Comput Stat (2012) 27:381–392
DOI 10.1007/s00180-011-0261-5

ORIGINAL PAPER

Boxplot for circular variables

Ali H. Abuzaid · Ibrahim B. Mohamed ·
Abdul G. Hussin

Received: 30 October 2008 / Accepted: 23 May 2011 / Published online: 23 June 2011

Abstract A boxplot is a simple and flexible graphical tool which has been widely used in exploratory data analysis. One of its main applications is to identify extreme values and outliers in a univariate data set. While the boxplot is useful for a real line data set, it is not suitable for a circular data set due to the fact that there is no natural ordering of circular observations. In this paper, we propose a boxplot version for a circular data set, called the circular boxplot. The problem of finding the appropriate circular boxplot criterion of the form $\nu \times CIQR$, where $CIQR$ is the circular interquartile range and ν is the resistant constant, is investigated through a simulation study. As might be expected, we find that the choice of ν depends on the value of the concentration parameter κ. Another simulation study is done to investigate the performance of the circular boxplot in detecting a single outlier. Our results show that the circular boxplot performs better when both the value of κ and the sample size are larger. We develop a visual display for the circular boxplot in S-Plus and illustrate its application using two real circular data sets.

Keywords Circular boxplot · Boxplot · Resistant constant · Outlier · Overlapping

1 Introduction

A visual display is a useful and an informative technique for describing a data set. It includes a histogram, a pie chart, a Q-Q plot and a boxplot. Tukey (1977) developed the boxplot, a simple and flexible graphical tool in exploratory data analysis. It entails

A. H. Abuzaid · I. B. Mohamed (✉)
Institute of Mathematical Sciences, University of Malaya, 50603 Kuala Lumpur, Malaysia
e-mail: imohamed@um.edu.my

A. G. Hussin
Center for Foundation Studies in Science, University of Malaya, 50603 Kuala Lumpur, Malaysia

the measurements of the smallest value, the lower quartile Q_1, the median ϕ, the upper quartile Q_3 and the largest value. One of its main applications is to identify extreme values and outliers in a univariate data set.

Extensive research has been conducted on the use of the boxplot in the labelling of outliers. To identify outliers in a real line data set, most studies use 1.5 as the value for the resistant constant ν in the boxplot criterion $\nu \times IQR$, where IQR is the interquartile range. In other words, any observations smaller than $Q_1 - 1.5 \times IQR$ or greater than $Q_3 + 1.5 \times IQR$ are labelled as "outliers". Hoaglin et al. (1986) investigated the performance of the boxplot for outlier labelling by considering different values of ν. The value $\nu = 1.5$ is considered to be the best choice in avoiding masking problems while $\nu = 3$ is considered to be extremely conservative. On the other hand, Ingelfinger et al. (1983) suggested the use of $\nu = 2$ while Sim et al. (2005) demonstrated that the choice of $\nu = 1.5$ or $\nu = 3$ was in general inappropriate for normal samples and was completely inappropriate for skewed distributions. The discussion above signifies the importance of choosing the most suitable value of ν for different data sets with different underlying distributions.

Our focus here has directions in 2-dimensions. In many diverse applications, data are measured in degrees or radians; for instance, wind directions and animal navigation. Such data are known as circular data. Fisher (1993) noted that circular plots have existed since 1858, when Florence Nightingale drew the plot of mortality data in the British Army during the Crimean War. Such a plot is also known as a rose diagram or wind rose diagram. In addition, Graedel (1977) used the boxplot to describe wind speed in different sectors of a wind rose diagram. However, in general, the boxplot is not directly applicable to a circular data set. In the literature, no special boxplot framework for a circular data set has been found. In this paper, we address the problem by proposing a boxplot for a circular data set which we will call a circular boxplot. We describe in detail the construction of this circular boxplot and develop subroutines in an S-Plus environment to display it. The circular boxplot can be used to detect outliers in circular samples. To date, there are several methods available to detect outliers in circular data sets. Collett (1980) presented four different numerical tests of discordance in circular data, namely the C, D and L statistics and an improved version of the M statistic originally proposed by Mardia (1975). Recently, Abuzaid et al. (2009) proposed the A statistic based on the summation of the circular distances from the point of interest to all other points while Abuzaid et al. (2008) used numerical and graphical tools to detect outliers in a circular regression model. However, the circular boxplot is simpler and more appealing compared to the other outlier detection techniques described above.

This paper is organized as follows. The next section discusses the proposed construction of the circular boxplot. Simulation and numerical studies are carried out in Sect. 3 to estimate the appropriate values of ν. Then, in Sect. 4, we investigate the power of performance of the circular boxplot for different values of ν and sample sizes. Two numerical examples are discussed in Sect. 5. The first example gives an application of the circular boxplot to the frogs' directions data set as given in Ferguson et al. (1967), whereas the second utilizes the plot to identify outliers in a circular regression based on the resulting circular residuals discussed in Hussin et al. (2004).

2 Summary statistics for constructing the circular boxplot

2.1 Median direction of circular data

Fisher (1993) defined the median direction of a circular data set as an axis which divides the data into two equal groups. The median direction is taken as the observation ϕ which minimizes the summation of circular distances

$$d(\phi) = \pi - \sum_{i=1}^{n} |\pi - |\theta_i - \phi||,$$

where θ_i is the circular observation ($i = 1, \ldots, n$). See also Jammalamadaka and SenGupta (2003) for this particular measure of circular distance. It follows that for any c, $(\theta_1 + c, \ldots, \theta_n + c)$ has a median direction $\phi + c$. Consequently, we show that

$$\begin{aligned} d(\phi + c) &= \pi - \sum_{i=1}^{n} |\pi - |\theta_i + c - \phi - c|| \\ &= \pi - \sum_{i=1}^{n} |\pi - |\theta_i - \phi|| = d(\phi). \end{aligned}$$

In other words, the median direction ϕ is rotationally equivariant. Mardia and Jupp (2000) defined the median direction for a set of circular observations $\theta_1, \ldots, \theta_n$ as any point ϕ where half of the data lie in the arc $[\phi, \phi + \pi)$ and the other points are nearer to ϕ than to $\phi + \pi$. In the case of prior knowledge about the circular distribution, Mardia (1972) defined the median direction ϕ as the solution of

$$\int_{\phi}^{\phi+\pi} f(\theta) d\theta = 0.5,$$

where $f(\theta)$ is the probability density function of θ. Here, we employ the definition proposed by Fisher (1993) to obtain the median direction.

2.2 Quartiles of circular variables

Mardia (1972) defined the first and third quartile directions Q_1 and Q_3 as any solution of

$$\int_{\phi-Q_1}^{\phi} f(\theta) d\theta = 0.25 \quad \text{and} \quad \int_{\phi}^{\phi+Q_3} f(\theta) d\theta = 0.25$$

respectively. In most cases, the circular distribution is unknown. To date, no published literature has been found on nonparametric estimations of Q_1 and Q_3 for circular data.

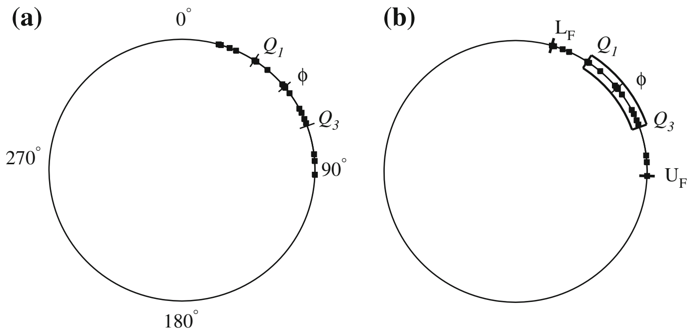

Fig. 1 The proposed structure of the *circular boxplot*

However, it seems sensible to estimate Q_1 and Q_3 by classifying the sample observations into two groups based on their locations with respect to the sample median direction. Subsequently, Q_1 can be considered as the median of the first group and Q_3 as the median of the second.

If the value of Q_1 is larger than that of Q_3, we simply interchange their labels. For simplicity and to avoid confusion caused by the localizations of Q_1 and Q_3, the rotatable property of circular data is used to ensure that the mean is in the zero direction, that is, we subtract the estimated mean direction of the circular sample from each observation in the sample. For a circular data set $\theta_1, \ldots, \theta_n$, the mean direction $\bar{\theta}$ can be calculated using the following equations:

$$\bar{\theta} = \begin{cases} \tan^{-1}(S/C) & \text{for } S \geq 0, \quad C > 0, \\ \tan^{-1}(S/C) + \pi & \text{for } C < 0, \\ \tan^{-1}(S/C) + 2\pi & \text{for } S < 0, \quad C \geq 0, \end{cases}$$

where $S = \sum_{i=1}^{n} \sin\theta_i$ and $C = \sum_{i=1}^{n} \cos\theta_i$.

The above rotation is helpful in identifying Q_1 and Q_3 in a more consistent way, that is, we can now assume that $Q_1 - \bar{\theta} \in [0, \pi]$ and $Q_3 - \bar{\theta} \in [\pi, 2\pi]$. The robustness of the mean direction is a useful property which gives a fair assurance that the existence of any possible outlier will not have much effect on the estimated mean direction (see Wehrly and Shine (1981)). Figure 1a shows the quartiles for the simulated circular data from the von Mises distribution with a mean direction of $\pi/4$ and a concentration parameter of $\kappa = 4$. The first quartile $Q_1 = 33°$, the median direction $\phi = 50°$ and the third quartile $Q_3 = 69°$.

2.3 Circular interquartile range and fences

Analogous to the interquartile range for the real line case, a circular interquartile range ($CIQR$) is required in order to construct the circular boxplot. After the rotation of sample observations, the $CIQR$ can be obtained by the following formula;

$$CIQR = 2\pi - Q_3 + Q_1.$$

Consequently, the lower and upper fences, L_F and U_F respectively, can be identified using $L_F = Q_1 + \nu \times CIQR$ and $U_F = Q_3 - \nu \times CIQR$. For a highly concentrated data set, it is possible to have quartiles and mean directions at the same point so that $CIQR = 0$.

Figure 1b illustrates a particular example of the proposed circular boxplot for a symmetric simulated circular data set. An example of a circular boxplot for asymmetric data is given in Fig. 5. In the following section, numerical and simulation studies are carried out in order to determine the appropriate values of ν.

3 Estimation of the resistant constant

3.1 Simulation and numerical studies

In the real line case, the $1.5 \times IQR$ criterion is commonly used to construct a boxplot. However, discussions on the appropriate values of the constant ν are still ongoing. For recent examples, see the work of Hoaglin et al. (1986); Ingelfinger et al. (1983) and Sim et al. (2005). In our case, it is not sensible to utilize the values of ν used for the linear boxplot due to the bounded range of the circle. This property gives rise to a high possibility of overlapping lower and upper fences for large values of ν and small values of κ.

Hoaglin et al. (1986) used different measures to investigate the behaviour of boxplots. In this paper, we employ one of their measurements to estimate the appropriate value of the constant ν based on the statistic $B(\nu, n)$. The statistic gives the probability that a sample of size n does not contain any observations outside the interval (L_F, U_F). Sim et al. (2005) also used the same statistic in their work.

In order to investigate the behaviour of circular variables with respect to five different summaries, namely, the median, Q_1, Q_3, L_F and U_F, a series of simulation studies are carried out. For each combination of sample size n and concentration parameter κ, 3,000 samples are generated from the von Mises distribution $VM(\mu, \kappa)$. The sample sizes are between 5 and 200 while the values of the concentration parameter κ considered are $0.5, 1, 2, \ldots, 10$. Furthermore, various values of the resistant constant $\nu = 1, 1.2, 1.4, \ldots, 3$ and 3.5 are utilized in order to obtain L_F and U_F. The outcomes of the simulation studies are $CIQR$, Q_1, Q_3, L_F, U_F and $B(\nu, n)$.

In the following subsections we investigate the properties of the $CIQR$, the problem of overlapping fences and the behaviour of the statistic $B(\nu, n)$.

3.2 Relationship between $CIQR$ and κ

The $CIQR$ can be estimated from the cumulative distribution function of any statistical distribution, and is defined as $CIQR = x_{75} - x_{25}$, where x_{25} and x_{75} are the solutions of $\int_0^{x_{25}} f(x)dx = 0.25$ and $\int_0^{x_{75}} f(x)dx = 0.75$ respectively. Thus the cumulative distribution function of any known distribution can be used to construct the circular boxplot, whereas for an unknown distribution or during the data exploration stage, nonparametric methods will be helpful in identifying the median direction and the $CIQR$. Fisher (1993) stated the fact that the measure of spread of any circular distribution cannot be rescaled to a unit spread. This causes difficulties as there is no standard von Mises distribution analogous to a standard normal distribution. Consequently, it is rather difficult to find a functional relationship for the $CIQR$.

However, we attempt to find a relationship between the $CIQR$ and κ by utilizing the approximation theory of the von Misses distribution for large values of κ. Jammalamadaka and SenGupta (2001) stated that, for large κ, a circular random variable θ from VM(μ,κ) is approximated by a normal distribution with mean μ and variance $1/\kappa$. On the other hand, Fox (1997) stated that for a random sample from a normal distribution with mean μ and variance σ^2, the interquartile range IQR can be estimated by 1.349σ. Hence, in our case, we may conclude that $CIQR = 1.349/\sqrt{\kappa}$ for large κ.

3.3 The problem of overlapping lower and upper fences

The above problem is expected to occur for some values of ν because of the bounded range of circular variables. This gives rise to a messy structure of the circular boxplot and may lead to misidentification of outliers.

Result 1 *For a large concentration parameter κ and sample size $n \geq 10$, the upper and lower fences of the circular boxplot are subject to overlapping if*

$$\nu > \pi\sqrt{\kappa}/1.349 - 0.5$$

where ν is the resistant constant.

Proof For a large concentration parameter κ and a large sample size n, an overlapping problem arises if $Q_1 + \nu \times CIQR > \pi$. For symmetric samples with median direction 0, $Q_1 = CIQR/2$. Therefore,

$$(0.5 + \nu)CIQR > \pi.$$

However, from Sect. 3.2, $CIQR = 1.349/\sqrt{\kappa}$. Thus,

$$1.349(0.5 + \nu)/\sqrt{\kappa} > \pi.$$

Hence, an overlapping problem occurs if $\nu > \pi\sqrt{\kappa}/1.349 - 0.5$. □

As an example, when $\kappa = 4$, an overlapping problem is expected to occur if ν is larger than 4.2. The relationship between the outliers and the concentration parameter is discussed in Sect. 4.

3.4 Description of $B(\nu, n)$

Let $B(\nu, n)$ denote the probability of no observations outside the interval (L_F, U_F) for a von Mises sample of size n and resistant constant ν. When $n = 50$, the overlapping problem affects the behaviour of $B(\nu, n)$ as shown in Fig. 2.

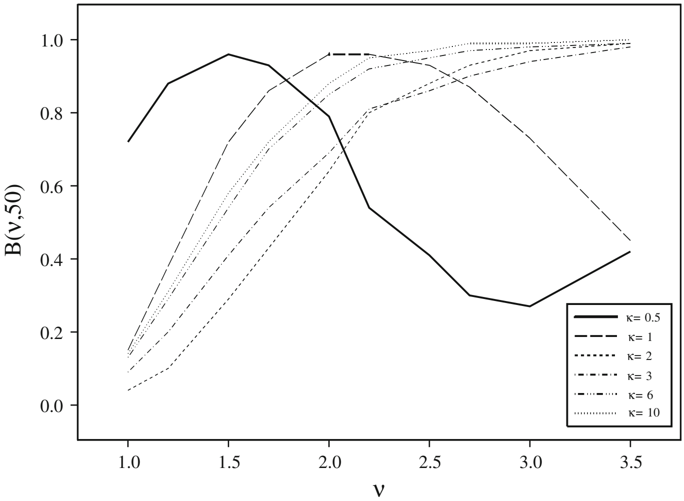

Fig. 2 Behavior of $B(\nu, 50)$ for simulated data

It is noticed that $B(\nu, 50)$ is a nonmonotone function of ν for $\kappa < 2$, while it is an increasing function of ν for $\kappa \geq 2$. In addition, for $\kappa > 3$, $B(\nu, 50)$ is only slightly affected by the increment of κ. Similar results are obtained for other values of n but are not shown here.

Simulation results of $B(\nu, n)$ are used to determine the values of ν. It is more informative to interpret the results of the simulation studies according to the *mod* of sample size n with respect to 4. Thus, the sample sizes can be clustered into one of 4 groups according to whether n has the form $4j, 4j + 1, 4j + 2$ or $4j + 3$, where $j \in \mathbf{N}$.

Figure 3 shows the values of ν for sample sizes $n < 56$ at 0.1, 0.05 and 0.01 significance levels and a large value of κ ($\kappa = 7$). Results for $n > 56$ are consistently similar to the results for $25 \leq n \leq 56$ for all significance levels, but are not shown here. It can be seen that the values of ν are a decreasing function of the significance level α. At 0.05 significance level, the values of ν seem to be stationary for $5 < n < 56$ with respect to the sample size n mod 4. The results suggest that it is appropriate to have $2.1 < \nu < 2.7$. A similar behaviour can be observed for $\alpha = 0.1$, provided $1.5 < \nu < 2.2$. The situation is different for $\alpha = 0.01$, where ν stays at 3.5 for $5 < n < 25$, and decreases to $\nu = 3$ for a larger sample size n. On the other hand, for $\kappa \leq 3$, the simulation results show that the convenient values of ν should be less than 2. In all cases, ν only increases slightly for a larger sample size. The results are not shown here.

4 Power of performance

The performance of discordance tests can be examined by using several measures. Barnett and Lewis (1978) stated that a good test should have (i) a high power, (ii) a high probability of identifying a contaminating value as an outlier when it is in fact an

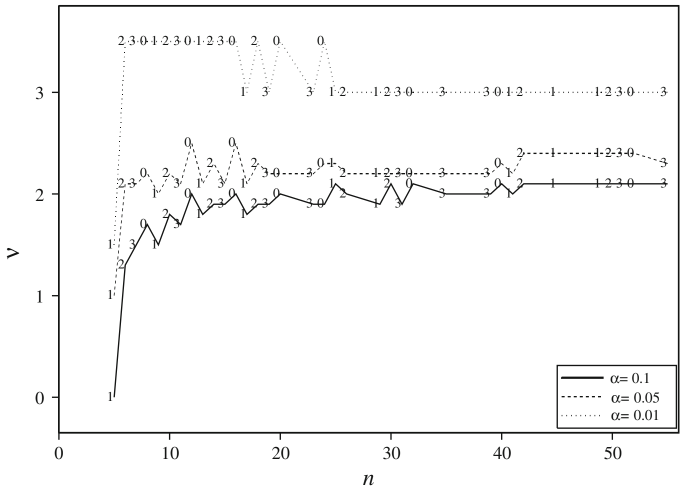

Fig. 3 Percentile points of the resistant constant ν for different sample sizes n and $\kappa = 7$

extreme value and (iii) a low probability of wrongly identifying a good observation as a discordant. David (1970) defined $P1$ as the power function, $P3$ as the probability of the contaminant point which is an extreme point but identified as a discordant, and $P5$ as the probability that the contaminant point is identified as a discordant given that it is an extreme point. Hence, a good test is expected to have (i) high $P1$, (ii) high $P5$ and (iii) low $P1 - P3$.

In order to study the performance of the circular boxplot, 3,000 samples based on different sample sizes $n = 5, 10, 15, 20, 60$ and 100, with concentration parameters $\kappa = 1, 5, 7$ and 10 are considered. Samples are obtained in such a way that (n–1) observations are generated from $VM(\alpha, \kappa)$ and the remaining one from $VM(\alpha + \lambda\pi, \kappa)$, where λ is the degree of contamination and $0 \leq \lambda \leq 1$.

Based on the simulation studies in Sect. 2.3, small values of ν $(1 \leq \nu \leq 2)$ are examined for a small concentration parameter ($\kappa = 1$), while larger values of ν $(2 \leq \nu \leq 2.7)$ are considered when κ is large ($\kappa = 5, 7$ and 10). The results show that $P1$, $P3$ and $P5$ are increasing functions of κ, n and λ.

Figure 4a gives the plot of $P1$ at $n = 60$ and $\nu = 2$ for different concentration levels. The power of performance of the circular boxplot is weaker for smaller values of κ. In fact, it is close to zero when $\kappa = 0.1$ or 0.5. This behaviour may be due, firstly, to the bounded range of the circle; and secondly, to the tendency of the observations to distribute themselves uniformly around the circumference of the circle when κ is small (see Fisher 1993). The tendency is stronger as κ gets closer to 0. This results in a weaker power of performance of the circular boxplot for $\kappa < 1$. Collett (1980) had similarly pointed out that it would be very difficult to identify an outlier in circular data sets with a small value of κ.

Figure 4b gives the plot of $P1$ at $\kappa = 10$ and $\nu = 2$ for different sample sizes. For a small sample size $n = 5$, the power of the circular boxplot does not exceed 50% at any κ or λ. It improves gradually as n increases.

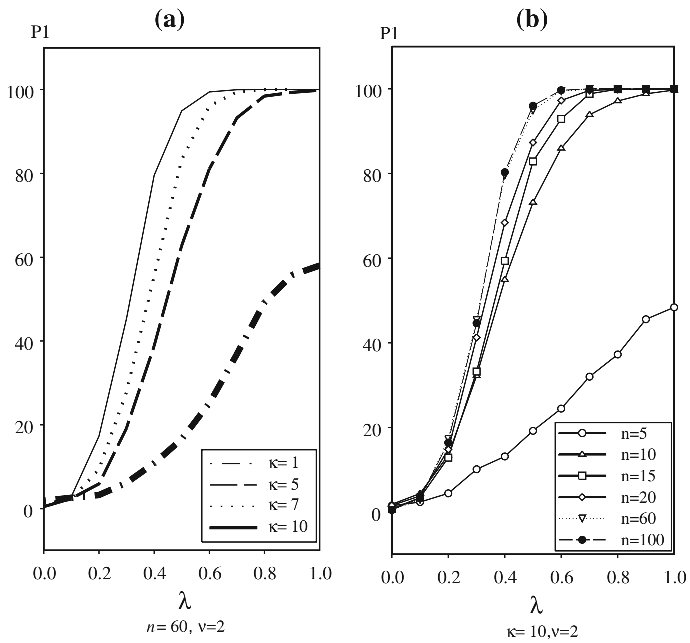

Fig. 4 Behaviour of the power of performance of the *circular boxplot*

Based on the discussions in the previous section, for a large κ, it is appropriate to use different values of ν between 2.0 and 2.7. Our studies show that the highest possible values of $P1$, $P3$ and $P5$ are obtained when $2.0 < \nu < 2.7$, which supports the results in Sect. 3.4. For a small κ, it seems sensible to use small values of ν $(1 < \nu < 2)$. Furthermore, it is found that $P1 - P3$ is always small and does not exceed 0.2% for all cases. Extensive simulation results are available from the authors upon request. For simplicity, we may fix $\nu = 1.5$ for $2 \leq \kappa \leq 3$, and $\nu = 2.5$ for $\kappa > 3$. These choices ensure that the problem of overlapping lower and upper fences of the circular boxplot does not occur. However, for $\kappa < 2$, it is rather difficult to find a circular boxplot with non-overlapping fences as the data are close to being uniform.

5 Numerical examples

Two examples are discussed in this section. The first is the frogs' directions data set by Collett (1980) and the second is the wind direction data set modelled by using a simple regression model for circular variables in Abuzaid et al. (2008).

Example 1 (*Frogs' directions data*) Ferguson et al. (1967) conducted an experiment to investigate the homing ability of a species of frogs. A total of 14 frogs was collected from the mud flats of an abandoned stream meandering near Indianola, Mississippi in the United States. After 30 h, the frogs were released and the directions taken by the frogs were recorded as follows:

$$104^\circ, 110^\circ, 117^\circ, 121^\circ, 127^\circ, 130^\circ, 136^\circ, 145^\circ, 152^\circ, 178^\circ, 184^\circ, 192^\circ, 200^\circ, 316^\circ.$$

Table 1 Observations detected using different values of ν for the frogs' data set

ν	L_F	U_F	Number of outliers	Outliers
1.0	50.0°	263.0°	1	316°
1.2	35.8°	277.2°	1	316°
1.5	14.5°	298.5°	1	316°
1.7	0.3°	312.7°	1	316°
2.0	339.0°	334.0°	0	–
2.2	317.7°	355.3°	0	–
2.5	303.5°	9.5°	0	–
2.7	289.3°	23.7°	0	–
3.0	268.0°	45.0°	0	–
3.5	232.5°	80.5°	0	–

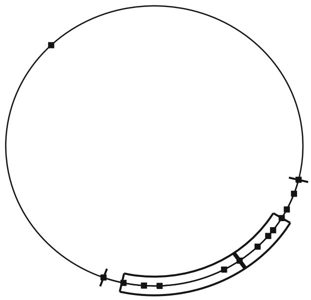

Fig. 5 *Circular boxplot* of the frogs' directions for $\nu = 1.5$

We find that the mean direction $\bar{\theta} = 145.97°$, the estimated concentration parameter $\hat{\kappa} = 2.18$, the median direction $\phi = 145°$, the first quartile $Q_1 = 121°$, the third quartile $Q_3 = 192°$ and the $CIQR = 71°$. Collett (1980) showed that the observation with value 316° was identified as an outlier by the D and M statistics, but not by the C statistic. Abuzaid et al. (2009) had also obtained the same result using the A statistic. For the construction of the circular boxplot, small values of ν should be considered since $\hat{\kappa}$ lies in the range [2,3].

Table 1 gives the observations detected as outliers in the frogs' directions data set for different values of ν. As expected, the value 316° is identified as an outlier when small values of $\nu = 1, 1.2, 1.5$ and 1.7 are used, whereas no outlier is identified for larger values of ν. Figure 5 shows the circular boxplot for $\nu = 1.5$, where the plot is obtained by using a special subroutine developed in an S-Plus environment.

Example 2 (*Wind direction data*) Two different techniques were used to measure wind directions along the Holderness coastline (the Humberside coast of the North Sea, United Kingdom), namely by using an anchored buoy (y) and *HF* radar (x). Altogether 129 observations were recorded in radians over 22.7 days. The data were fitted by using a simple circular regression model proposed by Hussin et al. (2004). The fitted model for the data is

$$\hat{y}_i = 0.165 + 0.973x_i \ (mod\, 2\pi).$$

Table 2 Observations detected using different values of ν for wind data set

ν	L_F	U_F	Number of outliers	Outliers
1.0	0.561	5.765	14	15,18,38,43,48,68,70,95,98,99,100,109,111,123
1.2	0.633	5.693	12	18,38,43,48,68,70,95,98,99,100,111,123
1.5	0.741	5.586	6	38,43,70,99,100,111
1.7	0.813	5.514	4	38,43,70,111
2.0	0.921	5.406	3	38,43,111
2.2	1.029	5.298	2	38,111
2.5	1.101	5.226	2	38,111
2.7	1.173	5.154	2	38,111
3.0	1.281	5.046	2	38,111
3.5	1.461	4.866	2	38,111

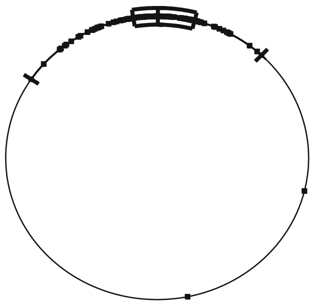

Fig. 6 *Circular boxplot* of circular residuals of the wind data set for $\nu = 2.5$

The estimated circular residuals have $\bar{\theta} = 0.017, \hat{\kappa} = 7.34, \phi = 0.0072$, $Q_1 = 0.202$, $Q_3 = 6.125$ and $CIQR = 0.360$. By using numerical and graphical methods, Abuzaid et al. (2008) identified the numbers 38 and 111 as outliers. The circular boxplot is used to identify possible outliers in a circular regression via the circular residuals. Since $\hat{\kappa} = 7.34$ is considered large, we can use values of ν larger than 2.

Table 2 gives the observations detected as outliers in the wind direction data set for different values of ν. As expected, the numbers 38 and 111 are identified as outliers for all values of ν. For smaller values of ν, many other observations are identified as outliers as well. Figure 6 shows the circular boxplot of circular residuals for $\nu = 2.5$.

6 Discussion

The boxplot has been used extensively in exploratory data analysis. In this paper, we propose a structure of the boxplot for circular variables. We specify the formulae for finding the median direction and the first and third quartiles, and overcome the problem of determining the lower and upper fences arising from the bounded range of the circle. It is shown that the values of the concentration parameter plays a significant role in the construction of the circular boxplot.

Several interesting results are found from the simulation studies in Sect. 3. Firstly, there exists a functional relationship between the $CIQR$ and a large κ. Secondly, we identify a condition where the problem of overlapping lower and upper fences may occur. Third, we are able to recommend that different values of ν be used to identify possible outliers in the circular variable. For samples with $\kappa > 3$, it is appropriate to use $2 < \nu < 2.7$, whereas for samples with $2 \leq \kappa \leq 3$, ν should be between 1 and 2. For simplicity, we fix $\nu = 1.5$ for $2 \leq \kappa \leq 3$ and $\nu = 2.5$ for $\kappa > 3$ since the choices ensure the nonoverlapping of the fences.

In the illustrations, we show that the circular boxplot is able to identify possible outliers in the frogs' directions data set and our result is similar to that of Collett (1980). Furthermore, in the case of the circular regression model, the numbers 38 and 111 are identified as outliers by the circular boxplot, which is consistent with Abuzaid et al. (2008).

The circular boxplot is therefore a useful graphical tool for an exploratory analysis of circular data as well as an alternative technique for identifying outliers in univariate circular samples.

Acknowledgments The authors are most grateful to the Associate Editor, referees and Prof S. Rao Jammalamadaka (University of California Santa Barbara) for their thorough reading and valuable suggestions which led to a substantial improvement of the article.

References

Abuzaid AH, Hussin AG, Mohamed IB (2008) Identifying single outlier in linear circular regression model based on circular distance. J Appl Probab Stat 3(1):107–117

Abuzaid AH, Mohamed IB, Hussin AG (2009) A new test of discordancy in circular data. Commun Stat Simul Comput 38(4):682–691

Barnett V, Lewis T (1978) Outliers in statistical data. Wiley, New York

Collett D (1980) Outliers in circular data. Appl Stat 29(1):50–57

David HA (1970) Order statistics. Wiley, New York

Ferguson DE, Landerth HF, Mckeown JP (1967) Sun compass orientation of the northern cricket frog. Acris Crepitans Anim Behav 15:45–53

Fisher NI (1993) Statistical analysis of circular data. Cambridge University Press, London

Fox J (1997) Applied regression analysis, linear models, and related models. Sage Publication, Singapore

Graedel TE (1977) The wind boxplot: an improved wind rose. J Appl Meteorol 16:448–450

Hoaglin DC, Iglewicz B, Tukey JW (1986) Performance of some resistant rules for outlier labeling. J Am Stat Assoc 81(396):991–999

Hussin AG, Fieller NRJ, Stillman EC (2004) Linear regression for circular variables with application to directional data. J Appl Sci Technol 8(1&2):1–6

Ingelfinger JA, Mosteller F, Thibodeau LA, Ware JH (1983) Biostatistics in clinical medicine. Macmillan, New York

Jammalamadaka SR, SenGupta A (2001) Topics in circular statistics. World Scientific Press, Singapore

Jammalamadaka SR, SenGupta A (2003) Linear statistical models: an integrated approach. World Scientific Press, Singapore

Mardia KV (1972) Statistics of directional data. Academic Press, London

Mardia KV (1975) Statistics of directional data. J R Stat Soc B 37:349–393

Mardia VK, Jupp PE (2000) Directional data 2nd. Wiley, London

Sim CH, Gan FF, Chang TC (2005) Outlier labeling with boxplot procedures. J Am Stat Assoc 100(470):642–652

Tukey JW (1977) Exploratory data analysis. Addison-Wesley, Reading

Wehrly T, Shine EP (1981) Influence curves of estimates for directional data. Biometrika 68:334–335