Semi-parametric and non-parametric methods for directional data

José Antonio Carnicero Carreño

May 12, 2011

Acknowledgements

It is a pleasure to thank the many people who made this thesis possible.

Foremost, it is difficult to overstate my gratitude to my Ph.D. supervisors, Dr. Maria Concepción Ausín Olivera and Dr. Michael Peter Wiper, for giving me the confidence to explore my research interests and the guidance to avoid getting lost in my explorations. With their enthusiasm, their inspiration, and their great efforts to explain things clearly and simply, the helped to make statistics fun for me. Throughout my thesis-writing period, we have solved several problems and they threw enough research questions my way to keep me busy for the rest of my life.

I would also like to thank the many people who have taught me statistics and mathematics: my high school math teacher (Ana Sanz), my undergraduate teachers at UC3M, especially Francisco Marcellán who give me my first class, Javier Prieto, Juan Romo, Daniel Peña and Regina Kaiser, the teaching assistants at UC3M, especially Sara López, Andrés Alonso and Pedro Galeano, and my graduate teachers, especially Froilán Martinez Dopico, Arturo de Pablo, Esther Ruiz and Rosario Romera. For their kind assistance, giving wise advise, helping with various applications, and so on.

I wish to thank to my many student colleagues providing a stimulating and fun environment in which to learn and grow. I am especially grateful to Alejandro Rodríguez, Santiago Pellegrini, André Santos, Ester Gonzalez, Alba Franco and Ana Laura Badagian.

I am also grateful to the secretaries in the statistics department, Gema García, Francisco García-Saavedra and Susana Linares for helping me in many different ways.

On a personal level, I wish to thank to my friends, Rubén Rodrigo, Ignacio Mahillo and Andrés Martínez, for their camaraderie, entertainment and emotional support.

I wish to thank my entire family for providing a loving environment for me. My brothers, my sister in law, some uncles and cousins (especially Anne-Marie Rodríguez) were particularly supportive.

Lastly, and most importantly, I wish to thank my parents, Josefa Victoria Carreño and

Vicente Carnicero. They bore me, raised me, supported me, taught me and loved me. To them I dedicate this thesis.

Resumen

... los datos direccionales proporcionan una amplia variedad de problemas abiertos y constituyen un área inagotable para desarrollar nuevos métodos estadísticos y herramientas de inferencia.

Jammalamadaka y otros (2001).

La mayoría de los trabajos en estadśtica se centran en datos lineales. Sin embargo, en muchos casos, se observan datos en forma de direcciones o tiempos de ocurrencia. Este típo de datos reciben el nombre de datos direccionales. Los datos direccionales habituales se pueden clasificar en datos circulares, que se pueden representar como puntos en el circulo unidad, datos toroidales, cilíndricos o esféricos. Ejemplos típicos de datos circulares son las rutas de migración de aves, estudiadas en biología (véase p.e. Batschelet, 1981), o la hora del día en la que llegan los pacientes a la sala de urgencias de un hospital. Ejemplos de datos toroidales, o circulares-circulares, se pueden encontrar en la estructura de los ángulos de los aminoácidos y ejemplos de datos cilíndricos, o circulares-lineales, en la dirección y fuerza del viento. Por último, un ejemplo de datos esféricos se tiene al medir la posición de los observatorios en los que se ha registrado más de cierta cantidad de lluvia durante el último año.

Los datos direccionales poseen algunas características intrínsecas que requieren técnicas de análisis especiales. Por ejemplo, en el caso circular, es evidente que los ángulos 10^o y 350^o están muy próximos, pero la media lineal de estos ángulos es 180^o, cuando en realidad es 0^o. Debidas a éstas y otras características, principalmente topológicas, una gran mayoría de las técnicas habituales que se aplican a los datos en la recta real resultan inaplicables a este tipo de datos y, por tanto, se han desarrollado numerosos modelos y técnicas especificas en la literatura para este tipo de datos.

La mayoría de los trabajos sobre datos direccionales en estadística se han centrado en modelos paramétricos. Sin embargo, casi todos estos modelos son unimodales y/o simétricos,

de modo que en muchas situaciones reales los datos no se pueden ajustar adecuadamente usando estos modelos. Por tanto, resulta muy interesante el desarrollo de otros procedimientos alternativos más flexibles y con menos restricciones paramétricas. Hasta ahora, la literatura existente sobre modelos no paramétricos o semiparamétricos es muy escasa reduciéndose, básicamente, a procedimientos basados en métodos kernel en el caso no paramétrico y basados en mixturas de modelos paramétricos en el caso semiparamétrico. El principal objetivo de esta Tesis es por tanto el desarrollo de nuevos modelos no paramétricos y semiparamétricos que sean apropiados para datos direccionales. Esta Tesis esta organizada del modo siguiente.

En primer lugar, en el Capítulo 1, se revisan las principales características de los datos direccionales y los principales modelos probabilísticos usados para estos datos. En primer lugar se considera el caso más sencillo de la modelización circular y luego se estudian los modelos circulares-circulares, circulares-lineales y esféricos. En cada caso, se muestran las principales características de estos datos y algunos de los modelos habituales. También, se presentan los métodos de inferencia habituales para el caso univariante y bivariante incluyendo los procedimientos no paramétricos.

En el Capítulo 2, se presenta la primera clase de modelos propuesta, que son distribuciones circulares basadas en polinomios de Bernstein circulares. Esta propuesta se basa en extender el procedimiento de aproximación de densidades basado en polinomios de Bernstein de Vitale (1975) al caso de datos circulares mostrando que el modelo propuesto conserva las buenas propiedades del modelo original. En este Capítulo, se introducen en primer lugar los polinomios de Bernstein que se usan habitualmente para la interpolación de funciones definidas sobre un intervalo cerrado y se muestran las tasas de convergencia de dicha interpolación. Posteriormente, se muestra cómo pueden aproximarse las funciones de distribución y de densidad de una variable lineal usando los polinomios de Bernstein.

A continuación se describe cómo adaptar el polinomio de Bernstein definido en un intervalo cerrado de la recta real al círculo y se muestra cómo, bajo ciertas condiciones, la función de densidad obtenida satisface las condiciones de densidad circular introducidas en el Capítulo 1. También se obtienen los momentos trigonométricos en una forma cerrada, y mediante un ejemplo basado en la distribución coseno, se ilustra cómo la densidad circular obtenida mediante polinomios Bernstein converge a la verdadera densidad y la convergencia del origen.

La principal contribución de este Capítulo es el desarrollo de un procedimiento de estimación no paramétrico para datos circulares basado en el desarrollado por Vitale (1975). Para ello, se propone realizar unas ciertas correcciones que permitan conseguir que la den-

sidad circular esté bien definida. Además, se muestra que a pesar de estas correcciones, la tasa de convergencia que es igual a la obtenida en Vitale (1975) para datos lineales y que coincide además con la tasa de convergencia que se obtiene mediante métodos Kernel. El procedimiento de estimación propuesto se ilustra con un conjunto de datos correspondientes a los delitos cometidos en Chicago el día 11 de mayo de 2007.

Por último, en este Capítulo, se exponen algunas conclusiones y posibles extensiones del modelo. Estas últimas consisten en usar distribuciones beta definidas de una forma similar en el círculo que, bajo ciertas restricciones, son mixturas de densidades circulares. Este modelo constituiría una alternativa semiparamétrica al modelo no paramétrico propuesto.

En el Capítulo 3, se construyen dos modelos que generalizan el polinomio de Bernstein circular al caso bivariante mediante el uso de cópulas Bernstein: un modelo circular-circular y un modelo circular-lineal.

En primer lugar, se introduce la generalización de los polinomios de Bernstein al caso bivariante de una forma similar a la expuesta en el Capítulo 2 y se revisan la definición, conceptos básicos y principales tipos de cópulas.

Posteriormente, se muestra mediante un sencillo ejemplo que no es posible adaptar el polinomio de Bernstein bivariante al caso circular-circular ya que no existe un origen que satisfaga las condiciones de continuidad en dos dimensiones. De este modo, se motiva el uso de cópulas como herramienta alternativa para poder estimar densidades bivariantes. Como es bien sabido, la principal ventaja de las cópulas es que permiten separar la estructura de dependencia de las distribuciones marginales, lo cual hace posible el uso de los métodos de estimación de distribuciones circulares univariantes desarrollados en el Capítulo 2.

Siguiendo la misma línea de estimación de modelos, la principal aportación de este Capítulo es el desarrollo de un procedimiento de estimación no paramétrico basado en cópulas Bernstein empíricas (véase Sancetta y otros, 2004), que constituyen estimadores no paramétricos de la estructura de dependencia entre variables continuas. Al igual que en el caso univariante, es necesario hacer ciertas correcciones para obtener distribuciones continuas. Sin embargo, se muestra que dichas correcciones preservan la uniformidad asintótica de las distribuciones marginales, dando lugar a modelos bien definidos.

A continuación se ilustran los procedimientos de estimación propuestos usando dos conjuntos de datos reales de carácter medioambiental. Uno sobre las direcciones del viento registradas por dos boyas de la NOAA situadas en el océano Atlántico y otro sobre las direcciones del viento y la cantidad de lluvia registrada en el observatorio de Somió en Gijón.

Se finaliza el Capítulo mostrando algunas conclusiones y extensiones de los modelos.

En el Capítulo 4, se construyen dos nuevos modelos bivariantes, uno para datos circulares-circulares y otro para datos esféricos haciendo uso de otra clase de polinomios, los polinomios trigonométricos. Tales polinomios están basados en sumas de términos sinusoidales y el principal problema cuando se construyen funciones de densidad basados en polinomios de este tipo es garantizar que sean no negativas. En el caso univariante, el teorema de Fejér-Riesz (véase Fejer, 1915) proporciona las condiciones necesarias para ello y este fue el resultado usado por Fernández-Durán (2004) para proponer un método de aproximación de densidades circulares univariantes. En el Capítulo 4, se usan los resultados de Geronimo y otros (2004) para generalizar este modelo al caso de datos circulares-circulares y esféricos. En particular, para la extensión de datos esféricos, se sigue una idea original basada en Merilees (1973) y Boer y otros (1975) para realizar una transformación (no biyectiva) de la esfera en el toro. Los procedimientos para datos circulares-circulares y esféricos se ilustran con dos conjuntos de datos reales. El capítulo acaba con algunas conclusiones y extensiones de los modelos propuestos.

Preface

... directional statistics provides an inquisitive reader with many open research problems and is a fertile area for developing new statistical methods and inferential tools.
Jammalamadaka et al (2001).

Most work in statistics is centred on linear data. However, in many cases, the observed data can be represented as directions, or periodic occurrence times. Data of this type are called directional data. Typical directional data types are circular data, that can be represented as points on a unit circle, toroidal data, cylindrical data or spherical data. Examples of circular data are the departure directions of birds, or the 24 hour clock arrival times of patients in a hospital. An example of toroidal, or circular-circular data can be found in the structure of the dihedral angles of amino acids and an example of cylindrical or circular-linear data can be found as the direction and strength the wind in various locations. Finally, an example of spherical data is the position of the observatories which registered more than a certain amount of rain during the last year.

Directional data have some intrinsic characteristics which require special statistical tools. For example, in the circular case, it is clear that the angles 10^o and 350^o are very close, but the linear mean of these angles is 180^o, while the true mean is 0^o. Thus, it is clear that linear statistical measures such as the mean are not appropriate summaries of these data. Also, due to the shape constraints of directional data, standard probabilistic models are not usually appropriate for such data and special models have been developed for such data.

Most work in directional statistics up to now has concentrated on parametric models. However, the great majority of such models are unimodal and / or symmetric, so that many real data samples cannot be fitted appropriately by such models. Therefore it is interesting to concentrate on alternative, more flexible approaches with fewer parametric assumptions. Up to now, there has been relatively little work on non-parametric or semi-parametric models, except for kernel density based approaches in the non-parametric case and approaches based on mixtures of parametric models in the semi-parametric case. The

main objective of this thesis is thus to develop new non-parametric and semi-parametric models which are appropriate for analyzing directional data. This thesis is organized as follows.

Firstly, in Chapter 1, we review the main characteristics of directional data and the main probabilistic models used for these data. Firstly, we consider the simplest case of circular modeling and then we study circular-circular, circular-linear and spherical models. In each case, we show the main characteristics of the data and some of the most typical models. We also present the usual, classical inference tools for both the univariate and multivariate case and comment on non-parametric approaches to directional data modeling.

In Chapter 2, we present our first class of proposed models, that is, circular Bernstein polynomial distributions. Our approach is based on extending the well known, Bernstein polynomial density approximation of Vitale (1975) to the case of circular data and showing that our model preserves the good properties of the original model.

In this chapter, we introduce the Bernstein polynomials which are usually used for interpolating functions defined on a closed interval and show the convergence rate of this interpolation. Then, following Vitale (1975), we show how to approximate the distribution and density functions of a linear random variable using Bernstein polynomials.

Next, we describe how to adapt linear Bernstein polynomials distributions to the circle and show how, under certain restrictions, the density functions that are obtained satisfy the required properties of circular densities introduced in Chapter 1. Also, we derive the trigonometric moments in a closed form and, using an example based on the cosine distribution, we illustrate that the approximated density converges to the theoretical density and the convergence of the origin.

The main contribution of this chapter is the construction of a nonparametric estimation approach for circular data based on the one developed by Vitale (1975). In particular, we propose to apply certain corrections in order to obtain well defined circular density functions. Moreover, we show that even with these corrections, the convergence rate is the same that the obtained by Vitale (1975) for linear data and it is also the same convergence rate that is obtained using Kernel methods. We illustrate the proposed estimation procedure with a real data set on the occurrence of crimes perpetrated in Chicago on May 11, 2007.

We finish the chapter with various conclusions and possible extensions of the model. These extensions are based on the use of beta distributions in a similar way for the circle such that, under certain restrictions, correspond to mixtures of circular density functions. This model is a semi-parametric alternative to the proposed model.

In Chapter 3, we then construct two models which generalize the circular Bernstein polynomial to the bivariate case using Bernstein copulas: that is a circular-circular model and a circular-linear model.

Firstly, we introduce the generalization of Bernstein polynomials to the bivariate case in a similar way to the considered in Chapter 1 and we review the definition, basic concepts and main types of copulas.

Then, we show using a simple example that it is not possible to adapt the bivariate Bernstein polynomial to the circular-circular case since there do not exist an origin that satisfies property of continuity in two dimensions. Due to this fact, we motivate the use of copulas as an alternative tool to estimate bivariate densities. As is well known, the main advantage of copulas is that they make possible to separate the dependence structure from the marginal densities and this allows us to make use of the estimation methods for univariate circular distributions developed in Chapter 2.

Using a similar approach to the developed in the previous Chapter, the main contribution of Chapter 3 is the construction a nonparametric estimation procedure based on empirical Bernstein copulas (see Sancetta et al, 2004), which are nonparametric estimators of the dependence structure in continuous random variables. As considered for the univariate case, we need to make certain corrections to obtain continuous distributions. However, we will show that these corrections preserve the asymptotic uniformity of the marginal distributions, such that both models the circular-circular and the circular-linear models are well defined.

Our proposed estimation procedures are illustrated using two environmental data sets, based on wind directions and rainfall levels. We finish with some conclusions and extensions of the models.

Finally, in Chapter 4, we construct two new bivariate models, a circular-circular model and a spherical model, making use of another class of polynomials, that is the trigonometric polynomials. Such polynomials are based on sums of sinusoidal terms and the key problem when forming density functions based on polynomials of this type is to ensure that they are non-negative. In the univariate case, the well-known Fejér-Riesz theorem, (Fejér, 1915) gives the necessary conditions for this to occur and this result was used to derive a circular density approximation in Fernández-Durán (2004). Here, we use results in Gerónimo et al (2004) to generalize this model to the case of circular-circular and spherical data. In particular, for the extension to spherical data, we have followed an original idea based on Merilees (1973) and Boer et al (1975) to make a (non-bijective) transformation of a sphere into a torus. Our procedures for circular-circular and spherical data are illustrated with two real data sets and

we finish the chapter with some conclusions and extensions of our proposed models.

Contents

1 Introduction **1**

1.1 Circular statistics . . . 1

1.1.1 Circular data . . . 2

1.1.2 Circular random variables . . . 5

1.2 Multivariate directional data . . . 8

1.2.1 Circular-circular models . . . 8

1.2.2 Circular-linear distributions . . . 13

1.2.3 Spherical distributions . . . 14

1.3 Density estimation . . . 18

1.4 Semi-parametric and non-parametric approaches . . . 19

1.4.1 Semi-parametric models . . . 19

1.4.2 Non-parametric approaches . . . 20

2 Circular Bernstein polynomial distributions **23**

2.1 Bernstein polynomials . . . 24

2.2 Bernstein polynomial distributions . . . 26

2.2.1 Non-parametric inference via Bernstein polynomials . . . 27

2.3 The circular Bernstein polynomial distribution . . . 28

2.4 Trigonometric moments . . . 31

2.5 Model fitting . . . 35

2.5.1 Tied data . . . 38

2.6 Illustrations . . . 39

2.6.1 A theoretical example: the cosine data . . . 39

2.6.2 Behaviour of the origins $\hat{\nu}$ . . . 39

2.6.3 A practical example: the Chicago crimes data . . . 40

2.7 Extensions . . . 41

2.7.1 Mixtures of scaled, shifted, beta distributions . . . 42

2.7.2 Other extensions . . . 45
2.8 Conclusions . . . 46

3 Bernstein copulas and applications **47**
3.1 Bivariate Bernstein polynomials . . . 48
3.1.1 Bivariate Bernstein polynomial distributions . . . 49
3.2 Copulas . . . 50
3.2.1 The empirical and empirical Bernstein copulas . . . 52
3.3 Circular-circular distributions . . . 53
3.3.1 Bivariate Bernstein polynomial distributions . . . 53
3.3.2 Modeling using the Bernstein copula . . . 56
3.4 Circular-linear distributions . . . 59
3.5 Illustrations . . . 61
3.5.1 Circular-circular data: wind directions . . . 61
3.5.2 Circular-linear data: wind directions and rainfall . . . 63
3.6 Conclusions and extensions . . . 67

4 Distributions based on Fourier series **69**
4.1 Fourier series and trigonometric polynomials . . . 70
4.1.1 Univariate Fourier series and their properties . . . 70
4.1.2 Non-negative trigonometric polynomials: The Fejér-Riesz theorem . . . 72
4.1.3 Bivariate Fourier series . . . 73
4.1.4 The two dimensional Fejér-Riesz theorem . . . 75
4.1.5 Fourier series on the sphere . . . 79
4.2 Positive trigonometric polynomial distributions . . . 81
4.2.1 Circular distributions . . . 81
4.2.2 A circular-circular or toroidal distribution . . . 82
4.2.3 Spherical distributions . . . 84
4.3 Illustrations . . . 86
4.3.1 Circular circular data . . . 86
4.3.2 Spherical data . . . 89
4.4 Conclusions and extensions . . . 90

List of Figures

1.1 Linear statistics do not work 3
1.2 Rose plot (top left), circular dot plot (bottom left) and linear histogram (bottom right)of the cosine data. 4
1.3 Circular and linear density plots of the cosine density with $\mu = \pi$ and $\rho = 0.3$ 7
1.4 The bivariate von-Mises density. 11
1.5 The bivariate cosine density 12
1.6 The Mardia-Sutton density with $(\mu_0 = \pi, \kappa = 1, \rho = 0.5, \mu = 0, \sigma = 1)$ 14
1.7 Spherical co-ordinates 15
1.8 Probability density element of the Fisher-von Mises distribution with $\mu = (0, \pi), \kappa = 1$(equator) 17
1.9 True (solid line) and fitted maximum likelihood (dashed line) and method of moment (dotted line) based density estimates 19

2.1 Approximation of the cosine distribution (solid line) with a circular Bernstein polynomial with degrees 50 (dashed line), 100 (dotted line) and 300 (dash dot line). 30
2.2 Approximation of the cosine distribution (solid line) with circular Bernstein polynomials of order 50 with origins 0 (dashed line) and π (dotted line). . . . 31
2.3 True (solid line) and fitted (dashed line) densities for the cosine data 40
2.4 Kernel density estimates of the density of $\hat{\nu}$ for samples of size 100 (dotted line), 1000 (dashed line) and 10000 (solid line). 41
2.5 Roseplot (left hand side) and histogram of the Chicago crime data with fitted Bernstein polynomial (solid line) and kernel density (dashed line) approximations. 42
2.6 Histogram of the Chicago crime data with fitted mixture of SSB distribution with integers parameters 44

2.7 Histogram of the Chicago crime data with fitted mixture of SSB distribution with continuous parameters 45

3.1 Marginal densities of the wind directions at buoys MISM1 (top) and MDRM1 (bottom) 61
3.2 Density function for the circular-circular model, $f(\theta_1, \theta_2)$. 62
3.3 Conditional density function, $f(\theta_1|\theta_2)$ 63
3.4 Conditional density function, $f(\theta_2|\theta_1)$ 64
3.5 Marginal densities of wind direction: whole data set (top), dry days (middle), rainy days (bottom). 65
3.6 Fitted bivariate density function, f(x,$\theta|X > 0$) 66
3.7 Conditional Distribution functions, F(x$|\theta, X > 0$) given the two modal wind directions. 67
3.8 Contour plot of $F(X \leq x|\theta, X > 0)$ 68

4.1 Estimated densities f(θ_1, θ_2) 87
4.2 Densities of the circular-circular model using copulas and bivariate trigonometric polynomials 88
4.3 Dot plot of the data) 89
4.4 Rotationally symmetric spherical density approximation 90
4.5 Estimated density at 4 degrees longitude and 10 degrees latitude. 91

Chapter 1

Introduction

Directional statistics is a part of statistics that have special treatment. In this chapter we show its special characteristics in both the univariate and multivariate case. As we will see below, these directions must be converted to angles in order to be analyzed. We illustrate through an example that linear statistics do not work with this type of data and then we give the formulae for analyzing this type of data.

The models for directional data must acquire their intrinsic properties. We will show these properties, the typical examples of densities used for modelling directional data in one and two dimensions and the different approaches to estimate the densities in each case.

This chapter is organized as follows. In Section 1.1, we extend the concept of circular data and show a brief summary of its characteristics, trigonometric moments, typical models, etc. In Section 1.2, we generalize to the multivariate case focusing on the bivariate case, that are the torus, the cylinder and the sphere. Then, in Section 1.3, we comment briefly on model fitting for directional data models, concentrating on the classical statistical approach. We finish in Section 1.4 with some brief comments on semi-parametric and non-parametric modelling for directional data.

1.1 Circular statistics

In this section, we explore the characteristics of circular data and the most important models for data of this kind.

1.1.1 Circular data

Circular data are data that can be represented as points on the circumference of a unit circle or as a unit vector in the plane. Given a suitable choice of origin and a sense of rotation, then data of this type can be transformed to angles. An important characteristic is that different origins or a different sense of rotation provide different values for the same observation. A second important property is periodicity. Suppose that a observation θ is given. Then given a circuit of length $2p\pi$ around the circumference of the circle, for $p \in \mathbb{N}$, we return to the same point, that is $\theta = \mod(\theta + 2p\pi, 2\pi)$. The consequence is that methods for analyzing this type of data should take into account how to measure the distance between any two observations.

Sample circular moments

Given the special properties of circular data, as noted in Section 1.1.1, then it is not surprising that standard summary measures for linear data can often provide unreasonable results.

Example 1. *Consider a sample of two observations,* $\theta_1 = 15\,^\circ$ *and* $\theta_2 = 345\,^\circ$*. Then the linear average of these two data is simply* $180\,^\circ$*, but as we can see in Figure 1.1, this value is not a sensible summary of the location of the data.*

For computing the different sample statistics we must compute the sample trigonometric moments and use them as we will see below. The p'th sample trigonometric moment can be defined as follows:

Definition 1. *Let* $\theta_1, \ldots, \theta_n$ *be a random sample from a circular variable* Θ*. The p'th sample trigonometric moment is given by:*

$$m_p = \bar{C}_p + i\bar{S}_p$$

where $i = \sqrt{-1}$ *and*

$$\begin{aligned} \bar{C}_p &= \frac{1}{n}\sum_{j=1}^{n} \cos p\theta_j \quad \textit{and} \\ \bar{S}_p &= \frac{1}{n}\sum_{j=1}^{n} \sin p\theta_j. \end{aligned}$$

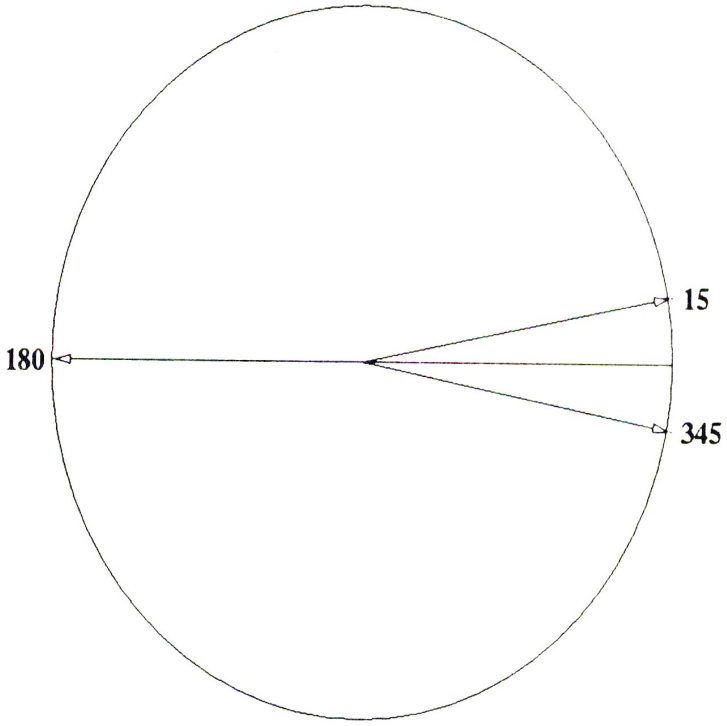

Figure 1.1: Linear statistics do not work

The most useful circular statistics are the sample circular mean direction and the mean resultant length which are measures of circular location and concentration respectively. These statistics are defined below.

Definition 2. *Given a sample of circular data, $\theta_1, \ldots, \theta_n$, the sample circular mean direction, $\bar{\theta}$, is defined as,*

$$\bar{\theta} = \begin{cases} \arctan\left(\bar{S}_1/\bar{C}_1\right), & \bar{C}_1 \geq 0, \\ \arctan\left(\bar{S}_1/\bar{C}_1\right) + \pi, & \bar{C}_1 < 0, \end{cases} \tag{1.1}$$

and the mean resultant length is,

$$\bar{R}_1 = \sqrt{\bar{C}_1^2 + \bar{S}_1^2}. \tag{1.2}$$

In contrast to the linear mean, the circular mean does provide a correct estimate of the average direction of the data. In particular, in the case of Example 1, it is easy to see that the circular mean direction is equal to $0\,^\circ$. Note also that when the data are perfectly uniformly distributed over the circle, then the circular mean will not exist, as in this case,

$C_1 = S_1 = 0$ and the mean resultant length will take the value 0. In contrast, when the data are concentrated on a single point, it is easy to see that the circular mean is equal to this point and the mean resultant length takes the value 1. Thus, low values of the mean resultant length correspond to disperse data and high values correspond to concentrated data. Note that many other circular statistics which measure different characteristics of the data have been developed. For a complete review, see e.g. Fisher (1993) and Jammalamadaka et al (2001).

Displaying circular data

There are several display options for this type of data. The most usual are the rose plot or angular histogram, the dot plot and, given an origin about which the circle is unwrapped, we obtain a linear histogram.

Example 2. *The following plots show 200 simulated circular data generated from a cosine distribution, see Example 4. As we can see, the sample is unimodal with a mode near to* π.

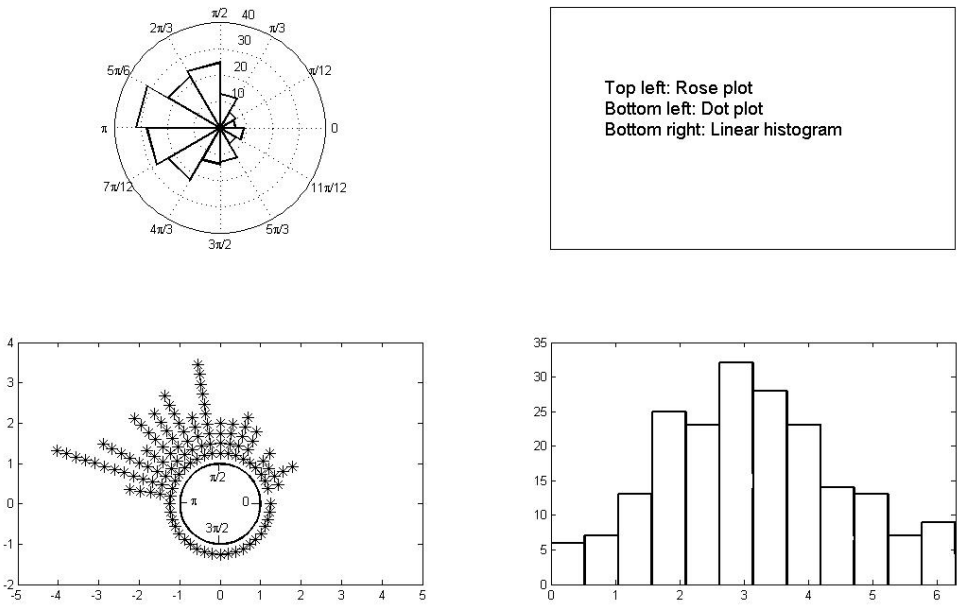

Figure 1.2: Rose plot (top left), circular dot plot (bottom left) and linear histogram (bottom right)of the cosine data.

1.1.2 Circular random variables

The density function of a circular random variable, Θ, is defined as follows:

Definition 3. *Let $f(\cdot)$ be the density function of a circular random variable Θ. The density function satisfies*

1. $f(\theta) \geq 0, \quad for \ -\infty \leq \theta < \infty,$
2. $f(\theta) = f(\theta + 2\pi j), \quad for\ all\ j \in \mathbb{Z},$
3. $\int_{\nu}^{\nu+2\pi} f(\theta)\, d\theta = 1, \quad for\ 0 \leq \nu < 2\pi.$

Thus, the density is a periodic, non-negative function with period 2π which integrates to 1 over any region of length 2π.

In order to define the distribution function, we must first select an origin ν. For such an origin, the distribution function is defined in the usual way as:

$$F^{\nu}(\theta) = \int_{\nu}^{\nu+\theta} f(\theta)\, d\theta, \quad \text{for } 0 \leq \theta < 2\pi. \tag{1.3}$$

Trigonometric moments

Trigonometric moments for circular random variables can be computed in a similar way to for circular data as seen in definition 1. Formally, we have the following definition.

Definition 4. *Let Θ be a circular random variable. The p'th trigonometric moment is given by:*

$$\begin{aligned} \mu'_p &= E[\cos p\Theta] + iE[\sin p\Theta] \\ &= C_p(\theta) + iS_p(\theta) \end{aligned}$$

For example, if we want to compute the mean direction and the mean resultant length for Θ, we must use the trigonometric moment of Θ in (1.1) and (1.2), respectively, such that the mean direction and the mean resultant length are given by:

$$\bar{\Theta} = \begin{cases} \arctan\left(S_1/C_1\right) & C_1 \geq 0 \\ \arctan\left(S_1/C_1\right) + \pi & C_1 < 0 \end{cases} \tag{1.4}$$

$$\bar{R}_1 = \sqrt{C_1^2 + S_1^2} \tag{1.5}$$

Parametric circular distributions

Many parametric families of circular distributions have been developed. The simplest circular model is the circular uniform distribution which is often used as a baseline with which to compare more complicated models.

Example 3 (The circular uniform distribution)**.** *This distribution assigns the same density to every point on the circle and thus, has density function:*

$$f(\theta) = \frac{1}{2\pi}, \quad \textit{for } 0 \leq \theta < 2\pi.$$

The mean resultant length of this distribution is 0 and the circular mean direction is undefined.

Some models have been developed specifically for circular data.

Example 4 (The cosine or cardioid distribution)**.** *The cosine or cardioid distribution (Jeffreys, 1948) is a two parameter distribution with density function given by:*

$$f(\theta) = \frac{1}{2\pi}\left\{1 + 2\rho\cos\left(\theta - \mu\right)\right\}, \quad \textit{for } 0 \leq \theta < 2\pi,$$

where $0 \leq \mu < 2\pi$ *is the circular mean and* $0 \leq \rho \leq 1/2$ *is the mean resultant length. This distribution converges to the circular uniform distribution when* $\rho \to 0$*. We shall use this density throughout the following chapter to illustrate various features of our proposed models.*

Figure 1.3 gives circular and linear density plots of the cosine density with $\mu = \pi$ *and* $\rho = 0.3$*. This is the density used to generate the data in Example 17.*

Example 5 (von Mises distribution)**.** *This model, developed in Langevin (1907) and von Mises (1918), is one of the most important circular distributions and has many similar properties to the normal distribution for linear data. The von Mises density is given by:*

$$f(\theta, \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)}\exp\left[\kappa\cos(\theta - \mu)\right], \quad \textit{for } 0 \leq \theta < 2\pi,$$

where $0 \leq \mu < 2\pi$ *is the mean direction, and* $0 \leq \kappa < \infty$ *is a concentration parameter and*

$$I_0(\kappa) = \frac{1}{2\pi}\int_0^{2\pi} exp\left[\kappa\cos(\phi - \mu)\right] d\phi,$$

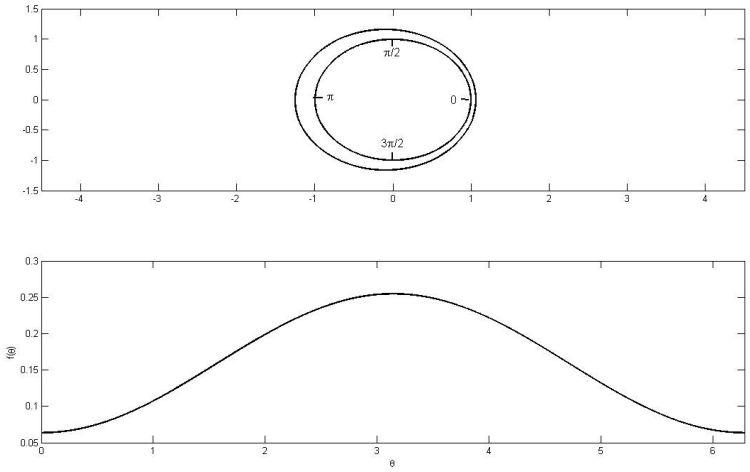

Figure 1.3: Circular and linear density plots of the cosine density with $\mu = \pi$ and $\rho = 0.3$

is the modified Bessel function of order 0. When $\kappa \to 0$*, this distribution approaches the uniform distribution and when* $\kappa \to \infty$*, it approaches a point mass at* μ*.*

Other approaches have been devised to translate standard distributions on the real line into circular models. One such class of models are wrapped distributions. Let $g(\cdot)$ be the density of a linear random variable X. Then, we can define a density for a circular variable θ by wrapping the real line onto the circle so that,

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

Example 6 (Wrapped exponential distribution)**.** *Consider an exponential density on the line, that is,*

$$g(x) = \lambda e^{-\lambda x}, \quad \textit{for } 0 \leq x < \infty$$

, where $\lambda > 0$*. Then, wrapping the exponential distribution as in Jammalamadaka et al (2004) onto the circle gives*

$$f(\theta) = \sum_{j=0}^{\infty} \lambda e^{-\lambda(\theta + 2\pi j)} = \frac{\lambda e^{-\lambda\theta}}{1 - e^{-2\pi\lambda}}, \quad \textit{for } 0 \leq \theta < 2\pi.$$

Other examples of this type of distribution are the wrapped normal (Schmidt, 1917), wrapped t family (Pewsey et al, 2007), wrapped skew-normal (Pewsey, 2000) and wrapped Laplace

(Jammalamadaka et al, 2004) distributions among others.

Another approach is to use offset distributions. Thus, we consider a bivariate random variable $\mathbf{X} = (X_1, X_2)$ with support in $\mathbb{R}^2$ and then transform to polar coordinates (R, θ) and integrate out over R to develop a circular distribution.

Example 7 (Offset normal distribution)**.** *Mardia (1972) assumes that* (X_1, X_2) *follow a bivariate normal random variable with each component having mean 0 and unit variance and correlation* $0 \leq \rho \leq 1$*. Then, this leads to a circular density of form*

$$f(\theta) = \frac{\sqrt{1-\rho^2}}{2\pi(1-\rho\sin 2\theta)}, \quad \textit{for } 0 \leq \theta < 2\pi.$$

For other approaches see e.g. Jammalamadaka et al (2001).

1.2 Multivariate directional data

In this section, we generalize the results of the previous sections to the multivariate setting. Here we shall consider three cases, that is circular-circular or toroidal distributions, i.e. joint distributions of two circular random variables, circular-linear distributions, i.e. joint distributions of one circular and one linear random variable and spherical distributions, i.e. distributions of variables with support defined on the surface of a sphere.

1.2.1 Circular-circular models

As we showed in the previous section, the circular data can be represented as points on the circumference of the unit circle. In this generalization, both components of the observed data can be represented on the circumference of a unit circle. This implies that circular-circular data can be represented on the surface of a torus that is the direct product of circles.

For displaying a circular-circular data set the unwrapped approach is simplest. This approach consists in choosing an origin point and unwrapping the two variables to construct a plane.

Sample trigonometric moments

The sample trigonometric moments can be defined as follows.

Definition 5. *Let $(\theta_{11}, \theta_{21}), \ldots, (\theta_{1n}, \theta_{2n})$ be a sample from a circular-circular variable (Θ_1, Θ_2). The (p_1, p_2)'th sample trigonometric moment is given by*

$$\begin{aligned}\frac{1}{n}\sum_{j=1}^{n} e^{i(p_1\theta_{1j}+p_2\theta_{2j})} &= \frac{1}{n}\sum_{j=1}^{n}\cos(p_1\theta_{1j}+p_2\theta_{2j}) + i\sin(p_1\theta_{1j}+p_2\theta_{2j}) \\ &= a_{p_1p_2} - c_{p_1p_2} + i(b_{p_1p_2}+d_{p_1p_2})\end{aligned}$$

where

$$\begin{aligned}a_{p_1p_2} &= \frac{1}{n}\sum_{j=1}^{n}\cos(p_1\theta_{1j})\cos(p_2\theta_{2j}) \\ b_{p_1p_2} &= \frac{1}{n}\sum_{j=1}^{n}\cos(p_1\theta_{1j})\sin(p_2\theta_{2j}) \\ c_{p_1p_2} &= \frac{1}{n}\sum_{j=1}^{n}\sin(p_1\theta_{1j})\cos(p_2\theta_{2j}) \\ d_{p_1p_2} &= \frac{1}{n}\sum_{j=1}^{n}\sin(p_1\theta_{1j})\sin(p_2\theta_{2j})\end{aligned}$$

Letting $p_2 = 0$ ($p_1 = 0$) in the above definition gives the marginal trigonometric moments of order p_1 (p_2) of the sample from Θ_1 (Θ_2). In contrast to the case of linear variables where there is a standard measure of correlation, many measures of correlation for circular variables have been proposed. For example, Fisher et al (1983) propose defining

$$r = \frac{\sum_{j=1}^{n-1}\sum_{k=j+1}^{n}\sin(\theta_{1j}-\theta_{1k})\sin(\theta_{2j}-\theta_{2k})}{\sqrt{\sum_{j=1}^{n-1}\sum_{k=j+1}^{n}\sin^2(\theta_{1j}-\theta_{1k})\sum_{j=1}^{n-1}\sum_{k=j+1}^{n}\sin^2(\theta_{2j}-\theta_{2k})}} \tag{1.6}$$

and show that this correlation coefficient takes values between -1 and 1 and, when there is no relation between the two sets of data takes the value 0. Alternative correlation coefficients are reviewed in e.g. Rivest (1982) or Jammalamadaka et al (2001).

Circular-circular random variables

Similar to the univariate case, the density function of a circular-circular variable must satisfy certain properties. We can define the density function of a circular-circular random variable as follows:

Definition 6. *Let $f(\cdot,\cdot)$ be the density function of a circular-circular random variable (Θ_1,Θ_2). The density function satisfies*

1. $f(\theta_1,\theta_2) \geq 0, \quad for\ -\infty \leq \theta < \infty,$
2. $f(\theta_1,\theta_2) = f(\theta_1 + 2j\pi, \theta_2 + 2k\pi), \quad for\ j,k \in \mathbb{Z},$
3. $\int_{\nu_1}^{\nu_1+2\pi}\int_{\nu_2}^{\nu_2+2\pi} f(\theta_1,\theta_2)\,d\theta_1\,d\theta_2 = 1, \quad for\ 0 \leq \nu_1,\nu_2 < 2\pi.$

In order to define the distribution function we must select an origin (ν_1,ν_2) such that the distribution function is given by

$$F^{\nu_1,\nu_2}(\theta_1,\theta_2) = \int_{\nu_1}^{\nu_1+\theta_1}\int_{\nu_2}^{\nu_2+\theta_2} f(\theta_1,\theta_2)d\theta_1 d\theta_2, \qquad \text{where } 0 \leq \theta_1,\theta_2 \leq 2\pi.$$

Definition 7. *The (p_1,p_2)'th trigonometric moment of a circular-circular variable (Θ_1,Θ_2) can be computing using*

$$\begin{aligned}
\alpha_{p_1p_2} &= \int_{\nu_1}^{\nu_1+2\pi}\int_{\nu_2}^{\nu_2+2\pi} f(\theta_1,\theta_2)\cos(p_1\theta_1)\cos(p_2\theta_2)d\theta_1 d\theta_2 \\
\beta_{p_1p_2} &= \int_{\nu_1}^{\nu_1+2\pi}\int_{\nu_2}^{\nu_2+2\pi} f(\theta_1,\theta_2)\cos(p_1\theta_1)\sin(p_2\theta_2)d\theta_1 d\theta_2 \\
\gamma_{p_1p_2} &= \int_{\nu_1}^{\nu_1+2\pi}\int_{\nu_2}^{\nu_2+2\pi} f(\theta_1,\theta_2)\sin(p_1\theta_1)\cos(p_2\theta_2)d\theta_1 d\theta_2 \\
\lambda_{p_1p_2} &= \int_{\nu_1}^{\nu_1+2\pi}\int_{\nu_2}^{\nu_2+2\pi} f(\theta_1,\theta_2)\sin(p_1\theta_1)\sin(p_2\theta_2)d\theta_1 d\theta_2
\end{aligned}$$

Example 8 (Bivariate von Mises distribution)**.** *One useful distribution on the torus is the bivariate von Mises model (Mardia, 1975). Its probability density function is proportional to:*

$$f(\theta_1,\theta_2) = exp^{\kappa_1\cos(\theta_1-\mu_1)+\kappa_2\cos(\theta_2-\mu_2)+(\cos\theta_1,\sin\theta_1)^T\mathbf{A}(\cos\theta_2,\sin\theta_2)}, \quad for\ 0 \leq \theta_1,\theta_2 < 2\pi,$$

where $0 \leq \mu_1,\mu_2 < 2\pi$, $0 \leq \kappa_1,\kappa_2 < \infty$ and where $\mathbf{A}$ is a 2×2 matrix. The marginal distributions of θ_1 and θ_2 are von Mises if and only if either $\mathbf{A} = \mathbf{0}$ (so that θ_1 and θ_2 are independent) or $\kappa_1 = \kappa_2 = 0$ and $\mathbf{A}$ is a multiple of an orthogonal matrix. Figure 1.4 shows the case where $(\mu_1 = \mu_2 = \pi, \kappa_1 = \kappa_2 = 1, A = 0)$.

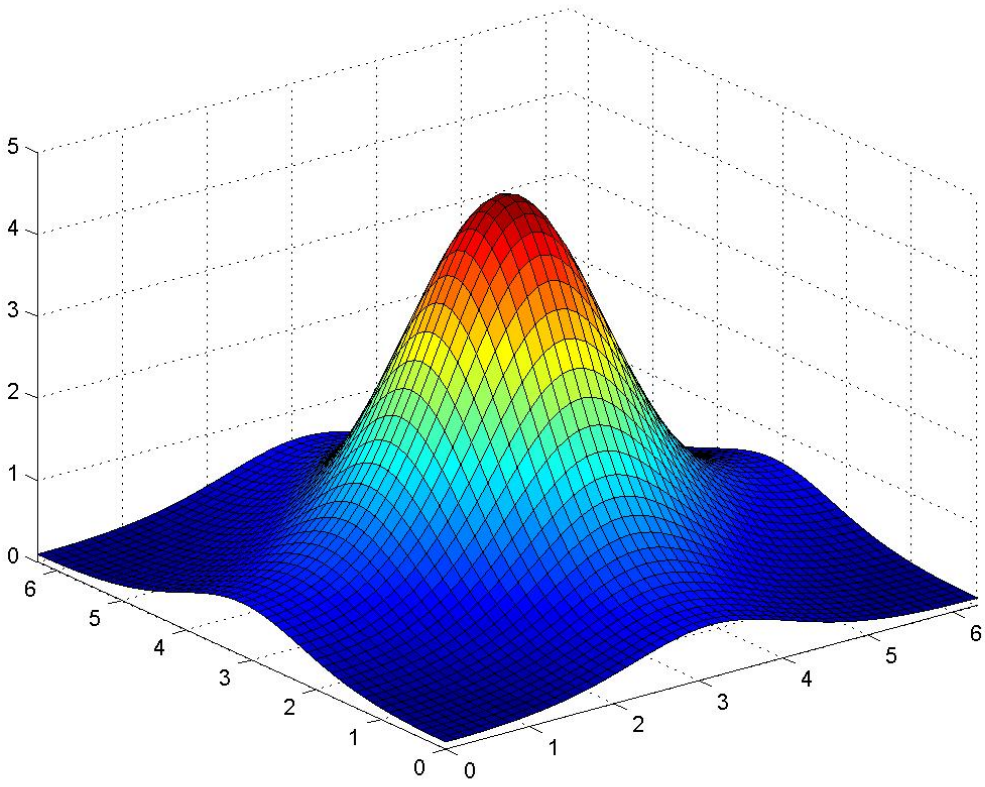

Figure 1.4: The bivariate von-Mises density.

Example 9 (Bivariate cosine distribution with helical symmetry)**.** *In this example we show how to implement a helical symmetry in the density of a distribution. The key step is that the mean direction of each component is just the value of the other component. Figure 1.5 shows the density of a circular-circular version of the cosine distribution with* $\rho = 0.5$ *given by:*

$$f(\theta_1, \theta_2) = \frac{1}{4\pi^2} \left(1 + \cos(\theta_1 - \theta_2)\right). \tag{1.7}$$

Observe that both marginal distributions are circular uniform distributions and the theoretical circular correlation is one but the conditional distribution is a cosine distribution. We can use any other circular distribution where the probability of a point depends on the distance from the mean direction, e.g. the von Mises distribution.

The general form of these type of distributions can be obtained multiplying the angles by a integer. Then, the general form of the cosine distribution with $\rho = 0.5$ *is given by:*

$$f(\theta_1, \theta_2) = \frac{1}{4\pi^2} \left(1 + \cos(j\theta_1 - k\theta_2)\right). \tag{1.8}$$

where $j, k \in \mathbb{N}$ *represent the number of modes of the marginal distributions.*

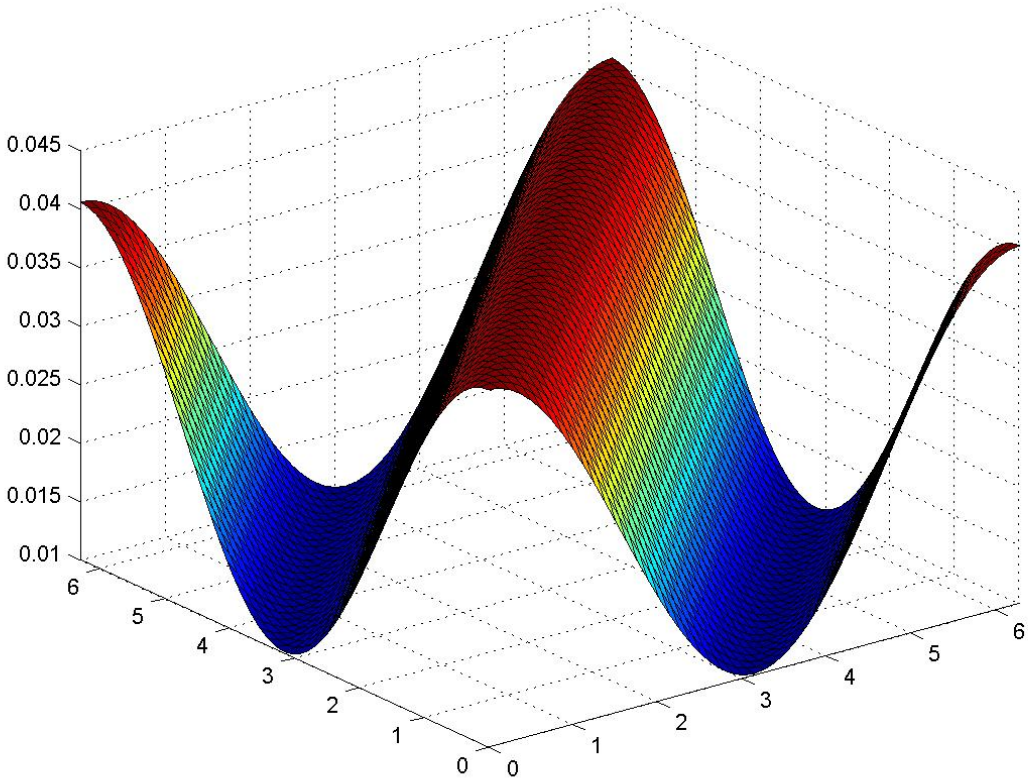

Figure 1.5: The bivariate cosine density

Other examples of circular-circular distribution are the uniform distribution on the torus, the wrapped bivariate normal distribution, (Jammalamadaka et al, 1988) and for maximum entropy circular-circular distributions (Kagan et al, 1973).

1.2.2 Circular-linear distributions

The second type of data we shall examine in this thesis is circular-linear data, where one component of the data is circular and the other component is linear. Given that circular data is represented on the circumference of the unit circle and linear data is represented on the real line, then we can represent this data on the surface of a cylinder. To display data of this type, in a similar way to the circular-circular model, we just need to unwrap the cylinder onto the plane.

Circular-linear random variables

We can define the density function of a circular-linear random variable as follows:

Definition 8. *Let $f(\cdot,\cdot)$ be the density function of a circular-linear random variable (Θ, X). The density function satisfies:*

1. $f(\theta, x) \geq 0, \quad -\infty \leq \theta, x < \infty,$
2. $f(\theta, x) = f(\theta + 2j\pi, x), \quad j \in \mathbb{Z},$
3. $\int_{-\infty}^{\infty} \int_{\nu}^{\nu+2\pi} f(\theta, x)\, d\theta dx = 1, \quad 0 \leq \nu < 2\pi.$

In order to define the distribution function we must select an origin ν such that the distribution function is given by

$$F^{\nu}(\theta, x) = \int_{-\infty}^{x} \int_{\nu}^{\nu+\theta} f(\theta, x)\, d\theta dx, \qquad \text{where } 0 \leq \theta \leq 2\pi, -\infty \leq x \leq \infty.$$

Example 10. *An example of cylindrical distribution can be found in Mardia et al (1978), where the density of (Θ, X) is given by:*

$$f(\theta, x) = \frac{1}{(2\pi I_0(\kappa)} e^{\kappa \cos(\theta - \mu_0)} \frac{1}{\sqrt{2\pi}\sigma_c} e^{-\frac{(x-\mu_c)^2}{2\sigma_c^2}}$$

where $\sigma_c^2 = \sigma^2(1-\rho^2)$ and $\mu_c = \mu + \sigma\kappa^{\frac{1}{2}}\left(\rho_1 \cos(\theta - \mu_0) + \rho_2 \sin(\theta - \mu_0)\right)$ and $\rho = \sqrt{\rho_1^2 + \rho_2^2}$ for $0 \leq \mu_0 < 2\pi$, $\kappa \geq 0$, $\mu \in \mathbb{R}$, $\sigma > 0$ and $0 \leq \rho \leq 1$. This implies that the marginal distribution of Θ is a von Mises distribution with parameters μ_0 and κ and the conditional distribution of X given θ is : $X|\theta \sim N(\mu_c, \sigma_c^2)$.

Figure 1.6 shows the density with parameters ($\mu_0 = \pi, \kappa = 1, \rho = 0.5, \mu = 0, \sigma = 1$).

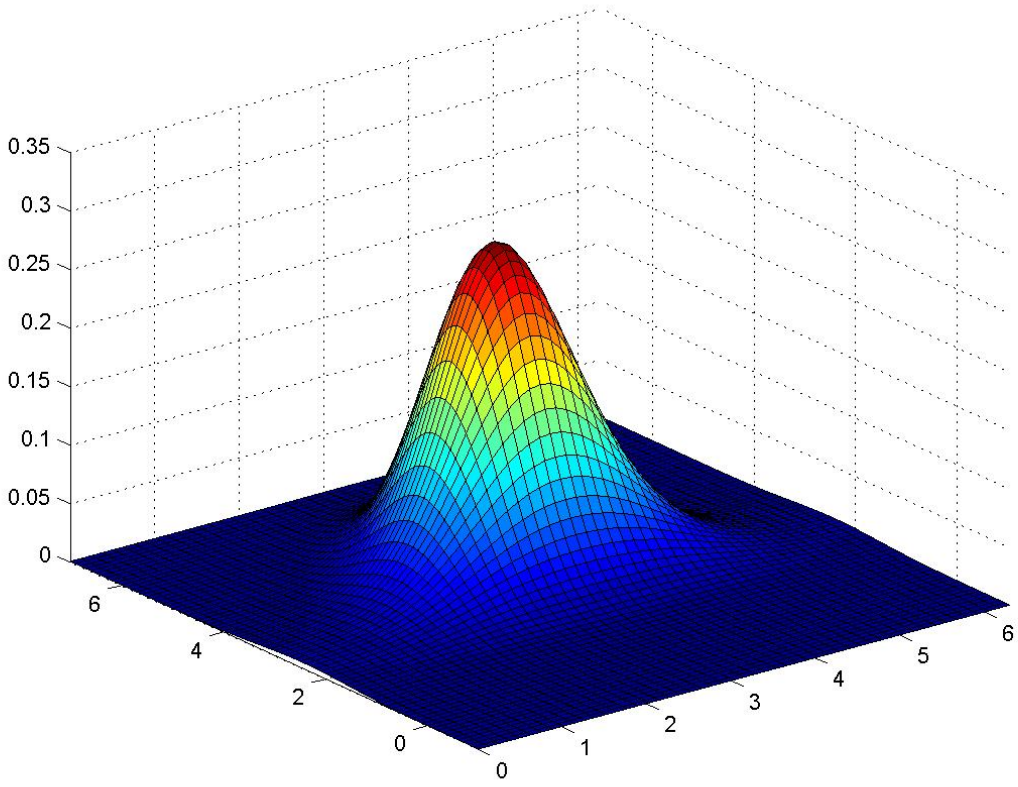

Figure 1.6: The Mardia-Sutton density with $(\mu_0 = \pi, \kappa = 1, \rho = 0.5, \mu = 0, \sigma = 1)$

Other examples of circular-linear distributions can be found in Batschelet (1981) and Johnson et al (1978). Also, suitable measures of circular-linear correlation are discussed in e.g. Johnson et al (1977).

1.2.3 Spherical distributions

Spherical data are data that take values on the surface of a sphere, with radius r. Thus, using linear co-ordinates, (x, y, z), about an origin at the centre of the sphere to represent data points, we have that such data satisfy the constraint $x^2 + y^2 + z^2 = r^2$. Such data can also be written in polar co-ordinates, (θ, ϕ), where $0 \leq \theta < 2\pi$ represents the longitude and

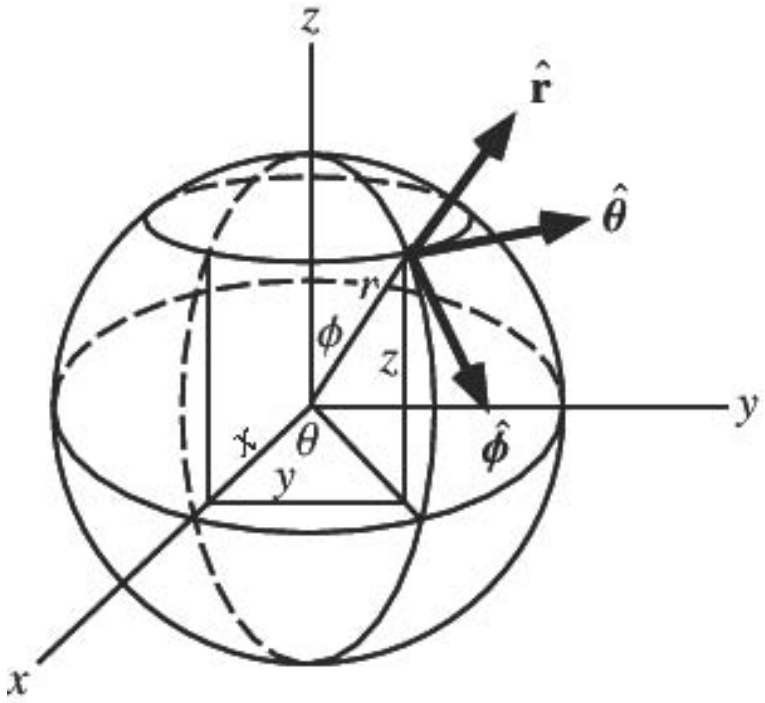

Figure 1.7: Spherical co-ordinates

$0 \leq \phi < \pi$ is the colatitude, or angle from the pole. Then we have:

$$x = r\cos\theta\sin\phi, \quad y = r\sin\theta\sin\phi, \quad z = r\cos\phi.$$

These two coordinate systems are illustrated in Figure 1.7, where the origin, $\theta = \phi = 0$, is the north pole.

In order to visualize spherical data, one possibility is simply to plot the surface of the sphere from various different viewpoints in order to see all aspects of the sphere, the so called stereographic projections. Another option is to project the sphere onto the plane using e.g. a Mercator projection or some other form of map projection, see e.g. Snyder (1997). However, projections of this type will always lead to some level of distortion of the data.

Spherical random variables

Without loss of generality, we shall consider variables defined on the sphere, $\mathbb{S}$, say with radius $r = 1$ fixed.

Consider a spherical random variable (Θ, Φ) where Θ represents the longitude with sup-

port $[0, 2\pi)$ and Φ represents the colatitude with support $[0, \pi)$. Assume that $f(\cdot, \cdot)$ is the density function of (Θ, Φ) so that $f(\theta, \phi) \geq 0$ for $[0 \leq \theta < 2\pi)$ and $[0 \leq \phi < \pi)$ and $\int_0^{2\pi} \int_0^{\pi} f(\theta, \phi)\, d\phi\, d\theta = 1$.

Note however that case needs to be taken when interpreting the density function. Firstly, it is important to note that if we consider the region $\theta < \Theta < \theta + d\theta$, $\phi < \Phi < \phi + d\phi$, then the area of this region is not $d\theta d\phi$, but is instead $\sin\phi d\theta d\phi$, which we can write as $d\$$, as it represents an element of the surface of the sphere. Then the *probability density element*, that is the proportion of values of (Θ, Φ) in the same small region, $d\$$ is given by $f(\theta, \phi)d\theta d\phi$, but this is not the same as $f(\theta, \phi)d\$$, and is instead equal to $h(\theta, \phi)d\$$ where

$$f(\theta, \phi) = h(\theta, \phi) \sin\phi. \tag{1.9}$$

Spherical distributions can be specified equivalently in terms of the density function $f(\theta, \phi)$ or in terms of the probability density element $h(\theta, \phi)d\$$.

To interpret these two concepts, consider the following illustration. Take the surface of the sphere and divide this in 1000 regions of equal area and similar form. If we choose one of them randomly, all of them are equally probable and this represents the probability density element. On the other hand, the number of regions close to the equator is greater than the number of regions near the poles. Then we are taking into account the element of surface like the density function.

The following examples illustrate two of the most well known spherical distributions.

Example 11 (Uniform distribution on the sphere)**.** *In this case, the probability density element is given by:*

$$h(\theta, \phi)d\$ = \frac{1}{4\pi} d\$$$

and therefore, the density function is:

$$f(\theta, \phi) = \frac{\sin\phi}{4\pi} \quad \textit{for } 0 \leq \theta < 2\pi \textit{ and } 0 \leq \phi < \pi.$$

Example 12 (von Mises-Fisher distribution)**.** *The most well known spherical distribution is the von Mises-Fisher distribution (Fisher, 1953) with probability density element given by:*

$$h(\theta, \phi)d\$ = C_f e^{\kappa(\sin\theta \sin\alpha \cos(\phi - \beta) + \cos\theta \cos\alpha)} d\$$$

where

$$C_f = \frac{\kappa}{4\pi \sinh \kappa}.$$

This distribution has two location parameters α and β which give us the point of rotationally symmetry and a concentration parameter κ about this point.

Similar to the von Mises distribution, when $\kappa \to 0$ the distribution tends to a uniform distribution on the sphere and when $\kappa \to \infty$ the distribution tends to a point distribution at (α, β).

Figure 1.8 shows the probability density element of the Fisher-von Mises distribution.

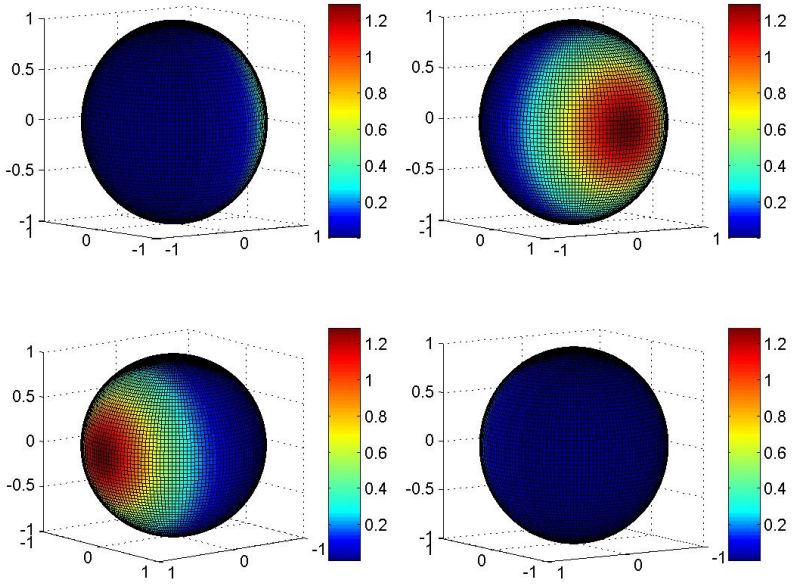

Figure 1.8: Probability density element of the Fisher-von Mises distribution with $\mu = (0, \pi), \kappa = 1$(equator)

Other well known examples of spherical models are the Watson distribution (Watson, 1965), the Kent distribution (Kent, 1982), the Wood distribution (Wood, 1982), the Bingham distribution (Bingham, 1964), the Arnold distribution (Arnold, 1941) and the Fisher-Watson distribution (Wood, 1988).

1.3 Density estimation

In this section we deal with the different methods for estimating the parameters of the distributions showed above.

From a parametric point of view, the univariate distributions can be estimated using maximum likelihood but only a few distributions have closed forms for the maximum likelihood estimators. In other cases, numerical techniques will typically be needed.

Example 13. *Suppose that we have a sample of data, say $\theta_1, \ldots, \theta_n$ generated from a von Mises distribution as in Example 5. Then, the log likelihood is given by:*

$$\log l(\boldsymbol{\theta}|\mu, \kappa) = -n \log 2\pi I_0(\kappa) + \kappa \sum_{j=1}^{n} \cos(\theta_j - \mu)$$

and differentiating with respect to μ gives:

$$\frac{\partial \log l}{\partial \mu} = \kappa \sum_{j=1}^{n} \sin(\theta_j - \mu)$$

and setting this to zero implies that $\tan \hat{\mu} = \frac{S_1}{C_1}$ where $S_1 = \sum_{j=1}^{n} \sin \theta_j$ and $C_1 = \sum_{j=1}^{n} \cos \theta_j$ so that $\hat{\mu}$ coincides with the sample circular mean direction. Note however that calculation of the maximum likelihood estimate of κ must be done numerically, see e.g. Mardia (1972).

Another possibility is to use method of moments techniques, e.g. approximating the theoretical circular mean and mean resultant length using their sample counterparts.

Example 14. *A cosine distribution was fitted to the data of Example 17 both by maximum likelihood and by method of moments approaches. In the first case, the parameter maximum likelihood estimates were $\hat{\mu} = 3.008$ and $\hat{\rho} = 0.326$ and in the second case, the sample circular mean, $\bar{\theta} = 2.974$, and mean resultant length, $\bar{R}_1 = 0.337$ were used as parameter estimates. Figure 1.9 shows circular and linear plots of the fitted densities.*

Both fitted densities are quite close to the true, generating cosine density.

Although the method of moments approach may often be more straightforward in practice, it will not always produce reasonable results. For example, in the case of the cosine distribution, if $\bar{R}_1 > 0.5$, then the method of moments approach cannot be applied directly as the mean resultant length, $\rho \leq 0.5$ for this distribution.

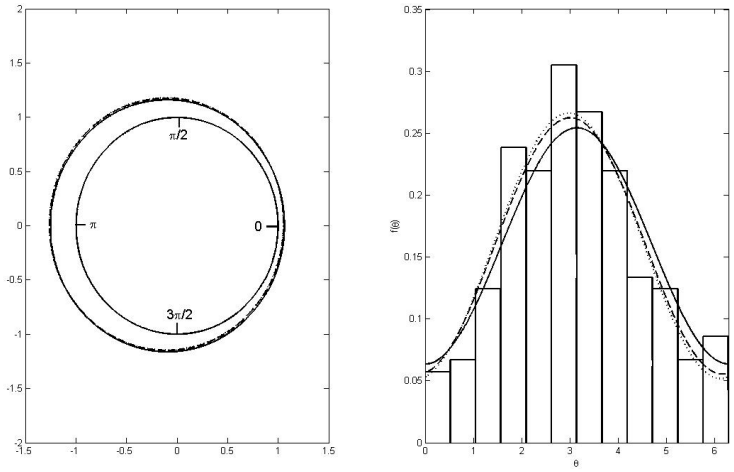

Figure 1.9: True (solid line) and fitted maximum likelihood (dashed line) and method of moment (dotted line) based density estimates

Bayesian approaches to inference for directional data have also been developed. However, from the Bayesian point of view, the main problem is the lack of conjugate priors. Most research has been focused on the von Mises distribution, see e.g. Guttorp et al (1988). For an example of Bayesian analysis for circular models other than the von Mises, see e.g. Coles (1998).

1.4 Semi-parametric and non-parametric approaches

Most of the models that have been developed for circular and other directional data are symmetric and unimodal, but in many cases, directional data may not share these characteristics. Therefore, semi-parametric or non-parametric techniques are often applied.

1.4.1 Semi-parametric models

A natural, semi-parametric extension of the basic models is to consider the use of mixtures of distributions, of form:

$$f_r(\theta) = \sum_{j=1}^{r} w_j f(\theta|\boldsymbol{\nu}_j) \tag{1.10}$$

where $f(\cdot|\boldsymbol{\nu})$ is a density with parameters $\boldsymbol{\nu}$ and the w_j are weights satisfying $0 < w_j \leq 1$ for $j = 1, \ldots, r$ and $\sum_{j=1}^{r} w_j = 1$. Clearly, as more mixture components are included, then

distributions of many different forms can be well approximated.

In the case of circular data, mixtures of von Mises distributions have been considered in a number of articles, see e.g. Stephens (1969), Mardia et al (1975), Spurr et al (1991), where a good comparison of the different methods that can be used to fit such models is given and Mooney et al (2003). More general mixture models where the restriction that the weights are non-negative is dropped have also been developed. In particular, Fernández Durán (2004) developed a model based on mixtures of sin and cosine terms as in Example 15. Generalizations of this approach to the case of circular-circular and spherical data modeling are considered in Chapter 4 of this thesis.

Example 15 (Non-negative trigonometric polynomials)**.** *Fernández-Durán (2004) proposed using distributions of form:*

$$f(\theta) = \frac{1}{2\pi} + \frac{1}{\pi}\sum_{j=1}^{k}\left(a_j \cos(j\theta) + b_j \sin(j\theta)\right)$$

where, to ensure that the distribution is non-negative, it is required that there exist coefficients $c_j \in \mathbb{C}$ *for* $j = 0, \ldots, k$ *such that* $\sum_{j=0}^{k} |c_j|^2 = \frac{1}{2\pi}$ *and* $a_j - ib_j = 2\sum_{l=0}^{k-j} c_{l+j}\bar{c}_l$ *for* $j = 1, \ldots, k$.

1.4.2 Non-parametric approaches

Non-parametric approaches to modelling directional data thus far have typically been based on kernel density estimation techniques, see e.g. Wand et al (1995). The usual kernel approach corresponds to using a reference density, K, which defines the form of smoothing and a smoothness parameter, $h > 0$ which governs the smoothness of the fitted density. For example, given a sample of circular data, $\theta_1, \ldots, \theta_n$, then a kernel density estimate of $f(\theta)$ can be constructed as

$$\hat{f}(\theta) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{\theta - \theta_i}{h}\right) \tag{1.11}$$

where K is a suitable circular density function, for example a von Mises density (Bai et al, 1988) or a quartic kernel (Fisher, 1989).

Multivariate directional data have also been modelled using kernel density methods. For example, spherical data modelling has been considered in e.g. Hall et al (1987), Bowman et al (1997) and Klemelä (2000) and kernel based approaches to circular-circular data are examined in Di Marzio et al (2011).

Typically, for kernel density methods, the choice of kernel function is not too important to the fitted distribution, whereas the choice of bandwidth is much more critical. Low values of the bandwidth lead to spiky density estimates and high values lead to overly smoothed estimates. Obviously, as the sample size increases, then the bandwidth should be decreased and it is well known that optimal bandwidth estimators of order $n^{-1/5}$ lead to density estimators $\hat{f}$ which converge at the optimal rate $O\left(n^{-4/5}\right)$. Optimal bandwidth choice for the circular case is considered in Taylor (2008) and Hall et al (1987) give some results for the spherical case.

Finally, we should note that there has been little work on alternative, non-kernel based nonparametric approaches to directional data. For an exception, see Baldi et al (2009) who model spherical data using spherical wavelets or needlets. In Chapters 2 and 3 of these thesis, we introduce an alternative approach.

Chapter 2

Circular Bernstein polynomial distributions

Bernstein polynomials have been widely studied in numerical analysis as an approach to interpolating functions defined on a closed interval. For a complete review see e.g. Lorentz (1986) and Devore et al (1993).

From the statistical point of view, starting from Vitale (1975), Bernstein polynomials have been used to approximate distribution functions defined on closed intervals. In the context of circular data analysis, an important problem with the standard approach is that the derived density estimators are not circular in that they do not guarantee continuity at the origin. The main objective of this chapter is to show that the Vitale approach can be generalized to the case of circular random variables by the use of a simple correction which preserves the convergence properties of the standard, linear estimator. The rest of this chapter is structured as follows.

In Section 2.1, we present a brief introduction to Bernstein polynomials and show how they can be used for approximating a function defined on a closed interval. Then, in Section 2.2, we review the use of Bernstein polynomials in the approximation of distribution and density functions. In Section 2.3, we define the circular Bernstein polynomial distribution and derive the conditions which must satisfy to be a well defined circular distribution. In Section 2.4, we derive the trigonometric moments and their recursive formula showing that the trigonometric moments have a closed form. In Section 2.5, we analyze how to fit the circular Bernstein polynomial distribution, deriving the necessary conditions for satisfying the circular properties. In Section 2.6, we illustrate various aspects of the fitting of the Bernstein polynomial distribution with both a theoretical and a practical example. In Section

2.7, we then present some parametric extensions of our basic model and in Section 2.8, we finish with some conclusions.

2.1 Bernstein polynomials

Bernstein polynomials were introduced by Bernstein (1912) in the context of providing a simple proof of Weierstrass' approximation theorem. They have been widely applied in numerical analysis as interpolating polynomials for continuous functions defined on a closed interval such as $[0, 1]$.

A Bernstein polynomial on the $[0, 1]$ interval can be defined via Definitions 9 and 10 as follows.

Definition 9. *The $k + 1$ Bernstein basis polynomials of degree k are defined as,*

$$P_{j,k}(x) = \binom{k}{j} x^j (1-x)^{k-j}, \text{ for } j = 0, \ldots, k \text{ and } 0 \leq x \leq 1.$$

In the context of probability, the basis functions are simply binomial probability mass functions. Thus, $P_{j,k}(x)$ is simply the probability of seeing j heads in k independent tosses of a coin with the probability of heads in a single toss equal to x.

Definition 10. *A linear combination of Bernstein basis polynomials, that is,*

$$B_k(x) = \sum_{j=0}^{k} b_{j,k} P_{j,k}(x),$$

is called a Bernstein polynomial of order k. The coefficients, $b_{j,k}$, are called Bernstein coefficients or Bézier coefficients.

Clearly, $B_k(x)$ can be expressed as a polynomial in x of degree less than or equal to k. More importantly, the Bernstein polynomial is a linear combination of Bernstein basis polynomials and Weierstrass' approximation theorem, see e.g. Lorentz (1986), guarantees that this basis is dense in the space of polynomials of degree k and form a partition of the unit interval, then we can use a Bernstein polynomial with a large enough value of k to approximate any continuous function on $[0, 1]$ to an arbitrary precision. The Bernstein polynomial approximation of order k for a function $g(x)$ is defined below.

Definition 11. *Let $g(\cdot)$ be a function defined on the closed interval $[0,1]$. Then the k'th order Bernstein polynomial approximation of $g(x)$ is given by:*

$$B_k^g(x) = \sum_{j=0}^{k} g\left(\frac{j}{k}\right) P_{j,k}(x).$$

It is well known that letting $k \to \infty$, the Bernstein polynomial approximation converges uniformly to the true function g. The exact convergence rate depends on the type of continuity of the function g. In particular, if g is absolutely continuous with finite second derivative at x, then,

$$B_k^g(x) - g(x) = \frac{g''(x)x(1-x)}{2k} + \frac{\delta(k)}{k}, \tag{2.1}$$

where $\delta(k) \to 0$ as $k \to \infty$. See e.g. Lorentz (1986) for more details.

Suppose now that g is a Lebesgue integrable function and let $G(x) = \int_0^x g(u)\,du$. Then, replacing $g\left(\frac{j}{k}\right)$ in the Bernstein polynomial approximation by, for instance by an integral mean of $g(x)$ over a small interval around the point j/k, we may use the derivative of the Bernstein polynomial to obtain a good estimate of $g(x)$ as follows.

Let $B_k^G(x)$ be the k'th order Bernstein polynomial approximation to G as in Definition 11. Then, calculating the derivative of $B_k^G(x)$, we obtain the so-called Kantorovich polynomials,

$$\begin{aligned} K_{k-1}^g(x) &= \frac{d}{dx} B_k^G(x) \\ &= \sum_{j=0}^{k-1} \binom{k}{j} x^j (1-x)^{k-1-j} k \int_{j/k}^{(j+1)/k} g(t)dt \qquad (2.2) \\ &= \sum_{j=1}^{k} \binom{k}{j} x^{j-1} (1-x)^{k-j} j \int_{(j-1)/k}^{j/k} g(t)dt, \qquad (2.3) \end{aligned}$$

which can be used as approximations to $g(x)$, see e.g. Theorem 2.1.1 of Lorentz (1986).

In particular, it is known that

$$\lim_{k\to\infty} k\left(K_{k-1}^g(x) - g(x)\right) = \frac{1}{2}\left((1-2x)g'(x) + x(1-x)g''(x)\right). \tag{2.4}$$

Finally, we should also comment that it is straightforward to generalize Bernstein polynomials to approximate functions on an arbitrary interval, say $[a,b]$ via a standard, linear transformation.

2.2 Bernstein polynomial distributions

In this section, we describe how Bernstein (Kantorovich) polynomials can be used to approximate a probability distribution (density) function defined on a closed interval.

Firstly, it is important to make some remarks about Kantorovich polynomials which are the key to the construction of the Bernstein density function.

The first remark is on the type of integrability, that is Lebesgue integrability, which allows us to construct different functions, for example, the empirical distribution. The second remark corresponds to the statistical perspective of the Kantorovich polynomials. As we can see in (2.2) the k term and in (2.3) the j term complete the integration constant of a beta distribution. Given that the sum of the integral terms for any partition of the $[0,1]$ interval is 1, this implies that the Kantorovich polynomial of a density function or function based on data is a mixture of beta distributions.

For example, consider a histogram of relative frequencies. Using the Kantorovich polynomials, the resulting density is a smoothed version of the histogram, see e.g. Vitale (1975).

Now, we can formally define the Bernstein polynomial approximation for a given distribution function defined on the $[0,1]$ interval.

Definition 12. *Let X be a random variable with support $[0,1]$ and continuous distribution function $F_X(\cdot)$. Then the Bernstein polynomial distribution function of order k is defined to be:*

$$B_k(x) = \sum_{j=0}^{k} F_X\left(\frac{j}{k}\right) \binom{k}{j} x^j (1-x)^{k-j}, \quad \text{for } 0 \leq x \leq 1 \text{ and } k \in \mathbb{N}. \tag{2.5}$$

Using the properties of the Bernstein polynomial described in the previous section, $B_k(x)$ converges uniformly to $F_X(x)$ as k goes to infinity. Differentiating, the associated Bernstein density function is given by

$$b_k(x) = \sum_{j=1}^{k} \left(F_X\left(\frac{j}{k}\right) - F_X\left(\frac{j-1}{k}\right) \right) \beta(x \mid j, k-j+1), \tag{2.6}$$

where $\beta(\cdot \mid a, b)$ is a beta density function:

$$\beta(x \mid a, b) = \frac{1}{B(a,b)} x^{a-1} (1-x)^{b-1}, \tag{2.7}$$

and $B(a,b) = (a-1)!(b-1)!/(a+b-1)!$, for $a, b \in \mathbb{N}$, is the beta function. The errors in these approximations can be derived from (2.1) and (2.4) respectively.

2.2.1 Non-parametric inference via Bernstein polynomials

Assume that a sample of data, $\mathbf{x} = (x_1, \ldots, x_n)$, are generated from an unknown, continuous distribution, F, with associated density function, f with support $[0, 1]$. Then a non-parametric approach to inference via Bernstein polynomial distributions was first developed by Vitale (1975). He suggests approximating the density, $f(x)$, by substituting the distribution function by the associated empirical distribution function in (2.6). This leads to the Bernstein density estimator,

$$\hat{b}_k(x) = \sum_{j=1}^{k} \left(\hat{F}_X\left(\frac{j}{k}\right) - \hat{F}_X\left(\frac{j-1}{k}\right) \right) \beta(x \mid j, k-j+1). \tag{2.8}$$

As demonstrated in Vitale (1975), recalling that the empirical cumulative distribution function converges uniformly (in n) to the true distribution and from (2.4), it can be seen that as long as the true density function, f_X, is twice differentiable, then the bias of this estimator is,

$$E\left[\hat{b}_k(x) - f_X(x)\right] = \frac{1}{2k}\left[(1-2x)f'_X(x) + x(1-x)f''_X(x)\right] + o\left(\frac{1}{k}\right), \tag{2.9}$$

where, in particular the estimator is free of boundary bias as the o term is uniform in $[0, 1]$. Furthermore, the variance of the estimator is,

$$V\left[\hat{b}_k(x)\right] = \frac{\sqrt{k}}{n}\frac{f(x)}{2\sqrt{\pi x(1-x)}} + o\left(\frac{\sqrt{k}}{n}\right), \tag{2.10}$$

Vitale (1975) further demonstrates that the optimal choice of k with respect to mean squared error is to set $k \to n^{2/5}$ and shows that in this case, the estimator converges at a rate $n^{-4/5}$, which is the same rate of convergence as the alternative, kernel based estimators.

Various extensions of Vitale's estimator have been proposed. For example, in a classical setting, Babu et al. (2002) approximate a distribution function defined on the unit interval using the Bernstein polynomial and then, they derive an estimator of the density function given the smoothness of the Bernstein polynomial. Also, Kakizawa (2004) proposes a kernel estimation method using the Bernstein polynomial modifying the kernel to avoid the boundary bias and the negative values of the kernel. Other relevant articles are Gawronski et al (1981), Gawronski (1985), Stadtmüller (1983), Leblanc (2009, 2010) and Kakizawa (2010).

In the Bayesian context, Petrone (1999a,b) developed an approach to inference via Bern-

stein polynomials based on the use of Dirichlet process prior distributions. The consistency of the Bernstein prior distribution was studied in Petrone et al (2002) and convergence rates of the posterior distribution were later studied in Ghoshal (2001) and Walker et al (2007).

2.3 The circular Bernstein polynomial distribution

In order to define a circular Bernstein polynomial density, it is necessary to keep in mind the properties of any circular density function and how to compute the distribution function of a circular variable, see Chapter 1.

Consider a circular random variable Θ with support $[0, 2\pi)$ and distribution function, $F_\Theta^\nu(\cdot)$, defined with respect to an origin ν. Then it is clearly straightforward to define a k'th order Bernstein polynomial approximation to the circular density function as,

$$f_k^\nu(mod(\theta+\nu, 2\pi)) = \frac{1}{2\pi}\sum_{j=1}^{k}\left(F_\Theta^\nu\left(\frac{2\pi j}{k}\right) - F_\Theta^\nu\left(\frac{2\pi(j-1)}{k}\right)\right)\beta\left(\frac{\theta}{2\pi}\middle| j, k-j+1\right). \tag{2.11}$$

However, for this to be a strictly continuous, circular density, it is necessary that,

$$F_\Theta^\nu\left(\frac{2\pi}{k}\right) = 1 - F_\Theta^\nu\left(\frac{2\pi(k-1)}{k}\right). \tag{2.12}$$

The following theorem guarantees the existence of at least one origin satisfying (2.12).

Theorem 1. *Let f be a density function for a continuous, circular random variable. Then there exists at least one point $\nu = \nu_k \in [0, 2\pi)$ such that for any $k \in \mathbb{N}$,*

$$\int_\nu^{\nu+\frac{2\pi}{k}} f(\theta)\,d\theta = \int_{\nu-\frac{2\pi}{k}}^{\nu} f(\theta)\,d\theta.$$

Proof. Define $G(\nu) = \int_\nu^{\nu+\frac{2\pi}{k}} f(\theta)\,d\theta - \int_{\nu-\frac{2\pi}{k}}^{\nu} f(\theta)\,d\theta$. If there exist two points, $0 \le \nu_1 \ne \nu_2 < 2\pi$ such that $G(\nu_1) \le 0$ and $G(\nu_2) \ge 0$, then by Bolzano's intermediate value theorem, there exists at least one point, $0 \le \nu_0 < 2\pi$ such that $G(\nu_0) = 0$. Otherwise, suppose that $G(\nu)$ is always positive. Then, we have,

$$\int_{\nu-\frac{2\pi}{k}}^{\nu} f(\theta)\,d\theta < \int_\nu^{\nu+\frac{2\pi}{k}} f(\theta)\,d\theta < \int_{\nu+2\pi-\frac{2\pi}{k}}^{\nu+2\pi} f(\theta)\,d\theta$$

which is impossible, as, due to the periodicity of f, we have that

$$\int_{\nu-\frac{2\pi}{k}}^{\nu} f(\theta)\,d\theta = \int_{\nu+2\pi-\frac{2\pi}{k}}^{\nu+2\pi} f(\theta)\,d\theta.$$

Similarly, G cannot always be negative and so the theorem is proved. □

The following example illustrates the convergence of the circular Bernstein polynomial density approximation as k increases in the case that the true distribution is a cosine distribution.

Example 16. *Consider a cosine distribution with $\mu = \pi$. Then this distribution is symmetric about $\nu = 0$ and therefore, taking this point as the origin, we have,*

$$F_{\Theta}^{0}(\theta) = \frac{1}{2\pi}\left\{\theta + \rho\sin(\theta-\pi)\right\}, \quad \textit{for } 0 \le \theta < 2\pi,$$

and therefore, for any given k, the circular Bernstein polynomial density can be computed explicitly.

Figure 2.1 shows the circular Bernstein polynomial density for $k = 50$ (dashed line), 100(dotted line) and 300(dash dot line) and the underlying cosine density function (solid line) with $\mu = \pi$ and $\rho = 0.3$. We have seen before that the circular Bernstein polynomial converges uniformly to f.

As k increases, the Bernstein polynomial approximation gets closer to the true density although we can see that convergence is slower in the neighbourhood of π.

Typically, there may be more than one origin satisfying the conditions of Theorem 1 and from now on we shall write $\boldsymbol{\nu}_k$ to represent the set of valid origins for a circular Bernstein polynomial approximation of order k.

Example 17. *Assume that Θ has a cosine distribution with circular mean μ, so that,*

$$f(\theta) = \frac{1}{2\pi}\left\{1 + 2\rho\cos(\theta-\mu)\right\}, \quad \textit{for } 0 \le \theta < 2\pi.$$

Then, the density function is symmetric about $\theta = \mu$ and $\theta = mod(\mu+\pi, 2\pi)$ so these are both valid origins for a circular Bernstein polynomial density approximation.

In fact, for any circular distribution which is symmetric about some point μ, then there are valid origins at μ and $\mod(\mu+\pi, 2\pi)$. In some cases, the number of valid origins can even be uncountable.

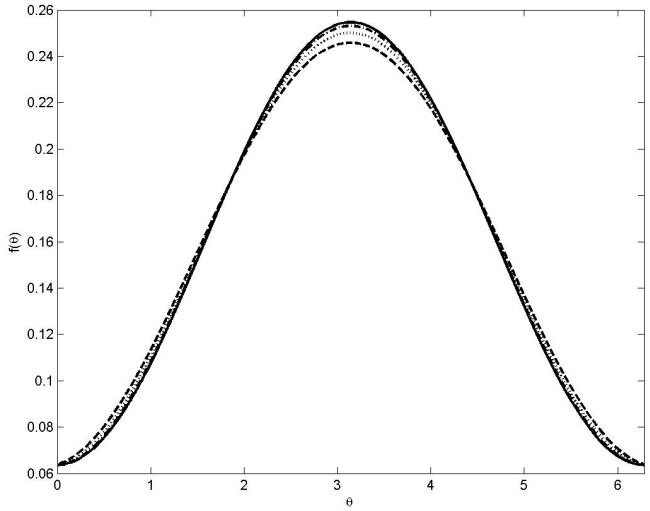

Figure 2.1: Approximation of the cosine distribution (solid line) with a circular Bernstein polynomial with degrees 50 (dashed line), 100 (dotted line) and 300 (dash dot line).

Example 18. *If Θ is uniformly distributed, then for any k, the set of valid origins is $\boldsymbol{\nu}_k = [0, 2\pi)$.*

It is also important to note that, for given k, the circular Bernstein polynomial approximations will typically be different with respect to the different origins.

Example 19. *Continuing from Example 16, it is clear that the circular Bernstein polynomial distribution based on the origin π can also be computed explicitly for any k. Figure 2.2 displays the Bernstein polynomial distributions for $k = 50$ about both the origins $\nu = 0$ and $\nu = \pi$. The approximation with origin π is closer to the underlying distribution in the neighbourhood of π but further away in the neighbourhood of 0.*

Finally, we should consider the convergence properties of circular Bernstein polynomials. Note first that an origin, ν_k, for the k'th order Bernstein polynomial satisfies

$$D(\nu_k|k) = F(\nu_k - 2\pi/k) + F(\nu_k + 2\pi/k) - 2F(\nu_k) = 0$$

and, letting $k \to \infty$, we have $\lim_{k\to\infty} k^2 D(\nu_k|k) = f' = 0$. This implies, that for large k, the valid origins approach the turning points of the density, f. In particular, we can see that in the case of symmetric distributions, then we always have at least two valid origins which do

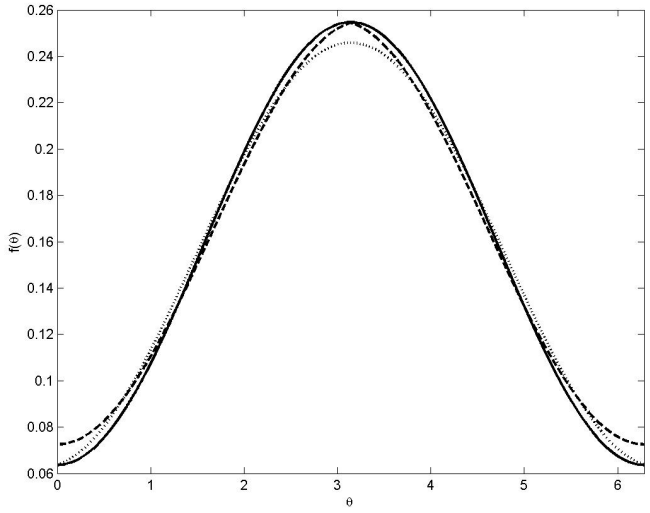

Figure 2.2: Approximation of the cosine distribution (solid line) with circular Bernstein polynomials of order 50 with origins 0 (dashed line) and π (dotted line).

not change for any k. Also, we have that

$$|f_k^{\nu_k}(\theta) - f(\theta)| \leq |f_k^{\nu_k}(\theta) - f_k^{\nu}(\theta)| + |f_k^{\nu}(\theta) - f(\theta)|,$$

where ν is such a turning point. Letting $k \to \infty$ and noting that the second term in this expression is the bias of the Bernstein polynomial approximation about ν, we see that the bias and convergence rate of this approximation will typically be different to the rate of the standard Bernstein polynomial approximation. More research needs to be done to assess the convergence conditions of the circular Bernstein polynomial distribution.

2.4 Trigonometric moments of the circular Bernstein polynomial distribution

Explicit formulae are available for the trigonometric moments of a circular Bernstein polynomial distribution. These can be derived from the following theorem.

Theorem 2. *The p'th order trigonometric moments of a circular Bernstein polynomial*

distribution are given by

$$\boldsymbol{\mu}'_p = \sum_{j=1}^{k} w_j \left(E[\cos 2\pi p \mathcal{B}_j] + iE[\sin 2\pi p \mathcal{B}_j] \right),$$

where

$$\omega_j = F_\Theta^\nu \left(\frac{j}{k} \right) - F_\Theta^\nu \left(\frac{j-1}{k} \right),$$

and where $\mathcal{B}_j$ is a beta random variable with density function $\beta(\cdot \mid j, k-j+1)$ as defined in (2.7) such that

$$\begin{aligned}
E[\cos(2\pi p \mathcal{B}_j)] &= \tfrac{1}{B(j,k-j+1)} \textstyle\sum_{r=0}^{k-j} (-1)^r \begin{pmatrix} k-j \\ r \end{pmatrix} I_p(j+r-1) && (2.13)\\
&= \tfrac{1}{B(j,k-j+1)} \textstyle\sum_{r=0}^{j-1} (-1)^r \begin{pmatrix} j-1 \\ r \end{pmatrix} I_p(k-j+r), && (2.14)\\
E[\sin(2\pi p \mathcal{B}_j)] &= \tfrac{1}{B(j,k-j+1)} \textstyle\sum_{r=0}^{k-j} (-1)^r \begin{pmatrix} k-j \\ r \end{pmatrix} J_p(j+r-1) && (2.15)\\
&= \tfrac{1}{B(j,k-j+1)} \textstyle\sum_{r=0}^{j-1} (-1)^{r+1} \begin{pmatrix} j-1 \\ r \end{pmatrix} J_p(k-j+r), && (2.16)
\end{aligned}$$

where $I_p(0) = J_p(0) = I_p(1) = 0$, $J_p(1) = -\frac{1}{2\pi p}$ and for $C = 2, 3, 4, \ldots$,

$$\begin{aligned}
I_p(C) &= \textstyle\sum_{c=1}^{\lfloor \frac{C}{2} \rfloor} (-1)^{c-1} \frac{C!}{(C-2c+1)!} \frac{1}{(2\pi p)^{2c}} && (2.17)\\
J_p(C) &= \textstyle\sum_{c=1}^{\lfloor \frac{C+1}{2} \rfloor} (-1)^{c} \frac{C!}{(C-2c+2)!} \frac{1}{(2\pi p)^{2c-1}}. && (2.18)
\end{aligned}$$

Proof. Let $I_p(j) = \int_0^1 \cos(2\pi p x) x^j \, dx$ and $J_p(j) = \int_0^1 \cos(2\pi p x) x^j \, dx$ respectively for $j = 0, 1, 2, \ldots$. Now,

$$\begin{aligned}
E[\cos(2\pi p\mathcal{B}_j)] &= \int_0^1 \cos(2\pi px)\frac{1}{B(j,k-j+1)}x^{j-1}(1-x)^{k-j}\,dx \\
&= \frac{1}{B(j,k-j+1)}\sum_{r=0}^{k-j}(-1)^r\left(\begin{array}{c} k-j \\ r \end{array}\right)\int_0^1 \cos(2\pi px)x^{j-1+r}\,dx \\
&= \int_0^1 \cos(2\pi py)\frac{1}{B(j,k-j+1)}(1-y)^{j-1}y^{k-j}\,dy \\
&= \frac{1}{B(j,k-j+1)}\sum_{r=0}^{j-1}(-1)^r\left(\begin{array}{c} j-1 \\ r \end{array}\right)\int_0^1 \cos(2\pi py)y^{k-j+r}\,dy,
\end{aligned}$$

which gives the expressions for (2.13) and (2.14). In a similar way, the expressions (2.15) and (2.16) can be derived, recalling that $\sin(2\pi-\theta) = -\sin(\theta)$. Thus, it only remains to demonstrate formulas (2.17) and (2.18) by induction.

Now observe that,

$$\begin{aligned}
I_p(0) &= \int_0^1 \cos(2\pi px)\,dx = 0, \\
J_p(0) &= \int_0^1 \sin(2\pi px)\,dx = 0, \\
I_p(1) &= \int_0^1 x\cos(2\pi px)\,dx = \frac{1}{2\pi p}\left[x\sin(2\pi px)\right]_0^1 - \frac{1}{2\pi p}\int_0^1 \sin(2\pi px)\,dx = 0, \\
J_p(1) &= \int_0^1 x\sin(2\pi px)\,dx = -\frac{1}{2\pi p}\left[x\cos(2\pi px)\right]_0^1 + \frac{1}{2\pi p}\int_0^1 \cos(2\pi px)\,dx = -\frac{1}{2\pi p}
\end{aligned}$$

Now consider $I_p(C)$. For $C \geq 2$,

$$\begin{aligned}
I_p(C) &= \int_0^1 x^C\cos(2\pi px)\,dx \\
&= \frac{1}{2\pi p}\left[x^C\sin(2\pi px)\right]_0^1 - \frac{C}{2\pi p}\int_0^1 x^{C-1}\sin(2\pi px)\,dx \\
&= -\frac{C}{2\pi p}\int_0^1 x^{C-1}\sin(2\pi px)\,dx \\
&= \frac{C}{(2\pi p)^2}\left[x^{C-1}\cos(2\pi px)\right]_0^1 - \frac{C(C-1)}{(2\pi p)^2}\int_0^1 x^{C-2}\cos(2\pi px)\,dx \\
&= \frac{C}{(2\pi p)^2} - \frac{C(C-1)}{(2\pi p)^2}I_p(C-2), \qquad (2.19)
\end{aligned}$$

and therefore $I_p(2) = \frac{2}{(2\pi p)^2}$ and $I_p(3) = \frac{3}{(2\pi p)^2}$ which satisfy (2.17). Assume now that the formula is valid for $c = 2, \ldots, C$. Then,

$$\begin{aligned}
I_p(C+2) &= \frac{C+2}{(2\pi p)^2} - \frac{(C+2)(C+1)}{(2\pi p)^2} I_p(C) \quad \text{from (2.19)} \\
&= \frac{C+2}{(2\pi p)^2} - \sum_{c=1}^{\lfloor \frac{C}{2} \rfloor} (-1)^{c-1} \frac{C!}{(C-2c+1)!} \frac{1}{(2\pi p)^{2c}} \quad \text{from the induction assumption} \\
&= \frac{C+2}{(2\pi p)^2} + \sum_{c=1}^{\lfloor \frac{C}{2} \rfloor} (-1)^{c+1-1} \frac{(C+2)!}{(C+2-2(c+1)+1)!} \frac{1}{(2\pi p)^{2(c+1)}} \\
&= \frac{C+2}{(2\pi p)^2} + \sum_{c=2}^{\lfloor \frac{C+2}{2} \rfloor} (-1)^{c-1} \frac{(C+2)!}{(C+2-2c+1)!} \frac{1}{(2\pi p)^{2c}} \\
&= \sum_{c=1}^{\lfloor \frac{C+2}{2} \rfloor} (-1)^{c-1} \frac{(C+2)!}{(C+2-2c+1)!} \frac{1}{(2\pi p)^{2c}},
\end{aligned}$$

which demonstrates (2.17).

Equally, we have the recurrence relation,

$$J_p(C) = -\frac{1}{2\pi p} - \frac{C(C-1)}{(2\pi p)^2} J_p(C-2) \tag{2.20}$$

which implies that $J_p(2) = -\frac{1}{2\pi p}$ and $J_p(3) = -\frac{1}{2\pi p} + \frac{3!}{(2\pi p)^3}$ which satisfy (2.18). Assuming the formula is valid for $c = 2, \ldots, C$ then,

$$\begin{aligned}
J_p(C+2) &= -\frac{1}{2\pi p} - \frac{(C+2)(C+1)}{(2\pi p)^2} J_p(C) \quad \text{from (2.20)} \\
&= -\frac{1}{2\pi p} - \frac{(C+2)(C+1)}{(2\pi p)^2} \sum_{c=1}^{\lfloor \frac{C+1}{2} \rfloor} (-1)^{c} \frac{C!}{(C-2c+2)!} \frac{1}{(2\pi p)^{2c-1}} \\
&\quad \text{from the induction assumption} \\
&= -\frac{1}{2\pi p} - \sum_{c=1}^{\lfloor \frac{C+1}{2} \rfloor} \frac{(C+2)!}{(C-2c+2)!} \frac{1}{(2\pi p)^{2c+1}} \\
&= \sum_{c=1}^{\lfloor \frac{C+2+1}{2} \rfloor} (-1)^{c} \frac{(C+2)!}{(C+2-2c+2)!} \frac{1}{(2\pi p)^{2c-1}},
\end{aligned}$$

which demonstrates (2.18) and proves the theorem. □

From an operational point of view, in order to minimize the computational time for computing the trigonometric moments, we suggest first computing the terms $J_p(C)$ and $I_p(C)$ for $C = 0, \ldots, k-1$ for the required values of p using (2.18) and (2.17) and storing the results. Then, the remaining formulae can be computed from (2.13) and (2.15). This is due to that the trigonometric moments of the circular Bernstein polynomial have a closed form.

2.5 Estimation for the circular Bernstein polynomial distribution

Suppose now that we have a sample of n data, $\{\theta_1, \ldots, \theta_n\}$, observed from a continuous, unknown, circular density, $f_\Theta(\theta)$. Then, for any origin, ν, a standard, linear Vitale estimator could be defined by,

$$\hat{b}_k^\nu(\theta+\nu) = \frac{1}{2\pi}\sum_{j=1}^{k}\left(\hat{F}_\Theta^\nu\left(\frac{2\pi j}{k}\right) - \hat{F}_\Theta^\nu\left(\frac{2\pi(j-1)}{k}\right)\right)\beta\left(\frac{\theta}{2\pi} \mid j, k-j+1\right), \qquad (2.21)$$

where $\hat{F}_\Theta^\nu(\cdot)$ is the empirical distribution function defined from ν.

However, similar to (2.12), this estimator will only be circular if

$$\hat{F}_\Theta^\nu\left(\frac{2\pi}{k}\right) = 1 - \hat{F}_\Theta^\nu\left(\frac{2\pi(k-1)}{k}\right),$$

and, in contrast to Theorem 1, it may be that there exists no origin $\nu \in [0, 2\pi)$ which leads to a circular density estimate. For example, in the case that $k = 2$, if an odd number of data are observed, then no origin, ν, satisfies $\hat{F}_\Theta^\nu(\pi) = 1/2$. Therefore, in order to produce a valid, circular, density estimator, a modification of the Vitale estimator is required. We propose the following procedure.

Firstly, for a given k, we need to select a suitable origin. To do this, define

$$d(\nu) = \hat{F}_\Theta^\nu\left(\frac{2\pi}{k}\right) + \hat{F}_\Theta^\nu\left(\frac{2\pi(k-1)}{k}\right) - 1, \qquad \text{for } 0 \le \nu < 2\pi, \qquad (2.22)$$

which measures the difference between the first and the last weights of the beta mixture density (2.21).

Now let $\hat{\boldsymbol{\nu}}_k = \arg\min |d(\nu)|$ be the set of origins which minimize the absolute distance

between the first and last weights.

Then for $\hat{\nu} \in \hat{\boldsymbol{\nu}}_k$, we can define a circular Bernstein polynomial estimator as

$$\begin{aligned}\hat{f}_k^{\hat{\nu}}(\theta) &= \frac{1}{2\pi}\left[\frac{1}{2}\left\{\hat{F}^{\hat{\nu}}\left(\frac{2\pi}{k}\right)+1-\hat{F}^{\hat{\nu}}\left(\frac{2\pi(k-1)}{k}\right)\right\}\beta\left(\frac{\theta}{2\pi}\middle|1,k\right)+\right.\\ &\quad \sum_{j=2}^{k-1}\left\{\hat{F}^{\hat{\nu}}\left(\frac{2\pi j}{k}\right)-\hat{F}^{\hat{\nu}}\left(\frac{2\pi(j-1)}{k}\right)\right\}\beta\left(\frac{\theta}{2\pi}\middle|j,k-j+1\right)+\\ &\quad \left.\frac{1}{2}\left\{\hat{F}^{\hat{\nu}}\left(\frac{2\pi}{k}\right)+1-\hat{F}^{\hat{\nu}}\left(\frac{2\pi(k-1)}{k}\right)\right\}\beta\left(\frac{\theta}{2\pi}\middle|k,1\right)\right].\end{aligned}$$

Thus, this estimator modifies the standard Bernstein polynomial estimator by averaging the first and last weights.

Firstly, we will show that for a random sample of size n from a strictly continuous distribution, then $D = \min|d(\nu)|$ is always less than or equal to $1/n$. In order to do this, we introduce the following generalization of a continuous function taken from Burgin (2010).

Definition 13. *A function $f : \mathbb{R} \to \mathbb{R}$ is called r-continuous at a point $a \in \mathbb{R}$ if $f(x)$ is defined at a and for any $\varepsilon > 0$ there is a $\delta > 0$ such that for any x with $|a - x| < \delta$, we have that $|f(x) - f(a)| < r + \varepsilon$.*

Thus, an r-continuous function is basically a non-continuous function with jumps smaller than a quantity r. The following result, taken from Burgin (2010), is Bolzano's intermediate value theorem for r-continuous functions,

Theorem 3 (Burgin 2010). *Let $f : [a, b] \to \mathbb{R}$ be an r-continuous function. If $f(a) < 0$ and $f(b) > 0$, then there is at least one point $c \in [a, b]$ such that $|f(c)| < r$.*

Now we can demonstrate the existence of an origin $\hat{\nu}$ such that $|d(\hat{\nu})| \leq 1/n$.

Theorem 4. *Let $\{\theta_1, \ldots, \theta_n\}$ be a random sample from a strictly continuous, circular, random variable, Θ, with density function $f_\Theta(\cdot)$. Then, for $k = 2, 3, \ldots$, there exists at least one point $\hat{\nu} \in [0, 2\pi)$ such that $|d(\hat{\nu})| \leq 1/n$.*

Proof. Write $d(\nu) = d_1(\nu) - d_2(\nu)$ such that $d_1(\nu) = \hat{F}_\Theta^\nu\left(\frac{2\pi}{k}\right)$ and $d_2(\nu) = 1 - \hat{F}_\Theta^\nu\left(\frac{2\pi(k-1)}{k}\right)$. For a sample from a strictly continuous density, then $d_1(\nu)$ and $d_2(\nu)$ are both step functions with steps of size $1/n$ and therefore, $d(\nu)$ is a step function taking steps of size $1/n$ or $2/n$ so that $d(\nu)$ is $2/n$-continuous.

Now, assume that there exist two points, $0 \leq \nu_1 \neq \nu_2 < 2\pi$ such that $d(\nu_1) < 0$ and $d(\nu_2) > 0$. Then from Theorem 3 there exists at least one point, $0 \leq \hat{\nu} < 2\pi$, such that $|d(\hat{\nu})| < 2/n$ and recalling that $\hat{d}(\nu)$ is a step function, we have that $|d(\hat{\nu})| \leq 1/n$.

On the contrary, suppose that $d(\nu)$ is always positive. Observe that $d_1(\nu) = \frac{1}{n}\sum_{i=1}^{n} I_{(\nu,\nu+2\pi/k]}(\theta_i)$ and $d_2(\nu) = \frac{1}{n}\sum_{i=1}^{n} I_{(\nu-2\pi/k,\nu]}(\theta_i)$ where $I(\theta)_{(a,b]}$ is an indicator function taking the value 1 if $\theta \in (a, b]$ and 0 otherwise. Then, we have that,

$$\sum\nolimits_{i=1}^{n} I_{\left(\nu-\frac{2\pi}{k},\nu\right]}(\theta_i) < \sum\nolimits_{i=1}^{n} I_{\left(\nu,\nu+\frac{2\pi}{k}\right]}(\theta_i) < \sum\nolimits_{i=1}^{n} I_{\left(\nu+2\pi-\frac{2\pi}{k},\nu+2\pi\right]}(\theta_i),$$

which is impossible, as we have that,

$$\sum\nolimits_{i=1}^{n} I_{\left(\nu-\frac{2\pi}{k},\nu\right]}(\theta_i) = \sum\nolimits_{i=1}^{n} I_{\left(\nu+2\pi-\frac{2\pi}{k},\nu+2\pi\right]}(\theta_i).$$

Equally, $d(\nu)$ cannot always be negative and so the theorem is proved. □

Finally, we can bound the difference between the circular and uncorrected Vitale estimators based on the origin $\hat{\nu}$.

Theorem 5. *Let $\{\theta_1, \ldots, \theta_n\}$ be a random sample from a strictly continuous, circular, random variable, Θ, with density function $f_\Theta(\cdot)$ and let $\hat{\nu} \in \hat{\boldsymbol{\nu}}_k$. Then we have that,*

$$\left|\hat{f}_k^{\hat{\nu}}(\theta) - \hat{b}_k^{\hat{\nu}}(\theta)\right| \leq \frac{k}{4\pi n}.$$

Proof. We have,

$$\hat{f}_k^{\hat{\nu}}(\theta) - \hat{b}_k^{\hat{\nu}}(\theta) = \frac{1}{4\pi}\left\{\hat{F}^{\hat{\nu}}\left(\frac{2\pi}{k}\right) + \hat{F}^{\hat{\nu}}\left(\frac{2\pi(k-1)}{k}\right) - 1\right\}\left(\beta\left(\frac{\theta}{2\pi}\middle| k, 1\right) - \beta\left(\frac{\theta}{2\pi}\middle| 1, k\right)\right),$$

and the maximum difference between the two estimators occurs when either $\theta = 0$ or $\theta = 2\pi$, so that,

$$\max\left|\hat{f}_k^{\hat{\nu}}(\theta) - \hat{b}_k^{\hat{\nu}}(\theta)\right| \leq \frac{k}{4\pi}\left|\hat{F}^{\hat{\nu}}\left(\frac{2\pi}{k}\right) + \hat{F}^{\hat{\nu}}\left(\frac{2\pi(k-1)}{k}\right) - 1\right| \leq \frac{k}{4\pi n},$$

from Theorem 4. □

Thus, as $n \to \infty$, for $k < \sqrt{n}$, the circular Bernstein polynomial approximation preserves the properties of the uncorrected approximation based on the same, random origin.

A number of comments are in order here. Firstly note that the previous theorems are valid for any origin, $\hat{\nu}$, in $\hat{\boldsymbol{\nu}}_k$ and that, typically, the cardinality of $\hat{\boldsymbol{\nu}}_k$ will be greater than 1. In order to define a unique origin, we suggest selecting the origin, $\hat{\nu} \in \hat{\boldsymbol{\nu}}_k$, which maximizes the (pseudo) log-likelihood estimate, $\sum_{i=1}^{n} \log \hat{f}_k^{\hat{\nu}}(\theta_i)$. Of course, many alternative criteria could also be used, e.g. maximizing the p value for some goodness of fit test.

Secondly, in order to choose k in practice, there are various possibilities. Here we propose increasing k until standard goodness of fit tests for circular densities such as those proposed in Watson (1961) or Kuiper (1960) do not reject the fitted density at a given significance level (e.g. 5%), k can be increased until the fitted model is accepted. Other techniques such as least squares cross validation, see e.g. Wand et al (1995) could also be used.

2.5.1 Tied data

If the underlying density function is not strictly continuous, then there may be multiple observations at the same point, and in this case, it is not necessarily true that an origin satisfying Theorem 1 exists. In practice, data will typically be observed rounded to a certain number of decimal places when the same problem also occurs. In such cases, we propose either adding a small, random jitter to the data to obviate the problem, or otherwise simply selecting the origin which minimizes the distance $|d(\nu)|$. In this latter case, as long as there is a true origin satisfying the conditions of Theorem 1, it is straightforward to show that, for fixed k, the difference between the Vitale and circular estimators converges almost surely to zero.

Theorem 6. *Let $\theta_1, \ldots, \theta_n$ be a sample from a circular distribution, f, such that an origin satisfying the conditions of Theorem 1 exists. Let $\hat{\nu}$ be chosen to minimize the distance $|d(\nu)|$. Then*

$$\left| \hat{f}_k^{\hat{\nu}}(\theta) - \hat{b}_k^{\hat{\nu}}(\theta) \right| \to 0 \quad \textit{almost surely.}$$

Proof. If $\hat{\nu}$ is an origin chosen to minimize $|d(\nu)|$, then

$$\begin{aligned} \left| \hat{f}_k^{\hat{\nu}}(\theta) - \hat{b}_k^{\hat{\nu}}(\theta) \right| &\leq \frac{k}{4\pi} \left| \hat{F}^{\hat{\nu}} \left(\frac{2\pi}{k} \right) + \hat{F}^{\hat{\nu}} \left(\frac{2\pi(k-1)}{k} \right) - 1 \right| \\ &\leq \frac{k}{4\pi} \left| \hat{F}^{\nu} \left(\frac{2\pi}{k} \right) + \hat{F}^{\nu} \left(\frac{2\pi(k-1)}{k} \right) - 1 \right| \end{aligned}$$

by construction, for any origin, $\nu \in \boldsymbol{\nu}_k$, satisfying the conditions of Theorem 1.

Now, for any n and $\nu \in \boldsymbol{\nu}_k$, using the triangle inequality, we have:

$$\begin{aligned}\left|\hat{f}_k^{\hat{\nu}}(\theta) - \hat{b}_k^{\hat{\nu}}(\theta)\right| &\leq \frac{k}{4\pi}\left\{\left|\hat{F}^{\nu}\left(\frac{2\pi}{k}\right) - F^{\nu}\left(\frac{2\pi}{k}\right)\right| + \right. \\ &\quad \left|\hat{F}^{\nu}\left(\frac{2\pi(k-1)}{k}\right) - F^{\nu}\left(\frac{2\pi(k-1)}{k}\right)\right| + \\ &\quad \left.\left|F^{\nu}\left(\frac{2\pi}{k}\right) + F^{\nu}\left(\frac{2\pi(k-1)}{k}\right) - 1\right|\right\} \\ &= \frac{k}{4\pi}\left\{\left|\hat{F}^{\nu}\left(\frac{2\pi}{k}\right) - F^{\nu}\left(\frac{2\pi}{k}\right)\right| + \right. \\ &\quad \left.\left|\hat{F}^{\nu}\left(\frac{2\pi(k-1)}{k}\right) - F^{\nu}\left(\frac{2\pi(k-1)}{k}\right)\right|\right\}\end{aligned}$$

which goes to zero almost surely as $n \to \infty$ by the Glivenko Cantelli theorem. □

2.6 Illustrations

In this section, we illustrate the use of the circular Bernstein polynomials from a theoretical point of view and from a practical point of view.

2.6.1 A theoretical example: the cosine data

Here, we fitted the circular Bernstein polynomial density of degree 15 to the data of Example 17. The fitted density is given in Figure 2.3, along with the true generating density of the data. Note that this was not rejected by the Watson or Kuiper tests at the 5% level.

It can be seen that the fitted Bernstein polynomial density estimator provides a comparable fit to the maximum likelihood and method of moments estimators of Example 14.

2.6.2 Behaviour of the origins $\hat{\nu}$

Clearly, before the sample is taken, for a given k, the origin $\hat{\nu}$ is a random variable. In standard statistical theory, it is typical that the distribution of a parameter estimator tends to cluster around a single point as the sample size increases. However, we have seen that for many circular distributions, there are multiple origins which lead to a valid circular Bernstein polynomial approximation. Therefore, we might speculate that the distribution of $\hat{\nu}$ will tend to centre on the set of valid origins. The following example explores the case of the cosine

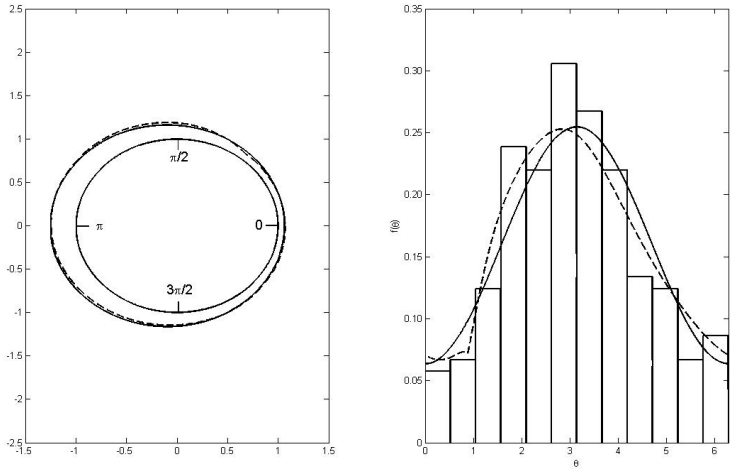

Figure 2.3: True (solid line) and fitted (dashed line) densities for the cosine data

distribution.

Example 20. *We simulated 1000 samples of size 100, 1000 and 10000 respectively from the cosine distribution with parameters* $\mu = \pi$ *and* $\rho = 0.3$*. Then, for each sample, we have calculated the origin for the Bernstein polynomial of order* $k = 10$*. Figure 2.4 gives circular kernel density estimates of the density of* $\hat{\nu}$ *for the three different sample sizes.*

It can be seen that for $n = 100$*, there is a large variation in the distribution of* $\hat{\nu}$ *whereas as* n *increases, then the sampled origins cluster into a bimodal distribution around the true origins* $\nu = 0$ *and* $\nu = \pi$ *with the main peak at* 0*.*

It is important to note that further research needs to be done in order to assess the convergence properties of the origins. In particular, it is necessary to study the conditions under which $\min_{\nu \in \boldsymbol{\nu}_k} |mod(\hat{\nu}_k - \nu, 2\pi)| \to 0$. Further research is underway on this problem.

2.6.3 A practical example: the Chicago crimes data

Here we consider data obtained from which `www.chicagocrime.org` which correspond to the twenty four hour clock times of 1297 crimes perpetrated in Chicago on May 11$^{\text{th}}$, 2007, obtained from `www.chicagocrime.org`.

A Bernstein polynomial density approximation of order $k = 20$ was fitted to these data. Also, for comparison a kernel density estimate based on a von Mises kernel and using an optimal bandwidth as in Taylor (2008) was fitted. Figure 2.5 shows a rose plot and a

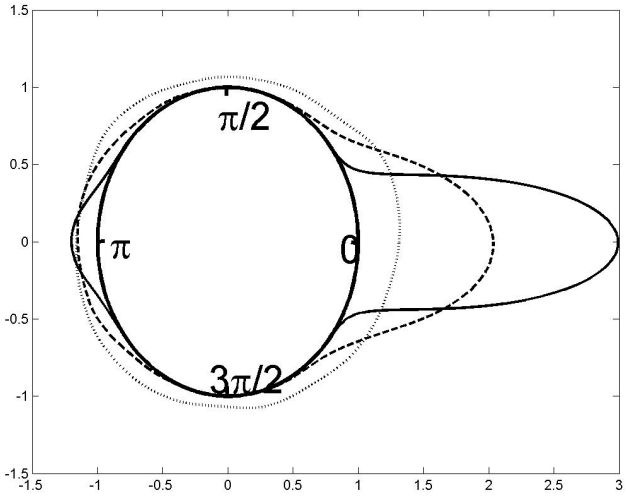

Figure 2.4: Kernel density estimates of the density of $\hat{\nu}$ for samples of size 100 (dotted line), 1000 (dashed line) and 10000 (solid line).

histogram of the data and the corresponding fitted densities. The data are plotted so that the origin is set to the fitted origin time, that is 03:59 hours.

The fitted and empirical mean direction and 17:45 hours and 17:42 hours and the fitted and empirical mean circular resultant lengths are 0.200 and 0.189 respectively. Thus, there is good agreement between the fitted and empirical moments.

Both the Kuiper and Watson tests were used to test the goodness of fit of the Bernstein polynomial density and in both cases, the model was not rejected at a 5% level.

2.7 Extensions

Various extensions of the approach introduced in this chapter can be considered. Firstly, it is important to explore in more detail the conditions for convergence of the circular Bernstein polynomial approximation as $k \to \infty$ and for convergence of the proposed estimator in this case.

Another possibility is to consider parametric models based on mixtures of beta distributions with both continuous and discrete parameters. These approaches are briefly reviewed below.

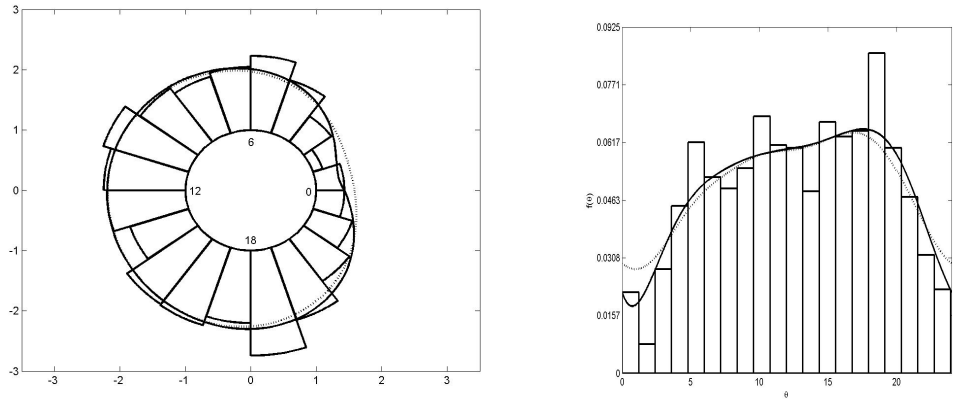

Figure 2.5: Roseplot (left hand side) and histogram of the Chicago crime data with fitted Bernstein polynomial (solid line) and kernel density (dashed line) approximations.

2.7.1 Mixtures of scaled, shifted, beta distributions

As a parametric alternative to the previous approach, we could consider the direct modeling of a circular variable by a mixture of scaled, shifted beta distributions.

Definition 14. *Let Θ be a random variable with support $[0, 2\pi)$. Then we shall say that Θ has a scaled, shifted, beta (SSB) distribution with parameters $0 \leq \nu < 2\pi$ and $\alpha, \beta \in \mathbb{N}$, denoted by $\Theta \sim \mathcal{SSB}(\nu, \alpha, \beta)$ if*

$$f(\theta) = \begin{cases} \frac{1}{(2\pi)^{\alpha+\beta-1}B(\alpha,\beta)}(2\pi + \theta - \nu)^{\alpha-1}(\nu - \theta)^{\beta-1} & \text{if} \quad 0 \leq \theta < \nu \\ \frac{1}{(2\pi)^{\alpha+\beta-1}B(\alpha,\beta)}(\theta - \nu)^{\alpha-1}(2\pi + \nu - \theta)^{\beta-1} & \text{if} \quad \nu \leq \theta < 2\pi \end{cases} \tag{2.23}$$

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

Clearly, we can represent the distribution of Θ as the distribution of a shifted, scaled, beta distributed random variable, so that $\theta = \mod(2\pi X + \nu, 2\pi)$ where X is a beta distributed random variable, as in (2.7).

A natural parametric generalization of the circular Bernstein polynomial model is to assume that Θ follows a mixture of SSB distributions as follows.

Definition 15. *A circular random variable Θ has a SSB mixture distribution with parameters $\mathbf{w} = (w_1, \ldots, w_k)$, where $w_i \geq 0$ for $i = 1, \ldots, k$ and $\sum_{i=1}^{k} w_i = 1$, $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_k)$, where $0 \leq \nu_i < 2\pi$, for $i = 1, \ldots, k$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)$ where $\alpha_i, \beta_i \in \mathbb{N}$ for*

$i = 1, \ldots, k$ *if*

$$f(\theta) = \sum_{j=1}^{k} w_j f(\theta|\nu_j, \alpha_j, \beta_j)$$

where $f(\cdot|\nu, \alpha, \beta)$ *is the* $\mathcal{SSB}(\nu, \alpha, \beta)$ *density function as in (2.23).*

Clearly, the trigonometric moments of the SSB mixture model can be derived explicitly from the results in Section 2.4.

Assume now that we observe a sample of data, $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_n\}$ to which we wish to fit a mixture of SSB distributions. Firstly, it is important to note that it is straightforward to calculate the maximum likelihood parameter estimates when the data are fitted by a single SSB distribution. For a given origin, ν, the maximum likelihood estimates of the beta parameters, α and β, may be calculated by first rescaling the data onto the data to the $[0, 1)$ interval and then calculating the unconstrained maximum likelihood estimates using, for example the `MATLAB` routine, `betafit`. Then the constrained maxima may be found by simply checking the likelihoods at the pairs of integer valued α and β around the unconstrained maxima. Finally, the global maxima may be estimated by using the same procedure over a grid of values of ν.

For fitting a k dimensional mixture distribution, then various approaches are possible. One possibility is to use the EM algorithm, see e.g. Dempster et al (1977), McLachlan et al (2000) but we have found that this procedure is very sensitive to the choice of initial values and thus, we prefer to use an alternative procedure based on direct likelihood maximization.

Firstly, when $k = 2$, or $k = 3$, it is possible to directly maximize the likelihood function in terms of the component weights, $\mathbf{w}$, by searching over a grid of values of $\boldsymbol{\nu}, \boldsymbol{\alpha}, \boldsymbol{\beta}$ although for higher dimensional mixtures, this procedure is too computationally intensive. In these cases, we use an iterative algorithm which automatically finds a (local) maximum. Given that we wish to fit a mixture of $k > 3$ terms, we assume that we have calculated the maximum likelihood estimates for the $k - 1$ dimensional mixture and proceed by setting the initial values, $(\nu_i, \alpha_i, \beta_i)$ for $i = 1, \ldots, k - 1$ equal to the maximum likelihood estimates for this model and then estimating all component weights and the parameters, $(\nu_k, \alpha_k, \beta_k)$ of the k'th mixture component by direct likelihood maximization using a grid search procedure. Then, the parameters of components $2, \ldots, k$ are fixed and the weights and $(\nu_1, \alpha_1, \beta_1)$ are re-estimated by direct likelihood maximization, the parameters of components, $1, 3, \ldots, k$ are fixed and the weights and $(\nu_2, \alpha_2, \beta_2)$ are re-estimated and so on until convergence. We find in practice that this algorithm usually converges to the global maximum likelihood,

due to the fact that the different beta components are well distinguished from each other due to the discreteness of the beta parameters. Thus, the main mixture terms are usually well identified in low dimensional mixtures and the effect of adding an extra component is essentially that of fitting a new component which well fits a small amount of the observed data.

Finally, in order to select the number of terms to include in the SSB mixture model, we use the Akaike information criterion (AIC), see Akaike (1974).

We have estimated a mixture of two SSB distributions for the Chicago crime data. Figure 2.6 show the data and the fitted density for this data. Note that the origin in this figure is the time $00:00$ hours.

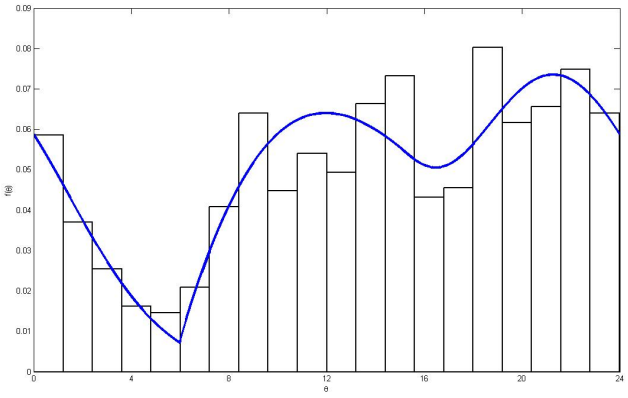

Figure 2.6: Histogram of the Chicago crime data with fitted mixture of SSB distribution with integers parameters

If we eliminate the constraint that the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ of the SSB must be integer vectors, we add more flexibility to the model. However, the closed form expressions for the trigonometric moments are lost.

As commented previously, for the model to be circular, the continuity property at the origin must be maintained. This implies that the SSB mixture model must be composed of a mixture of (a uniform density and) beta densities with parameters $\alpha_i, \beta_i \geq 1$.

We have estimated a mixture of two SSB distributions for the crime data using the first model. Figure 2.7 show the data and the fitted density for this data. As we can observe one of the terms has a parameter which is practically 1. There also appears to be little difference in the fit as compared to the previous model.

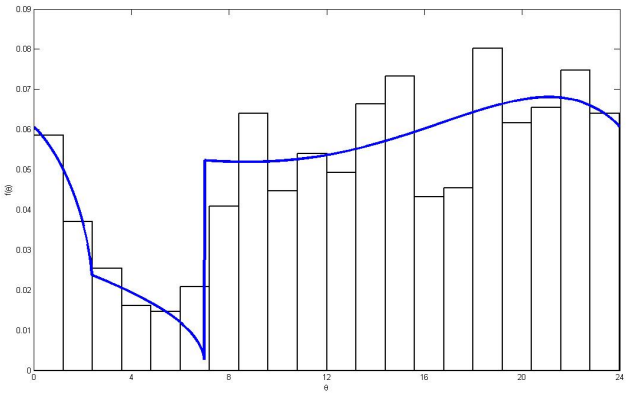

Figure 2.7: Histogram of the Chicago crime data with fitted mixture of SSB distribution with continuous parameters

2.7.2 Other extensions

Further extensions of the circular Bernstein polynomial model can be considered. Firstly, one possibility is to consider using kernel based estimators to substitute the empirical distribution function in the circular Bernstein polynomial estimator, as in Kakizawa (2004, 2010) for the linear case. Another possibility is to consider Bayesian estimation of the circular Bernstein polynomial model, based on generalizing the approach of Petrone (1999a,b).

Another extension is to extend the model to higher dimensions. For example, the d-torus is the direct product of d circles. Then, we must define the properties and theorems for these surfaces. However in the most of these cases the data used must be simulated and the resulting model is only interesting from a theoretical point of view.

From a practical point of view, we could also consider the sphere and adapt the procedure to this surface we must be careful with the Lebesgue measure for the sphere ($\sin\theta d\theta d\phi$) to define correctly the spherical Bernstein polynomial.

Finally, we could consider the estimation of bivariate circular circular and circular linear distributions with given marginal distributions via the use of Bernstein copulas, see e.g. Sancetta et al (2004), in which at least one marginal is estimated by a circular Bernstein polynomial model. This approach is implemented in the following chapter.

2.8 Conclusions

The main conclusions of this chapter can be summarized as follows. The Bernstein polynomial and the Kantorovich polynomials can approximate any continuous function defined on a closed interval with an arbitrary precision and in particular they can be easily adapted to the estimation of circular distributions.

In particular it is demonstrated that for a strictly continuous circular random variable, then a Bernstein-Kantorovich density approximation can always be derived. It is also shown that given a sample of circular data from a strictly continuous variable, then a simple, circular generalization of the Vitale (1975) density estimator can be derived and that this preserves the convergence properties of the standard Vitale estimator.

Finally, we have shown that various parametric extensions of the circular Bernstein polynomial model, can be developed and have illustrated them on a real data set.

Chapter 3

Bernstein copulas and applications to directional data

In the previous chapter we developed an approach to the non-parametric modeling of univariate circular data. Here, we shall extend this to the case of bivariate data. In particular, we shall consider two possibilities, that is the modeling of the bivariate distribution of two circular variables and the modeling of the joint distribution of a circular and a linear variable.

There have been a number of parametric approaches to analyzing distributions of this type, see e.g. Batschelet (1981), Mardia et al (1978), Johnson et al (1978), Kagan et al. (1973), Mardia (1975) and Jammalamadaka et al (1988).

To develop a nonparametric approach to these types of data, we combine nonparametric estimates of the marginal densities of the circular and linear components with the use of a modification of a class of nonparametric copulas, known as empirical Bernstein copulas, to model the dependence structure.

Copulas can be viewed as a tool for constructing a multivariate distribution in such a way that the individual marginal distributions can be defined separately from their dependence structure making use of Sklar's theorem (Sklar, 1973). In particular, empirical Bernstein copulas (see e.g. Sancetta et al., 2004), are based on multivariate Bernstein polinomials and have been used as smooth approximations of unknown copulas, as they are defined on the unit hypercube.

Along this chapter we show how circular-linear and circular-circular distributions can be constructed via empirical Bernstein copulas showing that for this type of data the generated distribution satisfy certain continuity constraints to be well behaved bivariate distributions. Following the orientation of this thesis, we have designed an algorithm which is

non-parametric in all its stages. As we describe below, we use a non-parametric estimation of the distribution function as the linear distribution and the circular Bernstein polynomial for the circular distribution. Observe that for the empirical Bernstein copula model the dependence structure estimated with the copula function depends on the data observed. Similar to the circular Bernstein polynomial, when we use data, we must to impose that the weights in the boundary cells are equal. Then, for constructing a well behaved bivariate distribution we must correct the weights. We will show that these corrections preserves the uniformness of the marginal distributions.

This chapter is organized as follows. In Section 3.1 we introduce the bivariate Bernstein polynomials. In Section 3.2 we show the concept of copula and summarize briefly the types of existing copulas. In Section 3.3 we show that the generalization of Theorem 1 in two dimension does not hold and motivate the use of copulas for constructing the bivariate distribution where both marginal distributions are circular. We define the circular-circular model based on copulas and describe how to estimate it in a non-parametric way. In Section 3.4 we define the circular-linear model based on copulas and describe how to estimate it in a non-parametric way. In Section 3.5 we illustrate these two models with environmental data and finalize this chapter showing the conclusions and possible extensions in Section 3.6.

3.1 Bivariate Bernstein polynomials

Here, we generalize the results of the previous chapter for the univariate Bernstein polynomials to the bivariate case, which is the basis for much of the work presented in the following sections.

Definition 16. *Let $k_1, k_2 \in \mathbb{N}$ and let $g(\cdot,\cdot)$ be a function of two variables, each with support $[0,1]$. Then the polynomials:*

$$B^g_{k_1,k_2}(x_1,x_2) = \sum_{j_1=0}^{k_1}\sum_{j_2=0}^{k_2} g\left(\frac{j_1}{k_1},\frac{j_2}{k_2}\right) P_{j_1,k_1}(x_1)P_{j_2,k_2}(x_2), \tag{3.1}$$

where $P_{j,k}(x)$ represents a Bernstein basis polynomial as defined in (9), are called the bivariate Bernstein polynomials of g.

As in the univariate case, see Section 2.1, it is well known that for a continuous function, g, then the Bernstein polynomials converge uniformly to g as $k_1, k_2 \to \infty$. For more details see e.g. Lorentz (1986).

3.1.1 Bivariate Bernstein polynomial distributions

As in the previous chapter, once we have introduced the bivariate Bernstein polynomial, we now illustrate how it can be used for modelling bivariate distribution functions. Clearly, a Bernstein polynomial approximation for a bivariate distribution function of two variables, each with support on the unit interval can be derived from Definition 16.

Definition 17. *Let* $\mathbf{X} = (X_1, X_2)$ *be a bivariate random variable with support in* $[0,1]^2$ *and continuous bivariate distribution function* $F_{\mathbf{X}}(\cdot)$*. Then the bivariate Bernstein polynomial approximation to* F *of order* $\mathbf{k} = (k_1, k_2)$ *is defined as:*

$$B_{\mathbf{k}}(x_1, x_2) = \sum_{j_1=0}^{k_1} \sum_{j_2=0}^{k_2} F_X\left(\frac{j_1}{k_1}, \frac{j_2}{k_2}\right) \prod_{i=1}^{2} \binom{k_i}{j_i} x_i^{j_i} (1-x_i)^{k_i - j_i}. \tag{3.2}$$

An approximation to the density function can then be derived by differentiating.

Definition 18. *For a bivariate distribution function,* $F_{\mathbf{X}}(\cdot,\cdot)$ *as in Definition 17, then the associated bivariate Bernstein density function of order* $\mathbf{k} = (k_1, k_2)$ *is:*

$$b_{\mathbf{k}}(x_1, x_2) = \sum_{j_1=1}^{k_1} \sum_{j_2=1}^{k_2} w_{j_1,j_2} \prod_{i=1}^{2} \beta(x_i | j_i, k_i - j_i + 1) \tag{3.3}$$

where, $\beta(\cdot|\cdot)$ *is the beta density function as given in (2.7) and,*

$$w_{j_1,j_2} = F_X\left(\frac{j_1-1}{k_1}, \frac{j_2-1}{k_2}\right) + F_X\left(\frac{j_1}{k_1}, \frac{j_2}{k_2}\right) - F_X\left(\frac{j_1-1}{k_1}, \frac{j_2}{k_2}\right) - F_X\left(\frac{j_1}{k_1}, \frac{j_2-1}{k_2}\right). \tag{3.4}$$

Observe that the weight, w_{j_1,j_2}, in (3.4) is the probability that an observation belongs to the region $\left(\frac{j_1-1}{k_1}, \frac{j_1}{k_1}\right] \times \left(\frac{j_2-1}{k_2}, \frac{j_2}{k_2}\right]$. Therefore, the bivariate Bernstein density is a smooth approximation of the true bivariate density function.

Finally, it is straightforward to see by integrating out that the marginal distribution of each variable is a univariate Bernstein polynomial as in (2.6). Thus, we have:

$$\begin{aligned} b_{k_1}(x_1) &= \int_0^1 b_{k_1,k_2}(x_1, x_2)\, dx_2 \\ b_{k_2}(x_2) &= \int_0^1 b_{k_1,k_2}(x_1, x_2)\, dx_1. \end{aligned} \tag{3.5}$$

Note finally that given sample (linear) data, a bivariate Bernstein polynomial density

estimate can be constructed either by substitution of the empirical cumulative distribution function for the theoretical distribution function in (3.3) or by using smoothed, kernel based estimators as in Tenbusch (1994).

3.2 Copulas

Often, we wish to model the joint distribution of two, or more, random variables in the case that the marginal distributions are known. The copula function provides the way to construct such a distribution.

Definition 19. *A bivariate copula is a joint distribution on* $[0,1]^2$ *such that both marginal distributions are uniform on the interval* $[0,1]$*. Specifically,* $C : [0,1]^2 \to [0,1]$ *is a bivariate copula if:*

1. *For all* $u \in [0,1]$*, then* $C(0,u) = C(u,0) = 0$*,*
2. *For all* $u \in [0,1]$*, then* $C(1,u) = C(u,1) = u$*,*
3. *For every* $u_1, u_2, v_1, v_2 \in [0,1]$*, where* $u_1 \leq u_2$ *and* $v_2 \leq v_2$ *then*

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0.$$

The final property in the above definition is the bivariate equivalent to the condition that a univariate distribution function is non-decreasing. Sklar's (1973) theorem, which is given below, shows that a given joint distribution function can be defined in terms of a copula function and the marginals.

Theorem 7 (Sklar 1973). *Let* F *be the distribution function of a bivariate random variable,* $\mathbf{X} = (X_1, X_2)$ *with marginal distribution functions* F_1 *and* F_2*. Then there exists a copula,* C*, such that*

$$F(x_1, x_2) = C(F_1(x_1), F_2(x_2)).$$

Conversely, for any univariate distribution functions, F_1 *and* F_2 *and any copula,* C*, the function* F *is the distribution function of a bivariate random variable with marginal distributions* F_1 *and* F_2*. Furthermore, if* F_1 *and* F_2 *are continuous, then* C *is unique.*

For continuous random variables X_1 and X_2 with marginal distribution functions F_1 and F_2 and copula function C, then clearly, the joint density function can be derived from

Theorem 7 as:

$$f\left(x_1, x_2\right) = c\left(F_1\left(x_1\right), F_2\left(x_2\right)\right) \prod_{i=1}^{2} f_i\left(x_i\right), \tag{3.6}$$

where the $f_1(\cdot)$ and $f_2(\cdot)$ are the marginal density functions of X_1 and X_2 respectively and c is the density function of the copula which is given by

$$c(u, v) = \frac{\partial^2}{\partial u \partial v} C(u, v). \tag{3.7}$$

Equally, the copula density function can be derived from the joint and marginal density and distribution functions as

$$c\left(u, v\right) = \frac{f\left(F_1^{-1}\left(u\right), F_2^{-1}\left(v\right)\right)}{f_1\left(F_1^{-1}\left(u\right) f_2\left(F_2^{-1}\left(v\right)\right)\right)}. \tag{3.8}$$

Various parametric forms for copulas have been considered as in the following examples.

Example 21. *The simplest copula is the product copula, $C(u, v) = uv$, when we have $F(x_1, x_2) = F_1(x_1)F_2(x_2)$. Therefore, in the case of continuous variables, this copula characterizes independence between X_1 and X_2.*

Two of the most popular approaches are the Gaussian and Archimedean copulas.

Example 22. *The Gaussian copula function is defined by*

$$C_\rho(u, v) = \Phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v))$$

where $u, v \in [0, 1]$ and $\Phi(\cdot)$ denotes the standard, normal cumulative distribution function and $\Phi_\rho(\cdot, \cdot)$ is the distribution function of a standard, bivariate normal random variable with correlation ρ.

Using this copula, one can define a bivariate distribution function with the same dependence structure as the standard bivariate normal distribution function but with non-normal marginal distributions.

Example 23. *An Archimedean copula is defined by*

$$C(u, v) = \Psi\left(\Psi^{-1}(u) + \Psi^{-1}(v)\right)$$

for a generator function, Ψ and its generalized inverse, Ψ^{-1}.

Many other parametric families of copulas have been developed. See e.g. Nelsen (1999) for a good review.

3.2.1 The empirical and empirical Bernstein copulas

When data are generated from an unknown underlying distribution function, the empirical data distribution can be transformed into an empirical copula so that the marginal distributions become uniform, see e.g. Nelsen (1999). Formally, given a sample, $\{(x_{11}, x_{21}), \ldots, (x_{1n}, x_{2n})\}$ from the joint distribution of (X_1, X_2), then the empirical copula function can be defined as follows.

Firstly, we transform the original data into a sample $(u_i, v_i) = (\hat{F}_{X_1}(x_{1i}), \hat{F}_{X_2}(x_{2i}))$, for $i = 1, \ldots, n$, where $\hat{F}_X(\cdot)$ and $\hat{F}_Y(\cdot)$ are consistent estimators of the true marginal distributions, $F_X(\cdot)$ and $F_Y(\cdot)$, respectively. The transformed values now form a sample on $[0, 1]^2$. Then, the empirical copula distribution function is defined as

$$\hat{C}_n(u, v) = \frac{1}{n} \sum_{i=1}^{n} I(u_i \leq u, v_i \leq v), \tag{3.9}$$

for $1 \leq i \leq n, 1 \leq j \leq n$.

Note that by construction, the empirical copula is a valid distribution function. However, it has marginals which are uniform only asymptotically as $n \to \infty$ and therefore is a valid copula only asymptotically. Clearly, the empirical copula is not a smooth function. A smoothed version can be obtained via the Bernstein polynomial approximation (see e.g. Sancetta and Satchell, 2004) as follows.

Given a sample, (u_i, v_i) for $i = 1, \ldots, n$, calculated by transforming the original data as above, then using Theorem 7, there exists a copula which can be approximated with the empirical Bernstein copula of order $\mathbf{k} = (k_1, k_2)$, which is defined as:

$$\hat{C}_B(u, v) = \frac{1}{n} \sum_{j_1=0}^{k_1} \sum_{j_2=0}^{k_2} \sum_{i=1}^{n} I\left(u_i \leq \frac{j_1}{k_1}, v_i \leq \frac{j_2}{k_2}\right) \binom{k_1}{j_1} u^{j_1} (1-u)^{k_1-j_1} \binom{k_2}{j_2} v^{j_2} (1-v)^{k_2-j_2}. \tag{3.10}$$

Clearly, using (3.6), the corresponding empirical Bernstein copula density is given by,

$$\hat{c}_B(u, v) = \sum_{j_1=1}^{k_1} \sum_{j_2=1}^{k_2} p_{j_1 j_2} \beta(u \mid j_1, k_1 - j_1 + 1) \beta(v \mid j_2, k_2 - j_2 + 1)$$

where

$$p_{j_1 j_2} = I\left(\frac{j_1 - 1}{k_1} < u_i \leq \frac{j_1}{k_1}, \frac{j_2 - 1}{k_2} < v_i \leq \frac{j_2}{k_2}\right). \tag{3.11}$$

As with the empirical copula, the empirical Bernstein copula is a copula in the asymptotic sense since,

$$\lim_{n\to\infty} \sum_{j_1=1}^{k_1} p_{j_1 j} = \frac{1}{k_2}, \qquad \text{for } j = 1, \ldots, k_2 \tag{3.12}$$

and

$$\lim_{n\to\infty} \sum_{j_2=1}^{k_2} p_{j j_2} = \frac{1}{k_1}, \qquad \text{for } j = 1, \ldots, k_1. \tag{3.13}$$

Further properties are examined in Sancetta et al (2004), Pfeifer et al. (2009) and Bouezmarni et al. (2010). In particular, using the properties of Bernstein polynomials, Sancetta et al (2004) demonstrate that in the case $k_1 = k_2 = k$, then the bias of the empirical Bernstein copula is $d/k + o(1/k)$ for some finite constant d and give the optimal choices for k under mean squared error loss.

In the following sections, we show how the Bernstein copula can be adapted for use in the case of circular-circular and circular-linear distributions.

3.3 Circular-circular distributions constructed via empirical Bernstein copulas

In this section, we shall develop an approach to Bernstein polynomial based approximation of the density of a bivariate, continuous, circular-circular, random variable, $\boldsymbol{\Theta} = (\Theta_1, \Theta_2)$. A natural possibility would be to try to see if bivariate Bernstein polynomial distributions, as outlined in Section 3.1.1 could be applied for some origin $\boldsymbol{\nu} = (\nu_1, \nu_2)$ such that the corresponding Bernstein polynomial approximation is properly circular.

3.3.1 Direct modeling via bivariate Bernstein polynomial distributions

Suppose that f is a strictly continuous, bivariate circular-circular density function. Then, given an origin, $\boldsymbol{\nu}$, from Definition 18, a Bernstein polynomial approximation of f of order

$\mathbf{k} = (k_1, k_2)$ is given by:

$$b_{\mathbf{k}}^{\boldsymbol{\nu}}(\theta_1, \theta_2) = \sum_{j_1=1}^{k_1} \sum_{j_2=1}^{k_2} w_{j_1,j_2} \prod_{i=1}^{2} \beta\left(\tfrac{\theta_i}{2\pi} \mid j_i, k_i - j_i + 1\right)$$

where

$$\begin{aligned} w_{j_1,j_2} = {} & F_{\boldsymbol{\Theta}}^{\boldsymbol{\nu}}\left(2\pi\frac{j_1-1}{k_1}, 2\pi\frac{j_2-1}{k_2}\right) + F_{\boldsymbol{\Theta}}^{\boldsymbol{\nu}}\left(2\pi\frac{j_1}{k_1}, 2\pi\frac{j_2}{k_2}\right) \\ & - F_{\boldsymbol{\Theta}}^{\boldsymbol{\nu}}\left(2\pi\frac{j_1-1}{k_1}, 2\pi\frac{j_2}{k_2}\right) - F_{\boldsymbol{\Theta}}^{\boldsymbol{\nu}}\left(2\pi\frac{j_1}{k_1}, 2\pi\frac{j_2-1}{k_2}\right). \end{aligned}$$

However, as in the univariate case, not all origins will give a valid, strictly continuous density approximation. $\boldsymbol{\nu}$ is a valid circular origin for a Bernstein polynomial distribution of order $\mathbf{k}$ if: $w_{1,j} = w_{k_1,j}$ for $j = 1, \ldots, k_2$ and $w_{j,1} = w_{j,k_2}$ for $j = 1, \ldots, k_1$. Note that in particular, we have $w_{1,1} = w_{1,k_2} = w_{k_1,1} = w_{k_1,k_2}$.

Unfortunately, it is easy to find counterexamples of strictly continuous distributions where no valid origin for the Bernstein polynomial approximation exists as the following.

Example 24. *Consider a bivariate cosine density defined by*

$$f(\theta_1, \theta_2) = \frac{1}{4\pi^2}\left(1 + 2\rho\cos(\theta_1 - \theta_2)\right), \qquad \textit{for } \theta_1, \theta_2 \in [0, 2\pi).$$

Then the corresponding distribution function with respect to the origin $(0, 0)$ *is*

$$F(\theta_1, \theta_2) = \frac{\theta_1\theta_2}{4\pi^2} + \frac{\rho}{2\pi^2}\left[\cos(\theta_1 - \theta_2) + 1 - \cos\theta_1 - \cos\theta_2\right], \qquad \textit{for } \theta_1, \theta_2 \in [0, 2\pi).$$

Given the origin, $\boldsymbol{\nu}$, then we can calculate the weights:

$$
\begin{aligned}
w_{11} &= \frac{1}{k_1 k_2} + \frac{\rho}{2\pi^2}\left(\cos\left(\nu_1-\nu_2\right) + \cos\left(\nu_1+\frac{2\pi}{k_1}-\nu_2-\frac{2\pi}{k_2}\right)\right.\\
&\quad \left. - \cos\left(\nu_1-\nu_2-\frac{2\pi}{k_2}\right) - \cos\left(\nu_1+\frac{2\pi}{k_1}-\nu_2\right)\right)\\
w_{k_1 1} &= \frac{1}{k_1 k_2} + \frac{\rho}{2\pi^2}\left(\cos\left(\nu_1-\frac{2\pi}{k_1}-\nu_2\right) + \cos\left(\nu_1-\nu_2-\frac{2\pi}{k_2}\right)\right.\\
&\quad \left. - \cos\left(\nu_1-\frac{2\pi}{k_1}-\nu_2-\frac{2\pi}{k_2}\right) - \cos\left(\nu_1-\nu_2\right)\right)\\
w_{1 k_2} &= \frac{1}{k_1 k_2} + \frac{\rho}{2\pi^2}\left(\cos\left(\nu_1-\nu_2+\frac{2\pi}{k_2}\right) + \cos\left(\nu_1+\frac{2\pi}{k_1}-\nu_2\right)\right.\\
&\quad \left. - \cos\left(\nu_1-\nu_2\right) - \cos\left(\nu_1+\frac{2\pi}{k_1}-\nu_2+\frac{2\pi}{k_2}\right)\right)\\
w_{k_1 k_2} &= \frac{1}{k_1 k_2} + \frac{\rho}{2\pi^2}\left(\cos\left(\nu_1-\nu_2\right) + \cos\left(\nu_1-\frac{2\pi}{k_1}-\nu_2+\frac{2\pi}{k_2}\right)\right.\\
&\quad \left. - \cos\left(\nu_1-\nu_2+\frac{2\pi}{k_2}\right) - \cos\left(\nu_1-\frac{2\pi}{k_1}-\nu_2\right)\right)
\end{aligned}
$$

In a first step we analyze when the weights w_{1k_2} and $w_{k_1 1}$.

$$
\begin{aligned}
w_{k_1 1} - w_{1 k_2} &= \cos\left(\nu_1-\frac{2\pi}{k_1}-\nu_2\right) + \cos\left(\nu_1-\nu_2-\frac{2\pi}{k_2}\right) - \cos\left(\nu_1-\frac{2\pi}{k_1}-\nu_2-\frac{2\pi}{k_2}\right)\\
&- \cos\left(\nu_1-\nu_2\right) - \cos\left(\nu_1-\nu_2+\frac{2\pi}{k_2}\right) - \cos\left(\nu_1+\frac{2\pi}{k_1}-\nu_2\right)\\
&+ \cos\left(\nu_1-\nu_2\right) + \cos\left(\nu_1+\frac{2\pi}{k_1}-\nu_2+\frac{2\pi}{k_2}\right)\\
&= \cos\left(\nu_1-\nu_2\right)\left(\cos\left(-\frac{2\pi}{k_1}\right) + \cos\left(-\frac{2\pi}{k_2}\right) - \cos\left(-\frac{2\pi}{k_1}-\frac{2\pi}{k_2}\right) - \cos\left(\frac{2\pi}{k_2}\right)\right.\\
&- \left.\cos\left(+\frac{2\pi}{k_1}\right) + \cos\left(+\frac{2\pi}{k_1}+\frac{2\pi}{k_2}\right)\right) + \sin\left(\nu_1-\nu_2\right)\left(\sin\left(\frac{2\pi}{k_1}\right) + \sin\left(\frac{2\pi}{k_2}\right)\right.\\
&- \left.\sin\left(+\frac{2\pi}{k_1}+\frac{2\pi}{k_2}\right) + \sin\left(+\frac{2\pi}{k_2}\right) + \sin\left(+\frac{2\pi}{k_1}\right) - \sin\left(\frac{2\pi}{k_1}+\frac{2\pi}{k_2}\right)\right)\\
&= 2\sin\left(\nu_1-\nu_2\right)\left(\sin\left(\frac{2\pi}{k_1}\right) + \sin\left(\frac{2\pi}{k_2}\right) - \sin\left(+\frac{2\pi}{k_1}+\frac{2\pi}{k_2}\right)\right)\\
&= 2\sin\left(\nu_1-\nu_2\right)\left(\sin\left(\frac{2\pi}{k_1}\right)\left(1-\cos\frac{2\pi}{k_2}\right) + \sin\left(\frac{2\pi}{k_2}\right)\left(1-\cos\frac{2\pi}{k_1}\right)\right)
\end{aligned}
$$

The difference is zero when $k_1 = k_2 = 2$ and for $k_1, k_2 > 2$ the second term is always positive and the difference is zero only when $\nu_1 = \nu_2$ or $\nu_1 = \nu_2 + \pi$. Now we compute the difference between w_{11} and $w_{k_1 1}$ using the constraint $\nu_1 = \nu_2$ found in the previous step. If we use $\nu_1 = \nu_2 + \pi$ the difference will have the opposite sign.

$$\begin{aligned}
w_{11} - w_{k_1 1} &= \cos(0) + \cos\left(\frac{2\pi}{k_1} - \frac{2\pi}{k_2}\right) - \cos\left(-\frac{2\pi}{k_2}\right) - \cos\left(\frac{2\pi}{k_1}\right) \\
&- \cos\left(-\frac{2\pi}{k_1}\right) - \cos\left(-\frac{2\pi}{k_2}\right) + \cos\left(-\frac{2\pi}{k_1} - \frac{2\pi}{k_2}\right) + \cos(0) \\
&= 2 + \cos\left(+\frac{2\pi}{k_1} - \frac{2\pi}{k_2}\right) + \cos\left(-\frac{2\pi}{k_1} - \frac{2\pi}{k_2}\right) - 2\cos\left(-\frac{2\pi}{k_1}\right) - 2\cos\left(-\frac{2\pi}{k_2}\right) \\
&= 2 + \cos\left(+\frac{2\pi}{k_1}\right)\cos\left(\frac{2\pi}{k_2}\right) + \sin\left(+\frac{2\pi}{k_1}\right)\sin\left(\frac{2\pi}{k_2}\right) \\
&+ \cos\left(\frac{2\pi}{k_1}\right)\cos\left(\frac{2\pi}{k_2}\right) - \sin\left(\frac{2\pi}{k_1}\right)\sin\left(\frac{2\pi}{k_2}\right) - 2\cos\left(-\frac{2\pi}{k_1}\right) - 2\cos\left(-\frac{2\pi}{k_2}\right) \\
&= 2 + 2\cos\left(+\frac{2\pi}{k_1}\right)\cos\left(\frac{2\pi}{k_2}\right) - 2\cos\left(-\frac{2\pi}{k_1}\right) - 2\cos\left(-\frac{2\pi}{k_2}\right) \\
&= 2\left(\cos\left(\frac{2\pi}{k_1}\right) - 1\right)\left(\cos\left(\frac{2\pi}{k_2}\right) - 1\right) > 0 \quad \textit{for } k_1, k_2 > 1.
\end{aligned}$$

Therefore, in the case of the bivariate cosine distribution, then no valid origin for a circular Bernstein polynomial density approximation exists.

3.3.2 Modeling using the Bernstein copula

As an alternative, we propose modeling a circular-circular distribution using circular Bernstein polynomial distributions to approximate the marginals and then adapting the standard Bernstein copula to approximate the underlying copula function.

Given a set of bivariate circular-circular data, $\{(\theta_{11}, \theta_{21}), \ldots, (\theta_{1n}, \theta_{2n})\}$, we use a two-step approach to estimation of the joint density.

Firstly, we obtain estimations of the marginal circular densities, $\hat{F}_{q_1}^{\hat{\nu}_1}(\theta_1)$ and $\hat{F}_{q_2}^{\hat{\nu}_2}(\theta_2)$, using the circular Bernstein estimator introduced in the previous chapter.

Secondly, using the sample of data in the unit square given by

$$\left\{(u_{11}, u_{12}) = \left(\hat{F}_{q_1}^{\hat{\nu}_1}(\theta_{11}), \hat{F}_{q_2}^{\hat{\nu}_2}(\theta_{21})\right), \ldots, (u_{1n}, u_{2n}) = \left(\hat{F}_{q_1}^{\hat{\nu}_1}(\theta_{1n}), \hat{F}_{q_2}^{\hat{\nu}_2}(\theta_{2n})\right)\right\},$$

we obtain the corresponding empirical Bernstein copula, $\hat{c}_B(\cdot, \cdot)$, as in (3.11), where in this

case, the weights are given by

$$p_{j_1 j_2} = \frac{1}{n}\sum_{i=1}^{n} I\left(\frac{j_1-1}{k_1} < \hat{F}_{q_1}^{\hat{\nu}_1}(\theta_{1i}) \leq \frac{j_1}{k_1}, \frac{j_2-1}{k_2} < \hat{F}_{q_2}^{\hat{\nu}_2}(\theta_{2i}) \leq \frac{j_2}{k_2}\right),$$

for $j_1 = 1, \ldots, k_1$ and $j_2 = 1, \ldots, k_2$.

The simple Bernstein copula will typically not lead to the obtention of a strictly continuous, bivariate density function. For the density to be continuous, it is required that:

$$p_{1j_2} = p_{k_1 j_2}, \qquad \text{for } j_2 = 1, \ldots, k_2, \tag{3.14}$$

$$p_{j_1 1} = p_{j_1 k_2}, \qquad \text{for } j_1 = 1, \ldots, k_1. \tag{3.15}$$

which will not in general be true. Therefore, similar to the procedure outlined in the previous chapter, we propose using the following corrections:

$$\begin{aligned}
\tilde{p}_{1j_2} &= \tilde{p}_{k_1 j_2} = \frac{p_{1j_2} + p_{k_1 j_2}}{2}, \qquad \text{for } j_2 = 2, \ldots, k_2 - 1, \\
\tilde{p}_{j_1 1} &= \tilde{p}_{j_1 k_2} = \frac{p_{j_1 1} + p_{j_1 k_2}}{2}, \qquad \text{for } j_1 = 2, \ldots, k_1 - 1, \\
\tilde{p}_{11} &= \tilde{p}_{1k_2} = \tilde{p}_{k_1 1} = \tilde{p}_{k_1 k_2} = \frac{p_{11} + p_{1k_2} + p_{k_1 1} + p_{k_1 k_2}}{4},
\end{aligned}$$

and $\tilde{p}_{j_1 j_2} = p_{j_1 j_2}$ for $j_1 \neq 1, k_1$ and $j_2 \neq 1, k_2$ which leads to a modified Bernstein copula, say $\tilde{c}_B(\cdot, \cdot)$.

An important result is that these corrections conserve the property that the marginal distributions are asymptotically, uniformly distributed on the $[0,1]^2$ interval so that the corrected copula approximation is still an asymptotic copula. In order to see this, observe that the initial matrix of weights,

$$\mathbf{P} = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1k_2} \\ p_{21} & p_{22} & \cdots & p_{2k_2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{k_1 1} & p_{k_1 2} & \cdots & p_{k_1 k_2} \end{pmatrix} \tag{3.16}$$

which verifies the limiting properties (3.12) and (3.13), is transformed into the corrected

matrix:

$$\tilde{\mathbf{P}} = \begin{pmatrix} \frac{p_{11}+p_{k_1 1}+p_{1k_2}+p_{k_1 k_2}}{4} & \frac{p_{12}+p_{k_1 2}}{2} & \cdots & \frac{p_{11}+p_{k_1 1}+p_{1k_2}+p_{k_1 k_2}}{4} \\ \frac{p_{21}+p_{2k_2}}{2} & p_{22} & \cdots & \frac{p_{2k_2}+p_{21}}{2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{p_{11}+p_{k_1 1}+p_{1k_2}+p_{k_1 k_2}}{4} & \frac{p_{k_1 2}+p_{12}}{2} & \cdots & \frac{p_{11}+p_{k_1 1}+p_{1k_2}+p_{k_1 k_2}}{4} \end{pmatrix} \tag{3.17}$$

which also verifies (3.12) and (3.13) because in the limit the sum by rows and columns is equal to k_1^{-1} and k_2^{-1}, respectively. For example, the sum of the first row of $\tilde{\mathbf{P}}$ is:

$$\begin{aligned} \lim_{n\to\infty} \sum_{j_2=1}^{k_2} \tilde{p}_{1,j_2} &= \lim_{n\to\infty} \left(\frac{p_{11}+p_{k_1 1}+p_{1k_2}+p_{k_1 k_2}}{4} + \frac{p_{12}+p_{k_1 2}}{2} + \cdots + \frac{p_{11}+p_{k_1 1}+p_{1k_2}+p_{k_1 k_2}}{4} \right) \\ &= \lim_{n\to\infty} \frac{p_{11}+p_{12}+\cdots+p_{1k_2}}{2} + \lim_{n\to\infty} \frac{p_{k_1 1}+p_{k_1 2}+\cdots+p_{k_1 k_2}}{2} = \frac{1}{k_1} \end{aligned}$$

and the same result is obtained for the last row. Also, for the first column, we obtain:

$$\begin{aligned} \lim_{n\to\infty} \sum_{j_1=1}^{k_1} \tilde{p}_{j_1 1} &= \frac{p_{11}+p_{k_1 1}+p_{1k_2}+p_{k_1 k_2}}{4} + \frac{p_{21}+p_{2k_2}}{2} + \cdots + \frac{p_{11}+p_{k_1 1}+p_{1k_2}+p_{k_1 k_2}}{4} \\ &= \lim_{n\to\infty} \frac{p_{11}+p_{21}+\cdots+p_{k_1 1}}{2} + \lim_{n\to\infty} \frac{p_{1k_2}+p_{2k_2}+\cdots+p_{k_1 k_2}}{2} = \frac{1}{k_2}. \end{aligned}$$

And the results for the remaining rows and columns follow in the same way.

Once we have shown that the empirical Bernstein copula is well defined with the corrections made and that we have constructed a strictly continuous circular-circular distribution, we can define the bivariate Bernstein density estimate as:

$$\hat{f}_B(\theta_1, \theta_2) = \tilde{c}_B\left(\hat{F}_{q_1}^{\hat{\nu}_1}(\theta_1), \hat{F}_{q_2}^{\hat{\nu}_2}(\theta_2)\right) \hat{f}_{q_1}^{\hat{\nu}_1}(\theta_1)\, \hat{f}_{q_2}^{\hat{\nu}_2}(\theta_2). \tag{3.18}$$

As commented earlier, the Bernstein copula captures the dependence structure between both variables. Therefore it is straightforward to evaluate the conditional distribution of θ_1 given θ_2 as follows:

$$\hat{f}_B(\theta_2|\theta_1) = \left[\sum_{j_1=1}^{k_1} \sum_{j_2=1}^{k_2} \tilde{p}_{j_1 j_2} \beta\left(\hat{F}_{q_1}^{\hat{\nu}_1}(\theta_1) \mid j_1, k_1 - j_1 + 1\right) \beta\left(\hat{F}_{q_2}^{\hat{\nu}_2}(\theta_2) \mid j_2, k_2 - j_2 + 1\right) \right] \hat{f}_{q_2}^{\hat{\nu}_2}(\theta_2)$$

As a final comment on the uniqueness of this model, Sklar's (1959) theorem establishes that the copula function is unique if the marginal densities are continuous. When we are

using continuous distributions defined on the real line, the cumulative distribution functions are defined in a unique form. However, in the case of circular distributions this property does not hold, except in the case of the circular uniform distribution. Thus, when different origins are selected, then different estimations of the copula will be made which will lead to slightly different approximations to the joint density.

3.4 Circular-Linear distributions constructed via empirical Bernstein copulas

In this section, we consider a second extension of the circular Bernstein polynomial, to the case of circular-linear distributions. As for the circular-circular case, we propose the use of nonparametric methods to estimate the marginal densities and the use of an adapted, empirical Bernstein copula to fit the dependence of the two variables.

Assume we have a sample of i.i.d. data, say $\{(\theta_1, x_1), \ldots, (\theta_n, x_n)\}$, generated from an unknown, circular-linear distribution.

Then, as previously, we can use a two-step estimation procedure for fitting the joint distribution where, in the first step, the marginal densities are estimated and then, in the second step, the dependence structure is estimated via Bernstein copulas.

Assuming that the marginal distribution of the circular variable may not be easily fitted by a parametric model, then the circular Bernstein estimator, $\hat{F}_k^{\hat{\nu}}(\theta)$, explained in Chapter 2, can be used to estimate this density. In the case of the linear model, any appropriate parametric or nonparametric approach could be applied. We shall write $\hat{F}_X(\cdot)$ for the estimated distribution function.

In the second step, we need to define an estimator of the copula. Given the sample of data in the unit square, $\left\{(u_{11}, u_{21}) = \left(\hat{F}_k^{\hat{\nu}}(\theta_1), \hat{F}_X(x_1)\right), \ldots, (u_{1n}, u_{2n}) = \left(\hat{F}_k^{\hat{\nu}}(\theta_n), \hat{F}_X(x_n)\right)\right\}$, we consider the empirical Bernstein copula given by,

$$\hat{c}_B(u_1, u_2) = \sum_{j_1=1}^{k_1} \sum_{j_2=1}^{k_2} p_{j_1 j_2} \prod_{i=1}^{2} \beta(u_i \mid j_i, k_i - j_i + 1), \tag{3.19}$$

where

$$p_{j_1 j_2} = \frac{1}{n} \sum_{i=1}^{n} I\left(\frac{j_1 - 1}{k_1} < \hat{F}_k^{\hat{\nu}}(\theta_i) \leq \frac{j_1}{k_1}, \frac{j_2 - 1}{k_2} < \hat{F}_X(x_i) \leq \frac{j_2}{k_2}\right), \tag{3.20}$$

for $j_1 = 1, \ldots, k_1$ and $j_2 = 1, \ldots, k_2$.

For the joint distribution to be strictly continuous, then it is necessary that

$$p_{1j_2} = p_{k_1 j_2}, \qquad \text{for } j_2 = 1, \ldots, k_2, \tag{3.21}$$

but in many cases, this condition will not be satisfied. Therefore, we propose using the following correction,

$$\tilde{p}_{1j_2} = \tilde{p}_{k_1 j_2} = \frac{p_{1j_2} + p_{k_1 j_2}}{2},$$

for $j_2 = 1, \ldots, k_2$, and $\tilde{p}_{j_1 j_2} = p_{j_1 j_2}$ for $j_1 \neq 1, k_1$ which ensures a strictly continuous circular-linear estimated density.

In a similar way to the circular-circular case, it can be easily demonstrated that this copula approximation preserves the asymptotic uniformity of the marginal distributions, as this can be seen as a particular case of the result of the previous section when one of the variables does not need correction.

Then, the bivariate Bernstein density estimate is :

$$\hat{f}_B(\theta, x) = \tilde{c}_B\left(\hat{F}_k^{\hat{\nu}}(\theta), \hat{F}_X(x)\right) \hat{f}_k^{\hat{\nu}}(\theta)\, \hat{f}_X(x) \tag{3.22}$$

The conditional density function of the linear variable given the value of the circular variable can be easily obtained from (3.22) as:

$$\hat{f}_B(x|\theta) = \sum_{j_1=1}^{k_1} \sum_{j_2=1}^{k_2} \tilde{p}_{j_1 j_2} \beta\left(F_k^{\hat{\nu}}(\theta) \mid j_1, k_1 - j_1 + 1\right) \beta\left(\hat{F}_X(x) \mid j_2, k_2 - j_2 + 1\right) \hat{f}_X(x)$$

and the conditional cumulative distribution function is:

$$\hat{F}_B(x|\theta) = \sum_{j_1=1}^{k_1} \sum_{j_2=0}^{k_2} \tilde{p}^*_{j_1 j_2} \beta\left(F_k^{\hat{\nu}}(\theta) \mid j_1, k_1 - j_1 + 1\right) \beta\left(\hat{F}_X(x) \mid j_2 + 1, k_2 - j_2 + 1\right)$$

where

$$\tilde{p}^*_{j_1 j_2} = \sum_{j=0}^{j_2} \tilde{p}_{j_1 j} \tag{3.23}$$

and $\tilde{p}^*_{j_1 0} = 0$.

3.5 Illustrations

In this section we illustrate both the circular-circular model and the circular-linear model. We shall use two practical examples based on weather data.

3.5.1 Circular-circular data: wind directions

In the first illustration we analyze data on wind directions observed from 2007 to 2009 at two buoys situated off the Atlantic coast of the USA at 43°47'0" N 68°51'18" W and 43°58'6" N 68°7'42" W with labels MISM1 y MDRM1 respectively which shall be referred as Θ_1 and Θ_2. These data are available from the National Data Buoy Center website at `http://www.ndbc.noaa.gov/`.

The two buoys have been chosen such that the distance between them is relatively small (33.35 nautical miles or 61.77 km) and they share several features, such as similar distance to land. Figure 3.1 shows circular Bernstein polynomial fits of the marginal densities of the wind directions at the two buoys. It can be seen that, as we might expect, the distribution of wind directions at the two sites are very similar.

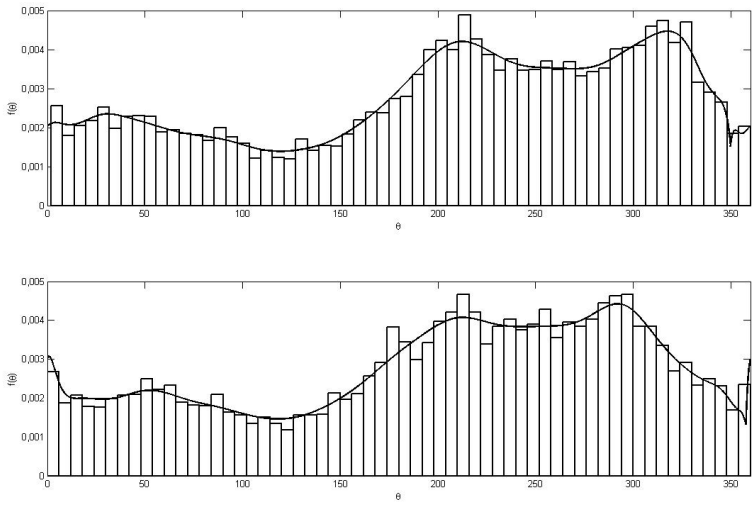

Figure 3.1: Marginal densities of the wind directions at buoys MISM1 (top) and MDRM1 (bottom)

The data have been cleaned to erase any missing values and after this, the data set contained 24807 observations.

Although, obviously, these data form a time series, we shall not consider time series modeling here and shall only concentrate on modeling the joint density of the wind directions without taking the time dependence into account. As a first step, we tested the hypothesis that the two variables were independent using the approach of Alvo (1998). This test rejected the hypothesis of independence at an $\alpha = 0.01$ level which implies that it makes sense to consider using our approach to estimate the joint density of the two variables.

Figure 3.2 shows the estimated density of the circular-circular model. For better visualization, we have moved the origins of the Bernstein polynomials to the center of the graph. As we can observe, there is a high correlation between the two variables and we may assume that the distributions of wind directions at both buoys are very similar.

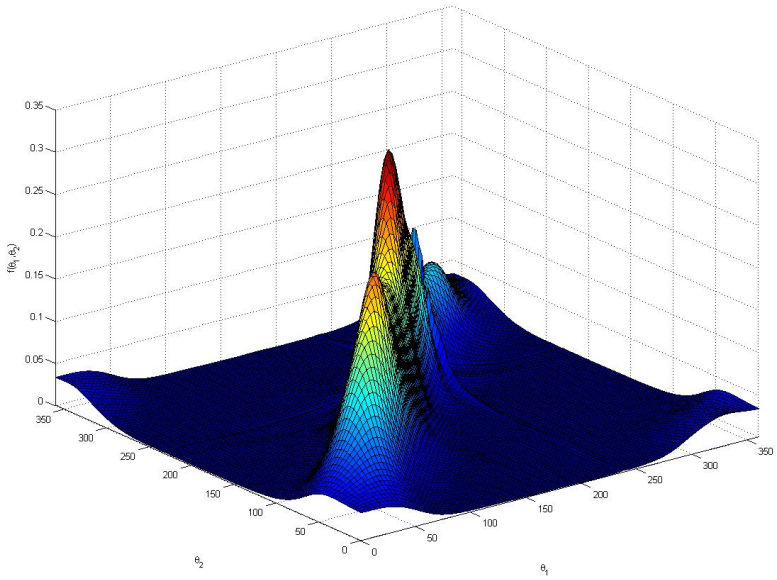

Figure 3.2: Density function for the circular-circular model, $f(\theta_1, \theta_2)$.

To illustrate this feature, we have computed the conditional densities for both buoys. Figure 3.3 shows the conditional distribution of $\Theta_1|\Theta_2$ and Figure 3.4 shows the conditional distribution of $\Theta_2|\Theta_1$. As can be seen in both graphs, we have the same density along the main diagonal of the graph. This indicates the behavior of one of the buoys as a function of the other. The distributions are useful from a operational point of view, suppose that two ships departing from the nearest ports to each of the buoys and one of them is temporarily disabled for maintenance. Observing the conditional probability we can assign a route in

which the fuel consumption of both boats can be optimized.

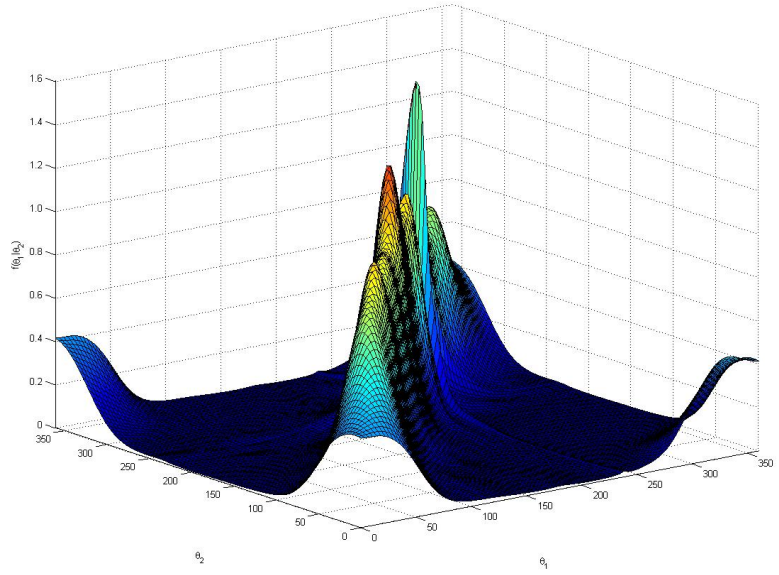

Figure 3.3: Conditional density function, $f(\theta_1|\theta_2)$

To summarize this example, we have modeled the wind direction in two nearby buoys showing the high correlated wind flow between these two sites under the lack of distorting elements such as mountains, etc. Observe that this correlation could be viewed as spatial correlation.

3.5.2 Circular-linear data: wind directions and rainfall

Many variables can influence the climate at a certain site, but from a meteorological point of view, most studies focus on wind direction and related variables such as rainfall.

For analyzing this relationship, we have chosen daily observations of rain and wind direction taken from 6/11/94 to 31/1/2009 at the observatory site at Somío, near Gijon in northern Spain, at latitude 43°32'17"N, longitude: 5°37'26"W and 30 metres above sea level. These data are available from `http://infomet.am.ub.es/clima/gijon/`.

The wind direction is measured in degrees from 0 to 359 and rainfall is measured on a grid of 0.2 liters per square meter.

At the site, rain was recorded on about 49.2% of the days. Figure 3.4 shows histograms

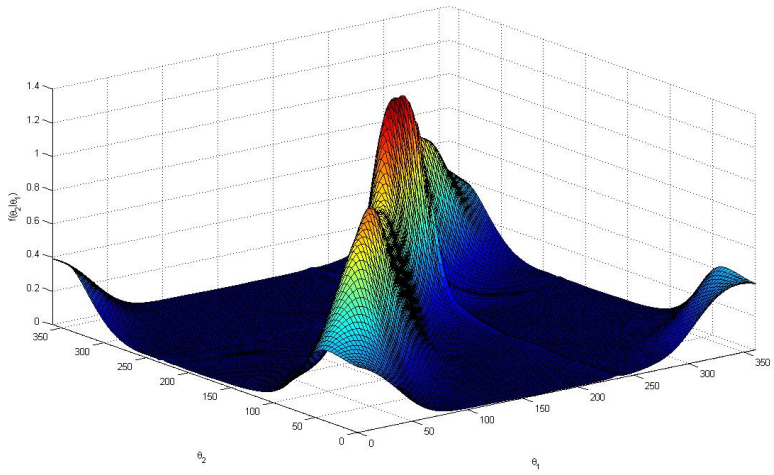

Figure 3.4: Conditional density function, $f(\theta_2|\theta_1)$

and circular Bernstein polynomial fits of the marginal density of the wind direction for the whole data set and conditioning on the weather being dry or rainy.

It can be seen that there is some difference between the estimated density fits. This may be explained by the fact that sea winds are often associated with rainfall, whereas winds coming from the land to the sea, are more regularly associated with dry weather in Spain. This suggests that it is sensible to model the joint density of the wind direction, Θ, and the level of rainfall, X, by conditioning as:

$$f(\theta, x) = f(\theta|X = 0)P(X = 0) + f(\theta, x|X > 0)P(X > 0).$$

Then, we can estimate $\hat{P}(X > 0) = 0.492$, $\hat{P}(X = 0) = 0.508$ and use a circular Bernstein polynomial, $\hat{f}_k^{\hat{\nu}}(\theta|X = 0)$ to estimate the density of Θ given that there is no rain, as in the middle diagram of Figure 3.5.

Finally, we can use the copula approach outlined in this chapter to estimate the joint density of X and Θ conditional on there being rain. In particular, we use the circular Bernstein copula density outlined in Chapter 2 to estimate the marginal density of Θ, as in the bottom diagram of Figure 3.5 and then apply a cubic spline smoothed density estimate of the marginal density of the level of rainfall and finally use the circular-linear copula outlined here to estimate the joint density.

Similar to the first illustration, we first carried out a hypothesis test for independence

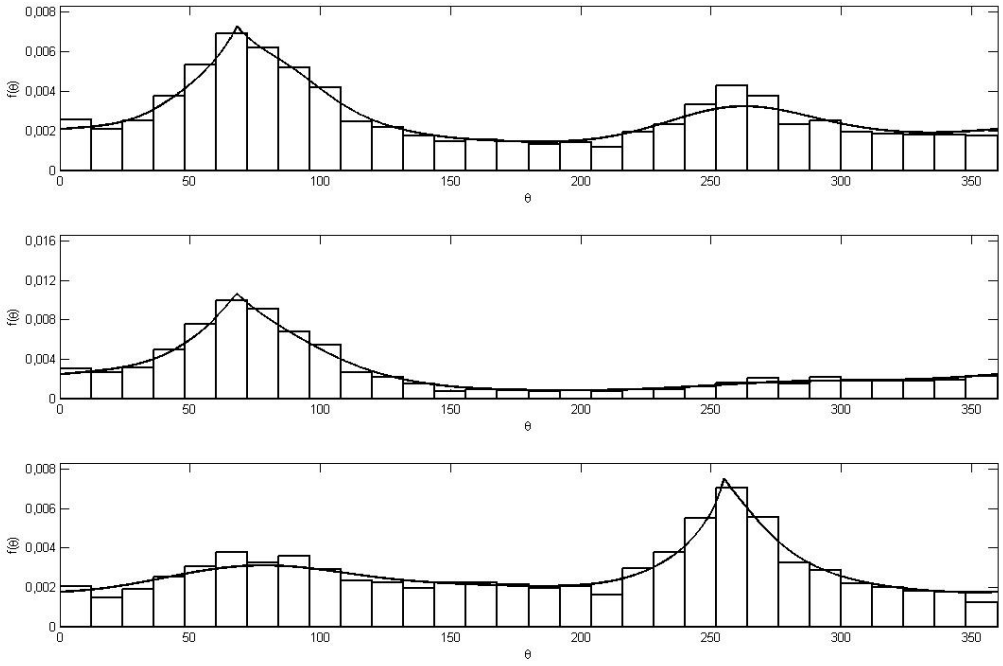

Figure 3.5: Marginal densities of wind direction: whole data set (top), dry days (middle), rainy days (bottom).

between the two variables. This was rejected at an $\alpha = 0.01\%$ level which suggests that there is dependence between the two sets of data.

In this case we have used the Bernstein polynomial of degree 51, with origin 255º approximately. We have constructed the figures with this origin represented as 0 and we increase the index clockwise.

Figure 3.6 shows the estimated joint density, where for better visualization we only illustrate the density for rainfall levels of up to 2mm. We have used cubic splines to interpolate the values of the function where there are not data (remember that the data analyzed in this example are discretized over a grid of 0.2) to obtain a smooth function. As we can observe in the figure there are two modes corresponding to East and West approximately. As we commented earlier, the orography of this part of Spain (the Cantabrian Mountains) implies that surface winds head East or West, when orographic rain is produced.

Figure 3.7 shows conditional distribution functions of the level of rainfall given that wind directions equal to the two modes of the previous figure. The solid line corresponds to the

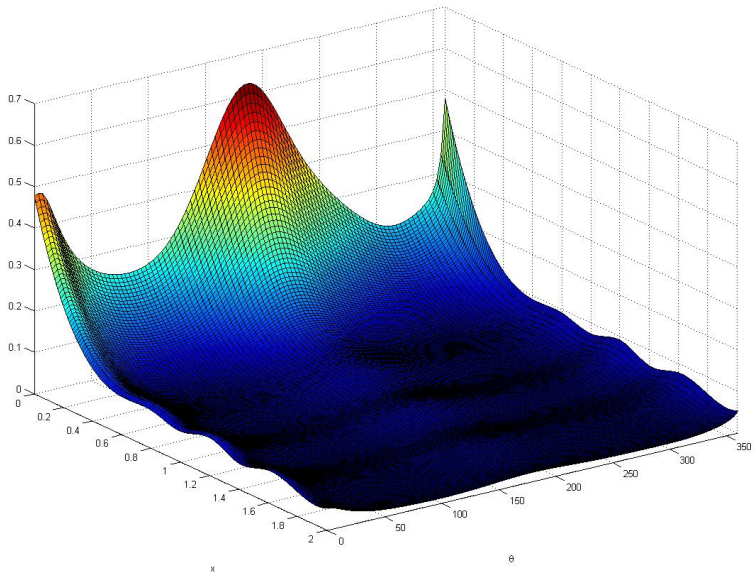

Figure 3.6: Fitted bivariate density function, f(x,$\theta|X > 0$)

origin (0) and the dotted line corresponds to the West (195º). As we can see in the figure the behavior is different due to that the cold fronts which produces intense rains termed frontal rains comes from West to East.

For a better visualization of this effect, Figure 3.8 shows a contour plot of the conditional distribution of the rainfall level given the wind direction. We show the lines corresponding to the 0.5, 0.75, 0.90 and 0.95 percentiles. Observing the graph, when the wind direction is East(0) there is more probability of heavy rain than when the wind comes from other directions.

To summarize this example, rainy days are very influenced by the orography of this part of Spain. Two types of rain predominate, that is firstly, orographic rain which happens when humid winds encounter a mountain range and producing small quantities of rain and secondly, cyclonic or frontal rains induced by low pressure systems which come from West to East and are associated with heavy rain. As we can observe in Figure 3.7, the conditional rainfall level distribution for the East direction has a heavy tail.

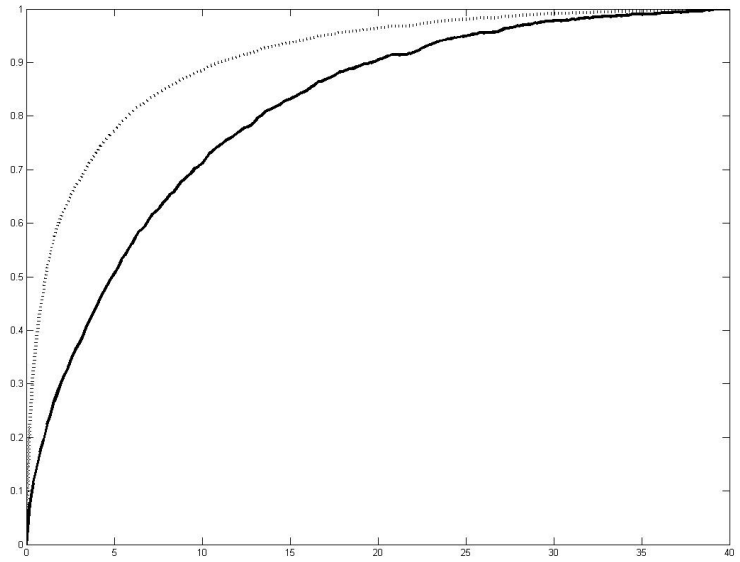

Figure 3.7: Conditional Distribution functions, F(x$|\theta, X > 0$) given the two modal wind directions.

3.6 Conclusions and extensions

In this chapter, we have illustrated how to use a nonparametric approach to construct a bivariate distribution with at least a component of circular type as an alternative to the usual, parametric models for these types of data. We have applied the empirical Bernstein copula as described in Sancetta et al (2004) and introduced appropriate corrections to guarantee that the constructed bivariate distribution is strictly continuous. Our approach has been illustrated with real data examples based on wind direction and rainfall data.

An important feature is that as in the univariate case, the fitted copulas depend on the chosen origins for estimating the marginal densities of the circular variables, although conditional on the origins, the copula is unique. Note also that although here we have used the circular Bernstein polynomials to estimate these marginal distributions, in the case of circular data, in principle any other approach could be used without altering the basic properties of the estimator.

Various extensions of our approach are possible.

Firstly, the selection of the values of $\mathbf{k} = (k_1, k_2)$ is an open problem. Here, we have chosen the recommendation of Sancetta et al (2004) which provides a relatively smooth fit,

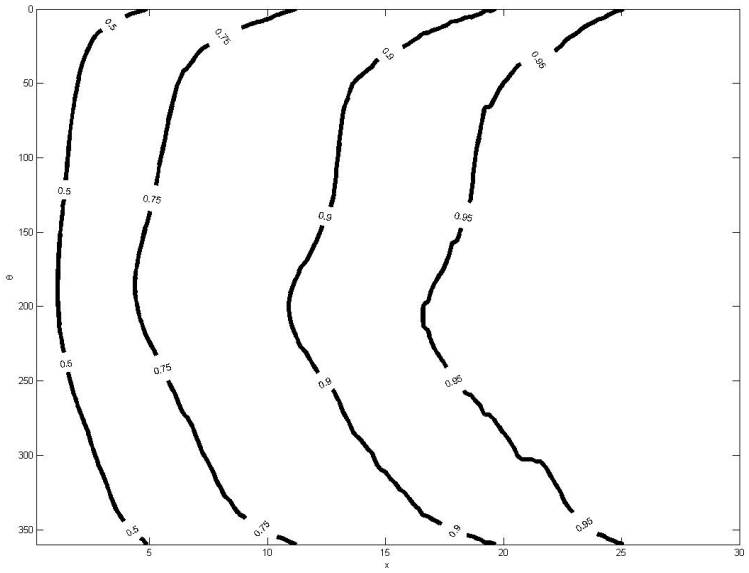

Figure 3.8: Contour plot of $F(X \leq x|\theta, X > 0)$

but it would be interesting to explore further possibilities. Secondly, it would be interesting to consider alternative approaches to non-parametric copula estimation. One procedure would be to use multidimensional, non-negative Fourier series as an approximating function.

Finally, it would be interesting to explore the possibility of using time varying copulas so that our approach could be incorporated in time series models for estimating wind directions. Also, it would be interesting to look at multivariate copulas so that other climatic variables could also be included

Chapter 4

Toroidal and spherical distributions based on positive Fourier Series

In the previous chapter we used Bernstein polynomials and Bernstein copulas to construct models for bivariate, circular-circular (toroidal) and circular-linear (cylindrical) random variables.

In this chapter we shall consider an alternative approach to constructing bivariate distributions via the use of another, well known approximating polynomial, that is the trigonometric polynomial which can be thought of as a partial sum of the terms in a Fourier series. We will use this approach to construct two distributions, firstly, a circular-circular distribution and secondly, a distribution defined on the surface of a unit sphere.

This chapter is organized as follows. In Section 4.1, we introduce Fourier series and non-negative trigonometric sums which are the most important tools for the results in this chapter. Then, in Section 4.2, we show how to construct distributions for directional data based on positive trigonometric sums. In particular, we briefly comment the approach of Fernández Durán (2004) for constructing circular distributions via non-negative trigonometric sums, and then we introduce new models for circular-circular and spherical random variables.

Illustrations are provided in Section 4.3 and then we finish with some conclusions and extensions in Section 4.4.

4.1 Fourier series and trigonometric polynomials

In this section we briefly review the basic results and properties of Fourier series and trigonometric polynomials. In particular, we concentrate on polynomials which have the property of being non-negative which implies that they can be applied as density functions for directional data. Proofs of results and further details are given in e.g. Walker (1988) and Stein et al (1971).

4.1.1 Univariate Fourier series and their properties

The Fourier series of a continuous, integrable, periodic function is defined as follows.

Definition 20. *Let $g(\cdot)$ be a real valued, periodic, function integrable in $[0, 2\pi]$. Let*

$$\begin{aligned} a_j &= \frac{1}{\pi}\int_0^{2\pi} \cos(j\theta)g(\theta)\,d\theta \quad \textit{for } j = 0, 1, 2, \ldots \\ b_j &= \frac{1}{\pi}\int_0^{2\pi} \sin(j\theta)g(\theta)\,d\theta \quad \textit{for } j = 1, 2, \ldots \end{aligned}$$

Then the infinite sum:

$$g^*(\theta) = \frac{a_0}{2} + \sum_{j=1}^{\infty} a_j \cos(j\theta) + b_j \sin(j\theta)$$

is called the Fourier series of g. The coefficients, a_j, b_j are called the Fourier coefficients.

Under certain regularity conditions, e.g. if g is square-integrable on $[0, 2\pi)$, then the Fourier series converges to the function almost everywhere. In particular, when the derivative of g is square integrable, then the Fourier series converges uniformly to g.

Fourier series can also be expressed in compact form using complex numbers. Thus, we can write

$$g^*(\theta) = \sum_{j=-\infty}^{\infty} g_j e^{ij\theta}, \tag{4.1}$$

where $g_j \in \mathbb{C}$ is defined as

$$g_j = \frac{1}{2\pi}\int_0^{2\pi} g(\theta)e^{-ij\theta}d\theta \tag{4.2}$$

and $i = \sqrt{-1}$.

From Definition 20, the coefficients, g_j in (4.1) are related to (a_j, b_j) in Definition 20 via

$$\begin{aligned} a_j &= g_j + g_{-j} \\ b_j &= i(c_j - c_{-j}) \\ c_j &= \begin{cases} \frac{1}{2}(a_j - ib_j) \\ \frac{1}{2}a_0 \\ \frac{1}{2}(a_{-j} + b_{-j}). \end{cases} \end{aligned}$$

Various useful properties of univariate Fourier series and their coefficients are given below.

Theorem 8. *The set of functions e^{ij}, $j \in \mathbb{N}$ form an orthonormal basis for the space of square integrable functions, $L^2[-\pi, \pi]$,*

$$\int_{-\pi}^{\pi} e^{ij\theta} e^{ik\theta} d\theta = 2\pi\delta_{jk} \tag{4.3}$$

where δ_{jk} is the Kronecker delta,

$$\delta_{jk} = \begin{cases} 1 & \text{if } j = k \\ 0 & \text{otherwise} \end{cases} . \tag{4.4}$$

Theorem 9 (Riemann-Lebesgue lemma)**.** *For an integrable function, $g(\cdot)$, then $\lim_{j\to\infty} a_j = \lim_{j\to\infty} b_j = 0$.*

The next two properties, known as Parseval's theorem and Plancherel's theorem, assume that $g(\cdot)$ is square integrable.

Theorem 10 (Parseval's theorem)**.** *If $g(\theta) \in L^2([0, 2\pi])$ then*

$$\sum_{j=-\infty}^{\infty} |g_j|^2 = \frac{1}{2\pi} \int_0^{2\pi} |g(\theta)|^2 d\theta \tag{4.5}$$

Theorem 11 (Plancherel's theorem)**.** *Given a sequence of coefficients d_j for $j = -\infty, \ldots, \infty$ satisfying:*

$$\sum_{j=-\infty}^{\infty} |d_j|^2 < \infty \tag{4.6}$$

then there is a unique function $g(\cdot) \in L^2([0, 2\pi))$ such that $g_j = d_j$ for $j = -\infty, \ldots, \infty$.

If (some of) the above properties are satisfied, then the function g can often be well approximated by truncating the Fourier series. These truncated Fourier series are called trigonometric polynomials.

Definition 21. *Let $g(\cdot)$ be a real valued, periodic, function integrable in $[0, 2\pi]$. Then the sum*

$$g_k^*(\theta) = \sum_{j=-k}^{k} g_j e^{ij\theta} \tag{4.7}$$

where the Fourier coefficients are calculated using (4.2) is called the trigonometric polynomial approximation of $g(\theta)$ of order k.

Parseval's theorem implies that the trigonometric polynomial is the optimal polynomial approximation in the sense of the following theorem.

Theorem 12. *The trigonometric polynomial $g_k^*(\theta)$ is the unique, best, trigonometric polynomial of degree k approximating $g(\theta)$, in the sense that, for any trigonometric polynomial approximation of degree k, say $R_k^*(\theta) \neq g_k^*(\theta)$, then*

$$\| g_k^*(\theta) - g(\theta) \|_2 < \| R_k^*(\theta) - g(\theta) \|_2$$

where the subscript 2 denotes the L^2 norm.

4.1.2 Non-negative trigonometric polynomials: The Fejér-Riesz theorem

In the statistical context of defining a trigonometric polynomial approximation of a density function, then it is important to ensure that the approximating function is non-negative. The following result due to Fejér (1915) and Riesz (1916) gives the conditions under which a trigonometric polynomial is non-negative.

Theorem 13 (Fejér-Riesz Theorem)**.** *A trigonometric polynomial $g_k^*(\theta)$ as in Definition 21 that assumes only non-negative real values for all real θ is expressible in the form*

$$g_k^*(\theta) = \left| p_k^*\left(e^{i\theta}\right) \right|^2$$

for some polynomial $p_k^*(z) = \sum_{j=0}^{k} p_j z^j$, where $z, p_j \in \mathbb{C}$ for $j = 0, \ldots, k$. In the non-singular case, the polynomial can be chosen so that it is stable, i.e. that it has no roots in the unit

disc, $\mathbb{D} = \{z : |z| < 1\}$ *and then is unique except for a multiplicative constant of modulus one.*

The exact form of a non-negative trigonometric polynomial can be found using results in e.g. Dimitrov (2002). There it is shown that non-negative trigonometric polynomials have the form:

$$|p_k^*(e^{i\theta})|^2 = \frac{a_0}{2} + \sum_{j=1}^{k}(a_j \cos(j\theta) + b_j \sin(j\theta)) \tag{4.8}$$

where:

$$a_0 = 2\sum_{k=0}^{n} |p_k|^2 \tag{4.9}$$

$$a_j - ib_j = 2\sum_{l=0}^{k-j} p_{j+l}\bar{p}_l. \tag{4.10}$$

4.1.3 Bivariate Fourier series and trigonometric polynomials

It is straightforward to generalize Definition 20 for a univariate Fourier series to the bivariate case.

Definition 22. *Let* $g(\cdot,\cdot)$ *be an integrable, periodic function of two variables and integrable in* $[0, 2\pi]^2$*, the infinite sum*

$$g^*(\theta_1, \theta_2) = \sum_{j,k\in\mathbb{Z}} g_{jk} e^{ij\theta_1} e^{ik\theta_2}$$

is called the Fourier series of $g(\theta_1, \theta_2)$*, where*

$$g_{jk} = \frac{1}{4\pi^2}\int_0^{2\pi}\int_0^{2\pi} g(\theta_1, \theta_2) e^{-ij\theta_1} e^{-ik\theta_2} d\theta_1 d\theta_2$$

are the Fourier coefficients of $g(\theta_1, \theta_2)$*.*

The properties of the bivariate Fourier series mirror those of the univariate series outlined earlier.

Theorem 14. *The set of functions* $e^{i(j\theta_1 + k\theta_2)}$*,* $j, k \in \mathbb{N}$ *is an orthonormal basis for the space*

of square integrable functions $L^2[0, 2\pi]^2$

$$\int_0^{2\pi} \int_0^{2\pi} e^{i(j\theta_1 + k\theta_2)} e^{i(l\theta_1 + m\theta_2)} \, d\theta_1 d\theta_2 = 4\pi^2 \delta_{jk,lm} \tag{4.11}$$

where δ *is the Kronecker delta*

$$\delta_{jk,lm} = \begin{cases} 1 & \text{if } j = l \text{ and } k = m \\ 0 & \text{otherwise} \end{cases} .$$

Theorem 15 (Riemann-Lebesgue lemma)**.** *Under the hypothesis that* $g(\theta_1, \theta_2)$ *is integrable, the Fourier coefficients tend to zero, that is*

$$\lim_{|j|,|k| \to \infty} g_{jk} = 0, \tag{4.12}$$

Under the assumption that $g(\cdot, \cdot)$ is square integrable, we also have bivariate versions of Parseval's theorem and Plancherel's theorem.

Theorem 16 (Parseval's theorem)**.** *If* $g(\theta_1, \theta_2) \in L^2([0, 2\pi]^2)$ *then*

$$\sum_{j,k \in \mathbb{Z}} |g_{nm}|^2 = \frac{1}{4\pi^2} \int_0^{2\pi} \int_0^{2\pi} |g(\theta_1, \theta_2)|^2 d\theta_1 d\theta_2 \tag{4.13}$$

Theorem 17 (Plancherel's theorem)**.** *Given the sequence of coefficients* d_{jk} *for* $j, k \in \mathbb{Z}$ *satisfying*

$$\sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} |d_{jk}|^2 < \infty \tag{4.14}$$

then there is a unique function $g(x, y) \in L^2([0, 2\pi]^2)$ *such that* $g_{jk} = d_{jk}$ *for* $j = -\infty, \ldots, \infty$ *and* $k = -\infty, \ldots, \infty$.

The truncated Fourier series give the trigonometric polynomials.

Definition 23. *Let* $g(\cdot, \cdot)$ *be an integrable, periodic function of two variables and integrable in* $[0, 2\pi]^2$*, then the finite sum*

$$g^*_{lm}(\theta_1, \theta_2) = \sum_{j=-l}^{l} \sum_{k=-m}^{m} g_{jk} e^{ij\theta_1} e^{ik\theta_2}$$

is the trigonometric polynomial of order (l, m), where the coefficients, g_{jk} are derived as in Definition 22.

As in the univariate case, Parseval's theorem implies that the trigonometric polynomials are optimal in the sense of the following result.

Theorem 18. *The trigonometric polynomial $g^*_{jk}(\theta_1, \theta_2)$ is the unique, best approximating, trigonometric polynomial of degree (j, k) in the sense that, for any other trigonometric polynomial of the same degree, say $R^*_{jk}(\theta_1, \theta_2) \neq g^*_{jk}(\theta_1, \theta_2)$, we have*

$$\| g^*_{jk}(\theta_1, \theta_2) - g(\theta_1, \theta_2) \|_2 < \| R^*_{j,k}(\theta_1, \theta_2) - g(\theta_1, \theta_2) \|_2 \tag{4.15}$$

where the subscript 2 denotes the L^2 norm.

Various results for the convergence rate of trigonometric polynomials can be derived which depend on the continuity properties of the underlying function. In particular, we have the following theorem.

Theorem 19. *If $g(\theta_1, \theta_2)$ belongs to $L^2([0, 2\pi)^2)$, then the bivariate trigonometric polynomial of Definition 23 converges to $g(\theta_1, \theta_2)$ in $L^2([0, 2\pi)^2)$, that is,*

$$\lim_{j,k\to\infty} \| g^*_{jk}(\theta_1, \theta_2) - g(\theta_1, \theta_2) \|_2 \to 0. \tag{4.16}$$

4.1.4 The two dimensional Fejér-Riesz theorem

In this section we show the conditions required for a bivariate trigonometric polynomial to be positive, see e.g. Geronimo et al (2004) for further details.

Generalizing Theorem 13, we might expect that a positive trigonometric polynomial, $g^*_{jk}(\theta_1, \theta_2)$ can be expressed as

$$g^*_{jk}(\theta_1, \theta_2) = \left| p^*_{jk}\left(e^{i\theta_1}, e^{i\theta_2}\right) \right|^2$$

where

$$p^*_{jk}(\theta_1, \theta_2) = \sum_{l=0}^{j} \sum_{m=0}^{k} p_{lm} e^{il\theta_1} e^{im\theta_2}$$

is a stable polynomial. However, this is not in general the case and further restrictions are necessary to provide a non-negative polynomial. The following lemma, Theorem 1.1.1 in Geronimo et al (2004) is needed for the main theorem.

Lemma 1 (Geronimo and Woerdeman, 2004)**.** *Assume that we are given complex numbers c_{lm} for $l = 0, \ldots, j$ and $m = 0, \ldots, k$. There exists a stable polynomial with no roots in $\mathbb{D}^2$,*

$$p^*(z, w) = \sum_{l=0}^{j} \sum_{m=0}^{k} p_{lm} z^l w^m \quad \text{for } z, w \in \mathbb{C}$$

with $p_{00} > 0$ and its spectral density function

$$h(z, w) = \left(p^*(z, w) \overline{p^* \left(\frac{1}{\bar{z}}, \frac{1}{\bar{w}} \right)} \right)^{-1}$$

has Fourier coefficients c_{lm} for $l = 0, \ldots, j$ and $m = 0, \ldots, k$ if and only if there exist complex numbers c_{lm} for $(l, m) \in \{1, \ldots, j\} \times \{-k, \ldots, -1\}$ such that the $(j+1)(k+1) \times (j+1)(k+1)$ doubly indexed Toeplitz matrix

$$\mathbf{\Gamma} = \begin{pmatrix} \mathbf{C}_0 & \cdots & \mathbf{C}_{-j} \\ \vdots & \ddots & \vdots \\ \mathbf{C}_j & \cdots & \mathbf{C}_0 \end{pmatrix},$$

where

$$\mathbf{C}_l = \begin{pmatrix} c_{l0} & \cdots & c_{l,-k} \\ \vdots & \ddots & \vdots \\ c_{lk} & \cdots & c_{l0} \end{pmatrix} \quad \text{for } l = -j, \ldots, j$$

and $c_{-l,-m} = \bar{c}_{lm}$ has the following two properties:

1. *$\mathbf{\Gamma}$ is positive definite.*
2. *The $(j+1)k \times (j+1)k$ submatrix of $\mathbf{\Gamma}$ obtained by removing scalar rows $1 + l(k+1)$ for $l = 0, \ldots, j$ and scalar columns $1, 2, \ldots k+1$ has rank jk.*

In this case one finds the column vector

$$(p_{00}^2 \, p_{00}p_{01} \cdots p_{00}p_{0k} \, p_{00}p_{10} \cdots p_{00}p_{20} \cdots\cdots p_{00}p_{jk})^T$$

as the first column of the inverse of $\mathbf{\Gamma}$.

The following result, Theorem 1.1.3 in Geronimo et al (2004), shows when the Fejér Reisz theorem can be generalized to the two dimensional case.

Theorem 20 (Geronimo and Woerdeman, 2004)**.** *Let $g^*_{jk}(\theta_1, \theta_2)$ be a positive, bivariate trigonometric polynomial on $[0, 2\pi)^2$ given by*

$$g^*_{jk}(\theta_1, \theta_2) = \sum_{l=-j}^{j} \sum_{m=-k}^{k} g_{lm} e^{il\theta_1} e^{im\theta_2}. \tag{4.17}$$

There exists a stable polynomial $p^*_{jk}\left(e^{i\theta_1}, e^{i\theta_2}\right)$ (with no roots in $\mathbb{D}^2$) such that

$$g^*_{jk}(\theta_1, \theta_2) = |p^*_{jk}\left(e^{i\theta_1}, e^{i\theta_2}\right)|^2, \tag{4.18}$$

if and only if the matrix $\mathbf{\Gamma}$, as in Theorem 1 constructed from the Fourier coefficients, c_{lm} of the reciprocal of g^*_{jk} satisfies condition 2 of Lemma 1. In that case, the polynomial is unique up to multiplication with a complex number of modulus 1.

Finally, we need to represent the coefficients of the bivariate trigonometric polynomial in terms of the coefficients of the stable polynomial $p^*_{jk}\left(e^{i\theta_1}, e^{i\theta_2}\right)$ by expanding (4.18). Thus, we have:

$$\begin{aligned}
|p^*_{jk}(e^{i\theta_1}, e^{i\theta_2})|^2 &= p^*_{jk}(e^{i\theta_1}, e^{i\theta_2})\overline{p^*_{jk}(e^{i\theta_1}, e^{i\theta_2})} = p^*_{jk}(e^{i\theta_1}, e^{i\theta_2})\overline{p^*_{jk}}(e^{-i\theta_1}, e^{-i\theta_2})) \\
&= \sum_{l_1=0}^{j} \sum_{m_1=0}^{k} p_{l_1 m_1} e^{il_1\theta_1} e^{im_1\theta_2} \sum_{l_2=0}^{j} \sum_{m_2=0}^{k} \overline{p}_{l_2 m_2} e^{-il_2\theta_1} e^{-im_2\theta_2} \\
&= \sum_{l_1=0}^{j} \sum_{m_1=0}^{k} \sum_{l_2=0}^{j} \sum_{m_2=0}^{k} p_{l_1 m_1} \overline{p}_{l_2 m_2} e^{i[(l_1-l_2)\theta_1 + (m_1-m_2)\theta_2]}
\end{aligned}$$

Now writing $g^*_{jk}(\theta_1, \theta_2) = |p^*_{jk}(e^{i\theta_1}, e^{i\theta_2})|^2$, we have:

$$\begin{aligned}
g^*_{jk}(\theta_1, \theta_2) &= a_{00} + \sum_{l=1}^{j} (a_{l0}\cos(l\theta_1) + b_{l0}\sin(l\theta_1)) + \sum_{m=1}^{k} (a_{0m}\cos(m\theta_2) + b_{0m}\sin(m\theta_2)) + \\
&\sum_{l=1}^{j} \sum_{m=1}^{k} (a_{lm}\cos(l\theta_1 + m\theta_2) + b_{lm}\sin(l\theta_1 + m\theta_2)) + \\
&(c_{lm}\cos(l\theta_1 - m\theta_2) + d_{lm}\sin(l\theta_1 - m\theta_2))
\end{aligned} \tag{4.19}$$

where

$$a_{00} = \sum_{l=0}^{j}\sum_{m=0}^{k} p_{lm}\overline{p}_{lm} = \sum_{l=0}^{j}\sum_{m=0}^{k} |p_{lm}|^2 \tag{4.20}$$

$$a_{l0} - ib_{l0} = 2\sum_{\varphi=0}^{j-l}\sum_{\psi=0}^{k} p_{\varphi+l,\omega}\overline{p}_{\varphi,\psi} \tag{4.21}$$

$$a_{0m} - ib_{0m} = 2\sum_{\varphi=0}^{j}\sum_{\psi=0}^{k-m} p_{\varphi,\psi+m}\overline{p}_{\varphi,\psi} \tag{4.22}$$

$$a_{lm} - ib_{lm} = 2\sum_{\varphi=0}^{j-l}\sum_{\psi=0}^{k-m} p_{l+\varphi,m+\psi}\overline{p}_{\varphi,\psi} \tag{4.23}$$

$$c_{lm} - id_{lm} = 2\sum_{\varphi=0}^{j-l}\sum_{\psi=m}^{k} p_{l+\varphi,\psi-m}\overline{p}_{\varphi,\psi} \tag{4.24}$$

Finally, we show how to compute the complex coefficients of the bivariate trigonometric polynomial in a easy way defining the stable polynomial $p^*_{jk}(\cdot,\cdot)$ in matrix form.

Let $\mathbf{P}$ be a double indexed Toeplitz matrix given by

$$\mathbf{P} = \begin{pmatrix} \mathbf{C}_0 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{C}_1 & \mathbf{C}_0 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{C}_n & \mathbf{C}_{n-1} & \cdots & \mathbf{C}_0 \end{pmatrix}$$

where

$$\mathbf{C}_l = \begin{pmatrix} p_{l,0} & 0 & \cdots & 0 \\ p_{l,1} & p_{l,0} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ p_{l,k} & p_{l,k-1} & \cdots & p_{l,0} \end{pmatrix} \quad \text{for } l = 1, \ldots, j,$$

where $p_{l,m}$ are the coefficients of the stable polynomial. Observe that the first column of $\mathbf{P}$ has the same lexicographic ordering as the stable polynomial $p^*_{jk}(z,w)$, see e.g. Geromimo et al (2004).

Then, if we compute $\mathbf{S} = \bar{\mathbf{P}}^T\mathbf{P}$, we obtain that the first column of $\mathbf{S}$ are the complex coefficients of the bivariate trigonometric polynomial with the same lexicographic ordering and the coefficients $(e^{i(lz-mw)})$ are the l'th element of the first row of the m'th block, that is

the $((m+1)j+1, k+1)$'th element of S.

4.1.5 Fourier series on the sphere

Let $g(\theta, \phi)$ be a continuous function defined on the surface of the unit sphere where $0 \leq \theta < 2\pi$ is the longitudinal angle and $0 \leq \phi < \pi$ is the colatitude. Then, we could try to use a direct Fourier series expansion of g, but it is known that this can exhibit problems such as the Gibbs phenomenon at the poles and lack of convergence see e.g. Boer et al (1975). Therefore, following Merilees (1973) and Boer et al (1975), an alternative approach can be based on transforming the sphere into a torus.

Thus, consider $(\theta, \phi) \in [0, 2\pi)^2$ and define the transformation

$$\tilde{g}(\theta, \phi) = \begin{cases} g(\theta, \phi) & \phi \in [0, \pi) \\ g(\theta + \pi, 2\pi - \phi) & \phi \in [\pi, 2\pi) \end{cases} \tag{4.25}$$

Clearly, $\tilde{g}$ is continuous and periodic with period 2π in both θ and ϕ and therefore, $\tilde{g}$ can be represented using a bivariate Fourier series,

$$\tilde{g}^*(\theta, \phi) = \sum_{j,k \in \mathbb{Z}} \tilde{g}_{jk} e^{ij\theta} e^{ik\phi}$$

where

$$\tilde{g}_{jk} = \frac{1}{4\pi^2} \int_0^{2\pi} \int_0^{2\pi} \tilde{g}(\theta, \phi) e^{-ij\theta} e^{-ik\phi} \, d\theta \, d\phi$$

as in Definition 22. Note that the imposed symmetry implies the following conditions on the coefficients:

$$\tilde{g}_{j-k} = (-1)^j \tilde{g}_{jk} \tag{4.26}$$

$$\tilde{g}_{-jk} = (-1)^j \bar{\tilde{g}}_{jk} \tag{4.27}$$

$$\tilde{g}_{-j-k} = \bar{\tilde{g}}_{jk} \tag{4.28}$$

Then, we can apply the results of Section 4.1.4 incorporating the symmetries in (4.26)-(4.28) to derive the form of the trigonometric polynomial approximation to $\tilde{g}$. Thus, consider the expression $\tilde{g}^*_{jk}(\theta, \phi) = \left| p^*_{jk}(\theta, \phi) \right|^2$. We have:

$$
\begin{aligned}
|p^*_{jk}(e^{i\theta}, e^{i\phi})|^2 &= \sum_{l=0}^{j}\sum_{m=0}^{k} p_{lm}\overline{p}_{lm} + \sum_{m=1}^{k}\left(\left(\sum_{l=0}^{j}\sum_{\psi=0}^{k-m} p_{l,m+\psi}\overline{p}_{l,\psi}\right)e^{i\omega\phi} + \left(\sum_{l=0}^{j}\sum_{\psi=0}^{k-m}\overline{p}_{l,m+\psi}p_{l,\psi}\right)e^{-i\omega\phi}\right) \\
&+ \sum_{l=1}^{j}\left(\left(\sum_{\varphi=0}^{j-l}\sum_{m=0}^{k} p_{\varphi+l,m}\overline{p}_{\varphi,m}\right)e^{i\nu\theta} + \left(\sum_{\varphi=0}^{j-l}\sum_{m=0}^{k}\overline{p}_{\varphi+l,m}p_{\varphi,m}\right)e^{-i\theta\nu}\right) \\
&+ \sum_{l=1}^{j}\sum_{m=1}^{k}\left(\left(\sum_{\varphi=0}^{j-l}\sum_{\psi=0}^{k-m} p_{\varphi+l,m+\psi}\overline{p}_{\varphi,\psi}\right)e^{i(\nu\theta+\omega\phi)} + \left(\sum_{\psi=0}^{k-m}\sum_{\varphi=0}^{j-l}\overline{p}_{\varphi+l,m+\psi}p_{\varphi,\psi}\right)e^{-i(\nu\theta+\omega\phi)}\right) \\
&+ \sum_{l=1}^{j}\sum_{m=1}^{k}\left(\left(\sum_{\varphi=0}^{j-l}\sum_{\psi=m}^{k} p_{\varphi+l,\psi-m}\overline{p}_{\varphi,\psi}\right)e^{i(\nu\theta-\omega\phi)} + \left(\sum_{\psi=m}^{k}\sum_{\varphi=0}^{j-l}\overline{p}_{\varphi+l,\psi-m}p_{\varphi,\psi}\right)e^{-i(\nu\theta-\omega\phi)}\right)
\end{aligned}
$$

which can be shown to be reducible to the real form:

$$
\begin{aligned}
|p^*_{jk}(e^{i\theta}, e^{i\phi})|^2 &= a_{00} + \sum_{l=1}^{j}\left(a_{l0}\cos(l\theta) + b_{l0}\sin(l\theta)\right) + \sum_{m=1}^{k} a_{0m}\cos(m\phi) && (4.29)\\
&+ \sum_{l=2,l\,even}^{j}\sum_{m=1}^{k}\left(a_{lm}\cos(l\theta)\cos(m\phi) + b_{lm}\sin(l\theta)\cos(m\phi)\right) && (4.30)\\
&+ \sum_{l=1,l\,odd}^{j}\sum_{m=1}^{k}\left(a_{lm}\cos(l\theta)\sin(m\phi) + b_{lm}\sin(l\theta)\sin(m\phi)\right) && (4.31)
\end{aligned}
$$

where

$$
\begin{aligned}
a_{00} &= \sum_{l=0}^{j}\sum_{m=0}^{k} p_{lm}\overline{p}_{lm} = \sum_{l=0}^{j}\sum_{m=0}^{k}|p_{lm}|^2 && (4.32)\\
a_{l0} - ib_{l0} &= 2\sum_{\varphi=0}^{j-l}\sum_{m=0}^{k} p_{l+\varphi,m}\overline{p}_{\varphi,m} && (4.33)\\
a_{0m} - ib_{0m} &= 2\sum_{l=0}^{j}\sum_{\psi=0}^{k-m} p_{l,m+\psi}\overline{p}_{l,\psi} && (4.34)\\
a_{lm} - ib_{lm} &= 4\sum_{\varphi=0}^{j-l}\sum_{\psi=0}^{k-m} p_{l+\varphi,m+\psi}\overline{p}_{\varphi,\psi} && (4.35)
\end{aligned}
$$

Finally, we can derive the trigonometric polynomial estimate of the original function, $g^*_{jk}(\theta,\phi)$ as $g^*_{jk}(\theta,\phi) = \tilde{g}^*_{jk}(\theta,\phi)$ for $0 \le \theta < 2\pi$ and $0 \le \phi < \pi$. Given that the conditions of

Theorem 1 are verified, the polynomial approximation is positive.

4.2 Distributions for directional variables based on positive trigonometric polynomials

In this section, we illustrate how the previous results may be used to construct densities for directional data. Firstly, we briefly mention the univariate case which was considered in Fernández Durán (2004) and then consider new models for circular-circular and spherical data.

4.2.1 Univariate circular trigonometric polynomial distributions

Fernández Durán (2004) showed how to apply the Fejér-Riesz theorem to derive a valid distribution for univariate, circular data.

Definition 24. *A circular variable, Θ, has a circular distribution based on non-negative trigonometric sums if it has density*

$$f(\theta|j,k,\mathbf{p}) = \frac{1}{2\pi} + \sum_{j=1}^{k} a_j \cos(j\theta) + b_j \sin(j\theta) \quad \textit{for } \theta \in [0, 2\pi),$$

where $\mathbf{p} = (p_0, \ldots, p_k)$ and $p_j \in \mathbb{C}$ for $j = 0, \ldots, k$ are such that

$$\begin{aligned} \sum_{j=0}^{k} |p_j|^2 &= \frac{1}{2\pi} \\ a_k - ib_j &= 2\sum_{l=0}^{k-j} p_{j+l}\bar{p}_l \end{aligned}$$

This distribution can be thought of as an alternative to the circular Bernstein polynomial distribution studied in Chapter 2. Both models are based on approximating polynomials and increasing the degree of these polynomials should provide a good approximation to a continuous circular distribution. Further properties are given in Fernández Durán (2004).

4.2.2 A circular-circular or toroidal distribution

Consider an unknown, strictly positive, continuous, circular-circular continuous density function $f(\theta_1, \theta_2)$ defined on the bitorus, $\theta_1, \theta_2 \in [0, 2\pi)^2$. Using the results of the previous subsection, this density function can be approximated by a bivariate trigonometric polynomial which can be factored as the sum of squares of the absolute value of the stable polynomials given in (4.19).

The first step is to normalize the bivariate trigonometric polynomial to have integral value one. If we integrate (4.19) over the domain $[0, 2\pi)^2$, we find

$$\int_0^{2\pi} \int_0^{2\pi} |p^*_{nm}(\theta_1, \theta_2)|^2 d\theta_1 d\theta_2 = 4\pi^2 a_{00}, \tag{4.36}$$

this implies from (4.20) that $a_{00} = \sum_{l=0}^{j} \sum_{m=0}^{k} |p_{lm}|^2 = \frac{1}{4\pi^2}$. Therefore, the resulting distribution is the following.

Definition 25. *Assume that we have a stable polynomial $p^*_{jk}(\cdot, \cdot)$ satisfying condition 2 of Lemma 1. Then, the associated circular-circular density function is:*

$$\begin{aligned} f(\theta_1, \theta_2 | j, k, p^*) &= \frac{1}{4\pi^2} + \sum_{l=1}^{j} (a_{l0} \cos(l\theta_1) + b_{l0} \sin(l\theta_1)) \\ &+ \sum_{m=1}^{k} (a_{0m} \cos(m\theta_2) + b_{0m} \sin(m\theta_2)) \\ &+ \sum_{l=1}^{j} \sum_{m=1}^{k} (a_{lm} \cos(l\theta_1 + m\theta_2) + b_{lm} \sin(l\theta_1 + m\theta_2)) \\ &+ \sum_{l=1}^{j} \sum_{m=1}^{k} (c_{lm} \cos(l\theta_1 - m\theta_2) + d_{lm} \sin(l\theta_1 - m\theta_2)) \end{aligned}$$

where the coefficients satisfy the conditions (4.20)-(4.24).

Trigonometric moments

It is easy to compute the trigonometric moments of this distribution. Given the orthogonality properties of the terms of a Fourier series, we have:

$$\begin{aligned}
E[\cos(l\theta_1)] &= 2\pi^2 a_{l0} && (4.37)\\
E[\sin(l\theta_1)] &= 2\pi^2 b_{l0} && (4.38)\\
E[\cos(m\theta_2)] &= 2\pi^2 a_{0m} && (4.39)\\
E[\sin(m\theta_2)] &= 2\pi^2 b_{0m} && (4.40)\\
E[\cos(l\theta_1)\cos(m\theta_2)] &= \pi^2 (a_{lm} + c_{lm}) && (4.41)\\
E[\cos(l\theta_1)\sin(m\theta_2)] &= \pi^2 (b_{lm} - d_{lm}) && (4.42)\\
E[\sin(l\theta_1)\cos(m\theta_2)] &= \pi^2 (b_{lm} + d_{lm}) && (4.43)\\
E[\sin(l\theta_1)\sin(m\theta_2)] &= \pi^2 (c_{lm} - a_{lm}) && (4.44)
\end{aligned}$$

Model fitting

Let $(\theta_{11}, \theta_{21}), \ldots, (\theta_{1n}, \theta_{2n})$ be a random sample from a circular-circular distribution. Then, in order to fit a trigonometric polynomial density approximation, we can use a restricted maximum likelihood estimation approach. Thus, first, we maximize the likelihood function,

$$\hat{p^*} = \arg\max \sum_{i=1}^{n} \log \ f^*_{jk}(\theta_{1i}, \theta_{2i}|p^*) \tag{4.45}$$

over the set of stable polynomials, p^*, subject to $f^*_{jk}(\theta_1, \theta_2|p^*) > 0$ for all (θ_1, θ_2) in $[0, 2\pi)^2$.

In other words, we must verify in each iteration of the optimization algorithm (i.e. linear search, sqp) that the movement direction in the set of stable polynomials generate a positive trigonometric polynomial. We suggest using the uniform distribution (i.e. $p_{00} = \frac{1}{2\pi}$) as starting point.

If the derived density approximation, $\hat{f}^*_{jk}(\theta_1, \theta_2)$ is positive for all (θ_1, θ_2) in $[0, 2\pi)^2$ then this is clearly a global maximum given the uniqueness of the stable polynomial.

Goodness of fit and model choice

The most practical tool for goodness of fit testing is to apply a simple χ^2 test. To do this, we must select an origin and reversing the operations the toroidal distribution is expanded

on a plane of size $[0, 2\pi]^2$. Then, we select two orders, m_1 and m_2 and divide the $[0, 2\pi]^2$ interval in $m_1 \times m_2$ equally sized cells and then, to do the test.

In order to select the order of the trigonometric polynomial to be applied, we suggest starting with the trigonometric polynomial approximation of lowest order that is not rejected using the goodness of fit test and then applying standard model choice criteria such as AIC or BIC to see if higher order polynomials provide better fits. Another possibility is to use the suggestions about the number of terms retained when the trigonometric polynomials have been used for data defined on the real line, see Hart(1985) and Diggle et al (1986) to try to find a good starting point.

4.2.3 Spherical distributions

The bivariate trigonometric polynomial outlined in Section 4.1.5 may be applied to derive a density function directly. However, it seems more natural in this context to estimate the function h, where

$$f(\theta, \phi) = h(\theta, \phi) \sin \phi.$$

This is the approach we shall take here.

First, we make the transformation $X = \cos \Phi$ when $dx = -\sin \phi \, d\phi$. As we have seen in Section 1.2.3, the surface element of the sphere arises in natural form when this change of variables is made. This transformation is used to construct maps using a cylinder tangent to the equator and is known as the Lambert cylindrical equal area projection, see e.g. Snyder (1997). Then we can define the function

$$H(\theta, x) = h(\theta, \cos \phi)$$

which takes values on $[0, 2\pi) \times (-1, 1]$. Finally, we transform this function into a function of a toroidal variable, (Θ, Y) using (4.25) and defining

$$\tilde{H}(\theta, y) = \begin{cases} H(\theta, y) & y \in (-1, 1] \\ H(\theta + \pi, 2 - y) & y \in (1, 3] \end{cases}$$

for $\theta \in [0, 2\pi)$ and $y \in (-1, 3]$. Finally, with suitable rescaling, a positive, trigonometric polynomial expansion of $\tilde{H}$ as outlined in Section 4.1.5 can be used to generate a spherical

density function. Thus, we obtain:

$$\begin{aligned}\tilde{H}(\theta, y) &= \frac{1}{4\pi} + \sum_{l=1}^{j} (a_{l0}\cos(l\theta) + b_{l0}\sin(l\theta)) + \sum_{m=1}^{k} a_{0m}\cos(my\pi/2) \\ &+ \sum_{l=2,leven}^{j}\sum_{m=1}^{k}(a_{lm}\cos(l\theta)\cos(my\pi/2) + b_{lm}\sin(l\theta)\cos(my\pi/2)) \\ &+ \sum_{l=1,lodd}^{j}\sum_{m=1}^{k}(a_{lm}\cos(l\theta)\sin(my\pi/2) + b_{lm}\sin(l\theta)\sin(my\pi/2))\end{aligned}$$

where a_{lm} and b_{lm} for $l = 0, \ldots, j$ and $m = 0, \ldots, k$ are given by (4.32)-(4.35).

As we can see in the previous set of equations, the functional form of the extension to the torus has two spheres projected onto the cylinder. Applying the previous result to the sphere, we must take a primary half (the $[0, 2\pi) \times [-1, 1)$ interval) and undo the change of variables to obtain a density function on the sphere given by:

$$\begin{aligned}f(\theta, \phi|\mathbf{p}) &= \hat{h}(\theta, \phi)\sin(\phi) \\ &= \sin(\phi)\left(\frac{1}{4\pi} + \sum_{l=1}^{j} (a_{l0}\cos(l\theta) + b_{l0}\sin(l\theta)) + \sum_{m=1}^{k} a_{0m}\cos(m\phi)\right. \\ &+ \sum_{l=2,leven}^{j}\sum_{m=1}^{k}(a_{lm}\cos(l\theta)\cos(m\phi) + b_{lm}\sin(l\theta)\cos(m\phi)) \\ &+ \left.\sum_{l=1,lodd}^{j}\sum_{m=1}^{k}(a_{lm}\cos(l\theta)\sin(l\phi) + b_{lm}\sin(l\theta)\sin(m\phi))\right)\end{aligned}$$

where $p_{lm} \in \mathbb{C}$ for $l = 0, \ldots, j$ and $m = 0, \ldots, k$ where a_{ij} and b_{ij} for $l = 0, \ldots, j$ and $m = 0, \ldots, k$ are given by (4.32)-(4.35).

Estimation

Let $(\theta_1, \phi_1, ..., \theta_n, \phi_n)$ be a random sample from a unknown spherical distribution Θ. The problem is to find a stable polynomial in order to construct the probability density function $\hat{f}_{jk}(\theta, \phi)$ using the procedure described above such that the probability density element of

the resulting spherical distribution maximize the log-likelihood of the sample.

$$\max \sum_{j=1}^{n} \log \hat{f}(\theta_j, \phi_j) \tag{4.46}$$

Observe that we are estimating the coefficients which maximize the log-likelihood function. Then, in each step of the optimization process we have to compute the log-likelihood on the primary region and and verify that h is positive or we have to compute the log-likelihood on the sphere and verify that the probability density function is positive.

As we commented previously, the stable polynomial is unique if $a_{00} > 0$. Then, the final step is to verify that (4.32) is satisfied which ensures that the integral of the probability density distribution is 1.

From a practical point of view, the torus version gives us an extra characteristic, multiplying the size of data by 2. Let us consider a small data set, say 100 observations. When we mirror the sphere, we have reflected the data and we have the double of observations, 200. Then we have more degrees of freedom given by the symmetries of the extended region to fit the model given that the number of parameters to be estimated are the same in both the model based on original region and the model based on the extended region.

We must observe that we are using an approximating polynomial. So, for large enough data sets we can approximate the underlying distribution independently of the properties of the distribution, i.e. rotational symmetry, bimodality, etc.

Similar to the toroidal distribution, there is no a general goodness of fit test for verifying if the estimated distribution fits well the data. However, one sensible approach as in Mardia et al (1999) is to use Lambert's projection to divide the sphere into regions of equal area and then apply a standard χ^2 goodness of fit test.

4.3 Illustrations

In this section we illustrate the circular-circular and spherical models described in this chapter with two real data sets.

4.3.1 Circular circular data

Here, we study the data on wind directions at two nearby buoys previously analyzed in Section 3.5.1. Bivariate trigonometric polynomial density estimators of various orders were

considered. In particular, Figure 4.1 shows the fitted density estimates based on using trigonometric polynomials of degrees (2,2) and (4,4) in the first row and (6,6) and (8,8) in the second row respectively.

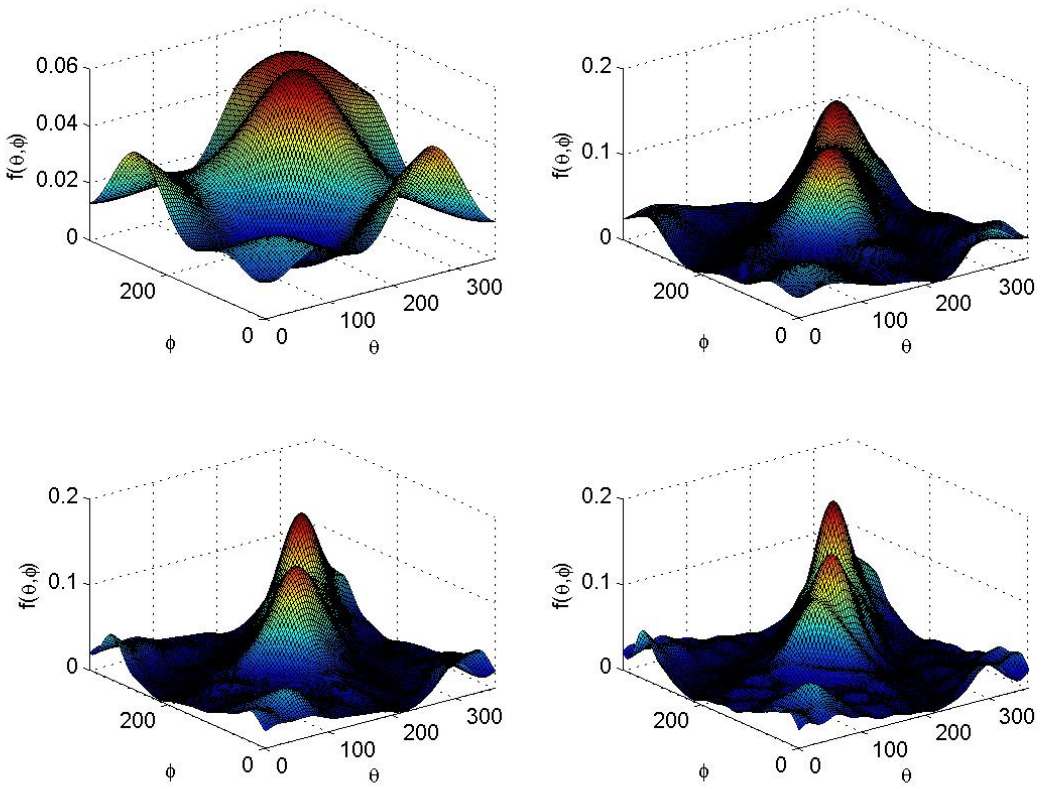

Figure 4.1: Estimated densities $f(\theta_1, \theta_2)$

As with the Bernstein polynomial based approaches, there is a tradeoff between smoothness and performance. When the degree of the trigonometric polynomial is increased, there is a tendency to overfit the data. In this case we have selected the bivariate trigonometric polynomial density of orders (8,8) as optimal model.

We also carried out chi-squared goodness of fit tests of the previous models by partitioning the domain into 14×14 and 15×15 equally sized regions. In all 4 cases, the null hypotheses that the sampled data were generated from the proposed distributions were not rejected at a 5% confidence level.

In order to compare the trigonometric polynomial density fit with the copula based

Bernstein polynomial density fit studied in Chapter 3, Figure 4.2 shows the circular-circular distribution obtained with both approaches; the empirical Bernstein copula (first row) and the optimal bivariate trigonometric polynomial of degrees (8,8) (second row). Both models give similar fits to the data and appear to capture the underlying distribution quite well.

As we can observe in the Figure 4.2 the bivariate trigonometric polynomial density is smoother than the circular-circular density using the empirical Bernstein copula. This is due that the selected model in the bivariate trigonometric polynomial density of order (8,8) has 162 parameters and each circular Bernstein polynomial density has a similar number of parameters when we add the number of weight of the copula we have a proportion of estimated parameters of approximately 1:8.

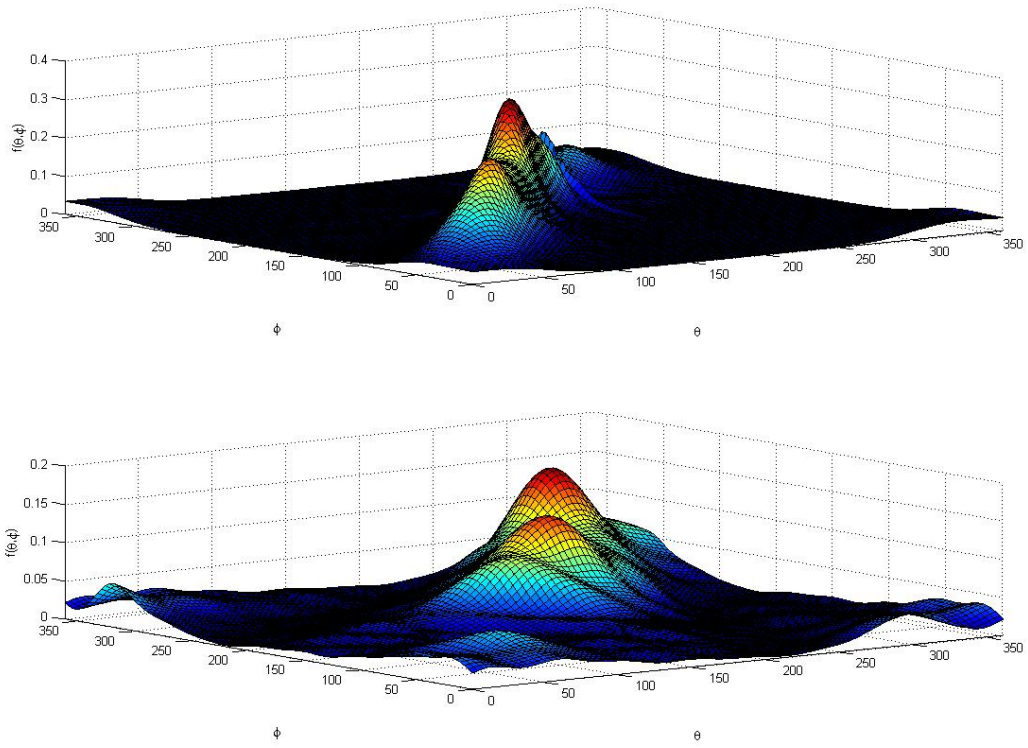

Figure 4.2: Densities of the circular-circular model using copulas and bivariate trigonometric polynomials

4.3.2 Spherical data

Here, we analyze a data set presented in Example 6.4 of Fisher et al (1987). The data correspond to 221 orientations of joint planes in Triassic sandstone at Wanganderry Lookout, New South Wales. Figure 4.3 shows the data defined on the sphere from four different viewpoints. As we can see in the figure there are several points where the data are more concentrated.

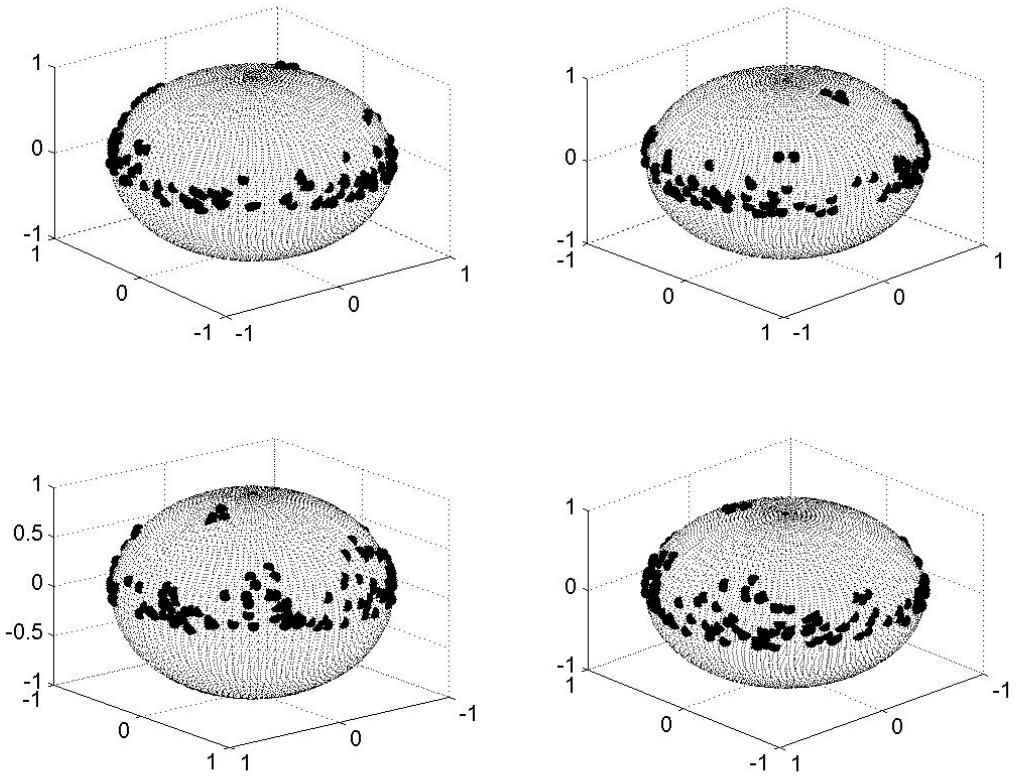

Figure 4.3: Dot plot of the data)

We applied the trigonometric polynomial based estimator of the spherical distribution using polynomials of various orders. We carried out goodness of fit tests on the models presented below and we considered only those models which were not rejected at a 5% significance level

Model selection was carried out using the AIC. Under this criterion, the optimal model

is rotationally symmetric as in Figure 4.4.

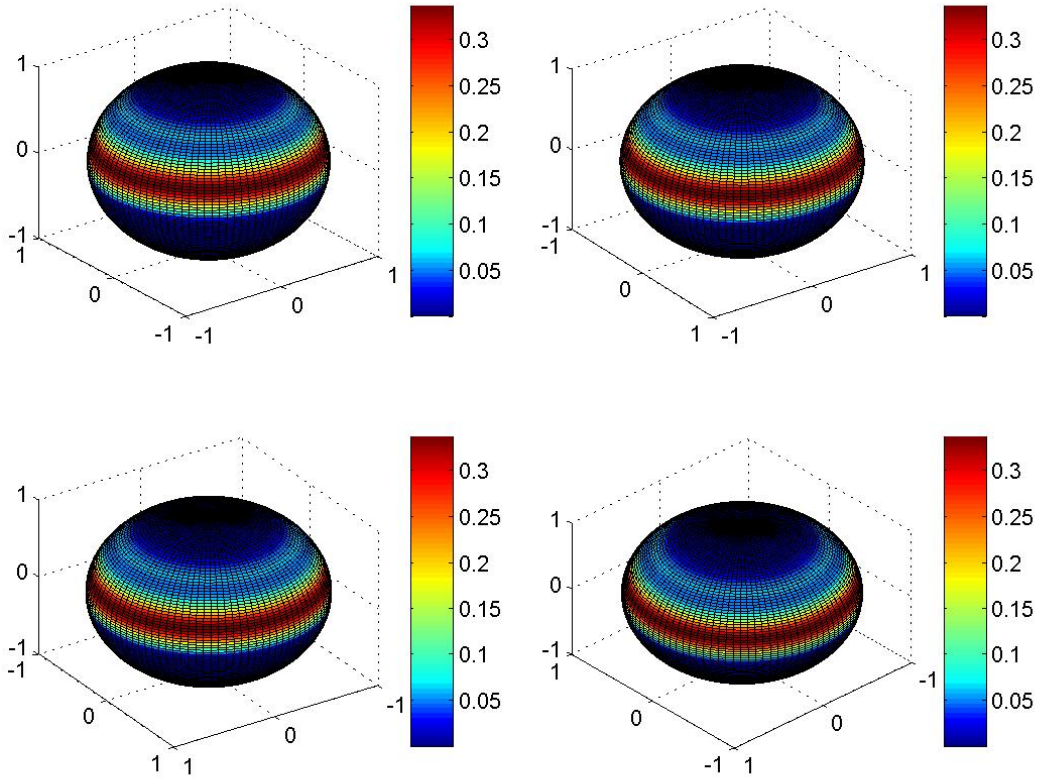

Figure 4.4: Rotationally symmetric spherical density approximation

As we can observe in Fisher et al (1987) the model of reference is a girdle form distribution which is rotationally symmetric and the number of parameters grow exponentially. This implies that the selected order of longitude is zero. When we increase this order we can capture the points on the sphere which are clustered and visualize the position the observations near the pole as can be seen in Figure 4.3.2.

4.4 Conclusions and extensions

Here, we have shown how bivariate Fourier series can be used to construct two new distributions defined on the surface of the bitorus, that is a circular-circular density model, and

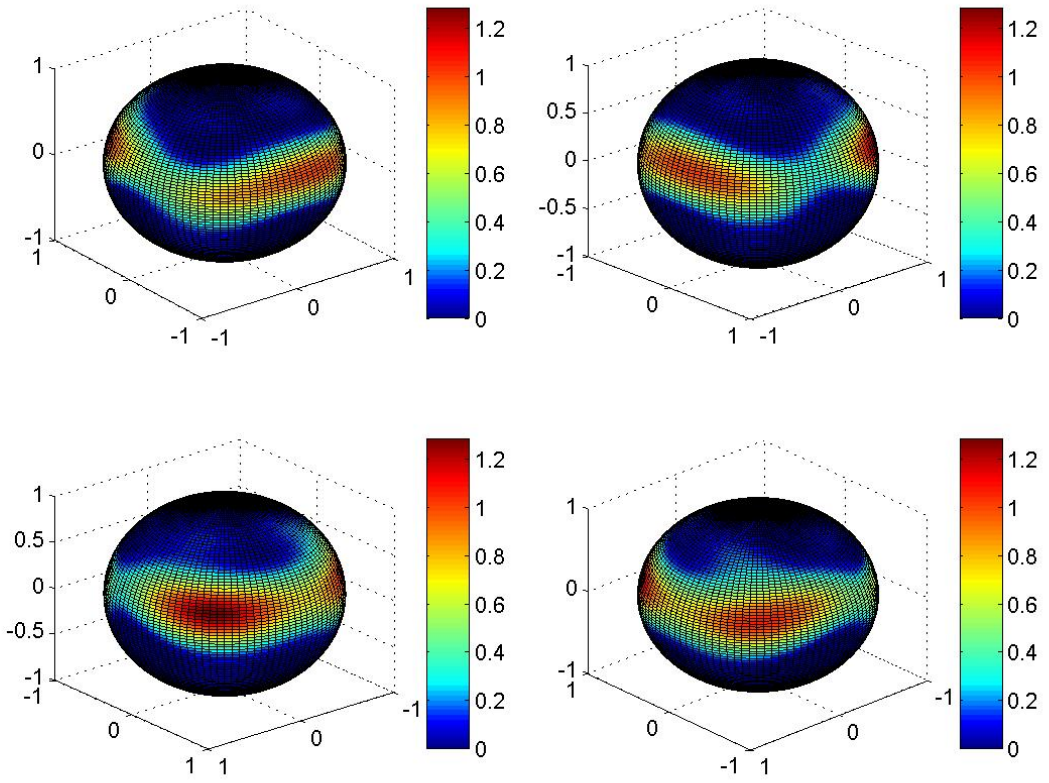

Figure 4.5: Estimated density at 4 degrees longitude and 10 degrees latitude.

on the sphere. These distributions are based on the sum of squares of the absolute value of stable polynomials. Our approach has been illustrated with real data examples.

Typically spherical data are modelled using spherical harmonics, see Müller (1966). Therefore, it would also be interesting to examine the possibilities of using positive spherical harmonic formulae to approximate density functions on the sphere.

Following the formulation of meteorological models, it would be interesting to extend these models to temporal data. An example, given in Geronimo et al (2004) is the construction of bivariate autoregressive filters using stable polynomials. Then, we can use a factorization of multivariate, positive, Laurent polynomials (Geronimo et al 2006) where three dimensions are used. We might factorize with respect to time and the other two dimensions to give us the distribution.

Bibliography

[1] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 6, 716–723.

[2] Alvo, M. (1998). On non-parametric measures of correlation for directional data. *Environmetrics*, **9**, 645–656.

[3] Arnold, K.J. (1941). *On spherical probability distributions*. Ph.D. Thesis, Massachusetts Institute of Technology.

[4] Babu, G.J., Canty, A.J. and Chaubey, Y.P. (2002). Application of Bernstein polynomials for smooth estimation of a distribution and density function. *J. Stat. Plan. Infer.*, **105**, 377–392.

[5] 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.

[6] Baldi, P., Kerkyacharian, G., Marinucci, D. and Picard, D. (2009). Adaptive density estimation for directional data using needlets. *Ann. Statist.*, **37**, 3362–3395.

[7] Batschelet, E. (1981). *Circular Statistics in Biology*. London: Academic Press.

[8] Bernstein, S. (1912). Démonstration du théorème de Weierstrass fondée sur le calcul des probabilities. *Comm. Soc. Math. Kharkov* **13**, 1-2.

[9] Bingham, C. (1964). *Distributions on the sphere and on the projective plane*. Ph.D. Thesis, Yale University.

[10] Boer, G.J. and Steinberg, L. (1975). Fourier series on spheres. *Atmosphere*, **1**, 180–191.

[11] Bouezmarni, T., Rombouts, J.V.K. and Taamouti, A. (2010). Asymptotic properties of the Bernstein density copula estimator for α-mixing data. *J. Multivariate Anal.*, **101**, 1–10.

[12] Bowman, A.W. and Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis*. Oxford: Universit Press.

[13] Burgin, M. (2010). Continuity in Discrete Sets. *Discussion paper*. Available from: `http://arxiv.org/abs/1002.0036v1`.

[14] Coles S. G. (1998) Inference for circular distributions and processes. *Statistic. Comput.*, **8**, 105–113.

[15] Dempster, A. P. Laird, N. M. and Rubin, D. B. (1977). *Journal of the Royal Statistical Society. Series B (Methodological)*, **39**, 1, 1–38.

[16] Devore, R. A. and Lorentz, G.G. (1993). *Constructive approximation*, Vol. 303. Berlin: Springer-Verlag.

[17] Diggle, P. J. and Hall, P. (1986). The selection of terms in an orthogonal series density estimator, *J. Am. Stat. Assoc.*, **81**, 230–233.

[18] Di Marzio, M., Panzera, A. and Taylor C.C. (2011). Kernel density estimation on the torus, *J. Stat. Plan. Infer.*, **141**, 2156–2173.

[19] Dimitrov, D.K. (2002). Extremal positive trigonometric polynomials. In *Approximation Theory: A volume dedicated to Blagovest Sendov*, (ed. B. Bojanov), 1–24, Sofia: Darba.

[20] Fejer, L. (1915). Über trigonometrische Polynome. *J. Reine Angew. Math.*, **146**, 53–82.

[21] Fernández-Durán, J.J. (2004). Circular distributions based on nonnegative trigonometric sums. *Biometrics*, **60**, 499–503.

[22] Fisher, R.A. (1953). Dispersion on a sphere. *Proc. R. Soc. A*, **217**, 295–305.

[23] Fisher, N.I. (1989). Smoothing a sample of circular data. *J. Struct. Geol.*, **11**, 775–778.

[24] Fisher, N.I. (1993). *Topics in Circular Data*. Cambridge: University press.

[25] Fisher, N.I. and Lee, A. (1983). A correlation coefficient for circular data. *Biometrika*, **70**, 327–332.

[26] Fisher, N.I., Lewis, T. and Embleton, B.J.J. (1987). *Statistical Analysis of Spherical Data*. Cambridge: University Press.

[27] Gawronski, W. (1985). Strong laws for density estimators of Bernstein type. *Period. Math. Hungar.*, **16**, 23–43.

[28] Gawronski, W. and Stadtmüller, U. (1981). Smoothing histograms by means of lattice and continuous distributions, *Metrika*, **28**, 155–164.

[29] Geronimo, J.S. and Lai, M-J. (2006). Factorization of multivariate positive Laurent polynomials. *J. Approx. Theory*, **139**, 327–345.

[30] Geronimo, J.S. and Woerdeman, H.J. (2004). Positive extensions, Fejér-Riesz factorization and autoregressive filters in two variables. *Ann. Math.*, 160, 839–906.

[31] Ghosal S. (2001) Convergence rates for density estimation with Bernstein polynomials. *Ann. Stat.*, **29**. 1264–1280.

[32] Guttorp P. and Lockhart R.A. (1988) Finding the location of a signal - a Bayesian analysis. *J. Amer. Statist. Assoc.*, **48**, 131–142.

[33] Hall, P., Watson, G.S. and Cabrera, J. (1987). Kernel density estimation for spherical data. *Biometrika*, **74**, 751–762.

[34] Hart, J.D. (1985). On the choice of trunction point in Fourier series density estimation, *J. Stat. Comput. Sim.*, **21**, 95–116.

[35] Jammalamadaka, S.R. and Kozubowski T.J. (2004). New families of wrapped distribution for modeling skew circular data. *Commun. Stat., Theor. M.*, **33**, 2059–2074.

[36] Jammalamadaka, S.R. and Sarma, Y.R. (1988). A correlation coefficient for angular variables. In Matusita, K., ed., *Statistical Theory and Data Analysis II*, 349–364, Amsterdam: North Holland.

[37] Jammalamadaka, S. R. and SenGupta, A. (2001) *Topics in Circular Statistics*. World Scientific Publishing.

[38] Jeffreys, H. (1948). *Theory of Probability* (2nd. ed.). Oxford: University Press.

[39] Johnson, R.A. and Wehrly, T.E. (1977). Measures and models for angular correlation and angular-linear correlation. *J. R. Stat. Soc. B*, **39**, 222–229.

[40] Johnson, R.A. and Wehrly, T.E. (1978). Some angular-linear distributions and related regression models. *J. Am. Stat. Assoc.*, **73**, 602–606.

[41] Kagan, A.M. Linnik, Y.V. and Rao, C.R. (1973). *Characterization Problems in Mathematical Statistics*. New York: Wiley.

[42] Kakizawa, Y. (2004). Bernstein polynomial probability density estimation. *J. Nonparametr. Stat.*, **16**, 709–729.

[43] Kakizawa, Y. (2010). A note on generalized Bernstein polynomial density estimators. *Statistical Methodology*, **8**, 136–153.

[44] Kent, T.J. (1982). The Fisher-Bingham distribution on the sphere. *J. R. Stat. Soc. B*, **44**, 71–80.

[45] Klemela, J. (2000). Estimation of densities and derivatives of densities with directional data. *J. Multivariate Anal.*, **73**, 18–40

[46] Kuiper, N. H. (1960). Test concerning random points on a circle. *Ned. Akad. Wet. Proc.*, **63**, 38–47.

[47] Langevin P. (1907) Magnétisme et théorie des électrons. *Ann. Chim. Phys.*, **5**, 71–127.

[48] Leblanc, A. (2009). Chung-Smirnov property for Bernstein estimators of distribution functions. *J. Nonparametr. Stat.*, **21**, 133–142.

[49] Leblanc, A. (2010). A bias-reduced approach to density estimation using Bernstein polynomials. *J. Nonparametr. Stat.*, **22**, 459–475.

[50] Lorentz G.G. (1986). *Bernstein polynomials*. New York: Chelsea Publishing Company.

[51] Mardia, K.V. (1972). *Statistics of Directional Data*. New York: Academic Press.

[52] Mardia, K.V. (1975). Statistics of directional data (with discussion). *J. R. Stat. Soc. B*, **37**, 349–393.

[53] Mardia, K.V. and Jupp, P.E. (1999). *Directional Statistics*. Chichester: Wiley.

[54] Mardia, K.V. and Sutton, T.W. (1975). On the modes of a mixture of two von Mises distributions. *Biometrika*, **62**, 699–701.

[55] Mardia, K.V. and Sutton, T.W. (1978). A model for cylindrical variables with applications. *J. R. Stat. Soc. B*, **40**, 229–233.

[56] McLachlan J. and Peel, D. (2000) *Finite Mixture Models*. New York: Wiley.

[57] Merilees, P.E. (1973). The pseudo-spectral approximation applied to the shallow water equations on a sphere. *Atmosphere*, **11**, 13–20.

[58] Mooney A., Helms P.J. and Jolliffe I.T. (2003). Fitting mixtures of von Mises distributions: a case study involving sudden infant death syndrome. *Comput. Stat. Data An.*, **41**, 505–513.

[59] Müller, C. (1966). *Spherical Harmonics*, Berlin: Springer.

[60] Nelsen R.B. (1999). *An Introduction to Copulas*. Berlin: Springer.

[61] Orszag, S.A. (1974). Fourier series on spheres. *Mon. Weather Rev.*, **102**, 56–75.

[62] Petrone, S. (1999a). Random Bernstein polynomials. *Scand. J. Stat.*, **26**, 373–393.

[63] Petrone, S. (1999b). Bayesian density estimation using random Bernstein polynomials. *Can. J. Stat.*, **27**, 105–126.

[64] Petrone, S. and Wassermann, L. (2002). Consistency of Bernstein polynomial posteriors. *J. R. Stat. Soc. B*, **64**, 79–100.

[65] Pewsey A. (2000) The wrapped skew-normal distribution on the circle. *Commun. Stat., Theor. M.*, **29**, 2459–2472.

[66] Pewsey A., Lewis, T and Jones M. C. (2007). The wrapped t family of circular distributions. *Aust. NZ. J. Stat.*, **49**, 79–91.

[67] Pfeifer. D., Straßburger, D. and Philipps, J. (2009) Modelling and simulation of dependence structures in nonlife insurance with Bernstein copulas. Paper presented at the 39th International ASTIN Colloquium, Helsinki. Available from `http://www.staff.uni-oldenburg.de/dietmar.pfeifer/PfStPh09rev4.pdf`.

[68] Riesz, F. M. (1916). Uber die Randwerte einer analytischen Funktion, *Quatrieme Congres des math. scandinaves*, 27–44.

[69] Rivest, L.P.(1982). Some statistical methods for bivariate circular data. *J. R. Stat. Soc. B*, **44**, 81–90.

[70] Rudin, W. (1963). The extension problem for positive-definite functions. *Illinois J. Math.*, **7**, 532–539.

[71] Sancetta, A. and Satchell, S. (2004). The Bernstein copula and its applications to modeling and approximating of multivariate distributions. *Econometric Theory*, **20**, 535–562.

[72] Schmidt, W. (1917). Statistiche Methoden beim Gefügestudium Kristalliner Schiefer. *Sitz. Kaiserl. Akad. Wiss. Wien, Math. - nat. Kl. Abt.*, **1**, **126**, 518–538.

[73] Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges, *Publications de l'Institut de Statistique de L'Université de Paris*, **8**, 229–231.

[74] Sklar, A. (1973). Random variables, joint distribution functions, and copulas. *Kybernetica*, **9**, 449–460.

[75] Snyder, J.P. (1997). *Flattening the Earth: Two Thousand Years of Map Projections.* Chicago: University Press.

[76] Spurr, B.D. and Koutbeiy, M.A. (1991). Comparison of various methods for estimating the parameters in mixtures of von Mises distributions. *Comm. Stat. Simul. Comp.*, **20**, 725–741.

[77] Stadtmüller, U. (1983). Asymptotic distributions of smoothed histograms. *Metrika*, **30**, 145–158.

[78] Stein E.M. and Weiss G. (1971). *Introduction to Fourier Analysis on Euclidean Spaces.* Princeton: University Press.

[79] Stephens, M. A. (1969) Techniques for directional data. *Technical Report*, **150**, Dept. of Statistics, Stanford University, Stanford, C. A.

[80] Taylor, C.C. (2008). Automatic bandwidth selection for circular density estimation. *Comput. Stat. Data An.*, **52**, 3493–3500.

[81] Tenbusch, A. (1994). Two-dimensional Bernstein polynomial estimators. *Metrika*, **41**, 233–253.

[82] Vitale, R.A. (1975). A Bernstein polynomial approach to density estimation. In *Statistical Inference and Related Topics, Vol 2,* (ed. Madan Lal Puri), 87–100. New York: Academic Press.

[83] Von Mises, R. (1918) Über die "Gunzzahligkeit" der Atomgewichte und Verwandte Fragen, *Phsikal. Z.* **19**, 490–500.

[84] Walker, J.S. (1988). *Fourier Analysis*. Oxford: University Press.

[85] Walker, S.G., Lijoi, A. and Prünster, I. (2007). On rates of convergence for posterior distributions in infinite-dimensional models. *Ann. Stat.*, **35**, 738–746.

[86] Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing*. London: Chapman and Hall.

[87] Watson, G. S. (1961). Goodness-of-fit tests on the circle. *Biometrika*, **48**, 109–114.

[88] Watson G.S. (1965). Equatorial distributions on a sphere. *Biometrika*, **52**, 193–201.

[89] Wood, A. (1982). A bimodal distribution on the sphere. *J. R. Stat. Soc. C*, **31**, 52–58.

[90] Wood, A. (1988). Some notes on the Fisher-Bingham family on the sphere. *Commun. Stat. Theor. M.*, **17**, 3881–3897.