A SEQUENTIAL TEST FOR THE MEAN DIRECTION APPLIED TO CIRCULAR DATA AND AN APPLICATION

Kadir Özgür PEKER[1] , Sevil BACANLI[2]

ABSTRACT: *In this study, a sequential test for circular data is considered. The theoretical details, operating characteristic function and average sample number function of the Sequential Probability Ratio Test for the von Mises distribution are given. Also, an illustrative application of the test is performed on a generated data. Finally, the results obtained from the application are evaluated and some interpretations of how to use this test are given.*

KEYWORDS : *Circular data, von Mises distribution, Mean direction, Sequential probability ratio test.*

DAİRESEL VERİLERDE ORTALAMA YÖN İÇİN ARDIŞIK TEST VE BİR UYGULAMA

ÖZET : *Bu çalışmada, dairesel verilerde ardışık test incelenmektedir. von Mises dağılımı için Olasılık Oranlarının Ardışık Testi'nin kuramsal tanımı, karakteristik işlem fonksiyonu ve ortalama örneklem sayısı fonksiyonu verilmiştir. Ayrıca, veri türetimi yoluyla, testin bir uygulaması yapılmıştır. Son olarak, uygulamada elde edilen sonuçlar değerlendirilmiş ve bu testin kullanımına ilişkin bazı yorumlar yapılmıştır.*

ANAHTAR KELİMELER : *Dairesel veri, von Mises dağılımı, Ortalama yön, Olasılık oranlarının ardışık testi.*

[1, 2] *Anadolu Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Yunusemre Kampüsü, ESKİŞEHİR.*
Hacettepe Üniversitesi, Fen Fakültesi,İstatistik Bölümü, Beytepe Kampüsü, ANKARA.

I. *INTRODUCTION*

Circular data are a large class of directional data, which are of interest in many fields, including biology, geography, geology, geophysics, medicine, oceanography and meteorology. For example, a biologist may be measuring the direction of flight of a bird, while a meteorologist may be interested in wind directions in a specific region. A set of such observations on directions is said to as directional data. Directional data can be conveniently represented as points on the circumference of a unit circle centered at the origin. Because of this circular representation, observations on such directions are also called circular data.

The most basic distribution on the circle is the uniform distribution. Other useful distributions are the cardioid, wrapped Cauchy and wrapped normal distributions. The von Mises distribution plays a key role in statistical inference on the circle, analogous to that of the normal distribution on the line.

A circular random variable θ is said to have a von Mises distribution if it has the probability density function:

$$f(\theta; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(\theta - \mu)} , \ 0 \leq \theta < 2\pi , \ 0 \leq \mu < 2\pi , \kappa \geq 0 \tag{1}$$

where μ and κ are parameters and $I_0(\kappa)$ denotes the modified Bessel function of the first kind and order zero, and is given by;

$$I_0(\kappa) = \sum_{r=0}^{\infty} \left(\frac{1}{2}\kappa\right)^{2r} (r!)^{-2} \tag{2}$$

If the concentration parameter κ is zero, probability density function will be uniform.

$$f(\theta) = (2\pi)^{-1} \quad , \quad 0 < \theta < 2\pi \tag{3}$$

The von Mises distribution is the most common distribution which is used in circular data analysis. Although it is used as the normal distribution on the circle, it does not have all the useful properties of the normal distribution. However, it still forms the basis on which directional data analysis is constructed.

In the second section, a single sample test of mean direction for the populations from von Mises distribution is examined for the case of the concentration parameter κ is known. In the third section, Sequential Probability Ratio Test (SPRT) of mean direction for the populations from von Mises distribution with a known concentration parameter κ is examined. Furthermore, the operating characteristic function and the average sample number for the test are described. In the fourth section, for the purpose of application of the tests, data is generated from von Mises distribution and working of the sequential test is introduced by a program. In the last section of the study, the evaluations are given.

II. A SINGLE SAMPLE TEST FOR THE MEAN DIRECTION OF A VON MISES DISTRIBUTION WHEN THE CONCENTRATION PARAMETER IS KNOWN

Let the population's distribution is von Mises and we shall deal with the problem of testing the hypothesis about mean direction is $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$. It is assumed that the concentration parameter κ is equal to a given value such as κ_0.

If $\kappa_0 \geq 2$, the circular standard error of the mean direction for the von Mises distribution is given by;

$$\sigma_{VM} = \frac{1}{\sqrt{n\rho\kappa_0}} \tag{4}$$

where ρ is the mean resultant length of the sample and can be calculated from κ_0 using Appendix-A. So, the test statistic is given by;

$$E_n = \frac{\sin(\overline{\theta} - \mu_0)}{\sigma_{VM}} \tag{5}$$

At α significance level, this value is being compared to the value which is obtained from standard normal distribution table.

4

- Test of $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$: If $\left| E_n \right| > z_{\alpha/2}$, reject H_0.

- Test of $H_0 : \mu = \mu_0$ against $H_1 : \mu < \mu_0$: If $\mu_0 - \pi < \overline{\theta} < \mu_0$ and $E_n < -z_\alpha$, reject H_0.

- Test of $H_0 : \mu = \mu_0$ against $H_1 : \mu > \mu_0$: If $\overline{\theta} < \mu_0 + \pi$ and $E_n > -z_\alpha$, reject H_0. [1].

III. SEQUENTIAL PROBABILITY RATIO TEST FOR THE MEAN DIRECTION APPLIED TO CIRCULAR DATA

III.1. Theoretical Definition

It will be assumed that observations of a variable θ are taken sequentially from a von Mises population with a mean direction μ_0 and concentration parameter κ.

Mean direction is the parameter which will be tested in a von Mises population and κ is known. So, we shall deal with the problem of testing the hypothesis; $H_0 : \mu = \mu_0$ versus $H_1 : \mu = \mu_1$.

In order to apply the SPRT, it is required to observe the likelihood function for the n-th observation where the circular observations, in order, are $\theta_1, \theta_2, \ldots, \theta_n$. For the von Mises distribution, the maximum likelihood ratio is defined as;

$$L_n = \prod_{i=1}^{n} \frac{f(\theta_i; \mu_1)}{f(\theta_i; \mu_0)} = \frac{\dfrac{1}{[2\pi I_0(\kappa)]^n} e^{\kappa \sum_{i=1}^{n} \cos(\theta_i - \mu_1)}}{\dfrac{1}{[2\pi I_0(\kappa)]^n} e^{\kappa \sum_{i=1}^{n} \cos(\theta_i - \mu_0)}} \qquad (6)$$

Then by taking logarithm and simplifying, equation (6) can be written as;

$$\ln L_n = \sum_{i=1}^{n} Z_i \qquad (7)$$

$$\ln L_n = 2\kappa \sum_{i=1}^{n} \sin(\theta_i - \nu_1)\sin(-\nu_2) \qquad (8)$$

where $v_1 = \dfrac{\mu_0 + \mu_1}{2}$ and $v_2 = \dfrac{\mu_0 - \mu_1}{2}$.

At each stage of the test process the value of $\sum\limits_{i=1}^{n} Z_i$ is computed and compared with $\ln A$ and $\ln B$ critical values which depend on type-1 (α) and type-2 (β) errors. A and B values are computed as $A = \dfrac{1-\beta}{\alpha}$, $B = \dfrac{\beta}{1-\alpha}$. Then, one of the following decision is made.

1- If $\sum\limits_{i=1}^{n} Z_i \leq \ln B$, the process is terminated with the acceptance of H_0 .

2- If $\sum\limits_{i=1}^{n} Z_i \geq \ln A$, the process is terminated with the rejection of H_0 .

3- If $\ln B < \sum\limits_{i=1}^{n} Z_i < \ln A$, the experiment is continued by taking an additional observation.

III.2. The Operating Characteristic Function and The Expected Sample Size

When μ is the test parameter for von Mises distribution, the approximate formula for the operating characteristic (OC) function $P(\mu)$, which is the probability of accepting H_0, can be obtained by considering expression;

$$E\left(\left[\frac{f(\theta;\mu_1)}{f(\theta;\mu_0)}\right]^h\right) = 1 \tag{9}$$

This expected value can be defined as;

$$\int_0^{2\pi}\left[\frac{f(\theta;\mu_1)}{f(\theta;\mu_0)}\right]^h f(\theta)\,d\theta = 1 \tag{10}$$

Then, an approximation to the OC function is given by the following equation;

$$P(\mu) = \frac{A^h - 1}{A^h - B^h} \tag{11}$$

6

where $\quad h = \dfrac{\sin(\mu - v_1)}{\sin v_2}$.

Acceptance probabilities are computed for various h values in linear data. Apart from linear data, minimum and maximum values of operating characteristic function are obtained in circular data. Differentiating OC function with respect to μ, we obtain $\mu = 90° + v_1$ and $\mu = 270° + v_1$, and these can be shown to be a minimum and maximum, respectively.

The main feature of sequential test, as separated from the current statistical test procedures, is that the sample size required by the test is not predetermined, but accepted as a random variable. Therefore, an expected value for the sample size can be obtained.

An approximation to the average sample number function $E(n;\mu)$, which is the expected number of observations, is given by;

$$E(n;\mu) = \frac{P(\mu)\ln B + \left[1 - P(\mu)\right]\ln A}{E(z;\mu)} \tag{12}$$

where

$$z = \ln\left[\frac{f(\theta;\mu_1)}{f(\theta;\mu_0)}\right] = \ln\left[\frac{\dfrac{1}{2\pi I_0(\kappa)}e^{\kappa\cos(\theta - \mu_1)}}{\dfrac{1}{2\pi I_0(\kappa)}e^{\kappa\cos(\theta - \mu_0)}}\right]$$

$$z = \kappa\left[\cos(\theta - \mu_1) - \cos(\theta - \mu_0)\right].$$

Then;

$$E(z;\mu) = E[\kappa(\cos\theta\,\cos\mu_1 + \sin\theta\,\sin\mu_1 - \cos\theta\,\cos\mu_0 - \sin\theta\,\sin\mu_0)]$$

is obtained. For a von Mises population;

$$E(\cos\theta) = A(\kappa) \quad, \quad E(\sin\theta) = 0 \quad, \quad A(\kappa) = I_1(\kappa)/I_0(\kappa)$$

Consequently,

$$E(z;\mu) = E\left[\kappa(\cos\theta(\cos\mu_1 - \cos\mu_0) + \sin\theta(\sin\mu_1 - \sin\mu_0))\right]$$

$$E(z;\mu) = A(\kappa)(\cos\mu_1 - \cos\mu_0)$$

$$E(z;\mu) = A(\kappa)\left[2\sin\left(\frac{\mu_1 + \mu_0}{2}\right)\sin\left(\frac{\mu_0 - \mu_1}{2}\right)\right]$$

$$E(z;\mu) = 2A(\kappa)\sin v_1 \sin v_2 \qquad (13)$$

is obtained [3], [4]. Substituting equation (13) in equation (12), the average sample number is;

$$E(n;\mu) = \frac{P(\mu)\ln B + [1 - P(\mu)]\ln A}{2A(\kappa)\sin v_1 \sin v_2} \qquad (14)$$

It is possible to compute maximum and minimum values of the average sample number in circular data. So, the average sample numbers, which are obtained when H_0 or H_1 is true in linear data, are computed for the maximum and minimum values in circular data. Differentiating the average sample number with respect to v_2 and setting that equal to zero gives;

$$E(n;\mu)_{min} = \frac{P(\mu)\ln B + [1 - P(\mu)]\ln A}{2A(\kappa)\sin v_1} \qquad (15)$$

As there is only one turning point which gives a minimum, the maximum will be at the ends of the range of v_2. This is the point $0°$ and it gives

$$E(n;\mu)_{max} = \infty \qquad (16)$$

This result is obvious since $v_2 = 0°$ the hypotheses are the same and it is impossible to distinguish between them [2].

IV. AN ILLUSTRATIVE APPLICATION

A data set, with 100 observations, is randomly generated from the von Mises distribution. For the selection process, the mean direction $\mu = 90°$ and the concentration parameter $\kappa = 20$ are chosen. The data set is given in Appendix-B and their raw data plot is shown in Figure 1.
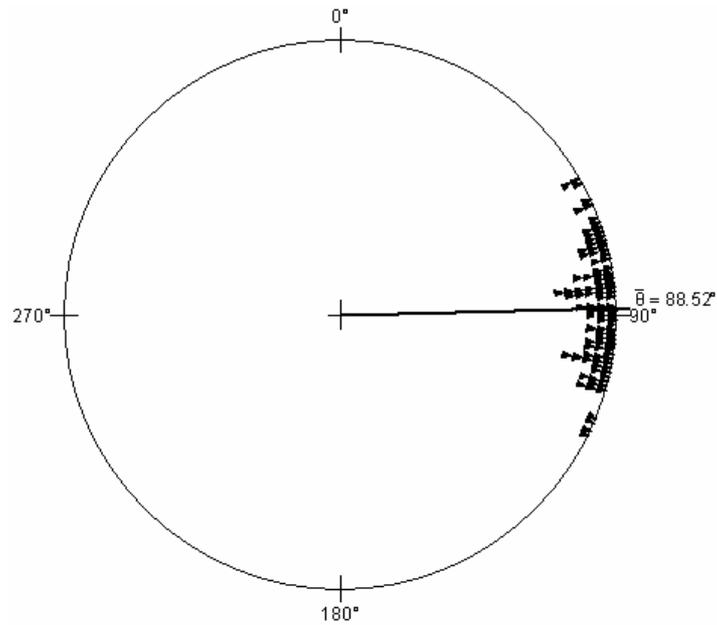
Figure 1. Circular raw data plot of the sample.

Descriptive statistics for the generated data are summarized in Table 1.

Table 1. Descriptive statistics for the sample.

Sample Size	100
Mean vector	88.52°
Length of mean vector	0.98
Concentration	20.89
Circular variance	0.02
Circular standard deviation	12.69°
Standard error of mean	1.27°

As seen from Table 1, data set follows a von Mises distribution with the parameters $\bar{\theta} = 88.52°$ and $\hat{\kappa} = 20.89$. First of all in here, a mean direction test, that is given in the second section of the study, for the case when the concentration parameter is known, is applied. The hypothesis to be tested is $H_0 : \mu = 90°$ against $H_1 : \mu \neq 90°$ and either 90° is accepted or the statistical difference is significant to reject H_0.

As $\kappa_0 = 20$ is known and $\kappa_0 \geq 2$, then according to equation (4), circular standard error of the mean direction from the von Mises distribution is;

$$\sigma_{VM} = \frac{1}{\sqrt{100(0.98)20}} = 0.0226$$

According to equation (5), the test statistic is calculated as follows;

$$E_n = \frac{\sin(88.52 - 90)}{0.0226} = -1.1428$$

As $|E_n| = 1.1428 < 1.9604 = z_{0.05/2}$, the null hypothesis is accepted for the %5 significance level.

In here, a sequential test for the mean direction that is given in the third section of the study for the case when the concentration parameter is known is applied for the data set. A new code is written in S-Plus package for this purpose. Using the code, the critical values and the stages of the test statistics, and result of the test can be seen depending on the type-1 and type-2 error probabilities for the given hypothesis.

Let the hypothesis to be tested are;

H_0: $\mu = 90°$

H_1: $\mu = 85°$.

Since $\kappa_0 = 20.89$, if α and β are chosen to be 0.05 and 0.10 respectively, the computer output for the sequential test is given as follows:

Table 2. The computer output for the sequential test.

lower limit --->	-2.2513	
upper limit --->	2.8904	
Observation No	Angle	Test Statistics
1	72.10	0.4840
2	78.40	0.7722
3	74.76	1.1741
4	98.75	0.8185
5	95.83	0.5545
6	104.37	0.0257
7	106.00	−0.5526
8	99.47	−0.9306
9	89.53	−0.9951
10	84.17	−0.8893
11	100.83	−1.3095
12	103.75	−1.8194
13	88.35	−1.8465
14	93.57	−2.0392
15	91.70	−2.1726
16	105.83	−2.7458

Null hypothesis is accepted.

After 16 observations iteration is stopped.

According to these results, from the sixteenth observation the lower limit is exceeded, so the process is stopped and the null hypothesis is accepted. As a result, the mean direction according to sequential test is 90°.

V. CONCLUSIONS

As it is seen from the test results, sequential test is more favorable from fixed sample size tests in point of sample size. Using sequential tests, particularly in researches which the data is obtained sequentially, is helpful.

Nowadays, the size of the sample is treated as a constant and predetermined in most of the studies. Sample is required to be predetermined in a test procedure which has a fixed sample size. In practice, this method is mostly caused loss of time and high costs. These difficulties can be eliminated when the sequential test is applied. An essential feature of the test is that the number of observations depends on the outcome of the observations and is, therefore, not determined in

advance, but a random variable. The process begins with an observation, required comparisons are made and it is continued until the decision about the hypotheses is made. Thus, it is provided a great saving in sample size.

When the data is collected sequentially, the measurement results are obtained serially. Therefore, using the tests which have a fixed sample size is not convenient. Testing cumulative data sequentially is a proper method. Hence, a sequential test which is improved for testing sequential circular data is examined in this paper. As it is seen from application results, instead of waiting to collect 100 observations for the simulated data, the decision about the hypothesis is made after 16 observations due to applying sequential test. Here, it is seen that the test can be concluded with how many observations on required error probabilities. If the efficiency is not adequate due to the test, it is easily concluded with changing the error probabilities.

REFERENCES

[1]   N.I. Fisher, "*Statistical analysis of circular data*", Cambridge University Press, Cambridge, Great Britain, 1993.

[2]   R.J. Gadsden., and G.K. Kanji., "Sequential analysis for angular data", *The Statistician*, 30, No. 2, 119-29, 1981.

[3]   R.J. Gadsden., ve G.K. Kanji., "Sequential analysis applied to circular data", *Commun. Statist.-Sequential Analysis*, 1(4), 304-314, 1982-83.

[4]   K.V. Mardia., "*Statistics of directional data*", Academic Press, London, England, 1972.

[5]   K.Ö. Peker., "*Dairesel Veriler ve Ardışık Testlerde Kullanımı*", Anadolu Üniversitesi Fen Bilimleri Enstitüsü Doktora Tezi, Eskişehir, 2002.

[6]   K.Ö.Peker., S. Bacanlı., "Dairesel Verilere Uygulanan Tanımlayıcı İstatistiksel Yöntemler ve Meteorolojik Bir Uygulama", *Anadolu Üniversitesi Bilim ve Teknoloji Dergisi*, Cilt/Vol.: 5 – Sayı/No: 2 : 343-349, 2004.

12

[7]   A. Wald., "*Sequential analysis*", John Wiley & Sons Inc., New York, USA, 1947.

[8]   G.B. Wetherill., "*Sequential methods in statistics*", Chapman & Hall, London, England, 1975.

Appendix-A.   The resultant lengths $\rho = A(\kappa)$ for the von Mises distribution.

κ	ρ	κ	ρ	κ	ρ
0.0	0.00000	3.5	0.84110	7.0	0.92553
0.1	0.04994	3.6	0.84616	7.1	0.92663
0.2	0.09950	3.7	0.85091	7.2	0.92770
0.3	0.14834	3.8	0.85537	7.3	0.92874
0.4	0.19610	3.9	0.85956	7.4	0.92975
0.5	0.24250	4.0	0.86352	7.5	0.93072
0.6	0.28726	4.1	0.86726	7.6	0.93168
0.7	0.33018	4.2	0.87079	7.7	0.93260
0.8	0.37108	4.3	0.87414	7.8	0.93350
0.9	0.40984	4.4	0.87732	7.9	0.93438
1.0	0.44639	4.5	0.88033	8.0	0.93524
1.1	0.48070	4.6	0.88320	8.1	0.93607
1.2	0.51278	4.7	0.88593	8.2	0.93688
1.3	0.54267	4.8	0.88853	8.3	0.93767
1.4	0.57042	4.9	0.89101	8.4	0.93844
1.5	0.59613	5.0	0.89338	8.5	0.93919
1.6	0.61990	5.1	0.89565	8.6	0.93993
1.7	0.64183	5.2	0.89782	8.7	0.94064
1.8	0.66204	5.3	0.89990	8.8	0.94134
1.9	0.68065	5.4	0.90190	8.9	0.94202
2.0	0.69777	5.5	0.90382	9.0	0.94269
2.1	0.71353	5.6	0.90566	9.2	0.94398
2.2	0.72803	5.7	0.90743	9.4	0.94521
2.3	0.74138	5.8	0.90913	9.6	0.94639
2.4	0.75367	5.9	0.91078	9.8	0.94752
2.5	0.76500	6.0	0.91236	10	0.94860
2.6	0.77545	6.1	0.91389	12	0.95730
2.7	0.78511	6.2	0.91536	15	0.96607
2.8	0.79404	6.3	0.91678	20	0.97467
2.9	0.80231	6.4	0.91816	24	0.97937
3.0	0.80999	6.5	0.91949	30	0.98319
3.1	0.81711	6.6	0.92078	40	0.98739
3.2	0.82375	6.7	0.92202	60	0.99163
3.3	0.82993	6.8	0.92323	120	0.99582
3.4	0.83570	6.9	0.92440	∞	1.00000

Appendix-B. Observations which are created from the von Mises distribution.

[1]	72.10	78.40	74.76	98.75	95.83	104.37	106.00	99.47	89.53	84.17
[11]	100.83	103.75	88.35	93.57	91.70	105.83	115.14	84.53	81.74	100.31
[21]	85.34	85.83	97.17	66.55	93.87	65.49	96.49	104.55	60.80	75.53
[31]	85.36	116.69	100.43	88.84	69.77	97.32	80.78	106.34	81.63	105.80
[41]	102.07	104.29	101.09	90.28	84.97	90.64	78.31	102.21	106.58	76.68
[51]	74.07	99.53	95.65	70.95	84.49	76.44	61.58	80.56	88.97	76.49
[61]	91.02	88.45	73.76	89.44	92.34	71.25	100.75	81.56	82.89	90.85
[71]	75.89	83.13	77.84	89.01	86.92	97.44	71.20	72.93	93.80	112.95
[81]	94.51	66.83	73.86	96.49	82.80	101.52	60.24	113.21	81.60	84.64
[91]	100.75	96.47	84.87	99.18	86.18	85.79	83.63	94.82	85.01	75.45