Research Journal of Applied Sciences 11 (9): 870-873, 2016
ISSN: 1815-932X

Robust Estimation of Circular Parameters

[1]Ehab A. Mahmood, [1]Habshah Midi and [2]Abdul Ghapor Hussin
[1]Department of Mathematics/Universiti Putra Malaysia,
43400 UPM Serdang, Selangor, Malaysia
[2]Faculty of Defence Science and Technology,
National Defence University of Malaysia, Kuala Lumpur, Malaysia

Abstract: Researchers interest to develop methods of robust estimation. These methods can be used when the data have outliers or not satisfy the condition of classical methods. However, few researchers suggest robust estimation of circular data. In this study, we propose robust estimation of circular variance and mean resultant length. The proposed robust estimation depends on extending trimmed procedure by find robust formula for trimming. Simulation results and practical example show that the proposed procedure for the circular variance and mean resultant length are better than classical methods for different ratios of outliers.

Key words: Outliers, robust statistics, circular variance, mean resultant length, resultant, length

INTRODUCTION

The problem of outliers is as old as statistics science. The occurrence of outliers in the statistical data may causes misleading of the outcome and the statistical analysis. There are many robust methods to estimate parameters in the linear data, for example Barnett and Lewis (1994), Rousseeuw and Leory (1987) and Maronna *et al*. (2006). However, very few methods have been suggested for circular data. In this study, we are interesting to find robust estimation of circular variance and mean resultant length.

Circular data are data that can be displayed on the circumference of a unit circle. They are used in many scientific fields such as meteorology, geology, physiology and psychology. They can be measured either in degrees when they are distributed in the interval (0- 360°) or in radians (0-2π). The mean resultant length $\bar{R}$ is a measure of the concentration of the circular observations at a specific point of the circumference of the circle. It is calculated using the following equation (Fisher, 1993):

$$\bar{R} = \sqrt{\bar{C}^2 + \bar{S}^2}$$

Where:

$(0 \leq \bar{R} \leq 1)$

$\bar{c} = cn^{-1}$

$\bar{s} = s\, n^{-1}$

$$s = \sum_{i=1}^{n} \{\sin(\vartheta_i)\},\ c = \sum_{i=1}^{n} \{\cos(\vartheta_i)\}$$

The $\bar{R} = 0$ is satisfied when the circular data are widely dispersed on the circumference ($\bar{c} = 0$; $\bar{s} = 0$). The $\bar{R}$ is satisfied when the circular data have a high concentration at a specific point ($\bar{c} + \bar{s} = 1$) (Mardia and Jupp, 2000). Circular variance cv which is a measure of the dispersion of the circular data. Smaller values for the circular variance refer to a higher concentration where ($v = 1 - \bar{R}$) (Fisher, 1993). Outlier in circular data can be defined as a value with large circular distances from the value to the two neighboring observations on a circumference of a circle (Rao, 1969). Researchers suggest two main procedures of dealing with existence of outliers in statistical data, identify outliers or apply robust statistics methods. The robust statistics aims to have high efficiency by down-weight effect of outliers.

In the literature, very few robust methods have been suggested to estimate circular parameters. Wehrly and Shine (1981) compared the robustness properties of the circular mean and the circular median. Laha and Mahesh (2011) studied the robustness of the circular mean and the trimmed circular mean for different mixtures of circular distributions. Kato and Eguchi (2016) suggested a procedure to estimate both the location and the concentration parameters simultaneously for the general case of the von Mises-Fisher distribution. To date, there is no work has been done to find robust estimator of the circular variance when the circular data have outliers.

Trimmed mean is one of the robust methods that is used to estimate location parameter in linear data. Statistical methods that are used in linear data can not be

Corresponding Author: Ehab A. Mahmood, Abdul Ghapor Hussin, Department of Mathematics/Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

used for circular data. This is because of the circular geometry theory. In this study, we improve trimmed procedure to estimate circular variance and mean resultant length.

MATERIALS AND METHODS

Proposed robust procedure: There are some circular distributions, the most famous is the von Mises distribution. It is called the normal distribution for the circular data which can be denoted by:

$$[vM(\mu, k)]$$

Where:

μ = The circular mean
k = The concentration parameter

The probability density function of the circular observations $\vartheta_1, \vartheta_2, ..., \vartheta_n$ that follow von Mises distribution is given by (Mardia and Jupp, 2000):

$$g(\vartheta,\mu,k) = \frac{1}{2\pi I_0(k)} e^{k\cos(\vartheta-\mu)}$$

Where:

$$I_0(k) = \frac{1}{2\pi}\int_0^{2\pi} e^{k\cos(\vartheta)} d\vartheta$$

is a modified Bessel function of order zero. The trimmed procedure aims to eliminate outliers from the data set. In linear data, the estimator is calculated by eliminating a proportion of the largest and smallest values where the proportion of eliminating α>(0.0.5) (Maronna *et al*., 2006). However, the circular data are bounded and the maximum values are close to the minimum values as illustration in Fig. 1.

So, we cannot trim the data set according to its largest and smallest values. Therefore, in this study, we try to find a robust circular formula to calculate the trimmed circular variance and trimmed mean resultant length when the circular data have some outliers. It is expected that outliers lie far away from the circular mean (Mardia and Jupp, 2000). However, the circular mean is not robust if there exist outliers. We therefore, propose to consider the circular distance between observations and the circular median as a measure to trim the circular data. (the circular median is more efficient than the circular mean when the circular data have outliers Ducharme and Milasevic, 1987). The circular distance is calculated according to the following formula (Jammalamadaka and Sengupta, 2001):

$$dist(i) = \pi - |\pi - |\vartheta i - med(\vartheta)||$$

where, $0 \leq dist(i) \leq \pi$.

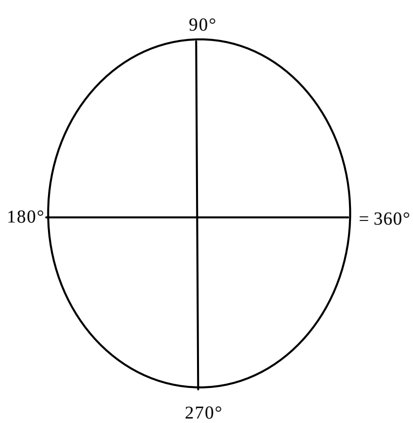

Fig. 1: Maximum and minimum values

RESULTS AND DISCUSSION

Simulation: We test the performance of five circular distances away from the circular median $\pi/6$, $2\pi/6$, $3\pi/6$, $4\pi/6$, $5\pi/6$, they are named Trim 1-5, respectively. The data are simulated from ϑ ~vM (0, k) for three samples n = 20, 60 and 100, respectively and using five values of the concentration parameter (k = 2, 4, 6, 8 and 10). The procedure is applied for clean circular data (without outliers) and with three ratios of contaminated data 5, 10 and 20%. About 10,000 samples are obtained for these combinations. The results of comparing classical methods with trimmed circular variance and trimmed mean resultant length are showed in Fig. 2 and 3, respectively for (n = 60).

By comparing with the results of circular variance of classical method for clean data, we notice that the values of classical method cause quite misleading for contaminated data especially with increasing ratio of contamination. Trimmed circular variance gives the best estimation according to the value of concentration parameter k. When k = 2, Trim 4 gives the best estimation for all ratios of contamination. Where, Trim 3 gives the best estimation for all ratios of contamination when k = 4. For other values of k, we notice Trim 2 have the best estimation. This is as we expected because for small values of k, the circular data will be more spread around the circumference of the circle. However, the circular data will more concentrate at circular mean for large value of k. As a results, the trimming depends on the value of k. We notice similar points for the mean resultant length.

For n = 20, we have got the best estimation for Trim 3 when k = 2 and Trim 2 for the otherwise. Trim 5 have the best estimation when k = 2 for n = 100. This is because of the data will be more spread around the circumference of a circle when increasing the sample size. Interested readers can request the corresponding author to provide more results.

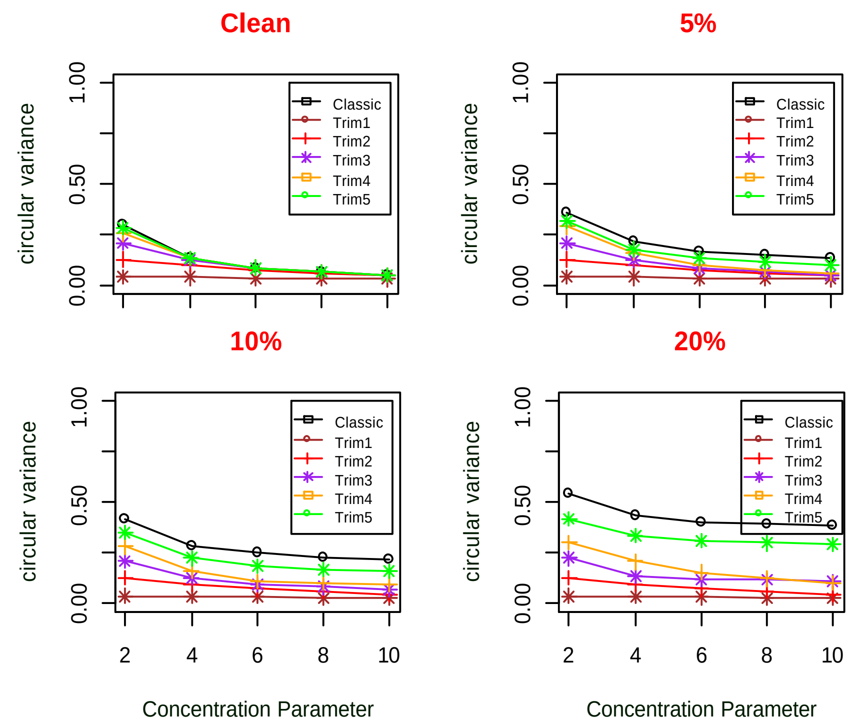

Fig. 2: Circular variance of clean and contaminated data (n = 60)

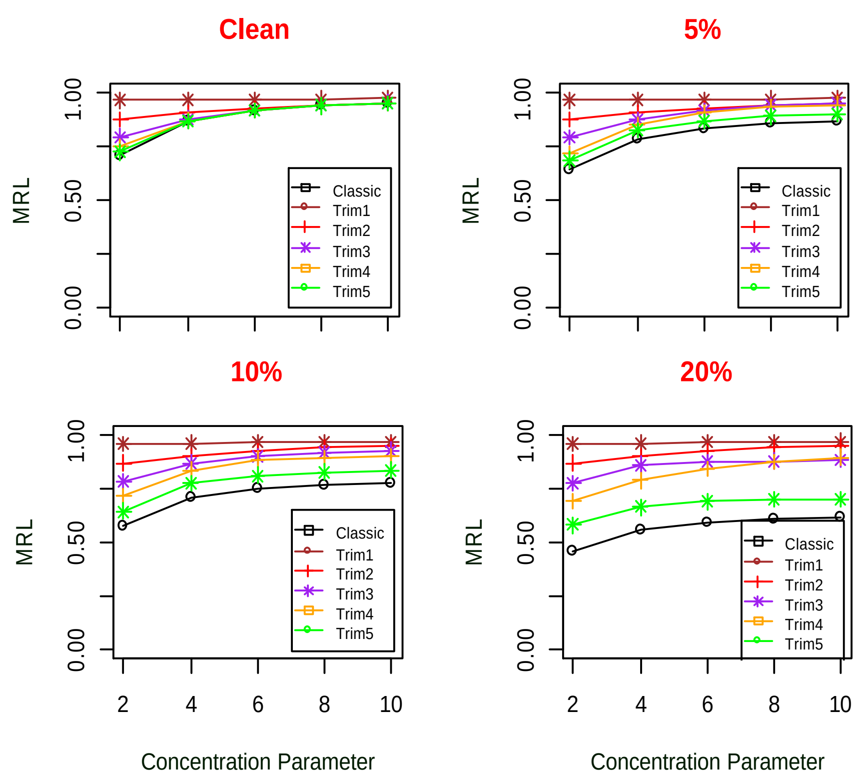

Fig. 3: Mean Resultant Length (MRL) of clean and contaminated data (n = 60)

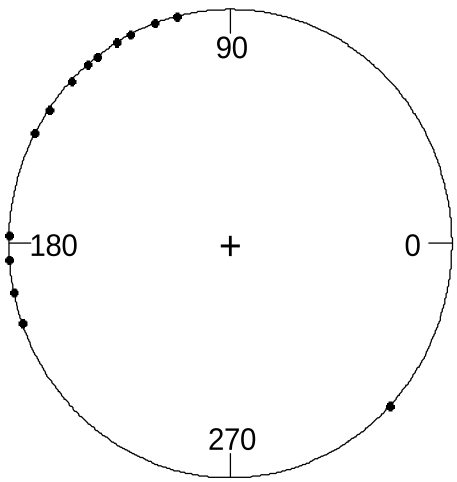

Fig. 4: Frog data set

Table 1: A comparison of the variance and the estimated *k* (frog data)

Variables	Original data	Delete outlier	Trim3
cv	0.275	0.143	0.143
MRL	0.725	0.857	0.857

Practical example: A number of frogs were collected from the mud flats of an abandoned stream meander near Indianola, Mississippi. After 30 h, fourteen of the frogs were released and the directions taken by them were recorded. These circular data have been tested by Collett (1980). He identified that the observation numbered 14 is an outlier. Figure 4 explains the spread of the frog data set on the circumference of a circle. We consider Trim3 because of the value of the estimation of k is >2. The estimation of the circular variance and the mean resultant length are given in Table 1. Trim3 gives the least value for the variance which is equal to the variance after deleting the outlier. Similarly, the results of the mean resultant length of Trim3 are the same as the estimates after the outlier was deleted.

CONCLUSION

In this study, we improve trimmed procedure to estimate the circular variance and the mean resultant length. The results confirmed that this procedure is successful with different ratios of outliers.

REFERNCES

Barnett, V. and T. Lewis, 1994. Outliers in Statistical Data. 3rd Edn., John Wiley and Sons, New York, ISBN: 0-471-93094-6, Pages: 584.

Collett, D., 1980. Outliers in circular data. Appl. Stat., 1: 50-57.

Ducharme, G.R. and P. Milasevic, 1987. Some asymptotic properties of the circular median. Commun. Stat. Theor. Methods, 16: 659-664.

Fisher, N.I., 1993. Statistical Analysis of Circular Data. 1st Edn., Cambridge University Press, Cambridge, New York, ISBN-10: 0521568900.

Jammalamadaka, S.R. and A. Sengupta, 2001. Topics in Circular Statistics. Vol. 5, World Scientific Publishing Co Pte Ltd., River Edge, New. Jersey, ISBN-13: 978-9810237783, pages: 319.

Kato, S. and S. Eguchi, 2016. Robust estimation of location and concentration parameters for the von Mises-Fisher distribution. Stat. Pap., 57: 205-234.

Laha, A.K. and K.C. Mahesh, 2011. SB-robustness of directional mean for circular distributions. J. Stat. Plann. Inference, 141: 1269-1276.

Mardia, K.V. and P.E. Jupp, 2000. Directional Statistics. 1st Edn., John Wiley, Chichester.

Maronna, R.A., R.D. Martin and V.J. Yohai, 2006. Robust Statistics, Theory and Methods. John Wiley and Sons Ltd, Hobokon, New Jersey, USA.,.

Rao, J.S., 1969. Some contributions to the analysis of circular data. Ph.D Thesis, Indian Statistical Institute, Calcutta, India.

Rousseeuw, P.J. and A.M. Leory, 1987. Robust Regression and Outlier Detection. 1st Edn., Wiley, New York, USA., ISBN-10: 0471852333, Pages: 352.

Wehrly, T.E. and E.P. Shine, 1981. Influence curves of estimators for directional data. Biometrika, 68: 334-335.