Session 2(b): Statistical Standard, Methodology and Application

Topic: In Approximate Distributions for Circular Variables

Nor Hafizah binti Moslim
Centre of Foundation Studies for Science, University of Malaya
Co-Authors: Yong Zulina Zubairi, Abdul Ghapor Hussin, Siti Fatimah Hassan

UNIVERSITY OF MALAYA

OUTLINE

Abstract

This study focuses on the approximation of von Mises distribution to the normal distribution. The von Mises distribution or also known as circular normal distribution has two parameters namely mean direction, μ and concentration parameter, κ . This distribution can be treated as linear variable for the sufficiently large value of the concentration parameter. This study will examine the asymptotic behavior of the concentration parameter with the suitable value of sample size, *n*. The simulation study is carried out and the Kolmogorov-Smirnov test is used to test the goodness of fit for three level of significance values. The study suggests that as sample size and concentration parameter increase, the percentage of samples follow the normality assumption increase.

Introduction

UNIVERSITY OF MALAYA

- Statistical data can be classified according to their distributional topologies. A linear data set can be represented on a straight line and for circular data, they can be represented by the circumference of a circle. The data is commonly measured in the range of $(0^\circ, 360^\circ]$ degree or $(0,2\pi]$ radian. It is worthwhile to note that statistical theories for straight line and circle are very different from one to another because the circle is a closed curve but line is not. The application of directional statistics can be found in the area of meteorology such as wind direction. Circular data can be found in many fields (Mardis, 1972); (Amos, 1974); (Mardia and Jupp, 2000).

- A circular random variable from a von Mises or Circular Normal distribution has a density function of

$$f(\theta) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(\theta-\mu)}, 0 \le \theta \le 2\pi$$

where $0 \le \mu \le 2\pi$ and $\kappa \ge 0$ are the parameters. $I_0(\kappa)$ is the modified Bessel function of order zero and can be defined as

$$I_0(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} e^{\kappa \cos(\theta-\mu)}$$

Introduction

CONT.

- This distribution is the 'natural' analogue on the circle of the normal distribution on the real line and has few similar characteristics with the normal distribution. (Dobson, 1978)

- It has been proved that the von Mises distribution can be approximated to the standard normal distribution for sufficiently large k (Jammalamadaka and Sengupta, 2001) ; (Fisher, 1993). As $\kappa \to \infty$,

$$\beta = \sqrt{\kappa}(\theta - \mu) \xrightarrow{d} N(0,1)$$

- Let $\beta = \sqrt{\kappa}(\theta - \mu)$. For large κ, $(\theta - \mu) = \frac{\beta}{\sqrt{\kappa}}$ is small and from the Taylor series expansion $\cos\theta \approx 1 - \frac{\theta^2}{2}$, it gives

$$\cos(\theta - \mu) = \cos\left(\frac{\beta}{\sqrt{\kappa}}\right)$$

$$\approx 1 - \frac{\beta^2}{2\kappa} \qquad (1)$$

Introduction

UNIVERSITY OF MALAYA

cont.

- Using the change of variable formula,

$$g(\beta)=\left|\frac{\partial\theta}{\partial\beta}\right| f\left(g^{-1}(\beta)\right)$$

$$=\frac{1}{\sqrt{\kappa}}\frac{e^{\kappa\cos\left(\frac{\beta}{\sqrt{\kappa}}\right)}}{2\pi I_0(\kappa)} \qquad (2)$$

- The Bessel functions, $I_0(\kappa)$ can be approximated to

$$I_0(\kappa)\sim\frac{e^{\kappa}}{\sqrt{2\pi\kappa}} \text{ for large } k. \qquad (3)$$

Introduction

UNIVERSITY OF MALAYA

cont.

- Thus, by substituting the Eq. (1) and (3) into Eq. (2),

$$g(\beta)=\frac{1}{\sqrt{\kappa}}\frac{e^{\kappa\cos\left(\frac{\beta}{\sqrt{\kappa}}\right)}}{2\pi\frac{e^{\kappa}}{\sqrt{2\pi\kappa}}}$$

$$=\frac{e^{\kappa\left(1-\frac{\beta^2}{2\kappa}\right)}}{e^{\kappa}\sqrt{2\pi}}$$

$$=\frac{1}{\sqrt{2\pi}}e^{\left(-\frac{\beta^2}{2}\right)}\sim N(0,1)$$

- From the substitution above, the distribution $g(\beta)$ is approximate to the standard normal distribution. The simulation study should be carried out to test the approximation (Abuzaid et al., 2012).

Simulation Study

- Simulation studies were conducted for fourteen different sample sizes, $n = 5$, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 150 and 200 respectively with various values of concentration parameter, $\kappa = 1$, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 15, 17, 20, 25, 30, 35 and 45 for three different levels of significance, $\alpha = 0.01, 0.05, 0.10$. The mean direction, μ is fixed at $\frac{\pi}{4}$. The 3000 simulation of sample data were generated from von Mises distribution. For each generated sample, get the value of $\beta = \sqrt{\kappa}(\theta - \mu)$. Use the Kolmogorov-Smirnov test to test the goodness-of-fit for each value of β under the three levels of significance.

- Table 1 shows the percentage of samples that follow Standard Normal Distribution, $N(0,1)$. The results were given under the level of significance, 0.01, 0.05, 0.10 for the first, second and third rows, respectively.

Simulation Result

UNIVERSITY OF MALAYA

	5	10	20	30	40	50	60	70	80	90	100	120	150	200
1	53.70	28.47	5.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	64.70	43.00	11.90	1.37	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	85.00	65.13	32.30	10.70	2.77	0.47	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	73.30	59.50	35.23	20.33	9.97	4.70	2.30	0.67	0.10	0.00	0.00	0.00	0.00	0.00
	81.40	73.23	48.50	32.93	18.93	10.90	6.50	2.90	1.17	0.40	0.03	0.00	0.00	0.00
	94.43	88.40	73.43	58.97	45.57	30.87	23.70	13.43	9.57	5.10	3.07	0.23	0.00	0.00
3	82.37	75.63	63.83	53.20	43.63	35.47	29.70	21.03	15.33	12.87	9.47	4.30	1.47	0.00
	89.40	84.77	76.37	65.83	57.83	49.93	43.57	34.03	27.00	22.90	18.83	10.23	4.03	0.57
	96.93	95.20	91.57	86.07	80.13	74.77	70.37	63.90	54.27	48.13	44.73	33.33	18.00	6.20
4	85.23	84.00	80.27	72.57	67.53	63.87	59.57	53.57	48.93	45.23	39.67	29.77	23.63	11.10
	91.80	91.43	88.03	83.30	78.70	76.23	72.93	67.90	64.40	60.30	54.53	44.27	35.53	21.07
	98.10	97.50	96.83	94.63	92.00	91.37	89.23	86.37	85.27	82.83	79.00	73.77	65.83	50.13
5	88.13	88.27	84.77	82.57	78.03	76.87	75.57	73.03	71.90	67.13	66.23	62.00	53.17	41.20
	94.03	94.13	91.77	89.57	86.60	86.60	85.40	82.80	81.60	79.03	78.40	75.40	66.70	55.90
	98.87	98.77	98.20	97.27	96.53	95.93	95.43	94.80	94.63	93.30	93.17	90.60	86.53	80.53
6	89.73	89.30	86.80	86.80	84.67	83.87	83.90	81.53	81.20	79.43	78.90	76.60	73.00	64.80
	94.50	94.77	93.23	92.67	92.23	91.07	91.17	89.67	89.10	88.20	87.30	85.63	83.47	76.73
	98.87	98.83	98.60	98.17	98.03	97.80	97.30	97.57	97.10	96.87	96.17	95.80	95.03	92.10
7	89.33	88.43	88.53	87.60	87.33	86.67	86.87	86.77	84.73	85.37	85.40	82.60	81.73	77.97
	94.20	94.13	94.63	93.67	93.07	92.40	92.87	92.37	91.60	91.77	92.57	90.20	90.37	86.67
	99.13	98.57	99.17	98.33	98.47	98.30	97.87	98.27	97.93	98.13	98.20	98.20	97.60	96.47

Simulation Result

cont.

	90.10	88.93	88.03	88.57	88.77	88.90	88.37	88.53	87.83	87.77	87.87	87.53	86.27	83.77
8	94.93	94.23	93.77	94.10	94.47	94.20	94.03	94.10	93.00	93.27	93.70	93.83	92.17	90.63
	99.17	98.60	98.63	98.80	98.63	98.50	98.73	98.67	98.63	98.70	98.93	98.10	97.93	97.63
	89.87	89.17	89.70	90.03	89.87	88.93	89.70	89.67	89.80	90.33	88.53	90.70	89.43	86.50
9	95.00	94.33	94.60	95.27	94.47	93.97	95.17	94.57	94.57	95.33	94.67	94.93	94.07	92.43
	99.27	98.70	98.83	99.20	98.90	99.00	99.10	99.00	99.03	99.13	98.90	99.00	98.70	98.37
	89.90	90.00	89.23	90.20	89.43	89.37	91.13	90.83	90.70	89.80	90.90	89.50	88.50	89.33
10	95.60	95.13	94.77	95.33	94.93	95.30	95.63	95.73	95.43	94.93	95.37	94.77	94.07	94.30
	99.07	99.17	98.97	98.83	99.10	98.83	99.00	99.20	99.07	98.97	99.27	98.60	98.70	98.77
	88.77	89.33	89.63	89.17	89.47	89.47	89.63	90.33	89.97	90.63	89.83	90.43	91.30	90.27
12	94.57	94.67	94.30	94.33	94.33	94.77	94.70	95.47	95.40	95.23	94.80	95.07	96.03	95.07
	99.10	98.97	98.80	98.83	98.97	99.17	99.13	99.30	99.23	99.23	98.97	99.07	99.20	99.07
	90.43	89.30	89.47	89.57	90.10	89.67	90.67	91.03	90.90	90.67	91.00	90.30	90.90	90.73
15	94.97	94.57	95.03	94.97	94.83	95.10	95.23	95.73	95.60	95.40	95.77	95.60	95.33	95.80
	99.07	98.93	98.90	98.80	99.13	98.97	99.17	99.27	99.33	99.03	99.00	99.40	99.17	99.17
	90.23	89.63	89.50	89.47	89.67	89.87	90.80	91.40	91.20	90.20	90.47	90.00	90.83	90.97
17	95.43	94.93	95.10	94.70	94.83	95.30	95.33	95.43	95.63	95.07	95.67	95.53	95.53	95.30
	99.03	98.90	98.90	98.67	99.13	98.97	99.20	99.43	99.20	99.23	99.10	99.23	99.20	99.30
	90.10	89.33	89.60	90.43	90.10	89.83	90.47	90.27	91.77	90.97	91.77	91.10	91.23	91.17
20	95.23	94.43	94.83	94.67	94.70	95.00	95.10	95.00	95.67	95.47	95.77	95.60	95.63	95.70
	99.07	99.07	98.97	98.73	98.97	98.70	99.20	98.83	99.07	99.10	99.03	99.00	98.93	99.37

Simulation Result

UNIVERSITY OF MALAYA

cont.

25	**89.97**	**88.80**	**90.07**	**89.60**	**89.17**	**89.70**	**90.43**	**90.47**	**90.83**	**90.37**	**90.53**	**90.87**	**90.57**	**90.37**
	94.93	94.67	94.87	94.50	94.03	94.50	95.27	95.07	95.73	95.90	95.43	95.07	95.43	95.27
	98.93	98.97	99.10	98.87	98.73	99.03	99.07	98.80	99.23	99.03	98.97	98.87	99.07	99.00
30	90.53	90.20	89.90	90.43	88.83	90.30	90.03	91.80	90.67	90.70	91.23	91.10	91.50	90.13
	95.47	94.43	95.27	95.47	93.97	95.23	94.87	96.13	95.77	95.33	96.03	95.57	95.93	94.70
	99.17	98.83	98.87	98.90	98.70	99.03	99.17	99.33	99.17	99.10	99.27	99.10	99.47	98.80
35	89.87	90.03	89.97	90.50	88.73	89.80	91.13	90.97	90.77	91.13	91.67	91.47	90.30	89.83
	94.73	94.53	94.93	95.47	94.70	94.67	95.47	95.73	95.47	95.50	95.83	95.63	94.43	94.83
	98.93	98.73	98.87	98.83	98.87	98.70	99.13	99.43	99.17	99.20	99.30	98.90	99.07	99.23
45	89.73	89.20	90.20	89.63	90.23	90.57	91.13	90.50	91.23	90.53	90.67	90.40	90.63	90.63
	95.60	94.70	94.93	94.47	94.97	94.73	95.63	95.40	95.53	95.17	95.37	95.43	95.57	95.67
	98.70	99.13	98.93	99.10	99.03	98.97	99.13	99.07	98.90	99.23	99.10	99.13	99.27	99.23

Simulation Result

cont.

- At 10% of significance level, for any fixed *n*, the percentage of samples that follow the standard normal distribution increase as the concentration parameter, k increase. This is there for 5% and 1%, respectively.

- Also, at any significance level, for any fixed $\kappa < 10$, increase in n results is decrease in the percentage of samples that follow the standard normal distribution. But for $\kappa \geq 10$, as n increase the percentage remains constant approximate at 90%, 95% and 99% for the significance level of 10%, 5% and 1%, respectively.

- Of the three levels of significance, $\alpha = 0.01$ shows most percentage of samples follow standard normal distribution for any value of concentration parameter, k and sample size, n. Small sample size ($n \leq 30$) shows more than 96% of samples follow standard normal distribution for $\kappa \geq 5$. Concentration parameter shows more than 96% of samples follow standard normal distribution for large sample size.

- More than 98% of generated samples follow standard normal distribution for any sample size when the concentration parameter, $\kappa \geq 9$.

Conclusion

- In summary, the generated samples of von Mises distribution for concentration parameter, $k \geq 5$ is approximated to the standard normal distribution. As the sample size, n and concentration parameter increase, the percentages of samples follow standard normal distribution increase regardless of the levels of significance values.

References

- Jammalamadaka, S. R. and Sengupta , A. Topics in Circular Statistics. London: World Scientific. 2001.
- Mardia, K. V. Statistics of Directional Data. London: Academic Press. 1972.
- Mardia, K. V. and Jupp, P. E. Directional Statistics. England: John Wiley & Sons Ltd. 2000.
- Abuzaid, A. H., Hussin, A. G., Rambli, A. and Mohamed, I. Statistics for a New Test of Discordance in Circular Data. Coomunications in Statistics – Simulation and Computation. 2012.41: 1882-1890.
- Amos D.E., Computation of Modified Bessel Functions and Their Ratios. Mathematics of Computation. 1974.28(125): 239-251.
- Dobson A.J., Simple Approximations for the Von Mises Concentration Statistic, Journal of the Royal Statistical Society, Series C (Applied Statistics). 1978.27(3): 345-347.
- Fisher N.1., Statistical Analysis of Circular Data, Cambridge University Press, Cambridge, 1993.

Thank You