Thailand Statistician
July 2006; 4: 3-12
http://statassoc.or.th
Contributed paper

Quasi Median on the Circle

Sauwanit Ratanaruamkarn
Department of Mathematics and Statistics, Faculty of Science and Technology, Surindra Rajabhat University, Surin, Thailand.
E-mail: sauwanit@srru.ac.th

Received: 1 August 2006
Accepted: 6 September 2006

Abstract

The specific properties of probability distributions on a circle require different definitions of several statistical concepts. For example, a median that can always be found for linear data not always exists on the circle. Several estimators of a circular median were proposed, lately by Otieno [6], Otieno and Anderson - Cook [7]. New median estimator was proposed (CQM) and it properties were studied. A simulation study was performed to study their properties in small samples from unimodal symmetric, and skewed distributions without and with contamination.

Keywords: circular median, circular variance, directional data, Mardia median, new median, von Mises distribution.

1. Introduction

Estimation of location parameters on a circle is different from those on the line. For example an average of 2 angles θ_1 = 15º and θ_2 = 345º is not 180º which points the wrong way, but it is 0º which obtained by applying vector algebra or trigonometry. Median on a circle is also different from those on the line. Mardia ([4], page 28) defined median on a circle as point P of any given points on the unit circle which has the properties that (1) half of the sample points are on each side of the diameter PQ through P and (2) the majority of the sample points are nearer to P than to Q, is called a median of the sample. For an odd size, the median is an actual observation while for an even size the median is the midpoint (circular mean) of the two consecutive observations. Such definition of a median does not guarantee uniqueness and therefore Mardia ([4], page 31) proves that the circular mean absolute deviation is minimum for median direction. We will refer to the median direction with the minimum circular mean absolute deviation (CMAD) as a Mardia median (MM). Otieno and Anderson-Cook [7] proposed a new median (NM) estimator defined as a circular mean of all candidate medians satisfy (1) and (2) above. Also Otieno [6] proposed three versions of a Hodges-Lehmann estimator on a circle which is analogues to the HL estimator on a line. We focus our attention on location parameter on the circle for symmetric von Mises distribution.

2. Quasi Median (Linear case)

This is a brief synopsis of quasi median techniques in the linear case. For more detail, the reader is referred to Hodges and Lehmann [2] or Cabrera, Maguluri and Singh [1].

Given a set of observations $y_1 < y_2 < \ldots < y_n$ are the order observations, Quasi Medians as an average of two order statistics of the same order when counting is done from both end:

$$M_r = \begin{cases} (Y_{m+1-r} + Y_{m+1+r})/2, n = 2m+1 \\ (Y_{m-r} + Y_{m+1+r})/2, n = 2m, \end{cases}$$

r is assumed fixed and 2r < n.

$$Var(M_r) = \begin{cases} \dfrac{1}{4f^2 n} - \dfrac{1}{16n^2 f^2}(g + 8r + 8), n = odd \\ \dfrac{1}{4f^2 n} - \dfrac{1}{16n^2 f^2}(g + 8r + 12), n = even \end{cases} \qquad (1)$$

Where $g = f'(0)/f^3(0)$, $f = f(0)$, f is a probability density.

3. Central Quasi Median for directional data

We observed that sample Mardia median for the small even size (n = 2m) is has almost the same circular variance as for the small odd size (n = 2m+1). As shown in Figure 1.

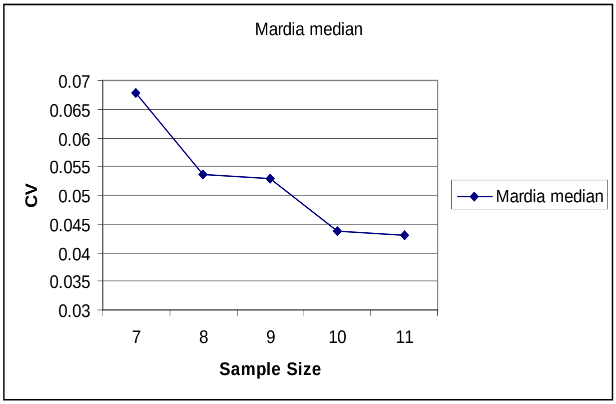
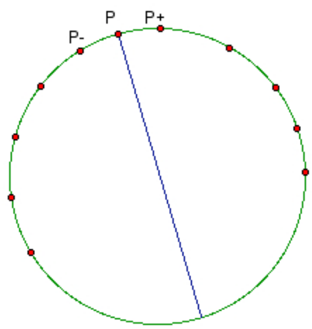


Figure 1. vM(0, 2) Based on 10,000 Repetitions

A slight modification is suggested, which eliminates this property, we proposed central quasi median as the linear combination of the two most central order statistics:



Suppose P is the Mardia median. P- and P+ are the observations closest to Mardia median on the left and right. The circ.mean (P-, P+) is the Central Quasi Median (CQM).This CQM alters the definition of the median very little. Retains the property of the median as a “middle point” of the data set, and the variance decrease with increasing sample size. As shown in Figure 2.

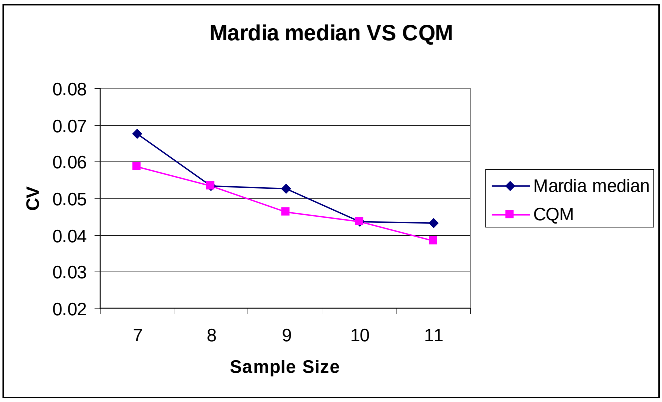


Figure 2. vM(0, 2) Based on 10,000 Repetitions, Small Sample Size

From figure 2, CQM has lower variance compare to Mardia median for small sample size.

4. Asymptotically Relative Efficiency for Large κ

Suppose α ~ vM(μ, κ). If β = $\sqrt{\kappa}$(α - μ). Then as $\kappa \to \infty$, β → N(0, 1) in distribution or for large κ , α ~ vM(μ, κ) can be approximated by ~ N(μ, 1/κ) [3]. This approximation is quite accurate for κ > 10 [5]. We will use the normal approximation to assess asymptotic variance for CQM. Suppose $\theta \sim N(\mu, \frac{1}{\kappa})$

$$f(\theta) = \frac{\sqrt{\kappa}}{\sqrt{2\pi}} e^{\frac{-\kappa(\theta-\mu)^2}{2}} \quad , \quad f''(\theta) = \frac{-\kappa^{\frac{3}{2}}}{\sqrt{2\pi}} e^{-\frac{\kappa(\theta-\mu)^2}{2}} [1-\kappa(\theta-\mu)^2] \quad ,$$

$$f^2(0) = \frac{\kappa}{2\pi} e^{-\frac{\kappa\mu^2}{2}}$$

$$f^3(0) = \frac{\kappa^{\frac{3}{2}}}{(2\pi)^{\frac{3}{2}}} e^{\frac{-3\kappa\mu^2}{2}} \text{, and g = } \frac{f''(0)}{f^3(0)} = -2\pi e^{\kappa\mu^2}[1-\kappa\mu^2]$$

And if $\theta \sim N(0, \frac{1}{\kappa})$ then

$$f(\theta)=\frac{\sqrt{\kappa}}{\sqrt{2\pi}}e^{\frac{-\kappa\theta^2}{2}}, \quad f''(\theta)=\frac{-\kappa^{\frac{3}{2}}}{\sqrt{2\pi}}e^{-\frac{\kappa\theta^2}{2}}[1-\kappa\theta^2],$$

$$f''(0)=\frac{-\kappa^{\frac{3}{2}}}{\sqrt{2\pi}} \tag{2}$$

$$f^2(0)=\frac{\kappa}{2\pi}, \tag{3}$$

$$f^3(0)=\frac{\kappa^{\frac{3}{2}}}{(2\pi)^{\frac{3}{2}}}, \tag{4}$$

and $$g = \frac{f''(0)}{f^3(0)}=-2\pi \tag{5}$$

Substitute (3) and (5) in (1) we get

$$MSE(\hat{\theta}_{CQM})=\begin{cases}\frac{\pi}{2n\kappa}\left(1-\frac{1}{4n}(-2\pi+8r+8)\right), n=odd\\ \frac{\pi}{2n\kappa}\left(1-\frac{1}{4n}(-2\pi+8r+12)\right), n=even\end{cases}$$

*Formula for ARE of circular mean, circular median (MM) and circular Hodges and Lehmann (HL2) were provided in Otieno [6].

Therefore $ARE_\theta(T_1,T_2) = MSE_\theta(T_2)/MSE_\theta(T_1)$

Clearly, T_1 is preferred over T_2 if $ARE_\theta(T_1,T_2) > 1$ with strictly inequality for at least one θ, or

$$ARE(\tilde{\theta}_{CQM},\hat{\theta}_{mean}) = \frac{MSE(\hat{\theta}_{mean})}{MSE(\tilde{\theta}_{CQM})}$$

$$=\begin{cases}\dfrac{\frac{1}{n\rho\kappa}}{\frac{\pi}{2n\kappa}(1-\frac{1}{4n}(-2\pi+8r+8))} & \text{,if n = odd}\\[2ex] \dfrac{\frac{1}{n\rho\kappa}}{\frac{\pi}{2n\kappa}(1-\frac{1}{4n}(-2\pi+8r+12))} & \text{,if n = even}\end{cases}$$

$$= \frac{2}{\rho\pi}$$

Since n $\to \infty$ and $\kappa \to \infty$, so $\rho = e^{\frac{-1}{2\kappa}} \to 1$, we obtain ARE $(\tilde{\theta}_{CQM}, \hat{\theta}_{mean}) = \frac{2}{\pi}$, therefore mean has more efficiency compare to CQM.

$$\text{ARE}(\tilde{\theta}_{CQM}, \hat{\theta}_{MM}) = \frac{MSE(\hat{\theta}_{MM})}{MSE(\tilde{\theta}_{CQM})}$$

$$= \begin{cases} \dfrac{\dfrac{\pi}{2n\kappa(1-e^{-2\kappa})^2}}{\dfrac{\pi}{2n\kappa}(1-\dfrac{1}{4n}(-2\pi+8r+8))} & \text{if n = odd} \\ \dfrac{\dfrac{\pi}{2n\kappa(1-e^{-2\kappa})^2}}{\dfrac{\pi}{2n\kappa}(1-\dfrac{1}{4n}(-2\pi+8r+12))} & \text{if n = even} \end{cases}$$

$$= \frac{1}{(1-e^{-2\kappa})^2}$$

Since n $\to \infty$ and $\kappa \to \infty$, we obtain ARE $(\tilde{\theta}_{CQM}, \hat{\theta}_{MM})$ = 1 or CQM has the same efficiency compare to MM.

$$\text{ARE}(\tilde{\theta}_{CQM}, \hat{\theta}_{HL2}) = \frac{MSE(\hat{\theta}_{HL2})}{MSE(\tilde{\theta}_{CQM})}$$

$$= \begin{cases} \dfrac{\dfrac{\pi}{3n\kappa}}{\dfrac{\pi}{2n\kappa}(1-\dfrac{1}{4n}(-2\pi+8r+8))} & \text{if n = odd} \\ \dfrac{\dfrac{\pi}{3n\kappa}}{\dfrac{\pi}{2n\kappa}(1-\dfrac{1}{4n}(-2\pi+8r+12))} & \text{if n = even} \end{cases}$$

= 2/3

Since n $\to\infty$, an even as well as odd sample size. We obtain ARE $(\tilde{\theta}_{CQM}, \hat{\theta}_{HL2})$ = 0.667, therefore HL2 has more efficiency compare to CQM.

5. Measure of Variation on the Circle

Three measures of variability are used to evaluate the performance of our estimators on the circle:

1. Circular Variance (CV) = 1 –r where r is the length of the mean vector and 0 < CV < 1.
2. Circular mean absolute deviation (CMAD)

CMAD(θ_0) = $\pi - \frac{1}{n}\sum_{i=1}^{n} | \pi - | \theta_i - \theta_0 ||$ where $\theta_1,\ldots,\theta_n$ are data points and θ_0 is the estimate of the preferred direction.

3. Circular median absolute deviation (CMEAD)

CMEAD(θ_0) = median($\pi - | \pi - | \theta_i - \theta_0 ||$)where $\theta_1,\ldots,\theta_n$ are data points and θ_0 is the estimate of the preferred direction.

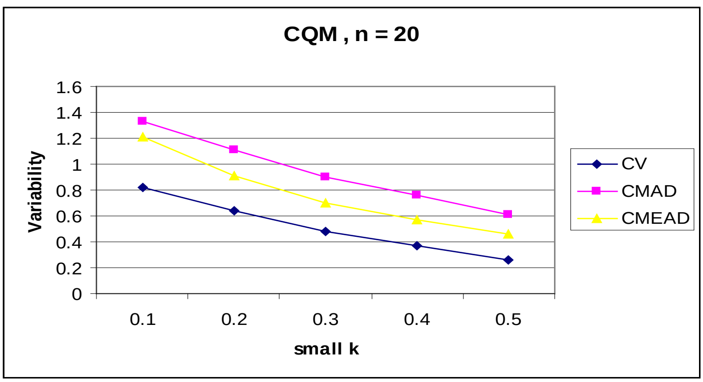


Figure 3. All Variability has the Same Trend Based on 10,000 Repetitions and Sample Size = 20.

From Figure 3, CV < CMAD < CMEAD.

6. Simulations Study

The CV's trends are similar for new median, Mardia median, CQM and HL2 for small and moderate k, for large k: HL2 has lowest CV while CQM is sandwiched between HL2 and NM, MM. For large k, MM=NM and have highest CV.

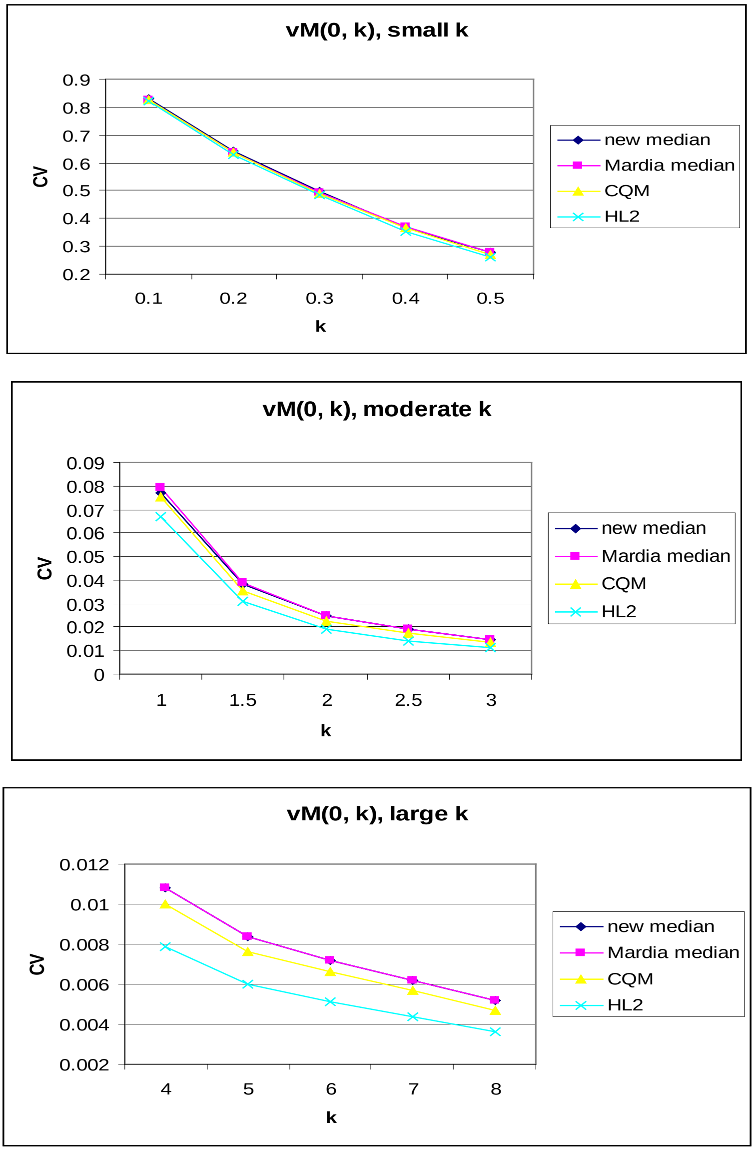


Figure 4. Compare CV for vM(0, κ), Sample Size = 19 Based on 10,000 Repetitions.

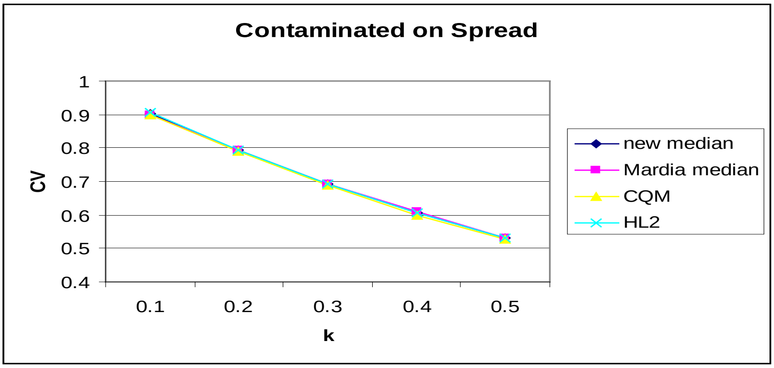


Figure 5. Effect of κ for Contaminated on Spread:(1-ε)vM(0, k)+ εU(-π, π), ε = 0.1, Sample Size = 7, Based on 10,000 Repetitions.

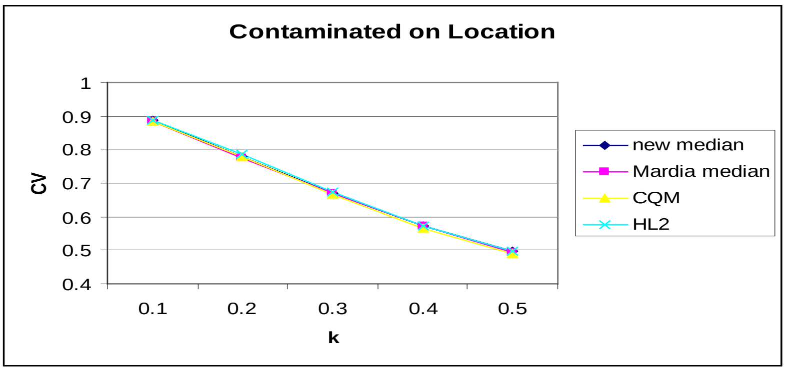


Figure 6: Effect of κ for Contaminated on Location: (1-ε) vM(0, k)+ ε vM(π/4, k), ε = 0.1, Sample Size = 7, Based on 10,000 Repetitions.

7. Summary

The CQM was proposed to improve existing estimators of a population median. Existing estimators either have a small variation (HL2) but require extensive computations or computationally simple (MM, NM) but have larger variability. Also properties of existing estimator were not studied in case of small samples and relatively small concentration of distribution. We cover variety of cases – small or different sample size, a whole range of concentration parameters for symmetric von Mises distribution.

Acknowledgement

I would like to thank my advisor, Dr. Magdalena Niewiadomska-Bugaj, Department of Statistics, Western Michigan University, Kalamazoo, MI 49008-1075 USA, whose continue advice and support made this paper possible.

References

[1] Cabrera, J., Maguluri, G., & Singh, K. An Odd Property of the Sample Median. *Statist. Probab.Lett*. **19**;1994: 349-354.

[2] Hodges, J.L.Jr., & Lehmann, E.L. On Medians and Quasi Medians. *Journal of the American Statistical Association.* **62**;1967: 926-931.

[3] Jammalamadaka, S. R., & SenGupta, A. Topics in Circular Statistics. World Scientific. New Jersey. 2001.

[4] Mardia, K.V. Statistics of Directional Data. London: Academic Press. 1972.

[5] Mardia, K.V., & Jupp, P.E. Directional Statistics. Wiley. ChiChester. 2000.

[6] Otieno, B. S. An Alternative Estimate of Preferred Direction for Circular Data. Ph.D. Dissertation, Dept. of Statistics, Virginia Polytechnic Institute and State University, Blacksburg Virginia. 2002.

[7] Otieno, B. S., & Anderson-Cook, C.M. A More Efficient Way of Obtaining a Unique Median Estimate for Circular Data. *Journal of Modern Applied Statistical Methods.* **2**;2003: 168-176.