DigitalCommons @WayneState

Journal of Modern Applied Statistical Methods

Volume 2 | Issue 1 Article 15

5-1-2003

A More Efficient Way Of Obtaining A Unique Median Estimate For Circular Data

B. Sango Otieno
Virginia Tech University, sango@vt.edu

Christine M. Anderson-Cook
Virginia Tech University, candcook@vt.edu

Journal of Modern Applied Statistical Methods
May, 2003, Vol. 2, No 1, 168-176

1538 – 9472/03/$30.00

A More Efficient Way Of Obtaining A Unique Median Estimate For Circular Data

B. Sango Otieno
Department of Statistics
Virginia Tech

C. M. Anderson-Cook
Department of Statistics
Virginia Tech

The procedure for computing the sample circular median occasionally leads to a non-unique estimate of the population circular median, since there can sometimes be two or more diameters that divide data equally and have the same circular mean deviation. A modification in the computation of the sample median is suggested, which not only eliminates this non-uniqueness problem, but is computationally easier and faster to work with than the existing alternative.

Key words: Preferred direction, circular median, uniqueness, robustness, local averaging

Introduction

Two common choices for summarizing the preferred direction are the mean direction and the median direction. (Fisher 1993, p. 30-36). The notion of preferred direction in circular data is analogous to the "center" of a distribution for data on a linear scale. The sample mean direction is frequently preferred for moderately large samples, because when combined with a measure of sample dispersion, it acts as a summary of the data suitable for comparison and amalgamation with other such information. An alternative, the sample median, can be thought of as balancing the number of observations on two halves of the circle.

Because there is no natural preferred direction for data that are uniformly distributed around the circle, it is natural and desirable that any measures of preferred direction are undefined if the sample data are equally spaced around the circle. In this paper, we consider estimating the preferred direction for a sample of unimodal circular data.

B. Sango Otieno is a Visiting Assistant Professor in the Department of Mathematics and Computer Science at The College of Wooster. E:mail: sango@vt.edu. C. M. Anderson-Cook is an Associate Professor in the Department of Statistics at Virginia Tech. Email: candcook@vt.edu

Ko and Guttorp (1988) showed that for a very wide class of families of distributions on S^{p-1}, the mean has infinite standardized gross error sensitivity; i.e., the asymptotic effect of a small contamination can be large compared with the dispersion. Hence, for the purposes of robust estimation, it is desirable to have a version of the sample median for circular data. As a nonparametric and robust estimate for the preferred direction of a distribution, the circular median has a different character from the sample circular mean as illustrated by different breakdown properties.

The sample median direction $\hat{q}$ of angles $q_1, \ldots, q_n$ is defined to be the point P on the circumference of the circle that satisfies the following two properties: (a) The diameter PQ through P divides the circle into two semi-circles, each with an equal number of observed data points and, (b) the majority of the observed data is closer to P than to the anti-median Q, See Mardia (1972, p. 28-30) or Fisher (1993, p. 35-36), for further details. For odd size samples, the medium is an observation, while for even sized samples, the median is the midpoint of two adjacent observations. Observations directly opposite each other do not contribute to the preferred direction, since these observations balance each other for all possible choices of medians. The procedure for finding the circular median has the flexibility to find a balancing point for situations involving ties, by mimicking the midranking idea for linear data. Potential median values are shown in Figure 1. For even samples, the candidate values are the midpoints of all neighboring observations, as shown in Figure 1a. For odd samples, the candidate values are the observations themselves, as in Figure 1b.

The circular median is rotationally invariant as shown by Ackermann (1997). Lenth (1981), and, Wehrly and Shine (1981) studied the robustness properties of both the circular mean and median using influence curves, and revealed that the circular mean is quite robust, in contrast to the sample mean on the real line. Durcharme and Milasevic (1987), show that in the presence of outliers, the circular median is more efficient than the mean direction. Many authors, including He and Simpson (1992), advocate the use of circular median as an estimate of preferred direction especially in situations where the data are not from the von Mises distribution.

A strategy to deal with non-unique circular median estimates is desired, especially for small samples, which are commonly encountered in circular data as is the case described below.

Consider the Frog data, given in Table 1 and shown in Figure 2, which relates the homing ability of Northern cricket frog, Acris crepitans, (Ferguson, et. al., 1967). For this data set, it is thought that the preferred direction for the population is 121^0 (where 0^0 is taken to be true North, and angles are measured in a clockwise direction), Collett (1980). The sample appears to be consist of a single modal group, with one observation which can be classified as an outlier. We wish to obtain the median as the point estimate of the preferred direction.

Notice that diameters P_1Q_1 and P_2Q_2 both divide the data evenly between the two semicircles, and hence both $P_1(133^0)$ and $P_2(140.5^0)$ satisfy the definition of a circular median. This implies that the median for this data set is not unique. A method for dealing with this non-uniqueness is the focus of this paper.

Methodology

To find a unique estimate of median, it is suggested to select the angle satisfying the median definition, such that it has the smallest circular mean deviation (Fisher, 1993, p. 35-36). The circular mean deviation is given by $d(\tilde{q}) = p - \frac{1}{n}\sum_{i=1}^{n}\left| p - |q_i - \tilde{q}| \right|$, where $\tilde{q}$ is the estimate of the preferred direction, and it is used as a measure of dispersion. Computing the circular median proposed by Mardia (1972, p. 28,31), henceforth referred to as "Mardia Median", occasionally leads to a non-unique estimate of the circular median since there can sometimes be two or more diameters that divide the data equally and have the same circular mean deviation.

In this section, we adapt the existing definition of circular median and propose that the estimate of the population circular median be the average (circular mean) of all angles satisfying the definition of median. This gives a unique estimate of the median, henceforth referred to as "New Median".

For the Frog data above, P_1 (133^0) and P_2 (140.5^0) are the two candidate sample medians. That is, the point estimate of the preferred direction based on Mardia Median can be taken to be either $P_1(133^0)$ or $P_2(140.5^0)$, since both have equal circular mean deviation of 0.650759. However, based on the new procedure, the point P (136.75^0) in Figure 2 is the circular mean of the two sample medians (P_1 & P_2). We conjecture that P will be more robust to rounding and will be a unique estimate since it involves local averaging, Cabrera et.al. (1994). Note that in this procedure, we eliminate the step of computing the circular mean deviation of candidate medians.

However, it is important to point out that if we treat $P_1(133^0)$ and $P_2(140.5^0)$ as equally good choices of median, since they have the same circular mean deviation, the circular mean deviation of P (136.75^0) is also 0.650759, hence it is the unique median. S-Plus functions for computing the circular mean direction, the Mardia Median and the New Median are given in the Appendix.

Figure 1: Original Observation o, Potential Median p

Figure 1a: Even sample size

Figure 1b: Odd sample size

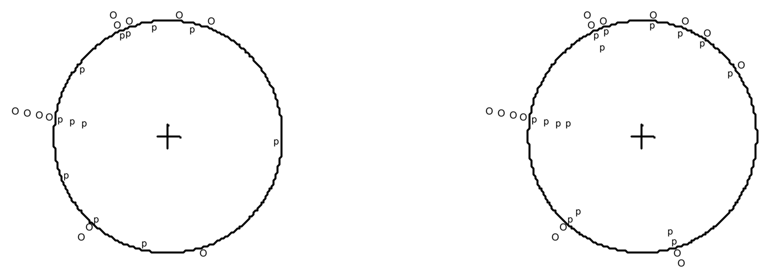
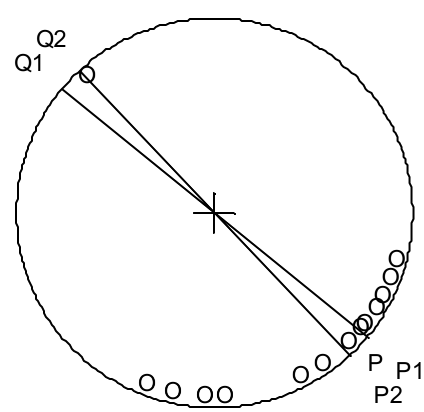


Table 1: Frog Data-Angles in degrees measured due North.

104	110	117	121	127	130	136
144	152	178	184	192	200	316

Figure 2: Homing Ability of Northern Cricket Frog



Results

Comparison of Mardia Median & New Median

To determine the relative performance of Mardia Median and the New Median, data was simulated from a von Mises (VM) distribution with probability density function $f(q) = [2pI_0(k)]^{-1} \exp[k \cos(q - m)]$, $0 \le q, m < 2p$ and $0 \le k < \infty$, Where m is the mean direction, k is the concentration parameter and

$$I_0(k) = (2p)^{-1} \int_0^{2p} \exp[kxos(f)]df = \sum_{j=0}^{\infty} - \frac{k^{2j}}{4^j j^2}$$

is the modified Bessel function of order zero.

Without loss of generality, the center of all the distributions considered was $m = 0$. Ten thousand samples each of sizes between 5 & 20 from the distributions with 6 dispersion values ranging from k =0.5 to 10 were obtained. The choice of sample size and dispersion values was based on the fact that non-uniqueness problems of the circular median are most common for small samples and large dispersions, so that is what we studied. For each sample, the sample circular medians (both Mardia Median and New Median) were computed.

The results were summarized using the following measures: 1) Circular mean $(\hat{m})$; and 2) circular variance $(1 - \hat{r})$ of the 10000 estimates obtained by solving the equations

$$\frac{1}{n}\sum_{i=1}^{n} \cos q_i = \hat{r}\cos(\hat{m}), \quad \frac{1}{n}\sum_{i=1}^{n} \sin q_i = \hat{r}\sin(\hat{m}),$$

where $\hat{r}$ is the sample resultant length; 3) the 95% Empirical Confidence Interval or the central 95% of the 10000 values; 4) Circular Mean Deviation (CMD) and 5) Circular Median Absolute Deviation (CMAD) given by $Median\left[\left|q_1 - \tilde{q}\right|, \ldots, \left|q_n - \tilde{q}\right|\right]$. Some of the simulation results are given in Tables 2 and 3.

Table 2, illustrates the effect of sample size on the two measures for $k = 2$. The measures appear unbiased, since the average of the point estimates is very close to zero, the true expected value. The confidence bands for the two medians are very similar and would be interchangeable for most required precision levels and become narrower as sample size increases for the two measures. The circular variances of the two medians, which could range between 1 for maximum variability to 0 for no variability, are consistently close over the whole range of sample sizes considered. Similarly, both the circular mean deviation (CMD), and the circular median absolute deviation (CMAD) are nearly the same for the two measures. These results were similar for other concentration parameters studied as well.

The effect of changing the concentration parameter on the two measures of preferred direction is illustrated in Table 3 for n = 20. Again, the two measures appear unbiased, and their confidence bands are very similar. The confidence bands become narrower as the concentration parameter increases for the two measures. The remaining measures for both medians are nearly identical for all possibilities. These results were similar for other sample sizes studied as well.

Note that computationally, the new procedure for obtaining the circular median is faster and simpler, since it eliminates the step of computing the circular mean deviation of each candidate median as opposed to Mardia Median. From the above results, we observe that the new procedure results in an estimate which minimizes the circular mean deviation relative to its counterpart, utilizing the benefits of local averaging.

Conclusion

For a fixed sample size and concentration, the Mardia Median and New Median give remarkably consistent results for all combinations of sample sizes and concentrations studied. Most strikingly, the two estimators, Mardia Median and New Median are approximately identical, which implies that either of the two can be used as an estimate of preferred direction. Computationally, the new measure is easier and faster to work with. Both Mardia Median and New Median are robust alternatives to the mean.

References

Ackermann, H. (1997). A note on circular nonparametrical classification. *Biometrical Journal, 5,* 577-587.

Cabrera, J., Maguluri, G. & Singh, K. (1994). An odd property of the sample median. *Statistics & Probability Letters 19,* 349-354.

Collett, D. (1980). Outliers in circular data. Applied Statistics, 29, 50-57.

Durcharme, G.R. & Milasevic, P. (1987). Some asymptotic properties of the circular median. *Communications in Statist. Theory and Methods, 16,* 163-169.

Ferguson, D.E., Landreth, H.F. & McKeown, J.P. (1967). Sun compass orientation of northern cricket frog, Acris crepitans. *Anim. Behav.,15,* 45-53.

Fisher, N.I. (1993). *Statistical analysis of circular data*. Cambridge University Press, Cambridge.

He, X. & Simpson, D.G. (1992). Robust direction estimation. *Annals of Statistics, 20, 1,* 351-369.

Ko, D. & Guttorp, (1988). Robustness of estimators for directional data. *Annals of Statistics, 16,* 609-618.

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

Mardia, K.V. (1972). *Statistics of Directional Data.* London: Academic Press.

Wehrly, T. & Shine, E.P. (1981). Influence curves for directional data. *Biometrika, 68*, 334-33.

Table 2. Mardia Median and New Median for VM(0, 2).

Sample Size	Measure	Point Estimate	Lower & Upper Confidence Limits	Circular Variance	Mean Deviation	Median Absolute Deviation
5	Mardia	0.001206	(-0.914198, 0.884683)	0.098107	0.559813	0.461589
	New	0.001347	(-0.913211, 0.889418)	0.098065	0.559152	0.461589
6	Mardia	-0.002618	(-0.77354, 0.790136)	0.075744	0.593154	0.484028
	New	-0.002350	(-0.774848, 0.787038)	0.075065	0.592542	0.484028
7	Mardia	0.004926	(-0.773052, 0.776042)	0.075079	0.597941	0.499424
	New	0.004867	(-0.771782, 0.778294)	0.075053	0.597611	0.499424
8	Mardia	-0.003863	(-0.700625, 0.658065)	0.059276	0.612813	0.507610
	New	-0.004103	(-0.699964, 0.65746)	0.058872	0.612625	0.507610
9	Mardia	-0.006341	(-0.69237, 0.673193)	0.059405	0.615008	0.515896
	New	-0.006230	(-0.693563, 0.668901)	0.059312	0.614815	0.515896

10	Mardia	-0.001831	(-0.62134, 0.631115)	0.049014	0.626990	0.524162
	New	-0.001734	(-0.619628, 0.631212)	0.048872	0.626892	0.524162
15	Mardia	0.000521	(-0.53107, 0.515293)	0.035605	0.641045	0.540889
	New	0.000580	(-0.531013, 0.515249)	0.03559	0.641003	0.540889
20	Mardia	0.000071	(-0.45413, 0.457305)	0.02582	0.651075	0.548252
	New	0.000010	(-0.453727, 0.455789)	0.025815	0.651067	0.548252

Table 3: Mardia Median and New Median for VM$(0, m)$, n = 20.

k	Measure	Point Estimate	Lower and Upper Confidence Limits	Circular Variance	Mean Deviation	Median Absolute Deviation
0.5	Mardia	-0.005483	(-1.796451 ,1.664871)	0.265584	1.189068	1.044356
	New	-0.010259	(-1.787609, 1.647442)	0.263658	1.178366	1.044356
1	Mardia	-0.002878	(-0.775017, 0.777624)	0.075995	0.959626	0.815823
	New	-0.003105	(-0.777569, 0.777397)	0. 076126	0.958215	0.815823
2	Mardia	0.000071	(-0.45413, 0.457305)	0. 02582	0.651075	0.548252
	New	0.000010	(-0.453727, 0.455789)	0. 025815	0.651067	0.548252
4	Mardia	-0.000058	(-0.296221, 0.285816)	0.010901	0.415821	0.350094
	New	-0.000058	(-0.296221, 0.285816)	0.010901	0.415821	0.350094
8	Mardia	0.000323	(-0.191746, 0.200085)	0.005015	0.280498	0.236698
	New	0.000323	(-0.191746, 0.200085)	0.005015	0.280498	0.236698
10	Mardia	-0.000812	(-0.176491, 0.169498)	0.003927	0.249753	0.211066
	New	-0.000812	(-0.176491, 0.169498)	0.003927	0.249753	0.211066

Appendix

A.1 cmed()

This function calculates circular median "New Median". It is a main program, one that the user will need to run. Input: data vector, x.

```
cmed<- function(x){
lenx <- length(x)
sx <- sort(x)
difsin <-c()
numties <-c()
if(lenx/2 == round(lenx/2)) {
# Checks if sample size is odd or even
# Computes median if sample size is even
posmed<- checkeven(x)
for(i in 1:length(posmed)) {
newx <- sx - posmed[i]
difsin[i] <-sum(round(sin(newx),10)> 0) - sum(round(sin(newx),10) < 0)
numties[i] <- sum(round(newx, 10) == 0)}
}
else
# Computes median if sample size is odd
posmed <- checkodd(x)
for(i in 1:length(posmed)) {
newx <- sx - posmed[i]
difsin[i] <- sum(round(sin(newx),10) >  0) - sum(round(sin(newx),10) < 0)
numties[i] <- sum(round(newx, 10)  == 0)}
}
# Checks for ties
cm <- c(posmed[round(difsin, 10)  == 0 | abs(difsin) >  numties])
circmed <- ave.ang(cm)
}
#takes into account if possible circmed are equidistant from mean
direction
circmed}
```

A.2 cmedM()

This function calculates Mardia Median. It is a main program, one that the user will need to run. Input: data vector, x.

```
cmedM <- function(x) {
lenx <- length(x)
sx <- sort(x)
sx2 <- c(sx[2:lenx], sx[1])
# Determines closest neighbors of a fixed observation
posmed <- rep(0, lenx)
difsin  <- rep(0, lenx)
numties <- rep(0, lenx)
med <- c()
if(lenx/2 == round(lenx/2)) {
\# Checks if sample is odd or even
posmed <- posmedf(x)
```

```
# Computes median if sample size is even
for(i in 1:length(posmed)) {
newx <- sx - posmed[i]
difsin[i]<- sum(round(sin(newx),10) > 0) - sum(round(sin(newx),10) < 0)
numties[i]<- sum(round(newx, 10)  == 0)}
}
else {
# Computes median if sample size is even
posmed <- checkodd(x)
for(i in 1:length(posmed)) {
newx <- sx - posmed[i]
difsin[i]<- sum(round(sin(newx),10) > 0) - sum(round(sin(newx),10) < 0)
numties[i]<- sum(round(newx, 10) == 0) }
}
# Checks for ties
cm <- c(posmed[round(difsin, 10) == 0 | round(abs(difsin),10) < numties])
for (i in 1:length(cm)) {
# Computes the circular mean deviation for candidate medians
med[i] <- meandev(x,cm[i]) }
circmed <- ave.ang(cm[round(med,10) == round(min(med),10)])
 }
 # Chooses the candidate medians with smallest circular mean deviations
and takes circular mean of them if more that one.
```

A. 3 ave.ang()

This function calculates circular mean direction. It is an internal function needed for the main programs. Input: data vector a.

```
ave.ang <- function(a) {
y <- sum(sin(a))
x <- sum(cos(a))
ifelse(round(x, 10) == 0 & round(y, 10) == 0, 9999, atan(y, x))}
# If both x and y are zero, then no circular mean exists, so assign it a
large number (9999).
```

A. 4 posmedf()

This function calculates all potential medians for even samples It is an internal function needed for the main programs. Input: data vector x.

```
posmedf <- function(x){
lenx <- length(x)
sx <- sort(x)
sx2 <- sx[c(2:lenx,1)]
# Determines closest neighbors of a fixed observation
posmed <- c()
for(i in 1:lenx) {
posmed[i]<- ave.ang(c(sx[i],sx2[i]))}
# Computes circular mean of two adjacent observations
posmed <- posmed[posmed ≠9999]
posmed }
```

A.5 checkeven()

This function checks if the number of possible medians is even. It is an internal function for the main programs. Input: data vector x.

```
checkeven<-function(x){
lenx <- length(x)
sx <- sort(x)
check <- c()
# Computes possible medians
posmed<- posmedf(x)
for(i in 1:length(posmed)){
#Takes posmed[i] as the center, i.e. draws diameter at posmed[i] and
counts observations on either side of the diameter
newx <-sx-posmed[i]
check[i]<-ifelse(sum(round(cos(newx),10)>0)<lenx/2, 9999,posmed[i])}
nposmed<- check[check≠ 9999]
nposmed }
```

A. 6 checkodd()

This function checks if the number of possible medians is odd. It is an internal function needed for the main programs. Input: data vector x.

```
checkodd <- function(x) {
lenx <- length(x)
sx <- sort(x)
check <- c()
posmed <- sx
# Each observation is a possible median
for (i in 1:length(posmed)) {
newx <- sx-posmed[i]
#Takes posmed[i] as the center, i.e. draws diameter at posmed[i] and
counts observations on either side of the diameter
check[i] <- ifelse(sum(cos(newx) > 0) > (lenx-1)/2, 9999,posmed[i]) }
nposmed <- check[check ≠9999]
nposmed }
```

A.7 meandev()

This function calculates circular mean deviation. It is an internal function needed for the main programs. Input: data vector x.

```
meandev <- function(x, teta) {
# Checks if circular mean exists
ifelse(teta == 9999, 9999, (pi - mean(round(abs(pi -
(abs(rangeang( x - teta)))), 10))))}
```

A.8 rangeang()

This function puts data in $\left(-p,p\right)$ range. It is an internal function needed for the main programs. Input: data vector x.

```
rangeang <-function(x) {
ang <-ifelse(x <  - pi, x + 2 * pi, x)
ang2<- ifelse(ang > pi, ang - 2 * pi, ang)
return(ang2)
```