Estimation of Trigonometric Moments for Circular Distribution of MA(p) Type by Using Binary Series

Yuichi Goto

Abstract. Directional statistics have received a great deal of interest in recent years, and a variety of distributions on the circle have been proposed. In this paper, we propose circular distributions of a moving average model of order p type which includes the cardioid distribution, and discuss estimation of trigonometric moments based on binary series. We give an explicit form of the root n consistent estimator based on clipped series, which enables us to construct an efficient estimator by the Newton–Raphson iterative method. We also show a robustness of the proposed estimator when the probability density function is contaminated with a noise term.

1 Introduction Directional statistics is an important field which deals with directional data. The history of directional statistics dates back to 1950s. Fisher (1953) had large influence and appealed the necessity of directional statistics. After that, many authors tackled the problem (see Mardia (1975); Watson (1983); Fisher et al. (1993)). In recent years, directional statistics has attracted attention because of Mardia and Jupp (2000).

Many distributions on the circle have been developed (e.g. uniform, cardioid, wrapped Cauchy, von Mises distribution). These distributions are closely related to the spectral density functions in time series with complex valued coefficients. For example, the spectral density of the autoregressive model of order 1, that of the moving average model of order 1, and that of the autoregressive model of order 2 correspond to wrapped Cauchy distribution, cardioid distribution, and the more flexible distribution proposed by Kato and Jones (2013), respectively.

Binary series are processes that each of realizations takes value 0 or 1. The execution time of methods based on clipped processes are significantly short, and estimation accuracy of methods based on 0-1 valued processes are high when the original processes are contaminated with outliers (see Bagnall and Janacek (2005), Kedem (1994, p.172), Goto and Taniguchi (2019), Goto and Taniguchi). Methods based on binary series have been applied to various fields including biology and linguistics. For example, analysis of vocal sounds of humpback whales of Kedem and Li (1989); speech discrimination of Panagiotakis and Tziritas (2005)); and emotion recognition using brain signals of Petrantonakis and Hadjileontiadis (2010).

Binary series have been studied by many researchers (see Rice (1944), Lomnicki and Zaremba (1955), McNeil (1967), Kedem (1980), Kedem (1994)). Rice (1944) gave a pioneer study in this field, and showed a relationship between correlations of Gaussian processes and correlations of a binary series generated by the Gaussian processes. Kedem (1980) showed the asymptotic normality of the estimator of autocorrelation based on clipped processes. In recent years, related to binary data, the categorical time series (see Fokianos and Kedem (2003)) and the quantile based spectra (e.g. Li (2014) and Dette et al. (2015)) have been developed. However, binary series in directional statistics have not yet been investigated.

2010 *Mathematics Subject Classification.* 62H11, 62H12 .

Key words and phrases. asymptotic normality; binary series; circular distribution; Cramér–Rao lower bound; efficient estimator; robustness; trigonometric moment.

In this paper, we propose a family of circular distributions of a moving average model of order p type, and discuss estimation of trigonometric moments based on binary series. We derive an explicit form of the root n consistent estimator. Although the estimator based on clipped series does not attain Cramér–Rao lower bound, it enables us to construct efficient estimator by the Newton–Raphson iterative method. We also show a robustness of the estimator when the true probability density function is contaminated with noise. The finite sample performance of proposed estimator is also investigated.

The paper is organized as follows: In Section 2, we introduce circular distributions of the moving average model of order p type and the estimator of trigonometric moments based on binary series for the proposed distribution. We show the asymptotic normality and compare the asymptotic variance with Cramér–Rao lower bound. In Section 3, we elucidate a robustness of the estimator when the probability density function is contaminated with noise. The finite sample performance of proposed estimator is investigated, and asymptotic normality of the proposed estimator is illustrated by computer simulation in Section 4. Finally, we conclude this paper with proofs of the theorems and the proposition in Sections 2 and 3.

2 Settings and Main Result In this section, we define a family of circular distributions of MA(p) type and propose a root n consistent estimator based on binary series. After that, we show the asymptotic normality and compare the asymptotic variance of the proposed estimator with Cramér–Rao lower bound.

Throughout this paper, we consider a family of circular distributions of MA(p) type whose probability density function is defined by

$$p(\theta) = \frac{1}{2\pi(1+\phi_1^2+\cdots+\phi_p^2)}|\phi(e^{i\theta})|^2 \tag{1}$$

where $\phi(z) = 1 + \phi_1 z + \phi_2 z^2 + \cdots + \phi_p z^p$ and $\phi_j \in \mathbb{R}$ for any j.

Let $\{\Theta_k : k \in \mathbb{N}\}$ be independent random variables with a common circular distribution defined by (1). From the residue theorem and symmetry of (1), the j-th sine and cosine moments can be obtained as

$$\begin{aligned}
&\mathrm{E}\{\sin(j\Theta_k)\} = 0 \quad \text{for } j \in \mathbb{Z}, \\
&\mathrm{E}\{\cos(j\Theta_k)\} = \begin{cases} \frac{\phi_j+\phi_{j+1}\phi_1+\cdots+\phi_p\phi_{p-j}}{1+\phi_1^2+\cdots+\phi_p^2} & \text{for } |j| \le p, \\ 0 & \text{for } |j| \ge p+1, \end{cases}
\end{aligned}$$

respectively. Then, the mean resultant length and the mean direction of $\{\Theta_k : k \in \mathbb{N}\}$ can be obtained as

$$\begin{aligned}
|\mathrm{E}\{e^{i\Theta_k}\}| &= \left|\frac{\phi_1+\phi_2\phi_1+\cdots+\phi_p\phi_{p-1}}{1+\phi_1^2+\cdots+\phi_p^2}\right|, \\
\arg \mathrm{E}\{e^{i\Theta_k}\} &= \begin{cases} 0 & \phi_1+\phi_2\phi_1+\cdots+\phi_p\phi_{p-1} > 0, \\ \pi & \phi_1+\phi_2\phi_1+\cdots+\phi_p\phi_{p-1} < 0, \\ \text{undefined} & \phi_1+\phi_2\phi_1+\cdots+\phi_p\phi_{p-1} = 0, \end{cases}
\end{aligned}$$

respectively. From Mardia and Jupp (2000, p.31), (1) can be written as

$$p(\theta) = \frac{1}{2\pi}\left(1+\sum_{j=1}^{p}\rho_j\cos(j\theta)\right), \tag{2}$$

where $\rho_j = 2(\phi_j + \phi_{j+1}\phi_1 + \cdots + \phi_p\phi_{p-j})/(1 + \phi_1^2 + \cdots + \phi_p^2)$. If we take $p = 1$, (2) is the well-known cardioid distribution (see Mardia and Jupp (2000, p.45)). Clearly, if $\phi_j = 0$ for any $j \in \{1, \ldots, p\}$, (2) is a uniform distribution. The proposed model (1) is generally non-identifiable. Actually, for $p = 2$ and $(\phi_1, \phi_2, \psi_1, \psi_2) := (0, -\frac{1}{2}, \pm\sqrt{\frac{1}{2}}, -1)$, we have $p(\theta; \phi_1, \phi_2) = p(\theta; \psi_1, \psi_2)$.

In this paper, we discuss the estimation problem of $\rho_1, \ldots, \rho_p$ of the proposed probability density function by using clipped series. Hereafter, we confine ourselves to the case that $(\phi_1, \ldots, \phi_p)$ satisfies $\phi_1 + \phi_2\phi_1 + \cdots + \phi_p\phi_{p-1} \geq 0$. Define $(\alpha_1, \alpha_2, \ldots, \alpha_p) \in \mathbb{R}^p$ such that $0 < \alpha_1 < \alpha_2 < \cdots < \alpha_p < \pi$. For each α_j, $j = 1, \ldots, p$, binary series $\{X_k^j\}$ are defined, for any $j = 1, \ldots, p$,

$$X_k{}^j = \begin{cases} 1 & -\alpha_j \leq \Theta_k \leq \alpha_j, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Applying the technique for the derivation of an orthant probability for normal distribution (see Kedem (1994, p.48)), we have the following equation

$$\begin{pmatrix} P(-\alpha_1 \leq \Theta_1 \leq \alpha_1) \\ P(-\alpha_2 \leq \Theta_1 \leq \alpha_2) \\ \vdots \\ P(-\alpha_p \leq \Theta_1 \leq \alpha_p) \end{pmatrix} = \begin{pmatrix} \frac{\alpha_1}{\pi} \\ \frac{\alpha_2}{\pi} \\ \vdots \\ \frac{\alpha_p}{\pi} \end{pmatrix} + \frac{1}{2\pi} \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{p1} & b_{p2} & \cdots & b_{pp} \end{pmatrix} \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{pmatrix},$$

where

$$\begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{p1} & b_{p2} & \cdots & b_{pp} \end{pmatrix} = \begin{pmatrix} \int_{-\alpha_1}^{\alpha_1} \cos\theta \mathrm{d}\theta & \int_{-\alpha_1}^{\alpha_1} \cos 2\theta \mathrm{d}\theta & \cdots & \int_{-\alpha_1}^{\alpha_1} \cos p\theta \mathrm{d}\theta \\ \int_{-\alpha_2}^{\alpha_2} \cos\theta \mathrm{d}\theta & \int_{-\alpha_2}^{\alpha_2} \cos 2\theta \mathrm{d}\theta & \cdots & \int_{-\alpha_2}^{\alpha_2} \cos p\theta \mathrm{d}\theta \\ \vdots & \vdots & \ddots & \vdots \\ \int_{-\alpha_p}^{\alpha_p} \cos\theta \mathrm{d}\theta & \int_{-\alpha_p}^{\alpha_p} \cos 2\theta \mathrm{d}\theta & \cdots & \int_{-\alpha_p}^{\alpha_p} \cos p\theta \mathrm{d}\theta \end{pmatrix}.$$

Here, we suppose the observed stretch $\{\Theta_1, \ldots, \Theta_n\}$ is available. We choose $(\alpha_1, \alpha_2, \ldots, \alpha_p) \in \mathbb{R}^p$ adequately so that $(b_{ij})_{i,j=1}^p$ is a nonsingular matrix, and substitute $\left(1/n \sum_{k=1}^n X_k{}^1, \ldots, 1/n \sum_{k=1}^n X_k{}^p\right)^{\mathrm{T}}$ for $\left(P(-\alpha_1 \leq \Theta_1 \leq \alpha_1), \ldots, P(-\alpha_p \leq \Theta_1 \leq \alpha_p)\right)^{\mathrm{T}}$. Then, the binary estimator $\left(\hat{\rho}_1, \ldots, \hat{\rho}_p\right)^{\mathrm{T}}$ can be defined as

$$\begin{pmatrix} \hat{\rho}_1 \\ \hat{\rho}_2 \\ \vdots \\ \hat{\rho}_p \end{pmatrix} = 2\pi \begin{pmatrix} b^{11} & b^{12} & \cdots & b^{1p} \\ b^{21} & b^{22} & \cdots & b^{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b^{p1} & b^{p2} & \cdots & b^{pp} \end{pmatrix} \begin{pmatrix} \frac{1}{n}\sum_{k=1}^n X_k{}^1 - \frac{\alpha_1}{\pi} \\ \frac{1}{n}\sum_{k=1}^n X_k{}^2 - \frac{\alpha_2}{\pi} \\ \vdots \\ \frac{1}{n}\sum_{k=1}^n X_k{}^p - \frac{\alpha_p}{\pi} \end{pmatrix},$$

where $(b^{ij})_{i,j=1}^p$ is the inverse matrix of $(b_{ij})_{i,j=1}^p$.

Before we derive the asymptotic distribution of the proposed estimator, we give some examples that $(b_{ij})_{i,j=1}^p$ is a nonsingular matrix for specific models.

Example 2.1. MA(2) case: if we take $\alpha_1 = \frac{\pi}{4}$ and $\alpha_2 = \frac{\pi}{2}$, then

$$(b_{ij})_{i,j=1}^2 = \begin{pmatrix} \sqrt{2} & 1 \\ 2 & 0 \end{pmatrix}, \quad (b^{ij})_{i,j=1}^2 = \begin{pmatrix} 0 & \frac{1}{2} \\ 1 & -\frac{1}{\sqrt{2}} \end{pmatrix}.$$

Example 2.2. MA(3) case: if we take $\alpha_1 = \frac{\pi}{4}$, $\alpha_2 = \frac{\pi}{2}$, and $\alpha_3 = \frac{3\pi}{4}$, then

$$(b_{ij})_{i,j=1}^3 = \begin{pmatrix} \sqrt{2} & 1 & \frac{\sqrt{2}}{3} \\ 2 & 0 & -\frac{2}{3} \\ \sqrt{2} & -1 & \frac{\sqrt{2}}{3} \end{pmatrix}, \quad (b^{ij})_{i,j=1}^3 = \begin{pmatrix} \frac{1}{4\sqrt{2}} & \frac{1}{4} & \frac{1}{4\sqrt{2}} \\ \frac{1}{2} & 0 & -\frac{1}{2} \\ \frac{3}{4\sqrt{2}} & -\frac{3}{4} & \frac{1}{4\sqrt{2}} \end{pmatrix}.$$

The following theorem shows that the asymptotic normality of the proposed estimator.

Theorem 2.1. *It holds that*

$$\sqrt{n} \begin{pmatrix} \hat{\rho}_1 - \rho_1 \\ \hat{\rho}_2 - \rho_2 \\ \vdots \\ \hat{\rho}_p - \rho_p \end{pmatrix} \Rightarrow N(0, \boldsymbol{V}),$$

where $\boldsymbol{V} = (v_{ij})_{i,j=1}^p$ *and*

$$\begin{aligned} v_{ij} = \; & 4\pi^2 \textstyle\sum_{s,k=1}^p b^{is} b^{jk} \{P(-\alpha_s \le \Theta_1 \le \alpha_s, -\alpha_k \le \Theta_1 \le \alpha_k) \\ & -P(-\alpha_s \le \Theta_1 \le \alpha_s) P(-\alpha_k \le \Theta_1 \le \alpha_k)\}. \end{aligned}$$

Next, we investigate whether our proposed method attains the Cramér–Rao lower bound or not. For simplicity, we confine ourselves to the case of circular distributions of MA(1) type.

Proposition 2.1. *The Cramér–Rao lower bound is given by*

$$\mathcal{I}^{-1}(\rho_1) = 1 - \rho_1^2 + \sqrt{1 - \rho_1^2}.$$

Proposition 2.1 enables us to compare the asymptotic variance of the proposed estimator with the Cramér–Rao lower bound. Thus, we have the following statement.

Remark 2.1. *The Binary estimator is not efficient.*

Actually, If we consider the case $\rho_1 = 1$, then it is easy to see that

$$(\text{Covariance of } \hat{\rho}_1) - \mathcal{I}^{-1}(\rho_1) > 0.$$

Remark 2.1 is not a preferable property of the estimator. However, from Hosoya and Taniguchi (1982, Theorem 5.1), we can construct an efficient estimator from $\hat{\rho}_1, \ldots, \hat{\rho}_p$ by the Newton–Raphson iterative method. In the next section, we show a robust property of the estimator when the true probability density function is contaminated.

3 Robustness of proposed estimator against contamination In the previous section, we showed that proposed estimator is root n consistent, and it enable us to construct the efficient estimator by the Newton–Raphson iterative method. In this section, we show our estimator is robust when the true probability density function is contaminated with noise. Let $q(\cdot)$ be a contaminated probability density function defined, for $\theta \in [-\pi, \pi]$ and some $\beta \in (0, \pi/2)$, as

$$q(\theta) = \begin{cases} p(\theta) & \text{if } -\pi + \beta \le \theta \le \pi - \beta, \\ cg(\theta) & \text{otherwise,} \end{cases}$$

where $p(\theta)$ is defined by (1), $g(\theta)$ is a non-negative function with $\int_{\pi-\beta}^{\pi+\beta} g(\theta)\mathrm{d}\theta > 0$, c is some constant such that $q(\theta)$ is probability density function. In the above setting, $cg(\theta)$ corresponds to a noise. Assume that the process $\{\Theta_k : k \in \mathbb{N}\}$ is misspecified, that is, the true model of $\{\Theta_k : k \in \mathbb{N}\}$ comes from $q(\theta)$, but we fit the process to $p(\theta)$.

Theorem 3.1. *If α_p and β satisfy $\alpha_p < \pi - \beta$, then the our estimator does not be influenced by the contamination.*

Thus, the proposed method is robust against the contamination of probability density.

4 Simulation Study In this section, we study finite sample performance of the proposed method, and confirm the asymptotic normality of the proposed estimator based on binary process. In this simulation, the circular distributions of MA(1) and MA(2) types are discussed. First, we illustrate finite sample performance. The procedure is the following; first, we generate random variables $\{U_i : i = 1, \dots, n\}$ ($n = 100, 300, 500, 1000$), which follows i.i.d. standard uniform distribution. Next, we compute $\{\Theta_i = 1 \dots, n\} := \{F^{-1}(U_i) : i = 1, \dots, n\}$, where F^{-1} is the generalized inverse of a distribution function of (1), which follows the circular distribution of MA(p) type for $p = 1, 2$. Then, we calculate the proposed estimators of ρ_1 and ρ_2 for the each set of parameters $\phi_1 = 0.4, 0.7, -0.5$ and angulars $\alpha_1 = \pi/4, \pi/2, 3\pi/4$ for MA(1) type distribution, and $(\phi_1, \phi_2) = (0.7, 0.4), (1.0, 0.7), (0.9, -0.3)$ and $(\alpha_1, \alpha_2) = (\pi/4, \pi/2), (\pi/2, 3\pi/4)$ for MA(2) type distribution. We iterate 1000 times and calculate mean absolute error, defined as $\mathrm{MAE_j} := \sum_{k=1}^{1000} |\hat{\rho}_j^{(k)} - \rho_j|/n$ for $j = 1, 2$, where $\hat{\rho}_j^{(k)}$ is the estimator of ρ_j of k-th iteration. Next, we calculate, for $n = 1000$, $\{\sqrt{n}(\hat{\rho}_1^{(k)} - \rho_1);\ k = 1, \dots, 10000\}$ and $\{\sqrt{n}(\hat{\rho}_1^{(k)} - \rho_1), \sqrt{n}(\hat{\rho}_2^{(k)} - \rho_2);\ k = 1, \dots, 10000\}$ for circular distributions of MA(1) type with $\phi_1 = 0.7$ and MA(2) type with $(\phi_1, \phi_2) = (0.7, 0.4)$, respectively to confirm the asymptotic normality of the proposed estimator. Then, we give the Q-Q plots in Figures 1, 2, and 3. We also provide the Kolmogorov–Smirnov test of normality to check the asymptotic normality of the proposed estimator. The null hypothesis is that $\{\sqrt{n}(\hat{\rho}_1 - \rho_1)\}$ follows the normal distribution for large n. For $n = 100$, 1000, 10000, $\{\sqrt{n}(\hat{\rho}_1^{(k)} - \rho_1); k = 1, \dots, 100\}$ and $\{\sqrt{n}(\hat{\rho}_1^{(k)} - \rho_1), \sqrt{n}(\hat{\rho}_2^{(k)} - \rho_2);\ k = 1, \dots, 100\}$ are calculated for circular distributions of MA(1) type with $\phi_1 = 0.7$ and MA(2) type with $(\phi_1, \phi_2) = (0.7, 0.4)$. Then, we compute p-value by using R-function ks.test() when $\{\sqrt{n}(\hat{\rho}_1^{(k)} - \rho_1); k = 1, \dots, 100\}$ regarded as a set of i.i.d. observations with respect to k. Note that, from the definition of binary estimator, we possibly have the exact same value $\hat{\rho}_j^{(k)} = \hat{\rho}_j^{(k')}$ for some k and $k'(\neq k)$. Therefore, we added a small perturbation to $\{\sqrt{n}(\hat{\rho}_1^{(k)} - \rho_1); k = 1, \dots, 100\}$ by R function jitter() in order to compute p-value (see Robert et al. (2010, p.17-18)).

The results are shown in Tables 1 and 2 and Figures 1, 2, and 3. Tables 1 and 2 show the proposed estimator works well, and the mean absolute errors get smaller as the sample size increases. In Table 1, for $\phi_1 = 0.4$ and 0.7 in MA(1) type model, MAE_1 is smallest when $\alpha_1 = 3\pi/4$ among $\alpha_1 = \pi/4,\ \pi/2,\ 3\pi/4$. On the other hand, for $\phi_1 = -0.5$ in MA(1) type model, MAE_1 is smallest when $\alpha_1 = \pi/4$ among three angulars. It is because MA(1) model with $\phi_1 = -0.5$ has a mean direction π. The mean directions of the proposed model are 0 in the other cases. In Table 2, MAE_1 are smaller than MAE_2. For better estimation of ϕ_2, the set of angulars $(\pi/2, 3\pi/4)$ is better than $(\pi/4, \pi/2)$. Regarding to estimation of ϕ_1, both sets of angulars $(\pi/2, 3\pi/4)$ and $(\pi/4, \pi/2)$ provide almost the same MAE_1. Figures 1 2, and 3 show that almost of all points are on the reference line, that is, we could confirm that our estimator has asymptotic normality. Moreover, for MA(1) model, the p-values of the KS test are obtained as 0.582, 0.987, 0.981 for $n = 100,\ 1000\ 10000$, respectively. For

MA(2) model, the p-values of the KS test for $\{\sqrt{n}(\hat{\rho}_1^{(k)} - \rho_1); k = 1, \ldots, 100\}$ are obtained as 0.528, 0.507, 0.718 and that for $\{\sqrt{n}(\hat{\rho}_2^{(k)} - \rho_2); k = 1, \ldots, 100\}$ are obtained as 0.990, 0.799, 0.989 for $n = 100,\ 1000\ 10000$, respectively. As a result, it shows that we cannot reject the null hypothesis in all cases we investigated.

Table 1: MAE for circular distributions of MA(1) type

ϕ_1	α_1	n	MAE_1	ϕ_1	α_1	n	MAE_1
0.4	$\pi/4$	100	0.175	0.7	$\pi/4$	100	0.171
		300	0.103			300	0.108
		500	0.076			500	0.077
		1000	0.054			1000	0.055
	$\pi/2$	100	0.115		$\pi/2$	100	0.098
		300	0.648			300	0.060
		500	0.050			500	0.044
		1000	0.036			1000	0.032
	$3\pi/4$	100	0.106		$3\pi/4$	100	0.071
		300	0.059			300	0.038
		500	0.046			500	0.030
		1000	0.034			1000	0.022
-0.5	$\pi/4$	100	0.087				
		300	0.052				
		500	0.041				
		1000	0.029				
	$\pi/2$	100	0.113				
		300	0.065				
		500	0.049				
		1000	0.034				
	$3\pi/4$	100	0.180				
		300	0.103				
		500	0.077				
		1000	0.055				

Table 2: MAE for circular distributions of MA(2) type

(ϕ_1, ϕ_2)	(α_1, α_2)	n	MAE_1	MAE_2
(0.7,0.4)	$(\pi/4, \pi/2)$	100	0.081	0.215
		300	0.045	0.121
		500	0.036	0.099
		1000	0.026	0.069
	$(\pi/2, 3\pi/4)$	100	0.083	0.094
		300	0.046	0.055
		500	0.037	0.041
		1000	0.026	0.029
(1.0,0.7)	$(\pi/4, \pi/2)$	100	0.060	0.222
		300	0.035	0.129
		500	0.028	0.096
		1000	0.020	0.070
	$(\pi/2, 3\pi/4)$	100	0.064	0.070
		300	0.036	0.039
		500	0.027	0.032
		1000	0.019	0.021
(0.9,-0.3)	$(\pi/4, \pi/2)$	100	0.115	0.210
		300	0.066	0.125
		500	0.051	0.095
		1000	0.035	0.069
	$(\pi/2, 3\pi/4)$	100	0.111	0.155
		300	0.067	0.094
		500	0.052	0.075
		1000	0.036	0.052

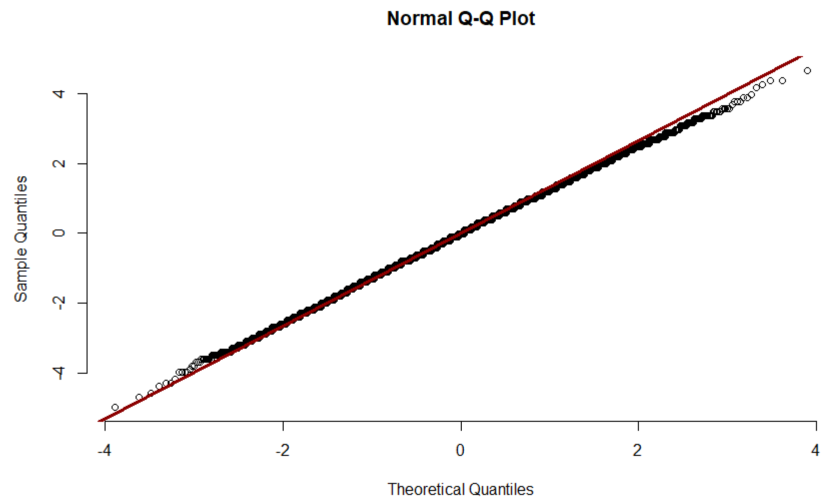

Figure 1: Q-Qplots of $\{\sqrt{n}(\hat{\rho}_1^{(k)} - \rho_1);\ k = 1, \dots, 10000\}$ for a circular distribution of MA(1) type with $\phi_1 = 0.7$ for $n = 1000$.

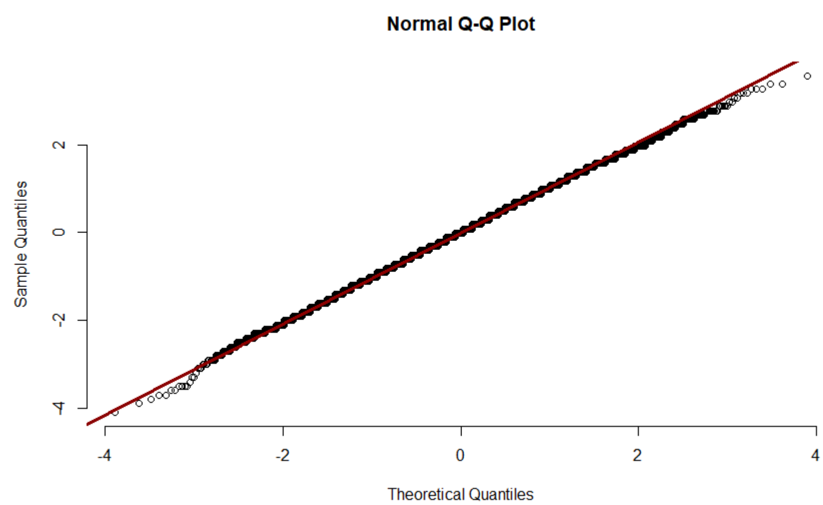

Figure 2: Q-Qplots of $\{\sqrt{n}(\hat{\rho}_1^{(k)} - \rho_1);\ k = 1, \dots, 10000\}$ for a circular distribution of MA(2) type $(\phi_1, \phi_2) = (0.7, 0.4)$ for $n = 1000$.

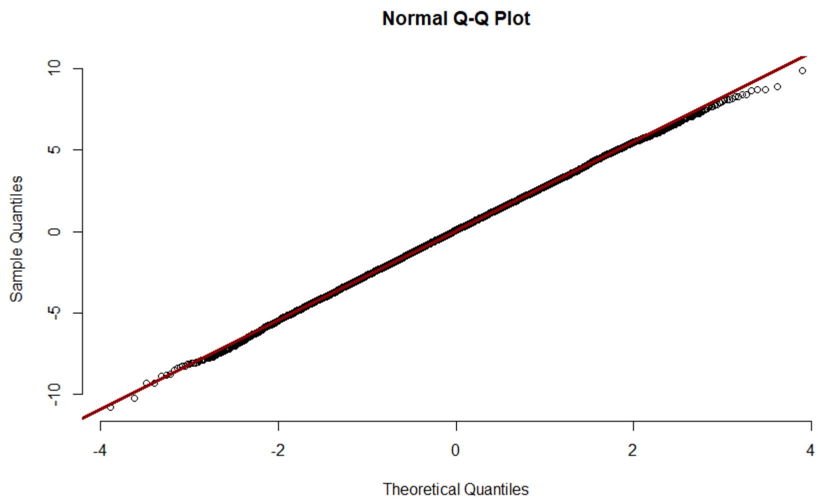

Figure 3: Q-Qplots of $\{\sqrt{n}(\hat{\rho}_2^{(k)} - \rho_2);\ k = 1, \ldots, 10000\}$ for a circular distribution of MA(2) type $(\phi_1, \phi_2) = (0.7, 0.4)$ for $n = 1000$.

5 Proof In this section, we provide the proofs of Theorems 2.1 and 3.1 and Proposition 2.1.

Proof of Theorem 2.1. First, we show the binary estimator is centered. For each $j \in \{1, \ldots, p\}$,

$$\mathrm{E}\{\sqrt{n}(\hat{\rho}_j - \rho_j)\} = \sqrt{n}2\pi(b^{j1}, \ldots, b^{jp}) \begin{pmatrix} \frac{1}{n}\sum_{k=1}^n \mathrm{E}{X_k}^1 - P(-\alpha_1 \leq \Theta_1 \leq \alpha_1) \\ \vdots \\ \frac{1}{n}\sum_{k=1}^n \mathrm{E}{X_k}^p - P(-\alpha_p \leq \Theta_1 \leq \alpha_p) \end{pmatrix} = 0.$$

Next, we evaluate the variance of estimator. For $i, j \in \{1, \ldots, p\}$,

$$\begin{aligned} &\mathrm{cum}\{\sqrt{n}(\hat{\rho}_i - \rho_i), \sqrt{n}(\hat{\rho}_j - \rho_j)\} \\ =&\frac{4\pi^2}{n} \sum_{s,k=1}^p b^{is} b^{jk} \sum_{v=1}^n \mathrm{cum}\{{X_v}^s, {X_v}^k\} \\ =&4\pi^2 \sum_{s,k=1}^p b^{is} b^{jk} \mathrm{cum}\{X_1^s, X_1^k\}. \end{aligned}$$

Finally, we elucidate the L-th order cumulant $(L \geq 3)$ of the binary estimator is of order $O(n^{-L/2+1})$. For $i_1, \ldots, i_L \in \{1, \ldots, p\}$,

$$\begin{aligned} &\mathrm{cum}\{\sqrt{n}(\hat{\rho}_{i_1} - \rho_{i_1}), \ldots, \sqrt{n}(\hat{\rho}_{i_L} - \rho_{i_L})\} \\ =&n^{L/2}(2\pi)^L \sum_{s_1, \ldots, s_L = 1}^p b^{i_1 s_1} \cdots b^{i_L s_l} \mathrm{cum}\{\frac{1}{n}\sum_{k=1}^n {X_k}^{s_1}, \ldots, \frac{1}{n}\sum_{k=1}^n {X_k}^{s_L}\} \\ =&n^{-L/2+1}(2\pi)^L \sum_{s_1, \ldots, s_L = 1}^p b^{i_1 s_1} \cdots b^{i_L s_l} \mathrm{cum}\{X_1^{s_1}, \ldots, X_1^{s_L}\} \\ =&O(n^{-L/2+1}), \end{aligned}$$

thus, we have the desired result. □

Proof of Proposition 2.1. It is sufficient to show the Fisher information $\mathcal{I}$, defined by

$$\mathcal{I}(\rho_1) = \int_{-\pi}^{\pi} \left(\frac{\partial}{\partial \rho_1} \log p(\theta) \right)^2 p(\theta) \mathrm{d}\theta,$$

becomes the following

$$\mathcal{I}(\rho_1) = \begin{cases} \frac{1}{2} & (\rho_1 = 0), \\ \frac{1}{\rho_1^2} \left(\frac{1}{\sqrt{1-\rho_1^2}} - 1 \right) & (0 < |\rho_1| < 1), \\ \infty & (\rho_1 = \pm 1). \end{cases}$$

First, for $\rho_1 = 0$, by a straightforward calculation. Second, the residue theorem yields the assertion when ρ_1 satisfies $0 < |\rho_1| < 1$. Third, for $\rho_1 = \pm 1$, it is easy to see the integral diverges. □

Proof of Theorem 3.1. For any $a_j (< \pi - \beta)$, $j = 1, \ldots, p$, we have

$$\int_{-a_j}^{a_j} q(\theta) \mathrm{d}\theta = \int_{-a_j}^{a_j} p(\theta) \mathrm{d}\theta,$$

from which the statement follows. □

6 Acknowledgement The author Y.G. is grateful to the referee for instructive comments and kindness. Y.G. also thanks Professor Masanobu Taniguchi and Doctor Yan Liu for their encouragements and comments. Y.G.'s research was supported by Grant-in-Aid for JSPS Research Fellow Grant Number JP201920060.

References

Bagnall, A. and Janacek, G. (2005). Clustering time series with clipped data. *Mach. Learn.* **58** 151–178.

Dette, H., Hallin, M., Kley, T., and Volgushev, S. (2015). Of copulas, quantiles, ranks and spectra: An L_1-approach to spectral analysis. *Bernoulli* **21** 781–831.

Fisher, N. I., Lewis, T., and Embleton, B. J. (1993). *Statistical Analysis of Spherical Data.* Cambridge university press.

Fisher, R. A. (1953). Dispersion on a sphere. *Proc. Roy. Soc. Lond. Ser. A* **217** 295–305.

Fokianos, K. and Kedem, B. (2003). Regression theory for categorical time series. *Statist. Sci.* **18** 357–376.

Goto, Y. and Taniguchi, M. Discriminant analysis based on binary time series. To appear in Metrika, 10.1007/s00184-019-00746-1.

Goto, Y. and Taniguchi, M. (2019). Robustness of zero crossing estimator. *J. Time Ser. Anal.* **40** 815–830.

Hosoya, Y. and Taniguchi, M. (1982). A central limit theorem for stationary processes and the parameter estimation of linear processes. *Ann. Statist.* **10** 132–153.

Kato, S. and Jones, M. (2013). An extended family of circular distributions related to wrapped Cauchy distributions via Brownian motion. *Bernoulli* **19** 154–171.

Kedem, B. (1980). *Binary Time Series.* New York: Dekker.

Kedem, B. (1994). *Time Series Analysis by Higher Order Crossings.* New York: IEEE press.

Kedem, B. and Li, T. (1989). Higher order crossings from a parametric family of linear filters. tech. rep. tr-89-47, Dept. Mathematics, Unvi. Maryland, College Park.

Li, T. H. (2014). *Time Series with Mixed Spectra*. CRC Press.

Lomnicki, Z. and Zaremba, S. (1955). Some applications of zero-one processes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **17** 243–255.

Mardia, K. V. (1975). Statistics of directional data. *J. Roy. Stat. Soc. Ser. B* **37** 349–371.

Mardia, K. V. and Jupp, P. E. (2000). *Directional Statistics* **494**. John Wiley & Sons.

McNeil, D. (1967). Estimating the covariance and spectral density functions from a clipped stationary time series. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **29** 180–195.

Panagiotakis, C. and Tziritas, G. (2005). A speech/music discriminator based on rms and zero-crossings. *IEEE Trans. Multimedia* **7** 155–166.

Petrantonakis, P. C. and Hadjileontiadis, L. J. (2010). Emotion recognition from EEG using higher order crossings. *IEEE Trans. Inform. Tech. Bio* **14** 186–197.

Rice, S. O. (1944). Mathematical analysis of random noise. *Bell Syst. Tech. J.* **23** 282–332.

Robert, C. P., Casella, G., and Casella, G. (2010). *Introducing Monte Carlo Methods with R*. Springer.

Watson, G. S. (1983). *Statistics on Spheres*. Wiley-Interscience.