Journal of Engineering and Applied Sciences 13 (16): 6612-6615, 2018
ISSN: 1816-949X

Robust Trimmed Mean Direction to Estimate Circular Location Parameter in the Presence of Outliers

[1]Habshah Midi, [2]Ehab A. Mahmood, [3]Abdul Ghapor Hussin, [1]Jayanthi Arasan
[1]Faculty of Science and Institute for Mathematical Research Universiti Putra Malaysia (UPM), 43400 Serdang, Selangor, Malaysia
[2]Department of Banking and Finance/University of Babylon, Iraq
[3]Faculty of Science and Institute for Mathematical Research Universiti Putra Malaysia (UPM)

Abstract: Mean direction is a good measure to estimate circular location parameter in univariate circular data. However, it is bias and cause misleading when the circular data has some outliers, especially with increasing ratio of outliers. Trimmed mean is one of robust method to estimate location parameter. Therefore in this study, it is focused to find a robust formula for trimming the circular data. This proposed method is compared with mean direction, median direction and M estimator for clean and contaminated data. Results of simulation study and real data prove that trimmed mean direction is very successful and the best among them.

Key words: Trimmed mean direction, mean direction, median direction, circular distance, M estimator, Malaysia

INTRODUCTION

It is well known that circular location parameter is an important measure to give an idea about the majority of the data set. Mean direction is successful to estimate the location parameter but it does not provide consistent as well as efficient estimator in the presence of outliers. Statistical data may has some observations that are inconsistent with the others and cause some problems of statistical analysis, these observations are defined as outliers. The existence of outliers in statistical data causes misleading of statistical results and parameters estimation (Barnett and Lewis, 1978). Therefore, researchers interest to identify outliers or to use robust methods.

The normal distribution for circular data is the von Mises distribution. In linear data, the normal distribution has some important properties. Similarly, in circular data, the von Mises distribution has some important properties. Therefore in this study, it is considered the von Mises distribution. It is a symmetric unimodal distribution with circular mean μ and concentration parameter k. The probability density function of the 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 0• μ<2• , k• 0 and I_0 denotes the modified Bessel function of the first kind and order 0 which can be defined as $I_0(k) = \frac{1}{2\pi}\int_0^{2\pi} e^{k\cos(\vartheta)}\,d\vartheta$. The mean direction of the circular observations is calculated by Jammalamadaka and SenGupta (2001):

$$\hat{\mu} = \begin{cases} \tan^{-1}\left(\frac{s}{c}\right) & \text{if } c>0,\ s \geq 0 \\ \frac{\pi}{2} & \text{if } c = 0,\ s>0 \\ \tan^{-1}\left(\frac{s}{c}\right)+\pi & \text{if } c<0 \\ \tan^{-1}\left(\frac{s}{c}\right)+2\pi & \text{if } c \geq 0,\ s < 0 \\ \text{Undefined} & \text{if } c = 0,\ s = 0 \end{cases} \quad (1)$$

where $s = \sum_{i=1}^{n}\{\sin(\vartheta_i)\}$, $c = \sum_{i=1}^{n}\{\cos(\vartheta_i)\}$. Median direction is defined as any angle • such that half of the data points lie in the arc •, •+• and the majority of the data points are nearer to • than to •+• (Mardia and Jupp, 2000).

In the literature, some researchers proposed methods to detect outliers or proposed robust methods in univariate circular data. Collett (1980) suggested four statistics, namely L, C, D and M, to detect an outlier in univariate circular data. Lenth (1981) adapted an M estimator method to estimate circular location parameters. He and Simpson (1992) recommended the use of the circular median as an estimate of the circular mean when the data do not follow the von Mises distribution. Jammalamadaka and SenGupta (2001) proposed three

Corresponding Author: Habshah Midi, Faculty of Science and Institute for Mathematical Research Universiti Putra Malaysia, (UPM) 43400, Serdang, Selangor, Malaysia

methods to detect outliers in univariate circular data. First, they use the P-P plot as way of detecting outliers in circular data. Second, used LMPI for the circular data that were obtained by mixing a wrapped stable distribution with a circular uniform distribution. Third, they proposed using a Likelihood Ratio Testing (LRT) approach to identify outliers in circular data. Abuzaid (2012) proposed many methods to identify outliers. He proposed to consider cluster analysis as a procedure to detect outliers in univariate circular data. In addition, he used the C and D statistics as numerical statistics as suggested by Collett (1980) and the boxplot as a graphical method to identify outliers. Laha and KC (2015) studied the robustness of the likelihood ratio, the circular mean and the circular trimmed mean test function to test hypotheses on the mean direction of two circular distributions: the von Mises and the wrapped normal distributions. However, in their simulation study, they assumed that the circular data had a single outlier and did not consider the problem when the circular data have many outliers. 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.

However, it is still important to propose a robust method to estimate circular location parameter when the circular data have outliers. Therefore, in this study, it is proposed to apply trimmed mean direction as a method to estimate circular location parameter by finding a formula to trim the circular data set.

MATERIALS AND METHODS

M-estimator: Lenth (1981) extended the M estimator to estimate a circular location parameter for circular data. Let $\bullet_1, \bullet_2, ..., \bullet_n$ be a random sample following the von Mises distribution with circular mean μ and concentration parameter k. Then, the $\bullet$ function is given by:

$$\sum_{i=1}^{n} \rho(t(\vartheta_i - \hat{\mu}; k) = \text{minimum} \quad (2)$$

where $t(\vartheta_i - \hat{\mu}; k)$ is a periodic function that in some sense standardizes the values of $(\vartheta_i - \hat{\mu})$ according to the concentration parameter k. By differentiating Eq. 2, we obtain:

$$-k\sum_{i=1}^{n} w_i \sin(\vartheta_i - \hat{\mu}) = 0$$

where:

$$w_i = w(t(\vartheta_i - \hat{\mu})) = \frac{\psi(t(\vartheta_i - \hat{\mu}; k))}{t(\vartheta i - \hat{\mu}; k)} \quad (3)$$

The $\bullet$ function is given by:

$$\psi_H(t) = \begin{cases} t & |t| \le c \\ c\,\text{sign}(t) & |t| > c \end{cases} \quad (4)$$

where c is a constant. The M estimator method is summarized as follows. First, set all $w_i = 1$. Second, compute the mean resultant length $\bar{R}_w$ according to its weight where:

$$\bar{R}_w = (\bar{C}_w^2 + \bar{S}_w^2)^{\frac{1}{2}}, \quad \bar{C}_w = \sum_{i=1}^{n} \frac{w_i \cos \vartheta_i}{\sum_{i=1}^{n} w_i},$$

$$\bar{S}_w = \sum_{i=1}^{n} \frac{w_i \sin \vartheta_i}{\sum_{i=1}^{n} w_i}$$

Then, estimate the concentration parameter k according to the following approximate formula:

$$\hat{k} = A^{-1}(\bar{R}_w) = [2(1-\bar{R}_w) + (1-\bar{R}_w)^2 (0.48798 - 0.82905\bar{R}_w - 1.3915\bar{R}_w^2)/\bar{R}_w]^{-1} \quad (5)$$

This approximation has an absolute error of <0.005 for $\bar{R}_w \ge 0.12$. Next, calculate new weights using Eq. 3. The iteration of this procedure continues until get convergence. Finally, the circular location parameter is estimated by $\cos\hat{\mu} = \bar{C}_w/\bar{R}_w$, $\sin\hat{\mu} = \bar{S}_w/\bar{R}_w$.

Trimmed mean direction: Trimmed mean is one of the robust methods that is used to estimate location parameter in linear data. It is calculated by eliminating a proportion of the largest and smallest values where the proportion of eliminating $\bullet$, $\bullet$ [0, 0.5) (Maronna *et al*., 2006). The main concept of circular data there is no maximum or minimum because $0 = 2\bullet$. So, statistical analysis that are used in linear data can not be applied for circular data because of the circular geometry theory. Therefore, cannot trim the data set according to its largest and smallest values. Therefore in this study, it is tried to propose a robust formula for trimming. It is expected that outliers lie far away from the circular mean (Mardia and Jupp, 2000). However, the circular median is more efficient than the circular mean when the circular data have outliers (Ducharme and Milasevic, 1987). Therefore, it is proposed to consider the circular distance between observations and the circular median as a measure to trim the circular data. Mahmood *et al.* (2017) proposed RCDu statistic to detect outliers in the univariate circular data. They suggested to calculate the circular distance $dist_{(i)}$ as follows: If ($0\bullet$ med$\bullet\bullet$):

$$dist_{(i)} = \begin{cases} |\vartheta_i \text{ med}| & \text{if } |\vartheta_i\text{-med}| \le \pi \\ 2*\pi-\vartheta_i+\text{med} & \text{if } |\vartheta_i\text{-med}| > \pi \end{cases} \quad (6)$$

If (• <med• 2 •):

$$dist_{(i)} = \begin{cases} |\vartheta_i \text{ med}| & \text{if } |\vartheta_i\text{-med}| \le \pi \\ 2*\pi\text{-med}+\vartheta_i & \text{if } |\vartheta_i\text{-med}| > \pi \end{cases} \quad (7)$$

the cut-off point is given as cut RCDu = max (dist). Hence, it is proposed to trim any circular data point has distance greater than the cutoff point.

RESULTS AND DISCUSSION

Simulation: In this study, simulation studies have been conducted to examine the performance of trimmed mean direction according to the proposed method for trimming. The data are simulated from von Mises distribution with mean direction 0 and five values of the concentration parameter (k = 2, 4, 6, 8 and 10) for two sample sizes n = 20 and 60 and. Three different cases are examined, clean data (without outliers), 5 and 10% of contamination. The circular observation • has been contaminated according to the following formula:

$$\vartheta_c = \vartheta+\lambda\pi \bmod(2\pi) \quad (8)$$

where • is the degree of contamination (0• • •1). Two measures are considered to test the performance. Bias $(\text{Estimated bias} = |\bar{\mu}\text{-}\mu|)$, circular mean deviation $(\text{CMD} = \pi\text{-}\frac{1}{n}\sum_{i=1}^{n}|\pi\text{-}|\vartheta_i\text{-}\mu|.$ The results are compared with mean direction, median direction and M estimator. The simulation is repeated 10000 times. The results for n = 20 and 60 are showed in Fig. 1 and 2, respectively. The results showed that the circular location parameters are unbiased except M estimator and there is no difference among CMD for the clean data. However, the mean direction, median direction and M estimator are biased of the estimation and CMD with presence of outliers and it is increasing with ratio of contamination. In contrast, the trimmed mean direction is unbiased of estimation and CMD with k>3 for all cases. Hence, the proposed formula for trimming is very successful to estimate circular location parameter for clean data and with presence of outliers in univariate circular data (Table 1).

Example: Data of direction of 14 frogs were collected from the mud flats of an abandoned stream meander near Indianola, Mississippi. This circular data has been tested

Table 1: Comparing measures of circular location parameters and CMD (frog data)

Parameters	Estimation	CMD
Mean direction		
Original data (w. outlier)	2.55	0.653
Modified data (w.o. outlier)	2.53	0.471
Median direction		
Original data (w. outlier)	2.45	0.651
Modified data (w.o. outlier)	2.37	0.479
M estimator		
Original data (w. outlier)	2.54	0.651
Modified data (w.o. outlier)	2.51	0.470
Trimmed mean direction		
Original data (w. outlier)	2.53	0.471
Modified data (w.o. outlier)	2.53	0.471

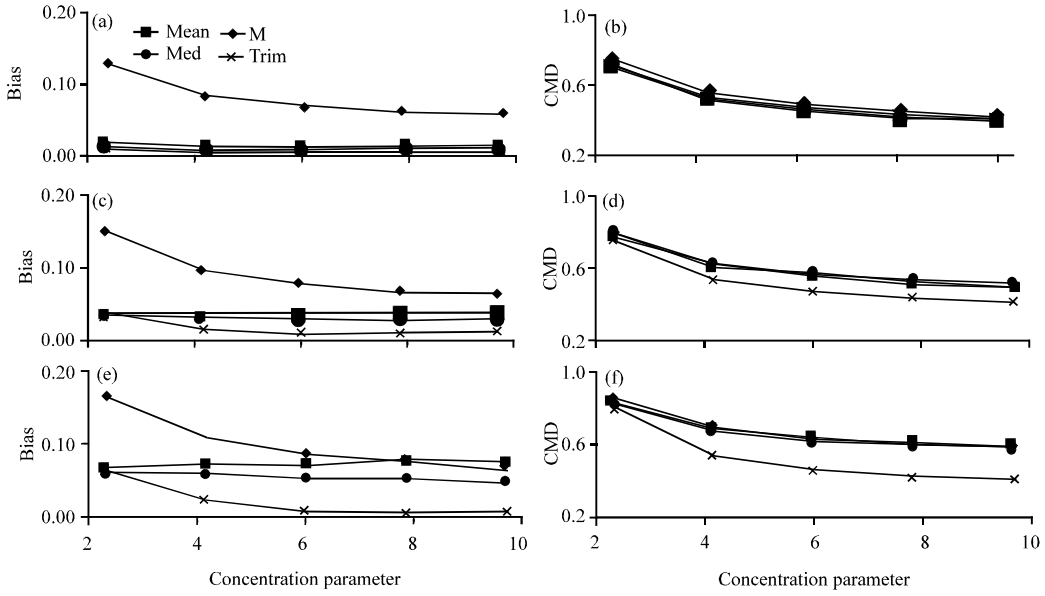

Fig. 1a-d): Bias and CMD of clean and contaminated with (5 and 10%) n = 20

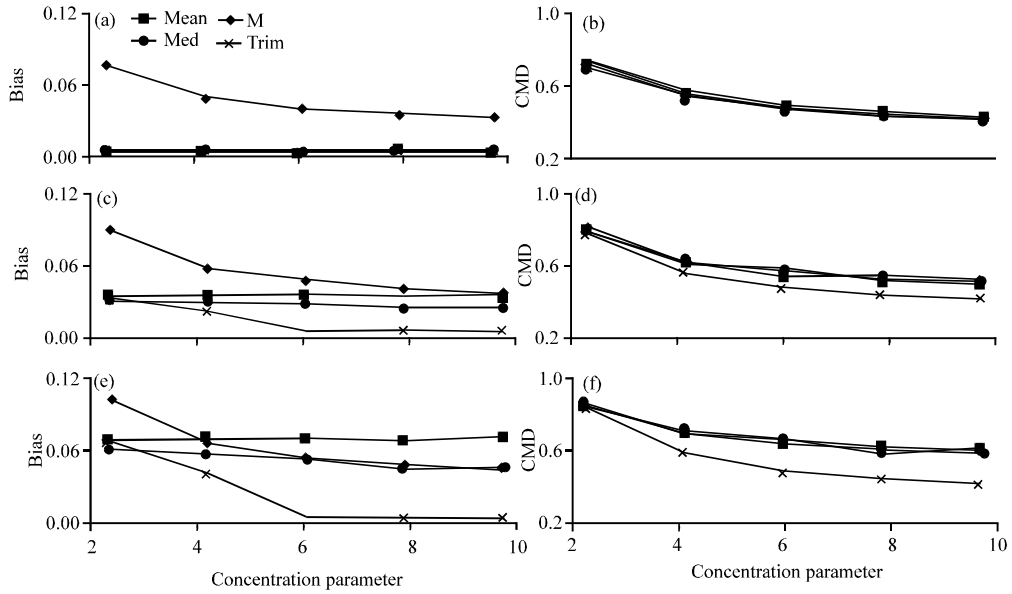

Fig. 2a-d): Bias and CMD of clean and contaminated with (5 and 10%) n = 60

by Collett (1980). He detected that, the observation numbered 14 is an outlier. The cut-off point 5% with concentration parameter $\hat{k} = 2.18$ is approximately 2.86. The estimation of mean direction with outlier, median direction, M estimator, mean direction without outlier and trimmed mean direction and CMD of them are given in Table 1. It is noticed that a significant difference between the results because of the effect of outlier. Besides, the results of Trim as the same as the results after delete outlier.

CONCLUSION

It is common in practice that circular data is contaminated with outliers. In the presence of outliers, the mean direction does not provide consistent and efficient estimator for the circular location parameter. To overcome this problem, some methods of detection outliers and robust methods have been proposed. In this study, it has suggested a robust formula for trimming to calculate trimmed mean direction. To examine the performance of the trimmed mean direction, it has conducted simulation studies for clean and contaminated data as well as it has applied for real data set. It has compared the results of mean direction, median direction and M estimator with the proposed method. It is found that the trimmed mean direction is very successful for all cases and it gives the best results according to the measures that paper depends on them.

REFERENCES

Abuzaid, A.H., 2012. Analysis of mother's day celebration via circular statistics. Philippine Statistician, 61: 39-52.

Barnett, V. and T. Lewis, 1978. Outliers in Statistical Data. John Wiley & Sons, Hoboken, New Jersey, USA., ISBN:9780471995999, Pages: 365.

Collett, D., 1980. Outliers in circular data. J. Royal Stat. Soc. Ser. C (Applied 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.

He, X. and D.G. Simpson, 1992. Robust direction estimation. Ann. Stat., 20: 351-369.

Jammalamadaka, S.R. and A. SenGupta, 2001. Topics in Circular Statistics. World Scientific Publishing Company, Singapore, ISBN-13: 978-9810237783, Pages: 350.

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 M. KC, 2015. Robustness of tests for directional mean. Stat., 49: 522-536.

Lenth, R.V., 1981. Robust measures of location for directional data. Technometrics, 23: 77-81.

Mahmood, E.A., S. Rana, H. Midi and A.G. Hussin, 2017. Detection of outliers in univariate circular data using robust circular distance. J. Mod. Appl. Stat. Methods, 16: 418-438.

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.