J. R. Statist. Soc. B (2010)
72, *Part* 5, *pp.* 655–672

A Markov process for circular data

Shogo Kato
Institute of Statistical Mathematics, Tokyo, Japan

[Received January 2009. Revised April 2010]

Summary. 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, in particular on the stationary process. The proposed model is compared with some existing Markov processes for circular data. Some statistical issues of the model, such as statistical inference, model selection and diagnostic checks, are considered. Finally, an application of the model to wind direction data is provided.

Keywords: Auto-regressive process; Circular auto-correlation; Circular time series; Möbius transformation; Wrapped Cauchy distribution

1. Introduction

Data which can be expressed as sets of observations on the circle arise in various areas of applications such as biology, meteorology and geology. On occasion, 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 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 auto-regressive 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 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 on the basis of 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

Address for correspondence: Shogo Kato, Institute of Statistical Mathematics, 10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan.
E-mail: skato@ism.ac.jp

 1369–7412/10/72655

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 time, say, n approach certain values as n increases. Or our model can also be used to fit stationary circular time series data. The proposed model can be applied to various kinds of time series including those which are uniformly dispersed on the circle or clustered in an arc. We present an application of our model to a time series of wind directions to illustrate an advantage of the model.

The subsequent sections are organized 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, discuss its interpretation 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 most attention to the stationary case of our model. A comparison with some existing Markov models is provided in Section 5. Statistical inference for the process, including parameter estimation, Fisher information and a hypothesis test, is discussed in Section 6. Then, in Section 7, we consider some methods for model selection and diagnostic checks. In Section 8 the proposed stationary model is fitted to a time series of wind directions to illustrate an advantage of our model. Proofs of lemmas 1, 2 and 4 can be found in earlier work (Kato, 2009a) or obtained from the author.

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

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

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

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

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. The transformation appears as a regression curve of the model of Kato *et al.* (2008). As discussed in section 2.1 of Kato *et al.* (2008), β can be interpreted as a parameter that attracts the points on the circle towards $\beta/|\beta|$, with the concentration of the points around $\beta/|\beta|$ increasing as $|\beta|$ increases. See also Jones (2004) for more details about this transformation. Following the convention in Kato *et al.* (2008), we call this mapping the Möbius circle transformation.

For convenience, write transformation (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:

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

2.2. Wrapped Cauchy distribution

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

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

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 $\arg(\phi)$ or $\phi/|\phi|$ is the mean direction for $\phi\neq 0$ and $|\phi|$ the mean resultant length of Z for $\phi\in\bar{D}$, where $\arg(z)$ denotes the complex argument of z taking values between $[-\pi,\pi)$ and $\bar{D}=D\cup\partial D$. As discussed in McCullagh (1996), $f(z;\phi)=f(z;1/\bar{\phi})$. 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 $|\phi|\to 1$, 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 equation (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), pages 51–52, and Jammalamadaka and SenGupta (2001), section 2.2.7, for book treatments of the wrapped Cauchy distribution.) The following properties hold for the wrapped Cauchy distribution:

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

$$Z_1\sim C^*(\phi_1), \quad Z_2\sim C^*(\phi_2), \quad |\phi_1|,|\phi_2|\leqslant 1, \quad Z_1\perp Z_2\Rightarrow Z_1Z_2\sim C^*(\phi_1\phi_2), \tag{5}$$

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

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

$$W_n=\frac{W_{n-1}+\beta}{\bar{\beta}W_{n-1}+1}\varepsilon_n, \qquad n=1,2,\ldots, \tag{7}$$

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.

The general class of Markov processes presented in theorem 1 does not assume that the error variables $\varepsilon_1, \varepsilon_2, \ldots$ are independent and identically distributed. In this paper we investigate a subclass of this general class. From now on, assume that $\varepsilon_1, \varepsilon_2, \ldots$ in theorem 1 are independent and identically distributed random variables which are independent of W_0 and are distributed as the wrapped Cauchy distribution $C^*(\varphi)$, $0 \leqslant \varphi < 1$. We call this subclass the Möbius Markov process.

3.2. Analogies to existing Markov processes

As will be shown later in the paper, there are some advantages to express the Möbius Markov process in terms of complex numbers in investigating some properties of the model. However, to discuss the relationship between the proposed model and existing Markov processes, it seems more helpful to represent our model in terms of angles between $-\pi$ and π. With this representation, we investigate how our model can be interpreted through the comparison with a circular Markov model of Fisher and Lee (1994) and the auto-regressive AR(1) process on the real line.

Let $\{W_n\}_{n=0}^{\infty}$ be the Möbius Markov process. Assume that $\Theta_n = \arg(W_n)$, $\delta_n = \arg(\varepsilon_n)$ $(n = 0, 1, \ldots)$ and, for simplicity of discussion, $\arg(\beta) = 0$. Then, after some algebra, equation (7) can be expressed as

$$\Theta_n = 2\tan^{-1}\{\lambda_c \tan(\Theta_{n-1}/2)\} \oplus \delta_n, \tag{8}$$

where $\lambda_c = (1-\beta)/(1+\beta)$ and '$\oplus$' indicates addition modulo 2π on the circle. The plausibility of the regression curve of expression (7) or (8) has been discussed by Downs and Mardia (2002) and Kato *et al.* (2008) in the context of circular–circular regression. In particular, regression curves displayed in Fig. 1 of Downs and Mardia (2002) and Fig. 1 of Kato *et al.* (2008) are helpful to see the visual relationship of the circular mean of Θ_n to θ_{n-1} for some selected values of λ_c or β.

Given the angular expression (8), it is easy to find the relationship between our model and a Markov model defined by equation (2.3) of Fisher and Lee (1994). They presented two classes of stochastic processes derived by using link functions. One of these models, which we refer to as an inverse auto-regressive IAR(1) model, is defined as follows. Let $\{\Theta_n^*\}_{n=0}^{\infty}$ be a sequence of $[-\pi, \pi)$-valued random variables. Then a special case of the IAR(1) model is of the form

$$\Theta_n^* = U^{-1}\{\lambda_c^* U(\Theta_{n-1}^*)\} \oplus \delta_n^*, \tag{9}$$

where $U(\cdot)$ is an odd monotone function mapping $(-\pi, \pi)$ onto the real line, and δ_n^* $(n = 1, 2, \ldots)$ is a $[-\pi, \pi)$-valued random sample from the von Mises distribution with mean direction 0 and concentration parameter κ. On putting $0 < \lambda_c^* \leqslant 1$ and $U(x) = \tan(x/2)$, we find that this process essentially has the same regression curve as model (8). The difference between two models is the error distribution of the model; our model assumes that the wrapped Cauchy distribution has an angular error, whereas the IAR(1) model adopts the von Mises distribution. As in the later discussion, this distinction makes considerable differences in properties.

The Möbius Markov process can be interpreted more clearly through comparison with the AR(1) model on the real line. It is well known that the AR(1) model on the real line takes the form

$$X_n = \lambda_l X_{n-1} + \gamma_n. \tag{10}$$

Although the proposed model and the AR(1) process are defined in different manifolds, two parameters, λ_c in equation (8) and λ_l in equation (10), can be interpreted in quite a similar manner. Both parameters represent the level of shrinkage of $\mathbb{R}$-valued variables, $\tan(\Theta_{n-1}/2)$ or X_{n-1}. Also, these two parameters influence stationarity of the processes. For the AR(1) process, it is usual to assume that $0<\lambda_l<1$ to ensure stationarity. As will be discussed in lemma 2, an analogous condition for our circular model, $0<\lambda_c<1$ or $0<\beta<1$, also guarantees stationarity. In addition, the case $-1<\lambda_c<0$, corresponding to $-1<\lambda_l<0$, can be reduced to the previous setting, i.e. $0<\lambda_c, \lambda_l<1$, by reflecting successive observations on the circle or line respectively.

However, there is a difference in interpretation between λ_c and λ_l. For example, model (8) with $\lambda_c>1$ corresponds to model (8) with $0<\lambda_c<1$ by rotating the angles $\{\Theta_n\}$ by π, and therefore this model is stationary on the circle. An analogous setting for the AR(1) process is given by $\lambda_l>1$, but this case is non-stationary on the real line.

So far, we have restricted the parameter β to be $\arg(\beta)=0$. Allowing β to be a complex number enables us to rotate successive observations on the circle by $\arg(\beta)$. This extension corresponds to introducing a mean parameter for the regression relationship on the real line.

3.3. Some properties

We discuss some properties of the proposed process. Throughout this section, we assume that $\{W_n\}_{n=0}^{\infty}$ is the Möbius Markov process expressed in a form of complex numbers as in Section 2.1.

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:

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

where $t\geqslant 0$ and $n\geqslant 1$. 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.

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

where

$$\phi_n(w_0)=\begin{pmatrix} \varphi & \beta\varphi \\ \bar{\beta} & 1 \end{pmatrix}^n \circ w_0, \qquad n\geqslant 0.$$

Thus, the conditional of W_n given $W_0=w_0$ is the wrapped Cauchy distribution for any n.

By theorem 2 of this paper and theorem 1 of Kato *et al.* (2008), the pth trigonometric moment of $W_n|(W_0=w_0)$ is given by

$$E(W_n^p|W_0=w_0)=\begin{cases} \phi_n^p(w_0), & p\geqslant 0, \\ \overline{\phi_n^{-p}(w_0)}, & p<0, \end{cases} \quad \text{for } p\in\mathbb{Z}. \tag{11}$$

In particular, $\arg\{\phi_n(w_0)\}$ 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. The proof is given in Kato (2009a), appendix A.

Lemma 1.

$$W_n|(W_0=w_0)\xrightarrow{\mathrm{d}} C^*(\phi_\infty) \qquad \text{as } n\to\infty,$$

where

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

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

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

where A ($\subset \partial D$) is a measurable set and, as in Section 2.2, μ denotes arc length on the circle.

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

See Kato (2009a), appendix B, for the proof of lemma 2. It is remarked that, unlike the stationary Gaussian AR(1) process on the real line, π is not a reversible distribution except for the cases $\beta = 0$ or $\varphi = 0$. This non-reversibility can be proved by showing that

$$f_\pi(w_t)\, p_{t+n|t}(w_{t+n}|w_t) \neq f_\pi(w_{t+n})\, p_{t|t+n}(w_t|w_{t+n})$$

for $\beta \neq 0$ and $\varphi \neq 0$, where f_π is the stationary wrapped Cauchy density $C^*(\phi_\infty)$, and $p_{v|u}$ is the transitional density from W_u to W_v which can be obtained from theorem 2. If $\beta = 0$ or $\varphi = 0$, then the stationary Möbius Markov process has uniform marginals, i.e. $f_\pi = 1/2\pi$, and it immediately follows that the time reversibility holds.

A Markov process is said to be *ergodic* if it is positive Harris recurrent and aperiodic. (See, for example, Meyn and Tweedie (1993), pages 116, 200 and 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 $n \to \infty$. For any $\beta \in D\backslash\{0\}$ and $n \geqslant 1$,

$$\mathrm{Re}\left\{\bar\beta \frac{E(W_n|W_0 = w_0)}{|E(W_n|W_0 = w_0)|}\right\} \geqslant \mathrm{Re}\left\{\bar\beta \frac{E(W_{n-1}|W_0 = w_0)}{|E(W_{n-1}|W_0 = w_0)|}\right\}.$$

Or, equivalently,

$$\cos[\arg\{\bar\beta\, E(W_n|W_0 = w_0)\}] \geqslant \cos[\arg\{\bar\beta\, E(W_{n-1}|W_0 = w_0)\}].$$

This inequality can be proved by using equation (11) and noting that

$$\begin{aligned} \cos[\arg\{\bar\beta\, \phi_n(w_0)\}] &= \cos[\arg\{|\beta|^2 + \bar\beta\, \phi_{n-1}(w_0)\} - \arg\{1 + \bar\beta\, \phi_{n-1}(w_0)\}] \\ &\geqslant \cos[\arg\{\bar\beta\, \phi_{n-1}(w_0)\}], \end{aligned}$$

which can be seen geometrically.

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

$$g(\lambda)=\begin{pmatrix} \xi_1 w_0-2\beta\varphi & \xi_2 w_0+2\beta\varphi \\ \xi_2-2\bar{\beta}w_0 & \xi_1+2\bar{\beta}w_0 \end{pmatrix}\circ\lambda, \qquad 0\leqslant\lambda\leqslant 1, \tag{12}$$

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. From equation (15) of lemma 1 given in Kato (2009a) it is easy to see that equation (12) holds. 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 centre and radius are given by

$$-\mathrm{i}\frac{(\xi_1+\xi_2)(1-\varphi)(w_0+\beta)}{2\,\mathrm{Im}(\xi_2\beta\bar{w}_0-\xi_1\bar{\beta}w_0)}$$

and

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

respectively. In equation (12), $g(\lambda)$ coincides with $\phi_n(w_0)$ when

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

Thus, the rate of convergence is

$$r(|\beta|,\varphi)\equiv\left|\frac{\lambda(n+1)}{\lambda(n)}\right|=\frac{1+\varphi-\surd\{(1-\varphi)^2+4\varphi|\beta|^2\}}{1+\varphi+\surd\{(1-\varphi)^2+4\varphi|\beta|^2\}}. \tag{13}$$

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 $r\to\varphi$ ($|\beta|\to 0$). Also, r is monotonically increasing as a function of φ. Here we obtain $r\to(1-|\beta|)/(1+|\beta|)$ ($\varphi\to 1$) and $r\to 0$ ($\varphi\to 0$). Thus the higher the value of φ is, the slower the convergence of the process.

3.4. Simulation

In this section 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

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

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

Fig. 1 displays simulations of the Möbius Markov process for specified values of φ and β. It explicitly shows that, as $n\to\infty$, the mean direction of the conditional of $\arg(W_n)$ given $W_0=w_0$ tends to $\arg(\beta)$, and this is mathematically validated by lemma 1. By comparing Figs 1(a) and 1(b), it seems that $|\beta|$ influences the speed of convergence r and the parameter of the limiting distribution $|\phi_\infty|$. Actually, as stated in Section 3.3, $|\phi_\infty|$ and the negative of the rate of convergence (13), $-r$, are monotonically increasing with respect to $|\beta|$. From Figs 1(a) and 1(c), 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 (13) which is monotonically decreasing. Finally, comparing Figs 1(a) and 1(d), we find that the parameter $\arg(\beta)$ controls the mean direction of the limiting distribution, i.e. $\arg(\phi_\infty)$.

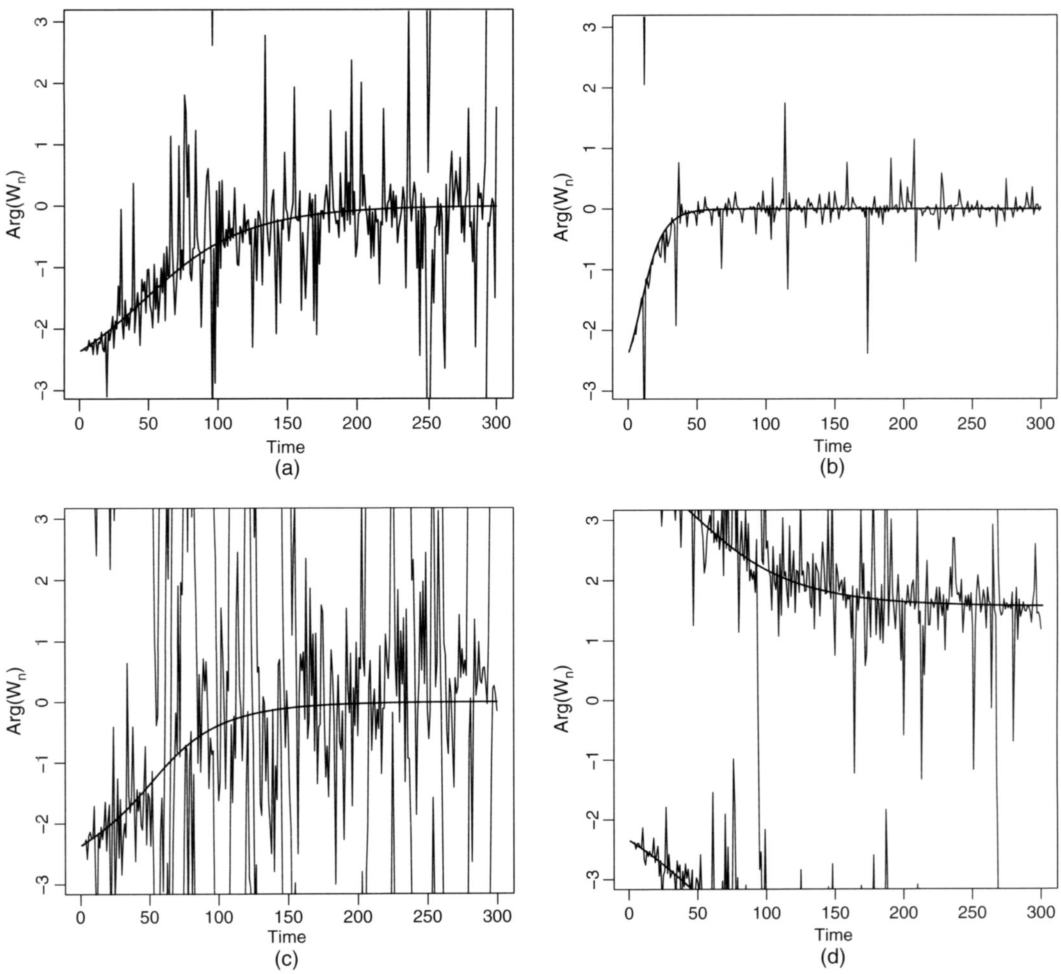

Fig. 1. Simulation of the Möbius Markov model $\{\arg(W_n)|W_0=\exp(-3\pi\mathrm{i})\}_{n=0}^{300}$ for some selected values of (β,φ) taken as (a) (0.01, 0.995), (b) (0.05, 0.995), (c) (0.01, 0.0985) and (d) (0.01i, 0.995): ———, $\arg[E\{W_n|W_0=\exp(-3\pi\mathrm{i}/4)\}]$

Summarizing the results in Section 3.3 and Fig. 1, the interpretation of each parameter is as follows. As stated in lemma 1 and as is clear from Fig. 1, the parameter $\arg(\beta)$ controls the mean direction of the limiting distribution, $\arg(\phi_\infty)$. 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 of ϕ_∞, 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 auto-correlation coefficient are considered. Finally we focus on the submodel of the stationary process with uniform marginals in Section 4.3.

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 distribution. The proof is clear from properties (4)–(6).

Lemma 3. Assume that $W_0 \sim C^*(\phi_0)$ where $\phi_0 \in \bar{D}$. Then $W_n \sim C^*\{\phi_n(\phi_0)\}$, $n \geqslant 0$.

We remark that, for the above model, the marginal of W_j and the conditional of W_j given $W_k = w_k$ $(j > k \geqslant 0)$ 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 the following lemma. The proof is given in appendix C of Kato (2009a).

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

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

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

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

where

$$Q = \frac{1}{|\beta|} \begin{pmatrix} \mathrm{Re}(\beta) & -\mathrm{Im}(\beta) \\ \mathrm{Im}(\beta) & \mathrm{Re}(\beta) \end{pmatrix},$$

$$a = \frac{(a_{00}a_{11} + a_{01}a_{10})\{1 - |\phi_\infty a_{10}/a_{11}|^2 + (|\phi_\infty| - |a_{10}/a_{11}|)^2\}}{2(a_{11}^2 - a_{10}^2)(1 + |\phi_\infty a_{10}/a_{11}|)}$$
$$+ \frac{(a_{00}a_{10} + a_{01}a_{11})(|\phi_\infty| - |a_{10}/a_{11}|)}{(a_{11}^2 - a_{10}^2)(1 + |\phi_\infty a_{10}/a_{11}|)},$$

$$b = \frac{\det(A)(1 - |\phi_\infty|^2 - |a_{10}/a_{11}|^2 + |\phi_\infty a_{10}/a_{11}|^2)}{2(a_{11}^2 - a_{10}^2)(1 + |\phi_\infty a_{10}/a_{11}|)},$$

$$A = \begin{pmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{pmatrix} = \begin{pmatrix} \varphi & |\beta|\varphi \\ |\beta| & 1 \end{pmatrix}^{m-n}.$$

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

Theorem 5. The Johnson and Wehrly (1977) correlation coefficient of (W_m, W_n) $(m > n)$ of the Markov process given in lemma 4 is

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

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 $\mathbf{V}_l$ $(l = m, n)$ is defined as in lemma 4. The correlation coefficients of Jupp and Mardia (1980) and Fisher and Lee (1983) are given by

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

and

$$\rho_{\mathrm{FL}} = \frac{\det\{E(\mathbf{V}_m\mathbf{V}_n')\}}{\surd[\det\{E(\mathbf{V}_m\mathbf{V}_m')\}\,\det\{E(\mathbf{V}_n\mathbf{V}_n')\}]} = \frac{4ab}{1 - |\phi_\infty|^4} \tag{14}$$

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 the auto-correlation coefficient for circular time series. Fig. 2 plots their auto-correlation coefficients for some selected values of parameters. All frames of Fig. 2 show that, as the lag between two variables increases, the auto-correlation between them decreases. Fig. 2 also implies that, the greater the value of φ, the greater the auto-correlation coefficient between W_n and W_{n+h}. Also, the larger the value of $|\beta|$, the greater the auto-correlation coefficient.

Compared with Fig. 2 of Fisher and Lee (1994), our model shows similar correlation patterns to their linked auto-regressive LAR(1) process when the concentration parameter φ of our model is large; both models show an 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.

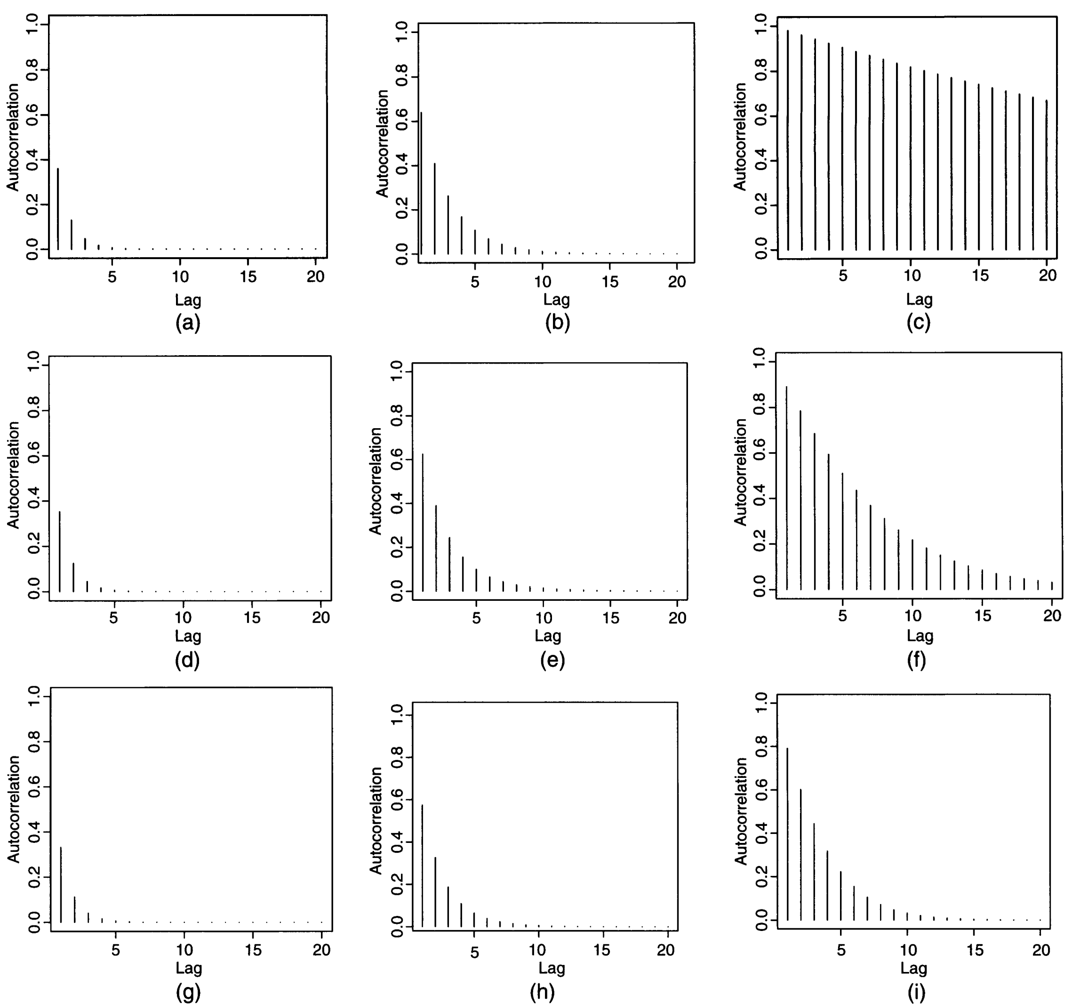

Fig. 2. Fisher and Lee (1983) auto-correlation coefficients for the stationary Markov process with parameters (a) $(\beta,\varphi)=(0,0.6)$, (b) $(\beta,\varphi)=(0,0.8)$, (c) $(\beta,\varphi)=(0,0.99)$, (d) $(\beta,\varphi)=(0.1,0.6)$, (e) $(\beta,\varphi)=(0.1,0.8)$, (f) $(\beta,\varphi)=(0.1,0.99)$, (g) $(\beta,\varphi)=(0.2,0.6)$, (h) $(\beta,\varphi)=(0.2,0.8)$ and (i) $(\beta,\varphi)=(0.2,0.99)$

In this section 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. Although 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 $\{W_n\}_{n=0}^{\infty}$ is the Möbius Markov process with $\beta=0$ and $W_0 \sim C^*(0)$. Then the circular uniform distribution $C^*(0)$ 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

$$f(w_m, w_n)=\frac{1}{4\pi^2}\frac{1-\varphi^{2(m-n)}}{|1-\varphi^{m-n}w_n\overline{w_m}|^2}, \qquad w_m, w_n \in \partial D. \tag{15}$$

This density is equivalent to that given by Kato (2009b), equation 4.1, which is obtained by 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 $C^*(\varphi^{m-n} w_m)$ distribution.

The auto-correlation coefficients for this submodel can be expressed in simple form as follows.

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 Möbius Markov process with $\beta = 0$ and $W_0 \sim C^*(0)$ are given by

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

All auto-correlations have the following properties:

(a) the auto-correlation is positive for any $m - n$;
(b) the greater the value of the lag $m - n$, the smaller the auto-correlation between W_m and W_n;
(c) as the lag $m - n$ tends to ∞, the auto-correlation tends to 0;
(d) as φ increases, the auto-correlation between W_m and W_n decreases.

Remember that the Fisher and Lee (1983) auto-correlation of this stationary process for some selected values of parameters is given in Figs 2(a)–2(c).

Furthermore, the following corollaries are obtained for the above model by using equation (4.3) and section 2.1 of Kato (2009b) respectively.

Corollary 2. Let $\{W_{1n}\}_{n=0}^{\infty}$ and $\{W_{2n}\}_{n=0}^{\infty}$ be independent Möbius Markov processes with $\beta = 0$ and $W_{10} \sim C^*(0)$ and $W_{20} \sim C^*(0)$ respectively. Then $\{W_{1n} W_{2n}\}_{n=0}^{\infty}$ is also a Möbius Markov process with $\beta = 0$, $W_{10} W_{20} \sim C^*(0)$, and the parameter φ replaced by φ^2.

Corollary 3. Let $\{B_t\}_{t \geqslant 0}$ be $\mathbb{C}$-valued Brownian motion starting at the origin. Suppose that 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 $n = 0, 1, \ldots$ and $0 < \varphi < 1$. Then a sequence of random variables $\{\varphi^n B_{\tau_n}\}_{n=0}^{\infty}$ is a Möbius Markov process with $\beta = 0$ and $B_{\tau_0} \sim C^*(0)$.

5. Comparison with existing Markov processes for circular data

As discussed in Section 3.2, the Möbius Markov process has some relationship with the IAR(1) model that was proposed by Fisher and Lee (1994). However, our process is not a submodel of their 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 distribution, and this property enables us to derive many of the results that were presented in Section 4 such as the limiting distribution, the stationary distribution and auto-correlation coefficient of the model.

The wrapped Cauchy and von Mises distributions 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 the IAR(1) 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 densities of the wrapped Cauchy and von Mises distributions show different behaviour, especially around modes and antimodes. When the mean resultant lengths of the densities are large, the wrapped Cauchy density takes greater values around the modes and antimodes and lesser values in between than the von Mises density (see Mardia (1972), section 3.4.9c).

Our stationary process with uniform marginals given in lemma 5 has a 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 model uses the von Mises distribution as an error distribution. As discussed before, the behaviour of the wrapped Cauchy density is different from that of the von Mises density if the distributions are highly concentrated. Therefore one can select which model to use via visual inspection of the histogram of the error or *Q*–*Q* plot. From the mathematical point of view, our model has some tractable properties as discussed in the previous paragraph. For example, the auto-correlation 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 that were 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 auto-correlation coefficient of the model can be expressed in relatively simple form. An advantage of the projected model is that the parameters of the model can be readily estimated by applying the EM algorithm. However, 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 (15).

6. Statistical inference for the process

6.1. Parameter estimation

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

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

Transform the observations and parameters by taking $w_n=\exp(\mathrm{i}\theta_n)$ and $\beta = r\exp(\mathrm{i}\nu)$, where $-\pi\leqslant\theta_n,\nu<\pi$ and $0\leqslant r<1$. Then the above function can be expressed as

$$\log\{L_q(r,\nu,\varphi|\theta_0)\} = C + T\log(1-\varphi^2) - \sum_{n=1}^{T}\log\{1-2\varphi\cos(\theta_n-\nu_n)+\varphi^2\}, \qquad (16)$$

where $\nu_n=\theta_{n-1}-2\arg[1+r\exp\{\mathrm{i}(\theta_{n-1}-\nu)\}]$ and C is a constant that does not depend on 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 ν 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

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

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

$$\log\{L_s(r,\nu,\varphi)\}=C+\log\{L_q(r,\nu,\varphi|\theta_0)\}+\log(1-|\phi_\infty|^2)-\log\{1-2|\phi_\infty|\cos(\theta_0-\nu)+|\phi_\infty|^2\}, \tag{17}$$

where

$$|\phi_\infty|=\frac{\varphi-1+\surd\{(1-\varphi)^2+4\varphi r^2\}}{2r}$$

and $L_q(r,\nu,\varphi)$ is given by equation (16). 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). In the general case of the stationary process, however, it seems that no explicit formulae for maximum likelihood estimates arise, and the score equations do not give rise to any simple iterative method of solution. Therefore, in practice, we resort to direct maximization with multiple restarts to check for multiple maxima.

6.2. Fisher information matrix

The Fisher information matrix is helpful to obtain confidence intervals of the maximum likelihood estimates. We consider the observed information for our stationary process. Each element of the observed information is of the form

$$\iota_{\alpha_1\alpha_2}=n^{-1}\frac{\partial\log(L_s)}{\partial\alpha_1}\frac{\partial\log(L_s)}{\partial\alpha_2},$$

where α_1 and α_2 are the parameters of the process. Then it is clear that the elements of the information matrix can be easily calculated by applying the following result:

$$\begin{aligned}\frac{\partial}{\partial r}\log(L_s)&=\sum_{n=1}^{T}\frac{2\varphi\sin(\theta_n-\nu_n)}{1-2\varphi\cos(\theta_n-\nu_n)+\varphi^2}\frac{\partial\nu_n}{\partial r}\\&\quad+2\left\{\frac{\cos(\theta_0-\nu)-|\phi_\infty|}{1-2|\phi_\infty|\cos(\theta_0-\nu)+|\phi_\infty|^2}-\frac{|\phi_\infty|}{1-|\phi_\infty|^2}\right\}\frac{\partial|\phi_\infty|}{\partial r},\\ \frac{\partial}{\partial\nu}\log(L_s)&=\sum_{n=1}^{T}\frac{2\varphi\sin(\theta_n-\nu_n)}{1-2\varphi\cos(\theta_n-\nu_n)+\varphi^2}\frac{\partial\nu_n}{\partial\nu}+\frac{2|\phi_\infty|\sin(\theta_0-\nu)}{1-2|\phi_\infty|\cos(\theta_0-\nu)+|\phi_\infty|^2},\\ \frac{\partial}{\partial\varphi}\log(L_s)&=-\frac{2T\varphi}{1-\varphi^2}+2\sum_{n=1}^{T}\frac{\cos(\theta_n-\nu_n)-\varphi}{1-2\varphi\cos(\theta_n-\nu_n)+\varphi^2}\\&\quad-2\left\{\frac{|\phi_\infty|}{1-|\phi_\infty|^2}+\frac{|\phi_\infty|-\cos(\theta_0-\nu)}{1-2|\phi_\infty|\cos(\theta_0-\nu)+|\phi_\infty|^2}\right\}\frac{\partial|\phi_\infty|}{\partial\varphi},\end{aligned}$$

where

$$\begin{aligned}\frac{\partial\nu_n}{\partial r}&=\frac{-2\sin(\theta_n-\nu)}{1+2r\cos(\theta_n-\nu)+r^2},\\ \frac{\partial\nu_n}{\partial\nu}&=\frac{2r\{r+\cos(\theta_n-\nu)\}}{1+2r\cos(\theta_n-\nu)+r^2},\\ \frac{\partial|\phi_\infty|}{\partial r}&=r^{-1}\{(1-\varphi)^2+4\varphi r^2\}^{-1/2}(1-\varphi)|\phi_\infty|,\\ \frac{\partial|\phi_\infty|}{\partial\varphi}&=(2r)^{-1}[1+(\varphi-1+2r^2)\{(1-\varphi)^2+4\varphi r^2\}^{-1/2}].\end{aligned}$$

For the Möbius Markov process which does not assume any initial distribution, elements of the observed Fisher information matrix can be obtained by removing the terms which include $|\phi_\infty|$ from the elements of the information matrix given above.

6.3. A hypothesis test

It might be of interest to see whether the proposed stationary process provides a better fit than a simpler Markov process $W_n = W_{n-1}\varepsilon_n$. To investigate this, we consider the test $H_0: r=0$ *versus* $H_1: r>0$. The log-likelihood ratio test gives the test statistic

$$T = -2\log\{\max(L_{s0})/\max(L_{s1})\},$$

where $\max(L_{s0}) = \max_\varphi\{L_s(r=0, \nu=0, \varphi)\}$ and $\max(L_{s1}) = \max_{r,\nu,\varphi}\{L_s(r, \nu, \varphi)\}$. Here $\max(L_{s0})$ can be obtained by using the algorithm of Kent and Tyler (1988) and $\max(L_{s1})$ is calculated from the result of maximum likelihood estimation that was discussed before. Under the null hypothesis, the statistic T is distributed as a χ^2-distribution with 2 degrees of freedom. (Note that the parameter under the null hypothesis actually takes a value on the centre of the two-dimensional unit disc, not on the boundary of the line, because r is the absolute value of a D-valued parameter β.) When T is significantly large, we reject the null hypothesis. The same test based on the quasi-log-likelihood L_q can be conducted with a similar approach.

7. Model selection and diagnostic checks

Fisher and Lee (1994), section 3.1, considered model selection for four families of models presented in their paper. Since our model is related to one of their families, the IAR model (see Section 3.2), our process can be compared with the other three families of theirs in a similar manner to their discussion on model selection for their four families. Therefore, as well as the IAR model, our process is also flexible and capable of capturing a wide range of behaviours. For example, our process is likely to provide a reasonable fit to data which tend to cluster around a certain direction. In addition it is also possible to apply our process to data which are uniformly distributed on the circle because our process has dispersed marginal distributions for small $|\beta|$ or φ (see lemma 1). In contrast, other models, e.g. a linked auto-regressive moving average process, are advocated for a time series whose mean directions keep moving and do not converge to a certain direction as time tends to ∞.

The difference between our process and an IAR model arises when the concentration parameters of the error distributions are large. In this situation our process provides a better fit to a time series whose observations tend to take close values to the transformed previous states and include some jumps. In contrast, the IAR model is more suitable for data which cluster in an arc and have few jumps. If the concentration parameters of angular errors are small, then our process and IAR model can be used interchangeably. A similar discussion holds when we compare our stationary submodel having uniform marginals with the model of Wehrly and Johnson (1980).

Some other measures for model selection are also available. For example, information criteria such as the Akaike information criterion AIC and Bayesian information criterion BIC are also helpful to choose an optimal process among fitted models. To compare the Möbius Markov process with our submodel with $\beta=0$, one can use the log-likelihood ratio test that was given in Section 6.3. Other tests based on large sample theory, such as the Wald test and score test, can also be used for this.

There are some known methods for the diagnostic checks of the circular models. One method is, as in linear time series, the inspection of the correlogram. The auto-correlation coefficients of our stationary process can be expressed in explicit form as in theorem 5, and it is possible to calculate them without involving any integration. Another way for diagnostic checking is a Q–Q-plot, which is helpful to see how successful the fit of the estimated errors is. We shall see more details about these methods through an application in the next section.

8. Application

We consider a time series of wind directions measured hourly at a weather station in Waldhof, Germany. These wind data are part of a larger data set that contains atmospheric observations from the Umweltbundesamt (German Federal Environment Agency). The full data set is available from the World Meteorological Organization World Data Centre for Greenhouse Gases, `http://gaw.kishou.go.jp/cgi-bin/wdcgg/accessdata.cgi?index=NGL653N00-UBAG&select=inventory` (wind directions WD, parameter O_3, in 2007). Here we discuss a time series of 96 wind directions measured hourly between 10 p.m. on May 23rd and 9 p.m. on May 27th, 2007, except for two missing observations. We interpolate each of the missing observations by the mean direction of two wind directions recorded 1 h earlier and later.

Fig. 3(a) plots the time series of the wind directions. This frame implies that, despite the fact that the observations take values in a wide range of the circle, there is an overall tendency that the most observations are recorded around π. From this frame and the discussion given in Section 7, it appears appropriate to fit our process and the IAR(1) model to the data. The sample auto-correlation coefficients proposed by Fisher and Lee (1994), equation (3.1), are plotted in Fig. 3(b), implying that most of the coefficients are positive and those values are not small when the lag is equal to or less than 3. This feature of the correlogram seems quite similar to that in Fig. 2(g). We fit our stationary process, its submodel having uniform marginals and two inverse models (9) of Fisher and Lee (1994) based on maximum likelihood. To maximize the likelihood functions, we adopt an optimization method, the PORT routine, which can be implemented by using the command `nlminb` in the statistical software R.

First, consider our stationary process. For the estimation of the parameters, we maximize the log-likelihood function (17). The estimated maximum log-likelihood and the maximum likelihood estimates of the parameters are given by $\log(L_s) = -117.8$, $\hat{r} = 0.170$, $\hat{\nu} = -3.11$ and $\hat{\varphi} = 0.712$. Hence, the estimated parameter of the limiting distribution is $\hat{\phi}_\infty = 0.348 \exp(-3.11\mathrm{i})$. As for the stationary submodel with uniform marginals presented in Section 4.3, we obtain the maximum log-likelihood and the maximum likelihood estimate as $\log(L_s) = -124.3$ and $\hat{\varphi} = 0.664$.

Next we discuss two inverse Markov, or IAR(1), models of Fisher and Lee (1994). As the first model, we define the link function U in equation (9) as $U(x;\nu) = \tan\{(x-\nu)/2\}$. Note that this model is not a stationary process in general. The estimated maximum log-likelihood and estimated parameters are $\log(L) = -130.1$, $\hat{\lambda}_c^* = 0.617$, $\hat{\nu} = 2.80$ and $\hat{\kappa} = 1.74$. The second model supposes that U is the probit link, namely $U(x;\nu) = \Phi^{-1}\{(x-\nu)/2\pi + 0.5\}$, where

$$\Phi(x) = \int_{-\infty}^{x} \frac{\exp(-t^2/2)}{\surd(2\pi)}\, \mathrm{d}t.$$

Then we obtain the maximum log-likelihood and estimated parameters as $\log(L) = -128.8$, $\hat{\lambda}_c^* = 1.25$, $\hat{\nu} = 0.0384$ and $\hat{\kappa} = 1.78$.

Since the numbers of the parameters for our stationary process and the two IAR models are the same, model selection based on some information criterion 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 stationary process proposed is best among these three models.

It would be of interest to see that our stationary process provides a significantly better fit than its submodel having uniform marginals. The log-likelihood ratio test provided in Section 6.3 yields the test statistic $T=-2\{-124.3-(-117.8)\}=13.0$ with $P(T\geqslant 13.0)\simeq 1.50\times 10^{-3}$. This test is highly significant and thus the null hypothesis, $r=0$, is rejected. Also, according to AIC and BIC, our stationary process provides a better fit than its submodel with $r=0$.

Fig. 3(c) shows a Q–Q-plot, a plot of quantiles of the estimated angular error distribution against those of the empirical error distribution. Fig. 3(c) suggests that the estimated error distribution provides a reasonable fit to the data set. The theoretical auto-correlation coefficient (14) for the fitted Markov model is displayed in Fig. 3(d). Both Fig. 3(d) and Fig. 3(b) appear to show that the auto-correlation is fairly large when the lag is not greater than 3.

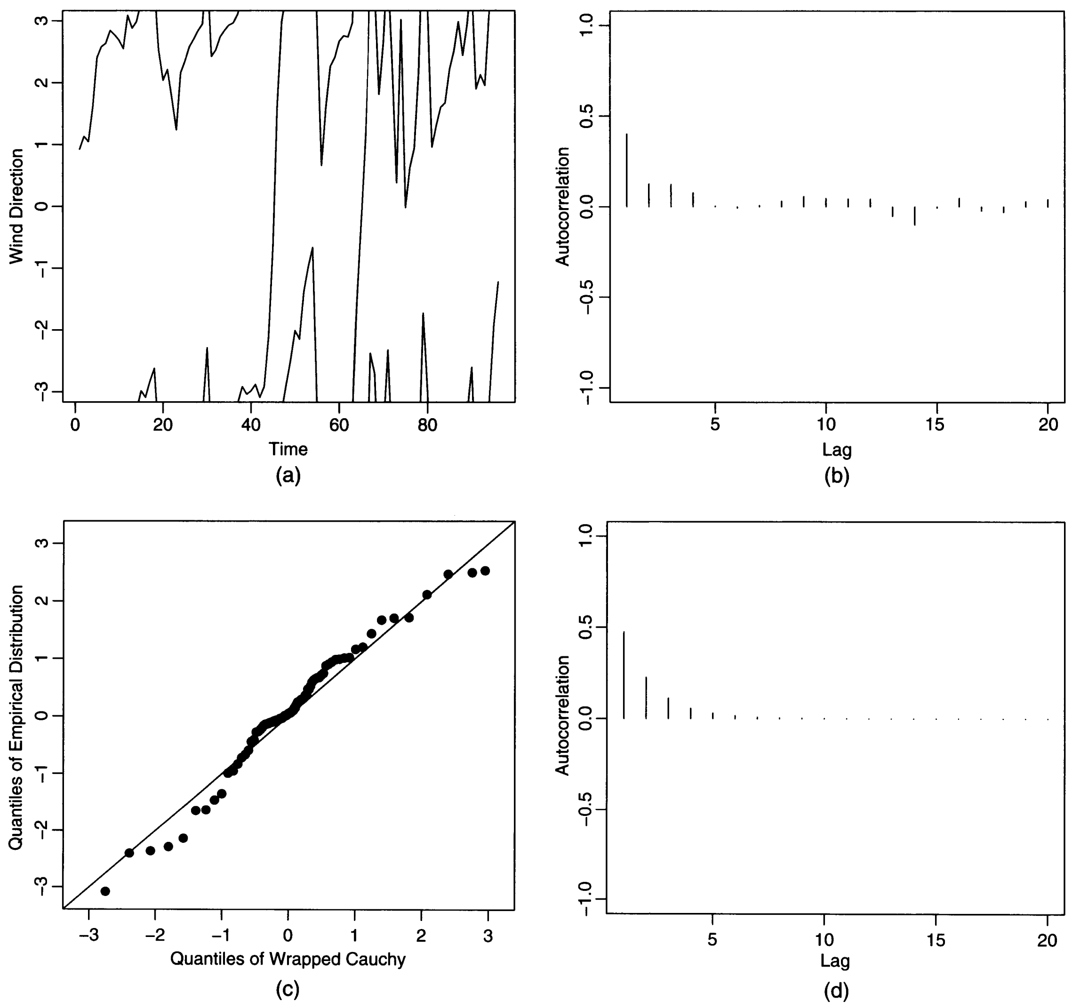

Fig. 3. (a) Series of 96 wind directions at Waldhof, Germany, measured hourly between 10 p.m. on May 23rd and 9 p.m. on May 27th, 2007, (b) Fisher and Lee (1983) sample auto-correlation coefficients, (c) Q–Q-plot for the Möbius Markov process where quantiles of the angular error distribution (x-axis) and of the empirical error distribution (y-axis) are plotted and (d) Fisher and Lee (1983) theoretical auto-correlation coefficients for the fitted process proposed

Acknowledgements

I am grateful to the Associate Editor and two referees for valuable comments which improved the quality of the paper. My gratitude also goes to Professor M. C. Jones and Professor K. Shimizu for helpful advice on the first draft of the paper. I thank the Umweltbundesamt (German Federal Environment Agency) and the World Meteorological Organization World Data Centre for Greenhouse Gases for providing the wind direction data that were used in Section 8. 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).

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.
Downs, T. D. and Mardia, K. V. (2002) Circular regression. *Biometrika*, **89**, 683–697.
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 circular 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. Statist.*, **13**, 325–347.
Jammalamadaka, S. R. and SenGupta, A. (2001) *Topics in Circular Statistics*. Singapore: World Scientific Publishers.
Johnson, R. A. and Wehrly, T. (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.
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. (2009a) A Markov process for circular data. *Research Memorandum 1086*. Institute of Statistical Mathematics, Tokyo.
Kato, S. (2009b) A distribution for a pair of unit vectors generated by Brownian motion. *Bernoulli*, **15**, 898–921.
Kato, S., Shimizu, K. and Shieh, G. S. (2008) A circular–circular regression model. *Statist. Sin.*, **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. (1972) *Statistics of Directional Data*. New York: Academic Press.
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.
Rudin, W. (1987) *Real and Complex Analysis*, 3rd edn. 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.