. \tag{5.2.7}$$

Now, from Equation (5.2.6), as $(\cos u, \sin u)$ is a point on $A(u)$, we have, $\sin^2 u = k\cos u + k^2/4$, i.e., taking positive sign, $k = 2(1-\cos u)$. Then from Equation (5.2.7), $x(1-\cos\rho) = k/2 = (1-\cos u)$. So, $x = (1-\cos u)/(1-\cos\rho)$. Thus, the ancillary subspaces $A(u)$ can be represented parametrically as:

$$\eta_1 = \cos\rho(1-\cos u)/(1-\cos\rho), \eta_2 = \sin\rho(1-\cos u)/(1-\cos\rho).$$

Consider next the local coordinates associated with the test. In a neighborhood of M the parabola can be regarded as the pair of two ancillary subspaces or pieces of the curves $A(-u)$ and $A(u)$. We introduce a local coordinate system v in each of these ancillary subspaces which is defined as the arc-length from the intersecting point of M and $A(u)$. Then, from the formula for arc-length,

$$v = \int_\rho^u \left[\{\eta_1'(\rho)\}^2 + \{\eta_2'(\rho)\}^2\right]^{1/2} d\rho,$$

where differentiation is w.r.t. ρ. Now,

$$\eta_1'(\rho) = (1-\cos u)\left[-\frac{\sin\rho}{(1-\cos\rho)} - \frac{\cos\rho\sin\rho}{(1-\cos\rho)^2}\right]$$

and

$$\eta_2'(\rho) = (1-\cos u)\left[-\frac{\sin\rho}{(1-\cos\rho)} - \frac{\cos\rho\sin\rho}{(1-\cos\rho)^2}\right].$$

So,

$$\begin{aligned} v &= \int_\rho^u (1-\cos u)\left[-\frac{\sin^2\rho + (1-\cos\rho)^2}{(1-\cos\rho)^4}\right]^{1/2} d\rho \\ &= \sqrt{2}\int_u^\rho \frac{(1-\cos u)}{(1-\cos\rho)^{3/2}} d\rho. \end{aligned}$$

C. Locally Most Powerful Unbiased (LMPU) Test

The LRT is not known to be optimal for small sample size, while the other test is a conditional test. Here, in this Section, we propose a different unconditional unbiased test for the hypothesis (5.2.5), namely the locally most powerful unbiased (LMPU) test. The concept of LMPU test goes back to Neyman and Pearson (1936) and is based on finding a test φ_0 out of all θ-level tests φ satisfying

$$\mu_{\varphi}(\mu)\,|_{\mu=0} = \theta \qquad \text{and} \qquad \mu'_{\varphi}(\mu)\,|_{\mu=0} = 0 \tag{5.2.8}$$

and that which maximizes the value of the second derivative of $\mu_{\varphi}(\mu)$ at $\mu = 0$, that is

$$\mu''_{\varphi_0}(\mu)\,|_{\mu=0} > \mu''_{\varphi}(\mu)\,|_{\mu=0} \,.$$

A LMPU test can be found by using generalized Neyman-Pearson Lemma (see Lehmann (1986), pp. 96-101). According to this lemma, any test with critical region as follows will be the LMPU test:

$$\frac{\partial^2}{\partial\mu^2}L(\theta^*,\ \mu)|_{\mu=0}+\left(\frac{\partial}{\partial\mu}L(\theta^*,\mu)|_{\mu=0}\right)^2 > k_1+k_2\left(\frac{\partial}{\partial\mu}L(\theta^*,\ \mu)|_{\mu=0}\right), \tag{5.2.9}$$

where $\theta^* = (\theta_1,\ \theta_m)$, $L(\theta^*,\ \mu)$ is the loglikelihood function of $(\theta_1,\ldots,\theta_m)$ and $k_1,\ k_2$ is determined by (5.2.8). Hence, we have,

$$\frac{\partial}{\partial\mu}L(\theta^*,\ \mu)|_{\mu=0} = \kappa\sum_{i=1}^{m}\sin(\theta_i-\mu)|_{\mu=0} = \kappa\sum_{i=1}^{m}\sin\theta_i$$

and

$$\frac{\partial^2}{\partial\mu^2}L(\theta^*,\ \mu)|_{\mu=0} = -\kappa\sum_{i=1}^{m}\cos(\theta_i-\mu)|_{\mu=0} = -\kappa\sum_{i=1}^{m}\cos\theta_i. \tag{5.2.10}$$

Therefore, plugging (5.2.10) in (5.2.9), the critical function of this LMPU test for the hypothesis (5.2.5) is given below:

$$\varphi_0(\theta^*,\ \mu) = \begin{cases} 1 & \text{if}(-\kappa\sum_{i=1}^{m}\cos\theta_i) + (\kappa\sum_{i=1}^{m}\sin\theta_i)^2 > k_1 + k_2(\kappa\sum_{i=1}^{m}\sin\theta_i) \\ 0 & \text{otherwise.} \end{cases} \tag{5.2.11}$$

We now show the following lemma which allows us to drop k_2.

Lemma 5.1 *$k_2 = 0$ in (5.2.11).*

Proof: Define $U = \kappa \sum_{i=1}^{m} \sin\theta_i$ and $V = -\kappa \sum_{i=1}^{m} \cos\theta_i + (\kappa \sum_{i=1}^{m} \sin\theta_i)^2$. Then, the critical region (5.2.11) becomes

$$\varphi_0(\theta^*,\ \mu) = \begin{cases} 1 & \text{if } V > k_1 + k_2 U \\ 0 & \text{otherwise.} \end{cases} \tag{5.2.12}$$

Since

$$\mu'_{\varphi_0}(\mu)|_{\mu=1} = 0 \text{ if and only if } E_{\mu=0}\left\{\varphi_0(\theta^*,\ \mu)U\right\} = 0, \tag{5.2.13}$$

$$\begin{aligned} \mu'_{\varphi_0}(\mu) &= \frac{d}{d\mu}\int_0^{2\pi} \varphi_0(\theta^*,\ \mu)L(\theta^*,\ \mu)d\theta^* \\ &= \int_0^{2\pi} \varphi_0(\theta^*,\ \mu)\frac{d}{d\mu}L(\theta^*,\ \mu)d\theta^* \\ &= E\left\{\varphi_0(\theta^*,\ \mu)U\right\}. \end{aligned}$$

Now, $(U,\ V)$ and $(U,\ -V)$ have the same distribution, since V is an even function under H_0. Define $R = \{(u,\ v) :\ u > 0\}$ and $R^* = \{(u,\ v) : (-u,\ v) \in R\}$, the latter is the reflection of R about $u = 0$. Then,

$$\begin{aligned} & (U(\theta_1,\ \ldots,\ \theta_m),\ V(\theta_1,\ \ldots,\ \theta_m)) \in R \\ \Longleftrightarrow\ & (-U(\theta_1,\ \ldots,\ \theta_m),\ V(\theta_1,\ \ldots,\ \theta_m)) \in R^*. \end{aligned}$$

Thus,

$$P_{\mu=0}\left\{(U,\ V) \in R\right\} = P_{\mu=0}\left\{(U,\ V) \in R^*\right\}. \tag{5.2.14}$$

It is sufficient to show that Equation (5.2.13) holds if and only if $k_2 = 0$ in (5.2.11). First of all, we assume $k_2 = 0$ in (5.2.11) and P^* is the joint probability measure of $(\theta_1, \ldots, \theta_m)$ under the null hypothesis. From (5.2.13) and the fact that u is an odd function,

$$E_{\mu=0}\left\{\varphi_0(\theta^*,\ \mu)U\right\} = \int_{\{v>k_1\}} u\ dP^* = 0.$$

Conversely, we assume (5.2.14) holds. We want to show that $k_2 = 0$. If not, suppose $k_2 > 0$. Then,

$$\begin{aligned} E_{\mu=0}\{\varphi_0(\theta^*,\ \mu)U\} &= \int_{\{v>k_1+k_2u\}} udP^* \\ &= \int_{\{v>k_1+k_2u;\ u>0\}} udP^* + \int_{\{v>k_1-k_2u;\ u<0\}} udP^* \\ &\quad + \int_{\{k_1+k_2u<v<k_1-k_2u;\ u<0\}} udP^* \end{aligned}$$

and

$$\int_{\{v>k_1+k_2u;\ u>0\}} udP^* + \int_{\{v>k_1-k_2u;\ u<0\}} udP^* = 0$$

(because u is an odd function and (5.2.14)). Thus, it implies

$$E_{\mu=0}\{\varphi_0(\theta^*,\ \mu)U\} = \int_{\{k_1+k_2u<v<k_1-k_2u;\ u<0\}} dP^* = 0.$$

Also, the integral is over the negative value of integrand u, hence, by Ash (1972), p. 47, we conclude that

$$P^*\{k_1 + k_2U < V < k_1 - k_2U,\ U < 0\} = 0 \ \Rightarrow \ k_2 = 0$$

because, if $k_2 > 0$ as by assumption, then $P^*\{k_1 + k_2U < V < k_1 - k_2U,\ U < 0\} > 0$. Since P^* is symmetric about $(0, \ldots, 0)$, we then arrive at a contradiction to (5.2.13). Similar arguments work for the case $k_2 < 0$. Consequently, we have shown that $k_2 = 0$ in (5.2.11). □

According to the above Lemma 5.1, the critical region of the LMPU test of the hypothesis (5.2.5) becomes

$$-\sum_{i=1}^{m} \cos\theta_i + \kappa\left(\sum_{i=1}^{m} \sin\theta_i\right)^2 > k_1 \iff -C + \kappa S^2 > k_1,$$

where k_1 satisfies (5.2.8), i.e. k_1 is decided by the level of significance. Under the null hypothesis, we need to find the asymptotic distribution of $-C + \kappa S^2$ since the critical region is

$$-C + \kappa S^2 > k_1 \iff -\frac{1}{m}C + m\kappa\left(\frac{S}{m}\right)^2 > k^* \iff -\bar{C} + m\kappa\bar{S}^2 > k^*.$$

Using the identities (3.4.6) and (3.4.7), and Lemma 3.1 on the function

$$g(\bar{C},\ \bar{S}) = -\bar{C} + \kappa m \bar{S}^2$$

we get

$$\sqrt{n}\left\{-\bar{C} + \kappa n \bar{S}^2 - (-\alpha_1 + \kappa n \beta_1^2)\right\} \xrightarrow{d} N(0,\ \sigma^2),$$

where

$$\sigma^2 = \frac{1}{2}\left(1 + \alpha_1 - 2\alpha_2^2\right) + \frac{1}{2}\left(1 - \alpha_2 - 2\beta_1^2\right)(2\kappa n \beta_1)^2.$$

So, under the null hypothesis $H_0 :\ \mu = 0$, the distribution of α is unimodal and symmetric about 0. For the CN distribution, we have

$$\alpha_1 = \frac{I_1(\kappa)}{I_0(\kappa)}, \qquad \alpha_2 = \frac{I_2(\kappa)}{I_0(\kappa)}, \qquad \beta_1 = \beta_2 = 0.$$

Thus, under H_0,

$$\sqrt{n}\left\{-\bar{C} + \kappa n \bar{S}^2 + \frac{I_1(\kappa)}{I_0(\kappa)}\right\} \xrightarrow{d} N\left(0,\ \frac{1 + \alpha_1 - 2\alpha_2^2}{2}\right),$$

giving us the asymptotic distribution of this LMPU test. The critical values k and k^* can be easily obtained for large n. Both these tests are locally most powerful for all sample sizes, unbiased and the first one has monotonic increasing power in $(0,\ \pi)$. These are also consistent tests.

5.2.2 Higher-Order Power Comparison

For the case of the two-sided alternatives, the test based on the MLE and the LR test discussed above and also the LMP test are unconditional tests. Except for the LMP test, which is optimal for all sample sizes in the sense of maximum local power, no small-sample property of the other two tests is known. However, using standard results, e.g., following Amari (1985), we get the following results on the deficiencies of the tests.

Result 5.1 *The third order power loss of the LMP test compared to the MLE test is*

$$\begin{aligned} L_1(t) &= 1.78\tau^2\xi(t)\left[1 - 1/(2\tau^2) - J(t)\right]^2 \\ L_2(t) &= 1.78\tau^2\xi(t)J(t)^2, \\ \textit{and } L_3(t) &= 1.78\tau^2\xi(t)\left[1/2 - J(t)\right]^2, \end{aligned} \tag{5.2.15}$$

where τ *is the upper* $100(\alpha/2)\%$ *point of the standard normal distribution,* $\xi(t) = (t/2)[\phi(\tau - t) - (\tau + t)]$ *and* $J(t) = 1 - t/[2\tau \tanh t\tau]$.

This result is a re-statement of Theorems 6.6, 6.7, and 6.8 of Amari using $\gamma^2 = 1.78$ from Section 5.2.1 above.

Case of κ Unknown

When κ is unknown, the principle of similarity or invariance does not lead to any reduction and hence no useful unconditional test is available. A conditional test may be derived, and even a conditional LMP test could be constructed, i.e. a LMP test obtained from the conditional distribution, which is free of the nuisance parameter κ. Again however, to be useful in practice, one needs to construct extensive tables corresponding to the values of the conditioning (continuous) random variable. One may of course use the usual LRT, although it is neither simple in form nor does it have any small sample optimal properties. An exact conditional test for this case is already described earlier. Next, we consider an asymptotically optimal test, the Neyman's C_α-test.

B. The C_α Test

Here, we derive Neyman's C_α test. Let $\phi = \ln f(\theta, \kappa)$. Then at $\mu = 0$,

$$\phi_\mu = \kappa \sin\theta, \phi_\kappa = \cos\theta - A(\kappa).$$

Assume $\kappa < K_0 < \infty$. Then straightforward computations establish that all the conditions for ϕ_μ and ϕ_κ to be Crámer functions are satisfied. Consider testing $H_0 : \mu = 0$ against $H_1 : \mu > 0$. The C_α-test is to reject when

$$Z_m^* = \sum_{i=1}^{m} \left\{\phi_\mu(\theta, \hat{\kappa}) - a_1^0 \phi_\kappa(\theta, \hat{\kappa})\right\} / \sqrt{m}\sigma_0(\hat{\kappa}) > \tau_\alpha, \tag{5.2.16}$$

where $\hat{\kappa}$ is any locally root-m consistent estimator of κ under H_0, $\sigma_0(\hat{\kappa})$ is the standard deviation of $\phi_\mu(\theta, \kappa) - a_1^0 \phi_\kappa(\theta, \kappa)$ under H_0 and evaluated at $\kappa = \hat{\kappa}$ and a_1^0 is the partial regression coefficient of ϕ_μ on ϕ_κ. One may, e.g., take $\hat{\kappa}$ as the MLE of κ under H_0, i.e.,

$$\hat{\kappa} = \max\left\{0; A^{-1}(C/m), C > 0\right\}.$$

Further, a_1^0 is seen to be 0 by direct computation. Also, $E(\delta^2 \ell / \delta\mu\delta\kappa) = 0$ under H_0, holds. Then, the numerator of Z_m^* reduces to $\kappa \sum \sin\theta_i$ and thus

$\sigma_0^2(\kappa)$ reduces to $\sigma_0^2(\kappa) = Var_{\mu=0}(\kappa \sin\theta) = \kappa A(\kappa)$. Thus (5.2.16) reduces to the simple form

$$Z_m = \sqrt{\hat{\kappa}} \sum_{i=1}^{m} \sin\theta_i / (mA(\hat{\kappa})^{1/2}) > \tau_\alpha, \tag{5.2.17}$$

which involves computing $\hat{\kappa}$. This may be avoided to give an even simpler but asymptotically equivalent test. Note that $\sigma_0^2(\kappa) = \kappa^2 E_0(\sin^2\theta)$ and $\sum_{i=1}^{m} \sin^2\theta_i/m$ is a consistent estimator of $E_0(\sin^2\theta)$. Then (5.2.17) reduces to

$$T_m = \sum_{i=1}^{m} \sin\theta_i / \left(\sum_{i=1}^{m} \sin^2\theta_i \right)^{1/2} > \tau_\alpha. \tag{5.2.18}$$

T_m is equivalent to Z_m in the sense that it has, by Slutsky's theorem, the same limiting distribution as that of Z_m. For any sequence $\mu^* = \{\mu_n\}$ such that $\mu_n\sqrt{n} \to \gamma$, the asymptotic value of the power of the test is given by

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\tau_\alpha} \exp\left\{-(t - \sigma_0(\kappa)\gamma)^2/2\right\} dt.$$

Among all tests, T_m^*, for $H_0 : \mu = 0$ with asymptotic level of significance α, whatever the sequence of alternatives $\mu_m > 0$ with $\mu_m \to \mu_0 = 0$, and whatever the fixed $\kappa > 0$,

$$\underline{\lim}\, [\text{Power}\{T_m(\mu_m, \kappa)\} - \text{Power}\{T_m^*(\mu_m, \kappa)\}] \geq 0.$$

The test T_m is in this sense an asymptotically locally most powerful test. For testing against two-sided alternatives, a test can be based on the same test statistic as above, with obvious modification of the critical region. Neyman has shown that this also yields an asymptotically locally most powerful test.

C. An Approximate Test for Large κ

Recall the results of Section 3.6 : if κ is large, then we have

$$2\kappa(n - v) = 2\kappa(n - R) + 2\kappa(R - v), \tag{5.2.20}$$

where the terms have the approximate χ^2 decomposition

$$\chi_n^2 = \chi_{n-1}^2 + \chi_1^2$$

with the two χ^2 s on the right independently distributed. So an approximate test may thus be based on

$$\frac{2\kappa(R-v)}{2\kappa(n-R)/(n-1)} = \frac{(n-1)(R-V)}{(n-R)} \simeq F_{1,n-1}.$$

5.2.3 Tests for the Concentration Parameter

Observe that when μ is known, CN distribution reduces to a one-parameter REF with the canonical parameter coinciding with the natural parameter of interest κ. When μ is unknown, we are back to the two-parameter REF. Further, μ may be treated as a 'location' parameter and principle of invariance or similarity may be exploited when it appears as a nuisance parameter. We want to test $H_0 : \kappa = \kappa_0$ (known) against one or two sided alternatives. In particular, testing for $H_0 : \kappa = 0$ versus $H_1 : \kappa > 0$ is equivalent to testing for circular uniformity and is of more central concern.

Case 1. μ known

(i) Test for $H_0 : \kappa = \kappa_0 \quad (\neq 0)$

When the mean is known, say μ_0, observe that the CN population reduces to a member of the one-parameter REF, with κ the parameter of interest, being also the canonical parameter. Here

$$V = \sum_{i=1}^{n} \cos(\theta_i - \mu_0),$$

is a complete sufficient statistic whose null distribution $f_{\kappa_0}(v)$ is known (see (3.3.12)). Thus the UMP test for the one-sided alternative, and the UMPU test for the two-sided alternative, are simply based on V. For instance, against the two-sided alternative $H_a : K \neq K_0$, we reject H_0 when

$$V \leq v_1 \text{ or } V \geq v_2$$

with

$$P_0(V \leq v_1) + P_0(V \geq v_2) = \alpha.$$

(ii) Test for $H_0 : \kappa = 0$ versus $H_1 : \kappa > 0$

The UMP test, which coincides with the LRT, is to reject H_0 when $V > v_0$, where v_0 is obtained from the null distribution $f_0(v)$. This can be done using the routine `v0.test` in `CircStats` and is illustrated in Example 5.1. The power function is also easily computed using the expression for $f_\kappa(v)$ given in (3.3.12).

Case 2. μ unknown

Tests for $H_0 : \kappa = \kappa_0$

Note that R is a maximal invariant statistic here. It is easy to show that the pdf of R, though not a member of the exponential family, has the monotone likelihood ratio property with respect to κ. It then follows that the UMP Invariant test is to reject when

$$R > r_\alpha. \tag{5.2.21}$$

For two-sided alternatives, the UMPUI test is based on R also, with a two-sided critical region. Since the exact distribution of R under H_0 is known, see (3.2.4), the exact cutoff points can in general be obtained from it.

Tests for $H_0 : \kappa = 0$ versus $H_0 : \kappa > 0$

Following the same approach as in the case of μ known, the LRT reduces to a test based on the sample resultant length R, and rejects the null hypothesis when

$$R > r_0,$$

which is the familiar Rayleigh's test. This can also be seen to be the UMPI test for this problem. The exact distribution of R under circular uniformity is given by $f_0(r)$ (see Equation (3.2.4)). The Rayleigh test can be implemented using the subroutine `r.test` in `CircStats` as illustrated in the following example.

Example 5.1 *Recall the data in Table 1.2 where the interest was in testing whether the atomic weights of elements were uniformly distributed versus the specific alternative that they have a CN distribution with mean of zero. The* `v0.test` *is most appropriate for this situation. We also apply the* `r.test` *which tests uniformity against a general CN alternative with unspecified mean.*

```
> r.test(x.rad)
$r.bar: [1] 0.766282

$p.value: [1] 1.807656e-07

> v0.test(x.rad)
$r0.bar: [1] 0.7237434

$p.value: [1] 5.348331e-08
```

Both these tests strongly reject the null hypothesis of uniformity.

5.3 Two or More Populations

5.3.1 Comparing Mean Directions and Approximate ANOVA

A. An Exact Conditional Test

Consider two circular normal populations with parameters $(\mu_1,\ \kappa_1)$ and $(\mu_2,\ \kappa_2)$ and the problem of testing the hypothesis $H_0 : \mu_1 = \mu_2$ under the assumption of a *common* but unknown concentration parameter $\kappa_1 = \kappa_2$. If R_1, R_2 denote the lengths of the individual resultants and R, the length of the resultant for the combined sample, then it is clear that

$$|R_1 - R_2| \le R \le R_1 + R_2.$$

However if $\mu_1 = \mu_2$ then $S \equiv R_1 + R_2 \simeq R$. Hence, an approximate test is to reject the hypothesis $H_0 : \mu_1 = \mu_2$ if $S > s_0$, where s_0 is chosen in such a way that

$$P_H^{\kappa}(S \ge s_0|R) = \alpha.$$

Such an s_0 cannot be obtained computationally from the conditional distribution of R_1, R_2 given R (see (3.5.4)) which is independent of κ.

B. Approximate Tests and ANOVA

Now, let us consider testing for the equality of two or more mean directions from independent CN populations with the same concentration parameter.

We assume this concentration parameter to be large enough for the approximations of Section 3.5 to be valid. We wish to test

$$H_0 : \mu_1 = \cdots = \mu_p,$$

where the μ_i's are the polar directions for the p CN populations with the same (but unknown) concentration parameter κ, i.e. from independent $CN(\mu_i, \kappa)$, $i = 1, \ldots, p$ populations. Let n_i denote sample size from the ith population ($i = 1, \ldots, p$) and let $n = n_1 + \cdots + n_p$ denote the size of the combined sample. Under H_0, let μ denote the common polar direction. Compute the lengths of the individual sample resultants $R_1, \ldots, R_p$ and let

$R =$ the length of the overall resultant based on all the n observations.

From our earlier discussion, while $(n_i - R_i)$ is a measure of within dispersion of the ith sample based on n_i observations measured from their estimated mean direction, $(n - R)$ is a measure of dispersion of the combined sample of n observations from a common estimated mean direction. We have the identity

$$\begin{aligned} n - R &= \left(n - \sum_{i=1}^{p} R_i\right) + \left(\sum_{i=1}^{p} R_i - R\right) \\ &= \left(\sum_{i=1}^{p} (n_i - R_i)\right) + \left(\sum_{i=1}^{p} R_i - R\right) \end{aligned}$$

corresponding to the breakdown that total dispersion = within sample dispersion + between sample dispersion.

As discussed before in Section 3.6,

$$2\kappa(n - R) = 2\kappa \sum_{i=1}^{p} (n_i - R_i) + 2\kappa \left(\sum_{i=1}^{p} R_i - R\right)$$

has the approximate χ^2 decomposition

$$\chi^2_{n-1} = \chi^2_{n-p} + \chi^2_{p-1}.$$

From this, we can construct the following approximate Analysis of Variance (ANOVA) table for the analysis of mean directions (see Watson and Williams (1956)):

Approximate ANOVA Table

Source	d.f.	SS
Between samples	$p-1$	$\sum_1^p R_i - R$
Within samples	$n-p$	$\sum_1^p (n_i - R_i)$
Total	$n-1$	$n-R$

leading to the F ratio

$$F = \frac{(\sum R_i - R)/(p-1)}{\sum(n_i - R_i)/(n-p)}.$$

Note that $R \leq \sum_1^p R_i$ always but however if $\sum_1^p R_i$ is significantly larger than R, then we have reason to suspect the null hypothesis and reject H_0. The test is to reject when

$$F \equiv \frac{(n-p)(V-R)}{(p-1)(n-V)} \quad > \quad F_0, \tag{5.3.1}$$

where F follows $F(p-1, n-p)$ when κ is large, say larger than 2. When $1 < \kappa < 2$, one may improve the approximation using

$$F' = F\left(1 + \frac{3}{8\hat{\kappa}}\right)$$

with the same F-distribution approximation. Here $\hat{\kappa}$ is the estimated overall concentration parameter. Simulations show that this approximation is satisfactory for $\kappa \geq 1$, which translates to $\bar{R} \geq 0.45$ in applications. The test rejects H_0, the null hypothesis of equality of mean directions, for large values of the calculated F value. Rao and Sengupta (1972) discuss a geological application of this approximate ANOVA. See also Upton (1976) and and Section 12.4 for extensions.

Example 5.2 *Schmidt-Koenig (1958) reports on an experiment on pigeon-homing where the internal clocks of 10 birds were reset by 6 hours clockwise while the clocks of 9 birds were left unaltered. It is predicted from sun-azimuth compass theory that the mean direction of the vanishing angles in the experimental group should deviate by about* 90^o *in the anti-clockwise direction with respect to the mean direction of the angles of the birds in the control group. The vanishing angles of the birds for this experiment are given below, measured (in degrees) in the clockwise sense:*
Control group: 75, 75, 80, 80, 80, 95, 130, 170, 210.
Experimental group: 10, 50, 55, 55, 65, 90, 285, 285, 325, 355.
Does the data support sun-azimuth compass theory?

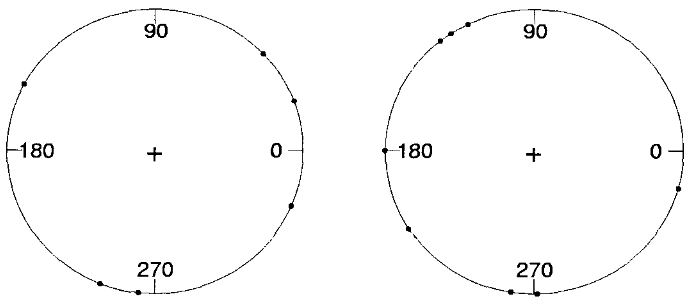


Figure 5.3: Control (left) and experimental (right) group of the pigeon-homing experiment.

In this case, n_1=9, n_2=10 and it can be verified that R_1=6.51, R_2=5.85 while the overall resultant for all the 19 observations is R=9.13. Thus the F-statistic takes the value 8.25. At 5% level of significance for a $F_{1,17}$ with critical value 4.45, we reject the null hypothesis that the mean directions are the same in both the experimental and control groups.

5.3.2 Tests for Concentration Parameters

We wish to test whether two independent CN populations $CN(\mu_i, \kappa_i)$, $i = 1, 2$ are equally concentrated, i.e. to test $H_0 : \kappa_1 = \kappa_2 = \kappa$, say, where κ is unknown. Let R_1 and R_2 be the resultants computed from two independent samples of sizes n_1 and n_2 taken from the above two populations, respectively.

A. An Exact Conditional Test

Consider the testing problem $H_0 : \kappa_1 = \kappa_2$. Since $\hat{\kappa}_i = A^{-1}(R_i/n_i)$, $i = 1, 2$ and $A^{-1}(\cdot)$ is a monotone function, we may use the test statistics

$$D = \left| \frac{R_1}{n_1} - \frac{R_2}{n_2} \right|$$

and reject H_0 if $D > d_0$. However, the distribution of D involves the unknown nuisance parameter κ, whereas the conditional distribution of R_1, R_2 and hence that of D for given R does not (see Equation (5.2.17)). Thus one can use the conditional test $P(D \geq d_0 | R) = \alpha$ utilizing the fact that $f_\kappa(R_1,\ R_2|R)$ is independent of κ (see (3.5.4)). If $n_1 = n_2$, the test can be equivalently based on the simple difference of the resultant lengths $D = |R_1 - R_2|$, given the overall resultant length, R.

B. An Approximate Test for Large κ

For large κ, we have $2\kappa(n - R) \simeq \chi^2_{n-1}$. Now, suppose we want to construct a confidence interval for κ, of the form

$$P[\kappa_L < \kappa < \kappa_U] = 1 - \alpha.$$

Then from χ^2-table with $(n-1)$ df, obtain lower and upper values χ^2_L and χ^2_U such that

$$P\left[\chi^2_L < \chi^2_{n-1} < \chi^2_U\right] = 1 - \alpha,$$

which could be taken with equal tails, for convenience. Since

$$P\left[\chi^2_L < 2\kappa(n - R) < \chi^2_U\right] = 1 - \alpha,$$

$$\kappa_L = \chi^2_L/\{2(n - R)\} \quad \text{and} \quad \kappa_U = \chi^2_U/\{2(n - R)\}$$

provide a $100(1-\alpha)\%$ confidence interval for the unknown κ.

C. An Approximate Test for Testing Equality of Concentrations

Using the approximation (see Appendix A)

$$\frac{I_1(\kappa)}{I_0(\kappa)} \doteq 1 - \frac{1}{2\kappa},$$

for large κ and the estimating equation for κ, namely

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

we have

$$1 - \frac{1}{2\kappa} \approx \frac{R}{n} \text{ or } \hat{\kappa} \approx \frac{n}{2(n-R)}.$$

In fact, it has been shown that $\hat{\kappa} = (n-1)/(2(n-R))$ gives a slightly better estimate of κ. Again for large κ, we have the approximate χ^2 distribution

$$\frac{2\kappa(n-R)}{n-1} \sim \frac{\chi^2_{n-1}}{n-1}.$$

Combining the above two results, we have

$$\frac{\kappa}{\hat{\kappa}} \sim \frac{\chi^2_{n-1}}{n-1}.$$

Hence, to test the hypothesis that $H_0 : \kappa_1 = \kappa_2$, one may use the approximate test statistic

$$F = \frac{\hat{\kappa}_1}{\hat{\kappa}_2} = \frac{(n_1 - R_1)(n_2 - 1)}{(n_2 - R_2)(n_1 - 1)} \sim F_{n_1-1,\ n_2-1}$$

which, under H_0, follows an F distribution with $(n_1 - 1)$ and $(n_2 - 1)$ dfs. For a q sample problem, an analogous technique can be followed; alternatively, one can apply Bartlett's M-test or Hartley's maximum F-ratio test.

Chapter 6

Tests for Uniformity

6.1 Introduction

The problem of testing for uniformity of a given set of directions i.e., their "randomness" or "isotropy", is of great importance in directional data analysis. For instance, astronomers have wondered whether the direction of the stars are uniformly distributed in the celestial horizon. von Mises (1918) wished to test the hypothesis that the atomic weights known till then, are integers subject to errors. Circular uniform distribution, as mentioned before, is the distribution with maximum entropy when the existence of a preferred direction is not assumed.

In oceanography, the direction of drift of icebergs and ice-floes has considerable impact on naval logistics as well as on a number of practical applications associated with transportation in ice-covered seas and in the design of off-shore structures. For such motions, the direction of movement has been assumed to follow the Circular Uniform density, because of its simplicity. This may be suspected in favor of an alternative circular distribution, say a member of the WSM family with a preferred direction of drift. As Watson (1982) put it – "It could well be that there is no preferred direction - for example, the pigeons may be unable to use any navigational clues and leave in random directions. A test for the stability of magnetization of rocks that left a formation (which would be magnetized uniformly) to be part of a conglomerate is that the direction of magnetization of pebbles in the latter is uniform. Thus tests here for uniformity are perhaps of more practical importance than on the line. .... the problem of testing uniformity arises more often here than on the line."

However, uniformity is a difficult property to verify and is certainly not something to assess by eye. There are many significance tests of uniformity, both nonparametric and against specific class of parametric alternatives. In testing uniformity against specific parametric models, which is what we discuss in this chapter, the alternative plays a crucial role in the construction of such tests. When an alternative parametric distribution or family is given, there is need for a more specific test which attains good power against such alternatives, than an omnibus test. And it would be quite appealing if such a test were to have good properties against a wide class of reasonable alternatives, i.e. it is 'robust' for a broader class. We take up the issue of nonparametric tests for uniformity in Chapter 7.

In this chapter, we show that for testing uniformity of circular data, Rayleigh's test based on R as well as the test based on V have many desirable properties such as local optimality, monotonicity and unbiasedness, admissibility, and consistency. Interestingly, these results not only hold against CN and Cardioid alternatives but also against the WS family with any α as well as to the WSM family. These tests can be easily implemented in practice using the routines `r.test` and `v0.test` of `CircStats`.

6.2 Uniformity against WS Alternatives

Consider testing uniformity against the WS family of distributions defined in (2.2.18), i.e. to test

$$H_0 : \rho = 0 \text{ against } H_1 : \rho > 0.$$

Here α, the index parameter for the stable family, may be unknown. We assume first that the location parameter μ_0 is known, which can then be taken to be $\mu_0 = 0$ without loss of generality. The case of μ_0 unknown can be tackled similarly using the maximal invariant statistic.

The case of known μ_0 has many interesting applications, as e.g. in testing for the von Mises data set on atomic weights being integers (see Table 1.2), $\mu_0 = 0$ is a natural alternative in such a case. For a given μ_0, tests based on

$$V = \sum_{i=1}^{n} \cos(\theta_i - \mu_0)$$

or equivalently (replacing $(\theta_i - \mu_0)(\text{mod } 2\pi)$ by θ_i), on

$$C = \sum_{i=1}^{n} \cos\theta_i$$

have been proposed. Recall that, $\bar{C} = C/n$ is the moment estimator of monotone function of the concentration parameter, the $A(\kappa)$ in CN or ρ in WS, etc. Further, the locally most powerful (LMP) test statistic for testing uniformity against CN or Cardioid distribution, also yields a test based on $\bar{C}$.

A similar result can be shown to extend to testing uniformity against mixture distributions like the CN-CU or Cardioid-CU distributions, even in the general case of unknown mixing proportion, p (see SenGupta and Pal (2001a)). The results here further extend the optimality robustness property of this test to all members of the WS-CU mixture, i.e. what we call the WSM family given in (2.2.22).

6.2.1 LMP Test Against the WS Family

Although the WS density for given α is a two-parameter distribution similar to the CN, it is not a member of the exponential family – either regular or curved, and the density is in the form of an infinite series. There is no non-trivial sufficient statistic for the parameters α, ρ and μ_0 and reduction through similarity and/or invariance cannot be achieved. Standard approaches fail and uniformly most powerful (UMP) tests do not exist. Moreover, H_0 specifies a boundary point on the parameter space for $\rho \in [0, 1]$ and thus the likelihood ratio test poses both theoretical and practical problems. The LMP test provides a simple yet robust test and has many exact and asymptotic optimality properties.

The concept of an LMP test goes back to Neyman and Pearson (1936) and enjoys wide popularity since it is easy to obtain. Let $\beta_\varphi(\rho) = E(\varphi_\rho(\theta))$ be the power function of the test φ at ρ. A test φ_0 is said to be a LMP test for testing $H_0 : \rho = 0$ against $H_1 : \rho > 0$ if for any other test φ of the same size α, $\exists\ \delta > 0$, such that $\beta_{\varphi_0}(\rho) > \beta_\varphi(\rho) \quad \forall\ \rho \in (0, \delta)$. That is, close to the null hypothesis, β_{φ_0} has maximum power among all other tests of the same size. Such tests are obtained by maximizing the slope of power function at the null hypothesis (see for instance Rao (1973), pp. 453-454).

Let $\theta_1, \cdots, \theta_m$ be i.i.d. observations from $f(\theta; \rho, \alpha, 0)$ as given in (2.2.18). Thus the LMP test is to reject H_0 when,

$$\sum_{i=1}^{n} \frac{d}{d\rho} \in f(\theta_i; \rho)|_{\rho=0} > k, \tag{6.2.1}$$

where k is a constant chosen to satisfy the prescribed level of significance. Applying this to the WS density in (2.2.18) we reject the null hypothesis of uniformity $H_0 : \rho = 0$ when

$$C = \sum_{i=1}^{n} \cos\theta_i > k. \tag{6.2.2}$$

Note that the test statistic is simple to compute, even though the density (2.2.18) looked formidable. Recall that this is also the Likelihood Ratio test for testing uniformity i.e., $\kappa = 0$ against the alternative of a CN distribution with 0 mean direction and concentration parameter $\kappa > 0$ and is indeed the Uniformly Most Powerful test for that problem.

The exact distribution of C for the uniform distribution, i.e. under our H_0 is the same as that of V given in equation (3.3.11) with $\mu = 0$. Critical values and P-values can be found using the routine *v0.test* in `CircStats`.

6.2.2 Monotonicity of Power Function

In the previous section we obtained the LMP test. But in non-regular exponential families there are specific examples (see, e.g., Chernoff (1951)) which demonstrate that the LMP test may perform badly. However, in this and the following sections, we establish that our LMP test possesses many optimal properties. The main result of this section, namely Theorem 6.1, establishes the monotonicity of its power function. The proof depends on some results from Fourier analysis and the method of induction. It is presented in the Appendix at the end of this chapter, after a series of lemmas.

Theorem 6.1 *The LMP test for testing uniformity against a specific member, of the wrapped stable family with known α, has a monotone increasing power function $\beta(\rho)$ for $\rho \in (0, h(\alpha))$, where $h(\alpha)$ is given by* $\exp[-2 \in 2/(2^\alpha)]$. *Moreover, for the wrapped Cauchy ($\alpha = 1$) and the wrapped normal ($\alpha = 2$) distributions, $\beta(\rho)$ is globally monotone increasing, i.e. has increased power for all $\rho \in (0, 1)$. The LMP test is locally unbiased for all members of WS family.*

Remark 6.1 *The result above stating the monotonicity of power in the range $\rho \in (0, h(\alpha))$ is perhaps a partial one and more might indeed be true. As the same result states, the monotonicity property holds for all values of $\rho \in$ (0, 1), when $\alpha = 1$ and $\alpha = 2$.*

6.2.3 Consistency and Other Optimal Properties

In this section, we establish some other optimal properties of this test, such as admissibility, null and optimality robustness, and asymptotic-consistency.

Theorem 6.2 *The LMP test in (6.2.2) is admissible, null robust, optimality robust and consistent for $\rho \in (0,1)$ against all members of the WS family, with unknown $\alpha \in (0,2]$.*

Proof: First consider the WS family. Admissibility is a consequence of the uniqueness of the LMP. Null-robustness holds since under H_0 the distribution of θ reduces to a uniform distribution free of α and optimality robustness holds since the same LMP test statistic, free of α and whatever be the distribution of θ, results in the WS family with any specified α.

We establish below global consistency, i.e. for $\rho \in (0,1)$, against any alternative distribution in WS, $\alpha \in (0,2]$. In view of the unimodality and symmetry of the WS distribution, and comparing the pdf (2.1.5) with (2.2.18), we have

$$\alpha_p^* = E(\cos(p\theta)) = \rho^{p^\alpha}, \quad \beta_p^* = E(\sin(p\theta)) = 0. \tag{6.2.3}$$

Let $\bar{C}_m = (\sum_{i=1}^m \cos\theta_i)/m$. Then by the Central Limit Theorem and Equation (6.2.3), for large m,

$$\bar{C}_m \sim N(\rho, (\rho^{2^\alpha} + 1 - \rho^2)/2m).$$

Under $H_0, \bar{C}_m \sim N(0, 1/2m)$, see Section 3.4. Hence for the asymptotic level of significance α, the cut-off point is, $\bar{C}_m = z_\alpha/(\sqrt{2m})$, where z_α is the upper 100α-percent point of $N(0,1)$. Thus,

$$\begin{aligned}
\beta(\rho) &= P_\rho(\bar{C}_m > c) \\
&\simeq 1 - \Phi[\{z_\alpha - \rho(2m)^{1/2}\}/(\rho^{2^\alpha} + 1 - \rho^2)^{1/2}] \\
&\to 1 \text{ as } m \to \infty, \quad \forall\, \rho \in (0,1) \text{ and } \forall\, \alpha \in (0,2],
\end{aligned}$$

where Φ is the distribution function of $N(0,1)$. Thus the test is consistent as claimed. □

Remark 6.2 *The test in (6.2.2) has the same properties stated in Theorem 6.2 against the CN and Cardioid distributions. Recall the CN density* $f_1(\theta; \mu_0, \kappa)$ *and the Cardioid density* $f_2(\theta; \mu_0, \rho)$ *defined in (2.2.6) and (2.2.3), respectively. The hypothesis of circular uniformity translates to* $H_0 : \kappa = 0$ *for the CN density and to* $H_0 : \rho = 0$ *for the Cardioid density, against the one-sided alternatives* $H_1 : \kappa > 0$ *and* $H_1 : \rho > 0$*, respectively. We assume, as before,* $\mu_0 = 0$.

Direct derivations of the LMP tests yield the C*- or* V*-test for both the above densities. Then the properties stated in the theorem follow by arguments similar to those made above for the WS family.*

Remark 6.3 *(a) Note that for* μ_0 *known,* $f_1(\theta; \mu_0, \kappa)$ *becomes a member of the regular exponential family. Thus the unique LMP* C*-test in (6.2.1) in fact becomes the UMP test.*

(b) It can be shown that the power functions of the C*-test are in fact globally monotone increasing against the CN* ($\kappa > 0$) *and Cardioid* ($0 < \rho < 1/2$) *distributions.*

(c) Similarly it can be shown that when μ_0 *is unknown, the LMP invariant test for all the above distributions, i.e. CN, Cardioid and WS family with unknown* $a \in (0, 2]$*, yields the same simple* R*-test which is to reject* H_0 *when* $R = \sum \cos(\alpha_i - \overline{\alpha}_0) > k$*. Thus the* R*-test possesses all the optimality properties similar to those as in Theorems 6.1 and 6.2 above against all such alternatives. Derivations are based on the distributions of the maximal invariant and are lengthy, although analytically interesting (see Chang (1991)).*

6.3 Uniformity versus WSM Alternatives

Mixture models arise naturally for directional data, possibly more often so than for the linear case. They are becoming increasingly popular, for example, for modeling bidirectional data peculiar to directional variables or for modeling situations similar to the linear case, as with possible outliers for circular data, e.g., Ducharme and Milasevic (1987), Guttorp and Lockhart (1988), etc. However, not much is known about constructing optimal tests, say for the hypothesis of no mixture. The asymptotic local minimaxity does not hold either for the likelihood ratio test (LRT) or for the Bayes test (see Feder (1968)). It is clear that the LRT is computationally quite involved. Moreover, even in the strongly identifiable case, Ghosh and Sen (1985) show that the asymptotic log LRT statistic is cumbersome, its asymptotic distribution intractable, and its optimal properties almost impossible to explore analytically. Finally, as we said before, the density function (2.2.22) does not admit of any reduction via sufficiency, similarity or invariance. Thus a globally best test is not expected and we are led to search for a locally optimal test.

Let $\theta_1, \ldots, \theta_m$ be i.i.d. observations from a member of the WSM family with probability density function given by (2.2.22). As before, we assume $\mu_0 = 0$ without loss of generality. Consider first the case when p is known, $p > 0$. Then testing the hypothesis of uniformity against the WSM family is equivalent to testing $H_0 : \rho = 0$ against $H_1 : \rho > 0$. For brevity, let us denote $f^*(\theta; p, \rho, a, \mu_0)$ by $f^*(\theta; \rho)$.

Again the LMP test rejects when

$$\sum_{i=1}^{n} \cos\theta_i > k,$$

where k is determined by the level condition. Thus this test is the same as that given in (6.2.2) for WS alternatives. Under the null hypothesis of no mixture, $f^*(\theta, \rho)$ is the uniform distribution on the unit circle and its distribution is as given in (3.3.11). The critical values or the P-value can be determined again by using the `CircStats`.

6.3.1 Monotonicity of the Power Function

As stated earlier, an LMP test may in general perform badly. We show that this is not the case for the C-test, by establishing some of its optimal properties, starting with the monotonicity property of the power function. If $Z = \sum_{i=1}^{m} \cos\theta_i$, then under the alternative of mixture model (2.2.22), one may have any s of the m terms coming from the uniform component and the other $(m - s)$ coming from the WS component. Let G_0 and G_ρ denote the d.f.'s of $\cos\theta$ when θ has the uniform and WS distribution respectively. Let $U_{s1}, \ldots, U_{ss}$ be i.i.d. with the distribution G_0 while the remaining $U_{ss+1}, \ldots, U_{sm}$ be i.i.d. with the distribution G_ρ. Then for given $s = 1, \ldots, m$, let

$$Z_{(s)} = \sum_{i=1}^{m} U_{si} = \sum_{i=1}^{s} U_{si} + \sum_{i=s+1}^{m} U_{si} = Z_{1s} + Z_{2s}, \text{say}.$$

Since the distribution of $Z_{(s)}$ is given by

$$G_{Z_{(s)}} = G_0^{s^*} * G_\rho^{(m-s)^*},$$

where G^{k^*} is the k-fold convolution of G with itself, the d.f. $H_\rho(x)$ of $\sum_{i=1}^m \cos\theta_i$ under the mixture alternative, is given by

$$H_\rho(x) = \sum_{s=0}^{m} C_s^m (1-p)^s p^{(m-s)} G_{Z_{(s)}}(x).$$

Theorem 6.3 *The LMP test defined above has a monotone increasing power function $\beta(\rho)$ for $\rho \in (0, h(\alpha))$, where $h(\alpha)$ is given in Theorem 6.1. $\beta(\rho)$ is globally monotone increasing against the WN-CU and WC-CU mixture distributions. This test is locally unbiased for all members of WSM family.*

Proof: The power function is given by

$$\beta(\rho) = P_\rho\left\{\sum_{i=1}^{m} \cos\theta_i > c\right\} = 1 - H_\rho(c).$$

If $\rho_1, \rho_2 \in (0, h(\alpha)), \rho_1 < \rho_2$ then by Lemma 6.4

$$G_{\rho_1}(c) = 1 - P_{\rho_1}\{\cos\theta > c\} > 1 - P_{\rho_2}\{\cos\theta > c\} = G_{\rho_2}(c).$$

Further, $G_{\rho_1}^{k^*}(c) > G_{\rho_2}^{k^*}(c)$ for all $k^* > 0$ as a consequence of convolution. So,

$$\begin{aligned}
\beta(\rho_1) &= 1 - \sum_{s=0}^{m} C_s^m (1-p)^s p^{(m-s)} G_0^{s^*} * G_{\rho_1}^{(m-s)^*}(c) \\
&< 1 - \sum_{s=0}^{m} C_s^m (1-p)^s p^{(m-s)} G_0^{s^*} * G_{\rho_2}^{m-s)*}(c) \\
&= \beta(\rho_2).
\end{aligned}$$

In view of Lemmas 6.5 and 6.6, the monotonicity property holds globally for the mixture model where the second component in (2.2.22) is a WN or a WC distribution. Local unbiasedness, for all members of the WSM family is a trivial consequence of the above proof. □

6.3.2 Consistency and Other Optimal Properties

Some further optimality properties of the above test follow.

Theorem 6.4 *The LMP test is admissible, null robust, optimality robust and consistent for $\rho \in (0,1)$ against all members of the WSM family, with unknown $\alpha \in (0,2]$, and known $p > 0$, and as well as against the CN-CU and Cardioid - CU mixture distributions.*

Proof: We begin with the consistency. Under the alternatives $H_1 : \rho > 0$, since

$$E(\cos\theta) = p\rho, \quad Var(\cos\theta) = \frac{1}{2}\left\{p\rho^{2^\alpha} + 1 - p^2\rho^2\right\},$$

$\bar{C}_m = \sum_{1=1}^{m}(\cos\theta_i)/m$ is asymptotically $N(p\rho, (p\rho^{2^\alpha} + 1 - p^2\rho^2)/(2m))$. In particular, under $H_0 : \rho = 0$, the asymptotic distribution of $\bar{C}_m$ is $N(0, 1/2m)$. Thus, the consistency follows, as do the admissibility and robustness by arguments similar to those used in Theorem 6.2.

Remark 6.4 *Remark similar to 6.3(b) on the global monotonicity of the power functions against CN-CU and Cardioid - CU mixture distributions, hold here also.*

6.3.3 Optimal Test with Unknown p

The parameter space of our mixture model can be expressed as

$$\Theta = \{(p,\rho) \ : \ (p,\rho) \ \in \ [0,1] \times [0,1)\}.$$

Since the WSM family (2.2.22) reduces to the uniform distribution if either $p = 0$ (no mixture) or $\rho = 0$ (the mixing distribution is also the uniform), we have a problem of identifiability. In other words, we can pass to the one-dimensional parameter space of H_0 by specifying either one of the coordinates in the two-dimensional space of H_1. Of course, one could try to get rid of the identifiability problem by a reparametrization. That is, redefine the Euclidean parameter space in a one-to-one mapping with the original mixture model. However, this cannot be done due to the lack of differentiability of the density function at the null hypothesis (Ghosh and Sen (1985), p. 791). Because of the identifiability problem, we can see that when H_0 is true, one cannot confine attention to the neighborhood of a single point or use the derivative of the density function with respect to the parameters around the null hypothesis. Thus the locally best test, for example the LMP test, is not available. Following a concept introduced in SenGupta (1991) for such situations, we obtain an L-optimal, an "optimal" test for mixture distributions, based on a Pivotal Parametric Product (P^3). For each fixed known p, such a test matches the power of the corresponding LMP test.

Note that H_0 is true iff either $p = 0$ or $\rho = 0$ or both $p = 0$ and $\rho = 0$, i.e. iff $p\rho = 0$. Since $E(\cos\theta) = p\rho$, we may take

$$\bar{C}_m = \frac{1}{m}\sum_{i=1}^{m}\cos\theta_i$$

to be the pivotal quantity P^3 for the hypothesis. Since $E(\bar{C}_m) = p\rho \geq 0$, it makes sense to reject the null hypothesis H_0 if $\bar{C}_m$ is large enough. Therefore, a test based on P^3 suggests to reject when,

$$\bar{C}_m > k^* \Leftrightarrow C > k,$$

where k is determined by the level condition.

Once more we obtain the same test as in (6.2.2), i.e. the LMP test for uniformity against WS and WSM (with known p) families. The monotonicity property of the power function in ρ for fixed p follows as in Theorem 6.1 and the consistency and null robustness follow as in Theorem 6.2. Thus, we get

Theorem 6.5 *For testing uniformity, the test given by (6.2.2) yields a robust statistic and is null robust against all members of the WSM family with unknown $\alpha \in (0, 2]$ and unknown $p \geq 0$ as well as against the CN-CU and Cardioid - CU mixture distributions. It is L-optimal for each fixed p, has monotonically increasing unbiased power as in Theorem 6.1 and is consistent.*

Remark 6.5 *Remarks similar to those in 6.1 on the global monotonicity of the power functions, for fixed p, against CN - CU and Cardioid - CU mixture distributions as well as on the L-optimality robustness of the R-test against all the above mixture distributions also hold here.*

Remark 6.6 *Theorem 6.2 may be viewed as one that establishes the "omnibus" null and optimality robustness of the simple C-test. Theorems 6.4 and 6.5 make the test even more appealing for applications to real life problems by establishing its validity and extending its optimality to mixture models, which allow one to incorporate possible outliers.*

6.4 LMP Invariant Test for Unknown μ

Consider now the case when μ is unknown. The pdf $f(\theta;\rho,\mu)$ of the WS(ρ,μ) can be rewritten as $f(\theta-\mu;\rho)$, and μ is a location parameter. The family of distributions $f(\theta;\rho,\mu)$ is invariant under the group of translations of the zero direction. We can utilize this property to find the locally most powerful invariant (LMPI) tests which depend on the gaps between successive points (see Ajne (1968), Rao (1969)).

Since invariance under changes of origin or rotational-invariance is equivalent to translation invariance, by standard arguments (see Lehmann (1986) pp. 289-290),

$$u_i = (\theta_i - \theta_n) \pmod{2\pi}; \qquad i = 1,\ldots,n-1, \tag{6.4.1}$$

is a maximal invariant statistic and any invariant test should be based on the likelihood function of $(u_1,\ldots,u_{n-1})$. This joint likelihood function of $(u_1,\ldots,u_{n-1})$ is given by

$$L^*(u_1,\ldots,u_{n-1};\rho) = \int_0^{2\pi} f(\theta_n;\rho)\prod_{i=1}^{n-1} f(\theta_n+u_i;\rho)d\theta_n. \tag{6.4.2}$$

Since the null and the alternative remain invariant under such rotations and the group acting on the parameter space is transitive, the most powerful invariant test is obtained by simply using the Neyman-Pearson Lemma. For any given ρ, it has the critical region

$$L^*(u_1,\ldots,u_{n-1};\rho) > k,$$

where k is a constant determined by the level of the test. Using (6.4.1) in (6.4.2) the critical region can be written in the form

$$L^*(u_1,\ldots,u_{n-1};\rho) \quad = \quad \int_0^{2\pi}\prod_{i=1}^{n} f(x+\theta_i;\rho)dx > k,$$

or

$$\int_0^{2\pi}\prod_{i=1}^{n}\frac{1}{2\pi}\left\{1+2\sum_{m=1}^{\infty}\rho^{m^\alpha}\cos m(x+\theta_i-\mu)\right\}dx > k.$$

If we now define, for $i=1,\ldots,n$,

$$A_i(\rho,x) = 2\sum_{m=1}^{\infty}\rho^{m^\alpha-1}\cos m(x+\theta_i-\mu),$$

then we can rewrite

$$f(x+\theta_i;\rho) = \frac{1}{2\pi}\{1+\rho A_i(\rho,x)\},$$

where

$$A_i(0,x) = 2\cos(x+\theta_i-\mu). \tag{6.4.3}$$

Since this is of the form

$$\int_0^{2\pi}\prod_{i=1}^n f(x+\theta_i;\rho,\mu)dx \;=\; \int_0^{2\pi}\left(\frac{1}{2\pi}\right)^n\prod_{i=1}^n\{1+\rho A_i(\rho,x)\}\,dx,$$

we can appeal directly to the results of Beran (1969). Expressing this product as a power series in ρ and noticing that

$$\begin{aligned}
\lim_{\rho\longrightarrow 0}\frac{1}{\rho}\int_0^{2\pi} h_1(\rho,x)dx &= \lim_{\rho\longrightarrow 0}\int_0^{2\pi}\sum_{i=1}^n A_i(\rho,x)dx \\
&= \sum_{i=1}^n\int_0^{2\pi}\lim_{\rho\longrightarrow 0}A_i(\rho,x)dx \\
&= \sum_{i=1}^n\int_0^{2\pi} 2\cos(x+\theta_i-\mu)dx \\
&= 0,
\end{aligned}$$

we see that the critical region of the LMPI test is of the form:

$$L^*(\theta_1,\ldots,\theta_n;\rho)>k \;\Rightarrow\; (2\pi)^{1-n}\left\{1+2\rho^2R^2+o(\rho^2)\right\}>k,$$

i.e., the lowest order random term is the coefficient of ρ^2, namely R^2. Thus the locally best invariant test is the Rayleigh test with the critical region

$$R^2>k.$$

Consequently, we have the result:

Theorem 6.6 *The locally most powerful invariant (LMPI) test for circular uniformity against the symmetric wrapped stable family of distributions is given by the critical region*

$$R^2>k.$$

The null distribution of this statistic R, the length of the resultant, is given in Equation (3.3.3). It is easy to see that the LMPI test for uniformity against the CN distribution is also the same Rayleigh test based on R^2. Also Bhattacharyya and Johnson (1969a) show that this same test is LMPI for testing uniformity against the offset normal distribution. Theorem 6.6 extends this optimality robustness of Rayleigh test to a much broader class, namely the WS family.

6.4.1 Monotonicity of the Power Function

In the previous section, we derived the LMPI test of uniformity against the WS family when the location parameter is unknown. We now develop some optimal properties for this case, in this and the following sections. We first show that the power function has monotonicity property for the concentration parameter in a certain range. Define the power function of the LMPI test in terms of ρ by

$$\beta(\rho) = P_\rho\left\{R^2 > k\right\}. \tag{6.4.4}$$

Remark 6.7 *The power function $\beta(\rho)$ is independent of μ, since the distribution of R^2 is independent of μ.*

In order to show the monotonicity of $\beta(\rho)$, we begin with just two observations and use mathematical induction.

Lemma 6.1 *For $n = 2$, the power function $\beta(\rho) = P_\rho\{\cos\theta^* > k'\}$, where $\theta^* = \theta_1 - \theta_2 \pmod{2\pi}$ and k' is in $(-1,\ 1)$. Then, $\beta(\rho)$ is an increasing function in ρ for ρ in the range $(0,\ h^*(\alpha))$, where*

$$h^*(\alpha) = \exp\left\{-\frac{\alpha\log 2}{2(2^\alpha - 1)}\right\}. \tag{6.4.5}$$

Proof: Let $\theta^* = \theta_1 - \theta_2 \pmod{2\pi}$ and $k' \in (-1,\ 1)$, then the power function is given by

$$\begin{aligned}\beta(\rho) &= P_\rho\left\{\cos\theta^* > k'\right\}\\ &= P_\rho\left\{0 < \theta^* < \arccos k'\right\} + P_\rho\left\{2\pi - \arccos k' < \theta < 2\pi\right\},\end{aligned}$$

where $\arccos k'$ is in $(0,\ 2\pi)$.

In view of Lemma 2.1, θ^* is WS(0, ρ^2). Using the symmetry and unimodality of WS(0, ρ^2), we can follow the same arguments as in Theorem 6.1 with ρ replaced by ρ^2. Thus the power function is an increasing function in $\rho \in (0,\ h^*(\alpha))$, where $h^*(\alpha)$ is given by (6.4.5). □

Remark 6.8 *As in Remark 5.1, the same conclusion of Lemma 6.1 can be extended to the whole range of $\rho \in (0,\ 1)$ for the wrapped Cauchy and the wrapped normal distributions.*

Theorem 6.7 *The power function of the LMPI test $\beta(\rho)$ given by (6.4.4) is an increasing function in $\rho \in (0,\ h^*(\alpha))$, where $h^*(\alpha)$ is given by (6.4.5).*

Proof: To establish this theorem, we follow the same arguments as we did in Theorem 6.1.

(i) When $n = 2$, the claim is true by Lemma 6.1.

(ii) When $n = 3$, the power function is

$$\beta(\rho) = P_\rho \{\cos\alpha_1 + \sin\alpha_2 + \cos\alpha_3 > k\},$$

where $\alpha_1 = \theta_1 - \theta_2 \pmod{2\pi}$, $\alpha_2 = \theta_1 - \theta_3 \pmod{2\pi}$ and $\alpha_3 = \theta_2 - \theta_3 \pmod{2\pi}$. Consider $X = \cos\alpha_1 + \cos\alpha_2$ and $Y = \cos\alpha_3$. As in Theorem 6.1, it is known that X has the stochastic ordering property and so does Y and hence it follows that $X + Y$ has the stochastic ordering property. Hence, we have shown that the hypothesis holds for $n = 3$.

(iii) We assume that the hypothesis is true for $n = k$.

(iv) It is left to show that the hypothesis is still true for $n = k + 1$. For the $n = k + 1$ case, it is the sum of $n(n-1)/2$ cos functions with arguments involving all possible pairs of differences among those $k + 1$ observations. From (iii), with $n = k$, note that

$$\begin{aligned} &\sum_{1\le i_1<i_2\le k+1} \cos(\theta_{i_1} - \theta_{i_2}) \\ =\ & \sum_{1\le i_1<i_2\le k} \cos(\theta_{i_1} - \theta_{i_2}) + \sum_{i=1}^{k+1} \cos(\theta_i - \theta_{k+1}) \\ =\ & X + Y, \qquad \text{say.} \end{aligned}$$

Now, $X = \sum_{1 \le i_1 < i_2 \le k} \cos(\theta_{i_1} - \theta_{i_2})$ has stochastic ordering in ρ for $\rho \in (0,\ h^*(\alpha))$. Also, $Y = \sum_{i=1}^{k+1} \cos(\theta_i - \theta_{k+1})$ has stochastic ordering by the same arguments. So, the power function $\beta(\rho) = P_\rho\{X + Y > k'\}$ is an increasing function in $\rho \in (0,\ h^*(\alpha))$. □

Remark 6.9 *Theorem 6.7 can be extended for all values of* $\rho \in (0,\ 1)$ *for the wrapped Cauchy and the wrapped normal distributions because of Remark 6.8.*

Simulations of the power function for several different values of α in the WS and different sample sizes indicate that the test performs very well. Also, these simulations suggest that the monotonicity property perhaps holds for $\rho \in (0,\ 1)$.

6.4.2 Consistency and Other Optimal Properties

We now examine other optimal properties, such as consistency and robustness. Consistency also follows from the results of Beran (1969) but can be seen more directly from the asymptotic distribution of the test statistic R when $\rho > 0$. Using results of Section 3.4.3 and Lemma 3.1 to the WS distribution, it can be seen that asymptotically, $E(\bar{R}) = \rho$ and

$$\mathrm{Var}(\bar{R}) = \frac{1}{2n\rho^2}\left\{\rho^2(1+2\rho^2) + \rho^{2^\alpha+2}\right\}$$

so that

$$\bar{R} \sim N\left(\rho,\ \frac{1}{2n\rho^2}\left\{\rho^2\left(1+2\rho^2\right) + \rho^{2^\alpha+2}\right\}\right). \tag{6.4.6}$$

However, under the null hypothesis $\rho = 0$, $2R^2/n$ has asymptotic χ^2 distribution with 2 degrees of freedom as we have seen in (3.4.2). Thus, given the significance level a and for n large enough, we have

$$P_{\rho=0}\left\{R^2 > k\right\} = P_{\rho=0}\left\{\frac{2R^2}{n} > \frac{2k}{n}\right\} = a,$$

so that $2k/n = \chi^2_{2a}$, implying $k = (n/2)\chi^2_{2a}$, where χ^2_{2a} is the a-th quantile of the χ^2_2 distribution. Therefore, the critical region of this test becomes

$$R^2 > \frac{m}{2}\chi^2_{2a}$$

or,

$$\bar{R} = \frac{R}{n} > \sqrt{\frac{\chi^2_{2a}}{2n}}.$$

Thus the power function, for large n, is

$$\begin{aligned}
\beta_n(\rho) &= P_\rho\left\{\frac{\bar{R}-\rho}{\sqrt{\frac{1}{2n\rho^2}\{\rho^2(1+2\rho^2)+\rho^{2\alpha+2}\}}} > \frac{\sqrt{\frac{\chi^2_{2a}}{2n}}-\rho}{\sqrt{\frac{1}{2n\rho^2}\{\rho^2(1+2\rho^2)+\rho^{2\alpha+2}\}}}\right\} \\
&\approx P_\rho\left\{Z > \frac{\sqrt{\frac{\chi^2_{2a}}{2n}}-\rho}{\sqrt{1}2n\rho^2\{\rho^2(1+2\rho^2)+\rho^{2\alpha+2}\}}\right\} \\
&= 1-\Phi\left(\sqrt{\frac{\chi^2_{2a}\rho^2}{\rho^2(1+2\rho^2)+\rho^{2\alpha+2}}} - \rho^2\sqrt{\frac{2n}{\rho^2(1+2\rho^2)+\rho^{2\alpha+2}}}\right)
\end{aligned}$$

which tends to 1, as $n \longrightarrow \infty$, proving the consistency. □

Theorem 6.8 *The LMPI given in Theorem 6.6 has the properties of consistency and robustness.*

Proof: The consistency of this test procedure has been established by the above arguments. The robustness is due to the fact that the same test statistic goes through for the whole family of the WS distribution. □

Remark 6.10 *For the wrapped Cauchy and the wrapped normal distributions, the unbiasedness property will hold as well because of Remark 6.1.*

6.5 Appendix: Proofs

In this section we prove Theorem 6.1 for which we need the following series of lemmas. The first of these is easy to verify and the second one is a standard result in Fourier series.

Lemma 6.2 *Let X and Y be two random variables defined on the same probability space $(\Omega, \Im, P)$ with joint density function $l_\theta(x,y)$ and marginal distribution functions $F_\theta(x)$ and $G_\theta(y)$, respectively. Suppose that $F_\theta(x)$ and*

$G_\theta(y)$ are absolutely continuous functions and θ is an unknown parameter in R. Suppose further that each X and Y has stochastic ordering property, that is,

$$\theta_1 < \theta_2 \Rightarrow F_{\theta_1}(x) > F_{\theta_2}(x) \text{ and } G_{\theta_1}(y) > G_{\theta_2}(y), \forall x, y \in R. \tag{6.5.1}$$

Then $(X + Y)$ also has stochastic ordering property.

□

Lemma 6.3 *(Wintner (1947)) Let (i) $\theta \in (0, \pi)$, (ii) $nb_n \longrightarrow 0$ as $n \longrightarrow \infty$ and (iii) $nb_n - (n+1)b_{(n+1)} > 0$ for all $n \geq 1$. Then*

$$\sum_{n=1}^{\infty} b_n \sin(n\theta)$$

is a positive function.

To prove the main result, we first begin with only one observation and then apply induction.

Lemma 6.4 *Define $g(\rho) = P_\rho\{\cos\theta > c\}$, $c \in (-1, 1)$ and θ be distributed as in (1.1). Then $g(\rho)$ is an increasing function in $\rho \in (0, h(\alpha))$, where $h(\alpha) = \exp[-\alpha \in 2/(2^\alpha - 1)]$, $\alpha \in [0, 2]$.*

Proof: We know

$$\{\cos\theta > c\} \equiv \{0 < \theta < \arccos c\} \cup \{2\pi - \arccos c < \theta < 2\pi\},$$

for $\arccos c \in (0, \pi)$. Furthermore, because of the unimodality and the symmetry of $f(\cdot)$, $g(\rho)$ can be rewritten as

$$\begin{aligned} g(\rho) &= 2P_\rho\{0 < \theta < \cos^{-1} c\} \\ &= \pi^{-1}\left\{\cos^{-1} c + 2\sum_{n=1}^{\infty}(\rho^{n^\alpha}/n)\sin n(\arccos c)\right\} \end{aligned}$$

for $0 < \arccos c < \pi$. Since the power series above is absolutely convergent, we can carry over the differentiation inside the summation. Hence,

$$g'(\rho) = (2/\pi\rho)\sum_{n=1}^{\infty} \rho^{n^\alpha} n^{\alpha-1} \sin n(\arccos c). \tag{6.5.2}$$

To show $g'(\rho) > 0$, it is sufficient to satisfy all the conditions of Lemma 6.3 with $b_n = \rho^{n^\alpha} n^{\alpha-1}$:

(i) By definition, $\arccos c \in (0, \pi)$;

(ii) $\lim_{n \longrightarrow \infty} nb^n = \lim_{t \longrightarrow \infty} t\rho^t (t = n^\alpha, \alpha > 0) = 0$, by L' Hospital's rule, since $0 < \rho < 1$;

(iii)

$$nb_n - (n+1)b_{(n+1)} > 0 \Leftrightarrow \rho^{(n+1)^\alpha - n^\alpha} < (n/(n+1))^\alpha. \qquad (6.5.3)$$

Since $[n/(n+1)]^\alpha \uparrow n$,

$$\inf \{[n/(n+1)]^\alpha : n = 1, 2, \ldots\} = 2^{-\alpha}.$$

And since $\rho^{(n+1)^\alpha - n^\alpha} \downarrow n$,

$$\sup_n \left\{\rho^{(n+1)^\alpha - n^\alpha} : n = 1, 2, \ldots\right\} = \rho^{2^\alpha - 1}.$$

Thus in Equation (6.5.3) we see that the supremum on the left-hand side is smaller than the infimum on the right-hand side when $\rho^{2^\alpha - 1} < 2^{-\alpha} \Leftrightarrow 0 < \rho < \exp[-\alpha \in 2/(2^\alpha - 1)] = h(\alpha)$. □

We next consider WN and WC separately.

Lemma 6.5 *Let θ follow a wrapped normal distribution, WN$(0, \rho)$. Then the power function based on one observation is a globally increasing function.*

Proof: The pdf of WN$(0, \rho)$ is $f(\theta; \rho, 2, 0)$ which has a unique mode at π and hence for $\theta \in (0, \pi)$

$$f'(\theta, \cdot) = -(1/\pi) \sum_{p=1}^{\infty} \rho^{p^2} n \sin(p\theta) < 0.$$

Then letting $\alpha = 2$ in (6.5.2) where $\arccos(c) \in (0, \pi)$, yields $g'(\rho) > 0$. □

Lemma 6.6 *Let θ follow a wrapped Cauchy distribution $WC(0, \rho)$. Then the power function based on one observation is a globally increasing function.*

Proof: Letting $\alpha = 1$ in (6.5.2), we have

$$g'(\rho) = (2/\pi\rho)\sum_{p=1}^{\infty} \rho^p \sin p(\arccos c), \text{ where } \arccos c \in (0, \pi).$$

Now, in order to show that $g'(\rho)$ is positive, we consider first the following series

$$\sum_{p=1}^{\infty}(\rho e^{i\theta})^p = \sum_{p=1}^{\infty} \rho^p \cos(p\theta) + i\sum_{p=1}^{\infty} \rho^p \sin(p\theta).$$

Then,

$$g'(\rho) = (2/\pi\rho)\mathrm{Im}\left(\sum_{p=1}^{\infty}(\rho e^{i\theta})^p\right), \text{ with } \theta = \cos^{-1} c.$$

Again,

$$\sum_{p=1}^{\infty}\left(\rho e^{i\theta}\right)^p = \left(\rho\cos\theta - \rho^2 + i\rho\sin\theta\right)/[1 - 2\rho\cos\theta + \rho^2].$$

Hence $g'(\rho) > 0$. □

Proof of Theorem 6.1: For $\alpha \in (0, 2)$, monotonicity of $\beta(\rho)$ over $\rho \in (0, h(\alpha))$ is established by the technique of mathematical induction on m. For $m = 1$, the claim is true by Lemma 6.4. For $m = 2$, the power function is, $\beta(\rho) = P_\rho(\cos\theta_1 + \cos\theta_2 > c)$. Let $X = \cos\theta_1$ and $Y = \cos\theta_2$. Each of X and Y has an absolutely continuous distribution function and has stochastic ordering property by Lemma 6.4. Therefore, in view of Lemma 6.2, it follows that, $X + Y = \cos\theta_1 + \cos\theta_2$ also has stochastic ordering property, i.e.,

$$\rho_1 < \rho_2 \Rightarrow P_{\rho_1}\{\cos\theta_1 + \cos\theta_2 > c\} < P_{\rho_2}\{\cos\theta_1 + \cos\theta_2 > c\}, \forall c \in R.$$

Hence, $\beta(\rho) \uparrow$ for $m = 2$ also.

Now assume that claim holds for $m = k$. Let

$$X = \cos\theta_1 + \cdots + \cos\theta_k \text{ and } Y = \cos\theta_{k+1}.$$

It follows from our assertion that X, and from Lemma 6.4 that Y, both have stochastic ordering. Then by using Lemma 6.2, the claim is true for $m = k + 1$. Hence, by induction, the lemma is true for all m.

Moreover, for the wrapped Cauchy and the wrapped normal distributions, in view of Lemmas 6.5 and 6.6 and arguing exactly as above, it follows that the monotone increasing property of $\beta(\rho)$ holds for all $\rho \in (0,1)$. Finally, since $h(\alpha) > 0 \ \forall \ \alpha \in (0,2)$, it follows that for all $\alpha \in (0,2)$, the LMP test for uniformity against WS family is (at least) locally unbiased. $\square$

Chapter 7

Nonparametric Testing Procedures

7.1 Introduction

In many practical situations, known parametric models like the von Mises or WS densities may not provide an adequate description of the data or the distributional information may be imprecise or altogether lacking. The search for methods which are robust for a large class of possible models, leads naturally to nonparametric or model-free inference. In linear inference, there are a number of considerations based on which one can justify an assumption of normality as for example when one is dealing with averages or when the samples are large enough. Unfortunately, there is no corresponding rationale for invoking the CN or WS models and thus the need for nonparametric methods is even stronger in directional data. This chapter is divided into 3 main sections dealing with the one-sample, two-sample and multi-sample tests.

7.2 One-Sample Problem and Goodness-of-fit

Let $\alpha_1, \ldots, \alpha_n$ be i.i.d. random variables with distribution function $F(\alpha)$ on $0 \leq \alpha < 2\pi$. As on the real line, one basic question that arises is the validity of the assumed model for the given data i.e., the so-called goodness of fit problem. We wish to test if the data is from a specified distribution function $F_0(\alpha)$ i.e., the null hypothesis

$$H_0 : F(\alpha) = F_0(\alpha),$$

where $F_0(\alpha)$ is a specified distribution function.

It is helpful to recall that in the linear case, apart from ad-hoc tests specific to given models, the goodness-of-fit tests can be broadly classified into three categories viz.

- those based on the χ^2 statistic,
- those that use the empirical distribution function (edf) and
- those based on the gaps between successive points or "spacings".

Since the χ^2 tests depend on how the cells are chosen, which in turn will depend on the choice of origin, they are not immediately applicable to circular data. Invariant versions of such chi-square tests are considered in Ajne (1968) and Rao (1972b). Similarly the class of tests based on the empirical distributions, like the Komogorov-Smirnov or the Cramer-von Mises tests, are also not directly applicable to circular data since their values again depend on the choice of the origin. We discuss invariant versions of such tests which are due to Kuiper (1960) and Watson (1961). Spacings tests are the only general class of goodness-of-fit tests that are directly applicable to circular data as much as they are for linear data. Indeed spacings form the maximal invariant statistic under changes in origin so that every rotationally invariant statistic that is useful for the circular context, can be expressed in terms of spacings. See Jammalamadaka (1984) for a survey article on nonparametric methods for directional data.

7.2.1 Tests Based on Empirical Distribution Functions

Recall that as in the linear case, if $\alpha_1, \ldots, \alpha_n$ are iid from $F(\alpha)$, the goodness-of-fit null hypothesis $H_0 : F(\alpha) = F_0(\alpha)$ when F_0 is continuous, can be translated into one of testing uniformity i.e. if $u_i = F_0(\alpha_i)$, then H_0 holds if and only if $u_1, \ldots, u_n$ are from the circular uniform distribution.

If $\alpha_{(1)} \leq \cdots \leq \alpha_{(n)}$ denote the ordered observations (with any starting point and any sense of rotation) corresponding to $\alpha_1, \ldots, \alpha_n$, the empirical distribution function (edf) is defined by

$$F_n(\alpha) = \begin{cases} 0 & \text{if } \alpha < \alpha_{(1)}, \\ i/n & \text{if } \alpha_{(i)} \le \alpha < \alpha_{(i+1)}, \\ 1 & \text{if } \alpha \ge \alpha_{(n)}. \end{cases} \tag{7.2.1}$$

One can define an edf for the circular case analogously with respect to any arbitrary origin, but the values taken by such an edf will depend on the choice of this origin as well as whether clockwise or anti-clockwise is taken as the positive direction. Thus in order to be able to use these for circular data problems, statistics such as the Kolmogorov-Smirnov, Cramer-von Mises should be modified so as to make them rotation invariant. It turns out that it is possible to make such invariant modifications of these classical tests.

Recall the one-sided Kolmogorov-Smirnov statistics given by

$$\begin{aligned} D_n^+ &= \sqrt{n}\sup_{\alpha}(F_n(\alpha) - F(\alpha)), \\ D_n^- &= \sqrt{n}\sup_{\alpha}(F(\alpha) - F_n(\alpha)) \\ &= -\sqrt{n}\inf_{\alpha}(F_n(\alpha) - F(\alpha)). \end{aligned}$$

The more common two-sided Kolmogorov-Smirnov statistic D_n can then be written as

$$D_n = \sqrt{n}\sup_{\alpha}|F_n(\alpha) - F(\alpha)| = \max(D_n^+,\ D_n^-).$$

Noticing that D_n^+ gains (or loses) just as much as D_n^- loses (or gains) due to a rotation, Kuiper (1960) suggested the statistic

$$V_n = \left(D_n^+ + D_n^-\right). \tag{7.2.2}$$

A computational form for V_n is given by

$$V_n = \sqrt{n}\left\{\max_{1\le i\le n}(U_{(i)} - \frac{i-1}{n}) + \max_{1\le i\le n}(\frac{i}{n} - U_{(i)})\right\},$$

where $U_{(i)} = F(\alpha_{(i)})$ in terms of the order statistics $\alpha_{(1)} \le \alpha_{(2)} \le \cdots \le \alpha_{(n)}$.

Proposition 7.1 *V_n is invariant under the change of origin.*

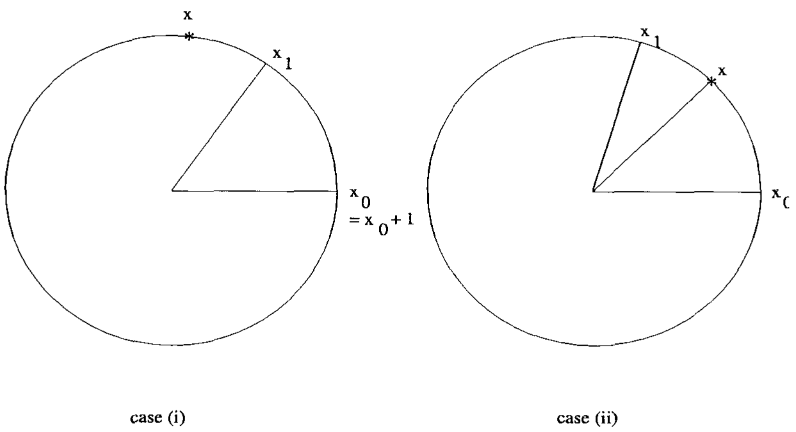


Figure 7.1: Value of the edf depends on choice of origin and sense of rotation.

Proof:

Assume for simplicity that the circle is of unit circumference by scaling $\alpha/2\pi$ as x. Let $G(x, x_0)$ and $G(x, x_1)$ denote distribution functions which are cumulated from the origin x_0 and a different origin x_1, respectively. Then from Figure 7.1, it is easy to see that

$$G(x, x_0) = \begin{cases} G(x, x_1) + G(x_1, x_0) & \text{if } x_1 \leq x \leq x_0 + 1 \text{ (case i)}, \\ G(x, x_1) + G(x_1, x_0) - 1 & \text{if } x_0 \leq x \leq x_1 \text{ (case ii)}. \end{cases} \tag{7.2.3}$$

If H denotes another distribution function, similar relations hold for H. Now, when x_0 is used as origin, let

$$D_n^+(0) = \sup_x [G(x, x_0) - H(x, x_0)]$$

and

$$D_n^-(0) = \sup_x [H(x, x_0) - G(x, x_0)] .$$

Similarly, when x_1 is used as the origin

$$D_n^+(1) = \sup_x [G(x, x_1) - H(x, x_1)]$$

and

$$D_n^-(1) = \sup_x [H(x, x_1) - G(x, x_1)] .$$

Denoting the two intervals by $I_1 = [x_1, x_0 + 1]$ and $I_2 = [x_0, x_1]$ (see Figure 7.1), we have

$$\begin{aligned}
D_n^+(0) &= \max\left\{\sup_{x\in I_1}(G(x,x_0) - H(x,x_0)), \sup_{x\in I_2}(G(x,x_0) - H(x,x_0))\right\} \\
&= \max\left[\sup_{x\in I_1}\{G(x,x_1) + G(x_1,x_0) - H(x,x_1) - H(x_1,x_0)\},\right. \\
&\quad \left.\sup_{x\in I_2}\{G(x,x_1) + G(x_1,x_0) - H(x,x_1) - H(x_1,x_0)\}\right] \\
&= \sup_{0\le x\le 1}[G(x,x_1) - H(x,x_1)] + \{G(x_1,x_0) - H(x_1,x_0)\} \\
&= D_n^+(1) + \delta, \qquad (7.2.4)
\end{aligned}$$

where $\delta = G(x_1, x_0) - H(x_1, x_0)$. That is when the origin is changed from x_0 to x_1, the value of D_n^+ changes by this amount δ. Similarly,

$$D_n^-(0) = D_n^-(1) - \delta. \qquad (7.2.5)$$

Adding the two equations (7.2.4) and (7.2.5), we get

$$D_n^+(0) + D_n^-(0) = D_n^+(1) + D_n^-(1)$$

which shows that $(D_n^+ + D_n^-)$ has the same value whether x_0 or x_1 is the origin i.e., the sum is invariant under changes in origin. In particular taking $G = F_n$ and $H = F_0$, we see that the Kuiper's statistic V_n is rotation invariant and hence useful as a test for circular data. □

The asymptotic distribution of V_n (see Kuiper (1960)) is given by

$$\begin{aligned}
P[V_n \ge v] &= \sum_{m=1}^{\infty} 2\left(4m^2v^2 - 1\right)e^{-2m^2v^2} \\
&\quad -\frac{8v}{3\sqrt{n}}\sum_{m=1}^{\infty} m^2\left(4m^2v^2 - 3\right)e^{-2m^2v^2} + o\left(\frac{1}{n}\right). \qquad (7.2.6)
\end{aligned}$$

Another classical alternative edf-based test for goodness-of-fit on the real line is provided by the Cramer-von Mises test, given by

$$C_n^2 = n\int_{-\infty}^{\infty}(F_n - F)^2\,dF.$$

Watson (1961) provided an invariant modification of this that is suitable for circular data. Watson's statistic is defined by

$$W_n^2 = \int_0^{2\pi} \left[(F_n - F) - \int_0^{2\pi} (F_n - F)\, dF \right]^2 dF \tag{7.2.7}$$

and can be written in the computational form

$$W_n^2 = \sum_{i=1}^{n} \left[\left(U_{(i)} - \frac{i-\frac{1}{2}}{n} \right) - \left(\overline{U} - \frac{1}{2} \right) \right]^2 + \frac{1}{12n}, \tag{7.2.8}$$

where $U_i = F(\alpha_i)$. Note that if the Cramer-von Mises statistic can be thought of as the "second moment" of $(F_n - F)$, Watson's statistic is similar to the expression for "variance". So that if the quantity $(F_n - F)$ changes by a constant δ due to a change of origin, as we have seen before in Equation (7.2.4), the variance will not change. This is formally proved in the next

Proposition 7.2 *W_n^2 is invariant under rotations i.e., shift of the zero direction.*

Proof: Let G and H be two cdf's as before. Then, using the notations as in the proof of Proposition 7.1 and from Equation (7.2.3),

$$\begin{aligned}
\int_0^1 G(x, x_0) dF(x) &= \left(\int_{x_1}^{x_0+1} + \int_{x_0+1}^{x_1+1} \right) G(x, x_0) dF(x) \\
&= \int_{x_1}^{x_0+1} \{ G(x, x_1) + G(x_1, x_0) \} \, dF(x) \\
&\quad + \int_{x_0+1}^{x_1+1} \{ G(x, x_1) + G(x_1, x_0) - 1 \} \, dF(x) \\
&= \int_0^1 G(x, x_1) dF(x) + \int_0^1 G(x_1, x_0) dF(x) \\
&\quad - \int_{x_0}^{x_1} dF(x) \\
&= \int_0^1 G(x, x_1) dF(x) + G(x_1, x_0) - \int_{x_0}^{x_1} dF(x).
\end{aligned} \tag{7.2.9}$$

A similar expression can be written for the distribution function H. Differencing the expressions and rearranging terms, we get

$$\begin{aligned} & \{G(x,x_0) - H(x,x_0)\} - \int \{G(x,x_0) - H(x,x_0)\}\, dF(x) \\ = \; & \{G(x,x_1) - H(x,x_1)\} - \int \{G(x,x_1) - H(x,x_1)\}\, dF(x). \end{aligned}$$

Again, taking $G = F_n$ and $H = F$, we see that the integrand of W_n^2 in (7.2.7) is independent of the rotational change and the proposition follows. □

The asymptotic distribution of W_n^2 is given by (Watson (1961))

$$\lim_{n \longrightarrow \infty} P\left[W_n^2 \geq w\right] = 2 \sum_{m=1}^{\infty} (-1)^{m-1} e^{-2m^2\pi^2 w}. \tag{7.2.10}$$

Critical values for the U_n^2 statistic are available in Lockhart and Stephens (1985). It has been shown that Kuiper's test V_n is more efficient in the sense of Bahadur efficiency than the Kolmogorov-Smirnov test D_n in situations where both are applicable as on the real line (Abrahamson (1967)).

7.2.2 χ^2 and Other Tests

Given n observations on the circumference of the circle, define

$$N(\alpha) = \text{ number of observations in the half-circle } [\alpha,\ \alpha + \pi).$$

Ajne (1968) suggested the statistic

$$N = \sup_{0 \leq \alpha < 2\pi} N(\alpha), \tag{7.2.11}$$

where clearly $0 \leq N \leq n$. Large values of N indicate clustering in one half of the circle and will lead us to reject the null hypothesis of uniformity. It has been pointed out by Rao (1969) and Bhattacharyya and Johnson (1969b) that this test is identical to the Hodges (1955) Bivariate Sign test. The same statistic was also considered by Daniels (1954) in connection with a regression problem.

The exact and asymptotic distribution of N, called the Ajne's test, is known and discussed below.

Exact Distribution of N

Corresponding to the n original observations denoted by crosses (x), denote the point on the diametrically opposite side of any 'x' by $\odot$. See Figure 7.2.

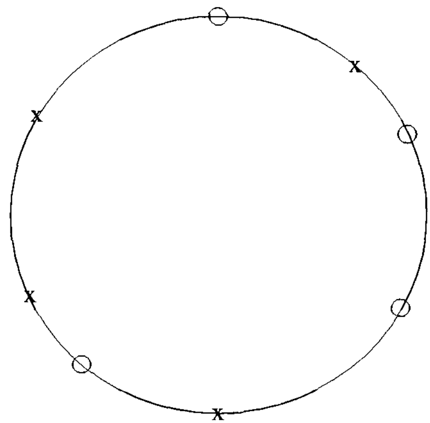


Figure 7.2: Configuration of original points x and the points $\odot$ diametrically opposite to them.

Now, with these additional $\odot$'s, there are $2n$ points on the circumference with every semicircle containing n points – some of these x's and the rest $\odot$'s. Define, for $i = 1, \ldots, n$,

$$X_i = \begin{cases} 1 & \text{if the symbol in the } i\text{th position of the first } n \text{ tuple is } \odot, \\ -1 & \text{if the symbol in the } i\text{th position of the first } n \text{ tuple is x.} \end{cases}$$

Under the hypothesis of uniformity, there are 2^n equally likely configurations.

$$P\left[X_i = +1\right] = P\left[X_i = -1\right] = \frac{1}{2} \qquad \forall i = 1, \ldots, n.$$

If K_i = number of x's in the ith configuration, $i = 1, \ldots, 2n$, then clearly

$$N = \sup_{\alpha} N(\alpha) = \max\left(K_1, \ldots, K_{2n}\right).$$

For $i = 1, \ldots, 2n$, observe that

$$K_{i+1} = K_i + X_i = K_1 + S_i,$$

where $S_i = X_1 + \cdots + X_i$ and $K_i + K_{n+i} = n$, as they together cover the whole circumference. Thus,

$$\begin{aligned} K_1 + K_{n+1} &= n \\ \Rightarrow K_{n+1} &= n - K_1 \\ &= K_1 + S_n \\ \Rightarrow K_1 &= (n - S_n)/2 \\ \Rightarrow K_{i+1} &= K_1 + S_i \\ &= \frac{n - S_n}{2} + S_i. \end{aligned}$$

For $K \geq [n/2] + 1$,

$$\begin{aligned} & P[N \leq K] \\ = \; & P\left[K_1, \ldots, K_{2n} \leq K\right] \\ = \; & P\left[n - K \leq \frac{n - s_n}{2}, \frac{n - S_n}{2} + S_1, \ldots, \frac{n - S_n}{2} + S_{n-1} \leq K\right] \\ = \; & P\left[n - K - \frac{n - S_n}{2} \leq 0, S_1, \ldots, S_{n-1} \leq k - \frac{n - S_n}{2}\right] \\ = \; & \sum_{s=n-2K}^{2K-n} P\left[S_n = s, n - K - \frac{n-s}{2} \leq S_1, \ldots, S_{n-1} \leq K - \frac{n-s}{2}\right]. \end{aligned}$$

Putting $v = (s + n)/2$, this probability can be written as

$$\sum_{v=n-K}^{K} P\left[S_n = 2v - n, \;\; v - K - 1 < S_1, \ldots, S_{n-1} < v - n + K + 1\right].$$

Now we apply the following classical result on random walks (see Feller (1968)).

Theorem 7.1 *Let $X_1, \ldots, X_n$ be iid with $P[X_i = 1] = p$ and $P[X_i = -1] = q = (1 - p)$. Define $S_i = \sum_{j=1}^{i} X_j$. Then, for $a > 0, \;\; b < 0$*

$$P\left[S_n = a + b, \quad b < S_1, \ldots, S_{n-1} < a\right]$$

is given by

$$\left[\binom{n}{\frac{a+b+n}{2}} \sum_{j=1}^{\infty} \binom{n}{c_j + a} + \sum_{j=1}^{\infty} \binom{n}{c_j - b} - 2 \sum_{j=1}^{\infty} \binom{n}{c_j}\right] p^{\frac{n+a+b}{2}} q^{\frac{n-a-b}{2}},$$

where $c_j = [(2j - 1)(a - b) + n]/2$.

From this we obtain

Theorem 7.2 *For $K \geq [n/2]+1$,*

$$\begin{aligned} P[N \geq k] &= \frac{2k-n}{2^{n+1}} \sum_{j=0}^{\infty} \binom{n}{j(2k-n)+k} \\ &= \frac{2k-n}{2} \sum_{j=0}^{n} b\left(k+j(2k-n); n, \frac{1}{2}\right). \end{aligned}$$

In particular if $k > 2n/3$,

$$P\left[N \geq k\right] = \frac{2k-n}{2^{n+1}} \binom{n}{k}.$$

From this one can also obtain the large sample result,

Theorem 7.3 *Let, for $c > 0$,*

$$N^* = \frac{2}{\sqrt{n}}\left(N - \frac{n}{2}\right).$$

Then,

$$\begin{aligned} \lim_{n \longrightarrow \infty} P\left[N^* \geq c\right] &= 4c \sum_{j=0}^{\infty} \phi\left(\overline{2j+1}c\right) \\ &= 2c\sqrt{2}\pi \sum_{j=0}^{\infty} e^{-(2j+1)^2 c^2/2}. \end{aligned}$$

An alternative test using $N(\alpha)$ is based on

$$A_n = \frac{1}{2\pi n} \int_0^{2\pi} \left(N(\alpha) - \frac{n}{2}\right)^2 d\alpha \tag{7.2.12}$$

since $n/2$ is the expected value in any half circle under uniformity. Since it is averaged over all possible α, it is easy to see that this statistic is invariant under rotations.

In terms of the "circular distance" between ith and jth observations, defined in (1.3.6), namely

$$d_0(x_i, x_j) = \begin{cases} (x_j - x_i) & \text{if } x_i < x_j \leq x_i + \pi, \\ 2\pi - (x_j - x_i) & \text{if } x_i + \pi < x_j \leq x_i, \end{cases}$$

A_n can be rewritten as the sum

$$
\begin{aligned}
A_n &= \frac{1}{2} - \frac{n}{4} + \frac{1}{n\pi} \sum_{i=1}^{n} \sum_{j=i+1}^{n} |\pi + x_j - x_i| \\
&= \frac{n}{4} - \frac{1}{n\pi} \sum_{i} \sum_{j} d_0(x_i, x_j) \,. \qquad (7.2.13)
\end{aligned}
$$

The asymptotic distribution of A_n was obtained by Watson (1967) and is given by

$$\lim_{n \longrightarrow \infty} P[A_n > a] = \sum_{m=1}^{\infty} \frac{4(-1)^{m-1}}{\pi(2m-1)} e^{-\frac{\pi^2 (2m-1)^2 a}{2}} . \qquad (7.2.14)$$

The statistic A_n is equivalent to a χ^2 statistic using the two cells $[\alpha, \alpha + \pi)$ and $[\alpha + \pi, \alpha + 2\pi)$ and averaging such a statistic over all starting points α. Rao (1972b) discusses dividing the circle into m equal arcs, counting the observed frequencies in each to form a chi-square statistic and averaging over it over all starting points α. One can also consider the maximum of such a chi-square statistic over all possible choices of the arbitrary origin but its distribution is difficult to obtain, even asymptotically. Another type of generalization is obtained by considering the "scanning window" to be an arc of arbitrary length t rather than half-circle. Rothman (1972) considers such a generalization. It is worth noting that the statistic N is a special case of a class of statistics known as "Scan statistics". See Glaz and Balakrishnan (1999) for more recent references on this topic.

7.2.3 Tests Based on Sample Arc-Lengths

Let $\alpha_1, \ldots, \alpha_n$ be any iid observations on a circle. In this section, we will take this circle to be of *unit circumference* for convenience. Instead, if the circle is taken to be of unit radius, one gets a circumference of length 2π. As pointed out before, testing the null hypothesis that these observations are uniformly distributed with no preferred direction, is of special importance in circular case. Order these observations with respect to the given zero direction and the sense of rotation, to get

$$0 \leq \alpha'_1 \leq \cdots \leq \alpha'_n \leq 1.$$

Define

$$D_i = \left(\alpha'_i - \alpha'_{i-1}\right), \qquad i = 1, \ldots, n, \tag{7.2.15}$$

where we take $\alpha'_0 = \alpha'_n - 1$. The unusual definition of $\alpha'_0 = \alpha'_n - 1$, makes D_1 the "natural" gap between the first and last ordered values that straddle the origin. Clearly

$$D_i \geq 0 \text{ and } \sum_{i=1}^{n} D_i = 1.$$

Under the null hypothesis of uniformity, $\{D_1, \ldots, D_n\}$ form a set of exchangeable random variables, i.e., for any k the density of $(D_{i_1}, \ldots, D_{i_k})$ is the same for any subset of indices $(i_1, \ldots, i_k)$. Now, under uniformity, $E(D_i) = 1/n$ so that from asymptotic considerations, we will refer to $\{nD_i\}$ as normalized spacings.

Rao (1969) defines

$$R = (1 - \max D_i),$$

which corresponds to the shortest arc on the circumference which contains all the n observations, as the "Circular Range" and provides its exact distribution. Clearly, a small value of the "Circular Range" indicates clustering and leads to the rejection of the hypothesis of uniformity. Its cdf is given by

$$F(r) = \sum_{k=1}^{\infty} (-1)^{k-1} \binom{n}{k} \left[(kr - (k-1))^{+}\right]^{n-1} \tag{7.2.16}$$

with the notation $x^{+} = x$ if $x > 0$ and is 0 otherwise. This is related to the distribution of the largest amplitude in harmonic analysis discussed by Fisher (1929). This test can be implemented using the routine `circ.range` in `CircStats` as we do with the von Mises data set on atomic weights.

Example 7.1 *Recall the data set in Table 1.2 from which von Mises was interested in testing if the atomic weights are whole integers i.e. if the preferred direction of their fractional parts is zero. Although this is not the most appropriate test, we apply the circular range test to see if the data might be considered as being uniform with no preferred direction.*

```
> x_c(rep(0,12),3.6,rep(36,6),72,108,108,169.2,324)

> x.rad_rad(x)
```

```
> x.rad

 [1] 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000

 [7] 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000

[13] 0.0628318 0.6283185 0.6283185 0.6283185 0.6283185 0.6283185

[19] 0.6283185 1.2566371 1.8849556 1.8849556 2.9530971 5.6548668


>  circ.range(x,T)

     range    p.value

1 4.584073 0.01701148
```

Based on this we have reason to reject uniform distribution as the appropriate model for this data. See also Example 7.2 later on for another illustration of the circular range.

More generally, one may consider the family of spacings statistics of the form

$$T_n = \frac{1}{n} \sum_{i=1}^{n} h\left(nD_i\right),$$

where h is a suitably chosen function. Various meaningful choices of h can be considered and for most such functions $h(\cdot)$, it has been shown (see for instance Pyke (1965), Sethuraman and Rao (1970) and Rao and Sethuraman (1975)) that T_n has an asymptotic normal distribution. For instance, the choice $h(x) = x^r$, $r > 0$, $r \neq 1$, corresponds to the test statistic

$$V_n(r) = \frac{1}{n} \sum \left(nD_i\right)^r .$$

We have

$$\sqrt{n}\left[V_n(r) - \Gamma(n+1)\right] \stackrel{d}{\Longrightarrow} N\left(0,\ \Gamma(2r+1) - (r^2+1)(\Gamma(r+1))^2\right).$$

For $r = 2$, $V_n(2)$ is known as the *"Greenwood Statistic"* and for large n, has the $N(2,\ 4/n)$ distribution. Taking $h(x) = \log x$, gives

$$L_n = \frac{1}{n}\sum_{i=1}^{n} \log(nD_i)$$

which has an asymptotic normal distribution with mean $-\gamma$ and variance $(\pi^2/6 - 1)/n$, where γ is the Euler Constant $(= 0.5772\cdots)$.

Another interesting choice is $h(x) = \frac{1}{2}|x - 1|$, so that we get the statistic

$$U_n = \frac{1}{2}\sum |D_i - 1/n| = \sum \max(D_i - 1/n, 0)\,. \tag{7.2.17}$$

In view of its being the sum of $\max(D_i - i/n,\ 0)$, it can be interpreted as the uncovered part of the circumference when n arcs of length $1/n$ are placed starting with each of the n observed points on the circle (see Rao (1969)). Clearly all the circumference will be covered resulting in $U_n = 0$ only when all the observations are equi-spaced. This test, sometimes referred to as *"Rao's Spacings test"* was introduced in the context of circular uniformity by Rao (1969) and is one of the few spacings statistics for which both its exact as well as asymptotic distributions are known. This test has been noted to have good power properties even when testing uniformity among data that is not unimodal. Adapted to the circular case where the circle has circumference 2π, the density function of U_n can be shown to be (cf. Rao (1976))

$$f_{U_n}(u) = (n-1)!\sum_{j=1}^{n-1}\binom{n}{j}\left(\frac{u}{2\pi}\right)^{n-j-1}\frac{g_j(nu)}{(n-j-1)!n^{j-1}} \tag{7.2.18}$$

for $0 \le u \le 2\pi(1 - 1/n)$, where

$$g_j(x) = \frac{1}{(j-1)!(2\pi)}\sum_{k=0}^{\infty}(-1)^k\binom{j}{k}\left[\left(\frac{x}{2\pi} - k\right)^{+}\right]^{j-1}$$

with the notation $x^+ = x$ if $x > 0$ and is 0 otherwise. An extended table of critical values of the U statistic can be found in Russell and Levitin (1995). This test can be implemented using the routine `rao.spacing` in `CircStats`. See Example 7.2.

It can be shown that for large samples,

$$\sqrt{n}\left(U_n - e^{-1}\right) \stackrel{d}{\Longrightarrow} N\left(0,\ 2e^{-1} - 5e^{-2}\right).$$

Remark 7.1 *The Asymptotic Relative Efficiency (A.R.E.) (or Pitman efficiency) of one sequence of tests against another is defined as the limit of the inverse ratio of sample sizes required for the two tests to attain the same power, at a sequence of alternatives which converge to the null hypothesis.*

Sethuraman and Rao (1970) show that among the spacing tests, $V_2(n)$ *is the most efficient in the Pitman sense. If its efficiency is taken to be one, then* $L(n)$ *and* $U(n)$ *have efficiencies of 0.38 and 0.57, respectively.*

Remark 7.2 *The Bahadur efficiency of a test sequence (see Bahadur (1960)) provides an alternate measure of test efficiency when there are several competing tests and is concerned with how fast the "attained level" or "P-value" goes to zero under the alternatives. It is concerned with how well the null hypothesis explains the sequence of test statistics when in fact the hypothesis is false. Rao (1972a) considers the Bahadur efficiencies of the different tests due to Rayleigh, Ajne, Watson, Kuiper and the spacings test* U_n*, for testing uniformity against CN alternatives. It is shown there that the limiting Bahadur efficiencies of Ajne's test A, Watson's test as well as the Rayleigh's test based on R are identical while the other tests have lower Bahadur efficiencies. However Rao (1969) shows by simulations that in small samples of the order of 10-20, the spacings test* U_n *performs reasonably well when compared to the Rayleigh test even against von Mises alternatives. It is important to note that while the spacings test(s) can be used even when the data is not unimodal, such is not the case with Rayleigh test. The nearly equal performance of Kuiper's, Watson's and Ajne's tests was also noted by Stephens (1969a) through a simulation study.*

Example 7.2 *13 homing pigeons were released singly in the Toggenburg Valley in Switzerland under sub-Alpine conditions (data quoted in Batschelet (1981)). They did not appear to have adjusted quickly to the homing direction but preferred to fly in the axis of the valley, indicating a somewhat bimodal distribution (see 7.3). The vanishing angles are given by the angles arranged in ascending order as follows: 20, 135, 145, 165, 170, 200, 300, 325, 335, 350, 350, 350 and 355. Do the birds indicate a preferred direction of flight?*

We now use the **CircStats** *package to run Watson's, Kuiper's and Rao's Spacing tests of uniformity.*

```
>pigeon
```

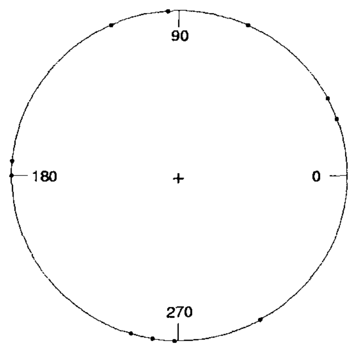


Figure 7.3: Circular plot for the Homing Pigeon data.

```
 [1]  20 135 145 165 170 200 300 325 335 350 350 350 355

>pigeon.rad_pigeon*pi/180
>pigeon.rad

 [1] 0.3490659 2.3561945 2.5307274 2.8797933 2.9670597 3.4906585
 [7] 5.2359878 5.6723201 5.8468530 6.1086524 6.1086524 6.1086524
 [13] 6.1959188

>circ.summary(pigeon.rad)

   n       rho   mean.dir
1 13 0.2156368 -0.2734986

> circ.range(pigeon.rad,T)
     range p.value
1 4.276057 0.12794

>rao.spacing(pigeon,alpha=0,rad=F)
```

```
        Rao's Spacing Test of Uniformity

Test Statistic = 161.92308 0.05 < P-value < 0.10

>watson(pigeon.rad,alpha=0,dist='uniform')

        Watson's Test for Circular Uniformity

Test Statistic: 0.137 P-value > 0.10

>kuiper(pigeon.rad,alpha=0)

        Kuiper's Test of Uniformity

Test Statistic: 1.5047 P-value > 0.15
```

All the tests indicate that there is not conclusive evidence to reject the hypothesis of uniformity and that the pigeons do not seem to have a preferred direction.

7.3 Two-Sample Problems

The general two-sample problem is to decide if two given circular data sets are from the same population or if they come from two different circular distributions. As in one sample case, one can proceed to modify and adapt the χ^2 goodness-of-fit tests as well as the empirical distribution function procedures, to two-sample problems.

7.3.1 Two-Sample Tests based on Edf's

Suppose there is a sample of n_1 observations say from F_1 and a second sample of n_2 observations say from F_2 with a total of $n = n_1 + n_2$ observations. For testing

$$H_0 : F_1 = F_2,$$

the two-sample analogue of the Kuiper test V_n is given by

$$V_{n_1,n_2} = \sqrt{\frac{n}{n_1 n_2}} \sup_x (F_{n_1}(x) - F_{n_2}(x)) + \sup_x (F_{n_2}(x) - F_{n_1}(x)) \qquad (7.3.1)$$

and the 2-sample analog of Watson statistic W^2 is given by

$$W^2_{n_1,n_2} = \frac{n_1 n_2}{n}\left[\int \left[F_{n_1}(x) - F_{n_2}(x) - \int (F_{n_1} - F_{n_2}) dF^*_n\right]^2 dF^*_n\right], \quad (7.3.2)$$

where $F^*_n(x) = (n_1 F_1 + n_2 F_2)/n$. It can be shown that if

$$n_1, n_2 \longrightarrow \infty,\ n_1/n_2 \longrightarrow \lambda,\ 0 < \lambda < 1,$$

the asymptotic distributions of these two-sample tests are the same as their one-sample counterparts given in (7.2.6) and (7.2.10), respectively. For instance,

$$\lim_{n_1,n_2\longrightarrow\infty} P\left[W^2_{n_1,n_2} > v\right] = 2\sum_{m=1}^{\infty}(-1)^{m-1}e^{-2m^2\pi^2 v}.$$

We apply Watson's two-sample test to the data in Example 7.2, to see if the experimental group and control group differ. This is done by using the program `watson.two` of the `CircStats` package:

```
>control

[1] 75 75 80 80 80 95 130 170 210

>experimental

[1] 10 50 55 55 65 90 285 285 325 355

>control.rad_control*pi/180
>control.rad

[1] 1.308997 1.308997 1.396263 1.396263 1.396263 1.658063
[7] 2.268928 2.967060 3.665191

>experimental.rad_experimental*pi/180
>experimental.rad

 [1] 0.1745329 0.8726646 0.9599311 0.9599311 1.1344640 1.5707963
 [7] 4.9741884 4.9741884 5.6723201 6.1959188
```

```
>circ.summary(control.rad)

  n       rho mean.dir
1 9 0.7230602 1.810058

>circ.summary(experimental.rad)

   n       rho  mean.dir
1 10 0.5844946 0.3295853

>watson.two(control.rad,experimental.rad,alpha=0)

       Watson's Two-Sample Test of Homogeneity

Test Statistic: 0.2982 0.001 < P-value < 0.01
```

7.3.2 Wheeler and Watson Test

Consider two independent samples of sizes n_1 and n_2, with $n = n_1 + n_2$ denoting the total number of observations. Denote observations from the first sample by x and those from the second sample by $\odot$ as in Figure 7.4. If F_1, F_2 denote the distribution functions of the first and second population, respectively, we wish to test the usual two-sample problem that $H_0 : F_1 = F_2$ for which Wheeler and Watson (1964) suggest the following procedure.

Adjust the observations in such a way that they become equidistant i.e., spacing between any two observations is $2\pi/n$. Because of being uniformly spread, the overall resultant vector $\boldsymbol{R} = \boldsymbol{0}$. However, since $\boldsymbol{R} = \boldsymbol{R}_1 + \boldsymbol{R}_2$, where $\boldsymbol{R}_i$ is the length of the resultant of the ith sample ($i = 1, 2$),

$$\boldsymbol{R}_1 = -\boldsymbol{R}_2.$$

Under the hypothesis that $F_1 = F_2$, the two samples should be well-mixed with each of the individual samples also being evenly spread over the circumference. This makes their resultant lengths small. For instance, if R_1 denotes the length of first sample resultant, the test procedure is to

$$\text{Reject} \quad H_0 \ \text{ if } \ R_1 \quad \text{is large.}$$

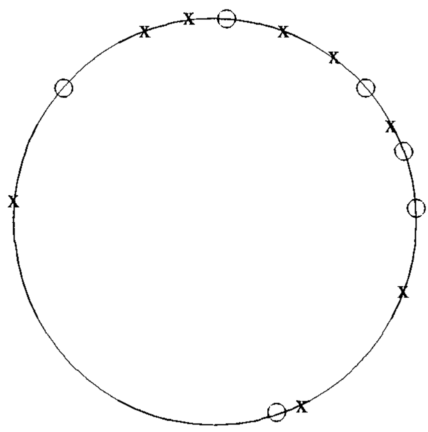


Figure 7.4: Observations from 2 samples.

Note that this test is equivalent to applying the Rayleigh test based on resultant length, to the so-called "uniform scores"

$$u_i = \frac{2\pi r_i}{n},$$

where $r_1, \ldots, r_{n_1}$ are the ranks of the first sample observations in the combined sample. Thus in large samples, the null distribution of the statistic $(2R_1^2/n_1)$ can be approximated by a χ_2^2, as discussed in (3.4.1).

7.3.3 Tests based on Spacing-Frequencies

Just as the spacings provide the maximal invariant under rotations for the one-sample problem, the so-called "spacing-frequencies," the frequencies of say the first sample observations that fall in between the gaps created by the second sample, are a maximal invariant for the two-sample case. More precisely, let $\alpha_1, \ldots, \alpha_m$ and $\beta_1, \ldots, \beta_n$ denote the two samples of sizes m and n respectively. If $\{\alpha_{(i)}\}$ denote the ordered values of the first sample (with any starting point and any sense of rotation), then we define "spacing-frequencies" as the counts

$$S_i = \text{the number of } \beta_j's \text{ that fall in } \quad [\alpha_{(i-1)}, \alpha_{(i)}) \quad i = 1, \ldots, m.$$

In Figure 7.4, these are the numbers of ⊙'s in between any two successive x's. Tests based symmetrically on these S_i's are clearly rotation invariant and the "circular run test" (cf. David and Barton (1962), pp. 132-136) and the test based on $\sum S_i^2$ suggested by Dixon (1940) are special cases of this form. Holst and Jammalamadaka (1980) consider general classes of such statistics and establish the asymptotic normality under some regularity conditions. They show in particular that the test based on $\sum S_i^2$ is asymptotically most efficient among the symmetric class of tests. Rao and Mardia (1980) provide some further power comparisons and discuss the special relevance of this class of tests for circular data. Observe that

$$S_i = r_i - r_{i-1} - 1 \quad \text{or} \quad r_i = \sum_{j=1}^{i} S_j + i$$

where r_i are the ranks defined in Section 7.3.2 so that tests based on the ranks can be expressed in terms of these spacing-frequencies.

7.4 Multi-Sample Tests

A simple extension of the idea contained in the Wheeler-Watson test to the multi-sample case leads to the following generalization. Let

$$\{\alpha_{ij} \mid j = 1, \ldots, n_i, i = 1, \ldots, q\}$$

be q independent samples, with the ith sample of size n_i coming from a circular cdf, F_i. The goal is to test if these populations are homogeneous with respect to their means and/or concentrations. One can re-arrange all the $n = n_1 + \cdots + n_q$ observations to be equi-spaced and obtain the "uniform scores" $\theta_{ij} = 2\pi r_{ij}/n$ corresponding to the n_i observations in the ith sample, where $\{r_{ij}, j = 1, \ldots, n_i\}$ are their (linear) ranks in the combined sample. If the groups are homogeneous, $\{\theta_{ij}, j = 1, \ldots, n_i\}$ are uniformly distributed and by Rayleigh's test, one expects $2R_i^2/n_i$ to be a chi-square, where

$$R_i^2 = \left(\sum \cos\theta_{ij}\right)^2 + \left(\sum \sin\theta_{ij}\right)^2.$$

Thus under the hypothesis of homogeneity, one can add these to get the statistic

$$2\sum_{i=1}^{q} R_i^2/n_i$$

which has a $\chi^2_{2(q-1)}$ distribution (see Mardia (1970)). For $q = 2$, this reduces to the Wheeler-Watson test described above.

7.4.1 Homogeneity Tests in Large Samples

One would often like tests which can specifically detect differences in the mean directions and/or differences in concentrations. Recall that the approximate ANOVA discussed in Section 5.3.1 is geared to detecting differences in mean direction but under the multiple assumptions that (i) the data is at least approximately CN, (ii) the concentration parameter is the same and (iii) that the value of this common concentration parameter is large. In many cases, one or more of these assumptions may be violated. In such situations the tests proposed in this section (see Rao (1967)) are appropriate. The only requirement to apply these homogeneity tests is that the data be unimodal so that R and $\overline{\alpha}_0$ represent the concentration and location measures and the sample size be reasonably large.

Again, let

$$\{\alpha_{ij} \mid j = 1, \ldots, n_i, i = 1, \ldots, k\}$$

be q independent samples, with the ith sample of size n_i. Let X_i and Y_i denote the means of cos and sin values for the ith sample i.e.,

$$X_i = \frac{\sum_{j=1}^{n_i} \cos\theta_{ij}}{n_i}, \qquad Y_i = \frac{\sum_{j=1}^{n_i} \sin\theta_{ij}}{n_i}.$$

A. Testing for Equality of Polar Vectors

A consistent estimator of $\tan\gamma_i$, where γ_i is the polar angle in the ith population, is provided by

$$T_i = \frac{Y_i}{X_i} \tag{7.4.1}$$

with the asymptotic estimated variance

$$s_i^2 = \frac{1}{n_i}\left(\frac{S_{ss}^{(i)}}{X_i^2} + \frac{Y_i^2 S_{cc}^{(i)}}{X_i^4} - \frac{2Y_i S_{cs}^{(i)}}{X_i^3}\right). \tag{7.4.2}$$

It can be seen from Proposition 3.1, that the asymptotic distribution of $\sqrt{n_i}(T_i - \tan\gamma_i)$ is normal and that (7.4.2) represents a consistent estimate of the true variance. Now consider the hypothesis

$$H_0: \ \tan\gamma_1 = \tan\gamma_2 = \cdots = \tan\gamma_k. \tag{7.4.3}$$

Since under the hypothesis, T_1, T_2, ..., T_k are independent and consistent estimators of the same quantity based on samples of size $n_1, n_2, ..., n_k$, respectively, we can use the "homogeneity statistic"

$$H_1 = \sum_{i=1}^{k} \frac{T_i^2}{s_i^2} - \left(\sum_{i=1}^{k} \frac{T_i}{s_i^2}\right)^2 \Big/ \left(\sum_{i=1}^{k} \frac{1}{s_i^2}\right). \tag{7.4.4}$$

This statistic measures the variability of the individual estimates from the common combined estimate. If these T_i's do not estimate the same quantity, the differences are reflected in H so that we reject the null hypothesis H_0 for large values of H. Under some very general conditions, this statistic H has a χ^2 distribution with $(k-1)$ degrees of freedom, under the hypothesis H_0 (Rao (1973), p. 323).

Rejection of H_0 leads us to conclude that the polar vectors are different. But on the other hand, even if H_0 is not rejected, it is possible that the true directions are different because our hypothesis does not distinguish between the pole and the antipole, since $\tan\gamma = \tan(\pi + \gamma)$. However such wide differences can easily be detected by a simple examination of the data, without needing a formal test. See also Yoshimura (1978). The procedure for applying this test is very simple and consists in getting the values of T_i from (7.4.1), s_i^2 from (7.4.2) and then computing H from (7.4.4). This test procedure can be implemented using `CircStats`. See Example 7.3.

B. Testing for Equality of Dispersions

Recall that the squared length of the vector resultant

$$U_i = X_i^2 + Y_i^2 \tag{7.4.5}$$

is a measure of concentration from the ith population. Its asymptotic estimated variance is given by

$$s_i^{*2} = \frac{4}{n_i}\left(X_i^2 S_{cc}^{(i)} + Y_i^2 S_{ss}^{(i)} + 2X_iY_iS_{cs}^{(i)}\right). \tag{7.4.6}$$

Under the hypothesis of equal dispersions, all these statistics U_1, U_2, ..., U_k are independent and consistent estimators of the same quantity so that we can use the test criterion H for homogeneity. Compute

$$H_2 = \sum_{i=1}^{k} \frac{U_i^2}{s_i^{*2}} - \left(\sum_{i=1}^{k} \frac{U_i}{s_i^{*2}}\right)^2 \Big/ \left(\sum_{i=1}^{k} \frac{1}{s_i^{*2}}\right), \tag{7.4.7}$$

which again is distributed as a χ^2 with $(k-1)$ degrees of freedom under the hypothesis of equality of dispersions.

Example 7.3 *These tests are applied to the data on the directions of paleocurrents analyzed in Sengupta and Rao (1967) . The approximate ANOVA discussed in Section 6.4.1 is not appropriate for two reasons: (1) the concentrations of the Upper, Middle and Lower Kamthi data sets are markedly different (see below, the result of the homogeneity test for concentrations), and also (2) since the CN model does not provide a good fit to any of these data sets. However the large sample tests of homogeneity described above are very appropriate as well as valid procedures to use. We use the* ***rao.homogeneity*** *routine from the* ***CircStats*** *package.*

```
>river.low.rad_river.low*pi/180
>river.mid.rad_river.mid*pi/180
>river.up.rad_river.up*pi/180
>river.rad_list(river.low.rad,river.mid.rad,river.up.rad)
>rao.homogeneity(river.rad,alpha=0)

Rao's Tests for Homogeneity

       Test for Equality of Polar Vectors:

Test Statistic = 53.18616 Degrees of Freedom = 2
P-value of test = 0

       Test for Equality of Dispersions:

Test Statistic = 23.83836 Degrees of Freedom = 2
P-value of test = 1e-05
```

Chapter 8

Circular Correlation and Regression

8.1 Introduction

In many situations, one comes across bivariate or multivariate data where some or all of the component variables are angular. For example, in studying bird migration, one may be interested in observing the direction from which the birds come as well as the direction of their return, resulting in observations on a torus i.e., (circle x circle); or one may record both the wind direction and flight direction of a migratory bird; or wind directions and drift direction of clouds, orientations of pebbles lying on foresets and foreset azimuths; or in medical studies of vector cardiograms, several variables are measured in angles. In such cases, one might be interested in questions of correlation or association between such variables as well as regression with the goal of predicting one variable given the other. Sometimes, some of the component variables may be linear so that we are interested in correlation and regression issues in such circular-linear cases as well.

This aspect of circular statistical analysis has received attention only relatively recently, as is the topic of checking the stochastic or statistical independence of two circular variables. In what follows, we will discuss some statistical aspects of bivariate data in which the observations are either circular or linear measurements.

8.2 A Circular Correlation Measure, ρ_c

In this section we consider the problem of measuring the association between two circular variables and developing related statistical inference, such as a significance test for the sample measure.

Let $(\alpha_1, \beta_1), \ldots, (\alpha_n, \beta_n)$ be a random sample of observations which are directions of two attributes both measured as angles with reference to the same zero direction and same sense of rotation. Let $f(\alpha, \beta)$ be the joint probability density function on the torus $0 \leq \alpha < 2\pi$, $0 \leq \beta < 2\pi$. Let μ and ν denote the mean directions of the two variables.

In linear analysis a measure of linear relationship between two variables X and Y is the classical product moment correlation coefficient, defined by

$$\rho(X, Y) = \mathrm{Cov}(X, Y)/\sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(Y)}.$$

This measure is natural and has the properties

$$\begin{aligned} -1 \leq \rho(X, Y) &\leq 1, \\ \rho(X, Y) &= \rho(Y, X), \\ \rho(aX + b, cY + d) &= (\mathrm{sgn}(a))(\mathrm{sgn}(c))\rho(X, Y). \end{aligned} \tag{8.2.1}$$

Further $\rho(X, Y) = 0$ if X and Y are independent although the converse is not true in general and if $\rho(X, Y) = \pm 1$, then $X = aY + b$ with probability one.

In dealing with circular variables α and β and trying to retain many of these properties, Jammalamadaka and Sarma (1988) define

$$\rho_c(\alpha, \beta) = \frac{E\left\{\sin(\alpha - \mu)\sin(\beta - \nu)\right\}}{\sqrt{\mathrm{Var}(\sin(\alpha - \mu))\mathrm{Var}(\sin(\beta - \nu))}} \tag{8.2.2}$$

as a measure of circular correlation coefficient. Notice that $E(\sin(\alpha - \mu)) = E(\sin(\beta - \nu)) = 0$ analogous to the fact that the first central moment in linear case is 0. See also Theorem 1.1 for the corresponding sample result. Thus $\sin(\alpha - \mu)$ and $\sin(\beta - \nu)$ can be taken to represent the deviations of α and β from their mean directions μ and ν and lead naturally to the correlation coefficient ρ_c. The following comments on ρ_c are in order.

(i) We may rewrite Equation (8.2.2) as

$$\rho_c(\alpha, \beta) = \frac{E\left[\cos(\alpha - \beta - \mu + \nu) - \cos(\alpha + \beta - \mu - \nu)\right]}{2\sqrt{E(\sin^2(\alpha - \mu))E(\sin^2(\beta - \nu))}}. \tag{8.2.3}$$

Since $E(\cos(\alpha-\mu))$ is a measure of concentration of α around the mean direction, the first term in the numerator represents how strongly the distribution of $(\alpha-\mu)-(\beta-\nu)$ is concentrated and this contributes to the positive part of the correlation. Similarly the second term measures the negative part of the correlation.

(ii) If μ or ν is not well defined or arbitrary because α or β has uniform distribution, then μ or ν is chosen in such a way that they maximize the terms $|E\cos(\alpha+\beta-\mu-\nu)|$ and $|E\cos(\alpha+\beta-\mu-\nu)|$ in the numerator individually. Thus whenever there is any ambiguity in the choice of the mean directions, they are chosen to yield the largest possible association in both positive and negative directions. This leads to the choice of μ and ν such that $(\mu-\nu)$ will stand for the mean direction of $(\alpha-\beta)$ and $(\mu+\nu)$, the mean direction of $(\alpha+\beta)$. Recall in this connection Remark 2.1 that the circular mean of the sum is not the sum of the circular means, in general.

The numerator then becomes the difference in the lengths of the mean vectors of $(\alpha-\beta)$ and $(\alpha+\beta)$. Writing

$$R_{\alpha\pm\beta} = |Ee^{i(\alpha\pm\beta)}|,$$

we have

$$\rho_c = \frac{(R_{\alpha-\beta} - R_{\alpha+\beta})}{2\sqrt{E(\sin^2(\alpha-\mu))E(\sin^2(\beta-\nu))}}. \tag{8.2.4}$$

(iii) It should be pointed out that in this case (of uniform marginals) the above correlation coefficient is equivalent to the one proposed by Rivest (1982) in the sense that they differ only in the denominators. In general the two are different and are arrived at from completely different considerations.

(iv) The definition of ρ_c as given above can be used even if (α,β) have supports which are less than the full circle.

(v) Observe that Pearson's product moment correlation in the linear case can be rewritten in the form

$$\frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = \frac{E(X_1-X_2)(Y_1-Y_2)}{\sqrt{E\{(X_1-X_2)^2\}E\{(Y_1-Y_2)^2\}}},$$

where (X_1, Y_1) and (X_2, Y_2) are i.i.d. as (X, Y). Fisher and Lee (1983) utilize this alternative form to define a circular correlation as

$$\rho = \frac{E\{\sin(\alpha_1-\alpha_2)\sin(\beta_1-\beta_2)\}}{\sqrt{E\{\sin^2(\alpha_1-\alpha_2)\}E\{\sin^2(\beta_1-\beta_2)\}}}.$$

However their definition and ρ_c are algebraically different and do not always yield the same value.

The circular correlation coefficient ρ_c satisfies the following properties.

- $\rho_c(\alpha, \beta)$ does not depend on the zero direction used for either variable;
- $\rho_c(\alpha, \beta) = \rho_c(\beta, \alpha)$;
- $|\rho_c(\alpha, \beta)| \leq 1$;
- $\rho_c(\alpha, \beta) = 0$ if α and β are independent although the converse need not be true;
- If α and β have full support, $\rho_c(\alpha, \beta) = 1$ iff $\alpha = \beta +$ const (mod 2π) and $\rho_c(\alpha, \beta) = -1$ iff $\alpha + \beta =$ const(mod 2π);
- $\rho_c(\alpha, \beta) \approx \rho(\alpha, \beta)$, the product moment correlation if α and β are concentrated in a small neighborhood of their respective mean directions and are measured in radians.

The *sample* correlation coefficient if $(\alpha_1, \beta_1), \ldots, (\alpha_n, \beta_n)$ is a random sample, is given by

$$r_{c,n} = \frac{\sum_{i=1}^{n} \sin(\alpha_i - \bar{\alpha})\sin(\beta_i - \bar{\beta})}{\sqrt{\sum_{i=1}^{n} \sin^2(\alpha_i - \bar{\alpha})\sin^2(\beta_i - \bar{\beta})}}, \tag{8.2.5}$$

where $\bar{\alpha}$ and $\bar{\beta}$ are the sample mean directions. The sample correlation is an estimate of ρ_c. When the joint distribution of (α, β) is not fully specified, we can use the sample measure $r_{c,n}$ for testing hypotheses about ρ_c when n is sufficiently large, using the following result.

Theorem 8.1 *The quantity* $\sqrt{n}(r_{c,n}-\rho_c)$ *converges in distribution to* $N(0,\sigma^2)$ *as* $n \longrightarrow \infty$*, where*

$$\begin{aligned}\sigma^2 &= \frac{\lambda_{22}}{\lambda_{20}\lambda_{02}} - \rho_c\left[\frac{\lambda_{13}}{\lambda_{20}\sqrt{\lambda_{20}\lambda_{02}}} + \frac{\lambda_{31}}{\lambda_{02}\sqrt{\lambda_{20}\lambda_{02}}}\right] \\ &\quad +\frac{1}{4}\rho_c^2\left[1+\frac{\lambda_{40}}{\lambda_{20}^2}+\frac{\lambda_{04}}{\lambda_{02}^2}+\frac{\lambda_{22}}{\lambda_{20}\lambda_{02}}\right],\end{aligned}$$

where

$$\lambda_{ij} = E\left\{\sin^i(\alpha-\mu)\sin^j(\beta-\nu)\right\} \qquad i,j = 0,1,2,3,4.$$

Proof: The proof follows by using Lemma 3.1 and considerable simplifications.

Corollary 8.1 *In particular under the hypothesis* $H_0 : \rho_c = 0$ *for large* n,

$$\sqrt{n}r_{c,n} \sim N\left(0, \frac{\lambda_{22}}{\lambda_{20}\lambda_{02}}\right)$$

and by Slutsky's theorem,

$$\sqrt{n}\sqrt{\frac{\widehat{\lambda_{20}}\widehat{\lambda_{02}}}{\widehat{\lambda_{22}}}}r_{c,n} \sim N(0,1),$$

where

$$\widehat{\lambda_{ij}} = \frac{1}{n}\sum_{k=1}^{n}\sin^i(\alpha_k-\bar{\alpha})\sin^j\left(\beta_k-\bar{\beta}\right).$$

Example 8.1 *Following are pairs of measurements on wind direction at 6:00 am and 12:00 noon (denoted by* θ *and* ϕ*, respectively), on each of 21 consecutive days, at a weather station in Milwaukee (from Johnson and Wehrly (1977)). We use the program* `circ.cor` *of the* `CircStats` *package to compute the circular correlation and check its significance.*

θ	356	97	211	232	343	292	157
ϕ	119	162	221	259	270	29	97
θ	302	335	302	324	85	324	340
ϕ	292	40	313	94	45	47	108
θ	157	238	254	146	232	122	329
ϕ	221	270	119	248	270	45	23

```
> theta_c(356,97,211,232,343,292,157,302,335,302,324,85,324,340,
+ 157,238,254,146,232,122,329)
> phi_c(119,162,221,259,270,29,97,292,40,313,94,45,47,108,221,
+ 270,119,248,270,45,23)
> rtheta_rad(theta)
> rphi_rad(phi)
> rtheta
 [1] 6.213372 1.692969 3.682645 4.049164 5.986479 5.096361
 [7] 2.740167 5.270894 5.846853 5.270894 5.654867 1.483530
 [13] 5.654867 5.934119 2.740167 4.153884 4.433136 2.548181
 [19] 4.049164 2.129302 5.742133
> rphi
 [1] 2.0769418 2.8274334 3.8571776 4.5204028 4.7123890 0.5061455
 [7] 1.6929694 5.0963614 0.6981317 5.4628806 1.6406095 0.7853982
 [13] 0.8203047 1.8849556 3.8571776 4.7123890 2.0769418 4.3284165
 [19] 4.7123890 0.7853982 0.4014257
> circ.cor(rtheta,rphi,T)
          r test.stat   p.value
1 0.2704648  1.214025 0.2247383
```

which shows that there is not a high degree of association between the two measurements.

8.2.1 ρ_c for Some Parametric Models

Wrapped Bivariate Normal

This distribution can be seen to have the joint c.f.

$$Q(m,n) = \exp\left\{ i\left(m\mu_1 + n\mu_2\right) - \frac{1}{2}\left(m\sigma_1^2 + 2mn\sigma_{12} + n\sigma_2^2\right)\right\},$$

where m and n are integers. After some calculations, it can be shown that

$$\rho_c(\alpha,\beta) = \frac{\sinh\sigma_{12}}{\sqrt{\sinh\sigma_1^2 \sinh\sigma_2^2}}.$$

Note that this is comparable to λ_2 of Johnson and Wehrly (1977) except that here the sign of association between α and β is naturally incorporated though ($\sinh\sigma_{12}$). Indeed, it has the same sign as the original correlation or covariance.

Remark 8.1 *It may also be noted that $\rho_c(\alpha, \beta) = 0$ if X_1 and X_2 are independent. However even if X_1 and X_2 are perfectly correlated in the linear sense, it does not follow that $\rho_c(\alpha, \beta) = 1$ unless $\sigma_1 = \sigma_2$. In this case, $\alpha = \beta + const \pmod{2\pi}$ from (5) of the properties listed above. Similarly when the linear correlation is -1 and $\sigma_1 = \sigma_2$, then $\rho_c = -1$.*

Estimation of ρ_c

In this model, the MLE of ρ_c is not easily tractable. However, by using the method of moments we arrive at the following estimates:

$$\begin{aligned}
\hat{\sigma}_1^2 &= -2\log\bar{x}, \\
\hat{\sigma}_2^2 &= -2\log\bar{y}, \\
\hat{\sigma}_{12} &= \log\frac{\bar{x}\bar{y}}{\bar{z}}, \\
\text{and } \hat{\rho}_c &= \frac{\sinh\ \hat{\sigma}_{12}}{\sqrt{(\sinh\ \hat{\sigma}_1^2 \sinh\ \hat{\sigma}_2^2)}},
\end{aligned}$$

where

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n}\cos\ \alpha_i, \qquad \bar{y} = \frac{1}{n}\sum_{i=1}^{n}\cos\ \beta_i, \qquad \bar{z} = \frac{1}{n}\sum_{i=1}^{n}\cos\ (\alpha_i + \beta_i).$$

The asymptotic distribution of $\hat{\rho}_c$ can be seen to be $N(\rho_c, \sigma^2)$, where the expression for σ^2 can be derived using Theorem 8.1. This result can be used for testing hypotheses about ρ_c in large samples.

A Model with Uniform Marginals

Suppose (α, β) is a random vector such that α has the uniform distribution and the conditional distribution of β given α is von Mises with p.d.f.

$$f(\beta|\alpha; \theta, a, \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp\kappa\left\{\cos(\beta - a\alpha - \theta)\right\},$$

where $\kappa \geq 0$, $0 \leq \theta < 2\pi$ and $a = \pm 1$ are unknown parameters. The marginal distribution of β will also be uniform on the circle and the joint distribution of α and β has the density function

$$f(\alpha, \beta; \theta, a, \kappa) = \frac{1}{4\pi^2 I_0(\kappa)} \exp\kappa\left\{\cos(\beta - a\alpha - \theta)\right\}.$$

The model $a = 1$ is considered by Johnson and Wehrly (1977). Since α and β have uniform distributions, μ and ν are arbitrary and will be chosen as described earlier in Comment (ii).

Suppose for definiteness $a = 1$. In the numerator of ρ_c,

$$\begin{aligned} E\{\cos(\alpha+\beta-\mu-\nu)\} &= E_\alpha E_{\beta|\alpha}\{\cos(\beta-\alpha-\theta+(2\alpha+\theta-\mu-\nu))\} \\ &= \frac{I_1(k)}{I_0(k)} E_\alpha\{\cos(2\alpha+\theta-\mu-\nu)\} \\ &= 0. \end{aligned}$$

Since μ and ν are arbitrary, they are chosen to maximize $E\{\cos(\alpha-\beta-\mu+\nu)\}$ which happens when $(\mu-\nu)$ is the mean direction of $\alpha-\beta$ which is $-\theta$. Then we have $E\{\cos(\beta-\alpha-\theta)\} = A(\kappa) = I_1(\kappa)/I_0(\kappa)$ and $E\{\sin^2(\alpha-\mu)\} = E\{\sin^2(\beta-\nu)\} = 1/2$, resulting in

$$\rho_c = A(\kappa).$$

Similarly, when $a = -1$, $\rho_c = -A(\kappa)$. The ML estimates of the parameters can be obtained as follows. Let $\delta_i^- = \beta_i - \alpha_i$, $\delta_i^+ = \beta_i + \alpha_i$, $\bar{\delta}^\pm$ and $R_{\delta^\pm}$ be the mean directions and resultant lengths. Then

$$\hat{a} = \begin{cases} 1 & \text{if } R_{\delta^-} > R_{\delta^+}, \\ -1 & \text{if } R_{\delta^-} < R_{\delta^+}, \end{cases}$$

$\hat{\kappa}$ and $\hat{\theta}$ are chosen to maximize $L(1, \kappa_+, \theta_+)$ and $L(-1, \kappa_-, \theta_-)$, $\hat{\kappa}_\pm$ is obtained from the equation

$$A(\hat{\kappa}_\pm) = \frac{1}{n} R_{\delta^\pm}$$

and $\hat{\theta}_\pm = \bar{\delta}^\mp$. After some more calculations we finally have the MLE as

$$\hat{\rho}_c(\alpha,\beta) = \hat{a}A(\hat{\kappa}) = \begin{cases} R_{\delta^-}/n & \text{if } R_{\delta^-} > R_{\delta^+}, \\ -R_{\delta^+}/n & \text{if } R_{\delta^-} < R_{\delta^+}. \end{cases}$$

Remark 8.2 *Mardia (1975) considers the case where α and β have uniform marginals and proposes* $r = \max(\bar{R}_{\delta^-}, \bar{R}_{\delta^+}) = |\hat{\rho}_c(\alpha,\beta)|$.

A test for $\rho_c = 0$ is considered next. This hypothesis is equivalent to $\kappa = 0$. Under this hypothesis α and β are independent and have uniform distribution on $(0, 2\pi)$. Hence both δ^+ and δ^- have uniform distribution on $(0, 2\pi)$ and are independent. Hence R_{δ^+} and R_{δ^-} are i.i.d. For testing against $\rho > 0, (\rho_c < 0)$ the test is based on R_{δ^-} (R_{δ^+}). So we can use Rayleigh's test for testing against the two-sided alternative $\rho_c \neq 0$, we use max $(R_{\delta^+},\ R_{\delta^-})$.

A test for $\rho_c = \rho_c^0$ can be converted into a hypothesis in terms of κ i.e., $H_0 : \kappa = \kappa_0$, and use the known distributions of the resultant lengths for circular normal samples.

8.3 Rank Correlation

In this section we define a nonparametric correlation between two circular variables, based on ranks. Although ranks are not well defined on a circle as we said before, the rank correlation measure which we propose is origin-invariant and provides a useful measure. Given $(\alpha_i, \beta_i), i = 1, \dots, n$, these are first converted to uniform scores i.e. one of these sets, say, $\alpha_1, \dots, \alpha_n$ are linearly ranked and then let r_i denote the rank of the β_i which corresponds to the ith largest α_i for $i = 1, \dots, n$. Thus the observations (α_i, β_i) are converted into ranks (i, r_i) and are replaced by the "uniform scores"

$$\varphi_i = \frac{2\pi i}{n} \text{ and } \psi_i = \frac{2\pi r_i}{n}, \quad i = 1, \dots, n.$$

The sample based rank correlation η_c is now defined as the circular correlation r_c between φ and ψ as given in Equation (8.2.5) i.e., $\rho_c(\varphi, \psi)$. Thus

$$\eta_c = \frac{\sum_{i=1}^n \sin(\varphi_i - \bar{\varphi}) \sin(\psi_i - \bar{\psi})}{\sqrt{\sum_{i=1}^n \sin^2(\varphi_i - \bar{\varphi}) \sum_{i=1}^n \sin^2(\psi_i - \bar{\psi})}}.$$

Since $\bar{\varphi}$ and $\bar{\psi}$ are not well-defined, using the means of uniform scores, we should deal with this as Comment (ii) following the definition (8.2.2). Noting that

$$\sum_{i=1}^{n} \sin^2\left(\frac{2\pi i}{n} - \bar{\varphi}\right) = \frac{n}{2}$$

for any choice of $\bar{\varphi}$, we may re-write

$$\eta_c = \frac{1}{n}\sum_{i=1}^{n} \cos\left(\varphi_i - \psi_i - (\bar{\varphi} - \bar{\psi})\right) - \frac{1}{n}\sum_{i=1}^{n} \cos\left(\varphi_i + \psi_i - (\bar{\varphi} + \bar{\psi})\right).$$

Now since $\bar{\varphi}$ and $\bar{\psi}$ are arbitrary, they are chosen to maximize the two terms on the r.h.s. individually. This occurs when $(\bar{\varphi}_i \pm \bar{\psi}_i)$ are the resultant directions of $(\varphi_i \pm \psi_i)$, $i = 1, \ldots, n$, respectively. The coefficient η_c then becomes the difference in the lengths of the resultants,

$$\eta_c = \frac{R_{\delta^-} - R_{\delta^+}}{n}.$$

This is clearly invariant under the choice of zero direction or sense of rotation and has the appropriate sign automatically attached to it. The nonparametric measure η_c bears the same relationship to the original measure ρ_c as does the Spearman's rank correlation, to the product-moment correlation. Mardia (1975) suggested $\max(R_{\delta^-}/n, R_{\delta^+}/n)$ and adding the sign corresponding to the larger resultant.

8.4 Other Measures of Circular Correlation

We shall briefly review some of the other measures of association that have been proposed in the literature.

(i) In the notations introduced before, Mardia (1975) proposes a circular correlation coefficient as

$$r^2 = \max\{D_+, D_-\},$$

where

$$n^2 D_\pm = \left\{\sum_{i=1}^{n} \cos(\alpha_i \mp \beta_i)\right\}^2 + \left\{\sum_{i=1}^{n} \sin(\alpha_i \mp \beta_i)\right\}^2.$$

Clearly $r^2 \leq 1$, being the concentration measure for the sums and/or differences in the angles and $D_+ = 1$ if the differences are constant while $D_- = 1$ if the sums are constant.

(ii) Johnson and Wehrly (1977) proposed the following measure of correlation

$$\rho_A = \sup_{\theta,\varphi \in [0,2\pi)} \rho\{\cos(\alpha - \theta), \cos(\beta - \varphi)\},$$

where ρ is the usual product moment correlation. This measure can be seen to be invariant under rotations of α and β and corresponds to the dominant canonical correlation between $\cos\alpha$ and $\cos\beta$. Further the independence of α and β implies $\rho_A = 0$. Also if α and β have support $[0, 2\pi)$, then $\rho_A = 1$ if and only if $\beta - \varphi \equiv \pm(\alpha - \theta)$ (mod 2π) for some θ and φ. However a linear relationship between α and β need not imply perfect correlation. For instance, if $\beta = 2\alpha$ (mod 2π) then $\rho_A = 0$.

(iii) Mardia and Puri (1978) propose

$$\rho^2 = \rho_{cc}^2 + \rho_{cs}^2 + \rho_{sc}^2 + \rho_{ss}^2,$$

where ρ_{cc} = corr $(\cos\alpha, \cos\beta)$, $\rho_{cs} = \text{corr}(\cos\alpha, \sin\beta)$ etc. with the corresponding sample version

$$r^2 = r_{cc}^2 + r_{cs}^2 + r_{sc}^2 + r_{ss}^2.$$

This is scale invariant and some distributional properties have been studied through simulations.

(iv) Stephens (1979) proposed the measure

$$\rho = \sup_H \; Tr(HA),$$

where

$$A = \begin{pmatrix} E(\cos\alpha\cos\beta) & E(\cos\alpha\sin\beta) \\ E(\sin\alpha\cos\beta) & E(\sin\alpha\sin\beta) \end{pmatrix}$$

and H is an orthonormal matrix. Since $A = E(\underset{\sim}{X}\;\underset{\sim}{Y}')$, where $\underset{\sim}{X}' = (\cos\alpha, \sin\alpha)$ and $\underset{\sim}{Y}' = (\cos\beta, \sin\beta)$, the idea is to measure how close the vectors $\underset{\sim}{X}$ can be brought to the vectors $\underset{\sim}{Y}$ by a rotation or an orthogonal transformation. If AA' is positive definite with eigenvalues λ_1, λ_2, then it can be seen that $\rho = \sqrt{\lambda_1} + \sqrt{\lambda_2}$. ρ is non-negative and $\rho \neq 0$ even when α and β are independent.

(v) Rivest (1982) proposes the smaller singular value of A (singular values of A are the non-negative square-roots of the eigenvalues of $A'A$), resulting in the measure

$$\rho = \frac{E\left\{\sin(\alpha - \mu_0)\sin(\beta - \nu_0)\right\}}{\max\left\{E\sin^2(\alpha - \mu_0), E\sin^2(\beta - \nu_0)\right\}}$$

provided the resultant lengths of $(\alpha+\beta)$ and $(\alpha-\beta)$ are non-zero.

This ρ is very similar to the ρ_c that we proposed and discussed earlier and shares many properties in common. See Comment (iii) made earlier.

8.5 Circular-Linear Correlation

In many situations one may observe a circular variable α say the wind direction and a linear variable x, say the humidity, amount of rain or the concentration of pollutants; or α may represent the direction of departure and x, the distance to the destination; or α may represent the time of birth within a cycle of 24 hours while x may represent the number of births.

Suppose we have a sample $(\alpha_1, x_1), \ldots, (\alpha_n, x_n)$. If a plot of x on the vertical axis versus α on the horizontal axis reveals that there is only one peak within the period, then one may use a sinusoidal regression curve represented by

$$X = M + A\cos(\alpha - \alpha_0), \tag{8.5.1}$$

where M represents the mean level, A the amplitude, and α_0 the so-called "acrophase angle", where the curve reaches its peak. One of the interesting problems here is the determination of the acrophase angle. We can rewrite Equation (8.5.1)

$$X = M + C_1\cos\alpha + C_2\sin\alpha,$$

where $C_1 = A\cos\alpha_0$, $C_2 = A\sin\alpha_0$ and then determine M, C_1 and C_2 by the method of least squares. From these estimates, we can determine A and α_0 as

$$\begin{aligned} A &= \sqrt{C_1^2 + C_2^2}, \\ \alpha_0 &= \arctan^* \frac{C_2}{C_1}. \end{aligned}$$

Alternatively, one may take $y_i = \cos(\alpha_i - \alpha_0)$, so that we have pairs (x_1, y_1), $\ldots$, (x_n, y_n) for which ordinary linear correlation coefficient can be computed. We may then choose α_0 as the one which maximizes the correlation coefficient. This is proposed by Mardia and Jupp (1980) and Johnson and Wehrly (1977).

A reasonable measure of correlation between a linear and angular variable is to take the multiple correlation between the linear variable X with the components $(\cos\alpha, \sin\alpha)$ corresponding to the angular variable α. Such a measure discussed in Mardia (1976) and Johnson and Wehrly (1977) is given by

$$r^2 = \frac{r_{xc}^2 + r_{xs}^2 - 2r_{xc}r_{xs}r_{cs}}{1 - r_{cs}^2},$$

where

$$\begin{aligned} r_{xc} &= \operatorname{corr}(x, \cos\alpha), \\ r_{xs} &= \operatorname{corr}(x, \sin\alpha), \\ r_{cs} &= \operatorname{corr}(\cos\alpha, \sin\alpha). \end{aligned}$$

When X and α are independent with X having a normal distribution, it may be seen that

$$\frac{(n-3)r^2}{1-r^2} \sim F_{2,n-3}.$$

The exact distribution of this statistic for a specific non-uniform alternative is discussed in Liddell and Ord (1978).

8.6 Circular-Circular Regression

Let (α, β) have joint pdf $f(\alpha, \beta)$, $0 < \alpha$, $\beta \leq 2\pi$. To predict β for a given α, consider the regression or conditional expectation of the vector $e^{i\beta}$ given α, say

$$E\left(e^{i\beta}|\alpha\right) = \rho(\alpha)e^{i\mu(\alpha)} = g_1(\alpha) + ig_2(\alpha). \tag{8.6.1}$$

Here $\mu(\alpha)$ represents the conditional mean direction of β given α and $0 \leq \rho(\alpha) \leq 1$ the conditional concentration towards this direction. Equivalently

$$\begin{aligned} E(\cos\ \beta|\alpha) &= g_1(\alpha), \\ E(\sin\ \beta|\alpha) &= g_2(\alpha) \end{aligned} \tag{8.6.2}$$

from which β is determined as

$$\mu(\alpha) = \hat{\beta} = \arctan^* \frac{g_2(\alpha)}{g_1(\alpha)}. \tag{8.6.3}$$

Predicting β in this way is optimal in the sense that it minimizes

$$E\left\{||e^{i\beta} - g(\alpha)||^2\right\}$$

and is similar to the least squares idea. This was proposed and explored in Jammalamadaka and Sarma (1993) and we now discuss this approach in detail.

In the absence of further specifications on the structure of $g_1(\alpha)$ and $g_2(\alpha)$, it is in general difficult to estimate these from the data. So $g_1(\alpha)$ and $g_2(\alpha)$ will be approximated by suitable functions. Since $g_1(\alpha)$ and $g_2(\alpha)$ are both periodic with period 2π, they may be expressed in terms of their Fourier series expansions. This leads to approximating $g_1(\alpha)$ and $g_2(\alpha)$ by trigonometric polynomials of suitable degree say m, i.e.

$$\begin{aligned} g_1(\alpha) &\approx \sum_{k=0}^{m} (A_k \cos k\alpha + B_k \sin k\alpha), \\ g_2(\alpha) &\approx \sum_{k=0}^{m} (C_k \cos k\alpha + D_k \sin k\alpha). \end{aligned} \tag{8.6.4}$$

Thus we have the following model, which is in fact a general linear model:

$$\begin{aligned} \cos\beta &\approx \sum_{k=0}^{m} (A_k \cos k\alpha + B_k \sin k\alpha) + \epsilon_1, \\ \sin\beta &\approx \sum_{k=0}^{m} (C_k \cos k\alpha + D_k \sin k\alpha) + \epsilon_2, \end{aligned} \tag{8.6.5}$$

where (ϵ_1, ϵ_2) is the error vector with mean vector $\mathbf{0}$ and dispersion matrix $\boldsymbol{\Sigma}$ which is unknown.

8.6.1 Estimation of Regression Coefficients

We now turn to the problem of estimation of (A_k, B_k, C_k, D_k), $k = 0, 1, \ldots, m$ and $\boldsymbol{\Sigma}$. Although $B_0 = D_0 = 0$, there are nearly four times as many parameters as m and thus small values of m should be considered. Let

$$
\begin{aligned}
Y_{1i} &= \cos\beta_i, \quad i = 1, \ldots, n, \\
Y_{2i} &= \sin\beta_i, \quad i = 1, \ldots, n, \\
\underset{\sim}{Y}^{(1)} &= (Y_{11}, \ldots, Y_{1n})', \\
\underset{\sim}{Y}^{(2)} &= (Y_{21}, \ldots, Y_{2n})', \\
\underset{\sim}{\epsilon}^{(1)} &= (\epsilon_{11}, \ldots, \epsilon_{1n})', \\
\underset{\sim}{\epsilon}^{(2)} &= (\epsilon_{21}, \ldots, \epsilon_{2n})'
\end{aligned}
$$

with the design matrix

$$
\underset{n\times(2m+1)}{X} = \begin{bmatrix} 1 & \cos\alpha_1 & \cos 2\alpha_1 & \cdots & \cos m\alpha_1 & \sin\alpha_1 & \cdots & \sin m\alpha_1 \\ \vdots & \vdots & \vdots & & \vdots & & \vdots & \\ 1 & \cos\alpha_n & \cos 2\alpha_n & \cdots & \cos m\alpha_n & \sin\alpha_n & \cdots & \sin m\alpha_n \end{bmatrix}
$$

and parameters,

$$
\begin{aligned}
\underset{\sim}{\lambda}^{(1)} &= (A_0, A_1 \ldots, A_m, B_1, \ldots, B_m)', \\
\underset{\sim}{\lambda}^{(2)} &= (C_0, C_1 \ldots, C_m, D_1, \ldots, D_m)'.
\end{aligned}
$$

Putting them together as a single vector,

$$
\begin{aligned}
\underset{\sim}{Y}^{*'} &= \left(\underset{\sim}{Y}^{(1)'} \vdots \underset{\sim}{Y}^{(2)'}\right), \\
\underset{\sim}{\epsilon}^{*'} &= \left(\underset{\sim}{\epsilon}^{(1)'} \vdots \underset{\sim}{\epsilon}^{(2)'}\right), \\
\underset{\sim}{\lambda}^{*'} &= \left(\underset{\sim}{\lambda}^{(1)'} \vdots \underset{\sim}{\lambda}^{(2)'}\right),
\end{aligned}
$$

the observational equations can be written as

$$
\underset{\sim}{\lambda}^* = \underset{\sim}{X}^* \underset{\sim}{\lambda}^* + \underset{\sim}{\epsilon}^*, \tag{8.6.6}
$$

where

$$
X^* = \begin{bmatrix} X & 0 \\ 0 & X \end{bmatrix}.
$$

From these the least squares estimates of λ^* can be obtained. However, because of the structure of X^* etc. the least squares estimates turn out to be

$$\hat{\underset{\sim}{\lambda}}^{(1)} = (X'X)^{-1}\ X'\underset{\sim}{Y}^{(1)}, \tag{8.6.7}$$

$$\hat{\underset{\sim}{\lambda}}^{(2)} = (X'X)^{-1}\ X'\underset{\sim}{Y}^{(2)}. \tag{8.6.8}$$

Remark 8.3 *If the values of α are taken to be equally spaced over the circle, the computations become much simpler. If for instance,*

$$\alpha_i = \frac{2\pi(i-1)}{n}, \quad i = 1, \dots, n,$$

then the LS estimates turn out to be

$$\begin{aligned}
\hat{A}_0 &= \frac{1}{n}\sum_{i=1}^{n}\cos\beta_i, \\
\hat{A}_j &= \frac{1}{n}\sum_{i=1}^{n}\cos\beta_i\cos j\alpha_i, \\
\hat{B}_j &= \frac{1}{n}\sum_{i}^{n}\cos\beta_i\sin j\alpha_i, \\
\hat{C}_0 &= \frac{1}{n}\sum_{i=1}^{n}\sin\beta_i, \\
\hat{C}_j &= \frac{1}{n}\sum_{i=1}^{n}\sin\beta_i\cos j\alpha_i, \\
\hat{D}_j &= \frac{1}{n}\sum_{i=1}^{n}\sin\beta_i\sin j\alpha_i.
\end{aligned}$$

Remark 8.4 *When several β's are observed corresponding to the same value of α, the matrix X' will not be of full rank and one has to use a generalized inverse like the Moore-Penrose inverse for $(X'X)$.*

Remark 8.5 *The covariance matrix Σ can be estimated as follows: let*

$$\begin{aligned} R_0(i,j) &= \underset{\sim}{Y}^{(i)\prime}\ \underset{\sim}{Y}^{(i)} - \underset{\sim}{Y}^{(i)\prime}\ X(X'X)^{-1}X'\underset{\sim}{Y}^{(j)} \\ &= \underset{\sim}{Y}^{(i)\prime}\ (I-M)\underset{\sim}{Y}^{(j)} \end{aligned}$$

and

$$R_0 = (R_0(i,j))_{2\times 2}.$$

Then $\hat{\Sigma} = [n-2(m+1)]^{-1}\mathbf{R_0}$ is an unbiased estimate of Σ. From this the standard errors of the estimators can be found.

8.6.2 Determination of m

A general problem in fitting any polynomial regression is the determination of the degree m of the polynomial. We give some details of this as this is useful in practice.

(i) A primary consideration in determining the degree is to assess the reduction one obtains in the error sum of squares when m is increased. The effect of augmenting the design matrix X on the residual sum of squares involves standard calculations and is done in our special case as follows (see Seber (1977)). Deciding to take $(m+1)$th degree trigonometric polynomial amounts to adding columns

$$(\cos(m+1)\alpha_1, \cdots, \cos(m+1)\alpha_n)' \text{ and } (\sin(m+1)\alpha_1, \cdots, \sin(m+1)\alpha_n)'$$

as $(m+2)$th and the last (i.e., $(2m+3)$rd) columns of X. The augmented matrix can be written as

$$X_{(1)} = (X \,\vdots\, W)P,\ W = \begin{pmatrix} \cos(m+1)\alpha_1 & \sin(m+1)\ \alpha_1 \\ & \vdots \\ \cos(m+1)\alpha_n & \sin(m+1)\ \alpha_n \end{pmatrix}$$

and P is a suitable permutation matrix to put the two columns of W in $(m+1)$th and $(2m+3)$rd places. The model is now

$$\begin{aligned} \underset{\sim}{Y}^{(1)} &= X_{(1)}\ \underset{\sim}{\lambda}^{(1)}_{(1)}\ \underset{\sim}{\epsilon}^{(1)}, \\ \underset{\sim}{Y}^{(2)} &= X_{(1)}\ \underset{\sim}{\lambda}^{(2)}_{(1)}\ \underset{\sim}{\epsilon}^{(2)}, \end{aligned}$$

where $\underset{\sim}{\lambda}_{(1)}^{(1)}$ and $\underset{\sim}{\lambda}_{(1)}^{(2)}$ are $(2m+3)\times 1$ vectors of coefficients, which can be estimated as before by

$$\hat{\underset{\sim}{\lambda}}_{(1)}^{(i)} = \left(X'_{(1)}\ X_{(1)}\right)^{-1} X'_{(1)}\ \underset{\sim}{Y}^{(i)}, \quad i=1,2.$$

Now

$$\underset{\sim}{X}'_{(1)}\ \underset{\sim}{X}_{(1)} = P' \begin{bmatrix} X'X & X'W \\ W'X & W'W \end{bmatrix} P.$$

We can write (see Rohde (1965))

$$\begin{aligned} &\begin{pmatrix} X'X & X'W \\ W'X & W'W \end{pmatrix}^{-1} \\ = &\begin{pmatrix} (X'X)^{-1} & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} (X'X)^{-1}X'W \\ -I \end{pmatrix} H^{-1} \begin{pmatrix} (X'X)^{-1}X'W \\ -I \end{pmatrix}', \end{aligned}$$

where

$$H = W'W - W'X(X'X)^{-1}X'W = W'(I-M)W.$$

The least squares estimators can then be written as

$$\begin{aligned} \hat{\underset{\sim}{\lambda}}_{(1)}^{(i)} &= P\left(X'_{(1)}\ X_{(1)}\right)^{-1} P\ P' \begin{pmatrix} X' \\ \cdots \\ W' \end{pmatrix} \underset{\sim}{Y}^{(i)} \\ &= P' \begin{pmatrix} (X'X)^{-1}X'\underset{\sim}{Y}^{(i)} - (X'X)^{-1}X'N(I-M)\ \underset{\sim}{Y}^{(i)} \\ H^{-1}W'(I-M)\ \underset{\sim}{Y}^{(i)} \end{pmatrix}, \end{aligned}$$

where $N = W\ H^{-1}\ W'$. The error sum of squares is now

$$\begin{aligned} &\underset{\sim}{Y}^{(i)\prime}\ \underset{\sim}{Y}^{(i)} - \underset{\sim}{Y}^{(i)\prime}(X\vdots W)P\,\hat{\lambda}_{(1)}^{(i)} \\ = &\ \underset{\sim}{Y}^{(i)\prime}(I-M)\underset{\sim}{Y}^{(i)} - \underset{\sim}{Y}^{(i)\prime}(I-M)N(I-M)\underset{\sim}{Y}^{(i)} \end{aligned}$$

on simplification.

Thus the reduction in the error sum of squares corresponding to $\hat{\lambda}^{(i)}$ by taking additional terms is

$$\underset{\sim}{Y}^{(i)\prime}(I-M)N(I-M)\underset{\sim}{Y}^{(i)} \geq 0. \tag{8.6.9}$$

In determining whether to take the terms of degree $(m+1)$, we first compute the reduction and if it is significantly large we decide to include the $(m+1)$th terms. When n is sufficiently large a test for this will be discussed a little later. Note that $(I-M)\underset{\sim}{Y}^{(i)}$ is already available and we need to find N for which a 2×2 matrix is to be inverted if additional terms are deemed necessary. The computations already carried out are adequate to get the new estimates $\hat{\lambda}^{(i)}_{(1)}$. This procedure can be repeated successively.

(ii) The extent to which the accuracy of prediction of β would be improved can be ascertained in a way similar to the linear analysis where the partial correlation ratio is considered. Recall that in linear analysis with linear predictors, the proportional reduction in the mean square error when $X_1, \ldots, X_p$ are used to predict Y rather than $X_1, \ldots, X_k$, $p > k$, is given by

$$\rho^2_{0(k+1,\ldots,p)(1,\ldots,k)} = \frac{\rho^2_{0(1,\ldots,p)} - \rho^2_{0(1,\ldots,k)}}{1 - \rho^2_{0(1,\ldots,k)}},$$

where

$$\rho^2_{0(1,\ldots,m)} = \underset{\sim}{\sigma}_0' \; C^{-1} \; \underset{\sim}{\sigma}_0 / \sigma_Y^2$$

is the square of multiple correlation coefficient with

$$\underset{\sim}{\sigma}_0 = Cov(Y, X_1, \ldots, X_m)', \qquad \underset{m\times m}{C} = Cov(X_1, \ldots, X_m).$$

An estimate is obtained by taking the sample moments. In the present case,

$$\rho_i^{*2} = \frac{\underset{\sim}{\hat{\lambda}}^{(i)'}(X'X)\;\underset{\sim}{\hat{\lambda}}^{(i)}}{\underset{\sim}{Y}^{(i)'}\;\underset{\sim}{Y}^{(i)}} = \frac{\underset{\sim}{Y}^{(i)'}\;M\,\underset{\sim}{Y}^{(i)}}{\underset{\sim}{Y}^{(i)'}\;\underset{\sim}{Y}^{(i)}} \quad \text{for } i = 1, 2.$$

Let

$$\rho_i^{*2} = \frac{\rho^{*2}_{i\;o(m+1)} - \rho^{*2}_{i\;o\;m}}{1 - \rho^{*2}_{i\;o\;m}}, \qquad \text{for } i = 1, 2$$

which represents the proportional reduction in the regression sum of squares. If these are "sufficiently large" we decide to include the higher order term also.

(iii) An alternate but equivalent approach is to look at the increase in the estimated concentration of the conditional distribution of β given α. Having computed the mth degree polynomial approximation $g_{im}\;(\alpha)$ say, the concentration parameter is given by

$$\rho_m\ (\alpha) = \{g_{1m}^2\ (\alpha) + g_{2m}^2\ (\alpha)\}^{1/2}.$$

Notice that $\rho_m(\alpha)$ is an increasing function of m and $\rho_m(\alpha)\ \uparrow\ \rho(\alpha)$ as $m \to \infty$ for any given α. Since $0 \leq \rho(\alpha) \leq 1$, $0 \leq \rho_m(\alpha) \leq 1\ \forall m$ this concentration is estimated by

$$\begin{aligned} \hat{\rho}_m^2 &= \frac{1}{n}\sum_{i=1}^{n}\{g_{1m}^2(\alpha_i) + g_{2m}^2(\alpha_i)\} \\ &= \frac{1}{n}\sum_{j=1}^{2} \underset{\sim}{Y}^{(j)'}\, M\, \underset{\sim}{Y}^{(j)} \\ &= \frac{1}{n}\sum_{j=1}^{2}\left\{\underset{\sim}{Y}^{(j)'}\ \underset{\sim}{Y}^{(j)} - R_0(j,j)\right\}. \end{aligned}$$

Working with the augmented matrix, it can be seen that the increase in the estimate of the square of concentration parameter is

$$\frac{1}{n}\sum_{j=1}^{2}\underset{\sim}{Y}^{(j)'}(I-M)N(I-M)\,\underset{\sim}{Y}^{(j)},$$

exactly the same as the reduction in error sum of squares obtained in Equation (8.6.9).

Asymptotic Tests for the Determination of m

The following proposition can be proved as in Section 2.4 of Schmidt (1976) and we omit the details.

Proposition 8.1 *Let*

$$S_i = \underset{\sim}{Y}^{(i)'}(I-M)N(I-M)\,\underset{\sim}{Y}^{(i)},$$

$$\Sigma = Cov(\underset{\sim}{\epsilon}) = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ & \sigma_2^2 \end{pmatrix},$$

and assume that $X'X/n \to Q$ finite, nonsingular as $n \to \infty$. Then,

$$(i) \quad X' \underset{\sim}{\epsilon}^{(i)}/\sqrt{n} \xrightarrow{\mathcal{L}} N(0, \sigma_i^2\, Q),$$

$$(ii) \quad \sqrt{n}\left(\hat{\underset{\sim}{\lambda}}^{(i)} - \underset{\sim}{\lambda}^{(i)}\right) \xrightarrow{\mathcal{L}} N(0, \sigma_i^2 Q^{-1}).$$

Since $R_0(i,i)/[n - \overline{2m+1}]$ is a consistent estimator of σ_i^2

$$\frac{n-(2m+1)}{R_0(i,i)}\left(\hat{\underset{\sim}{\lambda}}^{(i)}_{(n)} - \underset{\sim}{\lambda}^{(i)}\right)' (X'X)\left(\hat{\underset{\sim}{\lambda}}^{(i)} - \underset{\sim}{\lambda}^{(i)}\right) \xrightarrow{\mathcal{L}} \chi^2(2m+1).$$

□

Let ${}_m\underset{\sim}{\lambda}^{(i)\prime}_{(1)} = (\lambda^{(i)}_{(1),2m+2}\ \lambda^{(i)}_{(1),2m+3})$ be the last two components of $\lambda^{(i)}_{(1)}$. We have ${}_m\hat{\underset{\sim}{\lambda}}^{(i)}_{(1)} = H^{-1}W(I-M)\ \underset{\sim}{\lambda}$ which is unbiased for ${}_m\underset{\sim}{\lambda}^{(i)}_{(1)}$ and has dispersion matrix $\sigma_i^2 H^{-1}$. To test $H_0^{(i)} : \underset{\sim}{\lambda}^{(i)}_{(1)} = 0$, we can use

$$\begin{aligned} T^{(i)} &= \frac{n-(2m+1)}{R_0(i,2)}\ {}_m\hat{\underset{\sim}{\lambda}}^{(i)\prime}_{(1)}\ H_m\, \hat{\underset{\sim}{\lambda}}^{(i)}_{(1)} \\ &= n-(2m+1)\frac{\left[\underset{\sim}{\lambda}^{(i)\prime}(I-M)N(I-M)\ \underset{\sim}{\lambda}^{(i)}\right]}{\underset{\sim}{\lambda}^{(i)\prime}(I-M)\ \underset{\sim}{\lambda}^{(i)}} \\ &\sim \chi^2(2) \end{aligned}$$

under H_0. Thus for any of the three criteria proposed above to determine m, tests are based on $T^{(i)}$. We settle for degree m if $H_0^{(i)}$, $i = 1, 2$ are both accepted.

8.7 Circular-Linear Regression

Here a linear variable y depends on an independent circular variable α, e.g. wind direction may affect the amount of rainfall. Suppose we assume that independent variable has single period. When there are multiple periods corresponding modifications have to be made.

A simple model for the periodic phenomenon is suggested by

$$y = A_0 + A_1 \cos w(\alpha - \alpha_0), \tag{8.7.1}$$

where α has period say T. Here A_0 is the mean level, A_1 is the amplitude, w is the angular frequency $2\pi/T$, and α_0 the acrophase. The model can also be written as

$$Y = A_0 + A_1 \cos(w\alpha - \varphi), \quad \text{where } \varphi = w\alpha_0. \tag{8.7.2}$$

Note that φ is the point where y reaches its highest peak and is called acrophase. As before we have the linear model

$$y_i = A_0 + A_1 \cos(w\alpha_i - \alpha_0) + \epsilon_i$$

and determine the constants by the method of least squares. We can generalize this model to

$$y = A_0 + A_1 \cos(w\alpha - \varphi_1) + A_2 \cos(2w\alpha - \varphi_2) + \cdots + A_k \cos(kw\alpha - \varphi_k)$$

with multiple angular frequencies. Note that this algebraically reduces to the same computations as the circular-circular regression considered above. Here in addition to overall period $T = 2\pi/w$, there are smaller periods $T/2, T/3 \cdots$ to fit into the overall period. In this model, the number of terms k can be determined from considerations discussed before in the choice of m.

In cases when one peak and one trough per cycle are present but the oscillations are distorted from those of a sinusoidal curve so that peaks and troughs no longer follow each other at equal intervals, it is referred to as having skewed oscillations. The model

$$Y = A_0 + A_1 \cos(\psi + \nu \cos \psi), \tag{8.7.3}$$

where $\psi = (w\alpha - \varphi)$, has been used for representing moderately skew oscillations. Here a new parameter ν appears. For $\nu = 0$, we have the simple cosine curve. The more ν deviates from 0, the more skew the graph turns out to be. So ν is called the *parameter of skewness*. The parameter ν is generally limited to the interval $-30^o \leq \nu \leq 30^o$ (or the equivalent in radians) as for values of ν beyond this interval the graph will not be as smooth.

The angle $\psi = \psi_0$ is called the acrophase angle for which (8.7.3) reaches its maximum $A_0 + A_1$. To determine ψ_0,

$$\frac{dy}{d\psi} = -\sin(\psi + \nu \cos \psi)(1 - \nu \sin \psi) = 0.$$

Since the first factor vanishes when $\psi_0 + \nu \cos\psi_0 = 0$, ψ_0 can be solved to satisfy this equation. In this model, the distance from a peak to the consecutive trough is $\delta = \pi - 2\psi_0$.

Another kind of distortion in the sinusoidal oscillation occurs when in a cycle the peak appears between two troughs and is distorted to be sharp or flat and there is distortion in the troughs also. Such a situation is modelled by taking

$$Y = A_0 + A_1 \cos(\psi + \nu_1 \sin\psi). \tag{8.7.4}$$

Here the parameter of peakedness is restricted to $-\pi/3 \leq \nu_1 \leq \pi/3$ to avoid secondary peaks and troughs.

This situation where y is a periodic function depending on a circular variable α, can also be modeled using periodic splines. See Section 12.10.

8.8 Linear-Circular Regression

Finally we discuss the case when the explanatory or *independent variable X, is linear* and the *response variable θ, is circular*, e.g., the direction of a fault plane on displacement, etc. Here the dependence of θ on X can be modeled by either ensuring that its mean direction, concentration or both depended on X. Gould (1969) proposes the model

$$\mu = \mu_0 + \sum \beta_j\ x_j (\text{mod } 2\pi)$$

for the mean direction.

Johnson and Wehrly (1978) suggest a specific model for the joint distribution of θ and a linear variable x, with a completely specified marginal distribution $F(x)$. The conditional distribution of θ given x is given by $CN(\mu + 2\pi F(x), \kappa)$, a model which allows direct estimation of μ and κ by the method of ML estimation. Another choice would be to take θ to be $CN(\mu, \kappa x)$ conditional on $X = x$.

In general, not only the mean direction of θ given $X_1 = x_1, \ldots, X_k = x_k$ of explanatory variables, depends on $x_1, \ldots, x_k$ but also the dispersion or concentration of θ. Thus suppose $(\theta_1, \underset{\sim}{X}_1), \ldots, (\theta_n, \underset{\sim}{X}_n)$ is a set of independent observations, $\theta_i | X_i = x_i \sim CN(\mu_i, \kappa_i)$, $i = 1, \ldots, n$.

(i) First suppose $\theta_i \sim CN(\mu_i, \kappa)$ with the concentration not depending on

x_i, and assume

$$\mu_i = \mu + g\left(\underset{\sim}{\beta}' \underset{\sim}{x_i}\right), \quad i = 1, \ldots, n.$$

The function $g(\cdot)$ is called *link function* which maps $\mathbb{R}^k$ to $(-\pi, \pi)$. In order that μ has an interpretation as origin we will also assume $g(0) = 0$. A link function may have known form except for $\underset{\sim}{\beta}$. One can for example use

$$g(x) = 2\ \arctan(|x|^{\lambda} \cdot sgn(x)).$$

(ii) Given a specification of either μ or κ in terms of $\underset{\sim}{x}$, one can apply the ML methods to obtain the required estimates. Often such ML estimates requires iteratively re-weighted and numerical algorithms.

Example 8.2 *We use the program* `circ.reg` *of the* `CircStats` *package on the data given in Example 8.1, to analyze the circular regression of wind direction at noon-time as a function of the wind direction at 6:00 am.*

```
> theta_c(356,97,211,232,343,292,157,302,335,302,324,85,324,340,
+ 157,238,254,146,232,122,329)
> phi_c(119,162,221,259,270,29,97,292,40,313,94,45,47,108,221,
+ 270,119,248,270,45,23)
> rtheta_rad(theta)
> rphi_rad(phi)
> rtheta
 [1] 6.213372 1.692969 3.682645 4.049164 5.986479 5.096361
 [7] 2.740167 5.270894 5.846853 5.270894 5.654867 1.483530
 [13] 5.654867 5.934119 2.740167 4.153884 4.433136 2.548181
 [19] 4.049164 2.129302 5.742133
> rphi
 [1] 2.0769418 2.8274334 3.8571776 4.5204028 4.7123890 0.5061455
 [7] 1.6929694 5.0963614 0.6981317 5.4628806 1.6406095 0.7853982
 [13] 0.8203047 1.8849556 3.8571776 4.7123890 2.0769418 4.3284165
 [19] 4.7123890 0.7853982 0.4014257
> wind.clm_circ.reg(rtheta,rphi,order=1,level=0.05)
> wind.clm$rho
```

```
          [,1]
[1,] 0.502289
> wind.clm$fitted
        1        2        3        4         5        6         7
 1.058793 1.759751 4.343291 4.496017 0.9506717 5.738549 3.702566
        8        9       10         11        12        13       14
 6.270684 0.863580 6.270684 0.6971781 1.618877 0.697178 0.920412
       15       16       17        18        19        20        21
 3.702566 4.544043 4.700515 3.371617 4.496017 2.343416 0.7815045
> wind.clm$coef
                    [,1]      [,2]
(Intercept)  0.01441183 0.1145101
  cos.alpha  0.33465348 0.5494121
  sin.alpha -0.07478683 0.4821004
> wind.clm$pvalues
             p1        p2
[1,] 0.09971087 0.9474847
> wind.clm$A.k
[1] 0.5063665

message = "Higher order terms are not significant at the 0.05
level"
```

8.9 Testing Stochastic Independence of Circular Variables

In this section, we discuss a test of stochastic independence for measurements on a torus, i.e., bivariate circular data.

Suppose the observations $(X_1, Y_1), \ldots, (X_n, Y_n)$ are from a continuous d.f. (distribution function) $F(x, y)$ whose marginal d.f.'s are $F_1(x)$ and $F_2(y)$. The problem is to test the hypothesis

$$H_0 : F(x, y) = F(x)F(y) \;\; \forall (x, y).$$

With respect to the given origins for the two variates, let $F_{n1}(x)$, $F_{n2}(y)$ and $F_n(x, y)$ denote the empirical d.f.'s of the X's, Y's and (X, Y)'s, respectively.

That is

$$
\begin{aligned}
F_{n1}(x) &= \frac{1}{n}\sum_{i=1}^{n} I(X_i; x), \\
F_{n2}(y) &= \frac{1}{n}\sum_{i=1}^{n} I(Y_i; y), \\
F_n(x, y) &= \frac{1}{n}\sum_{i=1}^{n} I(X_i; x) I(Y_i; y),
\end{aligned}
$$

where

$$
I(s; t) = \begin{cases} 1 & \text{if } s \leq t, \\ 0 & \text{if } s > t. \end{cases}
$$

Distribution-free tests of the form

$$
\int_{-\infty}^{\infty} \int -\infty^{\infty} T_n^2(x, y) \mathrm{d}F_n(x, y),
$$

where

$$
T_n(x, y) = [F_n(x, y) - F_{n1}(x) F_{n2}(y)]
$$

were considered earlier by Blum et al. (1961). However since they are not invariant under different choices of the origins for X and Y, they are not applicable to the circular case. To circumvent this problem, Rothman (1971) suggested the modified statistic

$$
C_n = n \int_0^{2\pi} \int_0^{2\pi} Z_n^2(x, y) \mathrm{d}F_n(x, y),
$$

where

$$
\begin{aligned}
Z_n(x, y) &= \left[T_n(x, y) + \int_0^{2\pi} \int_0^{2\pi} T_n(x, y) \mathrm{d}F(x, y) \right. \\
&\quad \left. - \int_0^{2\pi} T_n(x, y) \mathrm{d}F_{n1}(x) - \int_0^{2\pi} T_n(x, y) \mathrm{d}F_{n2}(y) \right].
\end{aligned}
$$

The statistic C_n has the desired invariance property and its asymptotic distribution theory under the null hypothesis has been derived by Rothman (1971).

Let $\mathcal{F}$ denote the family of probability distributions on the circumference $[0,\ 2\pi)$ of the circle with the property $F(\alpha + \pi) = F(\alpha) + 1/2$ for all α.

For instance circular distributions with axial symmetry would be in this class. Rao and Puri (1977) proposed a test which is applicable to testing independence when the marginals F_1 and F_2 belong to this class $\mathcal{F}$. When dealing with axial data, each observed axis will be represented by both its antipodal points for the purposes of this test. The proposed test may also be applied for testing independence in the non-axial case. But in this case, corresponding to every observed direction, we should add its antipodal point also to the data thus doubling the original sample size. Thus from now on, the random sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ corresponds to the axial data with both ends represented as two distinct sample points or the "doubled" sample in the general non-axial case. This ensures that the marginal distributions F_1 and F_2 belong to class $\mathcal{F}$.

As before, let (X_1, Y_1), ..., (X_n, Y_n) denote a random sample of angular variates on the basis of which we wish to test the null hypothesis of independence. For any fixed x in $[0, 2\pi)$, let $N_1(x)$ denote the number of X_i's that fall in the half-circle $[x, x + \pi)$. Similarly, for any fixed y in $[0, 2\pi)$, let $N_2(x)$ denote the number of Y_i's that fall in the half-circle $[y, y + \pi)$. Also let $N(x, y)$ denote the number of observations (X_i, Y_i) that fall in the quadrant $[x, x + \pi) \times [y, y + \pi)$. Now defining the indicator variables

$$I_i(x) = \begin{cases} 1 & \text{if } X_i \in [x, x + \pi), \\ 0 & \text{otherwise} \end{cases} \tag{8.9.1}$$

and

$$\bar{I}_i(y) = \begin{cases} 1 & \text{if } Y_i \in [y, y + \pi), \\ 0 & \text{otherwise,} \end{cases} \tag{8.9.2}$$

we obtain

$$N_1(x) = \sum_{i=1}^{n} I_i(x), \quad N_2(y) = \sum_{i=1}^{n} \bar{I}_i(y), \tag{8.9.3}$$

$$N(x, y) = \sum_{i=1}^{n} I_i(x)\bar{I}_i(y).$$

If the hypothesis of independence holds, then we should have

$$N(x, y) = \frac{N_1(x)N_2(y)}{n}$$

by the usual arguments. Thus we define

$$N_n(x, y) = \frac{1}{\sqrt{n}} \left[N(x, y) - \frac{N_1(x)N_2(y)}{n} \right] \tag{8.9.4}$$

as a measure of discrepancy between the observed and expected (under the hypothesis of independence) frequencies. Since $T_n(x, y)$ depends specifically on the choices of x and y, we suggest the (invariant) statistic

$$\begin{aligned} T_n &= \int_0^{2\pi}\int_0^{2\pi} D_n^2(x,y)\frac{dx}{2\pi}\frac{dy}{2\pi} \\ &= \frac{1}{4\pi^2 n}\int_0^{2\pi}\int_0^{2\pi}\left[N(x,y) - \frac{N_1(x)N_2(y)}{n}\right]^2 dxdy \end{aligned}$$

for testing independence. The integrand $D_n^2(x, y)$ is much like the usual chi-square test for independence from a 2×2 table. We now derive a computational form for T_n in terms of the X and Y spacings.

In view of (8.9.1)–(8.9.3), we have

$$\begin{aligned} nD_n^2(x,y) &= \left[\left(1-\frac{1}{n}\right)\left(\sum_{i=1}^{n} I_i(x)\bar{I}_i(y) - \frac{1}{n}\left(\sum_{i\neq j}^{n} I_i(x)\bar{I}_i(y)\right)\right)\right]^2 \\ &= \left(1-\frac{1}{n}\right)^2\left\{\sum I_i\bar{I}_i + \sum I_{ij}\bar{I}_{ij}\right\} \\ &\quad -\frac{2}{n}\left(1-\frac{1}{n}\right)\left\{\sum I_i\bar{I}_{ij} + \sum I_{ij}\bar{I}_i + \sum I_{ij}\bar{I}_{ik}\right\} \\ &\quad +\frac{1}{n^2}\left\{\sum I_i\bar{I}_j + \sum I_{ij}\bar{I}_k + \sum I_i\bar{I}_{jk} + \sum I_{ij}\bar{I}_{kl}\right\}, \end{aligned}$$

where $I_{ij} = I_i(x)I_j(x)$, $\bar{I}_{ij} = \bar{I}_i(y)\bar{I}_j(y)$ and summations are over all distinct subscripts. It is easy to check that

$$\begin{aligned} &\textstyle\int_0^{2\pi} I_i(x)dx = \int_0^{2\pi} \bar{I}_i(y)dy = \pi, \\ &\textstyle\int_0^{2\pi} I_i(x)I_j(x)dx = \pi - D_{ij}, \\ &\textstyle\int_0^{2\pi} \bar{I}_i(x)\bar{I}_j(x)dx = \pi - \bar{D}_{ij}, \end{aligned}$$

where D_{ij} and $\bar{D}_{ij}$ are the "circular" distances between (X_i, X_j) and (Y_i, Y_j), respectively. Omitting the routine computations, we obtain

$$\begin{aligned}
T_n &= \frac{1}{4\pi^2 n}\Bigg[(n-1)\pi^2 - \frac{\pi}{n}\left\{\sum(\pi - D_{ij}) + \sum(\pi - \bar{D}_{ij})\right\} \\
&\quad + \left(1-\frac{1}{n}\right)^2 \left\{\sum(\pi - D_{ij})(\pi - \bar{D}_{ik})\right\} \\
&\quad - \frac{2}{n}\left(1-\frac{1}{n}\right)\left\{\sum(\pi - D_{ij})(\pi - \bar{D}_{ik})\right\} \\
&\quad + \frac{1}{n^2}\left\{\sum(\pi - D_{ij})(\pi - \bar{D}_{k1})\right\}\Bigg],
\end{aligned}$$

again the summations run over all the distinct subscripts. The statistic T_n, being a function of the circular distances $\{D_{ij}\}$ and $\{\bar{D}_{ij}\}$, is clearly invariant under rotations of either coordinates axis. Rao and Puri (1977) show that T_n has asymptotically the same distribution as the sum of squares of independent normal variables X_{km} with zero means and variances $(\pi^4 k^2 m^2)^{-1}$ for odd k and m. Thus the asymptotic characteristic function of T_n is given by

$$\begin{aligned}
\varphi(t) &= \prod_{\text{k odd}} \prod_{\text{m odd}} \left(1 - \frac{2it}{\pi^4 k^2 m^2}\right)^{-1/2} \\
&= \prod_{k=1}^{\infty} \prod_{m=1}^{\infty} \left(1 - \frac{2it}{\pi^4 (2k-1)^2 (2m-1)^2}\right)^{-2}.
\end{aligned}$$

This characteristic function can be formally inverted as in Rao (1972b). On the other hand by a result of Zolotarev (1961), if $F(x)$ denotes the asymptotic d.f. of T_n, then the upper tail probabilities relating to T_n may be approximated as follows:

$$\lim_{x\to\infty} \frac{1-F(x)}{\mathbf{P}[\chi_4^2 > \pi^4 x]} = \prod_{k=1}^{\infty} \prod_{m=1}^{\infty} \left(1 - \frac{1}{(2k-1)^2(2m-1)^2}\right)^{-2}, \quad (k,m) \neq (1,1),$$

where χ_4^2 denotes a random variable having the chi-square distribution with 4 degrees of freedom.

This page is intentionally left blank

Chapter 9

Predictive Inference for Directional Data

9.1 Introduction

In this chapter, we discuss the predictive density for a future observation when the given data is assumed to come from certain circular model like the von Mises. These predictive densities are shown to be "model-robust" i.e., have good performance even when used for other circular models. Predictive intervals corresponding to the highest posterior density regions are derived and illustrated with an example. The results stated here as well as extensions to the spherical case, are to be found in Jammalamadaka and SenGupta (1998).

If $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ is the "past" and has density $p_\theta(\boldsymbol{y})$, what can we say about the distribution of the "future" –a single value, a set of values or a function of these, say $\boldsymbol{z}$? Some claim that such prediction is the single most important aim of statistics. There are many approaches to finding predictive likelihoods and densities. See Butler (1986) and Bjornstad (1990) for reviews. Main approaches include (i) methods based on conditioning through sufficiency, (ii) Bayes predictive densities and (iii) maximum or profile likelihood methods.

Specifically, let $\boldsymbol{Y} = (Y_1, \ldots, Y_n) \sim p_\theta(\boldsymbol{y})$ and let $\boldsymbol{Y}' = (Y_1', \ldots, Y_m')$ denote future observations from the same model. The quantity of interest $\boldsymbol{Z}$ may be $\boldsymbol{Y}'$ itself or a function of $\boldsymbol{Y}'$. Let $p_\theta(\boldsymbol{y}, \boldsymbol{z})$ denote the pdf of $(\boldsymbol{Y}, \boldsymbol{Z})$. It can be argued that this is the likelihood of the unknowns $(\boldsymbol{Z}, \theta)$ given the observed $\boldsymbol{y}$ i.e.,

$$L_{\boldsymbol{y}}(\boldsymbol{z},\ \theta) = p_\theta(\boldsymbol{y},\ \boldsymbol{z}).$$

The goal is to eliminate the nuisance parameter(s) θ and obtain the predictive density of $\boldsymbol{Z}$ given $\boldsymbol{Y}$. Three such methods are:

(i) **Sufficiency Approach**: Suppose $T(\boldsymbol{Y},\ \boldsymbol{Z})$ is a minimal sufficient statistic for θ based on both $\boldsymbol{Y}$ and $\boldsymbol{Z}$. Then

$$p^{\text{cond}}(\boldsymbol{z} \mid \boldsymbol{y}) \propto \frac{p(\boldsymbol{y},\ \boldsymbol{z},\ \theta)}{p(T(\boldsymbol{y},\ \boldsymbol{z}),\ \theta)} = p(\boldsymbol{y},\ \boldsymbol{z} \mid T(\boldsymbol{y},\ \boldsymbol{z})),$$

where factorization criterion makes θ drop out of the last expression.

(ii) **Bayes Approach**:

$$p^{\text{Bayes}}(\boldsymbol{z} \mid \boldsymbol{y}) = \frac{\int p_\theta(\boldsymbol{z},\ \boldsymbol{y})\pi(\theta)d\theta}{\int p_\theta(\boldsymbol{y})\pi(\theta)d\theta},$$

where $\pi(\theta)$ is an assumed prior for θ.

(iii) **Maximization**:

$$p^{\max}(\boldsymbol{z}|\boldsymbol{y}) = \sup_\theta p_\theta(\boldsymbol{y},\ \boldsymbol{z}) = L_{\boldsymbol{y}}(\boldsymbol{z},\ \widehat{\theta}(\boldsymbol{y},\ \boldsymbol{z})),$$

where $\widehat{\theta}(\boldsymbol{y},\ \boldsymbol{z})$ is the MLE based on both $\boldsymbol{y}$ and $\boldsymbol{z}$.

We apply some of these approaches to circular data and obtain the predictive density of a future observation, given n past observations from the same population. We use the Bayesian approach in Section 9.4 while the rest of the paper uses the Sufficiency approach. For the circular case, we derive the predictive densities assuming that the underlying distribution is the von Mises or Circular Normal. Some generalizations that include the multimodal as well as axial distributions, are also considered. We show that conditional sufficiency argument of Fisher (1959) leads to exact solutions. We also observe that the Bayesian approach which often becomes intractable, leads to an elegant and computationally appealing representation in this case for large samples.

For CN data, we find that the predictive density is symmetric around the mean of the past data and is nearly CN although less concentrated. We also demonstrate through exact numerical computations as well as simulations that this predictive density derived under the assumption of CN population is robust against a large class of other symmetric unimodal circular distributions, the symmetric wrapped stable family (see density in (2.2.18)). We compute the Highest Posterior Density credible sets (HPDs) and illustrate the results with an example from a real-life data set.

9.2 Prediction Under a von Mises Model

We consider the data as having come from the most commonly used circular distribution – the von Mises distribution (or the Circular Normal distribution), CN(μ_0, κ) with the pdf

$$f(\theta; \mu_0, \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp\{\kappa \cos(\theta - \mu_0)\}, \qquad 0 \le \theta,\ \mu_0 < 2\pi, \qquad \kappa > 0.$$

Theorem 9.1 *Under the sufficiency approach,*

(i) The predictive density of θ_{n+1} is given by

$$g(\theta_{n+1}|\theta_1, \ldots, \theta_n) \propto \frac{1}{\psi_{n+1}(r_{n+1})}, \tag{9.2.1}$$

where $\psi_n(r)$ is as defined in (3.2.5).

(ii) $g(\cdot)$ above is a symmetric, unimodal density with its mode at $\overline{\theta}_n$, the MLE of μ_0 based on the past n observations, $(\theta_1,\ \theta_2,\ \ldots,\ \theta_n)$.

(iii) The predictive density, $g(\cdot)$, is proportional to the von Mises model, $f(\theta_{n+1};\ \overline{\theta}_n,\ 2\overline{R})$ for large n.

Proof: As we have seen before, in this model,

$$T_n = (C_n,\ S_n)$$

with

$$S_n = \sum_{i=1}^{n} \sin\theta_i \text{ and } C_n = \sum_{i=1}^{n} \cos\theta_i$$

is a minimal sufficient statistic whose distribution is given by (see Equation 3.3.5)

$$p(S_n,\ C_n) = \frac{1}{2\pi I_0^n(\kappa)} \exp\{\kappa(\nu S_n + \mu C_n)\}\, \psi_n(\sqrt{S_n^2 + C_n^2})$$

with $\nu = \sin\mu_0$, $\mu = \cos\mu_0$.

(i) Following the conditioning approach and using the minimal sufficient statistic $(C_{n+1},\ S_{n+1})$,

$$\begin{aligned} p^{\text{cond}}(\theta_{n+1}|\theta_1,\ \dots,\ \theta_n) &= \frac{p(\theta_1,\ \dots,\ \theta_{n+1})}{p(C_{n+1},\ S_{n+1})} \\ &= \left(\frac{1}{2\pi}\right)^n \frac{1}{\psi_{n+1}(\sqrt{C_{n+1}^2 + S_{n+1}^2})} \end{aligned} \tag{9.2.2}$$

after some simplification, which proves (i).

(ii) Noting

$$C_{n+1} = C_n + \cos\theta_{n+1},\ S_{n+1} = S_n + \sin\theta_{n+1}$$

and

$$C_n = R_n \cos\overline{\theta}_n,\ S_n = R_n \sin\overline{\theta}_n,$$

where $\overline{\theta}_n$ is the direction of the resultant vector $(C_n,\ S_n)$, the expression (9.2.2) is equal to

$$\left[(2\pi)^n \psi_{n+1}(\sqrt{R_n^2 + 1 + 2R_n\cos(\theta_{n+1} - \overline{\theta}_n)})\right]^{-1}$$

and the statement made in (ii) follows from this.

(iii) From Rayleigh's approximation in Equation (3.4.2), it follows that

$$\psi_n(r) \approx \frac{2}{n}\exp\left(-\frac{r^2}{n}\right)$$

for large n. Using this, it can be seen that the predictive density for large n is proportional to

$$\exp\left\{\left(-\frac{2r_n}{n+1}\right)\cos\left(\theta_{n+1} - \overline{\theta}_n\right)\right\}$$

which is essentially a von Mises distribution with center at $\overline{\theta}_n$, the estimate of μ based on the past. Indeed, $\hat{\kappa} = 2r_n/(n+1)$ can be seen to be the approximate MLE of the concentration parameter κ for smaller values of κ. The predictive density is less concentrated around $\overline{\theta}_n$ than the "plug-in" (but incorrect) method would suggest, namely a $\text{CN}(\hat{\theta}_n,\ \hat{\kappa}_{\text{MLE}})$.

□

9.3 Generalized von Mises-type Distributions

9.3.1 ℓ-Modal von Mises Distribution

A multi(ℓ)-modal von Mises density CNM(μ_0, κ, ℓ) may be obtained (see Chapter 2) from the von Mises density and has the pdf

$$p_M(\theta) = \frac{1}{2\pi I_0(\kappa)} \exp\left\{\kappa \cos \ell(\theta - \mu_0)\right\},\ 0 \leq \theta < 2\pi,\ 0 \leq \mu_0 < 2\pi/\ell. \tag{9.3.1}$$

For known ℓ, a sufficient statistic for $(\mu_0,\ \kappa)$ is $(S_\ell,\ C_\ell)$, where

$$S_\ell = \sum_{i=1}^{n} \sin \ell\theta_i \text{ and } C_\ell = \sum_{i=1}^{n} \cos \ell\theta_i.$$

Let

$$R_\ell^2 = S_\ell^2 + C_\ell^2.$$

We may thus apply the above approach here also, provided the joint density of $(S_\ell,\ C_\ell)$ can be derived. Towards this end, we have:

Lemma 9.1 *The joint density of $(S_\ell,\ C_\ell)$, where $\alpha_1, \ldots, \alpha_n$ are i.i.d. observations from the CNM(μ_0, κ, ℓ) distribution is:*

$$g(s_\ell,\ c_\ell) = \frac{1}{2\pi\ I_0^n(\kappa)} \exp(\kappa\nu s_\ell + \kappa\mu c_\ell)\mathcal{L}_n(r_\ell^2), \tag{9.3.2}$$

where $\nu = \sin\mu_0$, $\mu = \cos\mu_0$ *and*

$$\mathcal{L}_n(r_\ell^2) = \ell \int_0^\infty J_0(ur_\ell)\ J_0^n(u)udu = \ell\psi_n(r_\ell). \tag{9.3.3}$$

Proof: Let

$$x_\ell = r_\ell \cos \ell\theta,\ y_\ell = r_\ell \sin \ell\theta$$

be the rectangular coordinate representation of an arbitrary two-dimensional variable. If $\psi(t_1,\ t_2)$ represents its characteristic function, then by the inversion theorem, the pdf of $(x_\ell,\ y_\ell)$ is given by,

$$f(x_\ell,\ y_\ell) = \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp(-it_1x - it_2y)\psi(t_1,\ t_2)dt_1\ dt_2. \tag{9.3.4}$$

Writing $t_1 = \rho \cos \ell\phi$, $t_2 = \rho \sin \ell\phi$,

$$\psi(t_1, t_2) = E\left[\exp\{i\rho r_\ell \cos \ell(\theta - \phi)\}\right] \equiv \overline{\psi}_\ell(\rho, \phi),$$

say. The joint density of r_ℓ and θ is

$$\begin{aligned} & f_1(r_\ell, \theta) \\ = & \frac{r_\ell \ell}{(2\pi)^2} \int_0^{2\pi} \int_0^{\infty} \exp\{-i\rho\, r_\ell \cos \ell(\theta - \phi)\}\rho\, \overline{\psi}_\ell(\rho, \phi) d\rho\, d\phi. \end{aligned} \tag{9.3.5}$$

Then making a change of variables $(x_\ell, y_\ell) \to (r_\ell, \theta)$ in (9.3.5) and integrating out θ after interchanging the order of integration, the density of r_ℓ is given by

$$\begin{aligned} & f_1(r_\ell) \\ = & \frac{r_\ell \ell}{(2\pi)^2} \int_0^{\infty} \int_0^{2\pi} \left[\int_0^{2\pi} \exp\{-i\rho\, r_\ell \cos \ell(\theta - \phi)\} d\theta\right] \rho\, \overline{\psi}_\ell(\rho, \phi) d\phi\, d\rho \\ = & \frac{r_\ell \ell}{2\pi} \int_0^{\infty} \int_0^{2\pi} \rho J_0\left(\rho\, r_\ell\right) \overline{\psi}_\ell(\rho, \phi) d\phi\, d\rho. \end{aligned} \tag{9.3.6}$$

Now, let $\theta_1, \ldots, \theta_n$ be i.i.d. random variables following the uniform distribution on $[0, 2\pi)$. Then the c.f. of $(\cos \ell\theta, \sin \ell\theta)$ is

$$\overline{\psi}_\ell^*(\rho) = \overline{\psi}_\ell(\rho, \phi)|_{\phi,\ r_\ell = 1} \equiv J_0(\rho)$$

and from independence, the joint c.f. of (S_ℓ, C_ℓ) is $\left[\overline{\psi}_\ell^*(\rho)\right]^n$. Inverting this as in Equation (9.3.4) (see also Section 3.3.3), one obtains the pdf of R_ℓ as,

$$\begin{aligned} p\left(r_\ell\right) &= \frac{r_\ell \ell}{2\pi} \int_0^{\infty} \int_0^{2\pi} \rho\, J_0\left(-\rho r_\ell\right) \left[\overline{\psi}_\ell^*(\rho)\right]^n d\phi\, d\rho \\ &= \frac{r_\ell \ell}{2\pi} \int_0^{\infty} \int_0^{2\pi} \rho\, J_0\left(-\rho r_\ell\right) J_0^n(\rho) d\phi\, d\rho \\ &= r_\ell\, \mathcal{L}_n(r_\ell^2). \end{aligned} \tag{9.3.7}$$

Let

$$S_\ell = \sum_{i=1}^{n} \sin \ell\theta_i = R_\ell \sin \overline{x}_\ell,\ C_\ell = \sum_{i=1}^{n} \cos \ell\theta_i = R_\ell \cos \overline{x}_\ell.$$

From the discussion in Section 3.3 it is seen that for uniform samples, the distribution of $\overline{x}_\ell$ is uniform and $\overline{x}_\ell$ and R_ℓ are independent. Thus the joint pdf of $(\overline{x}_\ell, R_\ell)$ is

$$\frac{r_\ell \mathcal{L}_n(r_\ell^2)}{2\pi}, \quad 0 \le \overline{x}_\ell < 2\pi, \; R_\ell > 0. \tag{9.3.8}$$

Thus, the joint pdf of $(S_\ell,\ C_\ell)$ on transforming, is

$$g_0(s_\ell,\ c_\ell) = \frac{\mathcal{L}_n(r_\ell^2)}{2\pi}. \tag{9.3.9}$$

Now consider the general case where $\theta_1, \ldots, \theta_n$ are i.i.d. random variables following the multimodal von Mises distribution CNM $(\mu_0,\ \kappa,\ \ell)$. The joint pdf $g_\kappa(s_\ell,\ c_\ell)$ of $(S_\ell,\ C_\ell)$ is then

$$\begin{aligned} & g_\kappa(s_\ell,\ c_\ell) \\ = \ & \frac{1}{\{I_0(\kappa)\}^n} \exp\{\kappa(\nu s_\ell + \mu c_\ell)\} \\ & \times \int_{\{\theta:\ \sum_i \sin l\theta_i = s_\ell,\ \sum_i \cos l\theta_i = c_\ell\}} \prod_{i=1}^{n} (d\theta_i/2\pi). \end{aligned} \tag{9.3.10}$$

But the integral represents the density of $(C_\ell,\ S_\ell)$ when $\theta_1, \ldots, \theta_n$ is a random sample from the circular uniform distribution on $[0,\ 2\pi)$, which is given in (9.3.9). On substituting (9.3.9) into (9.3.10), (9.3.2) follows. □

We thus have the following:

Theorem 9.2 *The predictive density of θ_{n+1} is given by*

$$g\left(\theta_{n+1}|\theta_1, \ldots, \theta_n\right) \propto \frac{1}{\mathcal{L}_{n+1}(r_{\ell,n+1}^2)}.$$

Proof: This follows exactly as in part (1) of Theorem 9.1 by virtue of the above Lemma. □

Remark 9.1 *Arnold (1941) suggested an angular distribution of the von Mises type for a circular random variable restricted to only a semi-circular arc, i.e. $\theta \in (0, \pi]$. Its generalization to the t-angular case is given by*

$$g(\theta^*) = \frac{t}{2\pi I_0(\kappa)} exp\left\{\kappa \cos t(\theta^* - \mu_0)\right\}, \; 0 \le \theta^*, \; \mu_0 < 2\pi/t.$$

The predictive density of θ^* is equivalent to that in part (i) of Theorem 9.1 above as $\theta = t\theta^* \sim CN(\mu_0,\ \kappa)$, and t is known.

9.4 Bayesian Predictive Density

In this section we discuss Bayes prediction in the general context of directional vectors in p-dimensions, noting that $p = 2$ corresponds to the circular. The Bayes predictive density of a future observation u_{n+1} given the past $u_1, \ldots, u_n$ is given by

$$g(u_{n+1}|u_1,\ \ldots,\ u_n) = \int_\kappa \int_\mu p(u_{n+1}|\mu,\ \kappa)p(\mu,\ \kappa|u_1, \ldots,\ u_n)d\mu\, d\kappa. \quad (9.4.1)$$

Assume that the priors are given by

$$p(\mu) = C_p(\kappa_0)\exp\left(\kappa_0\mu'\mu_0\right)$$

and

$$p(\kappa) \propto \frac{1}{\{I_0(\kappa)\}^c}\exp(\kappa\Delta)$$

and they are independent. Defining

$$R = \|\sum_1^n u_i\|, \quad u_0 = \frac{1}{R}\sum_1^n u_i,$$

we have

$$\begin{aligned} g(u_{n+1}|u_1,\ \ldots,\ u_n) &\propto \int_0^\infty \int_0^{2\pi} C_p(\kappa)\exp(\kappa\mu' u_{n+1})C_p^n(\kappa)\exp(\kappa R\mu' u_0) \\ &\qquad \times C_p(\kappa_0)\exp(\kappa_0\mu_0'\mu)I_0^{-C}(\kappa)\exp(\kappa\Delta)d\mu d\kappa, \\ &= K\int_0^\infty C_p^{n+1}(\kappa)C_p^{-1}(r_2)I_0^{-C}(\kappa)\exp(\kappa\Delta)d\kappa, \end{aligned}$$

where

$$R_2 = \|\kappa(u_{n+1} + Ru_0) + \kappa_0\mu_0\|, \quad C_p(\kappa) = \frac{\kappa^{p/2-1}}{(2\pi)^{p/2}I_{p/2-1}(\kappa)}$$

and K is the normalizing constant. When needed, K can be obtained numerically. $g(u_{n+1}|u_1, \ldots, u_n)$ above simplifies somewhat if the prior is taken to be

$$p(\mu_1, \kappa) \propto \frac{I(\mu \in S_p)}{\sqrt{\kappa}},$$

i.e., a non-informative or vague prior for μ and Jeffrey's prior for κ, which turns out to be not a proper prior here.

9.4.1 Case of Large κ

In case there is a prior reason to believe that only large values of κ need to be considered, some further simplifications result. These are somewhat analogous to the standard Laplace-saddle point type approximations of the posterior.

Let

$$p(\kappa) \propto \frac{1}{I_0^C(\kappa)} \exp(\kappa\Delta) I(\kappa > M), \qquad M > 0, \text{ large and known.}$$

For large κ, it is known (see Appendix A) that for all $p \geq 0$,

$$I_p(\kappa) \simeq \frac{\exp(\kappa)}{\sqrt{2\pi\kappa}}.$$

Using this approximation, we get

$$C_p(\kappa) \simeq \frac{\kappa^{\frac{p-1}{2}} \exp(-\kappa)}{(2\pi)^{(p-1)/2}}.$$

Then,

$$\begin{aligned} & g(u_{n+1}|u_1, ..., u_n) \\ \propto & \int_M^\infty \frac{\kappa^{((n+1)(p-1)+c)/2}}{\|\kappa(u_{n+1} + Ru_0) + \kappa_0\mu_0\|^{(p-1)/2}} \\ & \times \exp\left[\kappa\left\{(\Delta - n - c - 1) + \|\kappa(u_{n+1} + Ru_0) + \kappa_0\mu_0\|\right\}\right] d\kappa. \end{aligned}$$

Specializing to $p = 2$, and choosing $\kappa_0 = 0$ i.e. uniform prior for μ, this simplifies to

$$
\begin{aligned}
&g(u_{n+1}|u_1, ..., u_n)\\
\propto &\int_M^{\infty} \frac{\kappa^{(n+c+1)/2}}{\sqrt{\kappa}||(u_{n+1} + Ru_0)||} e^{\kappa[(\triangle-n-c-1)+||u_{n+1}+Ru_0||]} d\kappa\\
= &\frac{1}{\sqrt{||u_{n+1} + Ru_0||}} \int_M^{\infty} \kappa^{\frac{n+c}{2}} e^{\kappa[(\triangle-n-c-1)+||u_{n+1}+Ru_0||]} d\kappa,
\end{aligned}
\tag{9.4.2}
$$

where B is the normalizing constant. Note that, taking $C = 0$ and $\triangle = -1$ yields $p(\kappa) \propto \exp(-\kappa)$. For large n, (9.4.2) behaves as a usual incomplete gamma integral, i.e.,

$$
\begin{aligned}
&g(u_{n+1}|u_1, \ldots, u_n)\\
= &\frac{B}{\sqrt{||u_{n+1} + Ru_0||}} \Gamma_{\frac{n+c+2}{2}} \left(\frac{M}{n + c + 1 - \triangle - ||u_{n+1} + Ru_0||} \right)
\end{aligned}
\tag{9.4.3}
$$

where B is the normalizing constant, i.e.,

$$
\begin{aligned}
&\frac{1}{B}\\
= &\int_{S_p} \frac{1}{\sqrt{||u_{n+1} + Ru_0||}} \Gamma_{\frac{n+c+2}{2}} \left(\frac{M}{n + c + 1 - \triangle - ||u_{n+1} + Ru_0||} \right) du_{n+1}.
\end{aligned}
$$

(9.4.3) is quite simple to calculate, especially for the CN population with $p = 2$. Such Bayes predictive densities can be numerically evaluated and provide sensible and useful results.

9.5 Robustness Against Symmetric Wrapped Stable Family

Exact computation of the predictive densities is done through numerical integrations and are plotted in Figures 9.1 and 9.2. The predictive density curve is super-imposed on the true CN curve in Figure 9.1 while Figure 9.2 illustrates what happens when the predictive density of Theorem 9.1, derived based on the CN distribution is used when the data is actually from the symmetric wrapped stable family with $\alpha = 1.5$. This and other results clearly

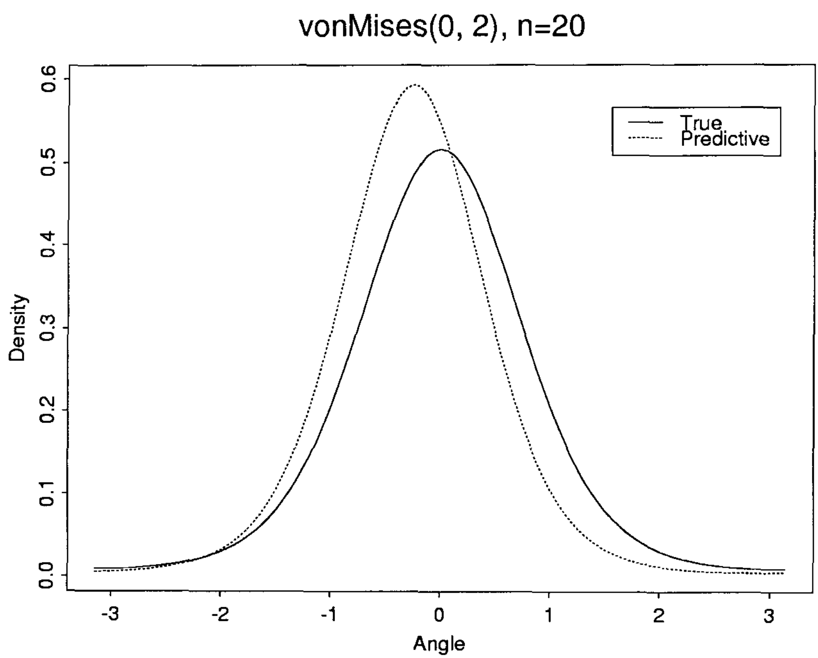


Figure 9.1: Predictive density for von Mises (0, 2), $n = 20$.

indicate the encouraging performance and model-robustness of the predictive density as derived above. As may be expected, these results on how close the predictive density is to the true density as well on model-robustness, improve considerably for larger sample sizes.

9.6 HPD Computation and An Example

Consider the following example (Batschelet (1981)) giving data on the time of day of major traffic accidents in a major city. The time (h) is converted to angle (η) (in degrees) by the formula $\eta = 15 \times h$. The $n = 21$ observations are given below in terms of degrees:
14, 47, 73, 109, 122, 150, 171, 182, 202, 214, 245, 251, 256, 260, 261, 272, 274, 284, 293, 313, 332.

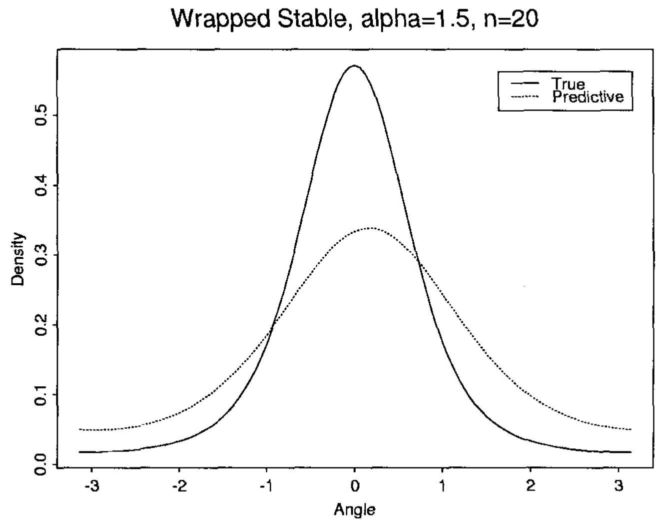


Figure 9.2: Predictive density for Wrapped Stable with $\alpha = 1.5$, $n = 20$.

Assuming a von Mises distribution of the data, we get $\hat{\mu} = 248.59^o$ which translates to 16hr 34min in time scale. Also, $\hat{\kappa} = 0.7021$ is the MLE of the concentration parameter. The 90% HPD credible set is given by $(104.50^o, 32.68^o)$ i.e. (6hr 58min, 2hr 11min) while the 75% HPD is given by (9hr 45min, 23hr 24min).

The large widths of the predictive intervals in this example can be explained by the fact that the data is not very concentrated and the sample size is rather small.

Chapter 10

Outliers and Related Problems

10.1 Introduction

In practice, the observed directions may contain one or more data points which appear to be peculiar, not representative or inconsistent relative to the main part of the data. Such data points may come from a model which differs from the one that generated the main part of the data. Recording errors, sudden short external shocks, sampling under abnormal conditions, etc. can give rise to such data. Such problems can be expected to occur rather infrequently but often enough to deserve consideration. This may include directional data, such as with wind directions, direction of movement of icebergs, propagation of cracks, periodic phenomena, etc. This is commonly referred to as the outlier, slippage, discordancy or spuriousity problem. The goal is to detect such outliers and to develop robust methods which take into account the possible existence of outliers. See for instance Chapter 11 of Barnett and Lewis (1994) for a survey.

The outlier problem in directional data is somewhat different from that in the linear case. Clearly, how *far* an observation is from the mean in directional data setup, should be judged by using the appropriate "Circular distance". Thus, unlike in the linear case, outliers here need not be too large or too small, but could be in the "central" part of the data.

Initially one may use a simple yet fast diagnostic tool for the detection of outliers. One such tool is a simple P-P plot of the data for the CN distribution which can be obtained using the routine `pp.plot` in the `CircStats` package. We then discuss two formal approaches to the outlier problem. In the first approach the original model is extended to accommodate outliers so that the extended model can provide a good fit to the entire data set. A formal test is performed to test for the presence of outliers. In the event that the presence of outliers cannot be ruled out, one should adopt this extended model for further inference. Ideally one should develop procedures for the original model which are robust against such extended models. In this approach, the outlier(s) may be thought of as appearing with a certain unknown probability from a distribution with a quite different functional form than the rest i.e., the extended model is a mixture model. In case the test for zero mixing proportion leads to rejection, it is desirable for the inference procedures developed for the original uncontaminated model to be robust against the mixture model. Such an approach may be ambitious as robust procedures are not always easy to develop. An alternate approach is to detect the outliers, simply delete these and proceed with the original model for the remaining data. In this approach, the outlier may be thought of arising due to a 'slip' in the value of a parameter of the underlying distribution. Here, we consider specifically the situation where there is a possible slip in the mean direction μ of an underlying $CN(\mu, \kappa)$ population. While the mixture models are discussed in Section 10.3, the slippage and outlier approach is covered in Section 10.4.

10.2 Diagnostic Tools

A fast and simple exploratory data analytic technique can be quite useful for initial identification of the outliers before embarking on the formal statistical tests. Even a plot of the data can sometimes be revealing. One can measure the "circular distance" as defined in Equation (1.3.6) between any observation and its neighbors on either side. The observation θ_j may be suspected to be an outlier if *both* of these are found to be 'large' indicating that this point is somewhat isolated, compared to the remaining values.

A circular P-P plot is another simple graphical way to detect outliers. For testing for a Circular Uniform distribution, one simply plots

$$(j/(n+1), \theta_{(j)}/2\pi), j = 1, ..., n,$$

where $\theta_{(j)}$ are the ordered observations with respect to any origin and sense of rotation. For the CN distribution, a P-P plot is easier to obtain and is done this way; First find the best-fitting CN distribution, say $\hat{F}(\theta; \hat{\mu}, \hat{\kappa})$ to the data, and then plot

$$(j/(n+1), \hat{F}(\theta_{(j)}; \hat{\mu}, \hat{\kappa})), j = 1, ..., n.$$

A Q-Q plot is harder to get but an approximate plot is obtained by calculating the sample quantiles $q_i = F^{-1}(i/(n+1); 0, \hat{\kappa})$, finding $z_i = \sin(\theta_i - \hat{\mu})/2$, $i = 1, ..., n$, and then plotting

$$\left(\sin(q_i/2), z_{(i)}\right), \ i = 1, ..., n,$$

where $z_{(1)}, ..., z_{(n)}$ are the ordered values. See Fisher (1993), p.53 who discusses an iterative procedure for finding q_i. Such a P-P plot can be obtained by using the program `pp.plot` in `CircStats`. If a few points on this circular P-P plot seems to be quite away from the 45° diagonal, one may suspect these to be possible outliers.

Example 10.1 *We wish to examine if the data on vanishing angles of pigeons from the Schmidt-Koenig experiment represented in Figure 1.2, follows a CN distribution and if so, whether there are any outliers:*

```
> bird_c(85, 135, 135, 140, 145, 150, 150, 150, 160, 185, 200,
 + 210, 220, 225, 270)
> pp.plot(bird)
        mu       kappa
1  4.612744 0.2092706
```

It is clear both from Figure 1.2 and the P-P plot above, that the data does not provide a good fit to a CN distribution.

10.3 Tests for Mixtures

10.3.1 Introduction

Consider now the case of the mixture model which is a common method of treating outliers. For example, a CN-CU mixture model has been used by Guttorp and Lockhart (1988) to provide a Bayesian solution of the problem of detecting the location of a downed aircraft from distress signals transmitted by it, subject to possible sudden disruptions in environmental and

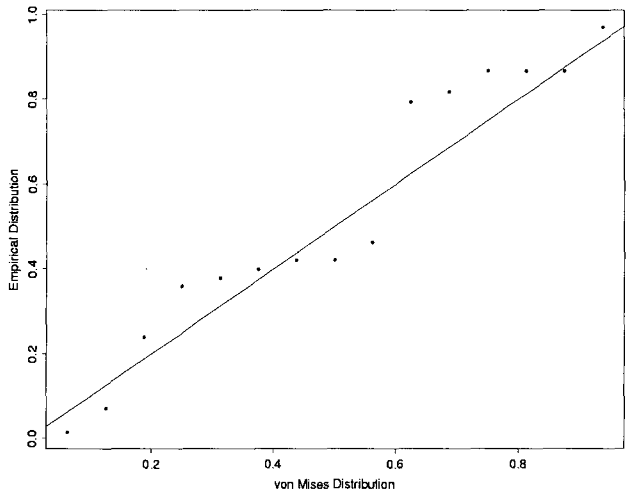


Figure 10.1: P-P plot of vanishing angles of pigeons released in Schmidt-Koenig experiment.

atmospheric conditions. We adopt here a member of the WSM family (see Section 2.2.9) to model circular data with possible outliers or contamination. By taking a WSM model, we have a rich variety of symmetric distribution while the parameter p, the "mixing proportion", provides an intuitive interpretation as the probability of an outlier.

Let

$$f(\theta;\ \alpha,\ \rho,\ \mu_0) = \frac{1}{2\pi} + \frac{1}{\pi}\sum_{k=1}^{\infty} \rho^{k^\alpha} \cos\{k(\theta - \mu_0)\} \tag{10.3.1}$$

denote the WS density as given in Equation (2.2.18) and

$$g(\theta; \alpha, \rho, \mu_0, p) = pf(\theta; \alpha, \rho, \mu_0) + (1-p)\frac{1}{2\pi} \tag{10.3.2}$$

the WSM. Here $0 \leq \theta,\ \mu_0 < 2\pi;\quad 0 \leq \rho \leq 1,\quad 0 < \alpha \leq 2$ and $0 \leq p \leq 1, q = (1-p)$. The density $g(.)$, when the contaminating distribution is CU, often occurs naturally as in the case of distress signals referred above. Other examples arise e.g., in connection with experiments on perception of a group of subjects, say insects, birds or human beings, for movements towards a given target direction. In Example 10.2, we see that a WS distribution gives a much better fit than a CN distribution and a WSM model gave an even better fit. Contamination of CU by WS distributions may also serve as models in human perception tests, e.g. in traffic engineering, where it is generally observed that most of the individuals tend to move randomly, save a few who have rather strong perception, say after undergoing some "training".

Here we derive optimal tests for possible outliers using the null hypothesis $H_0 : p = 0$ against $H_1 : p > 0$. Rejection of H_0 will suggest the model to be an isotropic distribution contaminated by some member of the WS family of distributions when p is small or the reverse scenario when p is large. The derivation of the LMPI test when μ is unknown but ρ and α are known is presented in Section 10.3.2. Although Beran (1969) considers a much more general formulation of the problem, we present below a more direct and explicit derivation of the optimal test and its properties for the WSM mixtures.

The asymptotic null and non-null distributions of the test statistic are derived in Section 10.3.3 and Section 10.3.4 demonstrate the global monotonicity of the power function and consistency of the test. Cut-off points and power are discussed in Section 10.3.3 and a short table of critical values given in Table 10.1. The non-regular situations arising when some or all of the parameters are unknown are dealt with in Section 10.3.5. An example justifying the proposed model and the applicability of the LMPI test is also given.

10.3.2 The LMPI Test

Consider the WSM model discussed in Section 2.2.9. For now we will assume that ρ and α are known while μ_0 and p are unknown. In practice either or both of ρ and α may be unknown, this is dealt with later on. Suppose $\theta_1, \theta_2, \ldots, \theta_n$ is a random sample of size $n (\geq 2)$ from a population with density given by Equation (10.3.2). We want to test the hypothesis

$$H_0 : p = 0 \quad \text{against} \quad H_1 : p > 0.$$

Denoting by

$$R_k^2 = \left(\sum_{j=1}^{n} \cos\ k\theta_j\right)^2 + \left(\sum_{j=1}^{n} \sin\ k\theta_j\right)^2,$$

the form of the LMPI test is derived in the following.

Theorem 10.1 *For known α and ρ, the LMPI test for $H_0 : p = 0$ against $H_1 : p > 0$ is given by rejecting H_0 when*

$$T = \sum_{k=1}^{\infty} (\rho^2)^{k^\alpha}\ R_k^2 > C,$$

where the constant C is to be determined from the size condition.

Proof: First note that the problem of testing H_0 against H_1 remains invariant under the change of location $\theta_i \to \theta_i + c \pmod{2\pi}$. Based on the maximal invariant $(\theta_1 - \theta_n, \theta_2 - \theta_n, \ldots, \theta_{n-1} - \theta_n)$, the most powerful invariant test for H_0 against a fixed $p > 0$ is given by the test statistic

$$T^*(p) = \int_0^{2\pi} \prod_{i=1}^{n} \left[\frac{p}{2\pi}\left\{1 + 2\sum_{k=1}^{\infty} \rho^{k^\alpha} \cos k(x + \theta_i)\right\} + \frac{q}{2\pi}\right] dx. \quad (10.3.3)$$

To get the LMPI test we expand $T^*(p)$ in powers of p and consider the lowest order random term. Now the right-hand side of Equation (10.3.3) is

$$\left(\frac{1}{\pi}\right)^n \int_0^{2\pi} \prod_{i=1}^{n} \left\{\frac{1}{2} + p\ Y_i(x)\right\} dx, \quad (10.3.4)$$

where

$$Y_i(x) = \sum_{k=1}^{\infty} \rho^{k^\alpha} \cos k(x + \theta_i).$$

The coefficient of p in (10.3.4) is, apart from a multiplicative constant,

$$\sum_{i=1}^{n} \int_0^{2\pi} Y_i(x)dx = \sum_{i=1}^{n} \left[\int_0^{2\pi} \left\{\sum_{k=1}^{\infty} \rho^{k^\alpha} \cos k(x + \theta_i)\right\} dx\right].$$

The signs of summation and integration in the square bracket of the above expression are seen to be interchangeable by virtue of an extended version of

Levi theorem for series of functions (see Theorem 10.26 of Apostol (1974), p. 269). Thus the coefficient of p is zero, while the coefficient of p^2 barring again a multiplicative constant, is

$$\begin{aligned}
&\sum_{i<j} \int_0^{2\pi} Y_i(x) Y_j(x) dx \\
= &\sum_{i<j} \int_0^{2\pi} \left[\sum_{k=1}^{\infty} \sum_{k'=1}^{\infty} \rho^{k^\alpha + k'^\alpha} \cos k(x+\theta_i) \ \cos k'(x+\theta_j) \right] dx \\
= &\ 2\pi \sum_{i<j} \sum_{k=1}^{\infty} (\rho^2)^{k^\alpha} \cos k(\theta_i - \theta_j)
\end{aligned}$$

again by interchanging the summation and integration and some algebraic manipulations. We can then write

$$T^*(p) = c_1 + c_2 p^2 \sum_{i<j} \sum_{k=1}^{\infty} (\rho^2)^{k^\alpha} \cos k(\theta_i - \theta_j) + o_p(p^2),$$

where c_1 and $c_2 (> 0)$ are constants. The LMPI test is therefore to reject when

$$T = \sum_{i<j} \sum_{k=1}^{\infty} (\rho^2)^{k^\alpha} \cos k(\theta_i - \theta_j) = \sum_{k=1}^{\infty} (\rho^2)^{k^\alpha} R_k^2 > C, \qquad (10.3.5)$$

as claimed. □

An easy consequence of this result is the

Corollary 10.1 *The LMPI test for $H_0 : p = 0$ against $H_1 : p > 0$ in the Cardioid-CU mixture family corresponds to the statistic T with $k = 1$ and therefore coincides with the Rayleigh's R test.*

10.3.3 Asymptotic Distribution of T

The asymptotic null and non-null distributions of the test statistic are now obtained by directly appealing to the multivariate Central Limit Theorem (CLT) and Lemma 3.1. They can also be obtained as a consequence of the results in Beran (1969).

Theorem 10.2 *Under H_0, the asymptotic distribution of $2T/n$ is the same as the distribution of*

$$\sum_{k=1}^{\infty}(\rho^2)^{k^\alpha}W_k,$$

where $\{W_k, k = 1, 2, \ldots\}$ is a sequence of independent χ^2 random variables with 2 d.f. *each.*

Proof: Straightforward calculation shows that under the null hypothesis of circular uniformity for θ,

$$\begin{aligned}
\mathrm{E}(\sin\ k\theta) &= \mathrm{E}(\cos\ k\theta) = 0;\\
\mathrm{Var}(\sin\ k\theta) &= \mathrm{Var}(\cos\ k\theta) = \tfrac{1}{2}, \forall n = 1, 2, \ldots;\\
\mathrm{Cov}(\sin\ k\theta, \sin\ k'\theta) &= \mathrm{Cov}(\sin\ k\theta, \cos\ k'\theta) = \mathrm{Cov}(\cos\ k\theta, \cos\ k'\theta) = 0;
\end{aligned}$$

for each $k, k' = 1, 2, \ldots$, $k \neq k'$. Consequently for any k, by multivariate CLT

$$\frac{1}{\sqrt{n}}\begin{pmatrix} \sum_{i=1}^n \cos\theta_i \\ \sum_{i=1}^n \sin\theta_i \\ \sum_{i=1}^n \cos 2\theta_i \\ \sum_{i=1}^n \sin 2\theta_i \\ \cdots\cdots\cdots \\ \sum_{i=1}^n \cos k\theta_i \\ \sum_{i=1}^n \sin k\theta_i \end{pmatrix} \xrightarrow{\mathcal{L}} N_{2k}\left(\mathbf{0}, \boldsymbol{\Sigma}\right),$$

where

$$\boldsymbol{\Sigma}_{(2k\times 2k)} = \frac{1}{2}\begin{pmatrix} I_2 & \mathbf{O} & \ldots & \mathbf{O} \\ \mathbf{O} & I_2 & \ldots & \mathbf{O} \\ \ldots & \ldots & \ldots & \ldots \\ \mathbf{O} & \mathbf{O} & \ldots & I_2 \end{pmatrix}.$$

This shows that the limiting distribution of $2T/n = 2\sum_{k=1}^{\infty}(\rho^2)^{k^\alpha}R_k^2/n$ is the same as the distribution of $\sum_{k=1}^{\infty}(\rho^2)^{k^\alpha}W_k$, where W_k, $k = 1, 2, \ldots$ is a sequence of independent χ^2 variables each with 2 *d.f.* □

First note that since T is invariant under change of location, one may take without loss of generality, $\mu_0 = 0$. After some routine algebra it follows that under the mixture alternative with density $g(\theta; \alpha, \rho, 0, p)$ and for any

$k = 1, 2, \ldots,$

$$\begin{aligned}
\mathrm{E}(\cos k\theta) &= p\rho^{k^\alpha}, \\
\mathrm{E}(\sin k\theta) &= 0, \\
\mathrm{Var}(\cos k\theta) &= \frac{1}{2}\left[1 + p\rho^{(2k)^\alpha}\right] - p^2(\rho^2)^{k^\alpha}, \\
\mathrm{Var}(\sin k\theta) &= \frac{1}{2}\left[1 - p\rho^{(2k)^\alpha}\right], \\
\mathrm{Cov}(\cos k\theta, \sin k\theta) &= 0.
\end{aligned}$$

It also follows that for any $k, l = 1, 2, \ldots,$ $k \neq l$,

$$\begin{aligned}
\mathrm{Cov}(\sin k\theta, \sin l\theta) &= \frac{p}{2}\left[\rho^{|k-l|^\alpha} - \rho^{(k+l)^\alpha}\right], \\
\mathrm{Cov}(\sin k\theta, \cos l\theta) &= 0, \\
\mathrm{Cov}(\cos k\theta, \cos l\theta) &= \frac{p}{2}\left[\rho^{|k-l|^\alpha} + \rho^{(k+l)^\alpha}\right] - p^2\rho^{k^\alpha + l^\alpha}.
\end{aligned}$$

For $r, s = 1, 2, \ldots,$ writing

$$\begin{aligned}
\sigma_{rs}^{(CC)} &= \mathrm{Cov}(\rho^{r^\alpha}\cos r\theta, \rho^{s^\alpha}\cos s\theta), \\
\sigma_{rs}^{(CS)} &= \mathrm{Cov}(\rho^{r^\alpha}\cos r\theta, \rho^{s^\alpha}\sin s\theta), \\
\sigma_{rs}^{(SS)} &= \mathrm{Cov}(\rho^{r^\alpha}\sin r\theta, \rho^{s^\alpha}\sin s\theta)
\end{aligned}$$

and noting that $\sigma_{rs}^{(CS)} = 0 \ \forall\ r$ and s, one sees again by multivariate CLT that, under the mixture alternative,

$$\frac{1}{\sqrt{n}}\begin{pmatrix}
\rho\sum_{i=1}^n \cos\theta_i \ - \ np\rho^2 \\
\rho\sum_{i=1}^n \sin\theta_i \\
\rho^{2^\alpha}\sum_{i=1}^n \cos 2\theta_i \ - \ np(\rho^2)^{2^\alpha} \\
\rho^{2^\alpha}\sum_{i=1}^n \sin 2\theta_i \\
\cdots\cdots\cdots\cdots\cdots\cdots \\
\rho^{n^\alpha}\sum_{i=1}^n \cos k\theta_i \ - \ mp(\rho^2)^{k^\alpha} \\
\rho^{k^\alpha}\sum_{i=1}^n \sin k\theta_i
\end{pmatrix} \xrightarrow{\mathcal{L}} N_{2k}\left(\mathbf{0}, \boldsymbol{\Sigma}^*\right),$$

where

$$\Sigma^*{}_{(2k\times 2k)} = \begin{pmatrix} \sigma_{11}^{(CC)} & 0 & \sigma_{12}^{(CC)} & 0 & \dots & \sigma_{1k}^{(CC)} & 0 \\ 0 & \sigma_{11}^{(SS)} & 0 & \sigma_{12}^{(SS)} & \dots & 0 & \sigma_{1k}^{(SS)} \\ \sigma_{21}^{(CC)} & 0 & \sigma_{22}^{(CC)} & 0 & \dots & \sigma_{2k}^{(CC)} & 0 \\ 0 & \sigma_{21}^{(SS)} & 0 & \sigma_{22}^{(SS)} & \dots & 0 & \sigma_{2k}^{(SS)} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \sigma_{k1}^{(CC)} & 0 & \sigma_{k2}^{(CC)} & 0 & \dots & \sigma_{kk}^{(CC)} & 0 \\ 0 & \sigma_{k1}^{(SS)} & 0 & \sigma_{k2}^{(SS)} & \dots & 0 & \sigma_{kk}^{(SS)} \end{pmatrix}.$$

Now using Lemma 3.1 as $k \longrightarrow \infty$, one obtains the asymptotic distribution of T which is given by the following

Theorem 10.3 *Under the alternative hypothesis H_1,*

$$\frac{1}{\sqrt{n}}\left(T - (pn)^2\sum_{k=1}^{\infty}(\rho^2)^{k^\alpha}\right) \xrightarrow{\mathcal{L}} N\left(0, \sigma^{*2}\right) \quad \text{as } m \longrightarrow \infty, \tag{10.3.6}$$

where

$$\sigma^{*2} = 4p^2\sum_{r=1}^{\infty}\sum_{s=1}^{\infty}(\rho^2)^{r^\alpha+s^\alpha}\sigma_{rs}^{(CC)}. \tag{10.3.7}$$

For any $p > 0$ under the alternative hypothesis H_1, exact distributions of the test statistic T under both the null and the alternative hypotheses are analytically intractable. We use simulations to obtain exact cut-off points and power values of this LMPI test for different parameter and sample size combinations. These are given in Table 10.1.

Remark 10.1 *In particular under the null hypothesis of circular uniformity, the characteristic function corresponding to the limiting distribution of $2T/n$, given in Theorem 10.2, is*

$$\varphi(t) = \left\{\prod_{k=1}^{\infty}\left(1 - 2(\rho^2)^{k^\alpha}it\right)\right\}^{-1},$$

a first order approximation of which is $1/(1-2Kit)$ with

$$K = \sum_{k=1}^{\infty}(\rho^2)^{k^\alpha}.$$

This is clearly the characteristic function of $K \cdot Z$, *where* Z *has a chi-square distribution with 2 d.f. The values of the cut-off points obtained using this approximation are quite close to those obtained by simulation for some many parameter combinations. For example, for* $m = 20$, $\rho = .5$ *and* $\alpha = 1.5$, *this asymptotic approximation gives the 5% cut-off point calculated as 16.21 as compared to the simulated value of 16.08 from Table 10.1.*

Remark 10.2 *For each parameter and sample size combination, the cut-off points in Table 10.1 were computed on the basis of 5000 observations, on the statistic* T. *The first 30 terms of the series were used to approximate the value of* T. *The power values were also obtained through simulations (see SenGupta and Pal (2001b)) and indicate encouraging performance for 'reasonable' parameter combinations even for samples of size 20. It is also noted that the power increases with* ρ *for each fixed* m, α *and* p. *This is expected because if* α *and* p *are fixed, the larger the value of* ρ, *the more the deviation of the density* g *from circular uniformity and this fact should be reflected by any reasonable test.*

10.3.4 Monotonicity and Consistency Properties

Monotonicity of power and consistency of test statistic T given in Equation (10.3.5) are proved in the following

Theorem 10.4 *For any fixed* ρ *and* α, *the test given in Equation (10.3.5) possesses a monotone power function in* $p \in [0, 1]$. *Further the test is also consistent.*

Proof: To prove the first part of the theorem we need the following two facts along with Lemma 6.2.

Lemma 10.1 *The function* $1 + 2\sum_{n=1}^{\infty} \rho^{n^\alpha} \cos n\theta$ *is decreasing in* $0 \leq \theta < \pi$ *and by symmetry, increasing in* $\pi \leq \theta < 2\pi$ *for any* $0 < \rho < 1$ *and* $0 < \alpha \leq 2$.

Lemma 10.2 *(Wintner (1947), p. 591): For* $0 < \delta < \pi$, *a set of sufficient conditions for the series* $\sum_{n=1}^{\infty} b_n \sin n\delta$ *to be positive is (i)* $nb_n \to 0$ *as* $n \to \infty$ *and (ii)* $nb_n > (n+1)b_{n+1}$, $\forall n = 1, 2, \ldots$.

Table 10.1: 5% and 1% Cut-off points of the LMPI test based on T.

m	ρ	α			
		0.5	1.0	1.5	2.0
10	.25	(2.25,3.30)	(1.86,2.74)	(1.81,2.73)	(1.78,2.59)
	.50	(12.94,16.81)	(8.40,11.48)	(7.53,10.73)	(7.26,10.56)
	.75	(53.74,65.48)	(25.12,35.99)	(19.78,27.92)	(17.00,24.90)
	.90	(124.97,157.95)	(64.86,79.08)	(39.55,52.67)	(30.52,43.53)
15	.25	(3.36,4.67)	(2.84,4.17)	(2.69,4.05)	(2.61,4.00)
	.50	(19.33,25.85)	(12.96,17.84)	(11.74,17.64)	(11.35,16.26)
	.75	(81.41,96.91)	(37.46,50.37)	(28.29,40.89)	(26.78,39.99)
	.90	(189.15,225.82)	(97.02,119.74)	(58.53,81.79)	(46.15,60.80)
20	.25	(4.44,6.29)	(3.81,5.84)	(3.77,5.69)	(3.59,5.45)
	.50	(25.82,35.90)	(17.48,24.73)	(16.08,23.05)	(14.77,21.71)
	.75	(110.06,133.34)	(51.53,65.28)	(37.80,51.91)	(35.61,50.21)
	.90	(250.27,300.33)	(127.81,159.79)	(78.08,101.63)	(63.26,85.76)
30	.25	(6.54,9.42)	(5.66,8.35)	(5.29,8.09)	(5.27,7.67)
	.50	(40.17,52.45)	(24.66,36.26)	(23.44,35.44)	(21.13,30.38)
	.75	(163.61,199.44)	(78.37,105.04)	(62.08,94.85)	(53.25,82.42)
	.90	(376.02,442.52)	(196.16,240.38)	(119.42,156.11)	(94.53,131.99)

Also, note that the pdf of $\gamma_{ij} = \theta_i - \theta_j \pmod{2\pi}$, where θ_i and θ_j are independently distributed as (10.3.2), can be written as

$$h(\gamma_{ij}; p) = \frac{1}{2\pi}\left\{1 + 2\sum_{n=1}^{\infty} p^2(\rho^2)^{n^\alpha}\cos n\gamma_{ij}\right\}, \quad 0 \le \gamma_{ij} < 2\pi.$$

By Lemma 2.1 and by Lemma 10.1 for given c, $\exists$ a $\delta \in (0, \pi)$ such that

$$\sum_{n=1}^{\infty}(\rho^2)^{n^\alpha}\cos n\gamma_{ij} > c \iff 0 < \gamma_{ij} < \delta \text{ or } 2\pi - \delta < \gamma_{ij} < 2\pi.$$

Hence for any $i, j (i < j)$,

$$\begin{aligned} P(p) &= P\left\{\sum_{n=1}^{\infty}(\rho^2)^{n^\alpha} \cos n\gamma_{ij} > c \mid p\right\} \\ &= 2\int_0^\delta h(\gamma_{ij}; p) d\gamma_{ij} \\ &= \frac{\delta}{\pi} + \frac{2p^2}{\pi} \int_0^\delta \sum_{n=1}^{\infty}(\rho^2)^{n^\alpha} \cos n\gamma_{ij}\ d\gamma_{ij}, \end{aligned}$$

and on differentiating with respect to p

$$\frac{\partial P}{\partial p} = \frac{4p}{\pi} \int_0^\delta \sum_{n=1}^{\infty}(\rho^2)^{n^\alpha} \cos n\gamma_{ij}\ d\gamma_{ij} = \frac{4p}{\pi}\sum_{n=1}^{\infty} b_n \ \sin\ n\delta$$

with $b_n = (\rho^2)^{n^\alpha}/n$, by interchanging again the order of summation and integration.

Thus $P(p)$ is increasing globally in $p \in [0, 1]$, by Lemma 10.2. Repeated application of Lemma 6.2 on the variables

$$\sum_{n=1}^{\infty}(\rho^2)^{n^\alpha} \cos n\gamma_{ij}$$

for all $i, j\ (i < j)$ then ensure the monotonicity of the power function.

To prove consistency, it suffices to check that

$$b(\theta; \alpha, \rho, \mu_0, p) = \int_0^{2\pi} \left[f(\theta; \alpha, \rho, x) - \frac{1}{2\pi}\right] g(x; \alpha, \rho, \mu_0, p) dx,$$

corresponding to any density $g(\theta; \alpha, \rho, \mu_0, p)$ under the alternative hypothesis is non-zero. Now

$$\begin{aligned} b(\theta; \alpha, \rho, \mu_0, p) &= \frac{1}{\pi}\int_0^{2\pi} \left[\sum_{n=1}^{\infty} \rho^{n^\alpha} \cos n(\theta - x)\right] \\ &\quad \times \left[\frac{1}{2\pi} + \frac{p}{\pi}\sum_{n=1}^{\infty} \rho^{n^\alpha} \cos n(x - \mu_0)\right] dx \\ &= \frac{p}{\pi}\sum_{n=1}^{\infty} (\rho^2)^{n^\alpha} \cos n(\theta - \mu_0), \end{aligned}$$

is non-zero. Consistency of the test now follows from Theorem 4 of Beran (1969). □

10.3.5 Non-regular Cases

The problem of testing for no contamination (or no outlier) in terms of $H_0 : p = 0$ poses non-regular situations when either or both of ρ and α are unknown. Since neither α nor ρ is a location or scale parameter, the usual approach of invariance or similarity does not apply. Further, neither an exact LMPI test, nor even an asymptotically LMP test, i.e., Neyman's C_α-test, can be constructed. This is because of the non-regular situation, whereby the nuisance parameter appears only under the alternative. We present below some techniques to deal with this situation.

For instance, when ρ is known but α is unknown, we encounter the non-regular problem of having the nuisance parameter α appearing only under the alternative. First consider the case when $\alpha \in [\epsilon, 2]$, where ϵ is a known small positive number. We can construct an optimal test for this case along the lines of Davies (1977). Observe that for each α, under H_0,

$$T'(\alpha) = \frac{2}{n}\sum_{k=1}^{\infty}(\rho^2)^{k^\alpha} R_k^2 \xrightarrow{\mathcal{L}} \sum_{k=1}^{\infty}(\rho^2)^{k^\alpha} W_k \text{ as } n \to \infty, \tag{10.3.8}$$

where $\{W_k, k = 1, 2, \ldots\}$ are as defined before, a sequence of independent χ^2 variables each with 2 *d.f.* The test then consists in rejecting H_0 for large values of $\sup_{\alpha\in[\epsilon,2]} T'(\alpha)$. Also observe that $T'(\alpha)$ is monotonically decreasing in α and hence

$$\sup_{\alpha\in[\epsilon,2]} T'(\alpha) = \frac{2}{n}\sum_{k=1}^{\infty}(\rho^2)^{k^\epsilon} R_k^2. \tag{10.3.9}$$

The significance probability of the test determined by the statistic (10.3.9), however, cannot be directly obtained from the results given in Davies (1977), since the asymptotic distribution considered there is that of a *single* χ^2. However the required significance probability can be calculated using the asymptotic distribution given in Equation (10.3.8) with α replaced by ϵ. From (2.10) of Beran (1969), it follows that the required probability can be expressed as

$$\sum_{k=1}^{\infty} \alpha_k \exp\left[-t/2(\rho^2)^{k^\epsilon}\right], \tag{10.3.10}$$

where

$$\alpha_k = \prod_{k \neq k'} \left[1 - (\rho^2)^{k^\epsilon - k'^\epsilon}\right]^{-1}$$

and t is the observed value of $\sup_{\alpha \in [\epsilon, 2]} T'(\alpha)$. When α is unrestricted and is in $(0, 2]$, the sup may be attained on the 'boundary' and an optimal solution is not readily available. Similar results follow when α is known and ρ is unknown but bounded above (and away from 1) and when ρ is unconstrained.

In the more general case, when both α and ρ are known to be strictly positive and bounded above by known constants, but otherwise unknown, a multiparameter generalization of the Davies test is called for. This seems to be an open problem.

Finally when all the parameters ρ, p as well as α and μ are unknown, the problem of testing for outliers becomes identical to that of testing for isotropy. The P^3 test is equivalent to one based on R^2. This test is L-optimal, robust and consistent as discussed in Chapter 5.

Example 10.2 *Batschelet (1981) p. 49, Fig. 2 depicted the orientation of ants towards a black target when released in a round arena – an experiment originally conducted by Jander (1957). We adapt the data to construct a grouped frequency distribution of angles with 36 classes of equal widths; the total frequency being 146.*

The tests for isotropy discussed in Sections 10.3.2 and 10.3.5 above are applied to these data assuming that WSM mixture model holds. Estimated values of the parameters are taken as the 'known' values in the respective cases to illustrate the application, although normally the 'known' values should be supplied in practical situations. Significance probability corresponding to the LMPI test has been computed using (10.3.10) of Section 10.3.5 with the determined value of α in place of ϵ. For the Davies-motivated test, the same formula has been used with $\epsilon = .1$, whereas the (asymptotic) significance probability for the Rayleigh's test has been calculated on the basis of a χ^2 distribution with 2 d.f.

Once uniformity is rejected, we search for a suitable model to fit the data. The parameters for fitting CN are estimated by the m.l. method and for the last two cases, the method of moments has been used, and these estimates are taken as the 'known' values. It is seen that all the three tests namely the LMPI, Rayleigh's R and Davis-type with $\epsilon = 0.1$, lead to rejection of the null hypothesis of circular uniformity of the data. It is also found that the

CN distribution does not give a good fit to the data, whereas a WSM mixture distribution provides an excellent fit.

10.4 Slippage Problem and Outliers

10.4.1 Introduction

The slippage problem is basically to detect whether any unspecified observation in a given random sample comes from a distribution different from that for all the other remaining observations. In a given sample the corresponding observation may manifest itself as an observation lying unduly away from the remaining set. However, such a manifestation may not always be apparent, save the knowledge of the possibility of its occurrence. One then needs both to infer on the occurrence and to identify the 'slipped' observation. This can also be viewed as a problem of "outlier" detection or that of "spuriousity". An outlier may occur in circular data due to e.g., a 'slip' in recording, a poorly 'trained'/'distracted' or highly 'brain-washed' individual in the test group, a physical 'distortion' in a segment of the otherwise homogeneous sampling site, a sudden 'shock' in the sampling environment, etc. In studies on homing abilities with biological subjects, unknown to the experimenter, the sample may draw a few from a differing 'guest' species, as could possibly be the case with Jander's ant data discussed in Section 10.3.5. A segment of the sandstone layer may be 'disorientated' due to geological or environmental disruptions which may give rise to possible outliers in the sample of paleocurrent orientations from that layer. Such a data set from Belford Anticline, New South Wales (Fisher and Lewis (1983)) is analyzed in Section 10.4.3. Related discussion can be found in Collett (1980), Bagchi (1987), Bagchi and Guttman (1990), Upton (1993) and Barnett and Lewis (1994). Unlike the preceding discussion on contamination or mixture formulation which helps to determine only the presence of outliers and not their detection, this alternate approach considered below aims at achieving both the results.

Next we consider the outlier problem in terms of a possible slip in the mean direction μ of a circular normal population $CN(\mu, \kappa)$. These results are from SenGupta and Laha (2000). Following a decision theoretic approach, in Section 10.4.2, we formulate a Bayes rule and the associated probabilities, which describe its performance. These results are contained in Theorems

10.5 and 10.6. Given i.i.d. observations, $\theta_1, \ldots, \theta_n$, the slippage problem is to test

$$H_0 : \theta_i,\ i = 1, \ldots, n \quad \text{are distributed as } CN(\mu_0, \kappa)$$

against

$$H_1^* : \text{For some } i, \theta_i \sim CN(\mu_1, \kappa) \text{ and } \theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_n \sim CN(\mu_0, \kappa).$$

The LRT statistics for testing H_0 vs H_1^* when the parameters μ_0 and μ_1 are both known or both unknown but κ is known, are presented in Theorem 10.7. While the exact null distribution of the LRT can be easily derived in the first case, simulations are needed to obtain the critical values for various κ in the later case. See Tables 10.3 and 10.4. These results are then illustrated through an example in Section 10.4.4. Finally, in Section 10.5, we provide brief comments on a simulation-based comparison of the various procedures for outlier detection.

10.4.2 A Decision Theoretic Approach

For $k \geq 3$, suppose $\theta_1, \ldots, \theta_k$ are independent $CN(\mu_i, \kappa)$ random variables where μ_i is either $= \mu_0$ or $= \mu_1$. Consider testing $H_0 : \theta_j, j = 1, ..., k$ are identically distributed as $CN(\mu_0, \kappa)$ against $H_i : \theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_k$ are identically distributed as $CN(\mu_0, \kappa)$ and θ_i is distributed as $CN(\mu_1, \kappa)$, $1 \leq i \leq k, \mu_1 > \mu_0$, μ_1, μ_0 and κ are all known. We are interested in finding the Bayes rule for the multiple decision problem of accepting one of the $k+1$ hypotheses $H_0, H_1, \ldots, H_k$ with respect to the prior distributions invariant under permutations of $H_1, \ldots, H_k$ using the loss function which assigns loss "0" if the correct hypothesis is accepted and loss "1" otherwise. The prior distributions invariant under permutations of $H_1, \ldots, H_k$ give equal weight to $H_1, \ldots, H_k$ and hence they are of the form τ_p, where

$$\begin{aligned} \tau_p(H_0) &= 1 - kp,\ 0 \leq p \leq 1/k, \\ \tau_p(H_i) &= p,\ 1 \leq i \leq k. \end{aligned}$$

Let

$$\Phi(\underline{\theta}) \equiv (\phi(1|\underline{\theta}), ..., \phi(k|\underline{\theta})),$$

be a generalized critical function or a multiple decision rule (Ferguson (1967)) with $\phi(i|\underline{\theta}), i = 1, ..., k$, taking values 0 or 1, and $\Sigma\phi(i|\underline{\theta}) = 1$. Thus Φ chooses

H_i when $\underline{\boldsymbol{\theta}} = \underline{\theta}$ is observed if $\phi(i|\underline{\theta}) = 1$. Let R_j be the likelihood ratio at $\boldsymbol{\theta}_j$ under μ_1 to μ_0, i.e.,

$$R_j = \frac{f(\boldsymbol{\theta}_j; \mu_1)}{f(\boldsymbol{\theta}_j; \mu_0)} = \exp\left[\kappa\left\{\cos(\boldsymbol{\theta}_j - \mu_1) - \cos(\boldsymbol{\theta}_j - \mu_0)\right\}\right].$$

The following result follows from the general theory given in Ferguson (1967) (pp. 299-306) after some simplifications. See also SenGupta and Laha (2000).

Theorem 10.5 *The Bayes rule with respect to τ_p for H_0 against $H_i, 1 \le i \le k$ is given by $\phi(0; \underline{\boldsymbol{\theta}}) = 0$ whenever $(1 - kp)/p < \max_j R_j$ and $\phi(i; \underline{\boldsymbol{\theta}}) = 0$ whenever either $R_i < \max_j R_j$ or $(1 - kp)/p > \max_j R_j, 1 \le i \le k$.*

□

Let

$$\begin{aligned} g_\eta(\theta) \;=\; & \Big[\exp\left\{\kappa\cos((\mu_1 - \mu_0)/2 + \pi - \arcsin(\ell n\ \theta))\right\} \\ & + \exp\left\{\kappa\cos((\mu_1 - \mu_0)/2 + \arcsin(\ell n\ \theta))\right\} \\ & + \exp\left\{\kappa\cos((\mu_1 - \mu_0)/2 + 3\pi - \sin^{-1}(\ell n\ \theta))\right\}\Big] \\ & \times \left[(2\pi I_0(\kappa))^2\ \left\{1 - (\ell n\ \theta)^2\right\}.\theta^2\right]^{-\frac{1}{2}}, \end{aligned} \tag{10.4.1}$$

for $1/e < \theta < e$ and

$$G(\eta) = \int_{-\infty}^{\eta} g_\eta(\theta) d\theta$$

be its cdf. The next theorem gives the performance of the Bayes rule given in Theorem 10.5 when H_0 and H_j are true.

Theorem 10.6 *In the framework of Theorem 10.5,*

(i) (a)

$$P\left(\phi(i; \underline{\boldsymbol{\theta}}) = 0 \mid H_0 \text{ is true}\right) = 1 - \frac{1}{k} + \frac{1}{k}G^k\left(\frac{1 - kp}{p}\right),$$

(b)

$$P\left(\phi(0; \underline{\boldsymbol{\theta}}) = 0 \mid H_0 \text{ is true}\right) = 1 - G^k\left(\frac{1 - kp}{p}\right).$$

(ii) For $1 \le j \le n$

(a)

$$
\begin{aligned}
& P\left(\phi(i;\underline{\boldsymbol{\theta}}) = 0 \mid H_j \textit{ is true}\right) \\
= \; & 1 + \int_{1/e}^{\frac{1-kp}{p}} G^{k-2}(w) G^{*}(w) dG(w) \\
& -(k-2) \int_{1/e}^{e} G^{k-3}(w)[1 - G(w)] G^{*}(w) dG(w) \\
& - \int_{1/e}^{e} G^{k-2}(w)[1 - G(w)] dG^{*}(w),
\end{aligned}
$$

where $i \neq j$,

(b)

$$
\begin{aligned}
& P\left(\phi(0;\underline{\boldsymbol{\theta}}) = 0 \mid H_j \textit{ is true}\right) \\
= \; & 1 - (k-1) \int_{1/e}^{\frac{1-kp}{p}} G^{k-2}(w) G^{*}(w) dG(w) \\
& - \int_{1/e}^{\frac{1-kp}{p}} G^{k-1}(w) dG^{*}(w),
\end{aligned}
$$

(c)

$$
\begin{aligned}
& P\left(\phi(j;\underline{\boldsymbol{\theta}}) = 0 \mid H_j \textit{ is true}\right) \\
= \; & \int_{1/e}^{\frac{1-kp}{p}} G^{k-1}(w) dG^{*}(w) + (k-1) \int_{1/e}^{e} G^{k-2}(w) G^{*}(w) dG(w).
\end{aligned}
$$

Proof:

(i) Observe that $R_j = \exp[2\kappa[\sin\{\boldsymbol{\theta}_j - (\mu_1 - \mu_0)/2]\} \sin(\mu_1 - \mu_0)/2$.

Define $\boldsymbol{\eta}_i = \exp\,[\sin\{\boldsymbol{\theta}_i - (\mu_1 + \mu_0)/2\}]$ for $i = 1, 2, \ldots, k$. Considering that branch of $\arcsin\theta$ which is monotone on $[(\pi/2), (3\pi/2)]$, and accordingly defining the inverse transformation from θ to η, after some calculations yield the density of η as $g_\eta(\eta_i)$ with $g_\eta(\cdot)$ as defined in Equation (10.4.1).

Define $W = \max_{\ell} \boldsymbol{\eta}_\ell$. Then,

$$P(\phi(i;\boldsymbol{\theta}) = 0) = P(\boldsymbol{\eta}_i < W) + P(\frac{1-kp}{p} > W) - P(\boldsymbol{\eta}_i < W < \frac{1-kp}{p}).$$

After some calculations we get,

$$P(\boldsymbol{\eta}_i < W < (1-kp)/p) = \frac{k-1}{k} G^k \left(\frac{1-kp}{p}\right)$$

from which (i) follows.

(ii) Let $\boldsymbol{\eta}_i$ be as in the proof of (i) and let G be its distribution function (d.f.) when $\boldsymbol{\theta}_i$ is $CN(\mu_0, \kappa)$ and let G^* denote its d.f. when $\boldsymbol{\theta}_i$ is $CN(\mu_1, \kappa)$. The proof here follows in the lines of the proof in (i), but with more tedious computations and after noting that under $H_i, \boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \ldots, \boldsymbol{\eta}_k$ are no longer i.i.d. In fact $\boldsymbol{\eta}_j$ has a different distribution than the rest. □

Remark 10.3 *The two probabilities given in (i) of Theorem 10.6 for the case of equal ignorance can be easily computed for the circular normal distribution. If prior information is asymmetric, these probabilities may still be similarly computed, but their expressions become quite complicated.*

10.4.3 The LRT

Consider now the likelihood ratio testing approach to this outlier problem. We want to test H_0 against H_1^* : There exists i, i unknown, such that $\boldsymbol{\theta}_i$ is distributed as $CN(\mu_1, \kappa)$ and $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ..., \boldsymbol{\theta}_{i-1}, \boldsymbol{\theta}_{i+1}, \ldots, \boldsymbol{\theta}_n$ are distributed as $CN(\mu_0, \kappa)$ and all are independent. Let $\mu_1 > \mu_0$. Let $H(\theta)$ be the distribution function of $\sin(\boldsymbol{\theta} - (\mu_0 + \mu_1)/2)$ under H_0.

Theorem 10.7 *Consider testing H_0 against H_1^*.*

(a) μ_0, μ_1 and κ are all known.

Let $H(\theta)$ be the distribution function of $\sin(\boldsymbol{\theta} - (\mu_0 + \mu_1)/2)$ under H_0. The LRT statistic is equivalent to

$$V = \max_{r} \sin(\boldsymbol{\theta}_r - \frac{\mu_0 + \mu_1}{2}).$$

The exact sampling distribution function of V under H_0 is given by $M(\theta) = [H(\theta)]^n$.

(b) *Only κ is known.*

Let ${}_0\hat{\mu}_0$ and ${}_1\hat{\mu}_0$ denote the maximum likelihood estimate of μ_0 under H_0 and H_1^*, respectively. Further let ${}_r f_1$ denotes the likelihood when there is a slip at $r, 1 \le r \le n$. Let $\hat{r}$ be that r for which ${}_r f_1$ attains its maximum. In testing H_0 against H_1^* the LRT statistic Λ is given by

$$\begin{aligned}\frac{1}{\Lambda} &= \exp\left[\kappa\left\{\left(\sum_{i\neq\hat{r}}\cos\boldsymbol{\theta}_i\right)(\cos\ {}_1\hat{\boldsymbol{\mu}}_0 - \cos\ {}_0\hat{\boldsymbol{\mu}}_0)\right.\right.\\ &\qquad \left.\left. + \left(\sum_{i\neq\hat{r}}\sin\boldsymbol{\theta}_i\right)(\sin\ {}_1\hat{\boldsymbol{\mu}}_0 - \sin\ {}_0\hat{\boldsymbol{\mu}}_0) + \cos\left(\boldsymbol{\theta}_{\hat{r}} - {}_0\hat{\boldsymbol{\mu}}_0\right)\right\}\right].\end{aligned}$$

Proof: The proof is fairly straightforward. For details the reader is referred to SenGupta and Laha (2000). □

Remark 10.4 *Observe that $H(\theta)$ can be evaluated numerically and hence the cutoff points for the LRT are readily available for case (a). However, the exact sampling distribution of LRT statistic is formidable - no closed form or even any analytic representation for it in small samples seems to be possible.*

We note that Collett (1980) considers a problem very similar to the above. He tests for no slippage versus a slippage alternative and derives a LRT statistic for it, which he calls the L-statistic. However the marked difference with our case is that, unlike here, he first uses a data based measure to detect the candidate for being the outlier. Subsequently formal tests are conducted to statistically validate the hypothesis that it is in fact an outlier. The choice of the measure may not be reasonable for even symmetric data sets where clusters may appear far from the mean direction. In the LRT procedure the detection and testing for the outlier is based on the LRT and is entirely probabilistic.

10.4.4 Simulations and An Example

We use simulations to obtain the null distribution of Λ (see Tables 10.3 and 10.4) as well as its power. The simulation results on null distribution are based on 5000 repetitions with sample size n, $n = 10, 20, 30$. For simulating the power of a particular observation is drawn from a circular normal population, $CN(\Delta, 1)$ with $\Delta = 20(20)180$ (in degrees) and repeating it 5000 times. Since the power function is symmetric about 180^o, the above computation is sufficient. From Tables 10.3 and 10.4, we observe that the null distribution of the test statistic gets increasingly concentrated with increase in sample size. Further the null distribution also gets concentrated around 0 as the value of κ increases.

Note that the assumptions involved in the large sample approximation for the loglikelihood as a Chi-square, are violated here. We obtain the power of the LRT with $\kappa = 1$ through extensive simulations. The power of this LRT, obtained through simulations, did not show any perceptible increase with increase in sample size.

The LRT exhibits encouraging power performance for somewhat small κ and the power increases to one with large κ. This is expected since with the higher values of the concentration parameter, the observations tend to be close together making it 'easier' for us to detect an outlier.

We illustrate the above tests with an example discussed in SenGupta and Laha (2000). The circular normal model gives acceptable fit to the dataset. We assume that the 'known' needed value of κ is $\hat{\kappa}$, the m.l.e. obtained from the data.

Example 10.3 *Fisher and Lewis (1983) give data from three samples of paleocurrent orientations from three bedded sandstone layers, measured on the Belford Anticline, New South Wales. The first of these data sets is (in degrees):*
284, 311, 334, 320, 294, 137, 123, 166, 143, 127, 244, 243, 152, 242, 143, 186, 263, 234, 209, 267, 315, 329, 235, 38, 241, 319,308, 127, 217, 245, 169, 161, 263, 209, 228, 168, 98, 278, 154, 279.
Our LRT picked up the observation with value 38 as an outlier with the computed value for $\Lambda = 0.1675$ *and with a P-value of 0.01. This may possibly be due to the fact that the segment of the sandstone layer corresponding to this outlier may have been disorientated by some abnormal geological event.*

Once an outlier has been detected as above, one may discard it and proceed with further statistical analyses with the original model say the CN, using the rest of the data set. Alternatively, one may use an extended model for the entire data set, say a contaminated or a mixture model that incorporates the outliers.

Table 10.2: 5% critical values for the LRT statistic for various κ

κ	n =10	n = 20	n = 30
0.5	0.3680	0.3679	0.3679
1.0	0.1357	0.1354	0.1354
1.5	0.0504	0.0499	0.0498
2.0	0.0199	0.0187	0.0185
4.0	0.0139	0.0062	0.0034
10.0	0.0210	0.0102	0.0064

Table 10.3: Percentiles of the null distribution of the LRT (for $\kappa = 1$)

Percentiles	n=10	n=20	n=30
5	0.1357	0.1354	0.1354
10	0.1367	0.1356	0.1354
25	0.1441	0.1371	0.1362
50	0.1786	0.1448	0.1398
75	0.2638	0.1707	0.1527
90	0.3884	0.2224	0.1774

10.5 Concluding Remarks

SenGupta and Laha (2000) discuss LRT and LMP tests for the slippage problem and provide a simulation based comparison of the various procedures for identifying an outlier. Compared were the L-statistic (Collett (1980)), M-statistic (Mardia (1975)), Bayes-Procedure, LRT and the LMP, using different criteria namely, the power of the procedure, the probability that an outlier is detected and correctly identified. The LMP appears to perform slightly better than the LRT when outliers of lesser severity are sought to be detected (i.e. μ_1 is small) which is to be expected, due to the nature of the LMP. However if we are interested in detecting outliers of moderate to large severity, the LRT performs much better than the LMP. Moreover, it was noted that the Bayes Procedure, the LRT and the LMP all perform better than the tests based on the L-statistic and the M-statistic.

Remark 10.5 *The discussion in this section was restricted to the simpler* *case when κ is known. The case where κ is unknown is quite challenging and no work in the classical framework appears to have been done. The outlier problem may also be considered in a Bayesian context as in Bagchi and Guttman (1990).*

Chapter 11

Change-point Problems

11.1 Introduction

Let $\alpha_1, \alpha_2, \ldots, \alpha_n$ denote a set of independent angular measurements observed say, as a time-ordered or space-ordered sequence. They may be wind directions over time, the vanishing angles at the horizon for a group of birds or the times of arrival at a hospital emergency room where the 24-hour cycle is represented as a circle. Such data may have one or more peaks and the goal is to check whether there has been a change in the preferred direction and/or concentration in the time-ordered or space-ordered data set, where the location of the change-point if any, is unknown. As an example, consider a meteorologist studying wind directions. Using previously gathered data, he/she might be interested in knowing if there has been a change in the mean direction of wind at some intermediate time point. Another interesting example is given in Lombard (1986) regarding evaluation of flares. Flares are launched upward attached to a projectile from a point O in a fixed direction. The quantity of interest is the latitude θ of the vector OP where P is the point at which the flare starts to burn. The stability of the flare-projectile assembly is reflected in an unchanging mean direction of θ.

Given the angular observations $\alpha_1, \ldots, \alpha_n$ we are interested in checking that for some unknown but fixed k, $(1 \leq k \leq n)$, $\alpha_1, \ldots, \alpha_k$ have the cdf F_1 while the next $(n-k)$ observations, $\alpha_{k+1}, \ldots, \alpha_n$ follow a different cdf $F_2 \quad (\neq F_1)$. Here, k is the point of change or the "change-point" for the data. If $k = n$, there are no observations from F_2, meaning all the observations are from the same population so that there is no change-point. We are interested in testing for the presence of a change-point i.e., the null-hypothesis of no-change, namely $H_0 : k = n$ versus $H_1 : 1 \leq k \leq n-1$.

Change-point problems have evoked considerable interest in the literature due to their application-oriented flavor. Page (1955), Sen and Srivastava (1975) and Horvath (1993) contain good discussions on the subject of change point problems on the line. The interest in change-point problems for directional data is relatively new. See for instance, Lombard (1986) and Csörgő and Horváth (1996), who consider this problem in non-parametric context. In this chapter, we discuss parametric tests for the change-point problem on the circle i.e., we assume that the two populations F_1 and F_2 are von Mises or CN with mean directions μ_1 and μ_2 and a common concentration parameter κ. Further details can be found in Ghosh et al. (1999).

Section 11.2 deals with the derivation of tests for changes in mean direction, both when κ is known and when it is unknown. In both cases, we use the generalized likelihood ratio method to derive tests for H_0 vs. H_1. We obtain the exact critical values of the test statistics through simulation. An alternative method with a Bayesian flavor, assumes that the change-point is equally likely to be at any one of the intermediate points. Hence, using a discrete uniform prior over the possible change-point values, we get an alternate test procedure. If we have further information about the possible point of change, we can build that into an appropriate prior on k and derive the corresponding Bayes procedure.

Simulated critical values of the test statistics, are presented in Section 11.2.3. They are provided as tables from which one can read the 5% and 1% critical values as well as nomograms for the 5% values. Section 11.2.4 provides a comparison of the powers of the two procedures, while in Section 11.2.5, we ascertain the model robustness of these procedures to the von Mises distributional assumption. These results indicate that although these tests are in some sense optimal for the CN data, they are indeed more widely applicable, being valid for most unimodal distributions. In Section 11.3 we consider changes in either the mean direction and/or in concentration and provide again 2 alternate test statistics as well as nomograms to obtain the simulated critical values for these tests. The tests statistics can be computed using the routine `change.pt` in `CircStats`. Section 11.4 provides a simple example illustrating these test procedures. We end with brief discussions on nonparametric and other approaches for the change-point problem.

11.2 Tests for Change in Mean Direction

We assume that the observations have a von Mises or Circular Normal distribution $CN(\mu,\ \kappa)$ with pdf as defined in (2.2.6). A change in mean direction at the kth observation implies that $\alpha_1,\ \ldots,\ \alpha_k \sim CN(\mu_1,\ \kappa)$ and $\alpha_{k+1},\ \ldots,\ \alpha_n \sim CN(\mu_2,\ \kappa)$.

11.2.1 κ Known Case

If $\boldsymbol{\theta} = (k,\ \mu_1,\ \mu_2)$ denotes the parameter vector for our problem, the parameter space is $\Omega = \{1,\ \ldots,\ n\} \times [-\pi,\ \pi) \times [-\pi,\ \pi)$, taking for convenience the range as $[-\pi,\ \pi)$. Technically, we might say that the hypothesis of no change corresponds to the change-point at n, so that under H_0, the parameter space becomes $\omega = \Omega \cap H_0 = \{n\} \times [-\pi,\ \pi) \times [-\pi,\ \pi)$. If the change in mean takes place at the point k, the likelihood function is given by

$$lik(\boldsymbol{\theta}) = \frac{1}{[2\pi I_0(\kappa)]^n} \exp\left[\kappa\left\{\sum_{i=1}^{k} \cos(\alpha_i - \mu_1) + \sum_{i=k+1}^{n} \cos(\alpha_i - \mu_2)\right\}\right]. \tag{11.2.1}$$

Let $(R_{1k},\ \overline{\alpha}_{1k})$, $(R_{2k},\ \overline{\alpha}_{2k})$ and finally $(R,\ \overline{\alpha}_0)$ denote the length and direction of the resultant of the subsets $(\alpha_1,\ \ldots,\ \alpha_k)$, $(\alpha_{k+1},\ \ldots,\ \alpha_n)$ and the combined sample $(\alpha_1,\ \ldots,\ \alpha_n)$, respectively. It can be easily verified that under H_0, the MLE of μ satisfies

$$\sum_{i=1}^{n} \sin(\alpha_i - \hat{\mu}_0) = 0. \tag{11.2.2}$$

Similarly, under H_1, for a given k, the MLEs of μ_1 and μ_2 satisfy

$$\sum_{i=1}^{k} \sin(\alpha_i - \hat{\mu}_{1k}) = 0 \quad \text{and} \quad \sum_{i=k+1}^{n} \sin(\alpha_i - \hat{\mu}_{2k}) = 0. \tag{11.2.3}$$

The solutions to Equations (11.2.2) and (11.2.3) are given by

$$\hat{\mu}_0 = \overline{\alpha}_0, \quad \hat{\mu}_{1k} = \overline{\alpha}_{1k} \quad \text{and} \quad \hat{\mu}_{2k} = \overline{\alpha}_{2k},$$

respectively. Since

$$\sum_{i=1}^{n} \cos(\alpha_i - \overline{\alpha}_0) = R \tag{11.2.4}$$

and

$$\sum_{i=1}^{k} \cos(\alpha_i - \hat{\mu}_{1k}) = R_{1k} \quad \sum_{i=k+1}^{n} \cos(\alpha_i - \hat{\mu}_{2k}) = R_{2k} \tag{11.2.5}$$

(see Theorem 1.1), the likelihood ratio becomes:

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

As usual, we reject H_0 if Λ_k is "sufficiently small". However, since we do not know the point of change k, we have 2 options: one is to treat k also as an unknown parameter and minimize the likelihood ratio Λ_k over k; the other is to take a Bayesian approach and average Λ_k over a prior distribution on $\{1, \ldots, n\}$. In the first case, the likelihood ratio test (LRT) of H_0 vs. H_1 would reject H_0 for small values of $\inf_k \Lambda_k$ or equivalently, reject H_0 for large values of

$$\Lambda = \sup_{k \in \{1, \ldots, n\}} (R_{1k} + R_{2k}) - R,$$

since κ is given. This leads to rejecting H_0 when

$$\Lambda > c, \tag{11.2.6}$$

where the cut-off point c is determined based on the significance level of the test. In all further discussion, we refer to this test as the supremum or "sup" test. On the other hand, for Bayesian analog, assuming a uniform prior on the possible values of k, we reject H_0 whenever

$$\frac{1}{n}\sum_{k=1}^{n}(R_{1k} + R_{2k}) - R > c', \tag{11.2.7}$$

where c' is also determined based on the significance level. We refer to this test as the average or "avg" test. The joint distribution of $(R_{1k},\ R_{2k})$ conditional on R, is known (see Equation (3.5.3)) for any fixed k and is indeed independent of κ. However, the exact distribution of the "sup" and "avg" statistics, which depend on the joint distribution of $(R_{1k},\ R_{2k})_{k=1}^{n}$, does not have a simple analytic form, partly because $(R_{1k},\ R_{2k})$ are not independent for different k. The only practical solution is to simulate their cut-off points for various combinations of κ and n, which is what we do. The results are presented in Tables 11.1 and 11.2 for both the 5% and 1% levels, as well as graphically in Figures 11.1 and 11.2 for the 5% level, from which one can read off the necessary critical values.

11.2.2 κ Unknown Case

While the discussion in the preceding subsection is enlightening in terms of procedures to be used, it is not often that one has a known κ. In practice, κ is also unknown so that we have a 4-dimensional parameter $\boldsymbol{\theta} = (k,\ \mu_1,\ \mu_2,\ \kappa)$ with the parameter space $\Omega = \{1,\ \ldots,\ n\} \times [-\pi,\ \pi) \times [-\pi,\ \pi) \times (0,\ \infty)$. Arguing as before, under H_0, the parameter space becomes $\omega = \Omega \cap H_0 = \{n\} \times [-\pi,\ \pi) \times [-\pi,\ \pi) \times (0,\ \infty)$.

The likelihood function for the data is given by (11.2.1) with the new parameter vector $\boldsymbol{\theta}$. It can be easily verified that under H_0, the MLEs of μ and κ satisfy (11.2.2) and

$$\frac{I_1(\hat{\kappa}_0)}{I_0(\hat{\kappa}_0)} = \frac{R}{n}. \tag{11.2.8}$$

Similarly, under H_1, for a given k, the MLEs satisfy Equations (11.2.3), and

$$\frac{I_1(\hat{\kappa}_k)}{I_0(\hat{\kappa}_k)} = \frac{R_{1k} + R_{2k}}{n}. \tag{11.2.9}$$

Again, the solutions to Equation (11.2.3) are given by $\hat{\mu}_{1k} = \overline{\alpha}_{1k}$ and $\hat{\mu}_{2k} = \overline{\alpha}_{2k}$, respectively. Hence, the likelihood ratio of the data (for a given k) is

$$\begin{aligned}
\Lambda_k &= \left[\frac{I_0(\hat{\kappa}_k)}{I_0(\hat{\kappa}_0)}\right]^n \exp\left[\hat{\kappa}_0 \sum_{i=1}^{n} \cos(\alpha_i - \overline{\alpha}_0)\right. \\
&\qquad \left. -\hat{\kappa}_k \left\{\sum_{i=1}^{k} \cos(\alpha_i - \overline{\alpha}_{1k}) + \sum_{i=k+1}^{n} \cos(\alpha_i - \overline{\alpha}_{2k})\right\}\right] \\
&= \left[\frac{I_0(\hat{\kappa}_k)}{I_0(\hat{\kappa}_0)}\right]^n \exp\left[\hat{\kappa}_0 R - \hat{\kappa}_k (R_{1k} + R_{2k})\right]
\end{aligned}$$

which gives

$$\begin{aligned}
\lambda_k &= -\log \Lambda_k \\
&= n\left\{\log(I_0(\hat{\kappa}_0)) - \log(I_0(\hat{\kappa}_k))\right\} + \left\{\hat{\kappa}_k(R_{1k} + R_{2k}) - \hat{\kappa}_0 R\right\} \\
&= n\left[\Psi(R_{1k} + R_{2k}) - \Psi(R)\right],
\end{aligned}$$

where

$$\Psi(t) = A^{-1}(t) - \log\left[I_0\left\{A^{-1}(t)\right\}\right]$$

and $A(\cdot)$ is defined as

$$A(t) = \frac{I_1(t)}{I_0(t)}.$$

As before, since k is unknown, we employ the "sup" and "avg" methods to come up with two test criteria, namely

$$\sup_k \Psi(R_{1k} + R_{2k}) - \Psi(R)$$

and

$$\frac{1}{n}\sum_{k=1}^{n} \Psi(R_{1k} + R_{2k}) - \Psi(R).$$

In both these cases, however, the null distribution of the test statistic depends on the unknown κ. To make these tests independent of κ, we condition them on the overall resultant length R. It can be checked that $\Psi(\cdot)$ is monotone in its argument, so that these two *conditional* tests are equivalent to the "sup" and "avg" tests given in Equations (11.2.6) and (11.2.7), respectively. Only difference from the κ known case is that the test statistic values are conditional on R. The critical values of these two conditional tests that we refer to as "rave" and "rmax" in the `change.pt` routine of `CircStats` are tabulated in Tables 11.3 and 11.4 respectively both for the 5% and 1% level as well as graphically in Figures 11.4 and 11.3 for the 5% level. The `change.pt` routine provides not only the value of "rmax" but also the point k at which such a maximum is attained so that if the change is indeed significant, one knows where it is most likely to have occurred. See Example 11.1.

11.2.3 Simulated Critical Values

As stated earlier, the tests proposed in the previous sections have no simple known distributional form. Thus, to obtain their cut-off values, we resort to large-scale Monte Carlo simulations. All the codes were written in the `C` Language with calls to the `IMSL/C/STAT` library for the random number generators. In particular, we made extensive use of the routine `imsls_f_random_von_mises` to generate all the von Mises random deviates.

For the κ known case, we sampled from von Mises distributions with center zero and concentration κ (i.e., $CN(0,\ \kappa)$). We considered $\kappa = 0.5(0.5)3(1)4$

Table 11.1: 5%(1%) critical values for the "avg" statistic, κ known.

n	κ						
	0.05	0.1	1.5	2.0	2.5	3.0	4.0
10	1.93 (2.63)	1.32 (2.04)	0.93 (1.44)	0.65 (1.01)	0.53 (0.84)	0.43 (0.67)	0.31 (0.47)
12	2.01 (2.75)	1.35 (2.11)	0.91 (1.44)	0.66 (1.03)	0.51 (0.77)	0.42 (0.65)	0.31 (0.48)
14	2.18 (2.97)	1.40 (2.05)	0.89 (1.32)	0.64 (1.00)	0.50 (0.77)	0.42 (0.66)	0.31 (0.50)
16	2.28 (3.21)	1.39 (2.08)	0.90 (1.41)	0.67 (1.04)	0.50 (0.78)	0.42 (0.67)	0.32 (0.50)
18	2.33 (3.30)	1.38 (2.07)	0.89 (1.36)	0.65 (0.99)	0.50 (0.76)	0.42 (0.63)	0.32 (0.51)
20	2.45 (3.43)	1.36 (2.16)	0.87 (1.32)	0.64 (0.99)	0.51 (0.80)	0.42 (0.66)	0.31 (0.47)
25	2.53 (3.77)	1.36 (2.08)	0.87 (1.34)	0.62 (0.94)	0.50 (0.79)	0.42 (0.67)	0.31 (0.49)
30	2.62 (3.85)	1.35 (2.08)	0.87 (1.32)	0.63 (0.99)	0.50 (0.77)	0.42 (0.66)	0.31 (0.50)
35	2.69 (4.03)	1.36 (2.09)	0.85 (1.33)	0.65 (0.99)	0.51 (0.77)	0.41 (0.65)	0.31 (0.48)
40	2.71 (4.13)	1.35 (2.10)	0.84 (1.32)	0.64 (0.98)	0.50 (0.78)	0.42 (0.65)	0.32 (0.49)
45	2.76 (4.32)	1.32 (2.10)	0.86 (1.34)	0.63 (0.99)	0.50 (0.75)	0.43 (0.67)	0.31 (0.49)
50	2.78 (4.14)	1.32 (2.00)	0.85 (1.34)	0.63 (0.99)	0.51 (0.78)	0.42 (0.65)	0.32 (0.50)

and n =10 (2) 20 (5) 50. At each (n, κ) combination, we did 100,000 simulations to obtain upper 5% points of the two tests. The results appear in Tables 11.1—11.2 and the corresponding nomograms are in Figures 11.1—11.2.

For the κ unknown case, we sampled from a conditional von Mises distribution, the conditioning event being the given length r of the resultant R. Each (r, n) combination results in a different distribution and we considered $n = 10(2)20(5)50$ and $r/n = 0.05(0.05)0.95$. Since the conditional sampling discards a lot of the random numbers for not meeting the conditioning criterion, sampling procedure was much slower compared to the unconditional case. Apart from using `imsls_f_random_von_mises`, we used `imsls_f_random_binomial` to draw the conditional samples. The results of 100,000 simulations appear in Tables 11.3 and 11.4 and the corresponding nomograms in Figures 11.3–11.4.

Casual examination of the nomograms suggest that high κ values for the unconditional case (and correspondingly, high r/n values for the conditional case) make the test statistics (and hence, the cut-off values) free of n. This is reflected by the almost horizontal lines in the corresponding figures. In particular, for the "avg" statistic in the κ known case, the $100(1-\alpha)\%$ cut-offs are approximated very well by $\chi^2_{1;\,\alpha}/2\kappa$ when $\kappa \geq 2$. This maybe explained by the fact that for large κ and any fixed k, $2\kappa(n-R)$ has an approximate χ^2_{n-1} distribution. Hence, for each k, we have

$$2\kappa\,(R_{1k} + R_{2k} - R) \sim \chi^2_1.$$

Table 11.2: 5%(1%) critical values for the "sup" statistic, κ known.

n	κ						
	0.05	0.1	1.5	2.0	2.5	3.0	4.0
10	4.19 (5.42)	3.36 (4.48)	2.50 (3.57)	1.94 (2.58)	1.61 (2.10)	1.32 (1.86)	0.94 (1.36)
12	4.58 (5.90)	3.49 (4.67)	2.57 (3.57)	1.97 (2.76)	1.63 (2.08)	1.32 (1.90)	0.95 (1.41)
14	5.01 (6.39)	3.65 (4.90)	2.62 (3.59)	1.98 (2.78)	1.64 (2.13)	1.39 (1.93)	1.00 (1.50)
16	5.36 (6.93)	3.77 (5.15)	2.73 (3.81)	2.00 (2.85)	1.69 (2.20)	1.36 (1.90)	1.02 (1.46)
18	5.56 (7.16)	3.90 (5.25)	2.79 (3.83)	2.03 (2.85)	1.70 (2.28)	1.43 (1.97)	1.04 (1.46)
20	5.91 (7.52)	3.94 (5.52)	2.72 (3.77)	2.00 (2.87)	1.73 (2.34)	1.41 (1.94)	1.04 (1.45)
25	6.30 (8.44)	4.12 (5.62)	2.83 (3.93)	2.05 (2.86)	1.78 (2.37)	1.45 (2.01)	1.07 (1.48)
30	6.80 (8.83)	4.24 (5.78)	2.91 (3.87)	2.13 (3.05)	1.78 (2.39)	1.50 (2.04)	1.08 (1.59)
35	7.23 (9.48)	4.37 (6.02)	2.96 (3.98)	2.21 (3.08)	1.85 (2.48)	1.49 (1.99)	1.11 (1.55)
40	7.44 (9.94)	4.44 (5.96)	2.97 (4.05)	2.26 (3.03)	1.83 (2.34)	1.52 (2.03)	1.12 (1.54)
45	7.69 (10.47)	4.41 (6.05)	3.03 (4.06)	2.26 (3.12)	1.82 (2.34)	1.56 (2.04)	1.13 (1.53)
50	7.95 (10.47)	4.55 (6.00)	3.02 (4.09)	2.27 (3.09)	1.84 (2.46)	1.57 (2.03)	1.16 (1.58)

Although averaging over k amounts to averaging dependent χ^2's, the result is well approximated by a χ_1^2 random variable. There does not seem to be any such easy explanation or approximation for the "sup" statistic. A similar phenomenon (albeit less pronounced), occurs for the κ unknown case. There, for high values of r/n, corresponding to high concentration, the cut-off values for the "avg" statistic do not change much for different n.

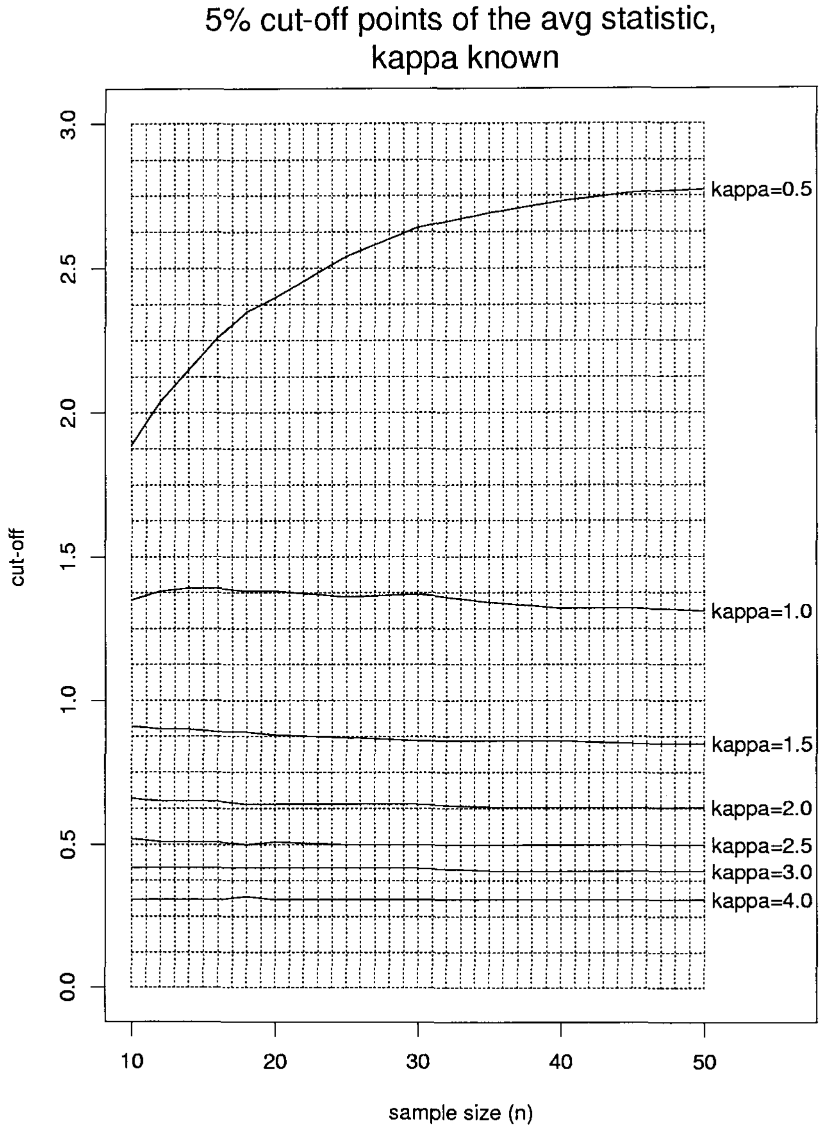


Figure 11.1: 5% critical values of the avg statistic, κ known.

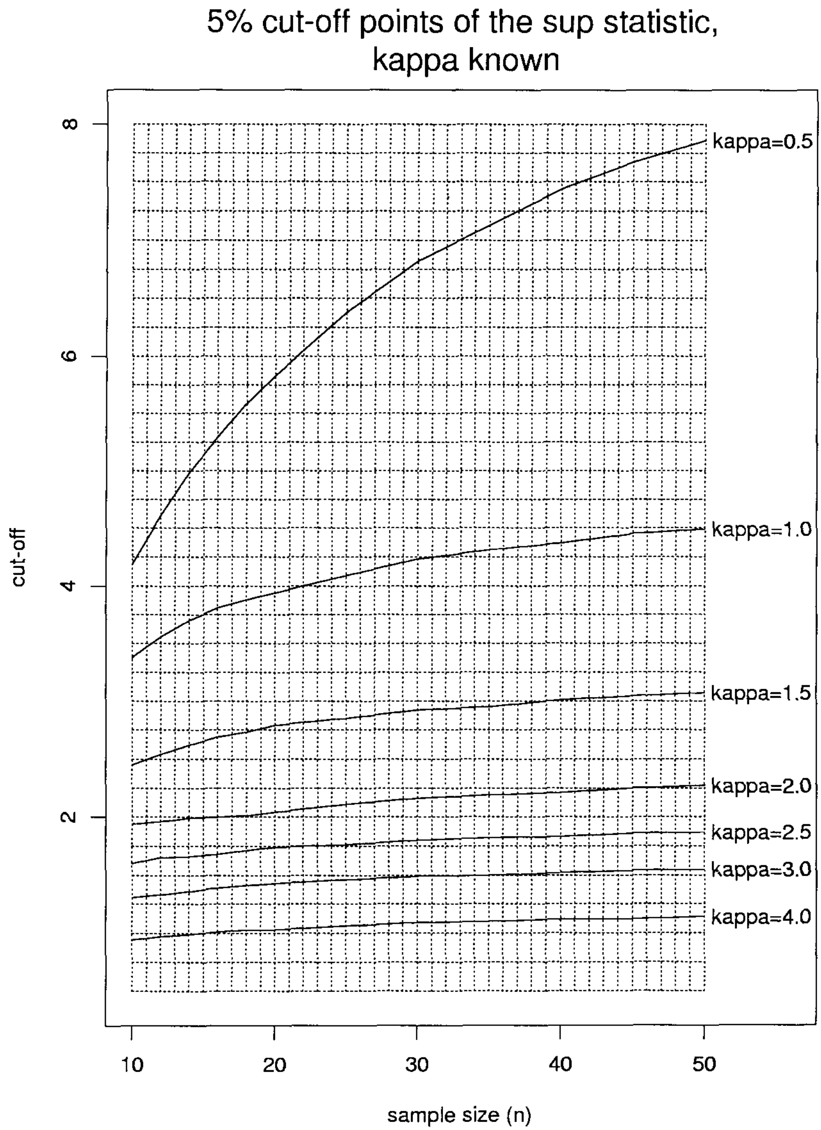


Figure 11.2: 5% critical values of the sup statistic, κ known.

Table 11.3: 5% (1%) critical values of "rmax" statistic, κ unknown.

n	r/n				
	0.05	0.10	0.15	0.20	0.25
10	5.86 (6.86)	5.44 (6.37)	5.01 (5.88)	4.65 (5.51)	4.25 (5.17)
12	6.47 (7.53)	5.92 (6.99)	5.44 (6.50)	4.99 (5.98)	4.58 (5.55)
14	7.00 (8.17)	6.39 (7.60)	5.80 (7.00)	5.31 (6.44)	4.84 (5.89)
16	7.51 (8.83)	6.79 (8.08)	6.14 (7.38)	5.59 (6.77)	5.09 (6.24)
18	7.91 (9.31)	7.15 (8.56)	6.46 (7.82)	5.84 (7.16)	5.30 (6.52)
20	8.33 (9.79)	7.49 (8.94)	6.75 (8.14)	6.06 (7.43)	5.48 (6.79)
25	9.31 (11.02)	8.27 (9.95)	7.34 (8.88)	6.58 (8.12)	5.85 (7.27)
30	10.16 (12.04)	8.93 (10.81)	7.87 (9.72)	7.01 (8.72)	6.22 (7.81)
35	10.90 (13.00)	9.50 (11.46)	8.30 (10.23)	7.31 (9.09)	6.46 (8.16)
40	11.60 (13.81)	10.02 (12.16)	8.71 (10.75)	7.59 (9.48)	6.68 (6.42)
45	12.27 (14.61)	10.50 (12.74)	9.06 (11.21)	7.86 (9.84)	6.89 (8.77)
50	12.86 (15.34)	10.91 (13.24)	9.36 (11.60)	8.10 (10.19)	7.09 (9.05)

n	r/n				
	0.30	0.35	0.40	0.45	0.50
10	3.90 (4.78)	3.62 (4.40)	3.37(4.02)	3.10 (3.76)	2.83 (3.49)
12	4.16 (5.14)	3.84 (4.77)	3.55 (4.36)	3.27 (3.95)	2.99 (3.69)
14	4.42 (5.44)	4.00 (5.01)	3.70 (4.61)	3.39 (4.21)	3.11 (3.85)
16	4.65 (5.71)	4.19 (5.24)	3.82 (4.80)	3.51 (4.40)	3.20 (3.98)
18	4.79 (5.98)	4.36 (5.46)	3.93 (4.97)	3.60 (4.54)	3.30 (4.12)
20	4.96 (6.18)	4.48 (5.66)	4.05 (5.12)	3.70 (4.68)	3.35 (4.22)
25	5.28 (6.61)	4.73 (5.98)	4.26 (5.46)	3.85 (4.94)	3.50 (4.47)
30	5.53 (6.94)	4.96 (6.35)	4.44 (5.70)	3.98 (5.12)	3.61 (4.64)
35	5.72 (7.32)	5.11 (6.51)	4.57 (5.85)	4.11 (5.29)	3.71 (4.78)
40	5.92 (7.55)	5.27 (6.72)	4.67 (6.04)	4.18 (5.46)	3.76 (4.85)
45	6.06 (7.78)	5.38 (6.97)	4.79 (6.19)	4.27 (5.58)	3.82 (4.99)
50	6.18 (7.97)	5.48 (7.06)	4.88 (6.32)	4.32 (5.64)	3.87 (5.08)

Table 11.3(contd).

n	r/n				
	0.55	0.60	0.65	0.70	0.75
10	2.59 (3.24)	2.30 (2.88)	2.01 (2.51)	1.94 (2.16)	1.83 (1.97)
12	2.75 (3.44)	2.46 (3.12)	2.15 (2.75)	1.97 (2.37)	1.89 (2.00)
14	2.83 (3.55)	2.55 (3.25)	2.25 (2.91)	1.99 (2.54)	1.92 (2.12)
16	2.94 (3.68)	2.64 (3.34)	2.34 (3.00)	2.03 (2.64)	1.94 (2.24)
18	3.01 (3.76)	2.72 (3.47)	2.41 (3.13)	2.09 (2.74)	1.95 (2.32)
20	3.08 (3.85)	2.78 (3.54)	2.46 (3.20)	2.14 (2.81)	1.96 (2.40)
25	3.21 (4.04)	2.88 (3.68)	2.56 (3.34)	2.25 (2.96)	1.98 (2.53)
30	3.30 (4.18)	2.97 (3.78)	2.64 (3.45)	2.32 (3.07)	2.00 (2.64)
35	3.37 (4.31)	3.03 (3.88)	2.71 (3.52)	2.37 (3.14)	2.02 (2.72)
40	3.43 (4.38)	3.10 (3.94)	2.76 (3.58)	2.42 (3.20)	2.07 (2.75)
45	3.47 (4.47)	3.14 (4.00)	2.80 (3.62)	2.45 (3.22)	2.10 (2.81)
50	3.53 (4.56)	3.16 (4.06)	2.83 (3.67)	2.48 (3.29)	2.13 (2.87)

n	r/n			
	0.80	0.85	0.90	0.95
10	1.53 (1.71)	1.15 (1.28)	0.76 (0.85)	0.39 (0.43)
12	1.69 (1.89)	1.28 (1.45)	0.86 (0.97)	0.43 (0.49)
14	1.78 (1.95)	1.41 (1.62)	0.93 (1.08)	0.47 (0.54)
16	1.83 (1.97)	1.51 (1.75)	1.01 (1.17)	0.50 (0.59)
18	1.86 (1.98)	1.60 (1.84)	1.08 (1.27)	0.54 (0.63)
20	1.88 (1.99)	1.67 (1.90)	1.14 (1.34)	0.57 (0.67)
25	1.91 (2.07)	1.77 (1.95)	1.29 (1.54)	0.64 (0.77)
30	1.92 (2.17)	1.82 (1.97)	1.41 (1.70)	0.71 (0.86)
35	1.94 (2.25)	1.84 (1.98)	1.52 (1.81)	0.77 (0.94)
40	1.94 (2.29)	1.86 (1.98)	1.61 (1.87)	0.83 (1.02)
45	1.95 (2.36)	1.87 (1.99)	1.66 (1.90)	0.88 (1.09)
50	1.96 (2.37)	1.88 (1.99)	1.70 (1.92)	0.93 (1.16)

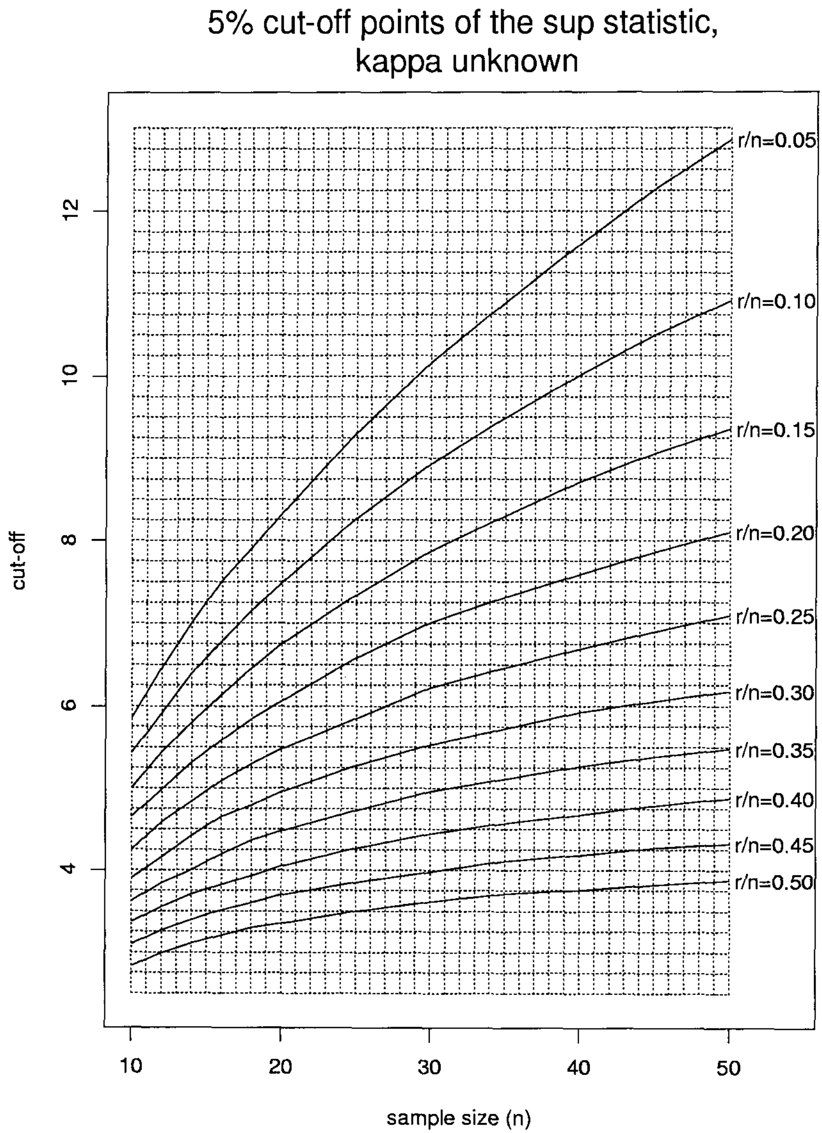


Figure 11.3: 5% critical values of the "rmax" statistic, κ unknown.

Table 11.4: 5% (1%) critical values of the "rave" statistic, κ unknown.

n	r/n				
	0.05	0.10	0.15	0.20	0.25
10	3.04 (3.54)	2.66 (3.17)	2.34 (2.83)	2.09 (2.59)	1.88 (2.34)
12	3.27 (3.87)	2.85 (3.43)	2.49 (3.08)	2.19 (2.75)	1.94 (2.49)
14	3.52 (4.20)	3.02 (3.70)	2.62 (3.28)	2.29 (2.90)	2.01 (2.59)
16	3.72 (4.48)	3.16 (3.89)	2.71 (3.42)	2.37 (3.03)	2.05 (2.70)
18	3.89 (4.69)	3.29 (4.09)	2.80 (3.56)	2.42 (3.16)	2.11 (2.78)
20	4.06 (4.92)	3.40 (4.27)	2.89 (3.68)	2.47 (3.23)	2.14 (2.86)
25	4.45 (5.45)	3.66 (4.61)	3.05 (3.92)	2.59 (3.43)	2.20 (2.96)
30	4.77 (5.88)	3.88 (4.93)	3.18 (4.18)	2.67 (3.58)	2.27 (3.10)
35	5.07 (6.28)	4.04 (5.23)	3.29 (4.32)	2.73 (3.69)	2.28 (3.15)
40	5.33 (6.60)	4.18 (5.40)	3.38 (4.49)	2.76 (3.75)	2.33 (3.23)
45	5.59 (6.99)	4.33 (5.61)	3.46 (4.62)	2.80 (3.84)	2.33 (3.30)
50	5.78 (7.26)	4.44 (5.75)	3.51 (4.70)	2.86 (3.96)	2.37 (3.35)

n	r/n				
	0.30	0.35	0.40	0.45	0.50
10	1.67 (2.12)	1.50 (1.93)	1.34 (1.76)	1.21 (1.60)	1.09 (1.46)
12	1.73 (2.25)	1.55 (2.04)	1.38 (1.83)	1.23 (1.65)	1.11 (1.51)
14	1.77 (2.33)	1.57 (2.09)	1.40 (1.90)	1.25 (1.71)	1.12 (1.54)
16	1.82 (2.41)	1.61 (2.15)	1.42 (1.94)	1.26 (1.74)	1.13 (1.57)
18	1.84 (2.48)	1.62 (2.21)	1.42 (1.98)	1.27 (1.78)	1.13 (1.59)
20	1.86 (2.51)	1.64 (2.24)	1.44 (1.99)	1.27 (1.79)	1.12 (1.60)
25	1.90 (2.63)	1.66 (2.31)	1.45 (2.07)	1.28 (1.82)	1.13 (1.65)
30	1.92 (2.68)	1.69 (2.37)	1.46 (2.10)	1.28 (1.85)	1.14 (1.64)
35	1.96 (2.77)	1.69 (2.41)	1.47 (2.12)	1.28 (1.87)	1.14 (1.66)
40	1.98 (2.81)	1.68 (2.43)	1.47 (2.12)	1.28 (1.89)	1.12 (1.67)
45	1.97 (2.83)	1.70 (2.45)	1.47 (2.17)	1.28 (1.90)	1.12 (1.67)
50	1.99 (2.86)	1.70 (2.49)	1.47 (2.18)	1.28 (1.90)	1.13 (1.69)

Table 11.4(contd).

n	r/n				
	0.55	0.60	0.65	0.70	0.75
10	1.00 (1.34)	0.89 (1.19)	0.80 (1.06)	0.70 (0.93)	0.60 (0.79)
12	1.01 (1.37)	0.90 (1.24)	0.80 (1.09)	0.71 (0.96)	0.61 (0.82)
14	1.01 (1.39)	0.90 (1.25)	0.79 (1.11)	0.70 (0.98)	0.61 (0.84)
16	1.01 (1.41)	0.90 (1.27)	0.80 (1.12)	0.70 (0.98)	0.60 (0.85)
18	1.01 (1.42)	0.90 (1.28)	0.80 (1.14)	0.70 (0.99)	0.60 (0.85)
20	1.01 (1.44)	0.89 (1.28)	0.79 (1.14)	0.70 (1.00)	0.60 (0.86)
25	1.01 (1.46)	0.90 (1.30)	0.79 (1.15)	0.69 (1.00)	0.60 (0.87)
30	1.00 (1.47)	0.89 (1.31)	0.78 (1.14)	0.69 (1.00)	0.59 (0.86)
35	1.01 (1.48)	0.88 (1.32)	0.78 (1.15)	0.68 (1.01)	0.58 (0.86)
40	1.00 (1.47)	0.88 (1.32)	0.77 (1.16)	0.68 (1.01)	0.58 (0.86)
45	0.99 (1.48)	0.88 (1.31)	0.77 (1.15)	0.67 (1.00)	0.57 (0.85)
50	0.99 (1.47)	0.88 (1.32)	0.77 (1.14)	0.67 (0.99)	0.57 (0.86)

n	r/n			
	0.80	0.85	0.90	0.95
10	0.50 (0.64)	0.39 (0.49)	0.26 (0.34)	0.14 (0.17)
12	0.50 (0.67)	0.39 (0.52)	0.27 (0.35)	0.14 (0.18)
14	0.51 (0.68)	0.40 (0.53)	0.28 (0.37)	0.14 (0.19)
16	0.50 (0.69)	0.40 (0.54)	0.28 (0.37)	0.14 (0.19)
18	0.50 (0.70)	0.40 (0.54)	0.28 (0.38)	0.14 (0.19)
20	0.50 (0.71)	0.40 (0.55)	0.28 (0.38)	0.15 (0.20)
25	0.49 (0.71)	0.39 (0.56)	0.28 (0.39)	0.15 (0.20)
30	0.49 (0.72)	0.39 (0.56)	0.28 (0.39)	0.15 (0.21)
35	0.49 (0.72)	0.38 (0.56)	0.27 (0.39)	0.15 (0.21)
40	0.48 (0.71)	0.38 (0.55)	0.27 (0.39)	0.15 (0.21)
45	0.48 (0.72)	0.38 (0.56)	0.27 (0.39)	0.15 (0.21)
50	0.48 (0.71)	0.37 (0.56)	0.27 (0.39)	0.15 (0.21)

11.2.4 Power Comparisons

Using Monte Carlo simulations, an extensive power comparison of the average and sup statistics was done in Ghosh et al. (1999). They study the effect of various factors including differing concentrations, sample sizes, placement of the point of change, and different values of $\Delta = |\mu_1 - \mu_2|$. The conclusions

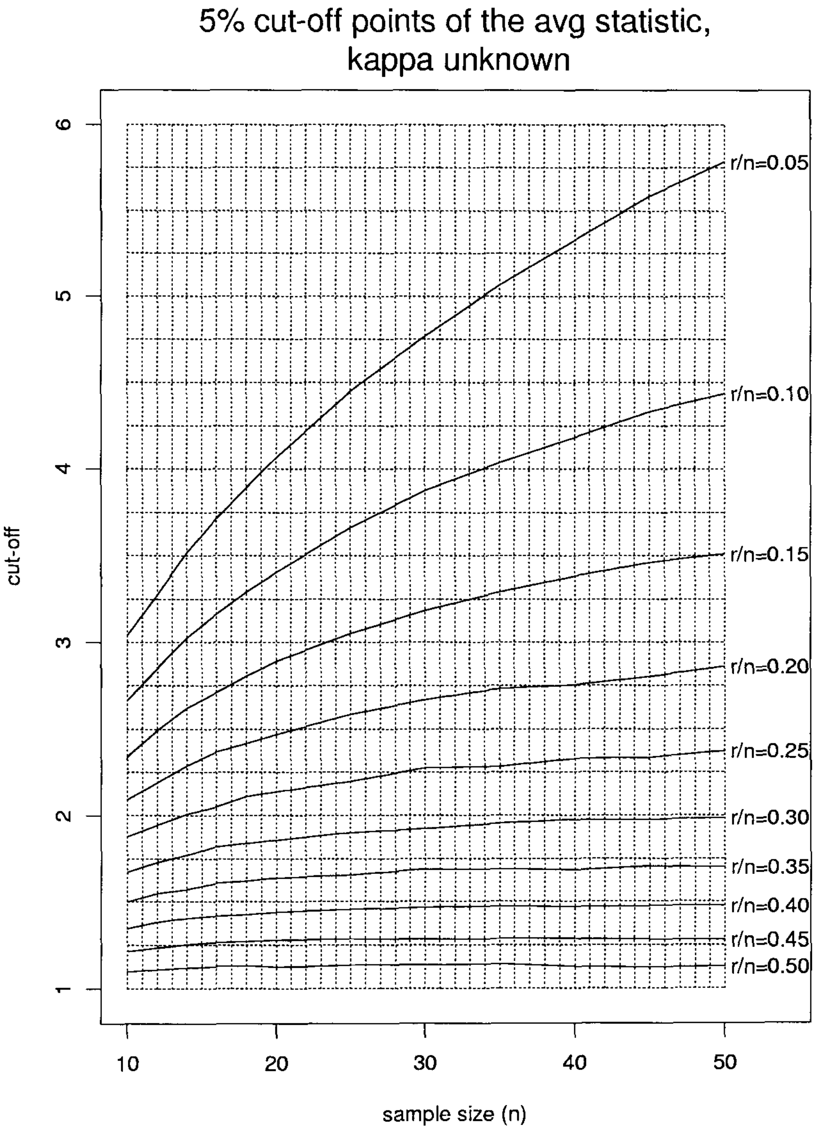


Figure 11.4: 5% critical values of the "rave" statistic, κ unknown.

are similar whether κ is known or unknown and they may be summarized as follows:

(i) Both the tests are "symmetric" in the change-point k in the sense that power at k is approximately equal to power at $n - k$, everything else being the same. This symmetry in our test procedures is to be expected.

(ii) As would be expected, the powers of both the statistics show an increasing trend as Δ increases, everything else being the same. The trend is not as pronounced for smaller values of the concentration parameter as in the case of larger values.

(iii) For high concentration parameter κ, the average statistic is more powerful than the sup statistic if the change-point is near the center (i.e. $k \approx n/2$). The maximum power reached here is almost 1 for both. On the other hand, if the change point is near the end ($k \approx 1$), the sup statistic becomes more powerful for high concentration. The max power reached in this case is about 50% for the sup statistic compared to 20% for the average statistic.

Hence, as intuition suggests, it is easier to detect a change-point in the middle than at the end. If one suspects change at the end points, the sup statistic is recommended.

(iv) For $k \approx n/4$, both the tests behave very similarly — having almost the same power curve, irrespective of n and κ.

(v) If there is a very low concentration in the data, none of the tests are useful since the maximum power obtained is as low as 15% even for $n = 20$. Thus, trying to detect changes in the mean direction when the distributions are nearly uniform, is as one should expect, not very fruitful.

(vi) For a given κ, power increases with n as long as the k/n ratio remains constant. Hence increase in sample size increases power.

Based on this discussion, we recommend the use of sup statistic if we have some previous knowledge that the change-point occurred at the end and concentration is high ($\kappa > 3.5$). If the change-point is suspected to be in the middle, and the concentration is moderate to high, we recommend using the average statistic. If the concentration is low ($\kappa < 2$), these statistics are not as useful. For detailed comparisons, the reader is referred to Ghosh et al. (1999).

11.2.5 Robustness

The test statistics proposed here as well as the critical values tabulated, are obtained assuming that the data comes from a von Mises model. One might ask how justified are these tests if one does not have the von Mises assumption, or more generally in the absence of any distributional assumption? With this goal of studying model robustness, Ghosh et al. (1999) conduct an extensive simulation study to see how well these tests perform, if for instance, the data come from a Wrapped Normal or a Wrapped Cauchy distribution. The general conclusion is that these tests are justified and valid when the underlying models are unimodal, and for which the resultant length and direction are relevant measures of concentration and mean direction. Based on the circular distance $d(\alpha, \beta)$ defined in Equation (1.3.7), we have seen that $(n - R)$ provides a good measure of dispersion for a set of n angular measurements. If there is a change in the mean direction at some k, then one expects the overall dispersion to be large compared to the dispersions of the two individual groups. Hence, a reasonable statistic to detect change is given by

$$(n - R) - \{(k - R_{1k}) + (n - k - R_{2k})\} = R_{1k} + R_{2k} - R.$$

This is similar to the idea used in the approximate ANOVA for circular data discussed in Section 5.3.1. When k is unknown, we can either average over all potential k or take supremum over k, resulting in the two suggested statistics. Thus although these test statistics are optimal for the von Mises context, they provide reasonable tests for a much broader class of models. The reader is referred to the article cited above for details.

11.3 Tests for Change in μ and/or κ

If the change-point is at a given k, the likelihood of the observed data is maximized for

$$\hat{\mu}_1 = \overline{\alpha}_{1k},\ \hat{\mu}_2 = \overline{\alpha}_{2k},\ \hat{\kappa}_1 = A^{-1}\left(\frac{R_{1k}}{k}\right) \ \text{ and } \ \hat{\kappa}_2 = A^{-1}\left(\frac{R_{2k}}{n-k}\right).$$

Here, $(\overline{\alpha}_{1k},\ R_{1k})$ denote the direction and length, respectively of the resultant of the unit vectors given by $\alpha_1,\ \ldots,\ \alpha_k$ while $(\overline{\alpha}_{2k},\ R_{2k})$ denote those for $\alpha_{k+1},\ \ldots,\ \alpha_n$. On the other hand, under the null hypothesis H_0 of no change, writing $\mu_1 = \mu_2(\equiv \mu)$, and $\kappa_1 = \kappa_2(\equiv \kappa)$, the likelihood is maximized when

$$\hat{\mu} = \overline{\alpha} \quad \text{and} \quad \hat{\kappa} = A^{-1}\left(\frac{R}{n}\right),$$

where $(\overline{\alpha},\ R)$ are the direction and length, respectively, of the resultant of all the n observations.

Thus, when k is known, the likelihood ratio becomes

$$\Lambda_k = \frac{[I_0(\hat{\kappa}_1)]^k [I_0(\hat{\kappa}_2)]^{n-k}}{[I_0(\hat{\kappa})]^n} e^{\hat{\kappa}\sum_{i=1}^n \cos(\alpha_i-\hat{\mu}) - \hat{\kappa}_1 \sum_{i=1}^k \cos(\alpha_i-\hat{\mu}_1) - \hat{\kappa}_2 \sum_{i=k+1}^n \cos(\alpha_i-\hat{\mu}_2)}.$$

After some algebra, we have

$$-\log \Lambda_k = \frac{k}{n}\Psi\left(\frac{R_1}{k}\right) + \frac{n-k}{n}\Psi\left(\frac{R_2}{n-k}\right) - \Psi\left(\frac{R}{n}\right).$$

When k is unknown, we can take either the "sup" and "avg" as was done before, giving rise to the 2 test statistics:

$$T_1 = \sup_{k\in\{2,\ \ldots,\ n-2\}} \left[\frac{k}{n}\Psi\left(\frac{R_1}{k}\right) + \frac{n-k}{n}\Psi\left(\frac{R_2}{n-k}\right) - \Psi\left(\frac{R}{n}\right)\right] \tag{11.3.1}$$

and

$$T_2 = \frac{1}{n-3}\sum_{k=2}^{n-2} \left[\frac{k}{n}\Psi\left(\frac{R_1}{k}\right) + \frac{n-k}{n}\Psi\left(\frac{R_2}{n-k}\right) - \Psi\left(\frac{R}{n}\right)\right]. \tag{11.3.2}$$

Note that unlike the previous cases, k runs from 2 through $n-2$. This is because, the concentration parameters being unknown, we need at least two observations to estimate them.

The test statistics so obtained are functions of the resultant lengths of the unit vectors, whose distributions under the null hypothesis depend on the common (unknown) concentration parameter κ. This can be made free of κ upon conditioning by the overall resultant length R, since the conditional distribution of $(R_1,\ R_2)$ given R is independent of κ (see Equation 3.5.4).

We now present the results of such conditional Monte-Carlo simulations in the form of nomograms for the modified statistics that we call

$$tmax = \sup_{k \in \{2, \ldots, n-2\}} \left\{ \frac{k}{n} \Psi\left(\frac{R_1}{k}\right) + \frac{n-k}{n} \Psi\left(\frac{R_2}{n-k}\right) \right\} \Bigg| R = r \qquad (11.3.3)$$

and

$$tave = \frac{1}{n-3} \sum_{k=2}^{n-2} \left\{ \frac{k}{n} \Psi\left(\frac{R_1}{k}\right) + \frac{n-k}{n} \Psi\left(\frac{R_2}{n-k}\right) \right\} \Bigg| R = r. \qquad (11.3.4)$$

The `change.pt` routine in `CircStats` computes both these test statistics and for the "tmax," provides not its value but also the point k at which such a maximum is attained so that if the change is indeed significant, one knows where it is most likely to have occurred. See Example 11.1.

The results of 100,000 simulations for $n = 10(2)20(5)30(10)50$ and $r/n = 0.05(0.05)0.95$, appear as nomograms in Figures 11.5 and 11.6.

11.4 An Example

To illustrate these change-point procedures, we revisit the data from Example 5.2.

Example 11.1 *We consider the combined time-ordered sample of 19 observations and test for the presence of a change-point, if any. Since κ is unknown, we use the conditional tests.*

```
> x_c(10,50,55,55,65,90,285,285,325,355)
> y_c(75,75,80,80,80,95,130,170,210)
> combo_c(x,y)
> rad(combo)
 [1] 0.1745329 0.8726646 0.9599311 0.9599311 1.1344640 1.5707963
 [7] 4.9741884 4.9741884 5.6723201 6.1959188 1.3089969 1.3089969
 [13] 1.3962634 1.3962634 1.3962634 1.6580628 2.2689280 2.9670597
 [19] 3.6651914
> change.pt(rad(combo))
   n       rho     rmax k.r     rave      tmax k.t      tave
1 19 0.4805735 3.523024  16 1.704912 0.5323069  16 0.4390563
```

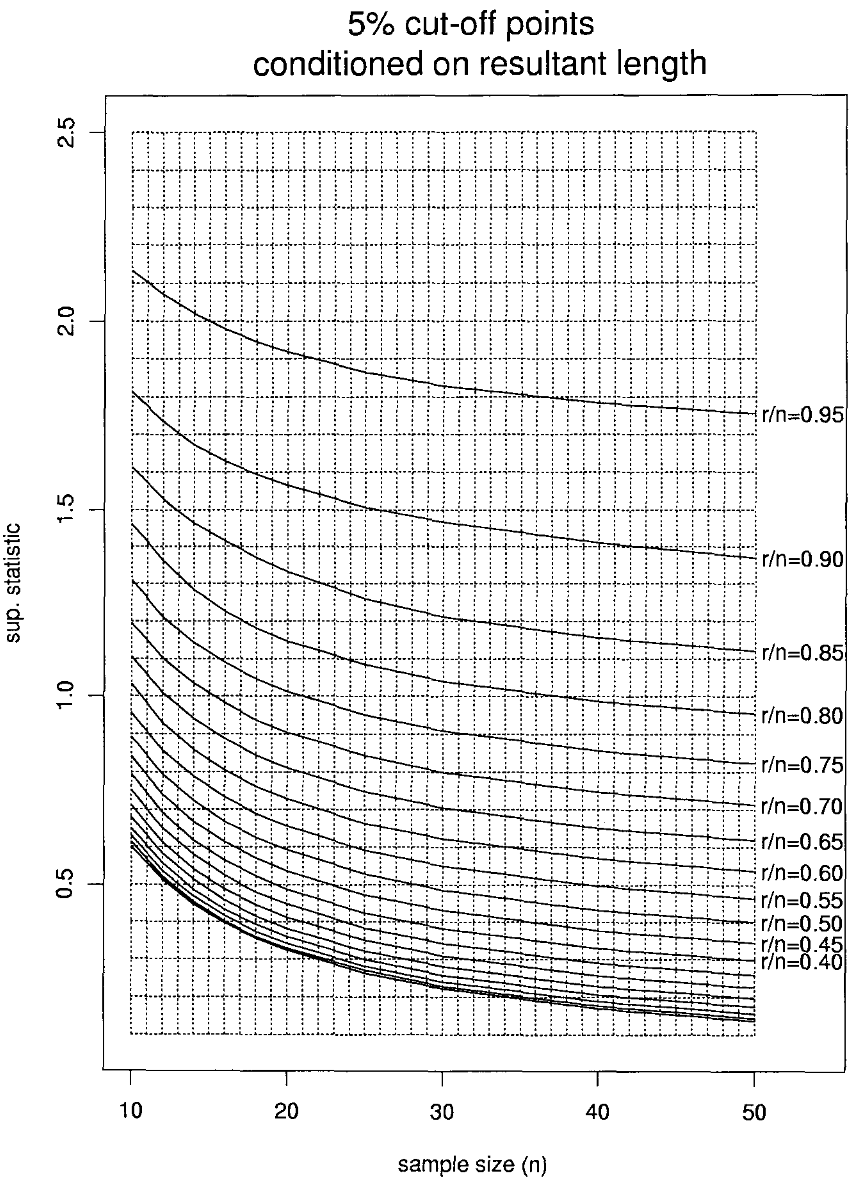


Figure 11.5: 5% critical values of "tmax" statistic, change in μ or κ or both.

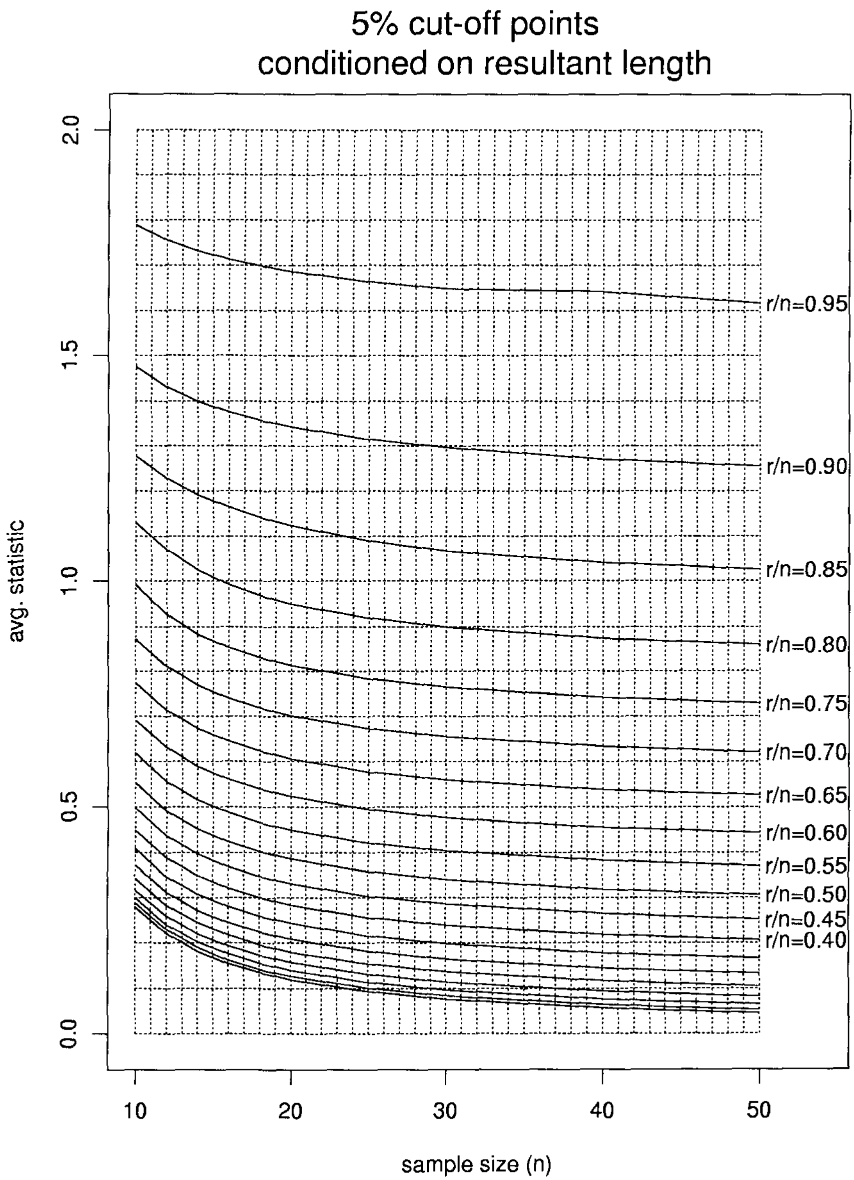


Figure 11.6: 5% critical values of "tave" statistic, change in μ or κ or both.

Calculations yield rho $= (r/n) = 0.48$ *corresponding to a* $\hat{\kappa} = 1.10$. *The values, rmax=3.523024 and rave=1.745029 are tests statistics for changes in the mean direction for* κ *unknown case. Consulting Tables 11.3 for "rmax" and 11.4 for "rave", we find that these values are both significant at 5%. When a possible change in* κ *is also suspected, the relevant test statistics are tmax= 0.5323 and tave= 0.4390, respectively. From Figures 11.3–11.4, we see that both tests reject the null hypothesis at 1% level, strongly indicating that there is a change. This is not surprising, in view of the conclusions reached earlier in Example 5.2. Finally, since both k.r and k.t are 16, they indicate this as the most likely place where such a (significant) change has occurred.*

11.5 Nonparametric Approaches

In this section we briefly review the nonparametric approaches discussed in Lombard (1986) and Csörgő and Horváth (1996). Let $\theta_1, \theta_2, \ldots, \theta_n$ be n independent observations. Denote by $r_1, r_2, \ldots, r_n$ the ranks of $\theta_1, \theta_2, \ldots, \theta_n$ i.e.

$$r_i = \text{the number of} \quad \theta' s \leq \theta_i, \ (i = 1, \ldots, n).$$

For any fixed integer k, $(1 \leq k < n)$, consider the two-sample problems in which $\theta_1, \theta_2, \ldots, \theta_k$ are supposed to have cdf F_1, $\theta_{k+1}, \ldots, \theta_n$, have cdf F_2 and $H_0 : F_1 = F_2$ is to be tested against $H_1 : F_1 \neq F_2$. As in the Wheeler-Watson two-sample test (see Section 7.3.2), consider the squared length of the resultant for the first k observations based on their uniform scores, namely

$$u_{k,n}^2 = \frac{\left(\sum_1^k \cos(2\pi r_i/n)\right)^2 + \left(\sum_1^k \sin(2\pi r_i/n)\right)^2}{k(n-k)}.$$

We reject H_0 for large values of this statistic. An intuitively sensible estimate $\hat{\tau}$, of a single change point is the k, for which the associated two-sample test is most significant i.e.,

$$\hat{\tau} = \min\left\{ j : u_{j,n}^2 = \max_{1 \leq k < n} u_{k,n}^2 \right\}.$$

Since testing the presence of single change point is equivalent to testing H_0 against $\bigcup_{1 \leq k < n} H_1$, a significance test could be based on some reasonable function of $u_{k,n}^2$, $(1 \leq k < n)$. Lombard (1986) proposes that H_0 be rejected for large values of the statistic

$$a_n^2 = 2 \sum_{1 \leq k < n} u_{k,n}^2$$

and show that for large values of n and w,

$$P\left\{a_n^2 > w | H_0\right\} \approx 3e^{-w}.$$

An alternative test, more sensitive to changes occurring in the central part of the series, can be obtained by introducing a weight factor $k(n-k)$ as in

$$c_n^2 = (\pi/n)^2 \sum_{1 \leq k < n} k(n-k) u_{k,n}^2.$$

For n and w large, one can again use the approximation

$$P\left\{c_n^2 > w | H_0\right\} \approx 2e^{-w}.$$

The procedure can be generalized to the case of multiple change points, the details of which can be found in the papers cited above.

11.6 Other Approaches

In this section, we discuss briefly some alternate approaches to addressing the change-point problem including semi-Bayesian and hierarchical Bayes approaches. See SenGupta and Laha (1999).

If one has some a priori information regarding the possible location of the change point, this can be quantified in the form of a prior distribution. Let p_i denote the probability that 'i' is the point of change, $1 \leq i \leq n-1$. Let p_n denote the probability of no change, $\sum_{i=1}^{n} p_i = 1$. Then the likelihood under H_0 is,

$$L_0(\underline{\theta}) = \frac{p_n}{\{2\pi I_0(\kappa)\}^n} \exp\left[\kappa \sum_{i=1}^{n} \cos(\theta_i - \mu_0)\right]$$

and that under H_1 is,

$$L_1(\underline{\theta}) = \frac{1}{(2\pi I_0(\kappa))^n} \sum_{i=1}^{n-1} p_i \exp\left[\kappa\left\{\sum_{j=1}^{i} \cos(\theta_j - \mu_0) + \sum_{j=i+1}^{n} \cos(\theta_j - \mu_1)\right\}\right].$$

Then the Bayes factor is,

$$\frac{L_1(\underline{\theta})}{L_0(\underline{\theta})} = \sum_{i=1}^{n-1} \frac{p_i}{p_n} \exp\left[\kappa\left\{\sum_{j=i+1}^{n} \cos(\theta_j - \mu_1) - \sum_{j=i+1}^{n} \cos(\theta_j - \mu_0)\right\}\right]$$

which after some simplification becomes

$$\frac{L_1(\underline{\theta})}{L_0(\underline{\theta})} = C\sum_{i=1}^{n-1} \frac{p_i}{p_n} \exp\left(\sum_{j=i+1}^{n} \sin\left(\theta - \frac{\mu_1+\mu_0}{2}\right)\right),$$

where C is a constant. A test for H_0 can then be based on this Bayes factor when μ_0 and μ_1 are known, and it is rejected if the Bayes factor is large.

Example: If $p_i \propto\ e^{-\lambda}\ \lambda^i/i!$, $1 \leq i \leq n, (\lambda$ known) then $p_i/p_n = n!\lambda^{i-n}/i!$, $1 \leq i \leq n$ and

$$\frac{L_1(\underline{\theta})}{L_0(\underline{\theta})} = C\sum_{i=1}^{n-1} \frac{\lambda^{i-n}}{i!}\ \exp\ \left(\sum_{j=i+1}^{n} \sin\left(\theta_j - \frac{\mu_1+\mu_0}{2}\right)\right),$$

where C is a constant. H_0 is rejected if $L_1(\underline{\theta})/L_0(\underline{\theta})$ is large.

On the other hand, if one cannot fully specify the prior parameters, a way out is the hierarchical Bayes approach (Berger (1985)) where we place a further prior on the prior parameters. For the change point we assume a prior on the change point. The prior is specified up to its form but its parameters are left unspecified. We then specify a prior on these unspecified parameters. Then to get the estimate of change point we merely find the posterior distribution of the change point given the data and find the posterior mode or a similar quantity of interest.

To illustrate this approach, suppose that the change point is a random variable denoted by $\boldsymbol{K}$. One may assume that the prior probability of the event $\boldsymbol{K} = k$ is given by

$$P(\boldsymbol{K} = k) \propto \binom{n}{k}\ p^k(1-p)^{n-k}, k = 1, \ldots, n.$$

Assume further that p is distributed uniformly over (0,1). Writing

$$h(k) = \exp\left[\kappa\left\{\sum_{j=1}^{k} \cos(\theta_j - \mu_0) + \sum_{j=k+1}^{n} \cos(\theta_j - \mu_1)\right\}\right], \qquad (11.6.1)$$

the posterior distribution of $\boldsymbol{K}$ is obtained as

$$P\left(\boldsymbol{K}=k|\theta_1\ldots,\theta_n\right)=h(k)\int_0^1\frac{\binom{n}{k}p^k(1-p)^{n-k}}{\sum_{\ell=1}^{n}\binom{n}{\ell}p^{\ell}(1-p)^{n-\ell}}dp,$$

whose mode, for instance provides the most likely change-point.

Remark 11.1 *Detecting changes just in the concentration parameter κ in a CN distribution are of importance and can be developed along similar lines. See for instance, Ghosh (2000) and SenGupta and Laha (1999) The ideas discussed in this chapter can be extended to the case when the data is either dependent in some fashion or when we deal with a directional process, say a time-series. Such extensions have many practical applications and need further investigation.*

Chapter 12

Miscellaneous Topics

12.1 Introduction

In previous chapters we discussed what we consider are some of the important areas of directional data. However there are many other interesting topics that we did not have a chance to explore. In this chapter we provide a brief discussion of some of these topics, hoping to provide the interested reader with leads for further more detailed research.

12.2 An Entropy-based Test for the CND

For a distribution function F with density function f on R^1, entropy is measured by the integral

$$H(f) = -\int_{-\infty}^{\infty} f(x)\,\log f(x)\,dx.$$

This may also be re-expressed as

$$H(f) = \int_0^1 \log\left\{\frac{d}{dp}F^{-1}(p)\right\} dp.$$

If $x_{(i)}$ denotes the ith order statistic, then for $m < n/2$, the differences $\{x_{(i+m)} - x_{(i-m)}\}$ are called the $2m$-step spacings (see Section 7.2.3 for a discussion on spacings) and a consistent nonparametric estimate of $H(f)$ is provided by

$$
\begin{aligned}
H_{mn} &= \frac{1}{n}\sum_{i=1}^{n}\log\left\{\frac{n}{2m}\left(x_{(i+m)} - x_{(i-m)}\right)\right\} \\
&= \log\left\{\frac{n}{2m}\prod_{i=1}^{n}\left[x_{(i+m)} - x_{(i-m)}\right]^{\frac{1}{n}}\right\}. \qquad (12.2.1)
\end{aligned}
$$

Consistency of H_{mn}, as $n \to \infty$, $m \to \infty$, and $m/n \to 0$, is proved by Vasicek (1976).

Lund and Jammalamadaka (2000) provide a statistic to test the goodness-of-fit of the von Mises distribution based on its maximum entropy characterization on the circle, subject to having a given mean direction and concentration. See Property G of the CND in Subsection 2.2.4.

To assess the goodness-of-fit of a von Mises distribution for a set of data, the sample entropy (12.2.1) can be compared to the actual entropy of the von Mises distribution, which is given by

$$
\begin{aligned}
H(f) &= -\int_{-\infty}^{\infty} f(\theta)\,\log f(\theta)\,d\theta \\
&= -\int_{-\infty}^{\infty} \frac{\exp\{\kappa\,\cos(\theta-\mu)\}}{2\pi I_0(\kappa)}\left[-\log\left(2\pi I_0(\kappa)\right) + \kappa\,\cos(\theta-\mu)\right]d\theta \\
&= \log\left[2\pi I_0(\kappa)\right] - \kappa\,A(\kappa) \\
&= \log\left[\frac{2\pi I_0(\kappa)}{\exp(\kappa\,A(\kappa))}\right]. \qquad (12.2.2)
\end{aligned}
$$

This can be consistently estimated by replacing the κ here by a consistent estimate $\hat{\kappa}$, say its MLE, obtaining

$$
H(\hat{f}) = \log\left[\frac{2\pi I_0(\hat{\kappa})}{\exp(\hat{\kappa}\,A(\hat{\kappa}))}\right]. \qquad (12.2.3)
$$

On the other hand, H_{mn} given in (12.2.1) is a consistent estimate of the entropy and can be no larger than that for the CN distribution. However, *under the null hypothesis* that the sample of observations $x_1, \cdots, x_n$ is from a von Mises distribution, it is a consistent estimate of the CN entropy $H(f)$, and therefore converges to (12.2.2). That is, under H_0,

$$
H_{mn} \xrightarrow{p} \log\left[\frac{2\pi I_0(\kappa)}{\exp(\kappa\,A(\kappa))}\right].
$$

We now consider the ratio of H_{mn} and $H(\hat{f})$ given in (12.2.1) and (12.2.3), respectively, to get

$$\begin{aligned} K_{mn} &= \frac{\exp\{H_{mn}\}\,\exp\{\hat{\kappa}\,A(\hat{\kappa})\}}{I_0(\hat{\kappa})} \\ &= \frac{n\,\exp\{\hat{\kappa}\,A(\hat{\kappa})\}}{2\,m\,I_0(\hat{\kappa})}\left[\prod_{i=1}^{n}\left(x_{(i+m)} - x_{(i-m)}\right)\right]^{\frac{1}{n}}. \end{aligned} \tag{12.2.4}$$

Then, under the null hypothesis,

$$K_{mn} \xrightarrow{p} \frac{\exp\{H(f)\}\,\exp\{\kappa\,A(\kappa)\}}{I_0(\kappa)} = 2\pi$$

as $n \to \infty$, $m \to \infty$ and $m/n \to 0$. However, samples coming from distributions that are not von Mises, will tend to have *lower* sample entropies, and thus smaller values of K_{mn}. The null hypothesis of circular normality is therefore rejected for sufficiently small values of K_{mn}.

Critical values of the test statistic, obtained by Monte Carlo simulation, are given in Table 12.1 for selected values of n and κ. For each of various sample sizes and values of κ, 5000 samples were generated from a von Mises distribution. From each sample, K_{mn} was calculated. A condensed table of 95th and 99th percentiles are given in Table 12.1. The choice of step size, m in the table of critical values was made depending on which m yielded the largest value for the test statistic K_{mn}. Some smaller scale simulations were performed to compare powers using various step sizes, and it was usually found that the step size yielding the largest critical values also yielded the best power. See also Dudewicz and van der Meulen (1981).

In practice, κ is generally unknown, in which case it is suggested to use the maximum likelihood estimate of κ to decide which row to use in the tabulated critical values. Lund and Jammalamadaka (2000) show that using a consistent estimate for κ in place of the unknown κ does not adversely affect the significance level of the test.

To calculate the statistic

$$K_{mn} = \frac{n\,\exp\{\hat{\kappa}\,A(\hat{\kappa})\}}{2\,m\,I_0(\hat{\kappa})}\left[\prod_{i=1}^{n}\left(x_{(i+m)} - x_{(i-m)}\right)\right]^{\frac{1}{n}},$$

we need to compute the $2m$-step spacings and a consistent estimate for κ. The maximum likelihood estimate $\hat{\kappa}$ of κ is consistent and is given by solving Equation (4.2.5) or using the routine `est.kappa` in `CircStats`.

Table 12.1: Critical values for testing at the 5% (1%) significance level.

	Sample Size				
κ	n = 20	n= 25	n = 30	n = 40	n = 50
Step Size	m = 3	m = 4	m = 4	m = 5	m = 5
0.20	4.11(3.81)	4.39(4.12)	4.57(4.30)	4.82(4.61)	4.99(4.83)
0.40	4.15(3.81)	4.40(4.12)	4.58(4.32)	4.84(4.59)	5.01(4.82)
0.60	4.17(3.83)	4.43(4.16)	4.61(4.37)	4.85(4.65)	5.03(4.86)
0.80	4.16(3.77)	4.44(4.14)	4.62(4.34)	4.89(4.68)	5.08(4.89)
1.00	4.19(3.83)	4.44(4.16)	4.66(4.36)	4.91(4.69)	5.08(4.90)
1.20	4.21(3.82)	4.44(4.13)	4.67(4.42)	4.92(4.72)	5.10(4.88)
1.40	4.23(3.92)	4.46(4.15)	4.66(4.38)	4.93(4.69)	5.11(4.89)
1.60	4.20(3.83)	4.46(4.13)	4.67(4.41)	4.93(4.68)	5.12(4.93)
1.80	4.21(3.88)	4.48(4.13)	4.65(4.40)	4.94(4.69)	5.12(4.91)
2.00	4.20(3.87)	4.44(4.13)	4.67(4.43)	4.94(4.72)	5.13(4.92)
2.20	4.22(3.86)	4.45(4.11)	4.66(4.40)	4.93(4.69)	5.12(4.93)
2.40	4.20(3.83)	4.45(4.14)	4.66(4.41)	4.92(4.67)	5.11(4.86)
2.60	4.20(3.85)	4.48(4.17)	4.65(4.39)	4.93(4.71)	5.12(4.93)
2.80	4.21(3.82)	4.45(4.14)	4.67(4.38)	4.94(4.68)	5.11(4.94)
$\geq$ 3.00	4.21(3.85)	4.46(4.16)	4.66(4.39)	4.92(4.69)	5.10(4.90)

Remark 12.1 *There are two ways to define the spacings on the circle. One method is to utilize the circularity of the observations, and taking*

$$x_{(i+m)} = 2\pi + x_{(i+m)(mod\ n)}$$

for values of $(i+m)$ *larger than* n*. This definition of spacings fully exploits the circularity of the data and makes* H_{mn} *invariant with respect to the zero location. However, in some cases, especially for large values of* κ*, it has been observed that such a circular definition of spacings leads to a rather poor approximation of the true entropy.*

In contrast, if the data is being treated as being on a line, one may use truncation and define

$$\begin{aligned} x_{(i+m)} &= x_{(n)} \text{ for } i+m > n \\ \text{and} \quad x_{(i-m)} &= x_{(1)} \text{ for } i-m < 1. \end{aligned}$$

However, under this definition of spacings, H_{mn} will not be invariant under rotations. One resolution to this problem is to define $x_{(0)}$ and $x_{(n)}$ such that the largest gap between adjacent observations occurs between $x_{(0)}$ and $x_{(n)}$ (see Figure 12.1). This is equivalent to finding the largest gap and cutting open

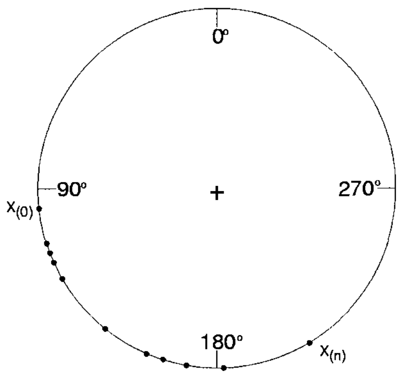


Figure 12.1: Defining the order statistics.

circle and is connected to the idea of circular range discussed in (7.2.16). Under this definition of the order statistics, the spacings will be invariant under rotations, and hence H_{mn} and K_{mn} will be as well.

A final remark about the calculation of the test statistic, K_{mn} concerns ties in the data set. As the statistic is a function of the product of $2m$-step spacings, any ties among $2m$ successive observations will set the test statistic equal to zero. It is therefore essential that the data set does not contain too many ties, if the entropy statistic is to be used. Theoretically, as the data is assumed to be from a continuous distribution, ties occur with probability zero. Of course, in practice it is not uncommon to encounter ties due to rounding and discretization of the data. One solution is to employ fixed-variability jittering of the data, as done in the context of density estimation.

Computationally, the entropy statistic, K_{mn}, is much less cumbersome than some of the alternatives such as Watson's U^2 (see Chapter 7), which requires the evaluation of the von Mises cdf. Since there is no closed form solution for the von Mises distribution function, this makes the computation of the U^2 statistic considerably more involved than the computation of K_{mn}. However, such computational edge would not be meaningful if the entropy statistic did not compete in terms of power. Lund and Jammalamadaka (2000) perform power comparisons between the entropy statistic and two competing statistics: Watson's U^2 and another density based, integrated squared error test statistic, which also requires evaluation of the von Mises distribution function. Although none of the statistics were uniformly better than the others, over the alternative distributions simulated form, it is shown there that U^2 performed best for a long-tailed mixture of von Mises distributions, while K_{mn} performed best for the half von Mises. K_{mn} was also shown to out-perform U^2 when the data are from a cardioid or from a triangular distribution.

Remark 12.2 *The maximum entropy property of the von Mises distribution is over the class of all probability distributions on the circle with given mean direction and concentration. Since the entropy statistic K_{mn}, is based on this characterization, it should not be used to distinguish between the von Mises distribution and a distribution which is outside this class, say e.g., the uniform distribution on $[0, 2\pi)$ whose mean direction is not defined and which has larger entropy than the von Mises. Instead, to test between the uniform and von Mises distributions, one would use the Rayleigh test or other tests described in Chapters 5 through 7.*

12.3 Classification and Discriminant Analysis

Consider the problem of classifying a new observation θ, into one of two distinct circular populations. Suppose we have observations from these two (identifiable) populations as $\{\theta_{ij},\ j = 1, \ldots, n_i,\ i = 1, 2\}$, which can be used as training samples. Morris and Laycock (1974) discuss the usual Fisher's discrimination rule for CN populations and discuss sampling distributions for related discriminant statistics as well as error probabilities. Collett and Lewis (1981) consider the problem of discriminating between the CN and WN distributions.

We now introduce a general discrimination rule based on a circular distance which may be used for any two parametric circular distributions. The basic idea is to find how far the new observation is, to either of the given samples and assign it to the population to which it is "closer."

Recall that for any two points on the unit circle, say (θ_i, θ_j), we defined (cf 1.3.7) the "circular distance"

$$d_{ij} = 1 - \cos(\theta_i - \theta_j) \tag{12.3.1}$$

which is non-negative, symmetric in its indices and is invariant under rotation. The average distance $d_i(\theta)$ of a new observation θ from the group i, is then given by

$$d_i(\theta) = 1 - \frac{1}{n_i}\sum_j \cos(\theta_{ij} - \theta). \tag{12.3.2}$$

Note that this is the same as the circular dispersion of the sample i around the given direction θ (cf. equation 1.3.8). Using the standard notations,

$$\overline{C}_i = \frac{1}{n_i}\sum_j \cos\theta_{ij}, \overline{S}_i = \frac{1}{n_i}\sum_j \sin\theta_{ij},\ \overline{R_i} = \sqrt{\overline{C}_i^2 + \overline{S}_i^2},\ \overline{\theta}_i = \arctan^* \frac{\overline{S}_i}{\overline{C}_i},$$

one may rewrite this as

$$d_i(\theta) = 1 - \overline{R_i}\cos(\overline{\theta_i} - \theta) = 1 - \frac{V_i}{n_i},$$

where V_i is the length of the projection of the ith sample resultant towards the direction θ. Our rule is to classify θ as belonging to population 1 if $d_1(\theta) < d_2(\theta)$ i.e., if

$$\frac{V_1}{n_1} > \frac{V_2}{n_2}.$$

For the case when the two samples sizes are the same and the concentrations are equal, this corresponds to classifying θ as belonging to that group whose sample mean direction is closer and is clearly very heuristic. For the CN model, the sampling distribution of this classification statistic can be obtained as was done in Roy (1999), who also provides an empirical comparison of the error probabilities of this rule with other competitors. Lund (1999) considers clustering problems for directional data.

12.4 Factorial Designs

Problems in design of experiments for circular data have been considered by several authors. The approximate ANOVA test due to Watson and Williams (1956), presented earlier in Section 5.3.1 for the equality of mean directions for several independent von Mises distributions with common unknown concentration parameter, is the simplest such example and corresponds to the case of one-way classified data. Except for the conditional arguments based on the distribution of the lengths of individual resultants given the overall resultant (see Equation (3.5.3) and the discussion there), no exact distribution theory is available even for the case of one-way ANOVA.

Unlike in the linear case, there are some difficulties associated with the elimination of the nuisance parameter κ which is not a scale parameter. The non-orthogonality of the components of variation as well as the non-availability of the exact distribution theory for an appropriate test statistic, have led to several alternate proposals and approaches to this problem. Recall that in this approximate ANOVA, the χ^2 distributions for the components of variation as well as the orthogonality of the components, are achieved only for large κ. The distributions of the statistics depend on the nuisance parameter κ when it is small, say $\kappa < 1$, and Upton and Fingleton (1989) review various tests for such situations. Alternatively, Fisher (1993) on the grounds of simplicity, advocates a randomization test for situations where the approximate ANOVA is deemed inappropriate, e.g. judging from the sample sizes and/or differences in $\hat{\kappa}$. Homogeneity tests discussed in Section 7.4.1 are also appropriate in this context, especially when dealing with large samples.

Factorial designs where differences in mean directions are of interest have been considered for instance, by Underwood and Chapman (1985), Chapman and Underwood (1992). They employ the randomization approach for inference, which though tedious, is expected to be more robust than the other approaches. Factorial designs where the directional spread or angular dispersion of a directional response is of interest, as in the automotive industry, have been considered in Anderson and Wu (1995).

These approaches have found straightforward generalizations to the more complex designs. Based on the approximate distribution theory, factorial designs and multiway ANOVA etc. have been considered for von Mises and other distributions. Extensions of the above approach were proposed among others by Stephens (1982) for nested 2-way designs, by Bijleveld and Commandeur (1987) to m-way designs and by Underwood and Chapman (1985) and for crossed designs. However, for crossed layouts and designs involving many factors, negative interactions which are difficult to interpret may occur frequently. Based on a large number of simulations using a variety of main and interaction effects and κ values, Anderson and Wu (1995) noted the prevalence of the number of observed negative interaction sums of squares and thus concluded that the extension of Underwood and Chapman (1985) "... does not appear suitable for analysis of designed experiments where interactions are of interest."

As we noted before, circular data may be represented as unit vectors $\mathbf{x}$, instead of their polar coordinate representations. This introduces the embedding of the circle S^1 as a subset of R^2, restricting $\mathbf{x}$ to be unit vectors, i.e. to the set $S^1 = \{\mathbf{x} : \mathbf{x}'\mathbf{x} = 1\}$. The usual L_2 distance may now be used for the $\mathbf{x}$ observations - i.e.,

$$d(\mathbf{x}, \mathbf{y}) = |\mathbf{x} - \mathbf{y}|^2 = |\mathbf{x}|^2 + |\mathbf{y}|^2 - 2\mathbf{x}'\mathbf{y} = 2(1 - \cos\theta),$$

where θ is the angle in between, and amounts to using the chord length as the measure of closeness (see Figure 1.8). The L_2 decomposition with the $\mathbf{x}_{ij}$ observations then yields sums of squares representing the corresponding sums of squared chord lengths. Harrison et al. (1986) and Harrison and Kanji (1988) use this approach and develop a test which rejects equality of mean directions for large values of the statistic,

$$F^* = \left(1 - \frac{.2}{\hat{\kappa}} - \frac{.1}{\hat{\kappa}^2}\right) \frac{\left(\sum n_i \bar{R}_i^2 - \bar{R}^2\right)/(p-1)}{\left(n - \sum n_i \bar{R}_i^2\right)/(n-p)}.$$

For large κ, F^* has $F_{p-1,n-p}$ as the approximate null distribution. This approach may be extended to multi-way classified circular data. Since the total sums of squares is decomposed into marginal sum of squares using a regular distance measure, the triangle inequality is obeyed and the partitioned sums of squares are all non-negative. Harrison et al. (1986) and Harrison and Kanji (1988) illustrate this for the case of a balanced two-way design, decomposing the total sum of squares as

$$SS_T = SS_R + SS_C + SS_I + SS_E$$

in standard notations, with the components on the right hand side representing the rows, columns, interaction and error, respectively.

Although this approach overcomes the problem of possible negative interactions, as Anderson and Wu (1995) point out, the above sums of squares reflect the combined location and dispersion measure of differences between groups, and not the location effects alone. A zero main effect then need not result when locations are all equal, but demands that both the mean directions and the lengths of the resultant vectors remain the same for all levels of the factors. This renders the above approach inappropriate for testing the homogeneity of mean directions alone.

Since ANOVA may be viewed as a particular case of regression analysis and of Generalized Linear Models, Fisher and Lee (1992) consider modeling the mean direction μ_i of von Mises distribution CN (μ_i, κ) using link function:

$$\mu_i = \mu + g\left(\beta' X_i\right).$$

One can construct an orthogonal array filled with columns of indicator variables for the different factor levels. The above regression then enables estimation of the β's and hence that of the factor effects. However, Anderson and Wu (1995) demonstrate that with the above suggested link function, interpretation of the results for factorial effects is difficult and hence this approach is not as suited for factorial experiments as might first appear.

12.4.1 Likelihood Ratio Tests

Given the lack of any exact or even asymptotic approach for all κ, one can fall back on the likelihood ratio test (LRT) for this situation (see Rao (1969)). For known κ, the LRT has a simple form and from the standard theory, converges to a χ^2_{p-1} distribution for large samples. An improvement in the convergence was achieved by Paula and Botter (1994) by incorporating a multiplicative factor to this LR statistic, yielding

$$2\kappa\left[1 - \frac{\dot{A}(\kappa)}{4\kappa}\left(\sum_{i=1}^{p}\frac{1}{n_i} - \frac{1}{n}\right)\right]\sum_{i=1}^{p} R_i\left(1 - \cos(\bar{\theta}_i - \bar{\theta})\right).$$

Anderson and Wu (1995) suggest replacing an unknown κ by its MLE $\hat{\kappa}$. They assert that this new statistic S^* also follows an asymptotic χ^2_{p-1} distribution and conclude through simulations, that S^* competes well with the approximate ANOVA test and is robust for wrapped Cauchy distributions. Both these approaches were found to be better than that of Harrison et al. (1986), Harrison and Kanji (1988), which may also not be robust in the presence of outliers and departures from von Mises distributions. Anderson and Wu (1996) extend this approach to the case of multi-way and factorial designs, exploring its performance under a variety of situations. When differences in exclusively the mean directional effects in the response are of interest, this approach is recommended. We should however note that this is *not* the LRT for the ANOVA problem - the actual LR statistic being a function of the different estimators of κ obtained separately under H_0 as well as under H_1.

Observe that ANOVA presumes the homogeneity of the concentration parameters of the von Mises distributions. The validity of this assumption can be verified using the homogeneity test statistic defined in Equation (7.4.7) for moderately large sample sizes and using the routine `rao.homogeneity` in `CircStats`. On the other hand, if the sample sizes are not large enough but the concentrations are high (say with $\bar{R} > 0.7$), one can invoke the normal approximation and do the straightforward Bartlett's test of homogeneity of variances for normal distributions. Finally, if these κ_i's are indeed heterogeneous, none of the above tests should be used. Then the LRT or the homogeneity test of Section 7.4.1 should be employed. A simplified adaptation of the LRT for this case is suggested by Watson (1983a), and is given by

$$T = 2\left(\sum_{i=1}^{p} \hat{\kappa}_i R_i - R_C\right),$$

where

$$R_C^2 = \left(\sum_{i=1}^{p} \hat{\kappa}_i R_i \cos\bar{\theta}_i\right)^2 + \left(\sum_{i=1}^{p} \hat{\kappa}_i R_i \sin\bar{\theta}_i\right)^2,$$

with a χ^2_{p-1} distribution under H_0. ANOVA for distributions other than von Mises have also been proposed, e.g. by Mardia and Spurr (1973) for ℓ-modal von Mises distributions and Anderson-Cook (2000) for one-way analysis of cylindrical data.

The problem of optimally allocating sample sizes in a hierarchical design with a fixed cost, has been considered in Rao and Sengupta (1970). In yet another direction, Wu (1997) considers the choice of the optimal location of the points to be sampled on a circle and develop Φ-optimal exact and approximate designs on a circle.

12.5 Bayesian Analysis

Attempts at Bayesian inference for circular data have not been as successful on the analytical front as they have been for the linear case, partly for lack of nice conjugate priors in the general case. In connection with the von Mises model, one of the early attempts is due to Schmitt (1969) who discussed the problem of the choice of appropriate priors for μ and κ, who cautions that unlike σ^2 for the normal distribution, κ for a von Mises distribution can actually assume the value zero. In the simplest of the cases where κ is known and we assume a uniform prior for μ, the posterior density is given by

$$p\left(\mu, \kappa | \theta_1, ..., \theta_n\right) \quad \propto \quad \frac{1}{I_0^n(\kappa)} \quad \exp\left[\kappa R \cos(\mu - \bar{\theta})\right].$$

Using numerical techniques, one can construct the Highest Posterior Density (HPD) credible sets for μ (see Lee (1989)). Mardia and El-Atoum (1976), Bagchi (1987) and Bagchi and Kadane (1991) consider the case of known κ while Lenth (1981) and Guttorp and Lockhart (1988) consider the case when κ is also unknown, using the posterior mode as an estimate for it. However, recall that when both μ and κ are unknown, the von Mises distribution is a member of a regular exponential family and hence a family of conjugate priors may be easily constructed. Such a general conjugate prior does not have a simple form. Guttorp and Lockhart (1988) suggest the prior given by

$$p(\mu, \kappa) \quad \propto \quad \frac{1}{I_0^c(\kappa)} \quad \exp\left[\kappa R_0 \cos(\mu - \mu_0)\right].$$

If $R_0 = 0$ this results in the uniform prior for μ. Further, if c is taken to be a non-negative integer, then μ_0 and R_0 can be given simple interpretations. Damien and Walker (1999) consider the posterior distribution

$$p(\mu, \kappa | \theta_1, ..., \theta_n) \quad \propto \quad \frac{1}{I_0^n(\kappa)} \quad \exp\left[\kappa R_n \ \cos(\mu - \mu_n)\right].$$

By the introduction of strategic latent variables, they represent the posterior distribution as one which has full conditional distributions of known types. A Gibbs sampler may then be conveniently implemented to simulate samples from this posterior, enabling one to pursue a full Bayesian analysis.

For the von Mises distribution, Bagchi and Guttman (1990) propose a Bayesian analysis of outliers. Bagchi and Kadane (1991) provide Laplace approximations to posterior moments. Zhong (1991) gives Bayes estimator for μ by using a general prior distribution for both μ and κ, thereby giving an exact approach and avoiding the first order approximation to the Bessel function used by Bagchi (1987) for the case of unknown κ. SenGupta and Maitra (1998) derive Bayes estimate of the mean directions of several independent von Mises distributions, and establish the admissibility of the simultaneous MLE of these mean directions. Zhong (1991) studies empirical Bayes methods for estimating μ by exploiting the fact that a CN distribution can be approximated by a wrapped normal and then estimating the hyper-parameters of the prior density function.

Bayesian analysis for circular models other than the von Mises, has not made much progress. Coles (1998) formulates a Bayesian model for the analysis of wrapped circular distributions, wherein the uncertainty in the missing wrapping coefficient values is averaged out to perform inference. Though such models are considerably more complex to deal with analytically, Markov Chain Monte Carlo algorithms should come in handy in dealing with them computationally.

12.6 Sequential Methods

There are many practical situations where it is desirable to update the decision with every incoming observation, i.e. sequentially, either in the temporal or in the spatial mode of collecting the circular data. Usually for simplicity of analysis it is assumed however that except for the natural sequencing, the observations are iid. In many practical situations, this may not be the case. For instance observations on the directions of imbalance in wheels as they are being produced one by one, may enable online identification of whether the process is going out of control. This is an example where a sequential test for circular data needs to be employed. Gadsden and Kanji (1981) and Gadsden and Kanji (1982) consider the construction of SPRTs for the mean direction of von Mises distributions with known and unknown, but large κ. Shepherd and Fisher (1982) cite an example from a coal mine, where the data are recorded at 20-meter intervals along a traverse of a coal seam and where a shift in the mean direction is indicative of a possible forthcoming hazard. This may be handled as a one-sided SPRT or a sequential change-point problem for circular data. They propose an exploratory plotting technique, adapted from the CUSUM plot, for arriving at some preliminary conclusions for this situation.

12.7 Shape Analysis

Shape analysis has many interesting links with the analysis of circular data. For instance, the "preshape" distributions which must be invariant under location and scale transforms, can be obtained from certain circular distributions- see for instance Engebretson and Beck (1978), Mardia and Dryden (1993).

An interesting application where circular data are used in conjunction with shape analysis is in a study of herding habits of dinosaurs cited in Small (1996). An excellent starting point for studying shape analysis, is Mardia and Jupp (2000), Chapter 14, with Small (1996) and Mardia and Dryden (1998) providing a more thorough coverage.

12.8 Stochastic Processes and Time Series

Unlike the iid situations that we have been dealing with so far, there are indeed many practical situations where a strong temporal or spatial dependence between successive data points exists. See for example Breckling (1989), Coles (1998), Craig (1988) and Lai (1994). A common example of this is provided by the hourly or daily mean wind directions at some location. Time series analysis of circular data is then needed for which measures of dependence and multivariate distributions for circular random variables are the usual prerequisites. Let $\{\mathbf{x}_t, t = 1, \ldots, T\}$ be a circular time series, represented in terms of the rectangular coordinates. The earliest attempt to measure the strength of dependence among such serial random variables may be attributed to Watson and Beran (1967) who use the serial coefficient,

$$\rho_{WB} = \frac{1}{T} \sum_{t=1}^{T} \mathbf{x}'_{t-1} \mathbf{x}_t.$$

Lai (1994) uses this measure to explore the correlation properties of some time series models. One such model is based on the construction of positive and negative autoregressive processes depending on whether ρ_{WB} is positive or negative. The process was constructed by adapting the approach of Acardi et al. (1987), viz. by using the tangent-normal decomposition of a distribution on $\mathbf{S}^{n-1}$ to the circular case. The parameters were estimated by the method of moments and sample paths for such processes examined by simulations.

A nice multivariate von Mises distribution whose marginals are von Mises (without being independent), is not readily available. Breckling (1989) introduces a conditional von Mises process and gives the likelihood equations for estimating the parameters. Lai (1994) supplements this work by studying the behavior of this von Mises process for different parameters and different order combinations and proposing a Lagrangian multiplier test and the Akaike Information Criterion (AIC) for identification and order determination of the process. Breckling (1989) also studies the properties of the first order wrapped autoregressive process which may be approximated by a first order von Mises process and provides several techniques for estimation of the autocovariance function. Using link functions, Fisher and Lee (1994) define a linked $ARMA(p,q)$ process and propose some new methods for fitting autoregressive models to circular data based on them. One such method, called the circular $AR(p)$ process, was constructed by specifying the conditional distribution of Θ_t given the history $\Theta_{t-i} = \theta_{t-i}, i = 1, 2, \ldots,$ as also a von Mises distribution $\mathrm{CN}(\mu_t, \kappa)$. Where the mean direction μ_t was modeled as,

$$\mu_t = \mu + g\left[\sum_{i=1}^{p} \alpha_i g^{-1}\left(\theta_{t-i} - \mu\right)\right]$$

for a suitably chosen link function $g(x)$. Craig (1988) examines several aspects of Markov processes on the circle. He establishes the existence of a unique stationary measure for given transition function arising out of a bivariate circular distribution and then examines the conditional behavior and the constraints imposed by stationarity upon the transition functions of the Markov models derived from various distributions including the wrapped bivariate normal distribution. This was also extended to define a model for ℓ-th order Markov process as follows,

$$f_{X_t|\mathbf{x}_1^{t-1}}(x_t|\mathbf{x}_1^{t-1}) = f_{X_t|\mathbf{x}_{t-1}^{t-1}}(x_t|\mathbf{x}_{t-1}^{t-1}) = \sum_{i=1}^{\ell} \lambda_j g_j(x_t|x_{t-j}),$$

where some initial distribution is chosen for $\mathbf{X}_1^{\ell}$ and g_j's are some bivariate conditional densities. Note that this formulation does not require any joint distribution higher than a bivariate one.

Wrapped distributions have also been used in the construction of models based on the ARMA family. Component-wise wrapping of any suitable multivariate distribution $g(.)$ on IR^p yields a multivariate distribution, say $f(.)$ on $(S^1)^p$, i.e. each marginal is a circular random variable. Similarly from Y_t, a real-valued stochastic process, one can derive a stochastic process on the circle by component-wise wrapping,

$$f_{(x_{t_1},\ldots,x_{t_p})} = \sum_{k^p \epsilon Z^p} f_{(Y_{t_1},\ldots,Y_{t_p})}(x_{t_1} + 2\pi k_1, \ldots, x_{t_p} + 2\pi k_p).$$

Breckling (1989) gives some exploratory techniques using the correlogram for such wrapped processes. Since the k_j's manifest as "missing" data, the EM algorithm was adopted by Fisher and Lee (1994) for the estimation of the model parameters. Coles (1998) suggests an alternative Bayesian approach for inference by constructing smoothed posterior density estimates of the associated parameters in the fit of a wrapped second-order autoregressive process with an underlying bivariate normal distribution for the errors. Craig (1988) studies such wrapped processes, in particular the wrapped Gaussian process, defining circular auto-correlation and auto-covariance functions to measure dependence and to perform identification of these models.

12.9 Density Estimation

Estimating the unknown density on the circle or a sphere based on n iid observations, has been studied by Hall et al. (1987) and Bai et al. (1988). See also Klemelä (2000) for results applicable to higher dimensions. All these authors use a kernel estimator, taking care to define the value of the estimator at a given point, to depend on the appropriate distance between that point and the observations. In the case of the circular densities, the appropriate distance can be the one given in Equation (1.3.7). Such density estimators take the form

$$\hat{f}_n(\mathbf{v}) = \frac{1}{n}\sum_{i=1}^{n} k_n(1 - \mathbf{v}'\mathbf{u}_i),$$

where $k_n(.)$ is an appropriate kernel. Note that for the circular case, $\mathbf{v}'\mathbf{u}_i = \cos\theta_i$ where θ_i is the angle between $\mathbf{v}$ and $\mathbf{u}_i$ (see Figure 1.8) so that closeness of the observation $\mathbf{u}_i$ to $\mathbf{v}$ is indicated by how small $(1-\mathbf{v}'\mathbf{u}_i)$ is. Asymptotic properties of these estimators for the spherical case and their rates of convergence are discussed in Hall et al. (1987) while Bai et al. (1988) consider conditions for point-wise, L_1-norm, strong and uniform strong consistency of such a kernel estimator. In related work, Beran (1979) uses such kernel estimators to construct robust estimators of the parameters in exponential models. The main ideas in circular density estimation, are essentially similar to those in the linear case and convergence rates are typically slow for many practical applications.

12.10 Periodic Smoothing Splines

Periodic smoothing splines are a variant of the basic smoothing spline estimators. Such estimators are useful when the mean response function is assumed to be smooth and periodic on an interval $[a, b]$. Without loss of generality, assume that $a = 0$ and $b = 1$, although modifications to the interval $[0, 2\pi)$ is clear. Periodic smoothing splines were introduced by Cogburn and Davis (1974). The analysis presented here is derived from Wahba (1975) and Rice and Rosenblatt (1981). Further references and applications of periodic smoothing splines can be found in Wahba and Wold (1975) and Wahba and Wendelberger (1980).

For simplicity of exposition, assume that

$$t_j = \frac{(j-1)}{n}, \qquad j = 1, \ldots, n$$

are n equi-spaced sample points, where observations $y_j, j = 1, \ldots, n$ are taken. Assuming that the mean response function is periodic and twice differentiable, we can formulate the *periodic smoothing spline* for this case as the minimizer of

$$\frac{1}{n}\sum_{j=1}^{n}(y_j - f(t_j))^2 + \frac{\lambda}{(2\pi)^4}\int_0^1 (f''(t))^2 dt \tag{12.10.1}$$

over all functions f which are twice differentiable and periodic over $[0, 1]$ and whose second derivative is square-integrable. The factor of $(2\pi)^4$ has been introduced just for notational convenience in the subsequent analysis.

To obtain the minimizer in (12.10.1), we use method of Fourier analysis. A function f satisfying the above-mentioned criteria admits the representation of the form

$$f(t) = \sum_{j=-\infty}^{j=\infty} b_j e^{2\pi i j t}$$

with Fourier coefficients

$$b_j = \int_0^1 f(t) e^{-2\pi i j t} dt.$$

Thus to find the minimizer in (12.10.1) it is sufficient to find its Fourier coefficients. Therefore we re-write the expression in (12.10.1) in terms of the Fourier coefficients of the function f. The minimizing function can then be obtained by minimization in terms of these coefficients. Consider the case when n is odd since the case where n is even can be treated analogously.

When n is odd, the vectors

$$\mathbf{x}_j = \left(1, e^{2\pi i j/n}, \ldots, e^{2\pi i j(n-1)/n}\right), \quad j = 0, \pm 1, \ldots, \pm(n-1)/2,$$

form an orthonormal basis for $\mathbb{R}^n$. Thus,

$$\mathbf{y} = \sum_{|j| \leq (n-1)/2} \tilde{\beta}_j \mathbf{x}_j$$

with

$$\tilde{\beta} = \frac{1}{n} \sum_{r=1}^{n} y_r e^{-2\pi i j(r-1)/n}.$$

In particular, this means that

$$y_k = \sum_{|j| \leq (n-1)/2} \tilde{\beta}_j e^{2\pi i j(k-1)/n}, \qquad k = 1, \ldots, n.$$

This same argument along with the Fourier series representation for f, shows that

$$f(t_k) = \sum_{|j| \leq (n-1)/2} c_j e^{2\pi i j(k-1)/n},$$

where

$$c_j = \frac{1}{n} \sum_{r=0}^{n-1} f\left(\frac{r}{n}\right) e^{-2\pi i j r/n} = \sum_{s=-\infty}^{\infty} b_{j+sn}.$$

By substituting the above expression for y_k and $f(t_k)$ in the MSE term of (12.10.1), we get

$$\frac{1}{n}\sum_{k=1}^{n}(y_k - f(t_k))^2 = \sum_{|j|\le(n-1)/2}\left|\tilde{\beta}_j - \sum_{s=-\infty}^{\infty} b_{j+sn}\right|^2. \tag{12.10.2}$$

By using the fact that $f(0) = f(1)$ and $f'(0) = f'(1)$ it can be seen that f'' has Fourier coefficients

$$b_j'' = \int_0^1 f''(t)e^{-2\pi ijt}dt = -(2\pi j)^2 b_j. \tag{12.10.3}$$

Now, using Parseval's relation and combining the expressions (12.10.2) and (12.10.3), we can express (12.10.1) as

$$\sum_{|j|\le(n-1)/2}\left\{\left|\tilde{\beta}_j - \sum_{s=-\infty}^{\infty} b_{j+sn}\right|^2 + \lambda\sum_{s=-\infty}^{\infty}(j+sn)^4\left|b_{j+sn}\right|^2\right\}. \tag{12.10.4}$$

This form of expression has the desired property of depending only on the Fourier coefficients of f.

Observe that (12.10.4) is a sum of n non-negative functions of the Fourier coefficients b_k and there is no overlap in the coefficients which appear in different terms. Thus, we may minimize each term in the sum separately. We skip the details of the optimization procedure which can be found in Section 6.3.1 of Eubank (1988).

Finally, we find the minimizing coefficients to be

$$b_{j+sn} = \frac{\tilde{\beta}_j}{(\lambda + r_j)(j+sn)^4}. \tag{12.10.5}$$

Thus the periodic smoothing spline estimator is now given explicitly, for n odd, by

$$\begin{aligned}\mu_{\lambda p}(t) &= \sum_{|j|\le(n-1)/2}\sum_{s=-\infty}^{\infty} b_{j+sn}e^{2\pi i(j+sn)t} \\ &= \sum_{|j|\le(n-1)/2}\frac{\tilde{\beta}_j}{(\lambda + r_j)j^4}e^{2\pi ijt} \\ &\quad + \sum_{|j|\le(n-1)/2}\frac{\tilde{\beta}_j}{(\lambda + r_j)}\sum_{s\ne0}\frac{1}{(j+sn)^4}e^{2\pi i(j+sn)}. \end{aligned} \tag{12.10.6}$$

Remark 12.3 *Since the periodic smoothing splines can be used to estimate the mean response function, it could provide a reasonable alternative to the circular regression ideas discussed in Chapter 8.*

Appendix A

Some Facts on Bessel Functions

We quote here some facts and results on Bessel functions that will be found useful in the text. These results and more can be found in Watson (1944) and Abramowitz and Stegun (1965), Chapter 9. The Bessel function of ν^{th} order, $J_\nu(z)$ is the solution to the differential equation

$$\frac{d^2x}{dz^2} + \frac{1}{z}\frac{dx}{dz} + \left(1 - \frac{\nu^2}{z^2}\right) x = 0$$

and occurs in mathematical physics in connection with vibrations of a uniformly stretched circular membrane, conduction of heat, etc. It has the integral representation

$$J_\nu(z) = \frac{(z/2)^\nu}{\Gamma(1/2)\Gamma(\nu + 1/2)} \int_0^\pi e^{iz\cos\theta} \sin^{2\nu}\theta d\theta \qquad (\nu > -\frac{1}{2})$$

and the series expansion

$$J_\nu(z) = \sum_{r=0}^{\infty} (-1)^r \frac{(z/2)^{\nu+2r}}{r!\Gamma(\nu + r + 1)}.$$

For a positive integer n, it has the alternate integral representation

$$J_n(z) = \frac{1}{\pi} \int_0^\pi \cos(n\theta - z\sin\theta) d\theta.$$

Also,

$$J_{-\nu}(z) = (-1)^\nu J_\nu(z),$$

and

$$\frac{d}{dz}\{z^{\nu}J_{\nu}(z)\} = z^{\nu}J_{\nu+1}(z), \qquad \nu > 0.$$

In particular,

$$\begin{aligned} J_0(z) &= \frac{1}{2\pi}\int_0^{2\pi} e^{iz}\cos\theta d\theta \\ &= \frac{1}{2\pi}\int_0^{2\pi} \cos(z\cos\theta)d\theta \\ &= \frac{2}{\pi}\int_0^{\pi/2} \cos(z\cos\theta)d\theta \end{aligned}$$

and

$$\frac{d}{dz}(zJ_1(z)) = zJ_0(z).$$

The *modified Bessel function of the first kind and order* ν (sometimes also called Bessel function of purely imaginary argument), is given by

$$\begin{aligned} I_{\nu}(z) &= i^{-\nu}J_{\nu}(iz) \\ &= \frac{1}{2\pi}\int_0^{2\pi} \cos\nu\theta e^{z\cos\theta}d\theta \\ &= \sum_{r=0}^{\infty}\frac{(z/2)^{\nu+2r}}{r!\Gamma(\nu+r+1)}. \end{aligned}$$

In particular

$$I_0(\rho) = \frac{1}{2\pi}\int_0^{2\pi} e^{\rho\cos\alpha}d\alpha$$

provides us the normalizing constant for the von Mises or Circular Normal distribution. Putting $a = \rho\cos\theta, b = \rho\sin\theta$ in the first integral, we have

$$\begin{aligned} &\frac{1}{2\pi}\int_0^{2\pi} e^{a\cos\alpha+b\sin\alpha}d\alpha \\ &= \frac{1}{2\pi}\int_0^{2\pi} e^{\rho\cos(\alpha-\theta)}d\alpha \\ &= I_0(\rho) = I_0(\sqrt{a^2+b^2}). \end{aligned} \tag{A.0.1}$$

The *modified Bessel function* has the properties

$$I_{-\nu}(z) = I_\nu(z),$$

$$I_{\nu-1}(z) + I_{\nu+1}(z) = 2I'_\nu(z),$$

$$I_{\nu-1}(z) - I_{\nu+1}(z) = \frac{2\nu}{z} I_\nu(z),$$

$$\frac{d}{dz} I_0(z) = I'_0(z) = I_1(z),$$

$$\begin{aligned} I''_0(z) = I'_1(z) &= \frac{1}{2}\left(I_0(z) + I_2(z)\right) \\ &= I_2(z) + \frac{I_1(z)}{z} \\ &= I_0(z) - \frac{I_1(z)}{z}. \end{aligned}$$

From this, it can be checked

$$\frac{d}{d\kappa}(\kappa I_1(\kappa)) = \kappa I_0(\kappa)$$

and

$$\frac{d}{d\kappa}\left(\frac{I_1(\kappa)}{\kappa}\right) = \frac{I_2(\kappa)}{\kappa}.$$

The ratio

$$A(\kappa) \equiv \frac{I_1(\kappa)}{I_0(\kappa)}$$

is a monotone increasing function of κ with

$$A'(\kappa) = 1 - \frac{A(\kappa)}{\kappa} - A^2(\kappa).$$

The following results provide some useful approximations for the Bessel functions, for large and small values of κ. For κ large,

$$\begin{aligned} I_p(\kappa) &= \frac{e^\kappa}{\sqrt{2\pi\kappa}}\{1 - \frac{(4p^2-1)}{8\kappa} + \frac{(4p^2-1)(4p^2-9)}{2(8\kappa)^2} \\ &\quad - \frac{(4p^2-1)(4p^2-9)(4p^2-25)}{6(8\kappa)^3} + \ldots\} \end{aligned}$$

so that in particular

$$I_0(\kappa) \sim \frac{e^\kappa}{\sqrt{2\pi\kappa}} \text{ for } \kappa \text{ large.}$$

and

$$A(\kappa) \sim (1 - \tfrac{1}{2k} - \tfrac{1}{8\kappa^2} - \ldots).$$

On the other hand, for κ small,

$$I_0(\kappa) \sim 1 + \frac{\kappa^2}{4} + \frac{\kappa^4}{64} + \ldots$$

and

$$A(\kappa) \sim \frac{\kappa}{2}(1 - \frac{\kappa^2}{8} + \frac{\kappa^4}{48} - \ldots).$$

Theorem A.1 (Weber's discontinuous factor theorem) *(cf. Watson (1944), p. 406, Equation 9)*

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} J_0(t\sqrt{a^2 + b^2 - 2ab\cos\theta})d\theta = J_0(at)J_0(bt).$$

$$\int_0^\infty J_0(at)J_1(bt)dt = \begin{cases} 0 & \textit{if } a > b, \\ \frac{1}{2b} & \textit{if } a = b, \\ \frac{1}{b} & \textit{if } a < b. \end{cases}$$

Theorem A.2 (Naumann addition formula) *(cf. Watson (1944), Section 11.2)*

$$J_0(\sqrt{(a^2 + b^2 - 2ab\cos\theta)}) = J_0(a).J_0(b) + 2\sum_{k=0}^{\infty} J_k(a).J_k(b).\cos k\theta.$$

Appendix B

How to Use the CircStats Package

B.1 Installation Instructions for the CircStats Library

B.1.1 Windows Version

(i) Decide on a directory in which to store the S-Plus library, for example

```
C:\Splus\library\circular\
```

(ii) The file named "`circular.exe`" is a self-extracting archive of the files making up the library. A 32-bit version of Windows is required (`Windows` 95, 98 or `NT`) to unpack the archive successfully. To extract the files from the archive, simply double click on the filename, from within the file manager window. You will be asked where to extract the files to. Select the directory from (i).

(iii) The `CircStats` Library is now available for use. Once S-Plus has been started, to access the functions in the library, issue the following command from the S-Plus prompt

```
library(circular, lib.loc="C:\\Splus\\library\\circular\\")
```

Note the double back-slashes.

(iv) Help files on all functions in the library are also available using the standard help interface for S-Plus for Windows. Once the library has been accessed according to (iii), help on a particular function, say "`function.name`", can be obtained by the S-Plus command

```
?function.name
```

B.1.2 UNIX Version

Tested on S-PLUS 3.4, Release 1 for Sun SPARC, SunOS 5.3. Users of S-PLUS 5 or later, using the new style of help files will have to complete the following instructions and then convert the help files according to the documentation provided with their versions of S-PLUS.

(i) Decide on a directory in which to store the S-PLUS library, for example

```
/SHOME/library/
```

(ii) Create a directory "`circular`" in this directory, and move the compressed archive file "`circular.tar.gz`" into the directory

```
"/SHOME/library/circular".
```

(iii) Unzip the compressed archive "`circular.tar.gz`", using the command

```
gunzip circular.tar.gz
```

(iv) Extract the files from the resulting archive "`circular.tar`" by issuing the command

```
tar xvf circular.tar
```

You should now have a "`README`"file, a directory "`.Data`" with the S-PLUS functions, and a directory "`/.Data/.Help`" with the associated help files for the functions.

(v) So that S-PLUS can locate the help files via the usual help interface, from within the "`/SHOME/library/circular`" directory, at the `UNIX` prompt type

```
Splus help.findsum .Data
```

(vi) The CircStats Library is now available for use. Once S-PLUS has been started, to access the functions in the library, issue the following command from the S-Plus prompt

```
library(circular, lib.loc="/SHOME/library")
```

The lib.loc option can be omitted if you have placed the circular library in the S-PLUS default location for libraries.

B.2 CircStats Functions

Table B.1 on page 294 gives a list of all the routines available in CircStats library.

A1	Ratio of First and Zeroth Order Bessel Functions
A1inv	Inverse of A1
change.pt	Change Point Test
circ.cor	Correlation Coefficient for Angular Variables
circ.disp	Circular Dispersion
circ.mean	Mean Direction
circ.plot	Circular Data Plot
circ.range	Circular Range
circ.reg	Circular/Circular Regression
circ.summary	Circular Summary Statistics
dcard	Cardioid Density Function
deg	Degrees
dmixedvm	Mixture of von Mises Distributions
dtri	Triangular Density Function
dvm	von Mises Density Function
dwrpcauchy	Wrapped Cauchy Density Function
dwrpnorm	Wrapped Normal Density Function
est.kappa	Estimate Kappa
est.rho	Mean Resultant Length
I.0	Zeroth Order Bessel Function of the First Kind
I.1	First Order Bessel Function of the First Kind
I.p	p^{th} Order Bessel Function of the First Kind
kuiper	Kuiper's Test
plot.edf	Plot Empirical Distribution Function
pp.plot	von Mises Probability-Probability Plot
pvm	Cumulative Probability for the von Mises Distribution
r.test	Rayleigh Test of Uniformity:General Unimodal Alternatives
rad	Radians
rao.homogeneity	Rao's Tests for Homogeneity
rao.spacing	Rao's Spacing Test of Uniformity
rcard	Random Generation from the Cardioid Distribution
rmixedvm	Random Generation from the Mixed von Mises Distribution
rose.diag	Rose Diagram
rtri	Random Generation from the Triangular Distribution
rvm	Random Generation from the von Mises Distribution
rwrpcauchy	Random Generation from the Wrapped Cauchy Distribution
rwrpnorm	Random Generation from the Wrapped Normal Distribution
rwrpstab	Random Generation from the Wrapped Stable Distribution
trig.moment	Trigonometric Moment
v0.test	Test of Uniformity:Alternative with Specified Mean Direction
vm.bootstrap.ci	Bootstrap Confidence Intervals
vm.ml	von Mises Maximum Likelihood Estimates
watson	Watson's Test
watson.two	Watson's Two-Sample Test of Homogeneity
wrpcauchy.ml	Wrapped Cauchy Maximum Likelihood Estimates

Table B.1: Functions available in CircStats.

Bibliography

ABRAHAMSON, I. G. (1967), Exact Bahadur efficiencies for the Kolmogorov-Smirnov and Kuiper one- and two-sample statistics, *Ann. Math. Statist.*, **38**, 1475—1490.

ABRAMOWITZ, M. AND STEGUN, I. A. (1965), *Handbook of Mathematical Functions*, Dover, New York.

ACARDI, L., CABRERA, J., AND WATSON, G. S. (1987), Some stationary Markov processes in discrete time for unit vectors, *Metron*, **45**, 115—133.

AJNE, B. (1968), A simple test for uniformity of a circular distribution, *Biometrika*, **55**, 343—354.

AMARI, S. (1985), *Differential Geometrical Methods in Statistics*, Springer, New York.

ANDERSON, C. M. AND WU, C. F. J. (1995), Measuring location effects from factorial experiments with a directional response, *Internat. Statist. Rev.*, **63**, 345—363.

ANDERSON, C. M. AND WU, C. F. J. (1996), Dispersion measures and analysis for factorial directional data with replicates, *Appl. Statist.*, **45**, 47—61.

ANDERSON-COOK, C. M. (2000), An industrial example using one-way analysis of circular-linear data, *Computational Statist. Data Anal.*, **33**, 45—57.

APOSTOL, T. M. (1974), *Mathematical Analysis*, Addison Wesley.

ARNOLD, K. J. (1941), *On Spherical Probability Distributions*, PhD thesis, MIT.

ASH, R. B. (1972), *Real Analysis and Probability*, Academic Press, New York.

BAGCHI, P. (1987), *Bayesian Analysis of Directional Data*, PhD thesis, Univ. of Toronto.

BAGCHI, P. AND GUTTMAN, I. (1990), Spuriosity and outliers in directional data, *J. Appl. Statist.*, **17**, 341—350.

BAGCHI, P. AND KADANE, J. B. (1991), Laplace approximations to posterior moments and marginal distributions on circles, spheres and cylinders, *Canad. J. Statist.*, **19**, 67—77.

BAHADUR, R. R. (1960), Stochastic comparison of tests, *Ann. Math. Statist.*, **31**, 276—295.

BAI, Z. D., RAO, C. R., AND ZHAO, L. C. (1988), Kernel estimators of density function of directional data, *J. Multivariate Anal.*, **27**, 24—39.

BARNETT, V. AND LEWIS, T. (1994), *Outliers in Statistical Data*, Wiley, Chichester, third edition.

BARTELS, R. (1984), Estimation in a bidirectional mixture of von Mises distributions, *Biometrika*, **44**, 777—784.

BASU, S. AND JAMMALAMADAKA, S. R. (2000), Unimodality of circular data- A Bayes test, In Balakrishnan, N., editor, *Advances on Methodological and Applied Aspects of Probability and Statistics*, volume 1, pages 141—153. Gordon and Breach Publishers, Amsterdam.

BATSCHELET, E. (1981), *Circular Statistics in Biology*, Academic Press, London.

BENFORD, F. (1938), The law of anomalous numbers, *Proc. Amer. Phil. Soc.*, **78**, 551—572.

BERAN, R. (1979), Exponential models for directional data, *Ann. Statist.*, **7**, 1162—1178.

BERAN, R. J. (1969), Asymptotic theory of a class of tests for uniformity of a circular distribution, *Ann. Math. Statist.*, **40**, 1196—1206.

BERGER, J. (1985), *Statistical Decision Theory and Bayesian Analysis*, Springer Verlag.

BHATTACHARYYA, G. K. AND JOHNSON, R. A. (1969a), On Hodge's bivariate sign test for uniformity of a circular distribution, *Biometrika*, **56**, 446—449.

BHATTACHARYYA, G. K. AND JOHNSON, R. A. (1969b), On Hodge's bivariate sign test for uniformity of a circular distribution, *Biometrika*, **56**, 446—449.

BIJLEVELD, C. C. J. H. AND COMMANDEUR, J. J. F. (1987), The extension of analysis of angular variation to m-way designs, PRM-87-03.

BJORNSTAD, B. J. F. (1990), Predictive likelihood: A review, *Statist. Sci.*, **5**, 242—265.

BLUM, J. R., KIEFER, J., AND ROSENBLATT, M. (1961), Distribution free test of independence based on the sample distribution function, *Ann. Math. Statist.*, **32**, 485—492.

BRECKLING, J. (1989), *The Analysis of Directional Time Series: Applications to Wind Speed and Direction*, Springer Verlag, New York.

BROWN, L. D. (1986), *Fundamentals of Statistical Exponential Families*, volume 9 of *Inst. Math. Statist. Lecture Notes*, Inst. Math. Statist., California.

BUTLER, R. (1986), Predictive likelihood inference with applications (with discussion), *J. Roy. Statist. Soc.*, **48**, 1—38.

CARSLAW, H. S. (1930), *Introduction to the Theory of Fourier's Series and Integrals*, Dover, New York, third edition.

CARTWRIGHT, D. E. (1963), The use of directional spectra in studying the output of a wave recorder on a moving ship, In *Ocean Wave Spectra*, pages 203—218. Prentice Hall, New Jersey.

CHANG, H. (1991), *Some Optimal Tests in Directional Data*, PhD thesis, Univ. of California - Santa Barbara, U. S. A.

CHAPMAN, M. G. AND UNDERWOOD, A. J. (1992), Experimental designs for analyses of movements by molluscs, *J. Moll. Stud.*

CHERNOFF, H. (1951), A property of some Type A regions, *Ann. Math. Statist.*, **22**, 472—474.

COGBURN, R. AND DAVIS, H. T. (1974), Periodic splines and spectral estimation, *Ann. Statist.*, **2**, 1108—1126.

COLES, S. G. (1998), Inference for circular distributions and processes, *Statist. Comput.*, **8**, 105—113.

COLLETT, D. (1980), Outliers in circular data, *Appl. Stat.*, **29**, 50—57.

COLLETT, D. AND LEWIS, T. (1981), Discriminating between the von Mises and wrapped normal distributions, *Austral. J. Statist.*, **23**, 73—79.

CRAIG, P. S. (1988), *Time Series Analysis for Directional data*, PhD thesis, Trinity College, Dublin.

CSÖRGŐ, M. AND HORVÁTH, L. (1996), A note on the change-point problem for angular data, *Statist. Probab. Lett.*, **27**, 61—65.

DAMIEN, P. AND WALKER, S. (1999), A full Bayesian analysis of circular data using the von Mises distribution, *Canad. J. Statist.*, **27**, 291—298.

DANIELS, H. E. (1954), A distribution-free test for regression parameters, *Ann. Math. Statist.*, **25**, 499—513.

DAVID, F. N. AND BARTON, D. E. (1962), *Combinatorial Chance*, Griffin and Company, London.

DAVIES, R. B. (1977), Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika*, **64**, 247—254.

DIXON, W. J. (1940), A criterion for testing the hypothesis that two samples are from the same population, *Ann. Math. Statist.*, **11**, 199—204.

DUCHARME, G. R. AND MILASEVIC, P. (1987), Some asymptotic properties of the circular median, *Comm. Statst.- Theor. Meth.*, **16**, 659—664.

DUDEWICZ, E. J. AND VAN DER MEULEN, E. C. (1981), Entropy based tests of uniformity, *J. Amer. Statist. Assoc.*, **76**, 967—974.

EFRON, B. (1975), Defining the curvature of a statistical problem, *Ann. Statist.*, **3**, 1189—1242.

ENGEBRETSON, D. C. AND BECK, M. E. (1978), On the shape of directional data sets, *J. Geophys. Res.*, **83**, 5979—5982.

EUBANK, R. L. (1988), *Spline Smoothing and Nonparametric Regression*, Marcel Dekker, New York.

FEDER, P. I. (1968), On the distribution of the log likelihood ratio statistic when the true parameter is near the boundaries of the hypothesis region, *Ann. Math. Statist.*, **39**, 44—55.

FELLER, W. (1968), *An Introduction to Probability Theory and Its Applications, Vol.I*, Wiley, New York, 3 edition.

FERGUSON, T. S. (1967), *Mathematical Statistics: A Decision Theoretic Approach*, Academic Press, New York.

FISHER, N. I. (1993), *Statistical Analysis of Circular Data*, Cambridge University Press, Cambridge.

FISHER, N. I. AND HALL, P. G. (1989), Bootstrap confidence regions for directional data, *J. Amer. Statist. Assoc.*, **84**, 996—1002.

FISHER, N. I. AND LEE, A. J. (1983), A correlation coefficient for circular data, *Biometrika*, **70**, 327—332.

FISHER, N. I. AND LEE, A. J. (1992), Regression models for an angular response, *Biometrics*, **48**, 665—677.

FISHER, N. I. AND LEE, A. J. (1994), Time series analysis of circular data, *J. Roy. Statist. Soc. B*, **56**, 327—339.

FISHER, N. I. AND LEWIS, T. (1983), Estimating the common mean direction of several circular of spherical distributions with different dispersions, *Biometrika*, **70**, 333—341.

FISHER, N. I., LEWIS, T., AND EMBLETON, B. J. J. (1987), *Statistical Analysis of Spherical Data*, Cambridge University Press, Cambridge.

FISHER, R. A. (1929), Tests of significance in harmonic analysis, *Proc. Roy. Soc. London, Ser A*, **125**, 54—59.

FISHER, R. A. (1953), Dispersion on a sphere, *Proc. Roy. Soc. London, Ser A*, **217**, 295—305.

FISHER, R. A. (1959), *Statistical Methods and Scientific Inference*, Oliver and Boyd, Edinburgh, second edition.

FULLER, M., LAJ, C., AND HERRERO-BERVERA, E. (1996), The reversal of the earth's magnetic field, *Amer. Sci.*, **84**, 552—561.

GADSDEN, R. J. AND KANJI, G. K. (1981), Sequential analysis for angular data, *Statistician*, **30**, 119—129.

GADSDEN, R. J. AND KANJI, G. K. (1982), Sequential analysis applied to circular data, *Seq. Anal.*, **1**, 305—314.

GATTO, R. AND JAMMALAMADAKA, S. R. (2000), Inference for wrapped symmetric α-stable circular models, to appear.

GHOSH, J. K. AND SEN, P. K. (1985), On the asymptotic performance of the log likelihood ratio statistic for the mixture model, In LeCam, L. M., editor, *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, volume II, pages 789—806. Wadsworth and Institute of Mathematical Statistics, California.

GHOSH, K. (2000), Detecting changes in the von Mises distribution, In Balakrishnan, N., editor, *Advances on Methodology and Applications of Probability and Statistics*. Gordon and Breach Publishers, Amsterdam.

GHOSH, K., JAMMALAMADAKA, S. R., AND VASUDEVAN, M. (1999), Change point problems for von Mises distribution, *J. Appl. Statist.*, **26**, 423—434.

GLAZ, J. AND BALAKRISHNAN, N. (1999), *Scan Statistics and Applications*, Birkhäuser.

GOULD, A. L. (1969), A regression technique for angular data, *Biometrics*, **25**, 683—700.

GREENWOOD, J. A. AND DURAND, D. (1955), The distribution of length and components of the sum of n random Unit vectors, *Ann. Math. Statist.*, **26**, 233—246.

GUMBEL, E. J., GREENWOOD, J. A., AND DURAND, D. (1953), The Circular normal distribution: theory and tables, *J. Amer. Statist. Assoc.*, **48**(261), 131—152.

GUTTORP, P. AND LOCKHART, R. A. (1988), Finding the location of a signal : a Bayesian analysis, *J. Amer. Statist. Assoc.*, **83**, 322—330.

HALL, P., S WATSON, G., AND CABRERA, J. (1987), Kernel density estimation with spherical data, *Biometrika*, **74**, 751—762.

HARRISON, D. AND KANJI, G. K. (1988), The development of analysis of variance for circular data, *J. Appl. Statist.*, **15**, 197—224.

HARRISON, D., KANJI, G. K., AND GADSDEN, R. J. (1986), Analysis of variance of circular data, *J. Appl. Statist.*, **13**, 197—223.

HILL, T. P. (1995), A statistical derivation of the significant-digit law, *Statist. Sci.*, **10**, 354—363.

HODGES, J. L. (1955), A bivariate sign test, *Ann. Math. Statist.*, **26**, 523—527.

HOLST, L. AND JAMMALAMADAKA, S. R. (1980), Asymptotic theory for families of two-sample nonparametric statistics, *Sankhyá, Ser.A*, **42**, 19—52.

HORVATH, L. (1993), The maximum likelihood method for testing changes in the parameters of normal observations, *Ann. Statist.*, **21**(2), 671—680.

Hrushesky, W. J. M., editor (1994), *Circardian Cancer Therapy,* CRC Press, Boca Raton.

JAMMALAMADAKA, S. R. (1984), Nonparametric methods in directional data analysis, In Krishnaiah, P. R. and Sen, P. K., editors, *Handbook of Statistics, Nonparametric Methods,* volume IV, pages 755—770. North Holland, Amsterdam.

JAMMALAMADAKA, S. R., BHADRA, N., CHATURVEDI, D., KUTTY, T. K., MAJUMDAR, P. P., AND PODUVAL, G. (1986), Functional assessment of knee and ankle during level walking, In Krishnan, T., editor, *Data Analysis in Life Science,* pages 21—54. Indian Statistical Institute, Calcutta, India.

JAMMALAMADAKA, S. R. AND SARMA, Y. R. (1988), A correlation coefficient for angular variables, In Matusita, K., editor, *Statistical Theory and Data Analysis II,* pages 349—364. North Holland, Amsterdam.

JAMMALAMADAKA, S. R. AND SARMA, Y. R. (1993), Circular Regression, In Matusita, K., editor, *Statistical Sciences and Data Analysis,* pages 109—128. VSP, Utrecht.

JAMMALAMADAKA, S. R. AND SENGUPTA, A. (1998), Predictive inference for directional data, *Statist. Prob. Lett.*, **40**, 247—257.

JANDER, C. (1957), Die optiche Richtungsorientierung der roten Waldameise (*Formica rufa* L.), *Z. vergl. Physiologie*, **40**, 162—238.

JEFFREYS, H. (1961), *Theory of Probability,* Oxford University Press, Oxford, 3rd edition.

JOHNSON, R. A. AND WEHRLY, T. E. (1977), Measures and models for angular correlation and angular-linear correlation, *J. Roy. Statist. Soc.*, **39**, 222—229.

JOHNSON, R. A. AND WEHRLY, T. E. (1978), Some angular-linear distributions and related regression models, *J. Amer. Statist. Assoc.*, **73**, 602—606.

KAGAN, A. M., LINNIK, Y. V., AND RAO, C. R. (1973), *Characterization Problems in Mathematical Statistics*, John Wiley, New York.

KARIYA, T. (1989), Equivariant estimation in a model with an ancillary statistic, *Ann. Statist.*, **17**, 920—928.

KENDALL, D. G. (1974), Pole-seeking Brownian Motion and bird navigation (with Discussion), *J. Roy. Statist. Soc.*, **36**, 365—417.

KENDALL, M. G. AND STUART, A. (1979), *The Advanced Theory of Statistics*, volume II, MacMillan, New York, 4 edition.

KENT, J. T., MARDIA, K. V., AND RAO, J. S. (1979), A characterization of uniform distribution on the circle, *Ann. Statist.*, **7**, 197—209.

KENT, J. T. AND TYLER, D. E. (1988), Maximum likelihood estimation for the wrapped Cauchy distribution, *J. Appl. Statist.*, **15**, 247—254.

KLEMELä, J. (2000), Estimation of densities and derivatives of densities with directional data, *J. Multivariate Anal.*, **73**, 18—40.

KLUYVER, J. C. (1906), A local probability theorem, *Ned. Akad. Wet. Proc., Ser. A*, **8**, 341—350.

KUIPER, N. H. (1960), Tests concerning random points on a circle, *Ned. Akad. Wet. Proc.*, **63**, 38—47.

LAI, M. K. (1994), *Some Results on the Statistical Analysis of Directional data*, Master's thesis, HongKong Univ.

LANGEVIN, P. (1905), Magnetisme et theorie des electrons, *Ann. Chim. Phys.*, **5**, 71—127.

LEE, P. M. (1989), *Bayesian Statistics: An Introduction*, Oxford University Press, New York.

LEHMANN, E. L. (1986), *Testing Statistical Hypotheses*, Wiley, New York, 2 edition.

LENTH, R. V. (1981), On finding the source of a signal, *Technometrics*, **23**, 149—154.

LÉVY, P. (1939), L'addition des variables aléatoires définies sur une circonférence, *Bull. Soc. Math. France.*, **67**, 1—41.

LIDDELL, I. G. AND ORD, J. K. (1978), Linear-circular correlation coefficients: some further results, *Biometrika*, **65**, 448—450.

LOCKHART, R. A. AND STEPHENS, M. A. (1985), Tests of fit for the von Mises distribution, *Biometrika*, **72**, 647—652.

LOMBARD, F. (1986), The change-point problem for angular data: a nonparametric approach, *Technometrics*, **28**, 391—397.

LUKACS, E. (1970), *Characteristic Functions*, Griffin, second edition.

LUND, U. (1999), Cluster analysis for directional data, *Communications in Statistics - Simulation and Computation*, **28**, 1001—1009.

LUND, U. AND JAMMALAMADAKA, S. R. (2000), An entropy based test for goodness-of-fit of the von Mises distribution, *J. Statist. Comput. Simulation*, **67**, 319—332.

MARDIA, K. V. (1970), A bivariate nonparametric c-sample test, *J. Roy. Statist. Soc., B*, **32**, 74—87.

MARDIA, K. V. (1972), *Statistics of Directional Data*, Academic Press, New York.

MARDIA, K. V. (1975), Statistics of directional data (with discussion), *J. Roy. Statist. Soc.*, **37**, 349—393.

MARDIA, K. V. (1976), Linear-circular correlation coefficients and rhythmometry, *Biometrika*, **63**, 403—405.

MARDIA, K. V. AND DRYDEN, I. L. (1993), Multivariate shape analysis, *Sankhyá, Ser. A*, **55**, 460—480.

MARDIA, K. V. AND DRYDEN, I. L. (1998), *Statistical Shape Analysis*, John Wiley and Sons, Chichester.

MARDIA, K. V. AND EL-ATOUM, S. A. M. (1976), Bayesian inference for the von Mises-Fisher distribution, *Biometrika*, **63**, 203—205.

MARDIA, K. V. AND JUPP, P. E. (1980), A general correlation coefficient for directional data and related regression problems, *Biometrika*, **67**, 163—173.

MARDIA, K. V. AND JUPP, P. E. (2000), *Directional Statistics*, John Wiley, Chichester.

MARDIA, K. V. AND PURI, M. L. (1978), A robust spherical correlation coefficient against scale, *Biometrika*, **65**, 391—396.

MARDIA, K. V. AND SPURR, B. D. (1973), Multisample tests for multimodal and axial circular populations, *J. Roy. Statist. Soc., B*, **35**, 422—436.

MARDIA, K. V. AND SUTTON, T. W. (1975), On the modes of a mixture of two von Mises distributions, *Biometrika*, **62**, 699—701.

MARDIA, K. V. AND SUTTON, T. W. (1978), A model for cylindrical variables with applications, *J. Roy. Statist. Soc., Ser. B*, **40**, 229—233.

MCCULLAGH, P. (1989), Some statistical properties of a family of continuous univariate distributions, *J. Amer. Statist. Assoc.*, **84**, 125—129.

MICHAEL, J. R., SCHUCANY, W., AND HAAS, R. W. (1976), Generating random variates using transformations with multiple roots, *American Statistician*, **30**, 88—90.

Morgan, E., editor (1990), *Chronobiology and Chronomedicine*, Peter Lang, Frankfurt.

MORRIS, J. E. AND LAYCOCK, P. J. (1974), Discrimination analysis of directional data, *Biometrika*, **61**, 335—341.

NELDER, J. A. AND MEAD, R. (1965), A simplex method for function minimization, *Computer J.*, **5**, 308—313.

NEYMAN, J. (1959), Optimal asymptotic tests of composite statistical hypotheses, In Grenander, U., editor, *Probability and Statistics*, pages 213—234. Wiley, New York.

NEYMAN, J. AND PEARSON, E. S. (1936), Contributions to the theory of testing statistical hypotheses, *Statist. Res. Mem.*, **1**, 1—37.

PAGE, E. S. (1955), A test for a change in a parameter occurring at an unknown point, *Biometrika*, **43**, 523—526.

PAULA, G. M. C. G. A. AND BOTTER, D. A. (1994), Improved likelihood ratio tests for dispersion model, *Int. Statist. Rev.*, **62**, 257—274.

PEARSON, K. (1905), The problem of the random walk, *Nature*, **72**, 294—342.

PEARSON, K. (1906), A mathematical theory of random migration, *Drapers' Company Research Memoirs*, Biometric Series #3.

POINCARÉ, H. (1912), Chance, *Monist*, **22**, 31—52.

PYKE, R. (1965), Spacings, *J. Roy. Statist. Soc., Ser. B*, **27**, 395—449.

RAIMI, R. A. (1969), The first digit problem, *Amer. Math. Monthly*, **76**, 521—538.

RAO, C. R. (1973), *Linear Statistical Inference and its Applications*, John Wiley & Sons, New York, 2nd edition.

RAO, J. S. (1967), Large sample tests for the homogeneity of angular data, *Sankhyá*, **28**, 172—174.

RAO, J. S. (1969), *Some Contributions to the Analysis of Circular Data*, PhD thesis, Indian Statistical Institute, Calcutta, India.

RAO, J. S. (1972a), Bahadur efficiencies of some tests for uniformity on the circle, *Ann. Math. Statist.*, **43**, 468—479.

RAO, J. S. (1972b), Some variants of chi-square for testing uniformity on the circle, *Zeitschrift fur Wahrscheinlichkeitstheorie and vervandt Gebiete*, **22**, 33—44.

RAO, J. S. (1976), Some tests based on arc-lengths for the circle, *Sankhyá*, **38**, 329—338.

RAO, J. S. AND MARDIA, K. V. (1980), Pitman efficiencies of some two-sample nonparametric tests, In Matsusita, K., editor, *Recent Developments in Statistical Inference and Data Analysis*, pages 247—254. North Holland, Amsterdam.

RAO, J. S. AND PURI, M. L. (1977), Problems of association for bivariate circular data and a new test of independence, In Krishnaiah, P. R., editor, *Multivariate Analysis*, volume IV, pages 513—522. North Holland, Amsterdam.

RAO, J. S. AND SENGUPTA, S. (1970), An optimum hierarchical sampling procedure for crossbedding data, *J. Geol.*, **78**(5), 533—544.

RAO, J. S. AND SENGUPTA, S. (1972), Mathematical techniques for paleocurrent analysis: Treatment of directional data, *Journal of int'l Assoc. Math. Geol*, **4**, 235—258.

RAO, J. S. AND SETHURAMAN, J. (1975), Weak convergence of empirical distribution functions of random variables subject to perturbations and scale factors, *Ann. Statist.*, pages 299—313.

RICE, J. AND ROSENBLATT, M. (1981), Integrated mean squared error of a smoothing spline., *J. Approx. Theory*, **33**, 353—369.

RIVEST, L. P. (1982), Some statistical methods for bivariate circular data, *J. Roy. Statist. Soc.*, **44**, 81—90.

ROHDE, C. A. (1965), Generalized inverses of partitioned matrices, *SIAM J. of Appl. Math*, **13**, 1033—1035.

ROTHMAN, E. D. (1971), Tests for coordinate independence for a bivariate sample on torus, *Ann. Math. Statist.*, **42**, 1962—1969.

ROTHMAN, E. D. (1972), Tests for uniformity of a circular distribution, *Sankhyá*, **34**, 23—32.

ROY, S. (1999), *On logistic and some new discrimination rules: characterizations, inference and applications*, PhD thesis, Indian Statistical Institute, Calcutta, India.

RUKHIN, A. L. (1972), Some statistical decisions about distributions on a circle for large samples, *Sankhyā Ser. A*, **34**, 243–250.

RUSSELL, G. S. AND LEVITIN, D. J. (1995), An expanded table of probability values for Rao's spacings test, *Communications in Statistics: Simulation and computation*, **24**, 879—888.

SAW, J. G. (1978), A family of distributions on the m-sphere and some hypothesis tests, *Biometrika*, **65**, 69—74.

SCHMIDT, P. (1976), *Econometrics*, Marcel Dekker, New York.

SCHMIDT-KOENIG, K. (1958), Experimentelle einflussnahme auf die 24-Stunden-Periodik bei brieftauben und deren auswirkungen unter besonderer Berūcksichtigung des heimfindevermōgens, *Z. Tierpsychol.*, **15**, 301—331.

SCHMIDT-KOENIG, K. (1963), On the role of loft, the distance and site of release in pigeon homing(the "cross-loft experiment"), *Biol. Bull.*, **125**, 154—164.

SCHMITT, S. A. (1969), *Measuring Uncertainty: An Elementary Introduction to Bayesian Statistics*, Addison Wesley, Chichester.

SEBER, G. A. F. (1977), *Linear Regression Analysis*, John Wiley, New York.

SEN, A. K. AND SRIVASTAVA, M. S. (1975), On tests for detecting a change of mean, *Ann. Statist.*, **3**, 98—108.

SENGUPTA, A. (1991), A review of optimality of multivariate tests, *Statist. Prob. Lett.*, **12**, 527—535.

SENGUPTA, A. AND LAHA, A. K. (1999), Change-point problems for the circular normal distribution., Submitted for publication.

SENGUPTA, A. AND LAHA, A. K. (2000), The slippage problem for circular normal distribution, *Austr. J. Statist.*, to appear.

SENGUPTA, A. AND MAITRA, R. (1998), On best equivariance and admissibility of simultaneous MLE for mean direction vectors of several Langevin distributions, *Ann. Inst. Statist. Math.*, **50**, 715—727.

SENGUPTA, A. AND PAL, C. (2001a), Optimal tests for no contamination in directional data, *J. Appl. Statist.*, **28**, 129—143.

SENGUPTA, A. AND PAL, C. (2001b), Optimal tests for no contamination in wrapped stable mixture family for directional data, *J. Appl. Statist.*, **28**, 129—143.

SENGUPTA, S. AND RAO, J. S. (1967), Statistical analysis of crossbedding azimuths from the Kamthi formation around Bheemaram, Pranhita Godavari Valley, *Sankhyá*, **28**, 165—174.

SETHURAMAN, J. AND RAO, J. S. (1970), Pitman efficiencies of tests based on spacings, In *Nonparametric Techniques in Statistical Inference*, pages 405—415. Cambridge University Press.

SHEPHERD, J. AND FISHER, N. I. (1982), Rapid method of mapping fracture trends in collieries, *Trans. Soc. Min. Eng. AIME*, **170**, 1931—1932.

SMALL, C. G. (1996), *The Statistical Theory of Shape*, Springer, New York.

SPURR, B. D. (1981), On Estimating the Parameters in Mixtures of Circular Normal Distributions., *Math. Geology*, **13**, 163 — 173.

STEPHENS, M. A. (1962), Exact and approximate tests for directions I, *Biometrika*, **49**, 463—477.

STEPHENS, M. A. (1963), Random walk on a circle, *Biometrika*, **50**, 385—390.

STEPHENS, M. A. (1969a), A goodness-of-fit statistic for the circle, with some comparisons, *Biometrika*, **56**, 161—168.

STEPHENS, M. A. (1969b), Techniques for directional data, Technical Report # 150.

STEPHENS, M. A. (1979), Vector correlation, *Biometrika*, **66**, 41—48.

STEPHENS, M. A. (1982), Use of the von Mises distribution to analyze continuous proportions, *Biometrika*, **69**, 197—203.

TITTERINGTON, D. M., SMITH, A. F. M., AND MAKOV, U. E. (1985), *Statistical Analysis of Finite Mixture Distributions*, Wiley, Chichester.

UNDERWOOD, A. J. AND CHAPMAN, M. G. (1985), Multifactorial analyses of directions of movements of animals, *J. Exp. Mar. Biol. Ecol.*, **91**, 17—43.

UPTON, G. J. G. (1976), More multisample tests for the von Mises distribution, *J. Amer. Statist. Assoc.*, **71**, 675—678.

UPTON, G. J. G. (1993), Outliers in circular data, *J. Appl. Statist.*, **20**(2), 229—235.

UPTON, G. J. G. AND FINGLETON, B. (1989), Spatial data analysis by example, In *Categorical and Directional Data*, volume 2. J. Wiley, New York.

VASICEK, O. (1976), A test of normality based on sample entropy, *J. Roy. Statist. Soc., Ser. B*, **38**, 54—59.

VON MISES, R. (1918), Über die "ganzzahligkeit" der atomgewichte und verwandte fragen, *Physikal. Z.*, **19**, 490—500.

WAHBA, G. (1975), Smoothing noisy data with spline functions, *Numer. Math.*, **24**, 383—393.

WAHBA, G. AND WENDELBERGER, J. G. (1980), Some new mathematical methods for variational objective analysis using splines and cross validation., *Monthly Weather Rev.*, **108**, 1122—1143.

WAHBA, G. AND WOLD, S. (1975), Periodic splines for spectral density estimstion: the use of cross validation for determining the degree of smoothing., *Commun. Statist.*, **4**, 125—141.

WATSON, G. N. (1944), *A Treatise on the Theory of Bessel Functions*, Cambridge University Press, Cambridge, second edition.

WATSON, G. S. (1961), Goodness-of-fit tests on the circle, *Biometrika*, **48**, 109—114.

WATSON, G. S. (1967), Another test for the uniformity of a circular distribution, *Biometrika*, **54**, 675—677.

WATSON, G. S. (1982), Directional data analysis, In *Encyclopedia of Statistical Sciences*, volume 2, pages 376—381. New York.

WATSON, G. S. (1983a), Large sample theory of the Langevin distribution, *J. Statist. Plan. Inference*, **8**, 245—256.

WATSON, G. S. (1983b), *Statistics on Spheres*, John Wiley, New York.

WATSON, G. S. AND BERAN, R. J. (1967), Testing a sequence of unit vectors for serial correlation, *J. Geophys. Res.*, **72**, 5655—5659.

WATSON, G. S. AND WILLIAMS, E. (1956), On the construction of significance tests on the circle and sphere, *Biometrika*, **43**, 344—352.

WEHRLY, T. E. AND JOHNSON, R. A. (1979), Bivariate models for dependence of angular observations and related Markov processes, *Biometrika*, **66**, 255—256.

WHEELER, S. AND WATSON, G. S. (1964), A distribution-free two-sample test on a circle, *Biometrika*, **51**, 256—257.

WHITTAKER, E. T. AND WATSON, G. N. (1944), *A Course in Modern Analysis*, Cambridge University Press, Cambridge.

WINTNER, A. (1933), On the stable distribution laws, *Amer. J. Math.*, **55**, 335—339.

WINTNER, A. (1947), On the shape of the angular case of Cauchy's distribution curves, *Ann. Math. Statist.*, **18**, 589—593.

WU, H. (1997), Optimal exact designs on a circle or a circular arc, *Ann. Statist.*, **25**, 2027—2043.

YOSHIMURA, I. (1978), On a test of homogeneity hypothesis for directional data, *Sankhyá*, **40**, 310—312.

ZHONG, J. (1991), *Some Contributions to the Spherical Regression Model*, PhD thesis, Univ. of Kentucky, Lexington, USA.

ZOLOTAREV, V. M. (1961), Concerning a certain probability problem, *Theory of Probability and Applications*, pages 201—204.

This page is intentionally left blank

Author Index

Subject Index

This page is intentionally left blank

Notations and Abbreviations Index

Series on Multivariate Analysis

TOPICS IN CIRCULAR STATISTICS

by **S Rao Jammalamadaka** *(University of California, Santa Barbara)* &
A SenGupta *(Indian Statistical Institute, Calcutta)*

This research monograph on circular data analysis covers some recent advances in the field, besides providing a brief introduction to, and a review of, existing methods and models. The primary focus is on recent research into topics such as change-point problems, predictive distributions, circular correlation and regression, etc. An important feature of this work is the S-plus subroutines provided for analyzing actual data sets. Coupled with the discussion of new theoretical research, the book should benefit both the researcher and the practitioner.

ISBN 981-02-3778-2

9 789810 237783

www.worldscientific.com

4031 hc