Sains Malaysiana 41(6)(2012): 769-778

On Discordance Test for the Wrapped Normal Data

(Ujian Tak Sejajar bagi Data Normal Balutan)

ADZHAR RAMBLI, SAFWATI IBRAHIM, MOHD IKHWAN ABDULLAH, ABDUL GHAPOR HUSSIN & IBRAHIM MOHAMED*

ABSTRACT

This paper focuses on detecting outliers in the circular data which follow the wrapped normal distribution. We considered four discordance tests based on M, C, D and A statistics. The cut-off points of the four tests were obtained and the performance of the detection procedures was studied via simulations. In general, we showed that the discordance test based on the A statistic outperforms the other tests in all cases. For illustration, the city of Kuantan wind direction data set was considered.

Keywords: A statistics; circular; discordance; outlier; wrapped normal distribution

ABSTRAK

Kajian ini lebih tertumpu kepada pengesanan nilai tersisih yang wujud di dalam data bulatan terutamanya data bertaburan normal balutan. Empat ujian sejajar berdasarkan statistik seperti M, C, D dan A dipertimbangkan. Titik potongan dan prestasi prosedur bagi mengesan nilai tersisih diperoleh melalui kajian simulasi. Secara amnya, dibuktikan bahawa ujian tak sejajar berdasarkan statistik A adalah lebih baik berbanding statistik yang lain untuk kesemua kes. Sebgaai contoh, data arah angin bandar Kuantan telah digunakan sebagai contoh.

Kata kunci: Bulatan; nilai tersisih; statistik A; taburan normal balutan

INTRODUCTION

Circular data can be visualized as being distributed on the circumference of a unit circle in the range of 0 to 2π radian. The data are commonly found in many scientific fields such as meteorology and biology where researchers are interested in studying direction of wind and direction of movement of animals, respectively. Due to the bounded range property of circular variables, special methods such as circular descriptive statistics, circular plots and goodness of fit tests are required to describe and model such data.

Several authors have comprehensively discussed the circular distributions including Jammalamadaka and SenGupta (2001), Mardia (1972) and Fisher (1993). Various distributions are available for circular data, for example, uniform distribution, wrapped Cauchy distribution, wrapped normal distribution, cardioid distribution, and others. Jammalamadaka and SenGupta (2001) reviewed the wrapped α stable distribution with the wrapped Cauchy and the wrapped normal distributions as the special cases. On the other hand, several bivariate circular distributions exist, such as the bivariate von Mises distribution, wrapped bivariate normal distribution and circular-linear distribution. The von Mises (*VM*) distribution (also known as the circular normal distribution) is the most commonly used and is a continuous probability distribution on a circle. The von Mises distribution may be thought of as a close approximation to the wrapped normal distribution, which is the circular analogue of the normal distribution.

As in the linear case, the existence of outliers in circular data is expected to affect the estimation of parameters and weaken the accuracy of forecast. Thus, it is very important that methods of identifying outliers in circular data are developed for proper handling of the data. Graphical and numerical methods are the most common tools used in investigating the existence of outliers in circular data. We consider four discordance tests to detect possible outliers in the circular data based on the *M, C, D* and *A* statistics. The later is proposed by Abuzaid et al. (2009) which has been shown to perform better than other methods for data from the von Mises distributions, except for small sample size. In this paper, we apply the tests on data from wrapped normal (*WN*) distributions.

With that view in mind, this paper is organized as follows: The following section describes properties of the wrapped normal distribution. This is followed by reviews on several discordance tests to detect the existence of outliers in circular univariate data. In the next section, we obtain the cut-off points and study the performance of each statistic by simulation studies to detect outliers in circular data that come from the wrapped normal distribution. We then apply the statistics on the real data set obtained from the Malaysian Meteorological Service Department in the last section.

WRAPPED NORMAL DISTRIBUTION

Jammalamadaka and SenGupta (2001) discussed the general wrapped α -stable distribution which is constructed by using the characteristic function of the α-stable of a real line. The characteristic function as given by Lukasc (1970) is:

$$\Psi(t)=\begin{cases}\exp\left\{-\tau^{\alpha}|t|^{\alpha}\left[1-i\beta\operatorname{sgn}(t)\tan\frac{\alpha\pi}{2}\right]+i\mu t\right\}, & if\,\alpha\in(0,1)\cup(1,2],\\ \exp\left\{-\tau|t|+i\mu t\right\}, & if\,\alpha=1,\end{cases}$$

where $\tau\geq 0, |\beta|\leq 1, 0<\alpha\leq 2$ while μ is a real number. The density function of a wrapped α -stable random variable for θ ∈ [0, 2π) is given by:

$$f(\theta)=\frac{1}{2\pi}+\frac{1}{\pi}\sum_{k=1}^{\infty}\exp\left\{-\tau^{\alpha}k^{\alpha}\right\}cos\left\{k(\theta-\mu)-\tau^{\alpha}k^{\alpha}\beta\tan\frac{\alpha\pi}{2}\right\},$$

when $\alpha\in(0,1)\cup(1,2]$, with μ conveniently redefined as μ (mod 2π). Note that although there is generally no closed form expression for the density of an α -stable distribution on the real line, we are able to write such density for the wrapped case, at least as an infinite series.

The particular case corresponding to $\beta = 0$ gives us the symmetric wrapped stable (SWS) family of circular densities, which we will simply refer to as wrapped stable (WS), given by

$$f(\theta)=\frac{1}{2\pi}+\frac{1}{\pi}\sum_{k=1}^{\infty}\rho^{k^{\alpha}}\quad cos\left\{k(\theta-\mu)\right\},$$

where $\rho = \exp(-\tau^{\alpha})$. We shall denote such distributions as $WS(\alpha, \rho, \mu)$. The special case with $\alpha = 2$ and $\beta = 0$ gives us the wrapped normal density with $\rho=exp\left(-\frac{\sigma^2}{2}\right)$. When $\alpha = 1$ and $\beta = 0$, it gives us the wrapped Cauchy density with $\rho = \exp(-\tau)$.

A wrapped normal distribution is obtained by wrapping a normal distribution around a unit circle. The normal distribution is denoted by $N(\mu_L, \sigma_L^2)$ where μ_L is the mean and σ_L^2 is the variance while the *WN* distribution is denoted by $WN(\mu, \rho)$, where μ is the mean direction and ρ is the measure of concentration parameter. Its probability distribution function is given by:

$$f(\theta)=\frac{1}{\sigma\sqrt{2\pi}}\sum_{k=1\infty}^{\infty}\exp\left[\frac{-(\theta-\mu-2k\pi)^2}{2\sigma^2}\right], \tag{1}$$

where σ^2 is the circular variance. From Whittaker and Watson (1944), an alternate and more useful representation of this density is:

$$f(\theta)=\frac{1}{2\pi}\left(1+2\sum_{k=1}^{\infty}\rho^{k^2}\cos k(\theta-\mu)\right)\quad 0\leq\theta<2\pi,\quad 0\leq\rho\leq 1.$$

The distribution is unimodal and symmetric about the value $\theta = \mu$. Unlike the von Mises distribution, the *WN* distribution possesses the additive property, that is, the convolution of two WN variables is also *WN*. Specifically, if $\theta_1 \sim WN(\mu_1, \rho_1)$ and $\theta_2 \sim WN(\mu_2, \rho_2)$ are independent, then $\theta_1 + \theta_2 \sim WN(\mu_1 + \mu_2, \rho_1\rho_2)$ (Jammalamadaka and SenGupta 2001). We study the case when the sample data follows the wrapped normal distribution with the mean direction μ and the concentration parameter κ.

In deciding whether a circular data set follows the von Mises (*VM*) distribution or the wrapped normal (*WN*) distribution, Kent (1976) highlighted the fact that both distributions are hardly distinguishable for κ < 0.1 or κ > 10. In this case, Kendall (1974) noted that for any analytical, computational and statistical purposes, the *WN* distribution is more convenient to be used in some cases and the *VM* distribution in other cases. Collet and Lewis (1981) concluded that a minimum sample size required in order to distinguish the two distributions is 200 via the classical discriminant approach.

DISCORDANCE TESTS IN DETECTING OUTLIERS

Suppose $\theta_1, \theta_2, \ldots, \theta_n$ are (i.i.d) circular observations located on the circumference of a unit circle. We consider four discordance tests based on the *C, D, M,* and *A* statistics to identify outliers in a univariate circular sample from the WN distribution.

C statistic. The mean resultant length of circular data set is given by $\bar{R}=\frac{R}{n}$, where $R=\sqrt{C^2+S^2}$ such that $C=\sum_{i=1}^{n}\cos\theta_i$ and $S=\sum_{i=1}^{n}\sin\theta_i$. By omitting the *i*th observation, the mean resultant length is given by $\bar{R}_{(-i)}=\frac{R_{(-i)}}{n-1}$. Collet (1980) proposed the test statistics as:

$$C=max_i\left\{\frac{\bar{R}_{(-i)}-\bar{R}}{\bar{R}}\right\}. \tag{2}$$

Values of *C* statistic will then be compared with the cut-off points for the corresponding sample size *n* and estimated concentration parameter κ. If *C* is larger than the cut-off point, we reject the null hypothesis so that the *i*th observation is identified as an outlier.

D statistic. The *D* statistic uses the relative arc length based on the ordered observation of a circular sample $\theta_{(1)}, \theta_{(2)}, \ldots, \theta_{(n)}$. Let T_i be the arc length between consecutive observations given by $T_i = \theta_{(i+1)} - \theta_{(i)}$, $i = 1, 2, \ldots, n$ and $T_n = 2\pi - \theta_{(n)} + \theta_{(1)}$. Define $D_i=\frac{T_i}{T_{i-1}}, i=1,2,\ldots,n$ and $T_0\equiv T_n$.

Let $D_k = \frac{T_k}{T_{k-1}}$ corresponds to the greatest arc containing a single observation θ_k. Note that D_k is two tailed. Collet (1980) suggested working in terms of:

$$D = \min (D_k, D_k^{-1}) \tag{3}$$

where $0 < D < 1$. The observation θ_k can be considered as an outlier if the value of D is larger than the cut-off point.

M statistic Mardia (1975) suggested a statistic of discordancy which is given by $M' = min_i \left\{ \frac{n-1-R_{(-i)}}{n-R} \right\}$. Later, Collett (1980) reformulated the statistic in terms of:

$$M = 1 - M' = max_i \left\{ \frac{R_{(-i)} - R + 1}{n-R} \right\} = \frac{R_q - R + 1}{n-R}, \tag{4}$$

where $R_q = \max_i\{R_{(-i)}\}$. He stated the asymptotic distribution of the M statistic for large values of κ. As the value of κ increases, the von Mises distribution will be approximated by a standard normal distribution. On the other hand, the M statistic can be approximated by $\frac{n(b^*)^2}{n-1}$, where $b^* = max_i \frac{|x_i - \bar{x}|}{\sum (x_i - \bar{x})^2}$ is the test statistic used to identify discordancy in normal data. Percentage points for are given in Pearson and Hartley (1966).

A statistic Rao (1969) defined the circular distance between θ_i and θ_j as:

$$d_{ij} = 1 \cos(\theta_i - \theta_j),$$

where d_{ij} is a monotone increasing function of $(\theta_i - \theta_j)$ and $d_{ij} \in [0,2]$. The summation of all circular distances of the point of interest θ_j to all other points is given by:

$$D_j = \sum_{i=1}^{n} \left(1 - cos\left(\theta_i - \theta_j\right)\right), i = 1,2,\ldots,n.$$

If the observation θ_j is an outlier, then the value of D_j will increase. Thus, the average circular distance given by $\frac{D_j}{n-1}$ can be used to identify possible outliers in the circular sample. Abuzaid et al. (2009) proposed the A statistic as:

$$A = max_j \left\{ \frac{D_j}{2(n-1)} \right\}, j = 1,2,\ldots,n, \tag{5}$$

where $A \in [0,1]$ is a linear measure. The average circular distance is divided by 2 in order to standardize the values of statistic A. The proposed statistic is based on the relative decrease in the summation of circular distances by omitting the point of interest θ_j.

CUT-OFF POINTS OF THE DISCORDANCE TESTS

The interest is on the use of the discordance tests to detect outliers in data generated from the wrapped normal distribution. Firstly, we have to obtain the cut-off points for each test. Thus, we design a simulation study in SPlus statistical package to find the percentage points of the null distribution of no outliers in the circular data set. We consider eleven values of measure of concentration parameter in the range of 0.1 to 0.975 and different sample sizes from 5 to 150. For each combination of n and ρ, we generate sample from $WN(\mu=0, \rho)$. All the test statistics in each generated random sample are calculated using statistics (2)-(5), respectively. We wish to estimate the percentage points of the discordance tests at the 10%, 5% and 1% upper percentiles when no outlier presents in the sample.

Tables 1 to 4 show the cut-off points of the four tests. Two main results are observed. Firstly, as the measure of concentration parameter increases, the cut-off points decreases for the three levels of percentiles. This is expected as the circular data are more concentrated with larger ρ resulting in a smaller difference between two largest values of the statistics. Secondly, as the sample size increases, the value of the cut-off point decreases. Again, this should be true as the sample size increases, the distance between the circular observations in circular plot become smaller.

PERFORMANCE OF THE DISCORDANCE TESTS

Collett (1980) applied selected measures to test the performances of several statistics to detect an outlier in circular sample. Here, we used similar measures to compare the performance of the tests. David (1970) and Barnett and Lewis (1978) stated that a good test should have: (1) a high power function; (2) a high probability of identifying a contaminating value as an outlier when it is in fact an extreme value, where an extreme value is defined as a point with the maximum circular deviation; and (3) a low probability of wrongly identifying a good observation as discordant.

Let P1 = 1 − β be the power function where β is the Type-II error; P3 the probability that the contaminant point is an extreme point and is identified as discordant; and P5 the probability that the contaminant point is identified as discordant given that it is an extreme point. A good test is expected to have (1) high P1, (2) high P5, and (3) low P1 − P3.

To study the performance of all discordance tests, we use 2000 simulations for different sizes of n and ρ. The samples were generated in SPlus Statistical Package in

TABLE 1. Table of cut-off points for the test based on the C statistic

n	Level of percentile	ρ 0.1	0.2	0.4	0.6	0.8	0.9	0.95	0.975
5	10%	1.711	1.577	1.171	0.686	0.241	0.109	0.050	0.024
	5%	2.379	2.195	1.581	0.866	0.313	0.133	0.062	0.030
	1%	4.867	4.306	2.776	1.622	0.506	0.216	0.092	0.044
10	10%	1.114	0.990	0.526	0.303	0.136	0.063	0.030	0.015
	5%	1.542	1.354	0.666	0.339	0.165	0.075	0.038	0.018
	1%	3.668	2.664	1.343	0.465	0.211	0.104	0.050	0.026
30	10%	0.551	0.373	0.156	0.097	0.052	0.026	0.013	0.006
	5%	0.739	0.496	0.177	0.102	0.058	0.030	0.014	0.008
	1%	1.398	1.132	0.233	0.111	0.070	0.037	0.019	0.010
50	10%	0.365	0.200	0.087	0.057	0.034	0.017	0.008	0.004
	5%	0.528	0.257	0.095	0.059	0.037	0.019	0.009	0.005
	1%	1.398	0.535	0.111	0.064	0.043	0.023	0.012	0.007
70	10%	0.286	0.134	0.060	0.040	0.024	0.013	0.007	0.003
	5%	0.409	0.162	0.065	0.041	0.026	0.014	0.007	0.004
	1%	0.954	0.309	0.073	0.043	0.030	0.018	0.009	0.005
90	10%	0.222	0.100	0.045	0.031	0.019	0.010	0.005	0.003
	5%	0.321	0.122	0.048	0.032	0.021	0.012	0.006	0.003
	1%	0.694	0.201	0.054	0.034	0.023	0.014	0.007	0.004
100	10%	0.202	0.089	0.040	0.028	0.018	0.010	0.005	0.002
	5%	0.283	0.107	0.042	0.029	0.019	0.011	0.005	0.003
	1%	0.544	0.179	0.046	0.030	0.021	0.013	0.007	0.003
150	10%	0.130	0.056	0.026	0.019	0.012	0.007	0.003	0.002
	5%	0.177	0.065	0.027	0.019	0.013	0.007	0.004	0.002
	1%	0.378	0.093	0.030	0.019	0.014	0.009	0.005	0.002

TABLE 2. Table of cut-off points for the test based on the M statistic

n	Level of percentile	ρ 0.1	0.2	0.4	0.6	0.8	0.9	0.95	0.975
5	10%	0.762	0.785	0.808	0.843	0.865	0.866	0.862	0.874
	5%	0.838	0.851	0.878	0.896	0.914	0.913	0.915	0.920
	1%	0.934	0.932	0.954	0.967	0.973	0.975	0.972	0.971
10	10%	0.349	0.374	0.443	0.506	0.552	0.581	0.574	0.587
	5%	0.389	0.415	0.498	0.560	0.603	0.645	0.634	0.644
	1%	0.487	0.519	0.602	0.692	0.709	0.744	0.742	0.744
30	10%	0.096	0.107	0.141	0.180	0.228	0.253	0.256	0.264
	5%	0.102	0.114	0.153	0.195	0.246	0.282	0.289	0.297
	1%	0.116	0.130	0.178	0.230	0.318	0.346	0.355	0.379
50	10%	0.054	0.060	0.080	0.110	0.150	0.163	0.169	0.178
	5%	0.057	0.063	0.085	0.116	0.164	0.179	0.189	0.198

cont.

TABLE 2 (cont.)

	1%	0.061	0.069	0.094	0.128	0.192	0.220	0.228	0.255
70	10%	0.037	0.042	0.056	0.079	0.108	0.125	0.131	0.133
	5%	0.038	0.043	0.059	0.083	0.118	0.138	0.147	0.149
	1%	0.041	0.047	0.063	0.093	0.134	0.169	0.178	0.183
90	10%	0.028	0.032	0.043	0.061	0.087	0.101	0.106	0.111
	5%	0.029	0.033	0.045	0.064	0.093	0.112	0.119	0.123
	1%	0.031	0.036	0.047	0.070	0.105	0.134	0.143	0.145
100	10%	0.025	0.029	0.038	0.055	0.080	0.093	0.098	0.103
	5%	0.026	0.029	0.040	0.057	0.085	0.103	0.107	0.112
	1%	0.028	0.031	0.043	0.062	0.096	0.124	0.135	0.138
150	10%	0.016	0.019	0.025	0.037	0.055	0.065	0.070	0.072
	5%	0.017	0.019	0.026	0.038	0.058	0.070	0.077	0.080
	1%	0.018	0.020	0.028	0.041	0.066	0.081	0.093	0.093

TABLE 3. Table of cut-off points for the test based on the the D statistic

n	Level of percentile					ρ			
		0.1	0.2	0.4	0.6	0.8	0.9	0.95	0.975
5	10%	0.872	0.857	0.811	0.587	0.277	0.171	0.105	0.071
	5%	0.937	0.931	0.898	0.737	0.337	0.198	0.127	0.082
	1%	0.990	0.984	0.984	0.909	0.487	0.275	0.169	0.109
10	10%	0.857	0.872	0.814	0.611	0.268	0.150	0.094	0.060
	5%	0.924	0.933	0.911	0.762	0.323	0.185	0.114	0.072
	1%	0.980	0.986	0.980	0.930	0.504	0.246	0.150	0.099
30	10%	0.872	0.875	0.814	0.695	0.259	0.132	0.078	0.051
	5%	0.937	0.926	0.902	0.847	0.335	0.165	0.097	0.061
	1%	0.988	0.985	0.981	0.966	0.494	0.226	0.133	0.093
50	10%	0.873	0.864	0.849	0.750	0.279	0.129	0.076	0.048
	5%	0.938	0.939	0.923	0.857	0.358	0.159	0.089	0.059
	1%	0.981	0.988	0.986	0.965	0.582	0.214	0.122	0.081
70	10%	0.873	0.869	0.861	0.758	0.252	0.125	0.073	0.046
	5%	0.931	0.939	0.934	0.876	0.334	0.152	0.089	0.055
	1%	0.989	0.985	0.985	0.965	0.555	0.218	0.127	0.075
90	10%	0.857	0.857	0.845	0.791	0.259	0.130	0.072	0.046
	5%	0.933	0.930	0.923	0.893	0.324	0.156	0.089	0.056
	1%	0.989	0.979	0.991	0.968	0.517	0.225	0.124	0.077
100	10%	0.864	0.861	0.870	0.779	0.264	0.127	0.071	0.045
	5%	0.930	0.930	0.930	0.882	0.341	0.160	0.084	0.054
	1%	0.984	0.984	0.980	0.974	0.541	0.219	0.120	0.075
150	10%	0.861	0.869	0.863	0.819	0.269	0.118	0.070	0.044
	5%	0.922	0.928	0.932	0.905	0.353	0.142	0.086	0.053
	1%	0.981	0.986	0.985	0.982	0.550	0.200	0.123	0.070

TABLE 4. Table of cut-off points for the test based on the A statistic

n	Level of percentile	ρ 0.1	0.2	0.4	0.6	0.8	0.9	0.95	0.975
5	10%	0.924	0.928	0.920	0.892	0.726	0.549	0.395	0.280
	5%	0.944	0.948	0.946	0.927	0.783	0.592	0.438	0.311
	1%	0.971	0.978	0.979	0.962	0.879	0.705	0.509	0.369
10	10%	0.859	0.872	0.893	0.888	0.751	0.569	0.419	0.296
	5%	0.880	0.890	0.914	0.911	0.802	0.612	0.454	0.322
	1%	0.913	0.927	0.949	0.943	0.869	0.691	0.515	0.386
30	10%	0.787	0.813	0.865	0.889	0.796	0.616	0.449	0.326
	5%	0.803	0.827	0.877	0.901	0.826	0.657	0.472	0.351
	1%	0.829	0.852	0.897	0.925	0.890	0.722	0.542	0.405
50	10%	0.760	0.790	0.851	0.889	0.826	0.638	0.468	0.339
	5%	0.774	0.802	0.860	0.898	0.860	0.671	0.494	0.366
	1%	0.793	0.824	0.878	0.914	0.909	0.729	0.548	0.417
70	10%	0.743	0.778	0.845	0.891	0.828	0.654	0.488	0.347
	5%	0.754	0.787	0.855	0.899	0.858	0.689	0.513	0.371
	1%	0.774	0.809	0.867	0.914	0.906	0.756	0.560	0.413
90	10%	0.735	0.771	0.841	0.892	0.838	0.667	0.492	0.361
	5%	0.744	0.781	0.848	0.899	0.865	0.700	0.517	0.380
	1%	0.767	0.802	0.861	0.910	0.906	0.759	0.574	0.421
100	10%	0.734	0.770	0.840	0.890	0.846	0.674	0.500	0.361
	5%	0.744	0.779	0.846	0.897	0.871	0.706	0.527	0.382
	1%	0.761	0.794	0.862	0.908	0.913	0.765	0.583	0.425
150	10%	0.720	0.760	0.833	0.892	0.857	0.682	0.510	0.368
	5%	0.729	0.768	0.839	0.897	0.880	0.708	0.536	0.386
	1%	0.746	0.779	0.852	0.908	0.926	0.767	0.595	0.427

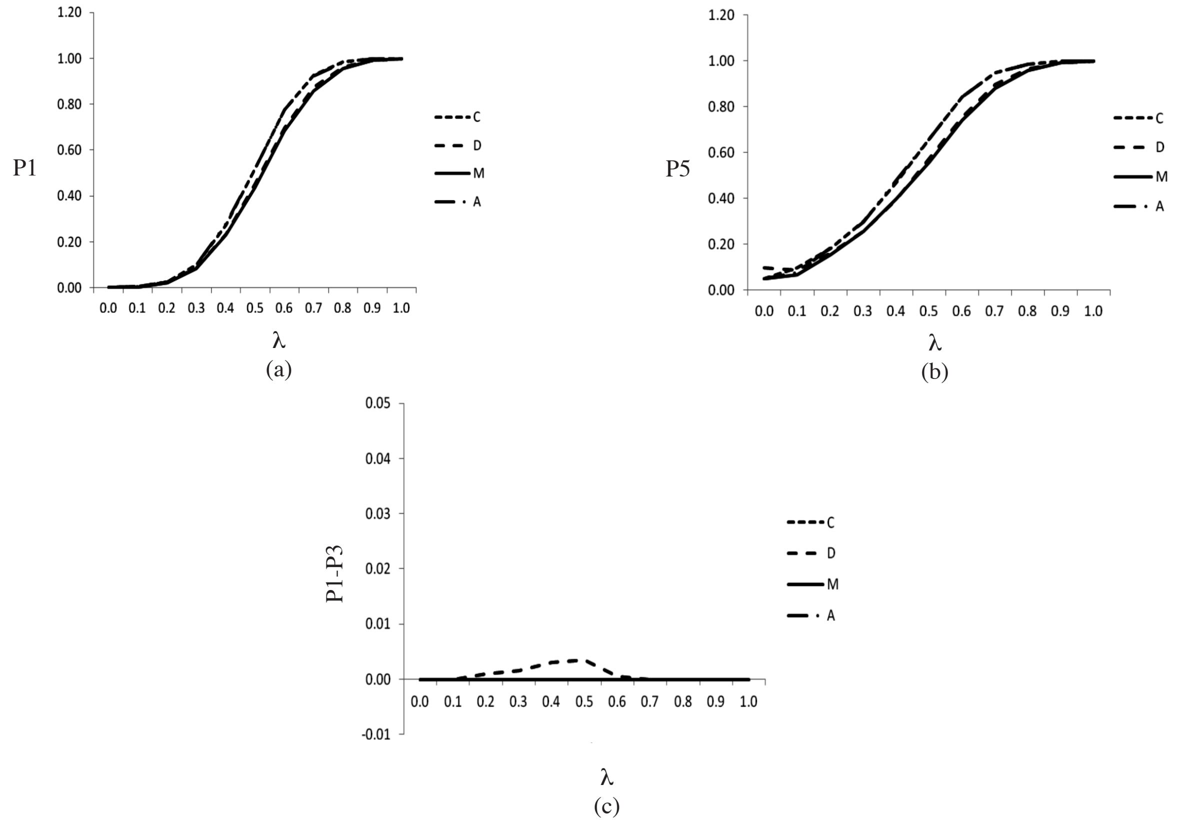

FIGURE 1. Performance of the statistics for n = 50 and ρ = 0.90

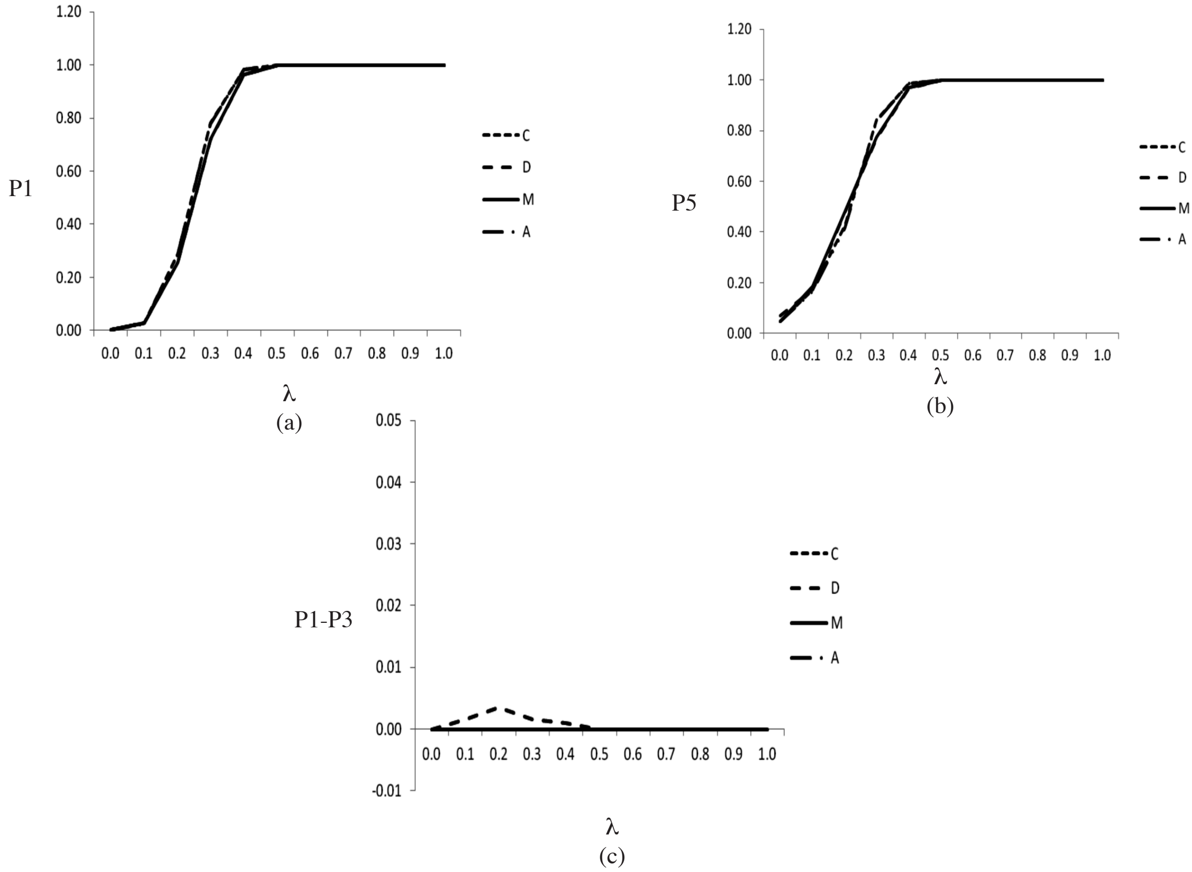

FIGURE 2. Performance of the statistics for n = 50 and ρ = 0.975

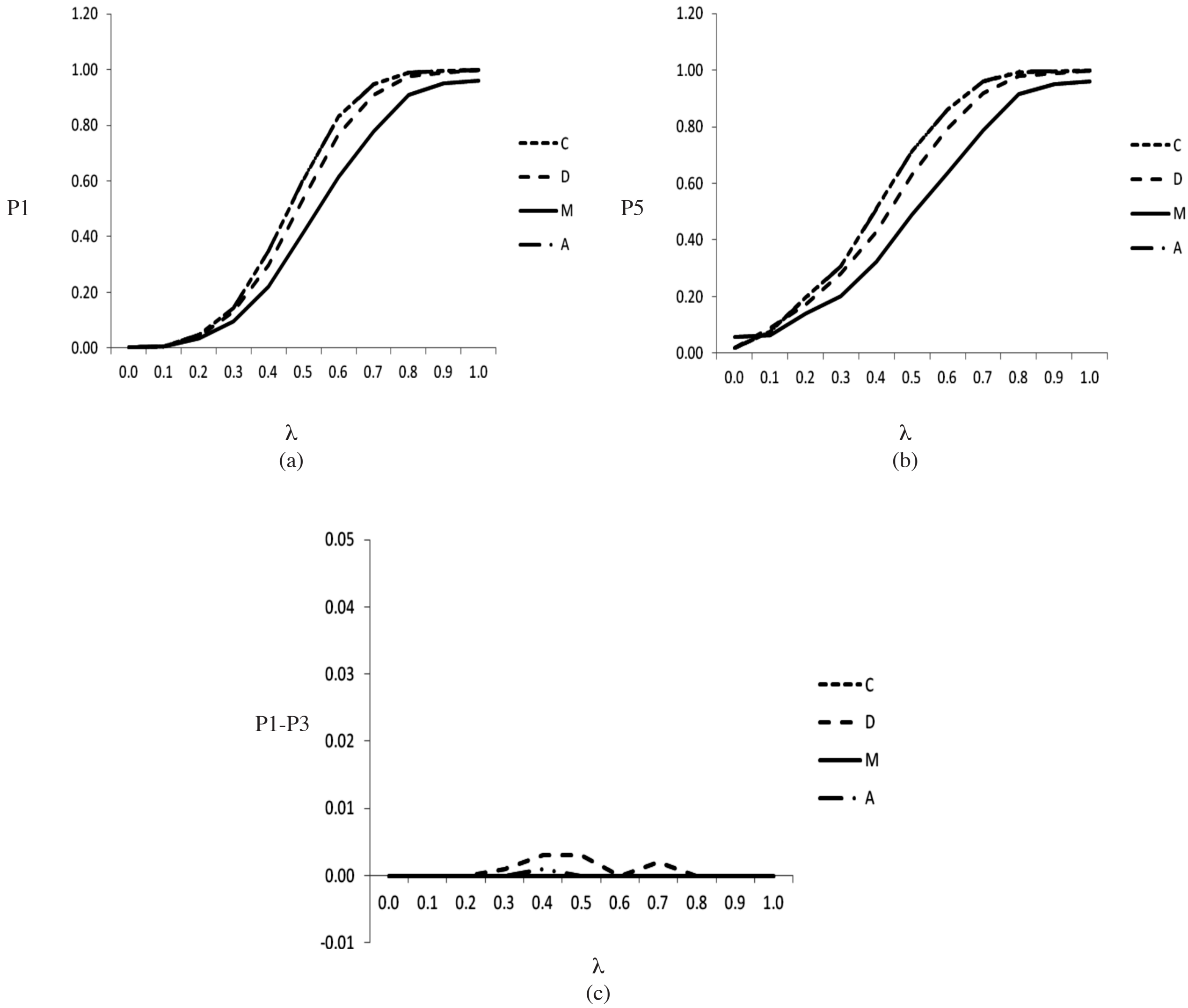

FIGURE 3. Performance of the statistics for n = 20 and ρ = 0.90

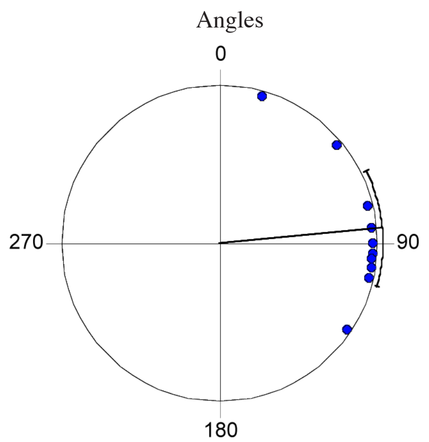

FIGURE 4. Circular plot of Kuantan wind data

such a way that ($n-1$) of the observations come from the $WN(0,\rho)$ and one observation from $WN(\alpha, \lambda\pi,\rho)$, where λ is the degree of contamination and $0 \leq \lambda \leq 1$. The *C*, *D*, *M* and *A* statistics in each random sample are then calculated.

Figures 1 to 3 illustrate the power of performance of the tests for different cases. Three main results were observed. Firstly, in Figure 1(a) and (b), for $n = 50$ and $\rho = 0.90$, the test based on the *A* statistic performs better than the others for all contamination levels λ since the P1 and P5 curves were always greater than the others. As in Figure 1(c), the four tests have low P1-P3 since the curves are almost 0. Secondly, from Figures 1(a) and 2(a), P1 reached the value 1 when $\lambda = 0.8$ and $\lambda = 0.6$, respectively. Similar results are observed for P5 as shown in Figures 1(b) and 2(b). These suggest that as ρ gets larger, the four tests showed better performance of detecting outliers at lower contamination level. Thirdly, from Figures 1(a) and 3(a), when *n* becomes smaller, the test based on the *A* statistic performs better than that of the *C* and *D* statistics, but performed much better than that of the *M* statistic. Similar results are observed for P5 as shown in Figure 1(b) and 3(b).

The four discordance methods have been investigated for the case when the data come from the von Mises distribution by Abuzaid et al. (2009). The cut-off points for tests based on the *C* and *D* statistics can be obtained from Collet (1980) while that of the *M* and *A* statistics are available in Mardia (1975) and Abuzaid et al. (2009), respectively. In summary, the results for the *VM* and *WN* are similar except for the case when *n* is small. In this particular case, the test based on the *A* statistic performs better than others in terms of P1 and P5 for the *WN* distribution but the test based on the *M* statistic performs better for the *VM* distribution.

TABLE 5. Kuantan wind direction data

Year	Mean Surface Wind Direction (radian)
1999	0.28707
2000	1.46071
2001	0.87509
2002	1.64563
2003	1.56786
2004	1.33478
2005	1.80266
2006	2.15736
2007	1.73430
2008	1.67275

TABLE 6. Descriptive statistics

Variable	Angles
Mean Vector (μ)	84.65°
Length of Mean Vector (r)	0.88
Concentration	3.33
Circular Variance	0.12
Circular Standard Deviation	28.45°
Standard Error of Mean	10.57°

TABLE 7. Results based on *C*, *M*, *D* and *A* statistics

Test	Test value	Cut-off point	Decision
C	0.07	0.08	Not an outlier
M	0.59	0.64	Not an outlier
D	0.13	0.19	Not an outlier
A	0.60	0.61	Not an outlier

Application

We considered the Kuantan wind direction data measured in unit radian from the year 1999 to 2008 as shown in Table 5 obtained from the Malaysian Meteorological Services Department. Table 6 gives the values of circular descriptive statistics for the data. The mean direction μ is 84.65° and the concentration parameter for this data is 3.33. We can conclude that the data sets are concentrated in the east direction. The circular plot of the data is given in Figure 4. We notice that there is one observation located a bit separated from the rest. Here, we have n =10 and ρ = 0.88. Table 7 gives the value of the test statistics, the cut-off point for n =10 and ρ = 0.9, as well as the decision for each statistic. It can be seen that the four tests do not identify the outlying observation as an outlier. Note, however, that the values of test statistics are very close to their respective cut-off points. Thus, it warrants further investigation on the observation.

Conclusion

In this paper, we have reviewed four discordance tests to identify the existence of outliers in circular data. The cut-off points for tests based on the C, M, D and A statistics for the wrapped normal distribution are obtained via simulation studies. We have compared the performance of the tests for the *VM* and the *WN* distributions. In general, the test based on the A statistic outperforms the other tests. As an illustration, we apply the statistics to identify the existence of outliers on the Kuantan wind direction data.

Acknowledgment

The authors wish to thank the editors and the referee for their valuable comments and suggestions.

References

Abuzaid, A.H., Mohamed, I.B. & Hussin, A.G. 2009. A new test of discordancy in circular data. *Communication in Statistics-Simulation and Computation* 38(4): 682-691.

Barnett, V. & Lewis, T. 1984. *Outliers in Statistical Data*. New York: John Wiley & Sons.

Collet, D. 1980. Outliers in circular data. *Applied Statistics* 29: 50-57.

Collet, D. & Lewis, T. 1981. Discriminating between the von Mises and wrapped normal distributions. *Australian Journal Statistics* 23(1): 73-79.

David, H.A. 1970. *Order Statistics*. New York and London: Wiley.

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

Jammalamadaka, S. R. & SenGupta, A. 2001. *Topics in Circular Statistics*. Singapore: World Scientific Press.

Kendall, D.G. 1974. Hunting quanta. *Trans. R. Soc. Lond*. A 276: 231-266.

Kent, J.T. 1976. Distributions, Processes and Statistics on "Spheres". Ph.D. Thesis, University of Cambridge.

Lukacs, E. 1970. *Characteristic Functions*. Griffin: Publisher

Mardia, K.V. 1972. *Statistics of Directional Data*. London: Academic Press.

Mardia, K.V. 1975. Statistics of directional data. *Journal of the Royal Statistical Society B* (37): 349-393.

Pearson, E.S. & Hartley, H.O. 1966. *Biometrika Tables of Statisticians*. Vol. 1, 3rd ed., London: Cambridge University Press.

Rao, J.S. 1969. Some Contributions to the Analysis of Circular Data. Ph.D. thesis, Indian Statistical Institute, Calcutta, India (Unpublished).

Whittaker, E.T. & Watson, G.N. 1944. *A Course in Modern Analysis*: Cambridge: Cambridge University Press.

Adzhar Rambli, Safwati Ibrahim, Mohd Ikhwan Abdullah & Ibrahim Mohamed*
Institute of Mathematical Sciences
University of Malaya
Kuala Lumpur
Malaysia

Abdul Ghapor Hussin
Centre for Foundation studies in Sciences
University of Malaya
Kuala Lumpur
Malaysia

*Corresponding author; email: imohamed@um.edu.my

Received: 7 July 2011
Accepted: 30 December 2011