Statistical Models for Data Which Include Angular Observations

Shogo Kato

September 2007

Contents

Chapter 1

Preliminaries

1.1 Introduction and summary

In a variety of scientific fields, observations are described as directions. In meteorology, for example, wind directions measured at a weather station are data of this type. Or biologists may be interested in the analysis of the gene locations of bacteria which can be expressed as directions. The directions of magnetic field in a rock sample are of interest for some geologists. A set of observations expressed as directions are referred to as *directional data*.

Since these directions can be expressed as points on the sphere, a set of directional observations is also called *spherical data*. In particular, two-dimensional spherical data is referred to as *circular data*. Statistics for these spherical data is called *directional statistics*, while the one which specialises in handling circular data is referred to as *circular statistics*.

There are enormous contexts which handle linear data, namely, a set of $\mathbb{R}^n$-valued observations. However, comparatively little work has been done if the observations are directions or points on the sphere. Some techniques for analysing this type of observations are summarized in the following monographs.

Mardia (1972) and Mardia and Jupp (2000) provided a theoretical overview of the statistics of directional data. Biological aspects of statistics of circular data were extensively studied in Batschelet (1981). Watson (1983) developed theoretical statistics on the general dimensional sphere. Fisher *et al.* (1987) and Fisher (1993) considered some techniques for the analysis of three-dimensional spherical data and circular ones, respectively. Jammalamadaka and SenGupta (2001) focused on the topics in circular statistics.

Although these monographs introduce some methods to analyse directional data, there are still various topics which have not been sufficiently investigated. For instance, there is not enough discussion on statistical models for multiple observations which include directional ones such as the data on the torus, cylinders or discs. Although scientists actually face a lot of datasets consisting of multiple angular or mixtures of linear and circular observations, there are only a limited number of works in statistics which tackle this problem.

Given this situation, in this thesis, we propose some new statistical models for the analysis of observations which include angular ones. The main topic of the thesis is to propose the following three models, based on the papers by Kato *et al.* (to appear), Kato and Shimizu (to appear) and Kato (submitted). First, a circular–circular regression model is proposed, which is a regression model in which both covariates and responses are angular variables. Second, some probability distributions for observations which include angular ones are provided. One of them is a class of tetravariate distributions with specified bivariate distributions, while the others are distributions defined on the cylinder which are obtained using the maximum entropy principle under some constraints on certain moments. The third model we introduce is a distribution for a pair of unit vectors which is generated by Brownian motion.

The subsequent chapters are organized as follows. In the rest of this chapter, we provide a brief overview of basic concepts of circular statistics and important models in directional statistics. First we introduce some key concepts such as trigonometric moments, mean direction and mean resultant length. Then we discuss some well-known distributions on the circle and sphere, namely, the von Mises, the generalized von Mises, the wrapped Cauchy and exit distributions.

The works in Chapters 2, 3 and 4 are based on the manuscripts by Kato *et al.* (to appear), Kato and Shimizu (to appear) and Kato (submitted), respectively. The particular emphasis is given to Chapters 2 and 3, which are the main chapters of the thesis. The other chapter, Chapter 4, provides a work which is related to the works in Chapters 2 and 3.

Chapter 2 provides a regression model in which both covariates and responses are angular variables. The regression curve is expressed as a form of the Möbius circle transformation. The angular error is assumed to follow a wrapped Cauchy or, equivalently, circular Cauchy distribution. A bivariate circular distribution is proposed to model our circular regression. Some properties of the regression, including estimation and testing procedures, are obtained. The proposed methods are applied to marine biology and wind direction data.

Chapter 3 discusses some stochastic models for dependence of observations which include angular ones. First, we provide a theorem which constructs four-dimensional distributions with specified bivariate marginals on certain manifolds such as two tori, cylinders or discs. Some properties of the submodel of the proposed models are investigated. The theorem is also applicable to the construction of a related Markov process, models for incomplete observations, and distributions with specified marginals on the disc. Second, two maximum entropy distributions on the cylinder are discussed. The circular marginal of each model is distributed as the generalized von Mises distribution which can be a symmetric or asymmetric, unimodal or bimodal shape. The proposed cylindrical model is applied to two datasets.

In Chapter 4, we propose a bivariate model for a pair of dependent unit vectors which is generated by Brownian motion. Both marginals have uniform distributions on the sphere, while the conditionals follow so-called "exit" distributions. Some properties of the proposed model, including parameter estimation, are investigated. Further study is given to the bivariate circular case by transforming variables and parameters into the form of complex numbers. Some desirable properties, such as multiplicative property and log-infinite divisibility, hold for this submodel. As a related topic, the proposed distribution is generalized so that both marginals have exit distributions. We also show how it is possible to construct distributions in the plane and on the cylinder by applying bilinear fractional transformations to the proposed bivariate circular model.

1.2 Basic concepts and distributions

Before we embark on the main topic, we briefly introduce some preliminary knowledge about directional statistics. First, we provide key concepts of circular statistics. Then we discuss some well known distributions on the circle and the sphere. The distributions we discuss here are the von Mises distribution, the wrapped Cauchy distribution, the generalized von Mises distribution and the exit distribution.

1.2.1 Basic concepts of circular statistics

Because of the geometrical difference between the line and the circle, some of the standard techniques for the linear data can not be directly applied to handle the circular data. Therefore it is necessary to develop statistical techniques for circular data, which are different from the ones for linear data.

In circular statistics, a random variable which takes values on the circle, say Θ, is often expressed in terms of radians as $[0, 2\pi)$ or $[-\pi, \pi)$. If Θ has the probability density function or the density, $f(\theta)$, then it satisfies the following properties:

1. $f(\theta) \geq 0 \quad$ a.e.,
2. $\displaystyle\int_0^{2\pi} f(\theta) d\theta = 1.$

The pth trigonometric moment of Θ, $\varphi_\Theta(p)$, is defined by

$$\varphi_\Theta(p) = E(e^{ip\Theta}), \quad p \in \mathbb{Z}.$$

In particular, the argument of the first trigonometric moment, i.e. $\arg\{\varphi_\Theta(1)\}$, is called *mean direction of* Θ, while the absolute value of that, $|\varphi_\Theta(1)|$, is referred to the *mean resultant length of* Θ.

Using the Fourier series theory, one can recover the density for Θ from its trigonometric moments on a natural assumption. Under a condition $\sum_{p=1}^{\infty} |\varphi_\Theta(p)|^2 < \infty$, the density for Θ is given by

$$f(\theta) = \frac{1}{2\pi} \sum_{p=-\infty}^{\infty} \varphi_\Theta(p) e^{-ip\theta}, \quad 0 \le \theta < 2\pi.$$

See Mardia and Jupp (2000, Sections 3.4.1 and 4.2.1) for details.

On occasions, a circular variable is expressed in the form of complex number via the transformation $Z = e^{i\Theta}$. Then the variable Z takes values on the unit circle in the complex plane. The trigonometric moments, mean direction and mean resultant length of Z are defined in a similar manner as in the $[0, 2\pi)$-valued random variable Θ.

Note that the circular variable can be expressed in the vectoral form using the transformation $X = (\cos\Theta, \sin\Theta)'$. It is clear that the random vector X takes values on the unit circle in $\mathbb{R}^2$.

1.2.2 Von Mises distribution

The von Mises distribution was introduced as a statistical model by von Mises (1918) and was discussed earlier by Langevin (1905) in the context of physics. The distribution is defined by the density

$$f(\theta) = \frac{1}{2\pi I_0(\kappa)} \exp\{\kappa \cos(\theta - \mu)\}, \quad 0 \le \theta < 2\pi, \tag{1.2.1}$$

where $\kappa \ge 0$, $\mu \in [0, 2\pi)$ and I_j denotes the modified Bessel function of the first kind and order j which is given by

$$I_j(z) = \frac{1}{2\pi} \int_0^{2\pi} \cos(j\theta) \exp\{z \cos\theta\}\, d\theta = \sum_{r=0}^{\infty} \frac{1}{\Gamma(r + j + 1)\, r!} \left(\frac{z}{2}\right)^{2r+j}, \quad z \in \mathbb{C}.$$

We write $\Theta \sim \mathrm{VM}(\mu, \kappa)$ if a random variable Θ has the density (1.2.1).

Assume Θ is a random variable from $\mathrm{VM}(\mu, \kappa)$. Then the pth trigonometric moment of Θ is given by

$$E(e^{ip\Theta}) = \frac{I_p(\kappa)}{I_0(\kappa)} e^{ip\mu}, \quad p \in \mathbb{Z}.$$

In particular, the mean direction and mean resultant length of Θ are given by μ and $I_1(\kappa)/I_0(\kappa)$, respectively.

The von Mises distribution is obtained as a maximum entropy distribution subject to certain moments. Let f be a density of the circular distribution with support $S = \{\theta \in [0, 2\pi)\,;\, f(\theta) > 0\}$. Then the density which maximizes the entropy

$$E = -\int_S \log\{f(\theta)\}\, f(\theta)\, d\theta,$$

subject to $E(\cos\Theta) = a,\ E(\sin\Theta) = b,\ a^2 + b^2 < 1$, corresponds to the density of the von Mises distribution.

Downs (1966) showed a method to generate a von Mises distribution through conditioning the bivariate normal distribution with some restriction on mean vector and variance-covariance matrix. Let X be a $\mathbb{R}^2$-valued random vector which obeys the bivariate normal distribution $\mathrm{N}(\mu, \sigma^2 I)$ where $\mu = (\xi\cos\eta, \xi\sin\eta),\ \xi, \sigma > 0,\ 0 \leq \eta < 2\pi$. Transform the bivariate random vector $X = (X_1, X_2)'$ into polar coordinates $(R, \Theta)'$ by putting $(X_1, X_2) = (R\cos\Theta, R\sin\Theta)$. Then the conditional distribution of Θ given $R = r$ is the von Mises distribution $\mathrm{VM}(\eta, r\xi/\sigma^2)$.

1.2.3 Generalized von Mises distribution

Rukhin (1972) discussed an asymmetric distribution on the circle as the distribution which has nontrivial sufficient statistics for location parameters on the circle with positive and continuous density. Preceding his work, the distribution originally appeared as a special case of the model proposed by Maksimov (1967). In directional statistics, this distribution is considered an extension of the von Mises distribution, which has the density

$$f(\theta) = C_1 \exp\left[\kappa_1 \cos(\theta - \mu_1) + \kappa_2 \cos\left\{2(\theta - \mu_2)\right\}\right], \quad 0 \le \theta < 2\pi, \tag{1.2.2}$$

where $\kappa_1, \kappa_2 \ge 0,\ 0 \le \mu_1, \mu_2 < 2\pi$ and C_1 is the normalizing constant

$$C_1^{-1} = 2\pi \left[I_0(\kappa_1) I_0(\kappa_2) + 2 \sum_{r=1}^{\infty} I_j(\kappa_2) I_{2j}(\kappa_1) \cos\left\{2j(\mu_1 - \mu_2)\right\} \right], \tag{1.2.3}$$

where I_j denotes the modified Bessel function of the first kind and order j. We denote the generalized von Mises distribution with density (1.2.2) by GVM$(\mu_1, \mu_2, \kappa_1, \kappa_2)$. This density can be used to represent symmetric or asymmetric, unimodal or bimodal shapes depending on the choice of $\mu_1, \mu_2, \kappa_1, \kappa_2$. When $\kappa_2 = 0$, generalized von Mises distribution reduces to the von Mises distribution VM(μ_1, κ_1). Yfantis and Borgman (1982) studied some properties of this model, including modes and parameter estimation.

The generalized von Mises is also obtained by conditioning a bivariate normal distribution without any restriction on the parameters. Let $X = (X_1, X_2)'$ be a $\mathbb{R}^2$-valued random vector which obeys the bivariate normal distribution $N_2(\mu, \Sigma)$ where $\mu = (\xi \cos\eta, \xi \sin\eta)'$, $\xi \ge 0,\ 0 \le \eta < 2\pi$,

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}, \quad \sigma_1, \sigma_2 > 0,\ -1 < \rho < 1.$$

We transform the random vector X by putting $(X_1, X_2) = (R\cos\Theta, R\sin\Theta)$, $R > 0,\ 0 \le \Theta < 2\pi$. Then the conditional distribution of Θ given $R = r$ is the generalized von Mises distribution GVM$(\mu_1, \mu_2, \kappa_1, \kappa_2)$ where

$$\cos\mu_1 = -\frac{r\xi(\rho\sigma_1\sigma_2 \sin\eta - \sigma_2^2 \cos\eta)}{\sigma_1^2\sigma_2^2(1-\rho^2)\kappa_1}, \quad \sin\mu_1 = -\frac{r\xi(\rho\sigma_1\sigma_2 \cos\eta - \sigma_1^2 \sin\eta)}{\sigma_1^2\sigma_2^2(1-\rho^2)\kappa_1},$$

$$\cos 2\mu_2 = -\frac{r^2(\sigma_2^2 - \sigma_1^2)}{4\sigma_1^2\sigma_2^2(1-\rho^2)\kappa_2}, \quad \sin 2\mu_2 = \frac{r^2\sigma_1\sigma_2\rho}{2\sigma_1^2\sigma_2^2(1-\rho^2)\kappa_2},$$

$$\kappa_1 = \frac{r^2}{2\sigma_1^2\sigma_2^2(1-\rho^2)} \sqrt{\left\{\frac{(\sigma_2^2 - \sigma_1^2)}{2}\right\}^2 + \rho^2\sigma_1^2\sigma_2^2},$$

$$\kappa_2 = \frac{r\xi}{\sigma_1^2\sigma_2^2(1-\rho^2)} \sqrt{(\rho\sigma_1\sigma_2 \sin\eta - \sigma_2^2 \cos\eta)^2 + (\rho\sigma_1\sigma_2 \cos\eta - \sigma_1^2 \sin\eta)^2}.$$

1.2.4 Wrapped Cauchy distribution

Let Θ be a random variable which takes values on the circle $[0, 2\pi)$. Then Θ has the wrapped Cauchy distribution or, equivalently, circular Cauchy distribution when the density for Θ is

$$f(\theta) = \frac{1}{2\pi} \frac{1-\rho^2}{1-2\rho\cos(\theta-\mu)+\rho^2}, \quad 0 \le \theta < 2\pi,$$

where $0 \le \rho < 1, \quad 0 \le \mu < 2\pi$. For convenience, we write $\Theta \sim WC(\mu, \rho)$ if a random variable Θ has the above density.

The distribution is derived as the wrapped distribution of the Cauchy distribution on the real line. Assume that a random variable X has the Cauchy distribution with density

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

where $-\infty < \mu < \infty$ and $\sigma > 0$. Then the wrapped distribution of the Cauchy distribution is given by $\Theta \equiv X \pmod{2\pi}$. After some algebra, it follows that $\Theta \sim WC(\mu, e^{-\sigma^2})$. See, for example, Jammalamadaka and SenGupta (2001) for the detailed calculation of the density.

McCullagh (1996) discussed the wrapped Cauchy distribution in the form of complex numbers. In this thesis we basically follow his representation of the wrapped Cauchy distribution. Let $\Theta \sim WC(\mu, \rho)$ and put $Z = e^{i\Theta}$ and $\phi = \rho e^{i\mu}$. Then the density of Z is

$$f(z) = \frac{1}{2\pi} \frac{|1-|\phi|^2|}{|z-\phi|^2}, \quad z \in \Omega, \tag{1.2.4}$$

where $|\phi| < 1, \ \Omega = \{z \in \mathbb{C}\,;\, |z| = 1\}$. McCullagh (1996) extended the parameter space of ϕ to $\mathbb{C} \backslash \Omega$. In this thesis, we further extend the domain of ϕ to $\mathbb{C}$ and define $Z = \phi$ for $\phi \in \Omega$. In the same way as McCullagh (1996), we denote the wrapped Cauchy distribution in (1.2.4) by $Z \sim C^*(\phi)$.

Here $\arg(\phi)$ is a mean direction and $|\phi|$ a mean resultant length of Z. The distribution is unimodal and symmetric about $z = \arg(\phi)$. When $|\phi|$ is equal to

0, the distribution is the uniform distribution on the circle. As $|\phi|$ tends to 1, the distribution approaches a point distribution with singularity at $Z = \phi$.

The properties of the wrapped Cauchy distribution have been investigated, for example, by Mardia (1972) and McCullagh (1996). The following hold for the wrapped Cauchy distribution:

(i) $Z \sim C^*(\phi) \implies \beta_0 Z \sim C^*(\beta_0 \phi), \quad \beta_0 \in \Omega,$

(ii) $Z_1 \sim C^*(\phi_1),\ Z_2 \sim C^*(\phi_2),\ Z_1 \perp Z_2,\ |\phi_1|, |\phi_2| \leq 1 \implies Z_1 Z_2 \sim C^*(\phi_1 \phi_2),$

(iii) $Z \sim C^*(\phi) \implies \dfrac{Z + \beta_1}{1 + \overline{\beta_1} Z} \sim C^* \left(\dfrac{\phi + \beta_1}{1 + \overline{\beta_1} \phi} \right), \quad \beta_1 \in \mathbb{C},$

(iv) $Z \sim C^*(\phi) \implies Z \sim C^*(1/\overline{\phi}).$

The properties (i) and (iii) show that if Z is distributed as a uniform distribution $C^*(0)$, then the Möbius circle transformation of Z generates the wrapped Cauchy distribution; i.e. $\beta_0 (Z + \beta_1)/(1 + \overline{\beta_1} Z) \sim C^* (\beta_0 \beta_1)$ where $\beta_0 \in \Omega$ and $\beta_1 \in \mathbb{C}$.

Note that (ii)–(iv) do not hold for the von Mises distribution.

1.2.5 Exit distribution

A random variable which takes values on the unit sphere in $\mathbb{R}^d$ is called a *d-dimensional spherical variable*. Some distributions for the spherical variable or, simply, spherical distributions are introduced in Mardia and Jupp (2000, Chapter 9). In this section we introduce a spherical distribution, i.e. the exit distribution, which is related to the distribution given in Chapter 4.

The exit distribution for the sphere appears, for example, in Durrett (1984, Section 1.10) as a distribution related to the Brownian motion. The exit distribution for the d-dimensional sphere, $\mathrm{Exit}_d(\eta)$, is of the form

$$f(x) = \frac{1}{A_{d-1}} \frac{1 - \|\eta\|^2}{\|x - \eta\|^d}, \quad x \in S^{d-1}, \tag{1.2.5}$$

where $\eta \in \{\zeta \in \mathbb{R}^d\,;\,\|\zeta\| < 1\}$, $S^{d-1} = \{x \in \mathbb{R}^d\,;\,\|x\| = 1\}$, A_{d-1} is a surface area of S^{d-1}, i.e. $A_{d-1} = 2\pi^{d/2}/\Gamma(d/2)$, and $\|\cdot\|$ is the Euclidean norm. This distribution is unimodal and rotationally symmetric about $x = \eta/\|\eta\|$, with the concentration being controlled by $\|\eta\|$. In particular, when $\|\eta\| = 0$, the distribution reduces to the spherical uniform. As $\|\eta\| \to 1$, it tends to a point distribution with singularity at $x = \eta$. It is noted that the exit distribution coincides with the wrapped Cauchy distribution for $d = 2$. Suppose $X = (X_1, X_2)' \sim \mathrm{Exit}_2(\eta)$ where $\eta = (\eta_1, \eta_2)'$. On putting $Z = X_1 + iX_2$, the distribution of Z is given by $Z \sim C^*(\eta_1 + i\eta_2)$.

The exit distribution is derived as follows. Let $\{B_t\}$ be $\mathbb{R}^d$-valued Brownian motion starting at $B_0 = \eta$ where $d \geq 2$ and $\eta \in \{x \in \mathbb{R}^d\,;\,\|x\| < 1\}$. Suppose that τ is the smallest time at which the Brownian particle hits the d-dimensional unit sphere, i.e. $\tau = \inf\{t \geq 0\,;\,\|B_t\| = 1\}$. Then $B_\tau \sim \mathrm{Exit}_d(\eta)$.

Chapter 2

Circular–Circular Regression Models

2.1 Introduction

Some regression models in which both covariates and responses take values on the circle have been proposed in the literature. Rivest (1997) provided a model for predicting the y-direction using a rotation of the "decentred" x-angle, which was applied to the prediction of the direction of earthquake displacement in terms of the direction of steepest descent. Downs and Mardia (2002) proposed a regression model in which the regression curve is expressed as a form of the Möbius transformation or tangent function, with application to data on circadian biological rhythms and wind directions. See Fisher (1993, p.168) for earlier works on circular–circular regression models.

The Möbius transformation is well known as a mapping which carries the complex plane onto itself. With some restrictions on the parameters, this mapping maps, for example, the unit circle onto itself or the unit circle onto the real line. One of the earlier works in directional statistics in which the Möbius transformation appeared was given by McCullagh (1996). In this paper he discussed the connection between the real Cauchy distribution and the wrapped or circular Cauchy distribution via the Möbius transformation. The Möbius transformation was also used in the link functions of regression models by Downs and Mardia (2002) and Downs (2003). Minh and Farnum (2003) induced some probabilistic models on the circle by using a bilinear transformation which maps the real line onto the unit circle and is related to the Möbius transformation in form. Jones (2004) proposed the Möbius distribution on the disc which is generated by applying the Möbius transformation to the symmetric beta or Pearson type II distribution. McCullagh (1989) and Seshadri (1991) transformed their distributions via a one-to-one mapping which has the same form as the Möbius transformation and maps the interval $(-1, 1)$ onto itself.

The wrapped Cauchy distribution was used as a statistical model by Mardia (1972, p.56) and Mardia and Jupp (2000, p.51). Its distributional properties and estimation were investigated by Kent and Tyler (1988) and McCullagh (1996). McCullagh (1996) showed that the wrapped Cauchy distribution is obtained by applying a bilinear transformation to the Cauchy distribution on the real line and is closed under the Möbius transformation. It has the additive property and a central limit theorem holds for this distribution (Kolassa and McCullagh, 1990).

In this chapter we propose a new circular–circular regression model and study some properties, including estimation and testing procedures, of this model. Its regression curve is expressed as a form of the Möbius circle transformation, and the angular error is distributed as a wrapped Cauchy distribution.

In Section 2.2 some properties of the proposed model, including its regression curve, are investigated. In addition, we compare our regression model with some existing models. A bivariate circular distribution, which could be useful for our regression model, is presented in Section 2.3. Next Section 2.4 considers parameter estimation, the Fisher information matrix, and a test of independence for the proposed model. In Section 2.5 our model is applied to marine biology and wind direction data.

2.2 Circular regression model

Let responses $Y_1, \ldots, Y_n$ be independent, and let $x_1, \ldots, x_n$ be nonstochastic covariates which take values on the unit circle, Ω, in the complex plane. In the proposed regression model, the conditional distribution of Y_j given $x_1, \ldots, x_n$ has the wrapped Cauchy distribution with mean direction $\arg\{v(x_j)\}$ defined in Section 2.2.1 and mean resultant length $\varphi \in [0, 1]$.

In Section 2.2.1 we define the regression curve v and investigate its properties. Some properties of the regression model are discussed in Section 2.2.2. Comparison with existing regression models is given in Section 2.2.3.

2.2.1 Regression curve

Suppose β_0 and β_1 are complex parameters with $\beta_0 \in \Omega$ and $\beta_1 \in \mathbb{C}$. The regression curve of the proposed regression model is defined by

$$v = v(x; \beta_0, \beta_1) = \beta_0 \frac{x + \beta_1}{1 + \overline{\beta_1} x}, \quad x \in \Omega, \tag{2.2.1}$$

where the mapping with $|\beta_1| \neq 1$ will be called a Möbius circle transformation, and this transformation is a one-to-one mapping which carries the unit circle onto itself.

The Möbius circle transformation is obtained by a composition of transformations of the following four types:

(1) Translations: $z \to z + b$,

(2) Rotations: $z \to az, \quad a \in \Omega$,

(3) Homotheties: $z \to rz, \quad r > 0$,

(4) Inversion: $z \to 1/z$.

Note that these transformations exhibit the action of the group on the complex plane, not on the circle. For $\beta_1 \neq 0$, v can be expressed as

$$v = \beta_0 \left(\frac{1}{\overline{\beta_1}} + \frac{\lambda}{\overline{\beta_1} x + 1} \right), \quad \lambda = \beta_1 - \frac{1}{\overline{\beta_1}}.$$

In (2.2.1) β_0 is evidently a rotation parameter, but the interpretation of β_1 is more complicated. However, the function of β_1 in (2.2.1) for $|\beta_1| < 1$ is revealed as follows. Assume, without loss of generality, that $\beta_0 = 1$. Then, for any $\beta_1 \in \mathbb{C}$ and any $x \in \Omega$, (2.2.1) implies that β_1 is the projection point for the straight line projection of $-x$ on the unit circle to the point v on the unit circle. From this fact, β_1 can be intuitively interpreted as the parameter that attracts the points on the circle toward $\beta_1/|\beta_1|$ with the concentration of points about $\beta_1/|\beta_1|$ increasing as $|\beta_1|$ increases. An exception is the point $x = -\beta_1/|\beta_1|$, which is invariant under the Möbius circle transformation for any $|\beta_1| < 1$.

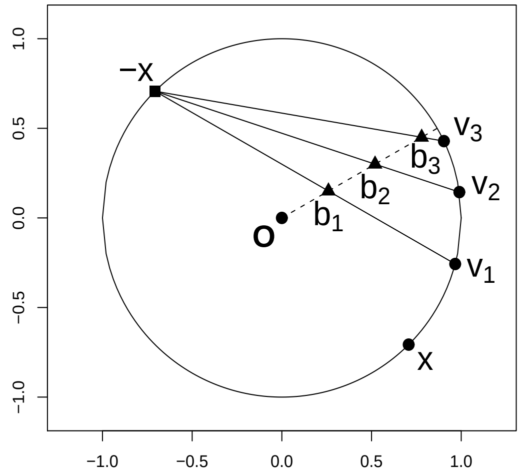
Figure 2.1(a) exhibits the behaviour of (2.2.1) for some specified values of β_1 for $|\beta_1| < 1$. Figure 2.1(a) explicitly shows that as $|\beta_1|$ approaches 1, v ($x \neq -\exp(\pi i/12)$) converges to a point $\beta_1/|\beta_1| = \exp(\pi i/12)$. It is also clear from the figure that as $|\beta_1|$ tends to 0, v approaches the identity mapping. When $|\beta_1| = 1$, the mapping (2.2.1) maps the unit circle onto the point β_1, i.e. $v = \beta_1$ for any x. For the case of $|\beta_1| > 1$, (2.2.1) can be expressed as

$$v = \beta_0 \frac{x + \beta_1}{1 + \overline{\beta_1} x} = \beta_0 \frac{\tilde{x} + \tilde{\beta}_1}{1 + \overline{\tilde{\beta}_1} \tilde{x}}, \tag{2.2.2}$$

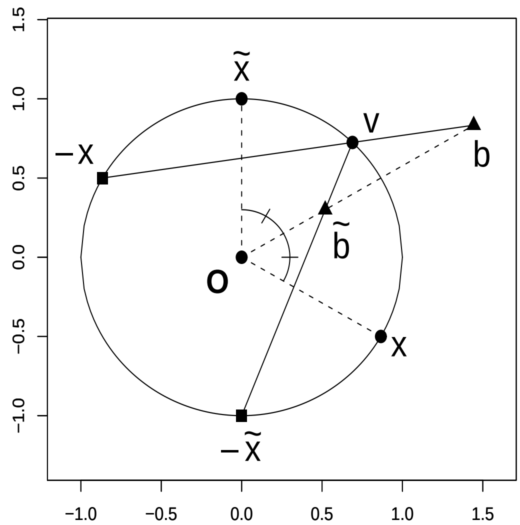
where $\tilde{x} = (\beta_1/|\beta_1|)\,(\beta_1 \overline{x}/|\beta_1|)$ and $\tilde{\beta}_1 = 1/\overline{\beta_1}$. The above expression (2.2.2) shows that the Möbius circle transformation with $|\beta_1| > 1$ consists of two types of transformations, namely, reflection and the Möbius circle transformation with $|\tilde{\beta}_1| < 1$, i.e.

$$x \longmapsto (\beta_1/|\beta_1|)\,(\beta_1 \overline{x}/|\beta_1|) \quad \text{and} \quad x \longmapsto \beta_0 (x + \tilde{\beta}_1)/(1 + \overline{\tilde{\beta}_1} x).$$

Figure 2.1(b) displays an example in which equation (2.2.2) holds for selected value of β_0, β_1 and x. The figure clearly shows the fact that the Möbius circle transformation with $|\beta_1| > 1$ is made up of the two transformations mentioned above.



(a)



(b)

Figure 2.1. a plot of $v(x;\beta_0,\beta_1)$ for regression curve (2.2.1) for $x = \exp(-\pi i/4)$ with $\beta_0 = 1$, $\arg(\beta_1) = \pi/6$ and: $|\beta_1| = 0.3$; $|\beta_1| = 0.6$; $|\beta_1| = 0.9$. Points on the plot are defined by $\mathrm{v}_j = v(x;1,\mathrm{b}_j)$, $\mathrm{b}_j = 0.3j\exp(\pi i/6)$, $j = 1,2,3$. (b) plot of $v, x, \tilde{x}, \beta_1, \tilde{\beta}_1$ for equation (2.2.2) for $\beta_0 = 1$, $x = \exp(-\pi i/6)$, $\beta_1 = 5\exp(\pi i/6)/3$. Parameters β_1 and $\tilde{\beta}_1$ are expressed as b and $\tilde{\mathrm{b}}$ on the plot, respectively.

2.2.2 Some properties of the proposed regression model

This subsection discusses some properties of the proposed regression model. For simplicity of expression, we consider a case in which a single pair of a covariate and a response are observed.

Let x be a covariate which takes values on the unit circle in the complex plane and let Y be a response. The complex parameters are $\beta_0 \in \Omega$ and $\beta_1 \in \mathbb{C}$. The proposed regression model is given by

$$Y = \beta_0 \frac{x + \beta_1}{1 + \overline{\beta_1} x}\, \varepsilon, \quad x \in \Omega, \tag{2.2.3}$$

where $\varepsilon \sim C^*(\varphi),\ 0 \leq \varphi \leq 1$. Here the restriction on the domain of φ is valid because the mean direction of the angular error should be 0 and $C^*(\varphi) = C^*(1/\overline{\varphi})$ holds for any $\varphi \in \mathbb{C}$. We have already discussed the interpretation of β_0 and β_1 in Section 2.2.1. The parameter φ is the concentration or precision parameter. If $\varphi = 1$, then covariates and responses are correlated without error. The smaller the value of φ the less concentrated the error variables. When $\varphi = 0$, the variable ε has a uniform distribution on the circle.

The conditional distribution of Y given x is

$$Y \mid x \sim C^*\left(\phi_{Y|x}\right) \quad \text{where } \phi_{Y|x} = \exp\left(i\mu_{Y.x}\right)\varphi \text{ and } \mu_{Y.x} = \arg\left(\beta_0 \frac{x + \beta_1}{1 + \overline{\beta_1} x}\right). \tag{2.2.4}$$

The following theorem holds for our regression model by applying well-known result in complex analysis. See Rudin (1987, Theorem 11.9) for the proof.

Theorem 2.1 *If $Y \sim C^*(\phi)$ where $|\phi| \leq 1$, then $E\{g(Y)\} = g(\phi)$ for any mapping g on the closed unit disc which is continuous on the closed unit disc and analytic on the open unit disc.*

Using the result we obtain the mean direction and the mean resultant length of $Y \mid x$,

$$\arg\{E(Y \mid x)\} = \mu_{Y.x} = \arg(\beta_0 x) - 2\arg\left(1 + \overline{\beta_1} x\right), \quad |E(Y \mid x)| = \varphi.$$

More generally, the kth trigonometric moment of $Y \mid x$ is

$$E\left(Y^k \mid x\right) = \phi_{Y|x}^k. \tag{2.2.5}$$

Since the wrapped Cauchy distribution is closed under rotation and the Möbius circle transformation (see properties (i) and (iii) in Section 2.2.2), we obtain

$$\gamma_0 \frac{Y+\gamma_1}{1+\overline{\gamma_1}Y} \,\bigg|\, (X=x) \sim C^*\left(\gamma_0 \frac{\phi_{Y|x}+\gamma_1}{1+\overline{\gamma_1}\phi_{Y|x}}\right), \tag{2.2.6}$$

where $\gamma_0 \in \Omega,\ \gamma_1 \in \mathbb{C}$. Because of the fact that the linear fractional transformations form a group under composition, the parameter of the wrapped Cauchy (2.2.6) can also be expressed as the linear fractional transformation

$$\gamma_0 \frac{\phi_{Y|x}+\gamma_1}{1+\overline{\gamma_1}\phi_{Y|x}} = \frac{a_{00}x+a_{01}}{a_{10}x+a_{11}},$$

where $a_{00} = \gamma_0(\beta_0\varphi + \gamma_1\overline{\beta_1}),\ a_{01} = \gamma_0(\gamma_1 + \beta_0\beta_1\varphi),\ a_{10} = \overline{\beta_1} + \overline{\gamma_1}\beta_0\varphi,\ a_{11} = 1 + \overline{\gamma_1}\beta_0\beta_1\varphi$.

Although property (2.2.6) is mathematically attractive, it is remarked here that the absolute values of the parameters in (2.2.6) depend on x and therefore homoscedasticity no longer holds. This formulation should be avoided unless heteroscedasticity is desired. To avoid this heteroscedasticity, one can transform Y to $W = \gamma_0\{(Y+\gamma_1)/(1+\overline{\gamma_1}Y)\}$ and then use the model set up by (2.2.3) and (2.2.4) for $W \mid x$. Similarly, the following property holds for the Möbius circle transformation of the covariate

$$Y \,\bigg|\, \gamma_0 \frac{x+\gamma_1}{1+\overline{\gamma_1}x} \sim C^*\left(\frac{b_{00}x+b_{01}}{b_{10}x+b_{11}}\right), \tag{2.2.7}$$

where $\gamma_0 \in \Omega,\ \gamma_1 \in \mathbb{C},\ b_{00} = \beta_0(1+\overline{\gamma_1}\beta_1)\varphi,\ b_{01} = \beta_0(\gamma_1+\beta_1)\varphi,\ b_{10} = \overline{\gamma_1}+\overline{\beta_1},\ b_{11} = 1 + \gamma_1\overline{\beta_1}$.

If we assume that x is an observed value of a random variable X which has the wrapped Cauchy distribution $C^*(\phi)$ and is independent of ε in (2.2.3), then the marginal distribution of Y is given by

$$Y \sim C^*\left(\beta_0 \frac{\phi+\beta_1}{1+\overline{\beta_1}\phi}\,\varphi\right). \tag{2.2.8}$$

The above is obvious from properties (i), (ii) and (iii) in Section 1.2.4.

2.2.3 Comparison with existing regression models

McCullagh (1996, Equation 28) proposed a regression model in which the error is assumed to follow a Cauchy distribution on the real line. Although his model looks similar to ours at first glance, his model and ours are different. His model is not circular–circular, but planar–linear regression model. In addition, our model is obtained neither by wrapping $Y \mid z$ nor by transforming $Y' = (1 + iY)/(1 - iY)$, which are the transformations to generate a wrapped Cauchy distribution from a Cauchy distribution on the real line.

Our proposed regression model also has some relationship with the models of Fisher and Lee (1992) and Downs and Mardia (2002). Fisher and Lee (1992) proposed a linear–circular regression model in which the link function is expressed as a form of tangent function. Tangent function is also used in the link function of the circular–circular regression model of Downs and Mardia (2002). After some algebra, it is shown that our regression curve corresponds to their link function. However our model is different from theirs. The major distinction is the distribution for the angular error. In their model the angular error assumes the von Mises distribution, whereas in our model we assume that the angular error is distributed as the wrapped Cauchy. Our model has some desirable properties that their model does not have such as Theorem 2.1 and properties (2.2.5)–(2.2.8).

2.3 Related bivariate circular distribution

To our knowledge, no bivariate angular distribution has been used to model circular–circular regression. We now provide a bivariate circular distribution which could be helpful in modelling our circular–circular regression. It has the density

$$f(x, y) = \frac{1}{(2\pi)^2} \frac{|1 - \varphi^2|}{\left|y - \phi_{Y|x}\right|^2} \frac{|1 - |\delta|^2|}{|x - \delta|^2}, \quad x, y \in \Omega, \tag{2.3.1}$$

where $|\delta| \neq 1$ and the other parameters are defined in the same way as in (2.2.3) and (2.2.4). The following properties hold for this distribution:

(1) $Y \,|\, (X = x) \sim C^* \left(\phi_{Y|x}\right)$,

(2) $Y \sim C^* \left(\beta_0 \dfrac{\delta + \beta_1}{1 + \overline{\beta_1}\delta} \varphi\right)$,

(3) $X \sim C^* (\delta)$.

Hence, the marginals and the conditional of Y given x are wrapped Cauchy distributions. The distribution (2.3.1) takes maximum (minimum) value for each x at $y = \exp(i\mu_{Y.x})$ $(\exp(-i\mu_{Y.x}))$. For $\beta_1 \in \Omega$, X and Y are independently distributed as $C^*(\delta)$ and $C^*(\beta_0\beta_1\varphi)$, respectively. The closer $|\beta_1|$ gets to 0, the closer $\exp(i\mu_{Y.x})$ is to being a pure rotation of X. For $\varphi = 0$, X and Y are independently distributed as $C^*(\delta)$ and the circular uniform distribution $C^*(0)$, respectively. The larger the value of φ, the greater the correlation between X and Y. See Fisher and Lee (1983) for a definition of circular correlation.

2.4 Estimation and test

2.4.1 Parameter estimation

Maximum likelihood estimation for the wrapped Cauchy distribution was investigated by Kent and Tyler (1988). However we cannot apply these results to the conditional distribution $Y \,|\, x$ directly, since the mean direction is a function of the covariate x. Therefore we need to obtain the maximum likelihood estimates of the wrapped Cauchy distribution in a different manner.

Let $Y_j \,|\, x_j$ $(j = 1, \ldots, n)$ be a set of random samples from the wrapped Cauchy distribution $C^*(\phi_{Y_j|x_j})$. The log-likelihood function for these samples is given by

$$\log L = C + \sum_{j=1}^{n} \left\{ \log\left|1 - \varphi^2\right| - 2\log\left|y_j - \beta_0(x_j + \beta_1)\varphi/(1 + \overline{\beta_1}x_j)\right| \right\}.$$

Transform the covariates and responses by equating $(x_j, Y_j) = (e^{i\theta_{x_j}}, e^{i\Theta_{Y_j}})$ and, for convenience, reparametrize $(\beta_0, \beta_1) = (e^{i\theta_0}, re^{i\theta_1})$ where $r > 0$, $0 \le \theta_0, \theta_1 < 2\pi$.

Then the log-likelihood function can be expressed as

$$\log L = C + n\log(1-\varphi^2) - \sum_{j=1}^{n} \log\left\{1 - 2\varphi\cos\left(\Theta_{Y_j} - \mu_{Y_j|x_j}\right) + \varphi^2\right\}, \qquad (2.4.1)$$

where $\mu_{Y_j|x_j} = \theta_0 + \theta_{x_j} - 2\arg\left\{1 + re^{i(\theta_{x_j} - \theta_1)}\right\}$.

If β_1 is known, the maximum likelihood estimates of θ_0 and φ are obtained by the recursive algorithm by Kent and Tyler (1988). The method of moments gives the estimators of θ_0 and φ as follows:

$$\hat{\theta}_0 = \arg\left(C_n + iS_n\right) \quad \text{and} \quad \hat{\varphi} = \frac{1}{n}\left|C_n + iS_n\right|,$$

where $C_n = \sum_{j=1}^{n} \cos[\Theta_{Y_j} - \theta_{x_j} + 2\arg\{1 + re^{i(\theta_{x_j} - \theta_1)}\}]$ and $S_n = \sum_{j=1}^{n} \sin[\Theta_{Y_j} - \theta_{x_j} + 2\arg\{1 + re^{i(\theta_{x_j} - \theta_1)}\}]$.

2.4.2 Fisher information matrix

Using the log-likelihood for $(\theta_0, r, \theta_1, \varphi)$ given by (2.4.1). We find that

$$-E\left(\frac{\partial^2}{\partial\theta_0\partial\varphi}\log L\right) = -E\left(\frac{\partial^2}{\partial r\partial\varphi}\log L\right) = -E\left(\frac{\partial^2}{\partial\theta_1\partial\varphi}\log L\right) = 0.$$

Hence, φ and (θ_0, r, θ_1) are orthogonal. The other elements of the Fisher information matrix are calculated as

$$\begin{aligned}
E\left\{\left(\frac{\partial}{\partial\theta_0}\log L\right)^2\right\} &= \frac{2n\varphi^2}{(1-\varphi^2)^2}, \\
E\left\{\left(\frac{\partial}{\partial r}\log L\right)^2\right\} &= \frac{2\varphi^2}{(1-\varphi^2)^2}\sum_{j=1}^{n}\left(\frac{\partial\mu_{Y_j|x_j}}{\partial r}\right)^2, \\
E\left\{\left(\frac{\partial}{\partial\theta_1}\log L\right)^2\right\} &= \frac{2\varphi^2}{(1-\varphi^2)^2}\sum_{j=1}^{n}\left(\frac{\partial\mu_{Y_j|x_j}}{\partial\theta_1}\right)^2, \\
E\left\{\left(\frac{\partial}{\partial\varphi}\log L\right)^2\right\} &= \frac{2n}{(1-\varphi^2)^2}, \\
E\left\{\left(\frac{\partial}{\partial\theta_0}\log L\right)\left(\frac{\partial}{\partial r}\log L\right)\right\} &= \frac{2\varphi^2}{(1-\varphi^2)^2}\sum_{j=1}^{n}\frac{\partial\mu_{Y_j|x_j}}{\partial r},
\end{aligned}$$

$$E\left\{\left(\frac{\partial}{\partial\theta_0}\log L\right)\left(\frac{\partial}{\partial\theta_1}\log L\right)\right\} = \frac{2\varphi^2}{(1-\varphi^2)^2}\sum_{j=1}^{n}\frac{\partial\mu_{Y_j|x_j}}{\partial\theta_1},$$
$$E\left\{\left(\frac{\partial}{\partial r}\log L\right)\left(\frac{\partial}{\partial\theta_1}\log L\right)\right\} = \frac{2\varphi^2}{(1-\varphi^2)^2}\sum_{j=1}^{n}\frac{\partial\mu_{Y_j|x_j}}{\partial r}\frac{\partial\mu_{Y_j|x_j}}{\partial\theta_1},$$

where

$$\frac{\partial\mu_{Y_j|x_j}}{\partial r} = \frac{-2\sin(\theta_{x_j}-\theta_1)}{1+2r\cos(\theta_{x_j}-\theta_1)+r^2}, \quad \frac{\partial\mu_{Y_j|x_j}}{\partial\theta_1} = \frac{2r\{r+\cos(\theta_{x_j}-\theta_1)\}}{1+2r\cos(\theta_{x_j}-\theta_1)+r^2}.$$

2.4.3 A test of independence

To investigate if the model (2.2.3) provides a better fit than the independence model, we test $H_0 : r = 1$ against $H_1 : r \neq 1$. The likelihood ratio test gives the test statistic as $T = -2\log(\max L_0/\max L_1)$, where $\max L_0 = \max_{\theta_0,\varphi} L\,(\theta_0,\varphi, r = 1, \theta_1 = 0)$ and $\max L_1 = \max_{\theta_0,r,\theta_1,\varphi} L\,(\theta_0, r, \theta_1, \varphi)$. Under the null hypothesis, T is asymptotically distributed as a chi-square distribution with two degrees of freedom. Here $\max L_0$ is easily obtained using the algorithm of Kent and Tyler (1988). We reject the null hypothesis when T is large.

Other large sample theories, such as Wald test and score test, could also be useful for inference for the proposed model.

2.5 Examples

Example 2.1 In a marine biology study by Dr. Robert R. Warner at University of California, Santa Barbara (Lund, 1999), whether the spawning time of a particular fish (T_S) depends on the time of the low tide (T_{LT}) is of interest. The data were gathered in St. Croix, the U.S. Virgin Islands. To study the dependence of T_S on T_{LT}, we converted the period 0 to 24 hours of T_S and T_{LT} to $[0, 2\pi)$. Paired T_S and T_{LT} are thus bivariate circular data, and they are plotted as circles in Figure 2.2. In the following, we apply model (2.4) to investigate whether and how T_S depends on T_{LT}.

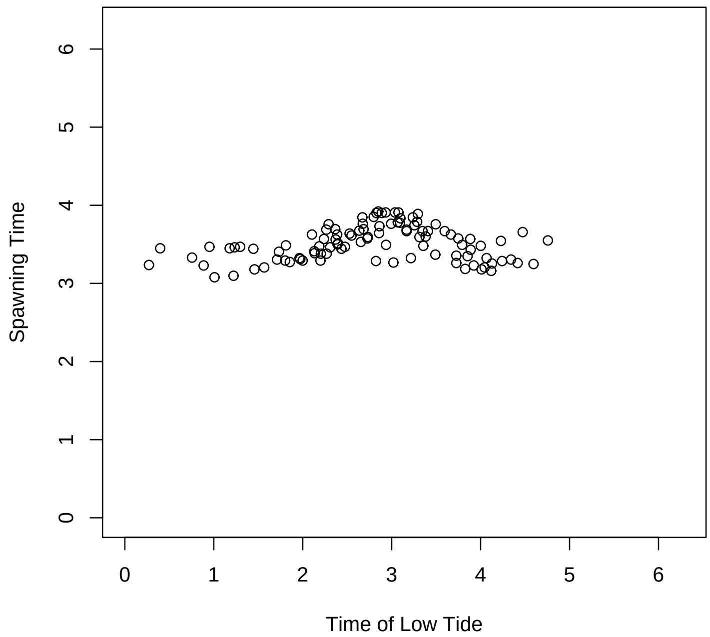


Figure 2.2. planar plot of the spawning time of certain fish and the time of low tide. Both times are converted into angles $[0, 2\pi)$.

The maximum likelihood estimates of the parameters are $\hat{\theta}_0 = 0.47$, $\hat{r} = 0.95$, $\hat{\theta}_1 = 3.06$ and $\hat{\varphi} = 0.87$. The maximum log-likelihood and AIC of the model are equal to -11.28 and 30.56, respectively. Approximate 90% confidence intervals for θ_0, r, θ_1 and φ are $(-0.11, 1.05)$, $(0.89, 1.00)$, $(2.46, 3.66)$ and $(0.84, 0.90)$ by the Fisher information matrix in Section 2.4.2. The test of independence for model (2.4) yields the test statistic $T = -2\{(-14.81) - (-11.28)\} = 7.06$. This test is highly significant and the assumption of independence is rejected. Circular distances of all observations lie in $[0, 0.25]$. Here the circular distance is defined by $d(Y, \hat{Y}) = 1 - \cos(Y - \hat{Y})$ where Y is a response and $\hat{Y}$ is a predictor in radians given by $\hat{Y} = \hat{\theta}_0 + x - 2\arg\{1 + re^{i(x-\hat{\theta}_1)}\}$.

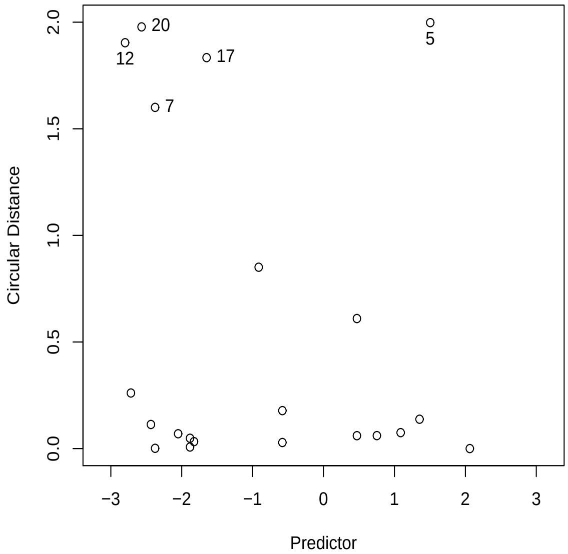
Example 2.2 The wind direction at 6 a.m. and 12 noon was measured each day at a weather station in Milwaukee for 21 consecutive days. (Johnson and Wehrly, 1977, Table 2). We use model (2.2.3) for regressing the wind direction at 12 noon on that at 6 a.m.

The maximum likelihood estimates of the parameters are $\hat{\theta}_0 = 1.27$, $\hat{r} = 0.53$, $\hat{\theta}_1 = 2.59$ and $\hat{\varphi} = 0.55$. The maximum log-likelihood and AIC of the model are -32.26 and 72.52, respectively. Approximate 90% confidence intervals for θ_0, r, θ_1 and φ are $(0.91, 1.63)$, $(0.31, 0.74)$, $(2.31, 2.87)$ and $(0.37, 0.73)$. Judging from the AIC, model (2.2.3) provides a better fit than the Downs and Mardia model, whose AIC is 74.56. The test of independence for the model (2.2.3) in Section 2.4.3 yields the test statistic as $T = -2\{(-38.48) - (-32.26)\} = 12.44$. This test is highly significant and the assumption of independence is rejected.

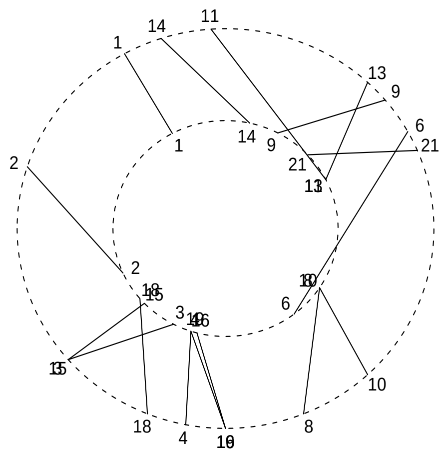
The plot of circular distances is given in Figure 2.3(a). The observation numbers of outliers are marked on the plot. Apart from five outliers, model (2.2.3) seems to provide a satisfactory fit to the data. Finally, the predictors and responses except for the outliers are plotted by observation numbers in Figure 2.3(b). The plots on the larger circle refer to the responses, while those on the smaller one are the predictors from model (2.2.3). A short line between the predictor and response means a good fit of the model to the observation. Judging from Figure 2.3(b), our model seems to provide satisfactory fit to the data. For the interpretation of how the responses are transformed through the Möbius circle transformation, see Section 2.2.1.

2.6 Discussion

Circular–circular regression is useful for analyzing bivariate circular data. Among existing regression models, the raison d'être of our model is its tractability and expandability. The tractability derives from the theory of the Möbius circle transformation and the wrapped Cauchy distribution. As discussed in Section 2.2.2 in this thesis, the wrapped Cauchy is related to the Möbius circle transformation, and



(a)



(b)

Figure 2.3. (a) plot of circular distance, and (b) plot of predictors and covariates, in which the predictors are plotted on the smaller circle whereas the responses are marked on the larger one.

thus enables us to obtain a number of desirable properties for our model. As for the expandability, our regression model could provide some topics to other related fields. For example, the related bivariate circular distribution, which is briefly discussed in Section 2.3, could be a possible field in which it is worth carrying out further research. It could be also interesting to investigate the properties of the regression model which has the angular error proposed by Jones and Pewsey (2005) instead of the wrapped Cauchy in (2.2.3). Their model includes the wrapped Cauchy and von Mises as special cases and might be used to discriminate in applications between these two distributions.

Chapter 3

Distributions for Angular Observations

3.1 Introduction

In directional statistics, various methods have been proposed in the literature to obtain distributions on manifolds such as the sphere, cylinder, disc or torus. A family of distributions with specified marginals on the cylinder was given by Johnson and Wehrly (1978). Wehrly and Johnson (1980) also proposed distributions on the torus constructed in a similar manner and studied a related Markov process on the circle and statistical inference for it. Shimizu and Iida (2002) proposed a Pearson type VII distribution on an arbitrary dimensional sphere by using scale mixtures of multivariate normal distributions. Jones (2002) proposed distributions on the disc with a single specified marginal. Another distribution on the disc was provided by Jones (2004), who generated the distribution through applying the Möbius transformation to the bivariate spherically symmetric beta or Pearson type II distribution. Multivariate distributions with specified conditionals on certain manifolds were discussed by SenGupta (2004). For other methods of obtaining multivariate distributions on certain manifolds, see Mardia and Jupp (2000) and Jammalamadaka and SenGupta (2001).

This chapter consists of two parts. The first provides four-dimensional distributions with specified bivariate marginals on certain manifolds such as two tori, cylinders or discs. These distributions provide models for observations which are represented as points on two bivariate manifolds. For example, in meteorology, two pairs of wind directions, which are measured at two locations at two points in time, are observations on two tori. Another example is given for two pairs of wind direction and speed, i.e., observations on two cylinders. Or observations may lie on two discs. Certain properties of the submodel of the proposed models are investigated. We also study a related Markov process on bivariate manifolds. Then models for incomplete observations are constructed. They can be applied in the modelling of two observations such as one on the circle and the other on the torus. Distributions with specified marginals on the disc are also discussed. The second part of the chapter discusses two distributions on the cylinder. The first distribution was originally proposed by Johnson and Wehrly (1978) as a maximum entropy distribution subject to constraints on certain moments. We investigate other properties of this distribution. The second distribution is a new distribution which is obtained as a maximum entropy distribution or a conditional distribution of a trivariate normal distribution. This distribution can be viewed as an extension of the distribution by Mardia and Sutton (1978). The common property of these two distributions is that the circular marginal of each model is distributed as Maksimov's (1967) generalized von Mises distribution. The distribution includes the von Mises distribution as a special case and has a unimodal or bimodal, symmetric or asymmetric shape depending on the choice of parameters. Occasionally, real circular data can exhibit bimodal. For example, data on the orientation of dragonflies of the genus Sympetrum with respect to the sun's azimuth as given in Example 1.6.2. of Batschelet (1981) are bimodal. Other datasets are asymmetrically distributed. Pewsey (2000) proposed a model for asymmetrically distributed data from a study of bird migration. When both a circular variable and linear one are considered jointly, the corresponding data are distributed on a cylinder with, potentially, an asymmetrical or bimodal (or both) circular component. The two distributions discussed may provide useful models for analyzing this kind of cylindrical data.

Subsequent subsections are organized as follows. Section 3.2 provides a theorem which constructs families of four-dimensional distributions on certain manifolds with specified bivariate marginal distributions. A related Markov process and models for incomplete observations are constructed. An investigation of some of the properties of a submodel of the proposed four-dimensional distributions is then presented. In addition distributions with specified marginals on the disc are constructed using the theorem from Section 3.2. They include the bivariate spherically symmetric beta distributions on the disc. In Section 3.3 the properties of the two distributions on the cylinder described in the preceding paragraph are studied, and the proposed model is applied to two cylindrical datasets. The chapter closes with an appendix which considers the modality of the cylindrical distribution and the derivation of the density of the other.

3.2 Distributions with specified marginal distributions

3.2.1 Definition of the proposed models

The following theorem provides continuous distributions on certain manifolds with specified bivariate marginal distributions. The theorem is applicable to the construction of distributions on two tori, cylinders and discs.

Theorem 3.1 *Let* (X_1, X_2) *have a specified density* $f_1(x_1, x_2)$ *and distribution function (d.f.)* $F_1(x_1, x_2)$ *on the support* $M_1 \subset \mathbb{R}^2$ *and* (Y_1, Y_2) *have a specified density* $f_2(y_1, y_2)$ *and d.f.* $F_2(y_1, y_2)$ *on the support* $M_2 \subset \mathbb{R}^2$*. Suppose that* $f_{11}(x_1)$ $(f_{21}(y_1))$ *is the marginal density of* X_1 (Y_1) *and* $F_{11}(x_1)$ $(F_{21}(y_1))$ *its d.f. Let* $g(\cdot)$ *be a density on the circle and* $h(\cdot, \cdot)$ *a density on the torus. Then*

$$p_{1\pm}(x_1, x_2, y_1, y_2) = 2\pi f_1(x_1, x_2) f_2(y_1, y_2)$$

$$\times g\left[2\pi\left\{\frac{1}{f_{11}(x_1)}\frac{\partial F_1(x_1,x_2)}{\partial x_1}\pm\frac{1}{f_{21}(y_1)}\frac{\partial F_2(y_1,y_2)}{\partial y_1}\right\}\right], \quad (3.2.1)$$

$$\begin{aligned} p_2(x_1,x_2,y_1,y_2) &= 4\pi^2 f_1(x_1,x_2)f_2(y_1,y_2) \\ &\times h\Bigg[2\pi\left\{F_{11}(x_1)\pm F_{21}(y_1)\right\}, \\ &\qquad 2\pi\left\{\frac{1}{f_{11}(x_1)}\frac{\partial F_1(x_1,x_2)}{\partial x_1}\pm\frac{1}{f_{21}(y_1)}\frac{\partial F_2(y_1,y_2)}{\partial y_1}\right\}\Bigg] \quad (3.2.2)\end{aligned}$$

are densities on $M_1\times M_2$*, where* $(x_1,x_2)\in M_1$*,* $(y_1,y_2)\in M_2$*. Both have the marginal distributions of* (X_1,X_2) $((Y_1,Y_2))$ *with density* $f_1(x_1,x_2)$ $(f_2(y_1,y_2))$*. Here the right hand side of (3.2.2) permits the four combinations of signs, i.e.* $(+,+)$*,* $(+,-)$*,* $(-,+)$ *and* $(-,-)$*.*

Proof It is clear that $p_{1\pm}(x_1,x_2,y_1,y_2)\ge 0$. We show

$$\int_{M_2}\int_{M_1} p_{1\pm}(x_1,x_2,y_1,y_2)dx_1dx_2dy_1dy_2 = 1.$$

Consider the integral

$$\begin{aligned} &\int_{M_2}\int_{M_1} p_{1\pm}(x_1,x_2,y_1,y_2)dx_1dx_2dy_1dy_2 \\ &= 2\pi\int_{M_2} f_2(y_1,y_2)\int_{M_1} f_1(x_1,x_2) \\ &\quad\times g\left[2\pi\left\{\frac{1}{f_{11}(x_1)}\frac{\partial F_1(x_1,x_2)}{\partial x_1}\pm\frac{1}{f_{21}(y_1)}\frac{\partial F_2(y_1,y_2)}{\partial y_1}\right\}\right]dx_1dx_2dy_1dy_2. \end{aligned} \quad (3.2.3)$$

Making the change of the variable $t=t(x_2)=2\pi\{\partial F_1(x_1,x_2)/\partial x_1\}/f_{11}(x_1)$, (3.2.3) is calculated as

$$\begin{aligned} &\int_{M_2}\int_{M_1} p_{1\pm}(x_1,x_2,y_1,y_2)dx_1dx_2dy_1dy_2 \\ &= \int_{M_2} f_2(y_1,y_2)\int_{\mathbb{R}}\int_0^{2\pi} g\left\{t\pm\frac{2\pi}{f_{21}(y_1)}\frac{\partial F_2(y_1,y_2)}{\partial y_1}\right\} f_{11}(x_1)dtdx_1dy_1dy_2 \\ &= \int_{M_2} f_2(y_1,y_2)\int_{\mathbb{R}} f_{11}(x_1)dx_1dy_1dy_2 \end{aligned}$$

$$
\begin{aligned}
&= \int_{M_2} f_2(y_1, y_2) dy_1 dy_2 \\
&= 1.
\end{aligned}
$$

From this result, it is obvious that the marginal density of (Y_1, Y_2) is $f_2(y_1, y_2)$. Similarly, we can show that the marginal density of (X_1, X_2) is $f_1(x_1, x_2)$.

By making the change of the variables $t_1 = t_1(x_2) = 2\pi\{\partial F_1(x_1, x_2)/\partial x_1\}/f_{11}(x_1)$ and $t_2 = t_2(x_1) = 2\pi\{F_{11}(x_1) \pm F_{21}(y_1)\}$, we can also show that

$$
\int_{M_2} \int_{M_1} p_2(x_1, x_2, y_1, y_2) dx_1 dx_2 dy_1 dy_2 = 1.
$$

The marginal density can be obtained using a similar approach to that used above. □

Example When $M_1 = M_2 = [0, 2\pi) \times \mathbb{R}$ for the distributions (3.2.1) and (3.2.2), distributions on two cylinders are constructed. For some examples of distributions on the cylinder, see Section 3.3. When $M_1 = M_2 = [0, 2\pi)^2$, distributions on two tori are constructed. Distributions on two tori or discs can be obtained in a similar manner.

It is not necessary for M_1 and M_2 to be the same manifolds. Actually, on specifying M_1 as a torus and M_2 a cylinder, we obtain families of distributions on the direct product of the torus and the cylinder. Similarly, distributions on the direct product of the torus and the disc can be constructed.

Remark As in Wehrly and Johnson (1980), families of distributions for a Markov process can be constructed using Theorem 3.1, as follows.

Let $\{X_i\}$, $X_i = (X_{i1}, X_{i2})'$, $i = 0, 1, \ldots$ be random variables taking values on $M \subset \mathbb{R}^2$ such that

$$
\begin{aligned}
p(x_0) &= f(x_{01}, x_{02}), \\
p(x_n | x_0, x_1, \ldots, x_{n-1}) &= p(x_n | x_{n-1}) \\
&= 2\pi f(x_{n1}, x_{n2})\; g\Bigg[2\pi\Bigg\{\frac{1}{f_1(x_{n1})}\frac{\partial F(x_{n1}, x_{n2})}{\partial x_{n1}} \\
&\qquad \pm \frac{1}{f_1(x_{n-1,1})}\frac{\partial F(x_{n-1,1}, x_{n-1,2})}{\partial x_{n-1,1}}\Bigg\}\Bigg], \quad n = 1, 2, \ldots,
\end{aligned}
$$

where $f(x_{n1}, x_{n2})$ denotes a density on M, $F(x_{n1}, x_{n2})$ its d.f., $f_1(x_{n1})$ the marginal density of X_{n1}, and $g(\cdot)$ a density on the circle. Then $\{X_i\}$ is a Markov process on M with initial distribution $p(x_0)$ and the stationary transition density $p(x_n|x_{n-1})$.

Similarly, we can construct an alternative Markov process using (3.2.2) as the stationary density.

Remark The following corollary provides models for two observations such as one on the circle and the other on the torus. These may be useful, in meteorology and the environmental sciences, for example, as models for a model for incomplete pairs of wind directions observed at two locations at two points in time.

Corollary 3.1 *Let* (X_1, X_2) *have a specified density* $f_1(x_1, x_2)$ *and d.f.* $F_1(x_1, x_2)$ *on the support* $M_1 \subset \mathbb{R}^2$ *and* Y *a specified density* $f_2(y)$ *and d.f.* $F_2(y)$ *on the support* $M_2 \subset \mathbb{R}$. *Suppose that* $f_{11}(x_1)$ *is the marginal density of* X_1. *Let* $g(\cdot)$ *be a density on the circle. Then*

$$\begin{aligned} p(x_1, x_2, y) &= 2\pi f_1(x_1, x_2) f_2(y) \\ &\quad \times g\left[2\pi\left\{\frac{1}{f_{11}(x_1)} \frac{\partial F_1(x_1, x_2)}{\partial x_1} \pm F_2(y)\right\}\right] \end{aligned}$$

is a density on $M_1 \times M_2$, *where* $(x_1, x_2) \in M_1$, $y \in M_2$. *It has the marginal distribution of* (X_1, X_2) (Y) *with the density* $f_1(x_1, x_2)$ $(f_2(y))$.

Note that Corollary 3.1 can be considered a special case of Theorem 3.1. Actually, on putting $M_2 = [0, 1) \times M$, $M \subset \mathbb{R}$, and $f_2(y_1, y_2) = f_2(y_2)$, $y_2 \in M$ in Theorem 3.1, we obtain Corollary 3.1.

3.2.2 Properties of the proposed models

In this subsection we investigate some properties of the distributions proposed in Theorem 3.1. We focus on distributions with density (3.2.1) and discuss some dependence properties for them. Clearly, if g in (3.2.1) is uniformly distributed, (X_1, X_2) and (Y_1, Y_2) are independent and distributed as $f_1(x_1, x_2)$ and $f_2(y_1, y_2)$, respectively.

When g is not uniform, (3.2.1) describes a distribution for which there is association between (X_1, X_2) and (Y_1, Y_2).

Here we discuss a submodel of the distribution with density (3.2.1) defined as follows. Suppose that X_2 and Y_2 take values on the circle $[0, 2\pi)$ and X_1 and Y_1 are defined on M_{11} and M_{21} ($M_{11}, M_{21} \subset \mathbb{R}$), respectively. Assume that the conditionals of $X_2|(X_1 = x_1)$ and $Y_2|(Y_1 = y_1)$ are unimodal and symmetric with mode at μ_1 and μ_2 ($0 \leq \mu_1, \mu_2 < 2\pi$), respectively. The function g in (3.2.1) is assumed to follow a cardioid distribution with density

$$g(\theta) = \frac{1}{2\pi}\left\{1 + 2\rho\cos(\theta - \mu)\right\}, \quad 0 \leq \theta < 2\pi;\ 0 \leq \rho \leq \tfrac{1}{2},\ -\pi \leq \mu < \pi. \qquad (3.2.4)$$

The d.f.'s of (X_1, X_2) and (Y_1, Y_2) are defined as

$$F_1(x_1, x_2) = \int_{-\infty}^{x_1}\int_A f_1(u_1, u_2)du_2du_1, \quad F_2(y_1, y_2) = \int_{-\infty}^{y_1}\int_B f_2(v_1, v_2)dv_2dv_1,$$

where $A = [\mu_1, x_2 + 2\pi k_1)$, $B = [\mu_2, y_2 + 2\pi k_2)$, $k_1 = [(\mu_1 - x_2)/(2\pi) + 1]$, $k_2 = [(\mu_2 - y_2)/(2\pi) + 1]$, and $[a] = \max\{n \in \mathbb{Z}\,;\, n \leq a\}$. We write $p_{1c\pm}$ to denote the density of this distribution.

The cardioid distribution is a unimodal distribution on the circle. It is symmetric about μ and takes its maximum value at $\theta = \mu$ and minimum value at $\theta = \mu + \pi$. If $\rho = 0$, the distribution is the circular uniform. As ρ increases the distribution becomes more concentrated around μ.

There are certain distributions which are suitable for (X_1, X_2) ((Y_1, Y_2)) in $p_{1c\pm}$. When (X_1, X_2) are distributed on the cylinder, the distribution of Johnson and Wehrly (1978) discussed in Section 3.3 and the one proposed in Theorem 1 of their paper might be useful. The conditionals of the circular component of these models, $X_2|(X_1 = x_1)$, are symmetric about a constant μ. As for distributions on the disc, the Möbius distribution (Jones, 2004) has a symmetric conditional for $X_2|(X_1 = x_1)$ with its mean direction not depending on x_1. The submodel of the distribution proposed by Mardia (1975, Equation 2.12) with $b = d = \mu = 0$ might be used for the distribution on the torus.

The following property holds for the model with density $p_{1c\pm}$.

Theorem 3.2 *Let (X_1, X_2, Y_1, Y_2) be a random vector which has the density $p_{1c\pm}$ and let $\mu = 0$. Then the following inequalities hold for (X_1, X_2, Y_1, Y_2):*

$$\begin{aligned}
&P(X_1 \in E_1,\, -x_2 < X_2 - \mu_1 \leq x_2,\, Y_1 \in E_2,\, -y_2 < Y_2 - \mu_2 \leq y_2) \\
&\qquad \geq P(X_1 \in E_1,\, -x_2 < X_2 - \mu_1 \leq x_2)\, P(Y_1 \in E_2,\, -y_2 < Y_2 - \mu_2 \leq y_2), \qquad (3.2.5) \\
&P(X_1 \in E_1,\, -x_2 < X_2 - \mu_1 + \pi \leq x_2,\, Y_1 \in E_2,\, -y_2 < Y_2 - \mu_2 + \pi \leq y_2) \\
&\qquad \geq P(X_1 \in E_1,\, -x_2 < X_2 - \mu_1 + \pi \leq x_2)\, P(Y_1 \in E_2,\, -y_2 < Y_2 - \mu_2 + \pi \leq y_2), \\
&P(X_1 \in E_1,\, -x_2 < X_2 - \mu_1 + \pi \leq x_2,\, Y_1 \in E_2,\, -y_2 < Y_2 - \mu_2 \leq y_2) \\
&\qquad \leq P(X_1 \in E_1,\, -x_2 < X_2 - \mu_1 + \pi \leq x_2)\, P(Y_1 \in E_2,\, -y_2 < Y_2 - \mu_2 \leq y_2), \\
&P(X_1 \in E_1,\, -x_2 < X_2 - \mu_1 \leq x_2,\, Y_1 \in E_2,\, -y_2 < Y_2 - \mu_2 + \pi \leq y_2) \\
&\qquad \leq P(X_1 \in E_1,\, -x_2 < X_2 - \mu_1 \leq x_2)\, P(Y_1 \in E_2,\, -y_2 < Y_2 - \mu_2 + \pi \leq y_2),
\end{aligned}$$

for any $E_1 \in \mathcal{B}(M_{11}),\ E_2 \in \mathcal{B}(M_{21}),\ 0 \leq x_2, y_2 \leq \frac{1}{2}\pi$, where $\mathcal{B}(M)$ denoting a Borel set of M.

Proof We prove inequality (3.2.5) for the density p_{1c+}. On setting

$$t_1(u_2) = \frac{2\pi}{f_{11}(u_1)} \frac{\partial}{\partial u_1} F_1(u_1, u_2 + \mu_1), \quad t_2(v_2) = \frac{2\pi}{f_{21}(v_1)} \frac{\partial}{\partial v_1} F_2(v_1, v_2 + \mu_2),$$

the left-hand side of (3.2.5) can be expressed as

$$\begin{aligned}
&\int_{E_1} \int_{E_2} \int_{-x_2+\mu_1}^{x_2+\mu_1} \int_{-y_2+\mu_2}^{y_2+\mu_2} p_{1c+}(u_1, u_2, v_1, v_2) dv_2 du_2 dv_1 du_1 \\
&\qquad = \frac{1}{4\pi^2} \int_{E_1} \int_{E_2} \int_{-s_{x_2}}^{s_{x_2}} \int_{-s_{y_2}}^{s_{y_2}} 1 + 2\rho \cos(t_1 + t_2) dt_2 dt_1 f_{11}(u_1) f_{21}(v_1) dv_1 du_1,
\end{aligned} \tag{3.2.6}$$

where s_{x_2} and s_{y_2} satisfy

$$s_{x_2} = \frac{2\pi}{f_{11}(u_1)} \frac{\partial}{\partial u_1} F_1(u_1, x_2 + \mu_1), \quad s_{y_2} = \frac{2\pi}{f_{21}(v_1)} \frac{\partial}{\partial v_1} F_2(v_1, y_2 + \mu_2).$$

From the symmetry and unimodality of $X_2|(X_1 = x_1)$ and $Y_2|(Y_1 = y_1)$, it follows that $0 \le s_{x_2}, s_{y_2} \le \pi$. Using

$$\int_{-s_{x_2}}^{s_{x_2}} \int_{-s_{y_2}}^{s_{y_2}} \cos(t_1 + t_2) dt_1 dt_2 = 4 \sin s_{x_2} \sin s_{y_2} \ge 0, \quad 0 \le s_{x_2}, s_{y_2} \le \pi,$$

then

$$\begin{aligned} &\int_{E_1} \int_{E_2} \int_{-x_2+\mu_1}^{x_2+\mu_1} \int_{-y_2+\mu_2}^{y_2+\mu_2} p_{1c+}(u_1, u_2, v_1, v_2) dv_2 du_2 dv_1 du_1 \\ &\quad \ge \frac{1}{4\pi^2} \int_{E_1} \int_{E_2} \int_{-s_{x_2}}^{s_{x_2}} \int_{-s_{y_2}}^{s_{y_2}} 1 \; dt_2 dt_1 f_{11}(u_1) f_{21}(v_1) dv_1 du_1 \\ &\quad = \int_{E_1} \int_{E_2} \int_{-x_2+\mu_1}^{x_2+\mu_1} \int_{-y_2+\mu_2}^{y_2+\mu_2} f_1(u_1, u_2) f_2(v_1, v_2) dv_2 du_2 dv_1 du_1. \end{aligned}$$

This equals the right-hand side of inequality (3.2.5).

The other inequalities for p_{1c+} and those for p_{1c-} are proved in a similar manner.

□

Remark If we assume that $\mu = \pi$, then the signs of all the inequalities are reversed.

Note that the results above are easily applied to the submodel of the bivariate circular distribution proposed by Wehrly and Johnson (1980) with density

$$f_{\pm}(\theta_1, \theta_2) = f_1(\theta_1) f_2(\theta_2) g\left[2\pi \left\{F_1(\theta_1) \pm F_2(\theta_2)\right\}\right], \quad 0 \le \theta_1, \theta_2 < 2\pi, \tag{3.2.7}$$

where f_1 and f_2 are circular densities which are unimodal and symmetric with mode at μ_1 and μ_2 ($0 \le \mu_1, \mu_2 < 2\pi$), respectively. The d.f. of f_1 (f_2) is given by F_1 (F_2) and is defined in a similar way as in $p_{1c\pm}$. Here we suppose that g is distributed as a cardioid distribution with $\mu = 0$. Then the following inequality holds for model (3.2.7).

$$\begin{aligned} &P(-\theta_1 < \Theta_1 - \mu_1 \le \theta_1, \ -\theta_2 < \Theta_2 - \mu_2 \le \theta_2) \\ &\qquad \ge P(-\theta_1 < \Theta_1 - \mu_1 \le \theta_1) \, P(-\theta_2 < \Theta_2 - \mu_2 \le \theta_2), \end{aligned}$$

for any $0 < \theta_1, \theta_2 < \pi/2$.

Similarly, the results analogous to the second through to the fourth inequalities in Theorem 3.2 can be derived.

Next, we discuss a test of independence of the distribution (3.2.1) by applying a result from Wehrly and Johnson (1980). Let $(X_{1j}, X_{2j}, Y_{1j}, Y_{2j})$ $(j = 1, \ldots, n)$ be a random sample from (3.2.1). Suppose that $f_1(x_1, x_2)$ and $f_2(y_1, y_2)$ are completely specified. Let g be the density of the von Mises distribution $\mathrm{VM}(\mu, \kappa)$

The test of independence can be expressed as a test of: $H_0 : \kappa = 0$ against $H_1 : \kappa > 0$. The likelihood ratio test is based on

$$T_n = \sum_{j=1}^{n} \cos\left(s_j - \hat{\mu}\right),$$

where s_j is given by

$$s_j = 2\pi \left\{ \frac{1}{f_{11}(x_{1j})} \frac{\partial F_1(x_{1j}, x_{2j})}{\partial x_{1j}} \pm \frac{1}{f_{21}(y_{1j})} \frac{\partial F_2(y_{1j}, y_{2j})}{\partial y_{1j}} \right\},$$

and $\hat{\mu}$ is given by $\hat{\mu} = \arg(C_n + iS_n)$ where $C_n = \sum_{j=1}^{n} \cos s_j$, $S_n = \sum_{j=1}^{n} \sin s_j$. According to Self and Liang (1987) (see also Shieh and Johnson (2005)), for sufficiently large n, $-2\log T_n$ is approximately distributed as a half and half mixture of zero and a chi-square random variable, i.e. $Z^2 I[Z > 0]$ where Z has a standard normal distribution and I is an indicator function.

If μ is known, this test is uniformly most powerful (see Jammalamadaka and SenGupta (2001, Section 5.2.3)).

3.2.3 Distributions on the disc

Jones (2002) proposed a class of distributions on the disc with a single specified marginal density $f_X(x)$ and a conditional distribution with density $f(y|x)$. The joint density is given, trivially, by $f(x, y) = f_X(x)f(y|x)$. Jones (2004) also provided another distribution on the disc, referred to him as the Möbius distribution on the disc. It is generated through applying the Möbius transformation to the bivariate spherically symmetric beta or Pearson type II distribution.

Classes of distributions with specified marginals on the cylinder and the torus were discussed in Johnson and Wehrly (1978) and Wehrly and Johnson (1980), respectively. In this subsection we discuss a class of distributions with two specified marginals on the disc, which could be useful for the marginals of the distribution we proposed in Section 3.2.1. The proposed model is defined, as follows, in a very similar way to that used in Wehrly and Johnson (1980).

Let $f_1(r)$ be a specified density on $[0,1)$, $f_2(\theta)$ a specified density on $[0,2\pi)$, and $F_1(r)$ and $F_2(\theta)$ their d.f.'s, respectively. Let $g(\cdot)$ be a density on the circle. Then

$$f(r,\theta) \;=\; 2\pi f_1(r) f_2(\theta) g\left[2\pi\left\{F_1(r) \pm F_2(\theta)\right\}\right], \quad 0 \le r < 1,\ 0 \le \theta < 2\pi, \tag{3.2.8}$$

is a density on the disc with the marginal densities $f_1(r)$ and $f_2(\theta)$.

This density can be obtained by applying Theorem 3.1. Actually, on putting $M_1 = [0,1)^2$, $M_2 = [0,1) \times [0,2\pi)$, $f_1(x_1,x_2) = f_1(x_2)$, $0 \le x_1, x_2 < 1$ and $f_2(y_1,y_2) = f_2(y_2)$, $0 \le y_1 < 1$, $0 \le y_2 < 2\pi$ in (3.2.1), we obtain distributions with the density (3.2.8). Distributions on the disc obtained in this manner include the bivariate spherically symmetric beta distribution with density

$$f(r,\theta) = \frac{\alpha}{\pi} r(1-r^2)^{\alpha-1}, \quad 0 \le r < 1,\ 0 \le \theta < 2\pi,$$

where $\alpha > 0$. It is obtained by setting $f_1(r) = 2\alpha r(1-r^2)^{\alpha-1}$, $f_2(\theta) = 1/(2\pi)$, and $g(t) = 1/(2\pi)$ in (3.2.8).

We next discuss certain dependence properties of a submodel of (3.2.8). A straightforward modification of Theorem 3.2 gives the following result.

Corollary 3.2 *Let g be the density of the cardioid distribution (3.2.4) and let $\mu = 0$. Suppose that $f_2(\theta)$ is symmetric about μ_1 ($\in [0,2\pi)$). Define the d.f.'s of $f_1(r)$ and $f_2(\theta)$ by $F_1(r) = \int_0^r f_1(u)du$ and $F_2(\theta) = \int_A f_2(v)dv$, respectively, where $A = [\mu_1, \theta + 2\pi k)$, $k = [(\mu_1 - \theta)/(2\pi) + 1]$. Then the following inequalities hold for (3.2.8):*

$$P(R \le r_1,\, -\theta < \Theta - \mu_1 \le \theta) \ge P(R \le r_1) P(-\theta < \Theta - \mu_1 \le \theta),$$

$$P(R \le r_1,\ -\theta < \Theta - \mu_1 + \pi \le \theta) \le P(R \le r_1)P(-\theta < \Theta - \mu_1 + \pi \le \theta),$$

$$P(R \le r_2,\ -\theta < \Theta - \mu_1 \le \theta) \le P(R \le r_2)P(-\theta < \Theta - \mu_1 \le \theta),$$

$$P(R \le r_2,\ -\theta < \Theta - \mu_1 + \pi \le \theta) \ge P(R \le r_2)P(-\theta < \Theta - \mu_1 + \pi \le \theta),$$

where $0 \le r_1 < s_r \le r_2 < 1,\ 0 \le \theta < \frac{1}{2}\pi,\ F_1(s_r) = \frac{1}{2}$.

We briefly discuss the relationship between the distribution with density (3.2.8) and copulas (e.g., Nelsen (1998) and Drouet Mari and Kotz (2001)). The distribution (3.2.8) can be viewed as a copula with density $c(u,v) = g\left\{2\pi(u \pm v)\right\},\ 0 \le u, v \le 1$, by transforming $U = F_1(R)$ and $V = F_2(\Theta)$. However copulas do not usually assume that either u or v is a periodic variable. Since g is a periodic function, it is natural to assume that one or more variables in u and v are periodic.

3.3 Distributions on the cylinder

3.3.1 A further study of the Johnson and Wehrly model

The purpose of this subsection is to discuss two distributions on the cylinder with a possibly asymmetric or bimodal marginal distribution for the circular component. The first model was proposed by Johnson and Wehrly (1978) and its property was briefly discussed in their paper. In this subsection we investigate its properties further. The second distribution on the cylinder is a generalization of the distribution by Mardia and Sutton (1978). We discuss the marginal and conditional distributions, maximum likelihood estimation and testing of this second distribution and apply it in the modelling of two datasets.

The model proposed by Johnson and Wehrly (1978) was based on the principle of maximum entropy subject to constraints on certain moments. In Theorem 2 in their paper, they introduced a distribution on the cylinder with density

$$f(\theta, x) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\kappa^2/(4\sigma^2)} C_2 \exp\left\{-\frac{(x-\lambda)^2}{2\sigma^2} + \frac{\kappa x}{\sigma^2}\cos(\theta - \mu)\right\}, \qquad (3.3.1)$$

where $0 \le \theta < 2\pi$, $-\infty < x < \infty$, $-\infty < \lambda < \infty$, $\kappa, \sigma > 0$ and $0 \le \mu < 2\pi$. As in Yfantis and Borgman (1982), C_2 is obtained from

$$C_2^{-1} = 2\pi \left\{ I_0\left(\frac{\kappa\lambda}{\sigma^2}\right) I_0\left(\frac{\kappa^2}{4\sigma^2}\right) + 2\sum_{j=1}^{\infty} I_j\left(\frac{\kappa^2}{4\sigma^2}\right) I_{2j}\left(\frac{\kappa\lambda}{\sigma^2}\right) \right\}. \tag{3.3.2}$$

Johnson and Wehrly (1978) showed that the conditional distribution of X given $\Theta = \theta$ is $N\left(\lambda + \kappa\cos(\theta - \mu), \sigma^2\right)$ and that of Θ given $X = x$ is a von Mises distribution with density

$$f(\theta|x) = \frac{1}{2\pi I_0(\kappa x/\sigma^2)} \exp\left\{\frac{\kappa x}{\sigma^2}\cos(\theta - \mu)\right\}, \quad 0 \le \theta < 2\pi.$$

The marginal density of Θ can be expressed as

$$f(\theta) = C_2 \exp\left[\frac{\kappa\lambda}{\sigma^2}\cos(\theta - \mu) + \frac{\kappa^2}{4\sigma^2}\cos\left\{2(\theta - \mu)\right\}\right], \quad 0 \le \theta < 2\pi. \tag{3.3.3}$$

This marginal density is actually the same as that of the generalized von Mises distribution $\mathrm{GVM}(\mu, \mu, \kappa\lambda/\sigma^2, \kappa^2/(4\sigma^2))$. The distribution with density (3.3.3) is symmetric about μ. It becomes unimodal or bimodal depending on the choice of the parameters. When $\lambda > \kappa > 0$, it is unimodal with mode at μ and antimode at $\mu + \pi$. When κ is larger than λ, the distribution is bimodal with modes at μ and $\mu + \pi$ and antimodes at $\mu + \pi \pm \arccos(\lambda/\kappa)$. When $\lambda = 0$, it reduces to a bimodal circular distribution and C_2 reduces to $C_2^{-1} = 2\pi I_0\left\{\kappa^2/(4\sigma^2)\right\}$.

The marginal density of X is given by

$$f(x) = C_3 \exp\left\{-\frac{(x - \lambda)^2}{2\sigma^2}\right\} I_0\left(\frac{\kappa x}{\sigma^2}\right), \quad -\infty < x < \infty, \tag{3.3.4}$$

where $C_3 = \sqrt{2\pi}/\sigma \exp\left\{-\kappa^2/(4\sigma^2)\right\} C_1$. The density (3.3.4) is generally asymmetric. As $\kappa \to 0$, it tends to the normal distribution with mean λ and variance σ^2.

When $\lambda = 0$, the density (3.3.4) can be expressed as

$$f(x) \quad = \quad \left\{\sqrt{2\pi}\sigma I_0\left(\frac{\kappa^2}{4\sigma^2}\right)\right\}^{-1} \exp\left(-\frac{x^2}{2\sigma^2} - \frac{\kappa^2}{4\sigma^2}\right) I_0\left(\frac{\kappa x}{\sigma^2}\right), \ -\infty < x < \infty. \tag{3.3.5}$$

When $\kappa^2 > 2\sigma^2$, (3.3.5) becomes bimodal, taking its maximum value at $x = \pm(\sigma^2/\kappa)B^{-1}(1 - 2\sigma^2/\kappa^2)$, where $B(x) = I_2(x)/I_0(x),\ x \geq 0$, and minimal value at $x = 0$. When $\kappa^2 \leq 2\sigma^2$, (3.3.5) is unimodal, taking maximum value at $x = 0$. See Appendix A for the proof.

3.3.2 An extension of the Mardia and Sutton model

The second distribution on the cylinder we discuss in this subsection is a generalization of the distribution by Mardia and Sutton (1978). Their distribution is obtained as a conditional distribution of a trivariate normal distribution with some restriction on parameters, or a maximum entropy distribution subject to constraints on certain moments. The marginal distribution of the circular component is von Mises. Here we propose a more flexible model which could be useful for cylindrical data where the marginal circular component possibly exhibiting asymmetry and/or bimodality of marginal circular component. The model is defined by the following theorem. The proof follows immediately from Theorem 13.2.1 of Kagan *et al.* (1973).

Theorem 3.3 *Let* (Θ, X) *have joint density*

$$f(\theta, x) = C_4 \exp\left[-\frac{\{x - \mu(\theta)\}^2}{2\sigma^2} + \kappa_1 \cos(\theta - \mu_1) + \kappa_2 \cos\{2(\theta - \mu_2)\}\right], \tag{3.3.6}$$

where $0 \leq \theta < 2\pi,\ -\infty < x < \infty,\ \sigma > 0,\ \kappa_1, \kappa_2 > 0,\ 0 \leq \mu_1 < 2\pi,\ 0 \leq \mu_2 < \pi,\ \mu(\theta) = \mu' + \lambda\cos(\theta - \nu),\ -\infty < \mu' < \infty,\ \lambda \geq 0$ *and* $0 \leq \nu < 2\pi$. *The normalizing constant* C_4 *is given by*

$$C_4^{-1} = (2\pi)^{3/2}\sigma\left[I_0(\kappa_1)I_0(\kappa_2) + 2\sum_{j=1}^{\infty} I_j(\kappa_2)I_{2j}(\kappa_1)\cos\{2j(\mu_1 - \mu_2)\}\right].$$

Then $f(\theta, x)$ *is the maximum entropy density on the cylinder subject to* $E(X^2)$, $E(X),\ E(X\cos\Theta),\ E(X\sin\Theta),\ E(\cos\Theta),\ E(\sin\Theta),\ E(\cos 2\Theta)$ *and* $E(\sin 2\Theta)$ *taking specified values consistent with expectation.*

The distribution with density (3.3.6) is also obtainable as a conditional distribution of a trivariate normal distribution without any constraints on the mean vector and covariance matrix. See Appendix B for details.

This distribution has the property that the conditional distribution of X given $\Theta = \theta$ is $N\left(\mu(\theta), \sigma^2\right)$ and the marginal distribution of Θ is the generalized von Mises distribution GVM$(\mu_1, \mu_2, \kappa_1, \kappa_2)$. As discussed in Section 1.2.3, this marginal distribution is symmetric or asymmetric, unimodal or bimodal, its shape depending on the choice of the parameter values. When $\kappa_2 = 0$, the distribution coincides with the one proposed by Mardia and Sutton (1978). The conditional distribution of Θ given $X = x$ is the generalized von Mises distribution GVM$(\nu_1, \nu_2, \lambda_1, \lambda_2)$ where $\lambda_1, \lambda_2, \nu_1$ and ν_2 satisfy

$$\lambda_1 \cos\nu_1 = \frac{\lambda}{\sigma^2}(x - \mu')\cos\nu + \kappa_1 \cos\mu_1,$$

$$\lambda_1 \sin\nu_1 = \frac{\lambda}{\sigma^2}(x - \mu')\sin\nu + \kappa_1 \sin\mu_1,$$

$$\lambda_2 \cos 2\nu_2 = -\frac{\lambda^2}{4\sigma^2}\cos 2\nu + \kappa_2 \cos 2\mu_2,$$

and

$$\lambda_2 \sin 2\nu_2 = -\frac{\lambda^2}{4\sigma^2}\sin 2\nu + \kappa_2 \sin 2\mu_2.$$

The marginal distribution of X has a complex form. When $\lambda = 0$, it is $N(\mu', \sigma^2)$. If $\lambda \neq 0$, then the distribution is asymmetric. Detailed properties are not easily derived.

Next, we discuss parameter estimation based on maximum likelihood. Let $(\Theta_i, X_i),\ i = 1, \ldots, n$ be a random sample from (3.3.6). Using a similar approach to that of Mardia and Sutton (1978), the maximum likelihood estimates of the parameters are given by

$$\begin{aligned}
\hat{\lambda} &= \frac{s_1}{1 - r_{23}^2}\left\{s_2^{-2}(r_{23}r_{13} - r_{12})^2 + s_3^{-2}(r_{23}r_{12} - r_{13})^2\right\}^{1/2}, \\
\cos\hat{\nu} &= \frac{s_1}{s_2\hat{\lambda}}\frac{r_{23}r_{13} - r_{12}}{r_{23}^2 - 1}, \quad \sin\hat{\nu} = \frac{s_1}{s_3\hat{\lambda}}\frac{r_{12}r_{23} - r_{13}}{r_{23}^2 - 1}, \quad \hat{\sigma}^2 = s_1^2(1 - r_{1.23}^2),
\end{aligned}$$

where

$$s_j^2 = \sum_{i=1}^{n}(x_{ji} - \overline{x}_j)^2/n, \quad r_{jk} = \sum_{i=1}^{n}(x_{ji} - \overline{x}_j)(x_{ki} - \overline{x}_k)/(ns_js_k),$$

$$r_{1.23}^2 = \frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2}, \quad \overline{x}_j = \sum_{i=1}^{n} x_{ji}/n, \quad j \neq k, \quad j = 1, 2, 3,$$

$$x_{1i} = x_i, \quad x_{2i} = \cos\theta_i, \quad x_{3i} = \sin\theta_i, \quad i = 1, \ldots, n.$$

Maximum likelihood estimation for the other parameters, i.e., $\kappa_1, \kappa_2, \mu_1$ and μ_2, is essentially the same as that for the generalized von Mises distribution GVM$(\mu_1, \mu_2, \kappa_1, \kappa_2)$. See Yfantis and Borgman (1982) for the details of maximum likelihood estimation for that distribution.

To investigate if there is dependence between x and θ, one can test $H_0 : \lambda = 0$ against $H_1 : \lambda > 0$. The likelihood ratio test approach leads the test statistic

$$T = -n \log\left(1 - r_{1.23}^2\right). \tag{3.3.7}$$

Because λ is on the boundary under H_0, the limiting distribution of T is not a chi-square distribution but an equally weighted mixture of zero and a chi-square random variable, $Z^2 I[Z > 0]$, as discussed in Section 3.2.2. The null hypothesis is rejected when T is large. Note that the test of independence is the same as that for the distribution proposed by Mardia and Sutton (1978).

Next we construct a test for the adequacy of the Mardia and Sutton submodel as a special case of our model, namely, $H_0 : \kappa_2 = 0$ versus $H_1 : \kappa_2 > 0$. The likelihood ratio test statistic is given by

$$T = -2 \log \frac{\max L_0}{\max L_1}, \tag{3.3.8}$$

where L_i is the likelihood under H_i, $i = 0, 1$. It is easy to get $\max L_1$ using the method to calculate the maximum likelihood estimates described above. $\max L_0$ is also easily obtained since the maximum likelihood estimators are given in (2.4)–(2.8) of Mardia and Sutton (1978). Since κ_2 is on the boundary of parameter space under the null hypothesis, the limiting distribution of T is that of $Z^2 I[Z > 0]$ as described above.

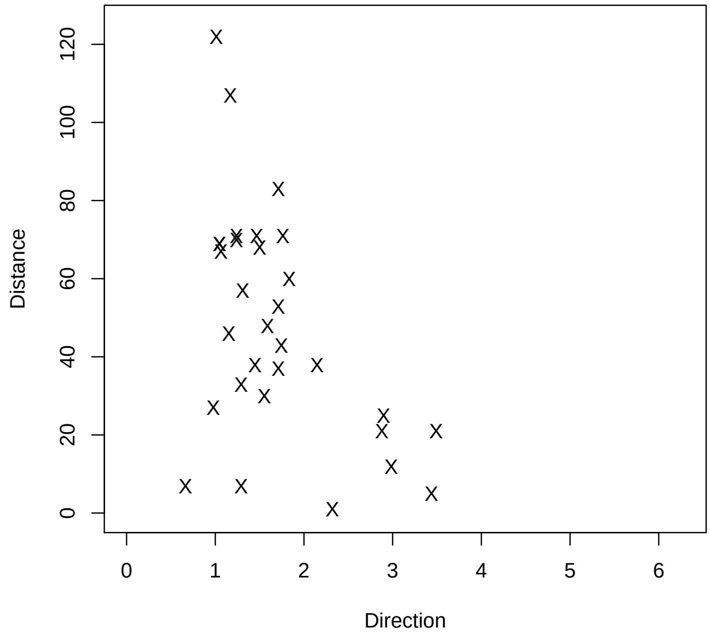


Figure 3.1. A planar plot of cylindrical data on the movements of blue periwinkles taken from Fisher (1993).

3.3.3 Examples

Example 3.1 As our first illustrative example, we consider a cylindrical dataset, $n = 30$, on the movements of blue periwinkles. The observations are directions (θ) and distances (x) moved by small blue periwinkles after they had been transplanted downshore from the height at which they normally live. The data are taken from Table B.20 of Fisher (1993).

A planar plot of the data, Figure 3.1, seems to show that there is dependence between the distances and directions. In fact the test of independence with test statistic (3.3.7) yields $T = -31(1 - r^2_{1.23}) = 4.68$ with $P \ll 0.05$. This test is highly significant and the assumption of independence is emphatically rejected.

Next we fit the model with density (3.3.6) and the Mardia and Sutton (1978)

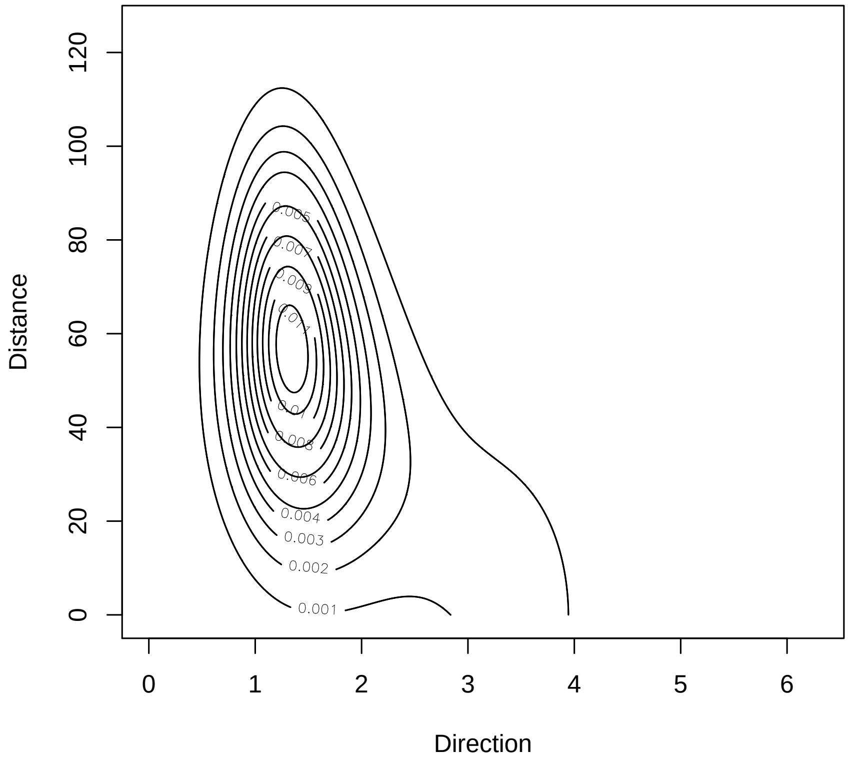


Figure 3.2. A contour plot of density (3.3.6) fitted to the blue periwinkles data from Fisher (1993).

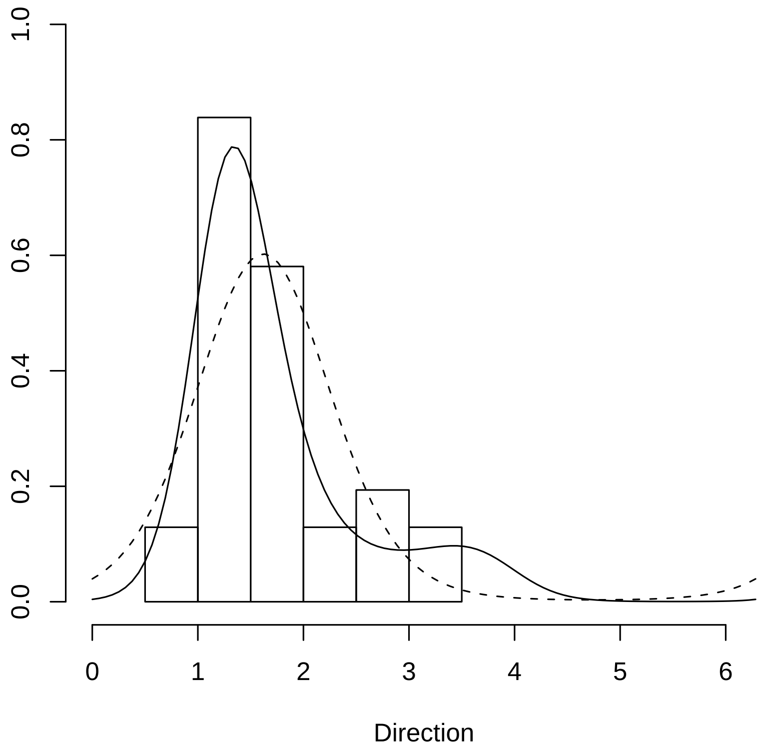


Figure 3.3. A histogram of the directions of movement of the blue periwinkles, and the fitted marginal circular densities for the full model (3.3.6) (solid line) and the Mardia and Sutton (1978) submodel (dashed line).

Table 3.1. Maximum likelihood estimates of the parameters and maximum log-likelihood and AIC values for the model with density (3.3.6) and the Mardia and Sutton model fitted to the blue periwinkles data from Fisher (1993).

Model	κ_1	κ_2	μ_1	μ_2	μ'	λ	σ	ν	log$L$	AIC
Model (3.3.6)	3.00	1.49	2.11	0.953	28.6	29.6	24.4	1.03	–170.7	357.5
Mardia & Sutton	2.59	–	1.62	–	28.6	29.6	24.4	1.03	–176.9	365.8

submodel and compare the results. Table 3.1 shows the maximum likelihood estimates of the parameters, log-likelihood and AIC values for the models. According to the AIC criterion, our full model gives a better fit than the Mardia and Sutton submodel. The test of the adequacy of the full model based on the test statistic in (3.3.8) results in a T-value of 12.3 with a corresponding $P \ll 0.001$. Clearly, here, there is significant improvement in fit using the full model (3.3.6) as compared to its Mardia and Sutton submodel. Figure 3.2 presents a contour plot of the fitted density for the full model (3.3.6). The figure seems to show a reasonable fit of our model to the dataset.

Finally we consider the marginal circular distribution of the dataset. Figure 3.3 displays a histogram of the circular data and the two fitted densities. It appears that the circular data are asymmetrically distributed. Also, the large-sample test of circular reflective symmetry of Pewsey (2002) yields a test statistic value of $T = \overline{b_2}/\text{var}(\overline{b_2})^{1/2} = -2.77$, with an associated p-value $P(T \leq -2.77) < 0.003$, and this test emphatically rejects the underlying symmetry. This result provides some evidence that the cylindrical model with the asymmetric circular marginal is a more appropriate one for fitting to this dataset.

Example 3.2 Another example consists of observations of January surface wind direction (θ) and temperature (x) at Kew at 12h GMT for the years 1956–60. The data are taken from Table 1 of Mardia and Sutton (1978) and a planar plot of

Table 3.2. Maximum likelihood estimates of the parameters and maximum log-likelihood and AIC values for the model with density (3.3.6) and the Mardia and Sutton submodel fitted to the data from Mardia and Sutton (1978).

Model	κ_1	κ_2	μ_1	μ_2	μ'	λ	σ	ν	log$L$	AIC
Model (3.3.6)	1.02	0.529	4.23	0.481	42.1	5.01	4.86	–2.78	–126.7	269.4
Mardia & Sutton	1.14	–	4.02	–	42.1	5.01	4.86	–2.78	–128.1	268.2

the dataset is given in Figure 1 of their paper. Table 3.2 shows the maximum likelihood estimates, log-likelihood and AIC values for both models. In this case, the AIC value for the Mardia and Sutton submodel is lower than that of our full model. There the Mardia and Sutton submodel is judged to be better, where the penalty for the two parameter increase is taken into account. Figure 3.4 exhibits a histogram of the circular data and the two fitted densities. The large-sample test of circular reflective symmetry of Pewsey (2002) finds no evidence that the underlying distribution is not reflectively symmetric. Also, a visual comparison of the two fitted densities suggest that little or no improvement in fit arised from the generalized von Mises distribution. Therefore, in this case, our conclusion is that the Mardia and Sutton submodel is the more suitable for these data.

Appendices

A Modality of the distribution with density (3.3.5)

The density (3.3.5) becomes unimodal or bimodal, depending on the values of κ and σ. That can be shown by differentiating (3.3.5) with respect to x, and equating to zero,

$$\begin{aligned} \frac{d}{dx}f(x) &= \left\{\sqrt{2\pi}\sigma I_0\left(\frac{\kappa^2}{4\sigma^2}\right)\right\}^{-1} \exp\left(-\frac{x^2}{2\sigma^2}-\frac{\kappa^2}{4\sigma^2}\right) \\ &\quad \times\left\{-\frac{x}{\sigma^2}I_0\left(\frac{\kappa x}{\sigma^2}\right)+\frac{\kappa}{\sigma^2}I_1\left(\frac{\kappa x}{\sigma^2}\right)\right\}=0. \end{aligned} \quad (3.3.9)$$

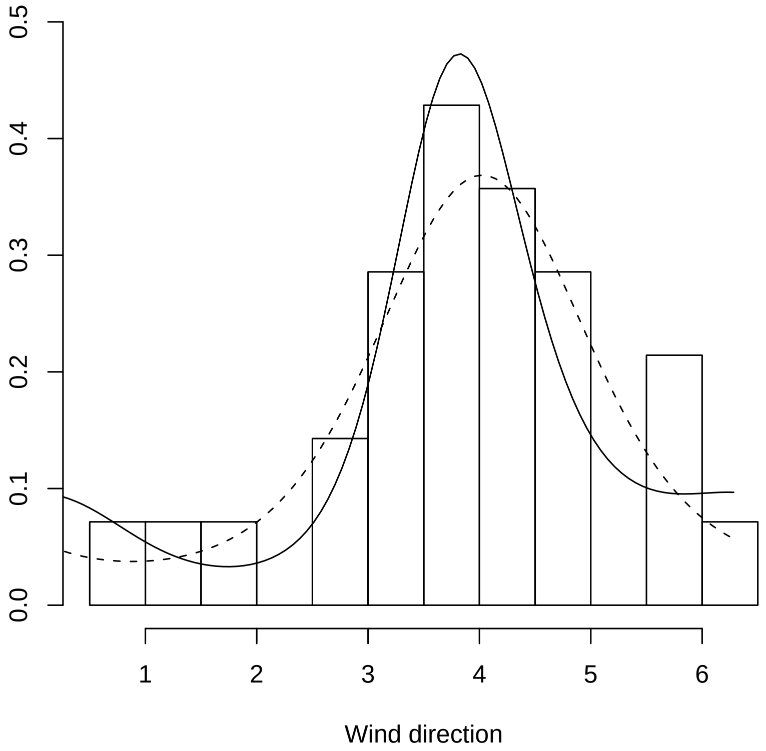


Figure 3.4. A histogram of the wind directions from Mardia and Sutton (1978), and the fitted marginal circular densities for the model with density (3.3.6) (solid line) and the Mardia and Sutton (1978) submodel (dashed line).

Using the fact that $I_0(x) - I_2(x) = 2I_1(x)/x$, (3.3.9) gives

$$I_2\left(\frac{\kappa x}{\sigma^2}\right) \Big/ I_0\left(\frac{\kappa x}{\sigma^2}\right) = 1 - \frac{2\sigma^2}{\kappa^2}.$$

Let $B(x)$ denote the ratio of the Bessel functions

$$B(x) = \frac{I_2(x)}{I_0(x)}, \quad x \geq 0.$$

Then $B(x)$ has the following properties:

(a) $0 \leq B(x) \leq 1, \quad x \geq 0$

(b) $\lim_{x \to +0} B(x) = 0, \quad \lim_{x \to \infty} B(x) = 1$

(c) $\dfrac{dB(x)}{dx} > 0$

The proof is as follows.

(a) It is obvious that $B(x) \geq 0$. Using the fact that $A(x)(= I_1(x)/I_0(x)) \geq 0, \ x \geq 0$, we have

$$B(x) = \frac{I_0(x) - (2/x)I_1(x)}{I_0(x)} = 1 - \frac{2}{x}A(x) \leq 1, \quad x > 0.$$

(b) Clearly $\lim_{x \to +0} B(x) = 0$. By using the fact that $\lim_{x \to \infty} A(x) = 1$ (See Jammalamadaka and SenGupta (2001)),

$$\lim_{x \to \infty} B(x) = \lim_{x \to \infty} \left\{1 - \frac{2}{x}A(x)\right\} = 1.$$

(c) The p variate von Mises-Fisher distribution on the unit sphere Ω^p in $\mathbb{R}^p$ has density

$$f(x) = \left(\frac{\kappa}{2}\right)^{p/2-1} \left\{\Gamma(p/2) I_{p/2-1}(\kappa)\right\}^{-1} \exp\left(\kappa \mu' x\right), \quad x \in \Omega^p,$$

where $\kappa \geq 0, \ \mu \in \Omega^p$. The Fisher information for the maximum likelihood estimator of κ (Mardia and Jupp, 2000, p. 199) is given by

$$-E\left[\frac{\partial^2}{\partial \kappa^2} \log f(X)\right] = A_p'(\kappa),$$

where $A_p(z) = I_{p/2}(z)/I_{p/2-1}(z)$. As the Fisher information is positive, we have

$$\begin{aligned}
\frac{d}{dz}\left\{\frac{I_2(z)}{I_0(z)}\right\} &= \frac{d}{dz}\left\{\frac{I_1(z)}{I_0(z)} \cdot \frac{I_2(z)}{I_1(z)}\right\} = \frac{d}{dz}\left\{A_2(z)A_4(z)\right\} \\
&= A_2'(z)A_4(z) + A_2(z)A_4'(z) \\
&> 0, \quad z > 0.
\end{aligned}$$

On using the properties of $B(x)$, it is shown that (3.3.5) is unimodal when $\kappa^2 > 2\sigma^2$ and bimodal otherwise.

B Derivation of density (3.3.6)

We noted that the density (3.3.6) can be obtained by conditioning a trivariate normal distribution without any constraints on the mean vector and covariance matrix. The exact derivation of the model is described as follows.

Let Y be a random vector which follows the trivariate normal distribution with mean vector $\eta = (\eta_1, \eta_2, \eta_3)'$ and covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 \\ \rho_{13}\sigma_1\sigma_3 & \rho_{23}\sigma_2\sigma_3 & \sigma_3^2 \end{pmatrix},$$

where $-\infty < \eta_i < \infty,\ \sigma_i > 0\ (i = 1, 2, 3),\ -1 < \rho_{23} < 1$ and $1 - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2 + 2\rho_{12}\rho_{13}\rho_{23} > 0$. We transform the trivariate random vector $Y = (X, X_1, X_2)' = (X, R\cos\Theta, R\sin\Theta)'$ where $R > 0,\ 0 \le \Theta < 2\pi$. Then the conditional distribution of (Θ, X) given $R = r$ can be shown to have density (3.3.6) by defining new parameters as

$$\mu(\theta) = \eta_1 + a_1\eta_2 + a_2\eta_3 + \lambda\cos(\theta - \nu), \quad \sigma^2 = \frac{\sigma_1^2\rho}{1 - \rho_{23}^2},$$

$$\kappa_1 \cos\mu_1 = r(b_1\eta_2 - b_2\eta_3), \quad \kappa_1 \sin\mu_1 = r(b_3\eta_3 - b_2\eta_2),$$

$$\kappa_2 \cos 2\mu_2 = \frac{r^2}{4}(b_3 - b_1), \quad \kappa_2 \sin 2\mu_2 = \frac{1}{2}b_2 r^2,$$

where

$$a_1 = \frac{\sigma_1}{\sigma_2}\frac{\rho_{13}\rho_{23} - \rho_{12}}{1 - \rho_{23}^2}, \quad a_2 = \frac{\sigma_1}{\sigma_3}\frac{\rho_{12}\rho_{23} - \rho_{13}}{1 - \rho_{23}^2},$$

$$b_1 = \frac{1}{\sigma_2^2(1-\rho_{23}^2)}, \quad b_2 = \frac{\rho_{23}}{\sigma_2\sigma_3(1-\rho_{23}^2)}, \quad b_3 = \frac{1}{\sigma_3^2(1-\rho_{23}^2)},$$
$$\lambda\cos\nu = -a_1 r, \quad \lambda\sin\nu = -a_2 r, \quad \rho = 1 - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2 + 2\rho_{12}\rho_{13}\rho_{23}.$$

The following properties hold between the new parameters and the original ones:

1. $\kappa_1 = 0 \iff \eta_2 = \eta_3 = 0$
2. $\kappa_2 = 0 \iff \sigma_2 = \sigma_3,\ \rho_{23} = 0$
3. $\lambda = 0 \iff \rho_{12} = \rho_{13} = 0$

The values taken by the parameters $\kappa_1,\ \kappa_2$ and λ range from 0 to infinity depending on the values of the original parameters. It is also clear that $\sigma > 0,\ 0 \leq \mu_1, \nu < 2\pi,\ 0 \leq \mu_2 < \pi$ and $-\infty < \mu' < \infty$.

Chapter 4

Distributions for a Pair of Unit Vectors

4.1 Introduction

In a variety of scientific fields, observations are described as pairs of d-dimensional unit vectors. In meteorology, for example, wind directions at the weather station in Milwaukee at 6 a.m. and noon (Johnson and Wehrly, 1977) are a data of this type with $d = 2$. Another example with $d = 3$ is consisting of the directions of magnetic field in a rock sample before and after some laboratory treatment (Stephens, 1979).

For the analysis of data of the type, various stochastic models have been proposed in the literature. Mardia (1975) provided a class of distributions for two unit vectors using the principle of maximum entropy subject to constraints on certain moments. Wehrly and Johnson (1980) proposed a family of bivariate circular distributions having specified marginals. Their submodel with von Mises marginals was studied by Shieh and Johnson (2005). Saw (1983) introduced bivariate families for pairs of dependent unit vectors, one of which is an offset distribution of the multivariate normal distribution with some restrictions on parameters. Rivest (1988) provided another model for two dependent unit vectors which is a generalization of the Fisher-von Mises distribution. A general class of bivariate distributions with exponential conditionals was proposed and discussed by Arnold and Strauss (1991) and Arnold *et al.* (1999), and a special case of their model defined on the two-dimensional torus was considered by SenGupta (2004). Recent work by Alfonsi and Brigo (2005) proposed new families of copulas based on periodic functions.

The main purpose of the chapter is to introduce a new distribution for a pair of dependent unit vectors which is generated by $\mathbb{R}^d$-valued Brownian motion. To our knowledge, distributions on this manifold have not previously been proposed based on Brownian motion. In this chapter, a new approach is taken to provide a tractable model. This method enables us to define a distribution with uniform marginals and derive some desirable properties.

Section 4.2 suggests a model for two dependent unit vectors and Section 4.3 investigates properties of the proposed model, including parameter estimation and a pivotal statistic. In Section 4.4 we focus on the bivariate circular case of the model and discuss its detailed properties. It is shown that some desirable properties, such as multiplicative property and log-infinite divisibility, hold for this submodel. In Section 4.5, generalizations of the proposed model are discussed. Also, related models on $\mathbb{R}^2$ and on the cylinder are constructed by applying bilinear fractional transformations to the proposed model.

4.2 Model for a pair of unit vectors

4.2.1 Definition of the proposed model

Let $\{B_t\,;\, t \geq 0\}$ be $\mathbb{R}^d$-valued Brownian motion with $d \geq 2$. Starting at $B_0 = 0$, a Brownian particle will eventually hit a d-sphere with radius ρ $(\in (0,1))$, and let τ_1 be the minimum time at which the particle exits the sphere, i.e. $\tau_1 = \inf\{t\,;\, \|B_t\| = \rho\}$ where $\|\cdot\|$ is the Euclidean norm. After leaving the sphere with radius ρ, the particle will hit a unit sphere first at the time τ_2, meaning $\tau_2 = \inf\{t\,;\, \|B_t\| = 1\}$. Then the proposed model is defined by the joint distribution of a random vector

$$\left(Q \frac{B_{\tau_1}}{\|B_{\tau_1}\|}, B_{\tau_2} \right),$$

where Q is a member of $O(d)$, the group of orthogonal transformations in $\mathbb{R}^d$. It is remarked here that the reason for multiplying Q by $B_{\tau_1}/\|B_{\tau_1}\|$ is to make the model more flexible without losing its tractability.

4.2.2 Probability density function

For convenience, write $(U, V) = (QB_{\tau_1}/\|B_{\tau_1}\|, B_{\tau_2})$. It is clear that (U, V) is a random vector for which each variable takes values on the unit sphere. The joint distribution of (U, V) has density

$$c(u, v) = \frac{1}{A_{d-1}^2} \frac{1 - \rho^2}{(1 - 2\rho u'Qv + \rho^2)^{d/2}}, \quad u, v \in S^{d-1}, \tag{4.2.1}$$

where $\rho \in [0, 1)$, $Q \in O(d)$. The domain of ρ is extended to include $\rho = 0$ so that the model includes the uniform distribution. We write $(U, V) \sim BS_d(\rho Q)$ if a random vector (U, V) has density (4.2.1). For the derivation of the density (4.2.1), see Appendix 4.5.2.

The parameter ρ influences the dependence between U and V. When $\rho = 0$, U and V are independent and distributed as the uniform distribution on the sphere, i.e. $c(u, v) = 1/A_{d-1}^2$ on $u, v \in S^{d-1}$. As ρ tends to 1, it can be shown that $P(\|U - QV\| < \varepsilon) \to 1$ for any $\varepsilon > 0$.

As is clear from the form of (4.2.1), $c(u, v)$ is a function of $u'Qv$, the inner product of u and Qv. From this fact, we easily find that the density (4.2.1) takes maximum (minimum) values for a given v at $u = Qv$ ($u = -Qv$). Thus the parameter Q controls the mode of the density. It is known that an orthogonal transformation Q involves two types of transformations, namely, rotation and/or reflection. In particular, when $d = 2$, these transformations can be expressed as

$$v \longmapsto \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} v \quad \text{and} \quad v \longmapsto \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} v,$$

where $0 \leq \theta < 2\pi$. If $\det Q = 1$, this transformation consists of only rotation. Otherwise, if $\det Q = -1$, the transformation is made up of a reflection together with a rotation.

4.3 Properties of and inference for the proposed model

4.3.1 Marginals and conditionals

One important feature of the proposed model is that it has well-known marginals and conditionals. Suppose $(U, V) \sim BS_d(\rho Q)$. The density for this random vector, (4.2.1), is $O(d)$-symmetric in the sense of Rivest (1984, Example 1). It follows then that the marginals of U and V are uniform distributions on S^{d-1} with density

$$f(x) = \frac{1}{A_{d-1}}, \quad x \in S^{d-1}.$$

Hence, model (4.2.1) can be viewed as a copula on $S^{d-1} \times S^{d-1}$. A difference between this special copula and the usual ones is the periodicity of the variables for this copula which the usual one does not assume.

Let U_j be the jth element of U, i.e. $U = (U_1, \ldots, U_j, \ldots, U_d)'$. It is known that the marginal of U_j has a distribution with density

$$f(u_j) = \frac{(1 - {u_j}^2)^{(d-3)/2}}{B\{\frac{1}{2}(d-1), \frac{1}{2}\}}, \quad -1 < u_j < 1,$$

where denote $B(\cdot, \cdot)$ the beta function. This model is U-shaped ($d = 2$), uniform ($d = 3$), or unimodal ($d \geq 4$). Note that $\frac{1}{2}(U_j + 1)$ has a beta distribution on $(0, 1)$, more specifically, $\mathrm{Beta}\{\frac{1}{2}(d-1), \frac{1}{2}(d-1)\}$.

Both conditional distributions of U given $V = v$ and V given $U = u$ are the exit distributions for the sphere, i.e. $U|(V = v) \sim \mathrm{Exit}_d(\rho Q v)$ and $V|(U = u) \sim \mathrm{Exit}_d(\rho Q' u)$.

It is worth remarking that the conditional of $W \equiv v'Q'U$ given $V = v$ has a family discussed by Leipnik (1947) and McCullagh (1989). As in the latter paper, write $X \sim H'(\theta, \nu)$ if the density of the random vector X is

$$f(x) = \frac{1 - \theta^2}{B(\nu + \frac{1}{2}, \frac{1}{2})} \frac{(1 - x^2)^{\nu - 1/2}}{(1 - 2\theta x + \theta^2)^{\nu + 1}}, \quad -1 < x < 1,$$

where $-1 < \theta < 1$ and $\nu > -\frac{1}{2}$. Then it follows that $W|(V = v) \sim H'\{\rho, (d-2)/2\}$.

4.3.2 Some properties

Here we investigate some of the properties of the model with density (4.2.1). The first is that the distribution is closed under orthogonal transformations:

$$(U, V) \sim BS_d(\rho Q) \implies (Q_1 U, Q_2 V) \sim BS_d(\rho Q_1 Q Q_2'), \quad Q_1, Q_2 \in O(d).$$

The next result is obtainable by applying a result which appears, for example, in Durrett (1984, Section 1.10).

Theorem 4.1 *Suppose that* (U, V) *is distributed as* $BS_d(\rho Q)$*. Let* f *be* C^2 *in* D *and continuous on* $\overline{D}$ *where* $D = \{\zeta \in \mathbb{R}^d \,;\, \|\zeta\| < 1\}$*. If* f *is harmonic, namely,*

$$\frac{\partial^2}{\partial x_1^2} f + \frac{\partial^2}{\partial x_2^2} f + \cdots + \frac{\partial^2}{\partial x_d^2} f = 0,$$

then $E\{f(V) \,|\, U = u\} = f(\rho Q' u)$ *and* $E\{f(U) \,|\, V = v\} = f(\rho Q v)$*.*

Using this fact, it is easy to show that $E\{f(U)\} = E\{f(V)\} = f(0)$.

Rivest (1984, Proposition 1) showed that the calculation of moments is simplified to some extent for a class of $O(d)$-symmetric distributions. This fact is helpful when obtaining the moments and correlation coefficient of the model, which we give in the following theorem.

Theorem 4.2 *Suppose* (U, V) *has density (4.2.1). Then*

$$E(U) = E(V) = 0, \quad E(UU') = E(VV') = d^{-1} I,$$

$$E(UV') = d^{-1} \rho Q. \tag{4.3.1}$$

The Jupp and Mardia (1980) coefficient of correlation, r^2*, is thus*

$$r^2 \equiv \mathrm{tr}(\Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1}) = d\rho^2,$$

where $\Sigma_{11} = E(UU') - E(U)E(U')$*,* $\Sigma_{12} = E(UV') - E(U)E(V')$*,* $\Sigma_{21} = \Sigma_{12}'$*, and* $\Sigma_{22} = E(VV') - E(V)E(V')$*.*

Note the simplicity of these moments and the correlation coefficient. See Appendix B for the proof.

The following result is useful to construct a pivotal statistic for (ρ, Q), which is discussed in Section 4.3.5. The proof is also given in Appendix B.

Theorem 4.3 *If* $(U, V) \sim BS_d(\rho Q)$, *then* $U'QV \sim H'\{\rho, \frac{1}{2}(d-2)\}$.

4.3.3 Random vector simulation

To generate a random vector having density (4.2.1), it is profitable to use the idea of the tangent-normal decomposition.

Let W be a random variable from $H'\{\rho, \frac{1}{2}(d-2)\}$, and let $d(X;\zeta) = (I - \zeta\zeta')X/\|(I-\zeta\zeta')X\|$ where $\zeta \in S^{d-1}$ and X a random vector having a uniform distribution on S^{d-1}. In other words, $d(X;\zeta)$ has a uniform distribution on the $(d-1)$-sphere, $S_\perp$, in $\mathbb{R}^d$ defined by $S_\perp = \{\eta \in \mathbb{R}^d;\ \|\eta\| = 1,\ \zeta'\eta = 0\}$. Then the conditional of U given $V = v$ can be decomposed into

$$U \,|\, (V = v) \stackrel{d}{=} WQv + (1 - W^2)^{1/2} d(X; Qv).$$

Given this, the generation of variates from (4.2.1) can be carried out using the following three steps: (i) Generate a random vector V which has a uniform distribution on S^{d-1}. This is achieved by using the method proposed by Tashiro (1977). (ii) Generate W, which has $H'\{\rho, \frac{1}{2}(d-2)\}$, as stated in Section 4 of McCullagh (1989). (iii) Finally, a random vector $d(X;Qv)$ distributed as a uniform distribution on $S_\perp$ is obtained in a similar manner as in Step (i), and one gets a variate from the conditional of U given $V = v$ as described in the preceding paragraph. Then the joint distribution of (U, V) is $BS_d(\rho Q)$.

4.3.4 Parameter estimation

Parameter estimation for multivariate distributions is often difficult. This is also the case for out model. However, one can discuss parameter estimation under certain conditions. Here we consider parameter estimation based on the method of moments and maximum likelihood.

First, the method of moments estimator is constructed from (4.3.1). Assume that (U_j, V_j) $(j = 1, \dots, n\,(\geq 2))$ is a random sample from density (4.2.1) with unknown parameters ρ and Q. Under the condition, $\mathrm{rank}(\sum_{j=1}^n U_j V_j') = d$, one can construct an estimator for the parameters based on the moment $E(UV')$. This is done by equating the theoretical and sample moments. Thus we obtain

$$\hat{\rho} = d \left| \det \left(\frac{1}{n} \sum_{j=1}^n U_j V_j' \right) \right|^{1/d} \quad \text{and} \quad \hat{Q} = \frac{d}{n\hat{\rho}} \sum_{j=1}^n U_j V_j'. \tag{4.3.2}$$

We note that although $\hat{Q}$ is an unbiased estimator of Q with $\det \hat{Q} = 1$, it is not necessarily an orthogonal matrix.

Next, we consider maximum likelihood estimation. Let (U_j, V_j) $(j = 1, \dots, n)$ be an iid sample from $BS_d(\rho Q)$, where Q is known and ρ is unknown. The log-likelihood for ρ is given by

$$l(\rho) = C + n \log(1 - \rho^2) - \frac{d}{2} \sum_{j=1}^n \log(1 - 2\rho u_j' Q v_j + \rho^2), \tag{4.3.3}$$

where C is a constant which does not depend on ρ. The derivative with respect to ρ is

$$\frac{\partial l}{\partial \rho} = \frac{-2n\rho}{1 - \rho^2} + d \sum_{j=1}^n \frac{x_j - \rho}{1 - 2\rho x_j + \rho^2},$$

where $x_j = u_j' Q v_j \in [-1, 1]$. From this expression, we find that the maximization of (4.3.3) with respect to ρ is essentially the same as that of $H'\{\rho, \frac{1}{2}(d-2)\}$ with respect to ρ.

4.3.5 Pivotal statistic

Suppose (U, V) is a $BS_d(\rho Q)$ random vector. Define a random variable

$$T(\rho, Q) = \frac{1 - (U'QV)^2}{1 - 2\rho U'QV + \rho^2}.$$

It is easy to see that $0 < T(\rho, Q) < 1$ a.s. for any ρ and Q. As shown in Theorem 4.3, $U'QV \sim H'\{\rho, \frac{1}{2}(d-2)\}$. Then by using equations (15.1.13) and (15.3.1) of Abramowitz and Stegun (1970), one obtains

$$E\left\{T(\rho, Q)^r\right\} = \frac{B\{r + \frac{1}{2}(d-1), \frac{1}{2}\}}{B\{\frac{1}{2}(d-1), \frac{1}{2}\}}.$$

Since these moments are equal to those of a beta distribution Beta$\{\frac{1}{2}(d-1), \frac{1}{2}\}$, it follows that $T(\rho, Q)$ is a pivotal statistic for (ρ, Q) having a Beta$\{\frac{1}{2}(d-1), \frac{1}{2}\}$ distribution almost surely. Because we know the exact distribution of $T(\rho, Q)$, confidence intervals for the parameters based on $T(\rho, Q)$ can be obtained in the usual way.

4.4 Bivariate circular case

4.4.1 Transformation of random vectors and parameters

So far we have considered properties of model (4.2.1) for the general dimensional case. The theme of this subsection is to specifically discuss the bivariate circular case of the proposed model which possesses some unique properties.

Suppose $(U, V) \sim BS_2(\rho Q)$. Then its density is expressed as

$$c(u, v) = \frac{1}{4\pi^2} \frac{1-\rho^2}{1 - 2\rho u'Qv + \rho^2}, \quad u, v \in S^1.$$

For further discussion, it is advantageous to transform the random variables and parameters by taking

$$(Z_U, Z_V) = (U_1 + iU_2, V_1 + iV_2) \quad \text{and} \quad \psi = \rho e^{i\theta},$$

where $U = (U_1, U_2)'$, $V = (V_1, V_2)'$, and θ is a constant satisfying

$$Q = \begin{pmatrix} \cos\theta & -\det Q \sin\theta \\ \sin\theta & \det Q \cos\theta \end{pmatrix}, \quad 0 \le \theta < 2\pi.$$

Then it follows that $|\psi| < 1$ and $Z_U, Z_V \in \Omega$. The density for (Z_U, Z_V) is given by

$$c(z_u, z_v) = \frac{1}{4\pi^2} \frac{1 - |\psi|^2}{|1 - \psi z_v z_u^{-\det Q}|^2}, \quad z_u, z_v \in \Omega. \tag{4.4.1}$$

If (Z_U, Z_V) has density (4.4.1) with $\det Q = 1$, we write $(Z_U, Z_V) \sim BC_+(\psi)$. Similarly, we write $(Z_U, Z_V) \sim BC_-(\psi)$ if (Z_U, Z_V) has density (4.4.1) with $\det Q = -1$.

Note that this transformation does not actually change the distribution. All have done is to express the random variables and the parameters in the form of complex numbers for the sake of further investigation of the distributions.

As already stated in Section 4.2.2, the marginals of Z_U and Z_V are circular uniform, whereas both conditionals of Z_U given $Z_V = z_v$ and Z_V given $Z_U = z_u$ are exit distributions for the circle, i.e., the wrapped Cauchy distributions. For brevity, we introduce the notation $C^*(\phi)$ derived from McCullagh (1996) which denotes the wrapped Cauchy distribution with density

$$f(z) = \frac{1}{2\pi} \frac{1 - |\phi|^2}{|z - \phi|^2}, \quad z \in \Omega;\ |\phi| < 1.$$

The relationship, $|\phi| = \|\xi\|$ and $\arg(\phi) = \arg(\xi_1 + i\xi_2)$ where $\xi = (\xi_1, \xi_2)'$, holds between the parameters of model (1.2.5) and those of the density above via a transformation $Z = X_1 + iX_2$. See McCullagh (1996) and Mardia (1972, pp.51-52) for further properties of the wrapped Cauchy distribution. For model (4.4.1), it is easy to show that $Z_U | (Z_V = z_v) \sim C^*(\psi z_v)$ and $Z_V | (Z_U = z_u) \sim C^*(\overline{\psi} z_u)$.

4.4.2 Some properties

To investigate other properties of the model, it is useful to calculate its moments. Assume that (Z_U, Z_V) has $BC_+(\psi)$. Then the moments for (Z_U, Z_V) are obtained, by applying Theorem 11.13 of Rudin (1987), as

$$E\left({Z_U}^j {Z_V}^k\right) = \begin{cases} \psi^j, & j = -k \geq 0, \\ \overline{\psi}^{-j}, & j = -k < 0, \\ 0, & \text{otherwise}, \end{cases} \quad \text{for } j, k \in \mathbb{Z}. \tag{4.4.2}$$

Similarly, we can obtain the moments for BC_-. According to Fourier series expansion theory, one can recover the density from these moments if the density f satisfies $f \in L^2(\Omega \times \Omega)$. See Dym and McKean (1972, Section 1.10) for details.

Using these results, the following properties are established. First, the BC_+ model has the multiplicative property:

$$(Z_{U1}, Z_{V1}) \sim BC_+(\psi_1) \perp (Z_{U2}, Z_{V2}) \sim BC_+(\psi_2)$$
$$\Longrightarrow (Z_{U1}Z_{U2}, Z_{V1}Z_{V2}) \sim BC_+(\psi_1\psi_2). \qquad (4.4.3)$$

Likewise, it can be shown that the BC_- model also has the multiplicative property. However, the convolution of BC_+ and BC_- is the uniform distribution, i.e.

$$(Z_{U1}, Z_{V1}) \sim BC_+(\psi_1) \perp (Z_{U2}, Z_{V2}) \sim BC_-(\psi_2)$$
$$\Longrightarrow (Z_{U1}Z_{U2}, Z_{V1}Z_{V2}) \sim BC_+(0).$$

In addition,

$$(Z_U, Z_V) \sim BC_\pm(\psi) \implies ({Z_U}^n, {Z_V}^n) \sim BC_\pm(\psi^n) \text{ for any } n \in \mathbb{N}.$$

As n tends to infinity, the distribution of $({Z_U}^n, {Z_V}^n)$ tends to a uniform distribution on the torus.

Furthermore, model (4.4.1) is log-infinitely divisible. This is proved as follows. Let $(Z_U, Z_V) \sim BC_\pm(\psi)$. Then for any positive integer n, the assumption that (Z_{Uj}, Z_{Vj}) $(j = 1, \ldots, n)$ is an iid sample from $BC_\pm(\sqrt[n]{\psi})$ yields

$$\left(\sum_{j=1}^{n} \log Z_{Uj}, \sum_{j=1}^{n} \log Z_{Vj}\right) \stackrel{d}{=} (\log Z_U, \log Z_V).$$

4.4.3 Random vector generator

In order to simulate a $BC_+(\psi)$ random vector, one could generate $\mathbb{R}^2$-valued Brownian motion and record the points of which the Brownian particle hits circles with radii ρ and 1. However, this algorithm is somewhat inefficient because we need to simulate Brownian motion at least up to the time at which the particle hits the unit circle. Another possibility is discussed in Section 4.3.3, but it too is less efficient than the method proposed below. The focus of this subsection is therefore to discuss an algorithm to simulate $BC_+(\psi)$ variates which we conclude to be more appealing than the aforementioned methods.

To obtain the random vector, we use the fact that the marginal of Z_U is circular uniform and the conditional of Z_V given $Z_U = z_u$ is wrapped Cauchy, specifically, $C^*(\overline{\psi} z_u)$. For the generation of a variate from a wrapped Cauchy distribution, we apply a result from McCullagh (1996) concerning the Möbius transformation of a circular uniform, namely that

$$Z \sim C^*(0) \implies \frac{Z+\beta}{1+\overline{\beta}Z} \sim C^*(\beta), \quad |\beta| < 1. \tag{4.4.4}$$

An algorithm for generating $BC_+(\psi)$ random vectors then involves the following steps:

Step 1: Generate uniform $(0, 1)$ random numbers U_1 and U_2.

Step 2: Put $Z_U = \exp(2\pi i U_1)$ and $Z_T = \exp(2\pi i U_2)$.

Step 3: Take $Z_V = \dfrac{\overline{\psi} Z_U + Z_T}{1 + \psi \overline{Z_U} Z_T}$.

Then the joint distribution of (Z_U, Z_V) is $BC_+(\psi)$. In Step 2, Z_U and Z_T are independent circular uniform random variables. In Step 3, because of property (4.4.4), the conditional distribution of Z_V given $Z_U = z_u$ is $C^*(\overline{\psi} z_u)$. Therefore it follows that $(Z_U, Z_V) \sim BC_+(\psi)$.

$BC_-(\psi)$ random vectors can be simulated using a very similar approach.

4.4.4 Parameter estimation

Here we consider parameter estimation for the $BC_+(\psi)$ model based on the method of moments and maximum likelihood. Although we discuss parameter estimation for the $BC_+(\psi)$ only here, it is possible to derive the estimates of the parameters for the $BC_-(\psi)$ model by a straightforward modification of the result below.

First, we consider method of moments estimation based on (4.4.2). Assume (Z_U, Z_V) is a $BC_+(\psi)$ random variable. As discussed in Section 4.4.2, its theoretical moments are given by (4.4.2). Suppose (Z_{Uj}, Z_{Vj}) $(j = 1, \ldots, n)$ is a random sample from the $BC_+(\psi)$ distribution. The method of moments estimator is obtained by equating the theoretical and sample moments. Thus we obtain

$$\hat{\psi} = \frac{1}{n} \sum_{j=1}^{n} Z_{Uj} \overline{Z_{Vj}}.$$

We note that this estimator is the same as (4.3.2), which is the method of moments estimator based on (4.3.1), if $\operatorname{rank}(\sum_{j=1}^{n} U_j V_j') = 2$.

Second, turning to the maximum likelihood estimation, it is obvious that the maximum likelihood estimator coincides with the method of moments estimator, i.e. $\hat{\psi} = Z_{U1}\overline{Z_{V1}}$ for a single observation, i.e. when $n = 1$. When n is large, the estimates must be obtained numerically. Note that the likelihood function can be written as

$$L(\psi) \propto \prod_{j=1}^{n} \frac{1 - |\psi|^2}{|z_{uj}\overline{z_{vj}} - \psi|^2}.$$

This expression suggests that maximum likelihood estimation for the $BC_+(\psi)$ model essentially coincides with that for the wrapped Cauchy distribution $C^*(\psi)$. Therefore we can obtain estimates by applying the algorithm of Kent and Tyler (1988).

4.5 Related models

4.5.1 Generalizations of model (4.2.1)

As described in Section 4.2.1, the model with density (4.2.1) is generated using Brownian motion starting at $B_0 = 0$. In this subsection we briefly discuss a distribution which is generated using Brownian motion starting at $B_0 = \xi$ $(\|\xi\| < \rho)$ instead of $B_0 = 0$. We define a random vector $(U, V) = (QB_{\tau_1}/\|B_{\tau_1}\|, B_{\tau_2})$ in the same way as that used in Section 4.2.1 except that we incorporate the new starting point. The resulting density for (U, V) is given by

$$f(u, v) = \frac{1}{A_{d-1}^2} \frac{1 - \rho^2}{(1 - 2\rho u'Qv + \rho^2)^{d/2}} \frac{\rho^2 - \|\xi\|^2}{(\rho^2 - 2\rho U'Q\xi + \|\xi\|^2)^{d/2}}, \quad u, v \in S^{d-1}. \tag{4.5.1}$$

The marginals and conditional distributions of V given $U = u$ are the exit distributions:

$$U \sim \mathrm{Exit}_d(\rho^{-1}Q\xi), \quad V \sim \mathrm{Exit}_d(\xi) \quad \text{and} \quad V|(U = u) \sim \mathrm{Exit}_d(\rho Q' u).$$

The conditional distribution of U given $V = v$ is not of the familiar form. This conditional distribution can be unimodal or bimodal and is generally skewed except for certain special cases such as $v = \pm\xi/\|\xi\|$. We remark here also that the bivariate circular case of model (4.5.1) is a submodel of the distribution briefly discussed in Section 2.3 as a model related to the circular–circular regression model.

Another generalization arises out of the use of the method discussed in Saw (1983). This method enables us to derive a distribution with prescribed marginals.

In the bivariate circular case, it might also be promising to apply the Möbius transformation to each variable. Let $(Z_U, Z_V) \sim BC_+(\psi)$ and define a random vector

$$(\tilde{Z}_U, \tilde{Z}_V) = \left(\frac{Z_U + \alpha_1}{1 + \overline{\alpha_1} Z_U}, \frac{Z_V + \alpha_2}{1 + \overline{\alpha_2} Z_V} \right), \quad |\alpha_1|, |\alpha_2| < 1.$$

Then, because of property (4.4.4), the marginals of $\tilde{Z}_U$ and $\tilde{Z}_V$ have wrapped Cauchy distributions $C^*(\alpha_1)$ and $C^*(\alpha_2)$, respectively. Another benefit of this extension is that its density has a simple and exact form, including the normalizing constant which does not involve any special functions.

4.5.2 Related distributions on $\mathbb{R}^2$ and on the cylinder

In previous subsections in this chapter, we have dealt with distributions for two directional observations. In this subsection, we provide models on two other manifolds, namely, $\mathbb{R}^2$ and the cylinder.

By applying bilinear fractional transformations to model (4.4.1), a distribution on $\mathbb{R}^2$ is constructed. Let (Z_U, Z_V) be distributed as $BC_-(\psi)$. Define a random vector (X, Y) as

$$X = i\frac{1 - Z_U}{1 + Z_U} \quad \text{and} \quad Y = i\frac{1 - Z_V}{1 + Z_V}.$$

Clearly, (X, Y) takes values in $\mathbb{R}^2$. It is straightforward to show that the joint density for (X, Y) is

$$f(x, y) = \frac{1}{\pi^2} \frac{\operatorname{Im}(\theta)}{|x + y + \theta(1 - xy)|^2}, \quad x, y \in \mathbb{R}, \tag{4.5.2}$$

where $\theta = i(1 - \psi)/(1 + \psi)$. Since $|\psi| < 1$, it is evident that $\operatorname{Im}(\theta) > 0$.

This model has the following properties:

$$X \sim C(i), \quad Y \sim C(i),$$

$$X|(Y = y) \sim C\left(\frac{\theta + y}{1 - \theta y}\right), \quad Y|(X = x) \sim C\left(\frac{\theta + x}{1 - \theta x}\right),$$

where the $C(\phi)$ notation is derived from McCullagh (1992) and denotes the Cauchy distribution on the real line with location parameter $\operatorname{Re}(\phi)$ and scale parameter $\operatorname{Im}(\phi)$. Thus the marginals and conditionals are members of the real Cauchy family. Further properties of model (4.5.2) are derived using the inverse transformations $Z_U = (1 + iX)/(1 - iX)$ and $Z_V = (1 + iY)/(1 - iY)$, which map the real line onto the unit circle in the complex plane.

A related distribution on the cylinder $\Omega \times \mathbb{R}$ is obtained in a similar fashion. Let (Z_U, Z_V) be $BC_+(\psi)$ distributed. Define a random vector

$$(Z_\Theta, X) = \left(Z_U, i\frac{1 - Z_V}{1 + Z_V}\right).$$

Then the marginals and conditionals of (Z_Θ, X) are

$$Z_\Theta \sim C^*(0), \quad X \sim C(i),$$

$$Z_\Theta|(X = x) \sim C^*\left(\frac{1 + ix}{1 - ix}\psi\right), \quad X|(Z_\Theta = z_\theta) \sim C\left(-i\frac{1 - \overline{z_\theta}\psi}{1 + \overline{z_\theta}\psi}\right).$$

Thus, the marginals are circular uniform and standard Cauchy, while the conditionals are the wrapped Cauchy and linear Cauchy distributions, respectively.

Appendices

A Derivation of density (4.2.1)

Let $c(u,v)$ be the joint density of $(U,V) = (QB_{\tau_1}/\|B_{\tau_1}\|, B_{\tau_2})$ which is defined in the same way as in Section 4.2.1. Note that the density can be expressed as

$$c(u,v) = f_U(u)g_{V|U}(v|u), \quad u,v \in S^{d-1},$$

where f_U is a density for the marginal of U and $g_{V|U}$ that for the conditional of V given $U=u$. Clearly, the marginal of U is distributed as the uniform distribution and thus $f_U(u) = 1/A_{d-1}$. Because of the Markov property of Brownian motion, the conditional of V given $U=u$ is essentially equivalent to the exit distribution for the sphere generated by Brownian motion starting at $B_0 = \rho Q'u$. (See Durrett (1984, Section 1.10)). The density for the exit distribution for the sphere is known to be

$$g_{V|U}(v|u) = \frac{1}{A_{d-1}} \frac{1-\rho^2}{\|v - \rho Q'u\|^d}, \quad v \in S^{d-1}.$$

Thus we obtain the density (4.2.1).

Density (4.5.1) is obtained by a straightforward modification of the above.

B Proofs of Theorems 4.2 and 4.3

Proof of Theorem 4.2 Since the marginals of U and V are uniformly distributed on the sphere, it is evident that $E(U) = E(V) = 0$ and $E(UU') = E(VV') = d^{-1}I$.

We show that $E(UV') = d^{-1}\rho I$. Because model (4.2.1) is $O(d)$-symmetric in the sense of Rivest (1988), calculation of $E(UV')$ is simplified by applying Proposition 1 of his paper to

$$E(UV') = \text{diag}\{E(R_jS_j)\}Q,$$

where $(R,S) \sim BS_d(\rho I)$, $R = (R_1,\ldots,R_d)'$, $S = (S_1,\ldots,S_d)'$. Consider the integral

$$E(R_1S_1) = \int_{S^{d-1}\times S^{d-1}} r_1 s_1 c(r,s) dr ds = \int_{S^{d-1}} \frac{r_1}{A_{d-1}} \int_{S^{d-1}} \frac{s_1}{A_{d-1}} \frac{1-\rho^2}{\|s-\rho r\|^d} ds dr.$$

Transforming S into $\tilde{S} = PS$ where P is a $d \times d$ orthogonal matrix such that $P = (r, p_2, \ldots, p_d)'$, $p_j = (p_{j1}, \ldots, p_{jd})' \in \mathbb{R}^d$, we have

$$\begin{aligned}
&\int_{S^{d-1}} \frac{r_1}{A_{d-1}} \int_{S^{d-1}} \frac{s_1}{A_{d-1}} \frac{1-\rho^2}{\|s-\rho r\|^d} ds dr \\
&\quad = \int_{S^{d-1}} \frac{r_1}{A_{d-1}} \int_{S^{d-1}} \frac{r_1 \tilde{s}_1 + \sum_{j=2}^d p_{j1} \tilde{s}_j}{A_{d-1}} \frac{1-\rho^2}{(1-2\rho\tilde{s}_1+\rho^2)^{d/2}} d\tilde{s} dr \\
&\quad = \int_{S^{d-1}} \frac{\tilde{s}_1}{dA_{d-1}} \frac{1-\rho^2}{(1-2\rho\tilde{s_1}+\rho^2)^{d/2}} d\tilde{s}.
\end{aligned}$$

The last equality results on using $E(R) = 0$ and $E(R_1^2) = d^{-1}$. Then, because if $X \sim H'(\theta, \nu)$, then $E(X) = \theta$, the above equation can be expressed as

$$\begin{aligned}
&\int_{S^{d-1}} \frac{\tilde{s}_1}{dA_{d-1}} \frac{1-\rho^2}{(1-2\rho\tilde{s_1}+\rho^2)^{d/2}} d\tilde{s} \\
&\qquad = \frac{1-\rho^2}{dA_{d-1}} \frac{2\pi^{(d-1)/2}}{\Gamma\{\frac{1}{2}(d-1)\}} \int_0^\pi \frac{\cos\theta \sin^{d-2}\theta}{(1-2\rho\cos\theta+\rho^2)^{d/2}} d\theta \\
&\qquad = \frac{1-\rho^2}{dB\{\frac{1}{2}(d-1), \frac{1}{2}\}} \int_{-1}^1 \frac{t\,(1-t^2)^{(d-3)/2}}{(1-2\rho t+\rho^2)^{d/2}} dt \\
&\qquad = \frac{\rho}{d}.
\end{aligned}$$

The other elements, $E(R_j S_j)$ $(2 \le j \le d)$, are calculated in a similar way. $\square$

Proof of Theorem 4.3 Consider $T \equiv U'QV$. The distribution function of T, F_T, is given by

$$\begin{aligned}
F_T(t) &= P(T \le t) = E_V\left\{P(U'Qv \le t \,|\, V = v)\right\} \\
&= E_{\tilde{V}}\left\{P(U'\tilde{v} \le t \,|\, \tilde{V} = \tilde{v})\right\},
\end{aligned}$$

where $\tilde{V} = QV$. Then transform $\tilde{U} = PU$ where $P \in O(d)$ such that $P = (\tilde{v}, p_2, \ldots, p_d)'$, $p_j \in \mathbb{R}^d$, and one obtains

$$\begin{aligned}
E_{\tilde{V}}\left\{P(U'\tilde{v} \le t \,|\, \tilde{V} = \tilde{v})\right\} &= \int_{S_{d-1}} \frac{1}{A_{d-1}} \int_{\substack{\tilde{u}_1 \le t \\ \tilde{u} \in S_{d-1}}} \frac{1}{A_{d-1}} \frac{1-\rho^2}{1-2\rho\tilde{u}_1+\rho^2} d\tilde{u} d\tilde{v} \\
&= \frac{1}{A_{d-1}} \frac{2\pi^{(d-1)/2}}{\Gamma\{\frac{1}{2}(d-1)\}} \int_{\substack{\cos\theta \le t \\ 0 \le \theta < \pi}} \frac{(1-\rho^2)\sin^{d-2}\theta}{(1-2\rho\cos\theta+\rho^2)^{d/2}} d\theta \\
&= \frac{1-\rho^2}{B\{\frac{1}{2}(d-1), \frac{1}{2}\}} \int_{-1}^t \frac{(1-x^2)^{(d-3)/2}}{(1-2\rho x+\rho^2)^{d/2}} dx.
\end{aligned}$$

Thus,

$$f_T(t) = \frac{dF_T}{dt}(t) = \frac{1-\rho^2}{B\{\frac{1}{2}(d-1), \frac{1}{2}\}} \frac{(1-t^2)^{(d-3)/2}}{(1-2\rho t+\rho^2)^{d/2}}. \quad \square$$

Acknowledgements

First and foremost, I would like to thank Professor Kunio Shimizu for his supervision over the entire work. My thanks also go to Professors Ritei Shibata, Makoto Maejima and Nobuo Shinozaki for various comments on the paper. Concerning my work in Chapter 2, I am grateful to Dr. Nicholas I. Fisher, Dr. Peter J. Thomson, Dr. Takeshi Kato, the Associate Editor of *Statistica Sinica* and two anonymous referees for their suggestions. My sincere gratitude goes to Dr. Arthur Pewsey and the referees for their advice on the work in Chapters 3 and 4.

Financial support for the research was provided, in part, by the Ministry of Education, Culture, Sport, Science and Technology in Japan under a Grant-in-Aid of the 21st Century Center of Excellence for Integrative Mathematical Sciences: Progress in Mathematics Motivated by Social and Natural Sciences.

Bibliography

Abramowitz, M. & Stegun, I.A. (1970). *Handbook of Mathematical Functions.* Dover Press, New York.

Alfonsi, A. & Brigo, D. (2005). New families of copulas based on periodic functions. *Commun. Statist.–Theory Methods* **34**, 1437–1447.

Arnold, B.C., Castillo, E. & Sarabia, J.M. (1999). *Conditional Specification of Statistical Models.* Springer, New York.

Arnold, B.C. & Strauss, D.J. (1991). Bivariate distributions with conditionals in prescribed exponential families. *J. R. Statist. Soc.* **B 55**, 365–375.

Batschelet, E. (1981). *Circular Statistics in Biology.* Academic Press, London.

Downs, T.D. (1966). Some relationships among the von Mises distributions of different dimensions. *Biometrika* **53**, 269–272.

Downs, T.D. (2003). Spherical regression. *Biometrika* **90**, 655–668.

Downs, T.D. & Mardia, K.V. (2002). Circular regression. *Biometrika* **89**, 683–697.

Drouet Mari, D. & Kotz, S. (2001). *Correlation and Dependence.* World Scientific Publ., New Jersey.

Durrett, R. (1984). *Brownian Motion and Martingales in Analysis.* Wadsworth, Belmont, CA.

Dym, H. & McKean, H.P. (1972). *Fourier Series and Integrals.* Academic Press, New York.

Fisher, N.I. (1993). *Statistical Analysis of Circular Data.* Cambridge University Press, Cambridge.

Fisher, N.I. & Lee, A.J. (1983). A correlation coefficient for circular data. *Biometrika* **70**, 327–332.

Fisher, N.I. & Lee, A.J. (1992). Regression models for an angular response. *Biometrics* **48**, 665–677.

Fisher, N.I., Lewis, T. & Embleton, B.J.J. (1987). *Statistical Analysis of Spherical Data*. Cambridge University Press, Cambridge.

Jammalamadaka, S.R. & SenGupta, A. (2001). *Topics in Circular Statistics.* World Scientific Publ., New Jersey.

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

Johnson, R.A. & Wehrly, T.E. (1978). Some angular-linear distributions and related regression models. *J. Amer. Statist. Assoc.* **73**, 602–606.

Jones, M.C. (2002). Marginal replacement in multivariate densities, with application to skewing spherically symmetric distributions. *J. Multivariate Anal.* **81**, 85–99.

Jones, M.C. (2004). The Möbius distribution on the disc. *Ann. Inst. Statist. Math.* **56**, 733–742.

Jones, M.C. & Pewsey, A. (2005). A family of symmetric distributions on the circle. *J. Amer. Statist. Assoc.* **100**, 1422–1428.

Jupp, P.E. & Mardia, K.V. (1980). A general correlation coefficient for directional data and related regression problems. *Biometrika* **67**, 163–173.

Kagan, A.M., Linnik, Yu.V. & Rao, C.R. (1973). *Characterization Problems in Mathematical Statistics*. Wiley, New York.

Kato, S. A distribution for a pair of unit vectors generated by Brownian motion, submitted for publication. (Technical Report version: KSTS/RR-06/010, Department of Mathematics, Keio University, Yokohama, 2006).

Kato, S. & Shimizu, K. Dependent models for observations which include angular ones. *J. Statist. Plann. Inference*, to appear.

Kato, S., Shimizu, K. & Shieh, G.S. A circular–circular regression model. *Statist. Sinica*, to appear.

Kent, J.T. & Tyler, D.E. (1988). Maximum likelihood estimation for the wrapped Cauchy distribution. *J. Appl. Statist.* **15**, 247–254.

Kolassa, J. & McCullagh, P. (1990). Edgeworth series for lattice distributions. *Ann. Statist.* **18**, 981–985.

Leipnik, R.B. (1947). Distribution of the serial correlation coefficient in a circularly correlated universe. *Ann. Math. Statist.* **18**, 80–87.

Lund, U.J. (1999). Least circular distance regression for directional data. *J. Appl. Statist.* **26**, 723–733.

Maksimov, V.M. (1967). Necessary and sufficient statistics for a family of shifts of probability distributions on continuous bicompact groups. *Teor. Veroyatn. Primen.*, **12**, 307–321 (in Russian). English translation: *Theory Probab. Appl.*, **12**, 267–280.

Mardia, K.V. (1972). *Statistics of Directional Data*. Academic Press, New York.

MARDIA, K.V. (1975). Statistics of directional data (with discussion). *J. R. Statist. Soc.* **B 37**, 349–393.

MARDIA, K.V. & JUPP, P.E. (1999). *Directional Statistics.* Wiley, Chichester.

MARDIA, K.V. & SUTTON, T.W. (1978). A model for cylindrical variables with applications. *J. R. Statist. Soc.* **B 40**, 229–233.

MCCULLAGH, P. (1989). Some statistical properties of a family of continuous univariate distributions. *J. Amer. Statist. Assoc.* **84**, 125–129.

MCCULLAGH, P. (1992). Conditional inference and Cauchy models. *Biometrika* **79**, 247–259.

MCCULLAGH, P. (1996). Möbius transformation and Cauchy parameter estimation. *Ann. Statist.* **24**, 787–808.

MINH, D.L.P. & FARNUM, N.R. (2003). Using bilinear transformations to induce probability distributions. *Commun. Statist.–Theory Methods* **32**, 1–9.

NELSEN, R.B. (1998). *An Introduction to Copulas.* Springer Verlag, New York.

PEWSEY, A. (2000). The wrapped skew-normal distribution on the circle. *Commun. Statist.–Theory Meth.* **29**, 2459–2472.

PEWSEY, A. (2002). Testing circular symmetry. *Canad. J. Statist.* **30**, 591–600.

RIVEST, L.-P. (1984). Symmetric distributions for dependent unit vectors. *Ann. Statist.* **12**, 1050–1057.

RIVEST, L.-P. (1988). A distribution for dependent unit vectors. *Commun. Statist.–Theory Methods* **17**, 461–483.

RIVEST, L.-P. (1997). A decentred predictor for circular–circular regression. *Biometrika* **84**, 717–726.

Rudin, W. (1987). *Real and Complex Analysis*. McGraw-Hill, New York.

Rukhin, A.L. (1972). Some statistical decisions about distributions on a circle for large samples. *Sankhyā Ser.* **A 34**, 243–250.

Saw, J.C. (1983). Dependent unit vectors. *Biometrika* **70**, 665–671.

Self, S.G. & Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Amer. Statist. Assoc.* **82**, 605–610.

SenGupta, A. (2004). On the construction of probability distributions for directional data. *Bull. Calcutta Math. Soc.* **96(2)**, 139–154.

Seshadri, V. (1991). A family of distributions related to the McCullagh family. *Statist. Probab. Lett.* **12**, 373–378.

Shieh, G.S. & Johnson, R.A. (2005). Inference based on a bivariate distribution with von Mises marginals. *Ann. Inst. Statist. Math.* **57**, 789–802.

Shimizu, K. & Iida, K. (2002). Pearson type VII distributions on spheres. *Commun. Statist.–Theory Meth.* **31**, 513–526.

Stephens, M.A. (1979). Vector correlation. *Biometrika* **66**, 41–48.

Tashiro, Y. (1977). On methods for generating uniform random points on the surface of a sphere. *Ann. Inst. Statist. Math.* **29**, 295–300.

Watson, G.S. (1983). *Statistics on Spheres*. Wiley, New York.

Wehrly, T.E. & Johnson, R.A. (1980). Bivariate models for dependence of angular observations and a related Markov process. *Biometrika* **67**, 255–256.

Yfantis, E.A. & Borgman, L.E. (1982). An extension of the von Mises distribution. *Commun. Statist.–Theory Meth.* **11**, 1695–1706.