Research Memorandum No. 1086 01/30/2009

A Markov process for circular data

Shogo Kato (Institute of Statistical Mathematics, Japan)

The Institute of Statistical Mathematics

4-6-7 Minami-Azabu, Minato-ku,
Tokyo, 106-8569, Japan

A Markov process for circular data

Shogo Kato

Institute of Statistical Mathematics
4-6-7 Minami-Azabu, Minato-ku, Tokyo, 106-8569, Japan
E-mail: skato@ism.ac.jp

Abstract

We propose a discrete-time Markov process which takes values on the unit circle. Some properties of the process, including the limiting behaviour and ergodicity, are investigated. Many computations associated with this process are shown to be greatly simplified if the variables and parameters of the model are represented in terms of complex numbers. A further discussion is given on some submodels, especially, on the stationary process. The proposed model is compared with some existing Markov processes for circular data. Statistical inference for the process is considered. Finally, an application of the model to wind direction data is provided.

Keywords: Circular autocorrelation; Circular time series; Möbius transformation; Robustness; Wrapped Cauchy distribution

1 Introduction

Data which can be expressed as sets of observations on the circle arise in a number of areas of applications such as biology, meteorology and geology. On occasions, circular observations appear in a time series context. For instance, a series of wind directions measured hourly at a weather station (Fisher and Lee, 1994) can be considered an example of time series of circular data. Other examples of circular time series are seen in Cameron (1983) and Breckling (1989, Part I).

For the analysis of this kind of data, some stochastic processes have been proposed in the literature. Wehrly and Johnson (1980) proposed a Markov process by applying a class of bivariate circular distributions with specified marginals. Breckling (1989, Chapter 6) proposed two stochastic processes, namely, the von Mises process and the wrapped autoregressive process, and fitted these models to time series of wind directions. Fisher and Lee (1994) discussed the models of Breckling (1989) and proposed new models based on a projection method and a link function concept. Hidden Markov models for circular time series were presented by Holzmann *et al.* (2006). See Fisher (1993, Chapter 7), Mardia and Jupp (1999, Section 11.5.2) and Jammalamadaka and SenGupta (2001, Section 12.8) for overviews of the time series models for circular data.

In this paper we provide a new discrete-time Markov process (Markov chain) on the circle. The model can be derived based on the regression idea of Kato *et al.* (2008), who provided a circular–circular regression model by adapting the Möbius circle transformation as a regression curve and the wrapped Cauchy distribution as an angular error.

As seen in some existing works such as McCullagh (1996) and Kato *et al.* (2008), the wrapped Cauchy distribution has some tractable features, one of which is related to the Möbius circle transformation. By applying these results, some desirable properties of the process, including the limiting behaviour and ergodicity, are obtained. To simplify many computations associated with the process, we represent the variables and parameters in terms of complex numbers.

The proposed model could be useful to describe a circular time series for which the mean direction and concentration of the state at the time, say, n approach certain values as n increases. Or our model can also be used to fit stationary circular time series data. In both situations the proposed model can be applied as a robust model because of the heavy tail the wrapped Cauchy distribution has. We present an application of our model to a time series of wind directions to illustrative an advantage of the model.

The subsequent sections are organised as follows. Section 2 provides some preliminary knowledge about the Möbius transformation and wrapped Cauchy distribution, which play an essential role in the proposed model. In Section 3 we propose a Markov process and investigate its properties. Also, we illustrate the interpretation of the parameters and the limiting behaviour of the process by simulating the proposed processes for specified values of the parameters. Section 4 concerns some submodels of the process proposed in the previous section. In particular we pay the most attention to the stationary case of our model. A comparison with some existing Markov models is in Section 5. Parameter estimation based on maximum likelihood and method of moments is briefly discussed in Section 6. In Section 7 the proposed stationary model is fitted to a dataset of wind directions to illustrate an advantage of our model. Finally, concluding remarks are made in Section 8.

2 Preliminaries

Before we embark on the main topic, we briefly introduce some preliminary knowledge about the Möbius transformation and the wrapped Cauchy distribution. This background is central to investigating the properties of the Markov process which we propose in Section 3.

2.1 Möbius transformation

The Möbius transformation is defined as

$$\mathcal{M}(x) = \frac{a_{00}x + a_{01}}{a_{10}x + a_{11}}, \quad x \in \mathbb{C}; \ \begin{pmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{pmatrix} \in \mathrm{GL}(2, \mathbb{C}), \tag{1}$$

where $\mathrm{GL}(2, \mathbb{C})$ is a group of 2×2 regular matrices of which each element is a complex number. The transformation is well known as a projection which maps the complex plane $\mathbb{C}$ onto itself. As seen in works such as McCullagh (1996) and Jones (2004), this transformation can play an important role in directional statistics.

In particular, we consider a subclass of the Möbius transformation with constraints

$$\begin{pmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{pmatrix} = \begin{pmatrix} 1 & \beta \\ \overline{\beta} & 1 \end{pmatrix}, \quad \beta \in D.$$

This transformation is a conformal mapping which maps the unit disc, $D = \{z \in \mathbb{C}\,;\, |z| < 1\}$, onto itself. In addition, the unit circle, $\partial D = \{z \in \mathbb{C}\,;\, |z| = 1\}$, is also mapped onto itself via the transformation. See Jones (2004) and Kato *et al.* (2008) for details about this transformation. Following the convention in the latter paper, we call this mapping the Möbius circle transformation.

For convenience, write (1) as

$$\frac{a_{00}x + a_{01}}{a_{10}x + a_{11}} = \begin{pmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{pmatrix} \circ x.$$

It is known that the following property holds for the Möbius transformation:

$$A \circ (B \circ x) = (AB) \circ x, \quad A, B \in \mathrm{GL}(2, \mathbb{C}). \tag{2}$$

2.2 Wrapped Cauchy distribution

A random variable Z is said to have the wrapped Cauchy distribution if it has the density

$$f(z; \phi) = \frac{1}{2\pi} \frac{\left|1 - |\phi|^2\right|}{|z - \phi|^2}, \quad z \in \partial D;\ \phi \in \mathbb{C} \setminus \partial D, \tag{3}$$

with respect to arc length on the circle μ. The model is also called the circular Cauchy distribution as seen in McCullagh (1996). In this paper we extend the domain of ϕ and define $Z = \phi$ for $\phi \in \partial D$. Here $\mathrm{Arg}(\phi)$ or $\phi/|\phi|$ is the mean direction for $\phi \neq 0$ and $|\phi|$ the mean resultant length of Z for $\phi \in \overline{D}$, where $\mathrm{Arg}(z)$ denotes the complex argument of z taking values between $[-\pi, \pi)$ and $\overline{D} = D \cup \partial D$. As discussed in McCullagh (1996), it is the case that $f(z; \phi) = f(z; 1/\overline{\phi})$. The distribution is unimodal and symmetric about $z = \phi/|\phi|$. 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 $Z = \phi$. In the same way as McCullagh (1996), we denote the wrapped Cauchy distribution in (3) by $Z \sim C^*(\phi)$.

The properties of the wrapped Cauchy distribution have been intensively investigated by McCullagh (1996). (See Fisher (1993, Section 3.3.4), Mardia and Jupp (1999, pp.51–52) and Jammalamadaka and SenGupta (2001, Section 2.2.7) for book treatments of the wrapped Cauchy distribution.) The following hold for the wrapped Cauchy distribution:

$$Z \sim C^*(\phi) \implies \alpha Z \sim C^*(\alpha\phi), \quad \alpha \in \partial D, \tag{4}$$

$$Z_1 \sim C^*(\phi_1),\ Z_2 \sim C^*(\phi_2),\ |\phi_1|, |\phi_2| \le 1,\ Z_1 \perp Z_2 \implies Z_1 Z_2 \sim C^*(\phi_1\phi_2), \tag{5}$$

$$Z \sim C^*(\phi) \implies \begin{pmatrix} 1 & \beta \\ \overline{\beta} & 1 \end{pmatrix} \circ Z \sim C^*\left\{\begin{pmatrix} 1 & \beta \\ \overline{\beta} & 1 \end{pmatrix} \circ \phi\right\}, \quad \beta \in \mathbb{C}. \tag{6}$$

3 A Markov process for circular data

3.1 Definition

Kato *et al.* (2008) proposed a circular–circular regression model by using the Möbius circle transformation as a regression curve and the wrapped Cauchy distribution as an angular error. In this paper we adapt their regression model to construct a discrete-time Markov process for circular data. A class of Markov processes is provided in the following theorem.

Theorem 1. *Let W_0 be a random variable or a constant which takes values on ∂D. Assume that $\{W_n\}_{n=1}^{\infty}$ is a sequence of random variables defined by*

$$W_n = \frac{W_{n-1} + \beta}{\overline{\beta} W_{n-1} + 1}\, \varepsilon_n, \quad n = 1, 2, \ldots,$$

where $\beta \in D$ and ε_n is a ∂D-valued random variable for any $n \in \mathbb{N}$. Then $\{W_n\}_{n=0}^{\infty}$ is a discrete-time Markov process which takes values on ∂D.

Proof. Clearly, the Markov property and time homogeneity hold for $\{W_n\}_{n=0}^{\infty}$. Since the Möbius circle transformation maps ∂D onto itself, it follows that $\{W_n\}_{n=0}^{\infty}$ takes values on ∂D. □

From now on, we assume that $\varepsilon_1, \varepsilon_2, \ldots$ in Theorem 1 are iid random variables which are independent of W_0 and are distributed as the wrapped Cauchy distribution $C^*(\varphi)$, $0 \leq \varphi < 1$. We call this process the Möbius Markov process.

3.2 Some properties

We discuss some properties of the proposed process. Throughout this subsection, we assume that $\{W_n\}_{n=0}^{\infty}$ is the Möbius Markov process.

From property (2), it follows that W_{t+n} can be expressed as a function of W_t, $\varepsilon_{t+1}, \ldots, \varepsilon_{t+n}$ as follows.

$$W_{t+n} = \left\{ \begin{pmatrix} \varepsilon_{t+n} & \beta\varepsilon_{t+n} \\ \overline{\beta} & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_{t+n-1} & \beta\varepsilon_{t+n-1} \\ \overline{\beta} & 1 \end{pmatrix} \cdots \begin{pmatrix} \varepsilon_{t+1} & \beta\varepsilon_{t+1} \\ \overline{\beta} & 1 \end{pmatrix} \right\} \circ W_t,$$

where $t \geq 0$ and $n \geq 1$. In the later discussion, we assume, without loss of generality, that $t = 0$. Using properties (2) and (4)–(6), the following theorem is readily established.

Theorem 2.

$$W_n \,|\, (W_0 = w_0) \sim C^* \left\{ \phi_n(w_0) \right\},$$

where

$$\phi_n(w_0) = \begin{pmatrix} \varphi & \beta\varphi \\ \overline{\beta} & 1 \end{pmatrix}^n \circ w_0, \quad n \geq 0.$$

Thus, the conditional of W_n given $W_0 = w_0$ is the wrapped Cauchy for any n. By using mathematical induction, it can be proved that the parameter of the above conditional can be expressed in another form as follows.

$$\begin{pmatrix} \varphi & \beta\varphi \\ \overline{\beta} & 1 \end{pmatrix}^n \circ w_0 = \left\{ \begin{pmatrix} \varphi & |\beta|\varphi \\ |\beta| & 1 \end{pmatrix}^n \circ \frac{\overline{\beta}}{|\beta|} w_0 \right\} \frac{\beta}{|\beta|} \tag{7}$$

This representation will be used later in the paper.

By the above theorem and Theorem 1 of Kato *et al.* (2008), the pth trigonometric moment of $W_n \,|\, (W_0 = w_0)$ is given by

$$E(W_n^p \,|\, W_0 = w_0) = \begin{cases} \phi_n^p(w_0), & p \geq 0, \\ \overline{\phi_n^{-p}(w_0)}, & p < 0, \end{cases} \quad \text{for } p \in \mathbb{Z}.$$

In particular, $\text{Arg}\{\phi_n(w_0)\}$ and $|\phi_n(w_0)|$ are the mean direction and the mean resultant length of $W_n \,|\, (W_0 = w_0)$, respectively.

Given a process as in Theorem 1, a natural question to address is the limiting behaviour of the process, which we describe in the following lemma. See Appendix A for the proof.

Lemma 1.

$$W_n \,|\, (W_0 = w_0) \xrightarrow{\text{d}} C^*(\phi_\infty) \quad \textit{as} \quad n \longrightarrow \infty,$$

where

$$\phi_\infty = \begin{cases} \dfrac{\varphi - 1 + \sqrt{\{(1-\varphi)^2 + 4\varphi|\beta|^2\}}}{2|\beta|} \dfrac{\beta}{|\beta|}, & \beta \in D \setminus \{0\}, \\ 0, & \beta = 0. \end{cases}$$

For convenience, we write π to denote the distribution and the density of $C^*(\phi_\infty)$. Also, define a transition kernel as

$$P^n(w, A) = \int_A \frac{1}{2\pi} \frac{1 - |\phi_n(w)|^2}{|z - \phi_n(w)|^2} \mu(dz),$$

where $A \,(\subset \partial D)$ is a measurable set.

Lemma 2. *The unique invariant distribution of the Möbius Markov process is given by* π*.*

Proof. See Appendix B. □

It is remarked that, in general, π is not a reversible distribution except for some special cases discussed in Section 4.3.

A Markov process is said to be *ergodic* if it is positive Harris recurrent and aperiodic. (See, for example, Meyn and Tweedie (1993, pp.116, 200, 230–231) for the definition of positive Harris recurrence and aperiodicity.)

Theorem 3. *The Möbius Markov process is ergodic.*

Proof. It is clear from Lemma 2 that the Möbius Markov process is π-irreducible and $\pi P = \pi$ holds. For each measurable set A with $\pi(A) = 0$, which is equivalent to the condition that A is a null set, it holds that $P(w_0, A) = 0$ for all $w_0 \in \partial D$. This means that $P(w_0, \cdot)$ is absolutely continuous with respect to π for all w_0. Hence Corollary 1 of Tierney (1994) implies that P is positive Harris recurrent. Aperiodicity of the process is clear from Lemma 1. □

Next we consider the orbit of a sequence of the parameters $\{\phi_n(w_0)\}_{n=0}^{\infty}$ which we already know converges to ϕ_∞ as n tends to infinity. For any $\beta \in D \setminus \{0\}$ and $n \geq 1$,

$$\text{Re}\left\{\overline{\beta} E(W_n \,|\, W_0 = w_0)\right\} \geq \text{Re}\left\{\overline{\beta} E(W_{n-1} \,|\, W_0 = w_0)\right\}.$$

The following theorem describes how a sequence of the parameters $\{\phi_n(w_0)\}$ approaches ϕ_∞.

Theorem 4. *Let $\{\phi_n(w_0)\}_{n=0}^{\infty}$ be a sequence of parameters defined in Theorem 2. Then $\{\phi_n(w_0)\}_{n=0}^{\infty}$ lies on the arc or on the line segment with equation*

$$g(\lambda) = \begin{pmatrix} \xi_1 w_0 - 2\beta\varphi & \xi_2 w_0 + 2\beta\varphi \\ \xi_2 - 2\overline{\beta} w_0 & \xi_1 + 2\overline{\beta} w_0 \end{pmatrix} \circ \lambda, \quad 0 \leq \lambda \leq 1, \tag{8}$$

where $\xi_1 = 1 - \varphi + \surd\{(1-\varphi)^2 + 4\varphi|\beta|^2\}$ and $\xi_2 = \varphi - 1 + \surd\{(1-\varphi)^2 + 4\varphi|\beta|^2\}$.

Proof. By looking at equation (15) as a function of $\lambda = (\lambda_1/\lambda_2)^n$, we obtain equation (8). Since the Möbius transformation maps the real line onto the circle or the line in the complex plane (see Rudin (1987, Section 14.3)), it follows that $g(\lambda)$ takes values on an arc or a line segment in the complex plane. □

In particular, $g(1) = w_0$ and $g(0) = \phi_\infty$. If $w_0 \neq \pm\beta/|\beta|$, $g(\lambda)$ takes values on the circle whose center and radius are given by

$$-i\,\frac{(\xi_1+\xi_2)(1-\varphi)(w_0+\beta)}{2\,\mathrm{Im}(\xi_2\beta\overline{w}_0 - \xi_1\overline{\beta}w_0)} \quad \text{and} \quad \left\{\left|\frac{(\xi_1+\xi_2)(1-\varphi)(w_0+\beta)}{2\,\mathrm{Im}(\xi_2\beta\overline{w}_0 - \xi_1\overline{\beta}w_0)}\right|^2 + \varphi\right\}^{1/2},$$

respectively. In equation (8), $g(\lambda)$ coincides with $\{\phi_n(w_0)\}_{n=0}^{\infty}$ when

$$\lambda = \lambda(n) = \left(\frac{1+\varphi-\surd\{(1-\varphi)^2+4\varphi|\beta|^2\}}{1+\varphi+\surd\{(1-\varphi)^2+4\varphi|\beta|^2\}}\right)^n, \quad n = 0, 1, \ldots$$

Thus, the rate of convergence is

$$r(|\beta|, \varphi) \equiv \left|\frac{\lambda(n+1)}{\lambda(n)}\right| = \frac{1+\varphi-\surd\{(1-\varphi)^2+4\varphi|\beta|^2\}}{1+\varphi+\surd\{(1-\varphi)^2+4\varphi|\beta|^2\}}. \tag{9}$$

From this, it follows that $|\beta|$ and φ influence the rate of convergence. Clearly, r is monotonically decreasing with respect to $|\beta|$. Hence, as $|\beta|$ increases, the process converges at the higher speed. In particular, $r \to 0$ ($|\beta| \to 1$) and $r \to \varphi$ ($|\beta| \to 0$). On the other hand, r is monotonically increasing as a function of φ. Here we get $r \to (1-|\beta|)/(1+|\beta|)$ ($\varphi \to 1$) and $r \to 0$ ($\varphi \to 0$). Thus the higher the value of φ, the slower the convergence of the process.

Finally, we discuss a method to make a manual plot of $\{\phi_n(w_0)\}_{n=0}^{\infty}$ in the following theorem. The proof is lengthy but straightforward, and therefore omitted.

Theorem 5. *Let Ω be a Riemann sphere in $\mathbb{C}\times\mathbb{R}$, i.e., $\Omega = \{(z,x) \in \mathbb{C}\times\mathbb{R}\,;\, |z|^2+x^2 = 1\}$. Suppose $\tilde{\phi}_n$ is the point where the line joining the north pole $N(=(0,1))$ and $(\phi_n(w_0), 0)$ crosses Ω. Let ω be the point where $-\tilde{\phi}_n$ is transformed via a straight-line projection through the point $(\beta, 0)$ onto the opposite side of Ω. Then the line from N through ω intersects the plane $\{(z, 1-\varphi)\,;\, z \in \mathbb{C}\}$ at the point $(\phi_{n+1}(w_0), 1-\varphi)$.*

3.3 Simulation

In this subsection we conduct further discussion about the interpretation of the parameters and the limiting behaviour of the process by simulating the Möbius Markov process for specified values of the parameters.

For simulation of a Möbius Markov process, it is necessary to generate random variables from the wrapped Cauchy distribution. A $C^*(\beta)$ random variable is generated by the following two steps:

STEP 1: Generate a uniform $(0, 1)$ random number U.

STEP 2: Put $Z = \begin{pmatrix} 1 & \beta \\ \bar{\beta} & 1 \end{pmatrix} \circ \exp(2\pi i U)$.

Then it follows from property (6) that Z has the wrapped Cauchy $C^*(\beta)$.

Fig. 1 displays simulations of the Möbius Markov process for specified values of φ and β. This figure explicitly shows that as n converges to infinity, the mean direction of the conditional of $\text{Arg}(W_n)$ given $W_0 = w_0$ tends to $\text{Arg}(\beta)$, and this is mathematically validated by Lemma 1. By comparing the first two frames of Fig. 1, it seems that $|\beta|$ influences the speed of convergence r and the parameter of the limiting distribution $|\phi_\infty|$. Actually, as stated in Section 3.2, $|\phi_\infty|$ and the negative of the rate of convergence (9), $-r$, are monotonically increasing with respect to $|\beta|$. From frames (a) and (c) of Fig. 1, it seems that the smaller the value of φ, the smaller the concentration of the limiting distribution. In addition, φ influences the rate of convergence (9) which is monotonically decreasing. Finally, comparing Fig. 1(a) and (d), we find that the parameter $\text{Arg}(\beta)$ controls the mean direction of the limiting distribution, i.e., $\text{Arg}(\phi_\infty)$.

As already discussed in Theorem 2, the conditional distribution of W_n given $W_0 = w_0$ is the wrapped Cauchy distribution. The parameters of this conditional distribution for specified values of β and φ are exhibited in Fig. 2. Note that sequences of parameters, $\{\phi_n(w_0)\}_{n=0}^{\infty}$, lie on circles as Theorem 4 shows. This figure is also helpful to interpret how the parameters influence the mean direction $\text{Arg}\{\phi_n(w_0)\}$ and mean resultant length $|\phi_n(w_0)|$. We omit the comparison between the four frames since this figure provides a very similar interpretation of the parameters to that in Fig. 1.

Summarising the results in Section 3.2 and these figures, the interpretation of each parameter is given as follows. As stated in Lemma 1 and as is clear from the two figures, the parameter $\text{Arg}(\beta)$ controls the mean direction of the limiting distribution, $\text{Arg}(\phi_\infty)$. The parameter $|\beta|$ determines the rate of convergence and the mean resultant length of the limiting distribution $|\phi_\infty|$. This interpretation is mathematically validated by Lemma 1 and the fact that r is monotonically decreasing as a function of $|\beta|$. In particular, $|\phi_\infty| \to 1$ as $|\beta| \to 1$. Also, φ influences the rate of convergence and the concentration of ϕ_∞ as shown in Lemma 1 and the fact that $r = r(\varphi)$ is monotonically decreasing for fixed $|\beta|$.

We note that although both φ and $|\beta|$ control the rate of convergence and the concentration, their roles are completely different. For example, $|\phi_\infty|$ is monotonically increasing with respect to φ or $|\beta|$, but r is monotonically decreasing as a function of φ whereas it is monotonically increasing with respect to $|\beta|$.

4 Submodels of the process

In this section we consider some submodels of the Möbius Markov process. The Markov process having the wrapped Cauchy initial distribution is discussed in Section 4.1. In Section 4.2 the stationary process and its autocorrelation coefficient are considered. Finally we focus on the submodel of the stationary process with uniform marginals in Section 4.3.

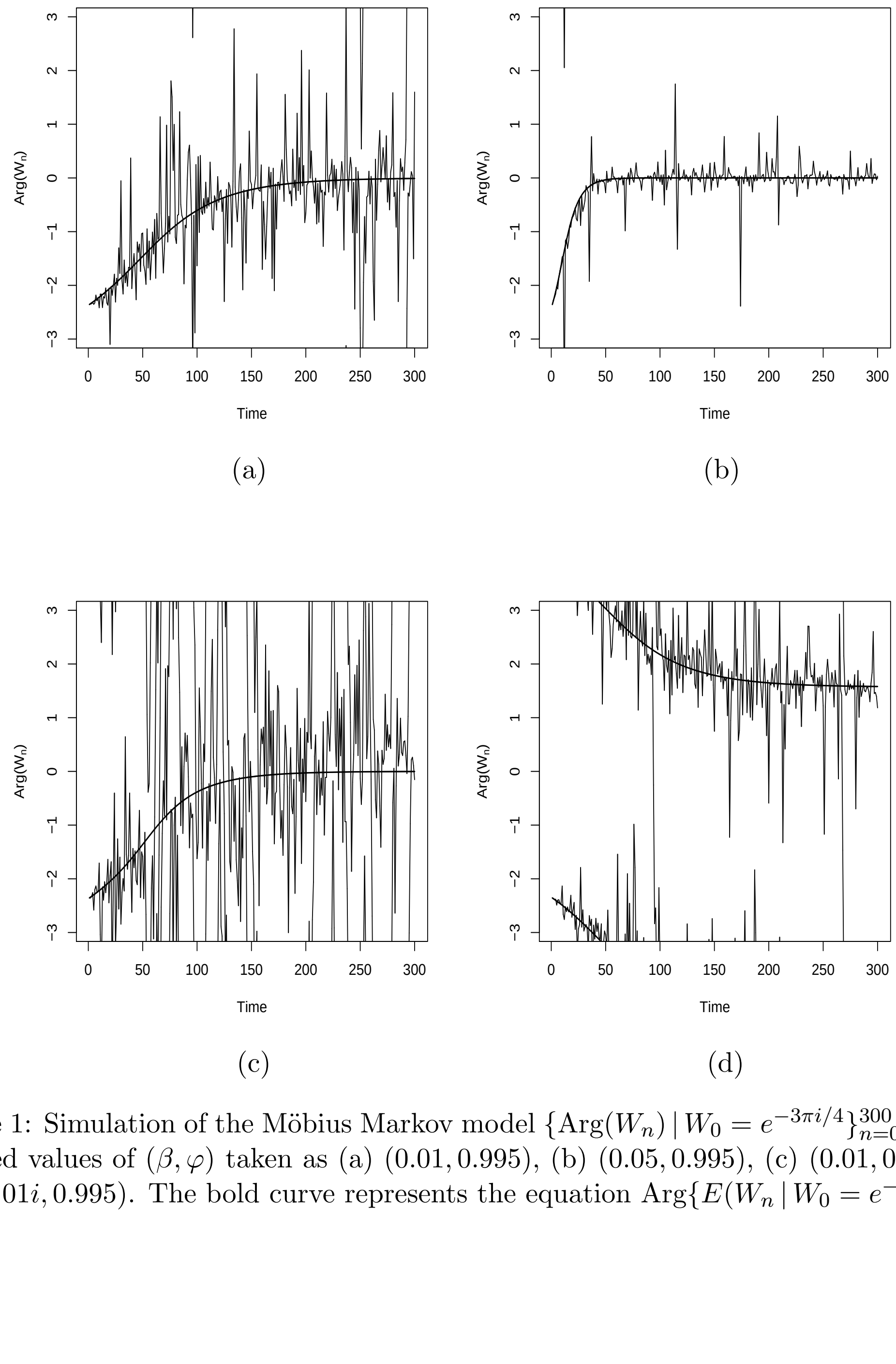

Figure 1: Simulation of the Möbius Markov model $\{\mathrm{Arg}(W_n) \,|\, W_0 = e^{-3\pi i/4}\}_{n=0}^{300}$ for some selected values of (β, φ) taken as (a) $(0.01, 0.995)$, (b) $(0.05, 0.995)$, (c) $(0.01, 0.985)$ and (d) $(0.01i, 0.995)$. The bold curve represents the equation $\mathrm{Arg}\{E(W_n \,|\, W_0 = e^{-3\pi i/4})\}$.

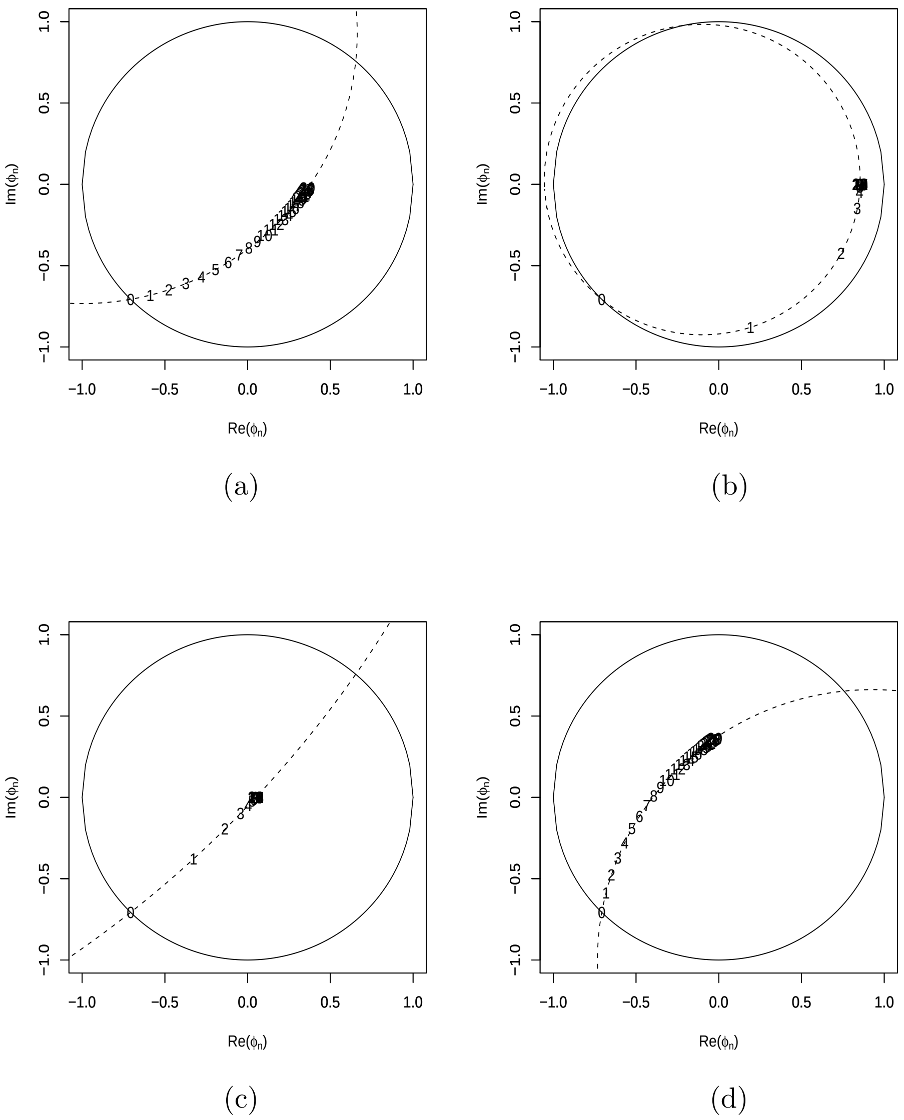

Figure 2: Plot of the parameters, $\{\phi_n(w_0)\}_{n=0}^{30}$, of the wrapped Cauchy distributions for the Möbius Markov process given in Theorem 2 for $w_0 = e^{-3\pi i/4}$ and the parameters: (a) $(\beta, \varphi) = (0.05, 0.9)$, (b) $(\beta, \varphi) = (0.5, 0.9)$, (c) $(\beta, \varphi) = (0.05, 0.5)$ and (d) $(\beta, \varphi) = (0.05i, 0.9)$. The dotted line represents the circle (8).

4.1 Wrapped Cauchy initial distribution

The following lemma shows that the Möbius Markov process has wrapped Cauchy marginals if the initial distribution follows the wrapped Cauchy. The proof is clear from properties (4)–(6).

Lemma 3. *Assume* $W_0 \sim C^*(\phi_0)$ *where* $\phi_0 \in \overline{D}$. *Then* $W_n \sim C^*\left\{\phi_n(\phi_0)\right\}, \, n \geq 0$.

Remark that, for the above model, the marginal of W_j and the conditional of W_j given $W_k = w_k$ $(j > k \geq 0)$ have the wrapped Cauchy distribution. Applying Theorem 4, it is easy to show that a sequence of parameters $\{\phi_n(w_0)\}_{n=0}^{\infty}$ given in Lemma 3 lies on the arc or on the line segment.

4.2 Stationary process

Here we focus on a submodel which can be obtained on putting $\phi_0 = \phi_\infty$ for the Markov process given in Lemma 3. As stated in Lemma 2, this process is strictly stationary. Some moments of the process are provided in following lemma. See Appendix C for the proof.

Lemma 4. *Let* $\{W_n\}_{n=0}^{\infty}$ *be the Möbius Markov process having the initial distribution* $W_0 \sim C^*(\phi_\infty)$. *Assume that* $\mathbf{V}_j = (Re(W_j), Im(W_j))'$ $(j = m, n)$. *Then the following equations hold*

$$E(\mathbf{V}_m) = E(\mathbf{V}_n) = Q\begin{pmatrix} |\phi_\infty| \\ 0 \end{pmatrix}, \quad E(\mathbf{V}_m \mathbf{V}_n') = Q\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} Q',$$

$$E(\mathbf{V}_m \mathbf{V}_m') = E(\mathbf{V}_n \mathbf{V}_n') = Q\begin{pmatrix} \frac{1}{2}(1+|\phi_\infty|^2) & 0 \\ 0 & \frac{1}{2}(1-|\phi_\infty|^2) \end{pmatrix} Q',$$

where

$$\begin{aligned}
Q &= \frac{1}{|\beta|}\begin{pmatrix} Re(\beta) & -Im(\beta) \\ Im(\beta) & Re(\beta) \end{pmatrix}, \\
a &= \frac{(a_{00}a_{11} + a_{01}a_{10})\{1 - |\phi_\infty a_{10}/a_{11}|^2 + (|\phi_\infty| - |a_{10}/a_{11}|)^2\}}{2(|a_{11}|^2 - |a_{10}|^2)(1 + |\phi_\infty a_{10}/a_{11}|)} \\
&\quad + \frac{(a_{00}a_{10} + a_{01}a_{11})(|\phi_\infty| - |a_{10}/a_{11}|)}{(|a_{11}|^2 - |a_{10}|^2)(1 + |\phi_\infty a_{10}/a_{11}|)}, \\
b &= \frac{(a_{00}a_{11} - a_{01}a_{10})\{1 - |\phi_\infty|^2 - |a_{10}/a_{11}|^2 + |\phi_\infty a_{10}/a_{11}|^2\}}{2(|a_{11}|^2 - |a_{10}|^2)(1 + |\phi_\infty a_{10}/a_{11}|)}, \\
\begin{pmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{pmatrix} &= \begin{pmatrix} \varphi & |\beta|\varphi \\ |\beta| & 1 \end{pmatrix}^{m-n}.
\end{aligned}$$

Note that these moments do not include any integrals or infinite sums. From this result, autocorrelation coefficients of this circular process are calculated as follows.

Theorem 6. *The Johnson and Wehrly (1977) correlation coefficient of* (W_m, W_n) $(m > n)$ *of the Markov process given in Lemma 4 is*

$$\rho_{JW} = \lambda^{1/2} = \frac{2\max\{|a - |\phi_\infty|^2|, |b|\}}{1 - |\phi_\infty|^2},$$

where λ *is the largest eigenvalue of* $\Sigma_{mm}^{-1}\Sigma_{mn}\Sigma_{nn}^{-1}\Sigma_{mn}'$*,* $\Sigma_{jk} = E(\mathbf{V}_j\mathbf{V}_k') - E(\mathbf{V}_j)E(\mathbf{V}_k')$ $(j,k = m,n)$*, and* $\mathbf{V}_\ell$ $(\ell = m,n)$ *is defined as in Lemma 4. The correlation coefficients of Jupp and Mardia (1980) and Fisher and Lee (1983) are given by*

$$\rho_{JM} = tr(\Sigma_{mm}^{-1}\Sigma_{mn}\Sigma_{nn}^{-1}\Sigma_{mn}') = \frac{4\{(a-|\phi_\infty|^2)^2+b^2\}}{(1-|\phi_\infty|^2)^2}$$

and

$$\rho_{FL} = \frac{det\{E(\mathbf{V}_m\mathbf{V}_n')\}}{\sqrt{[det\{E(\mathbf{V}_m\mathbf{V}_m')\}\, det\{E(\mathbf{V}_n\mathbf{V}_n')\}]}} = \frac{4ab}{1-|\phi_\infty|^4}, \tag{10}$$

respectively.

To compare our model with the models of Fisher and Lee (1994), here we consider the Fisher and Lee (1983) correlation coefficient which Fisher and Lee (1994) used as a measure of autocorrelation coefficient for circular time series. Fig. 3 plots their autocorrelation coefficients for some selected values of parameters. All frames of the figure show that as the lag between two variables increases, the autocorrelation between them decreases. This figure also implies that the greater the value of φ, the greater the autocorrelation coefficient between W_n and W_{n+h}. Also, the larger the value of $|\beta|$, the greater the autocorrelation coefficient.

Compared with Fig. 3 of Fisher and Lee (1994), our model shows similar correlation patterns to their linked autoregressive LAR(1) process when the concentration parameter φ of our model is large; Both models show exponential decay pattern as seen in the linear AR(1) model. Further comparison between our model and theirs will be given in the next section.

4.3 Stationary process with uniform marginals

Breckling (1989, Example 6.1) briefly considerd a stationary process with uniform marginals as a special case of the von Mises process. Fisher and Lee (1983) proposed a stationary process with uniform marginals by projecting two independent Gaussian processes.

In this subsection we discuss a submodel of our Markov process which has uniform marginals. The submodel can be derived on putting $\beta = 0$ in the Markov process given in Lemma 4. Althoug the model has a different form from the aforementioned models, it can be proved that the reversible distribution exists for our model. The proof is straightforward and therefore omitted.

Lemma 5. *Assume that* $\beta = 0$ *in the Markov process given in Lemma 4. Then a circular uniform,* $C^*(0)$*, is the reversible distribution of the process.*

Consider the two variables of the above process, (W_m, W_n) $(m > n)$. It is clear that the joint density for (W_m, W_n) is

$$f(w_m, w_n) = \frac{1}{4\pi^2}\frac{1-\varphi^{2(m-n)}}{|1-\varphi^{m-n}w_n\overline{w_m}|^2}, \quad w_m, w_n \in \partial D. \tag{11}$$

This density is equivalent to the one given by Kato (to appear, equation 4.1) which is obtained using Brownian motion. Note that, in this submodel, the conditional distribution of W_n given $W_m = w_m$ $(m > n)$ is also a wrapped Cauchy $C^*(\varphi^{m-n}w_m)$.

The autocorrelation coefficients for this submodel can be expressed in simple form as follows.

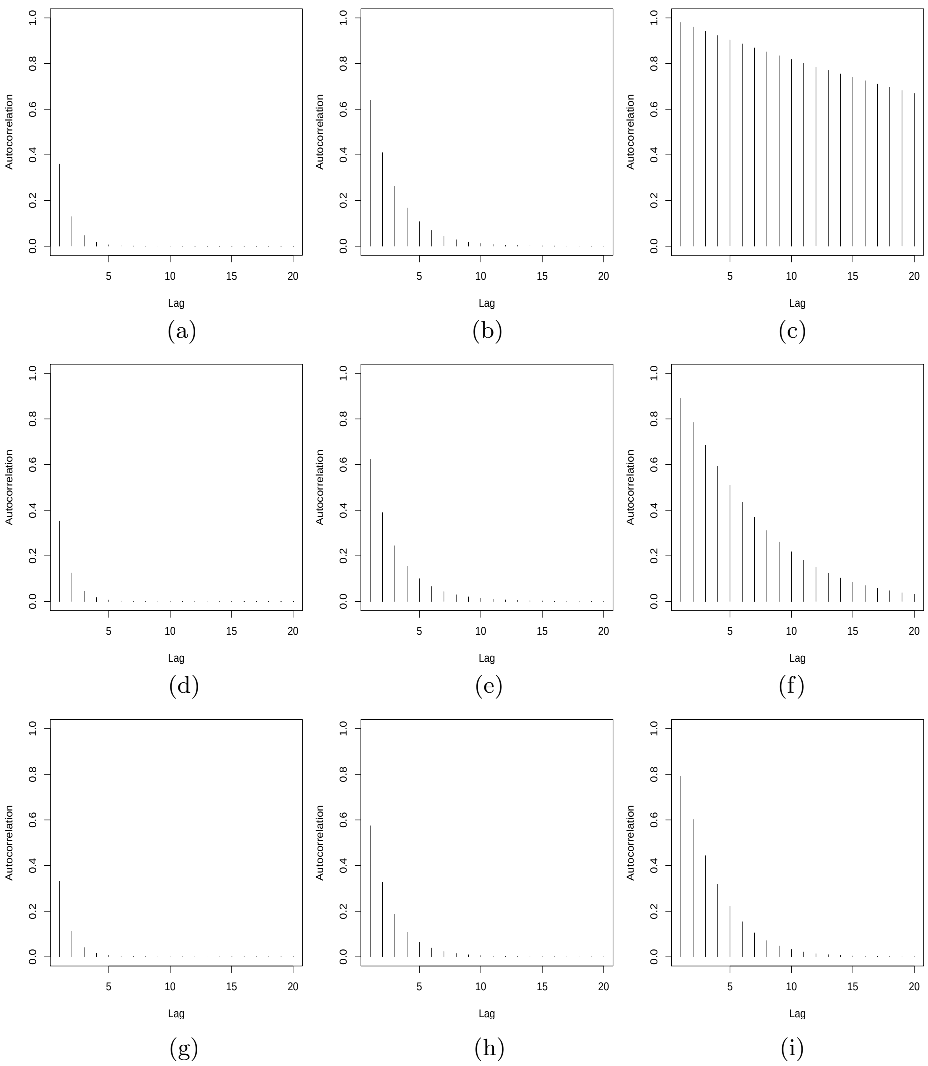

Figure 3: The Fisher and Lee (1983) autocorrelation coefficients for the stationary Markov process with parameters: (a) $(\beta, \varphi) = (0, 0.6)$, (b) $(\beta, \varphi) = (0, 0.8)$, (c) $(\beta, \varphi) = (0, 0.99)$, (d) $(\beta, \varphi) = (0.1, 0.6)$, (e) $(\beta, \varphi) = (0.1, 0.8)$, (f) $(\beta, \varphi) = (0.1, 0.99)$, (g) $(\beta, \varphi) = (0.2, 0.6)$, (h) $(\beta, \varphi) = (0.2, 0.8)$ and (i) $(\beta, \varphi) = (0.2, 0.99)$.

Corollary 1. *The Johnson and Wehrly (1977), Jupp and Mardia (1980), and Fisher and Lee (1983) correlation coefficients of* (W_m, W_n) $(m > n)$ *of the Markov process given in Lemma 5 are given by*

$$\rho_{JW} = \varphi^{m-n}, \quad \rho_{JM} = 2\varphi^{2(m-n)}, \quad \textit{and} \quad \rho_{FL} = \varphi^{2(m-n)}.$$

All autocorrelations have the following properties:

1. the autoccorelation is positive for any $m - n$,
2. the greater the value of the lag $m - n$, the smaller the autocorrelation between W_m and W_n,
3. as the lag $m - n$ tends to infinity, the autocorrelation tends to zero,
4. as φ increases, the autocorrelation between W_m and W_n decreases.

Remember that the Fisher and Lee (1983) autocorrelation of this stationary process for some selected values of parameters is given in Fig. 3(a),(b) and (c).

Furthermore, the following corollaries are obtained for the above model by using equation (4.3) and Section 2.1 of Kato (to appear), respectively.

Corollary 2. *Let* $\{W_{1n}\}_{n=0}^{\infty}$ *and* $\{W_{2n}\}_{n=0}^{\infty}$ *be independent Markov processes of the type given in Lemma 5. Then* $\{W_{1n}W_{2n}\}_{n=0}^{\infty}$ *is also a Markov process of the type given in Lemma 5 with the parameter* φ *replaced by* φ^2*.*

Corollary 3. *Let* $\{B_t\}_{t\geq 0}$ *be* $\mathbb{C}$*-valued Brownian motion starting at the origin. Suppose* τ_n *is the smallest time at which the Brownian particle hits a circle with radius* φ^{-n}*, i.e.,* $\tau_n = \inf\{t\,;\, |B_t| = \varphi^{-n}\}$ *where* $n = 0, 1, \ldots$*,* $0 < \varphi < 1$*. Then a sequence of random variables* $\{\varphi^n B_{\tau_n}\}_{n=0}^{\infty}$ *is a Markov process of the type given in Lemma 5.*

5 Comparison with existing Markov processes

Fisher and Lee (1994) proposed some stochastic processes for circular data. Among these models, a general class of the inverse autoregressive processes has some association with the general form of our Möbius Markov process given in Theorem 1. The inverse model denoted by IAR(p) in their paper is defined as follows. Let $\{\Theta_n\}$ be a sequence of $[-\pi, \pi)$-valued random variables. Assume that Θ_n given $(\Theta_{n-1}, \ldots, \Theta_{n-p}) = (\theta_{n-1}, \ldots, \theta_{n-p})$ has a von Mises distribution with concentration parameter κ and mean direction

$$\mu_t = \mu + g\{w_1 g^{-1}(\theta_{t-1} - \mu) + \cdots + w_p g^{-1}(\theta_{t-p} - \mu)\}, \tag{12}$$

where $\mu \in [-\pi, \pi)$ and $g(\cdot)$ is an odd monotone function mapping the real line onto the interval $(-\pi, \pi)$.

On putting $p = 1$, $0 < w_1 \leq 1$, and $g(x) = 2\arctan(x)$ in this model, one can obtain a Markov process given in Theorem 1 with $\beta = \{(1 - w_1)/(1 + w_1)\}\, e^{i\mu}$ and $\mathrm{Arg}(\varepsilon_n) \sim \mathrm{VM}(0, \kappa)$. However, the Möbius Markov process is not a submodel of this general class since our model assumes that the angular error has the wrapped Cauchy distribution, not the von Mises distribution. This distinction makes some differences in properties because of some desirable features of the wrapped Cauchy distribution and its relationship with the Möbius circle transformation as seen in Sections 2 and 3. For instance, the conditional distribution of W_{n+h} given $W_n = w_n$ has the wrapped Cauchy, and this property enables us to derive limiting distribution, the stationary distribution and autocorrelation coefficient of the model.

The wrapped Cauchy and von Mises are different in terms of shapes of the density in some situations. Both models are symmetric and unimodal distributions on the circle and these densities look similar when mean resultant lengths of these distributions are small. Therefore one can apply both IAR model and our process to a circular time series if the observations are dispersed. However, if the concentration parameters of both densities are large, then the wrapped Cauchy and von Mises have different 'tailweights', or the behaviour of the densities around the antimodes. When the mean resultant lengths of the densities are large, the wrapped Cauchy density has a heavier tail than the von Mises density has. This distinction suggests that the IAR model is more suitable if observations are highly concentrated and do not involve outliers. On the other hand, when robustness is desired, it would be expected that our model provides a better fit.

Our stationary process with uniform marginals given in Lemma 5 has relationship with the Markov model presented by Wehrly and Johnson (1980). They provided the model by applying a general class of bivariate circular distributions with specified marginals. A special case of their model is a Markov process which has uniform marginals and von Mises errors. The model has also been briefly considered by Breckling (1989, Example 6.1) as a submodel of the von Mises process. Their model and our submodel discussed in Lemma 5 are related in the sense that both models have uniform marginals. The difference is that we adopt the wrapped Cauchy error, whereas the existing one uses the von Mises as an error distribution. As discussed before, the tail behaviour of the wrapped Cauchy is different from that of the von Mises if the distributions are highly concentrated. Therefore one can select which model to use depending on how heavy contamination the dataset of interest includes. From the mathematical point of view, our model has some tractable properties as discussed in the previous paragraph. For example, the autocorrelation coefficients of this submodel can be expressed in simple form (see Corollary 1).

Our stationary process with uniform marginals also has some association with one of the models proposed by Fisher and Lee (1994). They presented a time series model by projecting two independent Gaussian processes. Both this projected model and our stationary process have the common advantage that the autocorrelation coefficient of the model can be expressed in relatively simple form. An advantage the projected model is that the parameters of the model can be readily estimated by applying EM algorithm. On the other hand, our model is attractive because it has clear dependence structure between W_n and W_{n+h} as one can see in the joint density (11).

6 Statistical inference for the process

Assume that $\{W_n\}_{n=0}^T$ is an observation from the Möbius Markov process with unknown β and φ. The quasi log-likelihood function for β and φ, $L_q(\beta, \varphi \,|\, w_0)$, is given by

$$\begin{aligned} \log L_q(\beta, \varphi \,|\, w_0) &= \log \{ f(w_n | w_{n-1}) f(w_{n-1} | w_{n-2}) \cdots f(w_1 | w_0) \} \\ &= C + T \log(1 - \varphi^2) - \sum_{n=1}^{T} \log \left| w_n - \frac{w_{n-1} + \beta}{\overline{\beta} w_{n-1} + 1} \varphi \right|^2 \end{aligned}$$

Transform the observations and parameters by taking $w_n = e^{i\theta_n}$ and $\beta = re^{i\mu}$, where $-\pi \leq \theta_n, \mu < \pi,\ 0 \leq r < 1$. Then the above function can be expressed as

$$\log L_q(r, \mu, \varphi \,|\, \theta_0) = C + T \log(1-\varphi^2) - \sum_{n=1}^{T} \log\left\{1 - 2\varphi\cos(\theta_n - \mu_n) + \varphi^2\right\}, \qquad (13)$$

where $\mu_n = \theta_{n-1} - 2\,\mathrm{Arg}\{1 + re^{i(\theta_{n-1}-\mu)}\}$ and C is a constant irrelevant to unknown parameters. Therefore the maximum likelihood estimation of the proposed process is essentially the same as that of the regression model of Kato *et al.* (2008).

It is clear from their context that, when r and μ are known, the estimates are obtained by the recursive algorithm of Kent and Tyler (1988). The method of moments estimator based on the first trigonometric moment can be obtained in closed form as follows

$$\hat{\varphi} = \frac{1}{T}\left|\sum_{n=1}^{T}\cos(\theta_n - \mu_n) + i\sum_{n=1}^{T}\sin(\theta_n - \mu_n)\right|.$$

As for the stationary process given in Lemma 4, it is easy to see that log-likelihood function is

$$\begin{aligned}\log L_s(r, \mu, \varphi) \quad &= \quad C + \log L_q(r, \mu, \varphi \,|\, \theta_0) + \log\left(1 - |\phi_\infty|^2\right) \\ &\qquad - \log\left\{1 - 2\,|\phi_\infty|\cos(\theta_0 - \mu) + |\phi_\infty|^2\right\}, \qquad (14)\end{aligned}$$

where $|\phi_\infty| = [\varphi - 1 + \surd\{(1-\varphi)^2 + 4\varphi r\}]/(2r)$ and $L_q(r, \mu, \varphi)$ is given by (13). If $r = 0$, the stationary process has uniform marginals, and therefore $\phi_\infty = 0$. In this case the maximum likelihood estimate of φ can be obtained, again, from Kent and Tyler (1988).

7 Application

We consider a time series of wind directions measured hourly at a weather station in Texas, U.S.A. The data are provided by NCAR/EOL under sponsorship of the National Science Foundation and accessible at `http://data.eol.ucar.edu/codiac/dss/id=85.034`. The original data contain hourly resolution surface meteorological data from the Texas Natural Resources Conservation Commission Air Quality Monitoring Network. Of all the data, we discuss a time series of 101 wind directions measured hourly at a weather station in Texas, which is denoted by `C1_1` in the dataset, from June 26 at 9 p.m. to July 1 at 1 a.m. in 2003.

Fig. 4(a) plots the time series of the wind directions. It seems from this figure that a robust model may be appropriate to fit the data since the dataset includes some outliers. The sample autocorrelation proposed by Fisher and Lee (1994, Equation (3.1)) is plotted in Fig. 4(b). We fit our stationary process given in Theorem 4 and two inverse models (12) of Fisher and Lee (1994) based on maximum likelihood. To maximise the likelihood functions, we adopt an optimisation method, which can be implemented using a command `nlminb` in `R`.

First, consider our stationary process. For the estimation of the parameters, we maximise the log-likelihood function (14). The estimated maximum log-likelihood and the maximum likelihood estimates of the parameters are given by $\log L = -41.5$, $\hat{r} = 0.205$, $\hat{\mu} = 0.323$ and $\hat{\varphi} = 0.861$, respectively. Hence, the estimated parameter of the limiting distribution is $\hat{\phi}_\infty = 0.649\exp(0.323i)$.

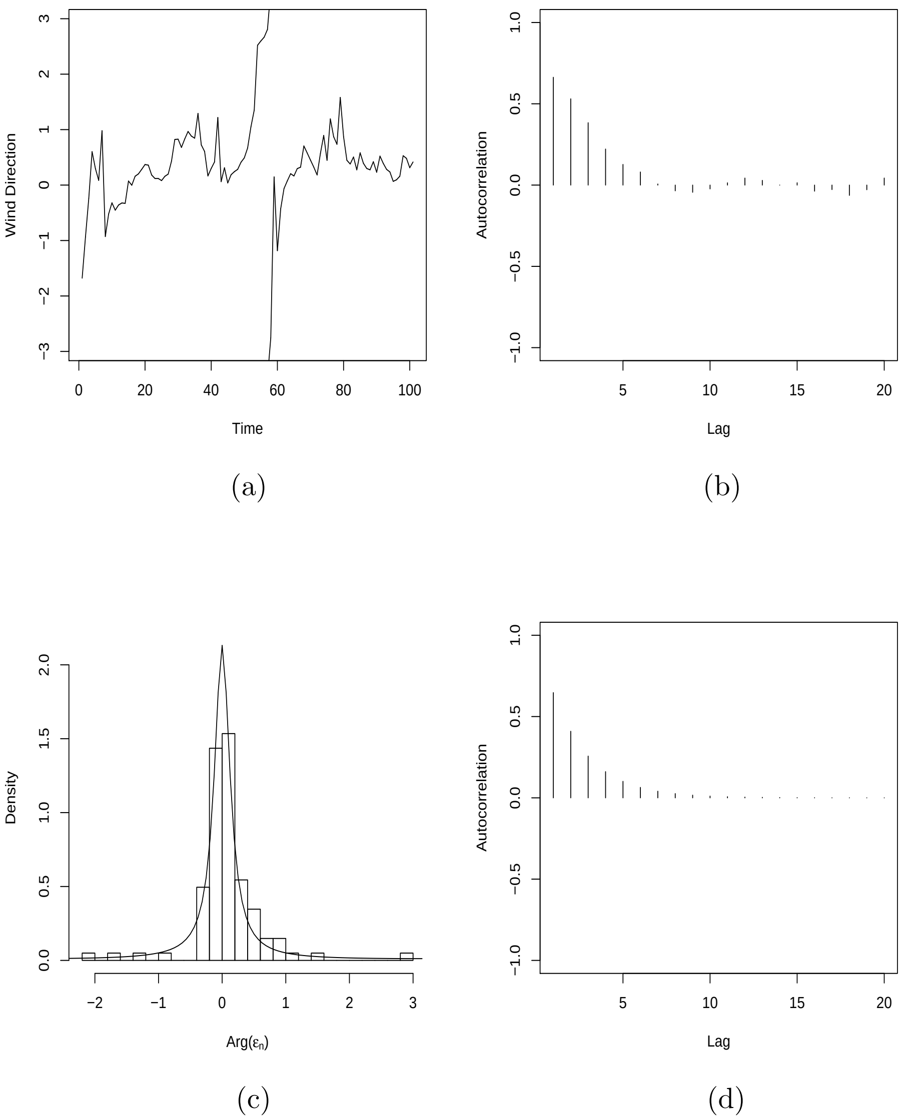

Figure 4: (a) the series of 101 wind directions hourly measured in Texas, U.S.A. from June 26 at 9 p.m. to July 1 at 1 a.m. in 2003, (b) the Fisher and Lee (1983) sample autocorrelation coefficients, (c) histogram of the estimated errors and the fitted density of the estimated error distribution of the proposed stationary process and (d) the Fisher and Lee (1983) theoretical autocorrelation coefficients for the fitted proposed process.

Next we discuss two inverse Markov, or IAR(1), models of Fisher and Lee (1994). As the first model, we define the link function g in (12) as $g(x) = 2\arctan(x)$. Note that this model is not a stationary process in general. The estimated maximum log-likelihood and estimated parameters are $\log L = -64.7$, $\hat{w}_1 = 1.00$, $\hat{\mu} = -1.68$ and $\hat{\kappa} = 5.31$. The second model supposes that g is the probit link, namely, $g(x) = 2\pi\{\Phi(x) - 0.5\}$, where $\Phi(x) = \int_{-\infty}^{x} \exp(-t^2/2)/\sqrt{(2\pi)}\, dt$. Then we obtain the maximum log-likelihood and estimated parameters as $\log L = -63.0$, $\hat{w}_1 = 1.14$, $\hat{\mu} = -2.59$ and $\hat{\kappa} = 5.48$.

Since the numbers of the parameters for the above three models are the same, model selection based on some information criteria such as AIC and BIC is essentially the same as the comparison of the maximum log-likelihood functions. Therefore, according to these criteria, we find that the proposed stationary process is best among three models. One reason our model fits better than the others could be the existence of some outliers seen around Time 55–60.

Fig. 4(c) shows a histogram of the estimated errors and the fitted wrapped Cauchy density. Here the estimated errors $\{\mathrm{Arg}(\hat{\varepsilon}_n)\}_{n=1}^{100}$ are $\mathrm{Arg}(\hat{\varepsilon}_n) = \theta_n - \theta_{n-1} + 2\,\mathrm{Arg}\{1 + \hat{r}e^{i(\theta_{n-1} - \hat{\mu})}\}$. From the histogram, it appears that our model provides a satisfactory fit to the dataset. It can also be confirmed from this frame that a probability distribution with a heavy tail such as the one we used here would be appropriate to model this time series.

The theoretical autocorrelation coefficient (10) for the fitted Markov model is displayed in Fig. 4(d). Comparing this frame with Fig. 4(b), our model provides a satisfactory result when the lag is less than 8. However if the lag is not less than 8, there is slight difference between these two autocorrelation coefficients.

In the example we consider a time series which involves some outliers. Since a robust model is desired for this dataset, we think that our model provides a better fit than the IAR models in terms of some information criteria. If we consider another time series which does not contain outliers, it is more likely that the IAR models, which adopt the von Mises error, are more suitable than the presented one.

8 Concluding remarks

Circle-valued processes can be used to analyse time series of circular data, which appear in various scientific fields such as meteorology and biology. The proposed Markov process has the virtues of being mathematically tractable as seen in properties in Section 3 such as easy interpretation of the parameters and clear limiting behaviour. Many of these properties can be derived because of some desirable features of the wrapped Cauchy distribution and its association with Möbius transformation. In practice our process can be used as a robust model even when observations include some outliers. The model is applicable to real data as we demonstrated in Section 7. Potential fields for future research include extension to an autoregressive process and construction of a hidden Markov process for circular data. It might be also interesting to investigate properties of a process which adopts a more flexible angular error distribution such as the ones provided by Jones and Pewsey (2005) and Pewsey *et al.* (2007).

Acknowledgements

I am grateful to Professors M.C. Jones and K. Shimizu for helpful comments on the paper. 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 for Young Scientists (B) (20740065).

Appendix A: Proof of Lemma 1

For $\beta = 0$, it is easy to see that $W_n \,|\, (W_0 = w_0) \sim C^*(\varphi^n w_0)$, and this converges to the circular uniform as n tends to infinity. For $\beta \neq 0$ and $\varphi > 0$, assume that $\mathrm{Arg}(\beta) = 0$. Using eigenvalue decomposition, we have

$$\begin{pmatrix} \varphi & \beta\varphi \\ \beta & 1 \end{pmatrix} = Q \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} Q^{-1},$$

where

$$\lambda_1 = \tfrac{1}{2}(1 + \varphi - A), \quad \lambda_2 = \tfrac{1}{2}(1 + \varphi + A),$$

$$Q = \begin{pmatrix} \varphi - 1 - A & 2\beta \\ \varphi - 1 + A & 2\beta \end{pmatrix}, \quad A = \sqrt{\{(1-\varphi)^2 + 4\beta^2\varphi\}}.$$

Then it follows that

$$\begin{aligned} &\begin{pmatrix} \varphi & \beta\varphi \\ \beta & 1 \end{pmatrix}^n \circ w_0 \\ &= \frac{\{\lambda_1^n(1-\varphi+A) + \lambda_2^n(\varphi-1+A)\}w_0 + 2(\lambda_2^n - \lambda_1^n)\beta\varphi}{2(\lambda_2^n - \lambda_1^n)\beta w_0 + \lambda_2^n(1-\varphi+A) + \lambda_1^n(\varphi-1+A)} \\ &= \frac{\{(\lambda_1/\lambda_2)^n(1-\varphi+A) + \varphi - 1 + A\}w_0 + 2\{1-(\lambda_1/\lambda_2)^n\}\beta\varphi}{2\{1-(\lambda_1/\lambda_2)^n\}\beta w_0 + 1 - \varphi + A + (\lambda_1/\lambda_2)^n(\varphi-1+A)}. \end{aligned} \tag{15}$$

Since $0 < \lambda_1 < 1$ and $\lambda_2 > 1$, it is clear that $(\lambda_1/\lambda_2)^n \to 0$ as $n \to \infty$. Thus

$$\begin{pmatrix} \varphi & \beta\varphi \\ \beta & 1 \end{pmatrix}^n \circ w_0 \longrightarrow \frac{(\varphi - 1 + A)w_0 + 2\beta\varphi}{2\beta w_0 + 1 - \varphi + A} \quad \text{as} \quad n \longrightarrow \infty.$$

After some algebra, one obtains

$$\frac{(\varphi - 1 + A)w_0 + 2\beta\varphi}{2\beta w_0 + 1 - \varphi + A} = \phi_\infty.$$

It follows from Theorem 2 and Lebesgue's dominated convergence theorem that $W_n \,|\, (W_0 = w_0) \xrightarrow{\mathrm{d}} C^*(\phi_\infty)$ as $n \to \infty$. In the case of $\mathrm{Arg}(\beta) \neq 0$, one can easily prove that the conditional of W_n given $W_0 = w_0$ also converges to $C^*(\phi_\infty)$ by combining the above result and equation (7). If $\varphi = 0$, then $W_n \,|\, (W_0 = w_0) \sim C^*(0)$ for any $n \geq 1$, and Lemma 1 holds.

Appendix B: Proof of Lemma 2

First, we show $\pi P = \pi$, namely,

$$\int_{\partial D} P(w, A)\pi(dw) = \pi(A),$$

for any measurable set $A\,(\subset \partial D)$. If $\varphi = 0$, this is obvious. When $\varphi > 0$, the following hold for πP:

$$\begin{aligned}(\pi P)(A) &= \int_{\partial D} P(w, A)\pi(dw) \\ &= \int_{\partial D} \left(\int_A \frac{1}{2\pi} \frac{1 - |\phi_1(w)|^2}{|z - \phi_1(w)|^2} \mu(dz) \right) \frac{1}{2\pi} \frac{1 - |\phi_\infty|^2}{|w - \phi_\infty|^2} \mu(dw) \\ &= \int_A \left(\int_{\partial D} \frac{1}{4\pi^2} \frac{1 - |\phi_1(w)|^2}{|z - \phi_1(w)|^2} \frac{1 - |\phi_\infty|^2}{|w - \phi_\infty|^2} \mu(dw) \right) \mu(dz). \qquad (16)\end{aligned}$$

By transforming

$$W' = \begin{pmatrix} 1 & -\phi_\infty \\ -\overline{\phi_\infty} & 1 \end{pmatrix} \circ W,$$

and using property (6), (16) can be expressed as

$$(\pi P)(A) = \int_A \left(\int_{\partial D} \frac{1}{4\pi^2} \frac{1 - |\phi_1'(w')|^2}{|z - \phi_1'(w')|^2} \mu(dw') \right) \mu(dz),$$

where

$$\phi_1'(w') = \begin{pmatrix} \varphi & \beta\varphi \\ \overline{\beta} & 1 \end{pmatrix} \circ \left\{ \begin{pmatrix} 1 & \phi_\infty \\ \overline{\phi_\infty} & 1 \end{pmatrix} \circ w' \right\}.$$

From equation (2) and the following relationship

$$\begin{pmatrix} \varphi & \beta\varphi \\ \overline{\beta} & 1 \end{pmatrix} \circ \phi_\infty = \phi_\infty,$$

it follows that

$$\phi_1'(w') = \left\{ \begin{pmatrix} 1 & \phi_\infty/\varphi \\ \overline{\phi_\infty}/\varphi & 1 \end{pmatrix} \circ w' \right\} \varphi.$$

Transform

$$W'' = \begin{pmatrix} 1 & \phi_\infty/\varphi \\ \overline{\phi_\infty}/\varphi & 1 \end{pmatrix} \circ W',$$

and note that

$$h(w'') = \frac{||w''|^2 - 1/\varphi^2|}{|w'' - z/\varphi|^2}, \quad w'' \in \overline{D},$$

is continuous on the closed unit disc and analytic on the open unit disc. Then, by Theorem 1 of Kato *et al.* (2008), $(\pi P)(A)$ reduces to

$$\begin{aligned}(\pi P)(A) &= \int_A \left(\int_{\partial D} \frac{1}{4\pi^2} \frac{||w''|^2 - 1/\varphi^2|}{|w'' - z/\varphi|^2} \frac{|1 - |\phi_\infty|^2/\varphi^2|}{|w'' - \phi_\infty/\varphi|^2} \mu(dw'') \right) \mu(dz) \\ &= \int_A \frac{1}{2\pi} \frac{||\phi_\infty|^2 - 1|/\varphi^2}{|\phi_\infty - z|^2/\varphi^2} \mu(dz) \\ &= \int_A \frac{1}{2\pi} \frac{1 - |\phi_\infty|^2}{|z - \phi_\infty|^2} \mu(dz) \\ &= \pi(A).\end{aligned}$$

Thus we obtain $\pi P = \pi$. Next we prove that the Möbius Markov process is π-irreducible. Since each support of the densities for π and $W_1 \,|\, (W_0 = w_0)$ is ∂D, it follows that if a measurable set A satisfies $\pi(A) > 0$, which is a condition equivalent to $\mu(A) > 0$, then $P(w_0, A) > 0$. Hence the Möbius Markov process is π-irreducible. Therefore, by Theorem 1 of Tierney (1994), π is the unique invariant distribution of the process.

Appendix C: Proof of Lemma 4

On putting $\tilde{\mathbf{V}}_j = (\mathrm{Re}(\tilde{W}_j), \mathrm{Im}(\tilde{W}_j))' = Q'\mathbf{V}_j$ $(j = m, n)$, it follows that

$$E(\mathbf{V}_m) = QE(\tilde{\mathbf{V}}_m) \quad \text{and} \quad E(\mathbf{V}_n) = QE(\tilde{\mathbf{V}}_n).$$

Since Q' is a rotation matrix which controls the mean direction of $\tilde{W}_\ell$ to be 0, we have $\tilde{W}_\ell \sim C^*(|\phi_\infty|)$ for any ℓ. Remember that a known result of the wrapped Cauchy distribution (see McCullagh (1996)):

$$Z \sim C^*(\phi) \quad \Longrightarrow \quad E(Z^p) = \phi^p. \tag{17}$$

On putting $p = 1$ in the above equation, we immediately obtain $E(\tilde{\mathbf{V}}_j) = (|\phi_\infty|, 0)'$. Therefore it follows that

$$E(\mathbf{V}_m) = E(\mathbf{V}_n) = Q\begin{pmatrix} |\phi_\infty| \\ 0 \end{pmatrix}.$$

Using equation (17) with $p = 0$ and 2, it is easy to show

$$E(\mathbf{V}_m\mathbf{V}_m') = E(\mathbf{V}_n\mathbf{V}_n') = Q\begin{pmatrix} \frac{1}{2}(1 + |\phi_\infty|^2) & 0 \\ 0 & \frac{1}{2}(1 - |\phi_\infty|^2) \end{pmatrix} Q'.$$

Next we consider $E(\mathbf{V}_m\mathbf{V}_n')$. It can be expressed as

$$E(\mathbf{V}_m\mathbf{V}_n') = QE(\tilde{\mathbf{V}}_m\tilde{\mathbf{V}}_n')Q'.$$

From equation (7), one obtains that the conditional distribution of $\tilde{W}_m$ given $\tilde{W}_n$ has the wrapped Cauchy $C^*(\tilde{\phi}_{m-n}(\tilde{w}_n))$, where

$$\tilde{\phi}_{m-n}(\tilde{w}_n) = \begin{pmatrix} \varphi & |\beta|\varphi \\ |\beta| & 1 \end{pmatrix}^{m-n} \circ \tilde{w}_n = \begin{pmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{pmatrix} \circ \tilde{w}_n$$

From this, $E\{\mathrm{Re}(\tilde{W}_m)\mathrm{Re}(\tilde{W}_n)\}$ can be expressed as

$$\begin{aligned} E\left\{\mathrm{Re}(\tilde{W}_m)\mathrm{Re}(\tilde{W}_n)\right\} &= \int_{\partial D} \mathrm{Re}(\tilde{\phi}_{m-n}(\tilde{w}_n))\mathrm{Re}(\tilde{w}_n)\frac{1}{2\pi}\frac{1 - |\phi_\infty|^2}{|\tilde{w}_n - |\phi_\infty||^2}d\mu(\tilde{w}_n) \\ &= \int_{-\pi}^{\pi} \mathrm{Re}\left(\frac{a_{00}e^{i\theta} + a_{01}}{a_{10}e^{i\theta} + a_{11}}\right)\cos\theta\,\frac{1}{2\pi}\frac{1 - |\phi_\infty|^2}{1 + |\phi_\infty|^2 - 2|\phi_\infty|\cos\theta}\,d\theta \\ &= \frac{1}{2\pi|a_{11}|^2}\int_{-\pi}^{\pi}\cos\theta\,\frac{(a_{00}a_{10} + a_{01}a_{11}) + (a_{00}a_{11} + a_{01}a_{10})\cos\theta}{1 + 2|a_{10}/a_{11}|\cos\theta + |a_{10}/a_{11}|^2} \\ &\quad \times\frac{1 - |\phi_\infty|^2}{1 + |\phi_\infty|^2 - 2|\phi_\infty|\cos\theta}\,d\theta. \end{aligned}$$

It is known in the theory of Fourier series expansion (see, for example, Mardia and Jupp (1999, p.51)) that

$$\frac{1-\rho^2}{1+\rho^2-2\rho\cos\theta}=1+2\sum_{j=1}^{\infty}\rho^j\cos(j\theta),\quad -1<\rho<1.$$

Applying this property, it holds for $0\le\rho_1,\rho_2<1$ that

$$\begin{aligned}
&\int_{-\pi}^{\pi}\cos\theta\,\frac{1-\rho_1^2}{1+\rho_1^2-2\rho_1\cos\theta}\cdot\frac{1-\rho_2^2}{1+\rho_2^2+2\rho_2\cos\theta}\,d\theta\\
&=\int_{-\pi}^{\pi}\cos\theta\left\{1+2\sum_{j=1}^{\infty}\rho_1^j\cos(j\theta)\right\}\left\{1+2\sum_{k=1}^{\infty}(-\rho_2)^k\cos(k\theta)\right\}d\theta\\
&=\int_{-\pi}^{\pi}\cos\theta d\theta+2\sum_{j=1}^{\infty}\rho_1^j\int_{-\pi}^{\pi}\cos\theta\cos(j\theta)d\theta+2\sum_{k=1}^{\infty}(-\rho_2)^k\int_{-\pi}^{\pi}\cos\theta\cos(k\theta)d\theta\\
&\quad+4\sum_{j,k=1}^{\infty}\rho_1^j(-\rho_2)^k\int_{-\pi}^{\pi}\cos\theta\cos(j\theta)\cos(k\theta)d\theta\\
&=2\rho_1\pi+2(-\rho_2)\pi+4\sum_{j=2}^{\infty}\rho_1^j(-\rho_2)^{j-1}\cdot\frac{\pi}{2}+4\sum_{j=1}^{\infty}\rho_1^j(-\rho_2)^{j+1}\cdot\frac{\pi}{2}\\
&=2\pi\left\{\rho_1-\rho_2-\frac{1}{\rho_2}\left(\frac{-\rho_1\rho_2}{1+\rho_1\rho_2}+\rho_1\rho_2\right)-\rho_2\frac{-\rho_1\rho_2}{1+\rho_1\rho_2}\right\}\\
&=\frac{2\pi(\rho_1-\rho_2)}{1+\rho_1\rho_2}.
\end{aligned}\tag{18}$$

Similarly,

$$\begin{aligned}
&\int_{-\pi}^{\pi}\cos^2\theta\,\frac{1-\rho_1^2}{1+\rho_1^2-2\rho_1\cos\theta}\cdot\frac{1-\rho_2^2}{1+\rho_2^2+2\rho_2\cos\theta}\,d\theta\\
&=\int_{-\pi}^{\pi}\frac{\cos2\theta+1}{2}\,\frac{1-\rho_1^2}{1+\rho_1^2-2\rho_1\cos\theta}\cdot\frac{1-\rho_2^2}{1+\rho_2^2+2\rho_2\cos\theta}\,d\theta\\
&=\frac{1-\rho_1^2\rho_2^2+(\rho_1-\rho_2)^2}{1+\rho_1\rho_2}.
\end{aligned}\tag{19}$$

From equations (18) and (19), it follows that

$$\begin{aligned}
E\left\{\mathrm{Re}(\tilde{W}_m)\mathrm{Re}(\tilde{W}_n)\right\}&=\frac{(a_{00}a_{10}+a_{01}a_{11})(|\phi_\infty|-|a_{10}/a_{11}|)}{(|a_{11}|^2-|a_{10}|^2)(1+|\phi_\infty a_{10}/a_{11}|)}\\
&\quad+\frac{(a_{00}a_{11}+a_{01}a_{10})\{1-|\phi_\infty a_{10}/a_{11}|^2+(|\phi_\infty|-|a_{10}/a_{11}|)^2\}}{2(|a_{11}|^2-|a_{10}|^2)(1+|\phi_\infty a_{10}/a_{11}|)}.
\end{aligned}$$

The other elements of $E(\tilde{\mathbf{V}}_j\tilde{\mathbf{V}}_k')$ are obtained in a similar manner as

$$\begin{aligned}
E\{\mathrm{Im}(\tilde{W}_m)\mathrm{Im}(\tilde{W}_n)\}&=\frac{(a_{00}a_{11}-a_{01}a_{10})(1-|\phi_\infty|^2-|a_{10}/a_{11}|^2+|\phi_\infty a_{10}/a_{11}|^2)}{2(|a_{11}|^2-|a_{10}|^2)(1+|\phi_\infty a_{10}/a_{11}|)},\\
E\{\mathrm{Re}(\tilde{W}_m)\mathrm{Im}(\tilde{W}_n)\}&=E\{\mathrm{Im}(\tilde{W}_m)\mathrm{Re}(\tilde{W}_n)\}=0.
\end{aligned}$$

Therefore we have

$$E(\mathbf{V}_j\mathbf{V}_k')=QE(\tilde{\mathbf{V}}_j\tilde{\mathbf{V}}_k')Q'=Q\begin{pmatrix}a&0\\0&b\end{pmatrix}Q'.$$

References

Breckling, J. (1989) *The Analysis of Directional Time Series: Applications to Wind Speed and Direction*. London: Springer.

Cameron, M. A. (1983) The comparison of time series recorders. *Technometrics*, **25**, 9–22.

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

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

Fisher, N. I. and Lee, A. J. (1994) Time series analysis of directional data. *J. R. Statist. Soc.* B, **56**, 327–339.

Holzmann, H., Munk A., Suster, M. and Zucchini, W. (2006) Hidden Markov models for circular and linear-circular time series. *Environ. Ecol. Stat.*, **13**, 325–347.

Jammalamadaka, S. R. and SenGupta, A. (2001) *Topics in Circular Statistics.* Singapore: World Scientific Publ.

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

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

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

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

Kato, S. A distribution for a pair of unit vectors generated by Brownian motion. *Bernoulli*, to appear. (Accepted manuscript is available at `http://isi.cbs.nl/bernoulli/future.htm`)

Kato, S., Shimizu, K. and Shieh G. S. (2008) A circular–circular regression model. *Statist. Sinica*, **18**, 633–645.

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

Mardia, K. V. and Jupp, P. E. (1999) *Directional Statistics.* Chichester: Wiley.

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

Meyn, S. P. and Tweedie, R. L. (1993) *Markov Chains and Stochastic Stability.* London: Springer-Verlag.

Pewsey, A., Lewis, T. and Jones, M. C. (2007) The wrapped t family of circular distributions. *Aust. N. Z. J. Statist.*, **49**, 79–91.

Rudin, W. (1987) *Real and Complex Analysis*, 3rd ed. New York: McGraw-Hill.

Tierney, L. (1994) Markov chains for exploring posterior distributions. *Ann. Statist.*, **22**, 1701–1762.

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