RESEARCH PAPER

IMPORTANCE OF CIRCULAR DATA IN SPORTS SCIENCE – A REVIEW

M. Srinivas, and S. Sambasiva Rao

Cite this article as: M. Srinivas, S. Sambasiva Rao, "Importance of Circular Data In Sports Science – A Review", Volume. 1, No.2, IJPEAAS (ISSN:2395-6895), pp.37-44.

IMPORTANCE OF CIRCULAR DATA IN SPORTS SCIENCE – A REVIEW

M. Srinivas[1†] and S. Sambasiva Rao[‡]

[†]Professor of Mathematics, [‡]Assistant Professor of Mathematics,
Department of Humanities and Sciences,
SVS Group of Institutions, Warangal-506015,
Telangana, India.
Email IDs: sri.srinivaswgl@gmail.com, ssrao.siginam@gmail.com

Abstract. The statistical analysis of angular data is typically encountered in biological and geological studies, among several other areas of research. Circular data is the simplest case of this category of data called directional data, where the single response is not scalar, but angular or directional. A statistical analysis pertaining to two dimensional directional data is generally referred to as "Circular Statistics". In this paper, an attempt is made to review various fundamental concepts of circular statistics and to discuss its applicability in sports science.

1. Introduction and Preliminaries

The majority of statistical techniques used in the analysis of human performance are linear, in which the assumptions are often easy to specify and provide good mathematical solutions for modeling a wide range of events. But many problems encountered in biological scenario do not lend themselves to strict linear representation (see [3] and [28]). It was observed that frequently cannot be modeled in a linear manner are data produced from circular scales. These variables are distinctive in the sense that data points are distributed on a circle instead of the traditional configuration of points on the real number line. Circular scales produce cyclic or periodic data that complicate traditional analytical procedures. The complexities found in evaluating circular data are largely a manifestation of the special interval level status the circular scale represents. Circular scales do not have a true zero point. That is, they are circular means that any designation of high or low or more or less is purely arbitrary.

The analysis of circular data is an odd corner of statistical science which many never visit, even though it has a long and curious history. Moreover, it seems that no major statistical language provides direct support for circular statistics.

The basic statistical assumption in circular statistics is that the data are randomly sampled from a population of directions. Observations arise either from direct measurement of angles or they may arise from the measurement of times reduced modulo some period and converted into angles according to the periodicity of time, such as days or years. They are commonly summarized as locations on a unit circle or as angles over a 360^0 or 2π radians range, with the endpoints of each range corresponding to the same location on the circle.

The key characteristic that differentiates circular data from data measured on a linear scale is its wrap-around nature with no maximum or minimum. That is, the "beginning" coincides with the "end", i.e., $0 = 2\pi$ and in general the measurement is periodic with θ being the same as $\theta + 2p\pi$ for any integer p. Differences between the theories of statistics on the line and on the circle can be attributed to the fact that the circle is a closed curve while the line is not. Thus, distribution functions, characteristic functions and moments on the circle have to be defined by taking into account the natural periodicity of the circle.

---

[1] Corresponding Author

Types of circular data

1. Vectorial (it consists of directed line segments in which both angle and direction associated with the point)
2. Axial (Data relating to the angular position of random lines which do not have a natural orientation associated with them)

We now provide formulae typically used in circular data analysis

1. Circular mean: $\overline{X} = \begin{cases} \frac{1}{n}\sum_{i=1}^{n}\cos\phi_i & \text{for ungrouped} \\ \frac{1}{n}\sum_{i=1}^{n} f_i\cos\phi_i & \text{for grouped} \end{cases}$

$$\overline{Y} = \begin{cases} \frac{1}{n}\sum_{i=1}^{n}\sin\phi_i & \text{for ungrouped} \\ \frac{1}{n}\sum_{i=1}^{n} f_i\sin\phi_i & \text{for grouped} \end{cases}$$

2. Length of mean vector: $r = \sqrt{\left(\overline{X}\right)^2 + \left(\overline{Y}\right)^2}$
3. Hence, average angle is $\cos\overline{\phi} = \frac{\overline{X}}{r}$ and $\sin\overline{\phi} = \frac{\overline{Y}}{r}$, and

$$\overline{\phi} = \begin{cases} \tan^{-1}(y/x) & \text{if } x > 0 \\ 180^0 + \tan^{-1}(y/x) & \text{if } x < 0 \\ 90^0 & \text{if } x = 0, y > 0 \\ 270^0 & \text{if } x = 0, y < 0 \\ \text{indeterminate} & \text{if } x = 0, y = 0 \end{cases}$$

4. A sample of n angles (in degrees) $a_1, a_2, \ldots a_n$ is to be summarized. [6] and [16] contain definitions of various summary statistics that are used for angular data.

$$C_p = \sum_{i=1}^{n}\cos(pa_i) \quad \text{and} \quad \overline{C_p} = \frac{C_p}{n}$$

$$S_p = \sum_{i=1}^{n}\sin(pa_i) \quad \text{and} \quad \overline{S_p} = \frac{S_p}{n}$$

$$R_p = \sqrt{C_p^2 + S_p^2} \quad \text{and} \quad \overline{R_p} = \frac{R_p}{n}$$

$$T_p = \begin{cases} \tan^{-1}\left(\frac{\overline{S_p}}{\overline{C_p}}\right) & \text{if } \overline{\mathrm{C_p}} > 0, \overline{S_p} > 0 \\ \tan^{-1}\left(\frac{\overline{S_p}}{\overline{C_p}}\right) + \pi & \text{if } \overline{\mathrm{C_p}} < 0 \\ \tan^{-1}\left(\frac{\overline{S_p}}{\overline{C_p}}\right) + 2\pi & \text{if } \overline{\mathrm{C_p}} > 0, \overline{S_p} < 0, \end{cases}$$

Where a =Resultant length

R=mean of the resultant length with limits 0 and 1

Circular variance = V=1-R

Mean direction=Theta=Measure of mean of individual angles, estimated by T
Circular dispersion= δ = (1-T)/(2R^2).

Major differences

1. A common problem in circular data is to estimate a preferred direction and its corresponding distribution. Common choices for summarizing the preferred direction are the sample circular mean, and sample circular median.
    - A common one-dimensional statistical problem is the estimation of a location parameter, and its corresponding distribution.
2. Linear descriptive statistics are limited in precisely characterizing the central tendency of circular distributions.
    - If we have a number of angular measurements on the circle, then the mean of those measures should give an estimate of the true population mean parameter.
3. Many of the problems associated with the use of traditional statistical methods for describing circular data emerge in statistical inference.
    - The usual parametric or non-parametric methods of statistical inference do not take into account scale circularity when it exists. Hence, these methods will be subject to serious, unknown and unrecognized errors in stated probabilities associated with Type **I** error rates, loss of statistical power, or both.
    - The statistical methods that minimize the interpretational risks associated with circular data analysis when certain distributional assumptions are met.
4. Failure to recognize the circularity of one or more variables in time series, regression, or correlation analysis may lead to overlooking important systematic relationships among variables.
    - The use of appropriate circular methods can assist in simplifying statistical relationships and improve the fit of data models.
5. Computing the median angle of unimodal circular data is similar to that for linear data.
6. If time is considered in the usual time-series analysis which is a linear variable, compared to situations where one is considering timing only within a cycle, which is most usefully treated as a circular variable.
7.

Type of Data	Deviation from mean	Sum of Deviations from Mean	Square of Deviations from mean	Variance
Linear	$\left(x_i - \bar{x}\right)$	$\sum\left(x_i - \bar{x}\right)$	$\left(x_i - \bar{x}\right)^2$	$\frac{1}{n}\sum\left(x_i - \bar{x}\right)^2 = s^2$
Circular	$\sin\left(\phi_i - \bar{\phi}\right)$	$\sin\left(\sum\left(\phi_i - \bar{\phi}\right)\right)$	$2\left[1-\cos\left(\left(\phi_i - \bar{\phi}\right)\right)\right]$	$\frac{1}{n}\sum 2\left[1-\cos\left(\phi_i - \bar{\phi}\right)\right] = 2(1-r)$

Circular Distribution

A circular distribution (CD) is that in which total probability is concentrated on the circumference of a unit circle. A set of identically independent random variables from such a distribution is referred to as a random sample from the CD. Two frequently used families of distributions for circular data include the von Mises and the Uniform distribution

The importance of the von Mises distribution is similar to the Normal distribution on the line (see [15]). It was introduced by von Mises (1918) to study the deviations of measured atomic weights from integral values. It is a symmetric unimodal distribution characterized by a mean direction μ, and concentration parameter k, with probability density function

$f(\theta) = \frac{1}{2\pi I_0(k)} e^{k\cos(\theta-\mu)}, \ \theta \geq 0, \mu < 2\pi, 0 \leq k < \infty$, where

$$I_0(k)=\frac{1}{2\pi}\int_0^{2\pi} e^{k\cos\theta}d\theta=\sum_{i=0}^{\infty}-\frac{1}{(i)^2}\left(\frac{k^2}{4}\right)^i$$

is the modified Bessel function of order zero. See [6] p.50 for a series expansion and methods for evaluating $I_o(k)$, k is a concentration parameter, which quantifies the dispersion. As k increase from zero, f(θ) peaks higher about μ.

The circular random variable θ is symmetric about a given direction μ if its distribution has the property f(μ + θ) = f(μ − θ), for all θ, where addition or subtraction is modulo (2π).

If k is zero, then f(θ) = 1/ 2π and the distribution is uniform with no preferred direction.

All directions are equally likely, hence this is also known as the Isotropic distribution. This distribution represents the state of no "preferred direction", since the total probability is spread out uniformly on the circumference of a circle.

The uniform distribution on the circle has the property that the sample mean direction and the sample length of the resultant vector are independent. Similar property is held by the normal distribution for linear data (see [12]).

The commonly used parametric model, the von Mises distribution, for analyzing directional data assumes unimodality and axial symmetry of a given data set. Since this is not always the case, the search for robust methods leads naturally to techniques which are non-parametric or are distribution free. The need for distribution-free methods is highly desirable in directional data analysis (see [17]).

2. Literature Review

Batschelet ([3]) has pioneered many of the principles in statistical circular methods; some of his work is no longer in print and thus not readily available. The definition of circular median is stated in [15], p. 28. For a detailed discussion of circular probability distributions we refer to see [8] p. 25--63.

The wrapped normal was introduced by [29] and later studied by [27], [13], and [7], [25] matched the first trigonometric moments of the von Mises and wrapped normal distributions.

The similarity of the two distributions has also been noted and to some extent explained by [9] , [10], [14] and [11]. Based on the difficulty in distinguishing the two distributions, [4] conclude that decision on whether to use a von Mises model or a wrapped normal model depends on which of the two is most convenient.

A large part of parametric statistical inference for circular data is derived based on one or two models and there has not been enough discussion on model-robustness, i.e., to justify their validity and use when the data is actually from another model (see [8]). As a result, modeling asymmetric data sets, which frequently occur in practice, provides some challenges because of the lack of appropriate models.

The standard texts on directional data are [15] and [6]. In [3], a less mathematical account of applications of circular data to the analysis of biological data is available and in [5], an account of methods for the analysis of spherical data is provided. The Authors (see [26], [16]) discuss both two and three dimensional data.

Inappropriate applications of linear methods to circular data are in the book by [28] p. 607, 624-625, who does not consider the fact that, the zero or positive direction in the circle is arbitrary. Many of his proposed methods are not rotationally invariant. Circular data is typically encountered in the following areas of research: migrating birds, human performance, military training, biomathematics (animal behavior studies on homing, migration, escape, and exploratory behavior), Biology (movements of migrating animals), Meteorology (winds), Geology (directions of joints and faults), Physics, Geomorphology (landforms, oriented stones), Spatial and temporal performance (navigation, work-system design, biological rhythms, and sleep), factor analysis, circadian physiology, atomic weights, Oceanography, Ozone concentrations, among others.

3. Circular data in Sports Science

In sports, the vision plays an important role in the control of far aiming tasks; its exact role is unclear with the purpose of scoring. In static far aiming tasks, like rifle shooting,

shooting free throws in basketball and playing billiards, the duration of the final fixation on the target before initiating the final movements correlates with expertise. Compared with non-experts, experts fixate their gaze at the target for longer before taking the actual shot, a phenomenon called ‘‘quiet eye’’. The prediction of offensive outcomes based on certain situations in the game. Some of these outcomes include, scoring, success of plays, and yardage gained. These predictions may help coaches make crucial decisions. In particular, in the basketball shooting, the goals were to find an average angle at which a player shoots the most often and has a high percentage of makes versus their average angle for missed shots. Data analysis involves check sheets, control charts, and Pareto diagrams. Examination of the charts should suggest causes that affect performance.

There has been no attention to the problem of circular data in sports science though there are several instances in various games where we encounter circular data. For instance, the strength of batsman in cricket to play shots in certain areas of cricket field follows strong directional patterns and one can analysis the bating style of a batsman based on this directional data using methods in circular statistics, like whether he/she is strong in hitting the ball more often in long off, midwicket and so on. One more example for using circular statistics analysis is due to Rita Ferraz De Oliveira et al. The authors examined the timing of optical information pick-up in basketball jump shooting using an intermittent viewing technique (see [31]). Expected shooters to prefer to look at the basket as late as possible under the shooting style used. Most shooters preferred to pick up optical information as late as possible given the adopted shooting style. One can conclude that, in dynamic far aiming tasks such as basketball jump shooting, late pick-up of optical information is critical for the successful guidance of movements.

4. Conclusions

Our purpose here is to point out that there exist circular data problems in sports science; there is a little research that has focused on the importance of circular data in sports science. Therefore, the coaches or sports science researchers must focus on the identification of data available to them regarding the games and differentiate them into linear and circular data for better understanding and analysis of outcome of the game and for the improvement in the performance of the player. To this end, the attempt made in this article may help to understand certain aspects of circular statistics.

References

[1]. Arnold, B.C., Castillo, E. and Sarabia, J.M. (1999), “Conditional Specification of Statistical Models”, Springer-Verlag, New York.

[2]. Arnold, B.C. and SenGupta, A. (2004), “Distributions with circular normal conditionals and tests for independence on the torus”, under revision for Annals of the Institute of Statistical Mathematics.

[3]. Batschelet, E. (1981), “Circular Statistics in Biology”, Academic Press. London

[4]. Collett, D., and Lewis, T. (1981), “Discriminating between the von Mises and wrapped normal distributions”, *Australian Journal of Statistics*, 23, 73-79.

[5]. Fisher, N.I., Lewis, T., and Embleton, B.J.J. (1987). “Statistical Analysis of Spherical data’, Cambridge University Press. Cambridge.

[6]. Fisher, N.I. (1993), “Statistical analysis of circular data”, Cambridge University Press, Cambridge.

[7]. Gumbel, E.J., Greenwood, J.A., and Durand, D. (1953), “The circular normal distribution Theory and Tables”, *Journal of the American Statistical Association.*, 48,131-152.

[8]. Jammalamadaka, S.R. and SenGupta, A. (2001), “Topics in Circular Statistics”, World Scientific, New Jersey.

[9]. Kendall, D.G. (1974a), “Hunting quanta”, *Trans. R. Soc. Lond.*, A 276, 231-266.

[10]. Kendall, D.G. (1974b), “Pole-seeking Brownian motion and bird navigation”, *J. R. Stat.Soc. B*, 36, 365-417.
[11]. Kent, J.T. (1976), “Distributions, Processes and Statistics on Spheres”, Ph.D Thesis. University of Cambridge. Cambridge.
[12]. Kent R.D., Netsell R.,and Abbs J.(1979), “Acoustic characteristics of dysarthria associated with cerebellar disease”, *Journal of Speech and Hearing Research*, 22, 627-648.
[13]. Levy, P. (1939). “L’addition des variable aleatoires defines sur une circonference”. *Bull. Soc. Math. France*, 67, pp. 1-41.
[14]. Lewis, T. (1974). Discussion on “Pole Seeking Brownian motion and bird navigation” by Kendall, D. G. *Journal of the Royal Statistical Society B* 36, 365-402.
[15]. Mardia, K.V. (1972) “Statistics of Directional Data”, Academic Press, New York.
[16]. Mardia, K.V., and Jupp, P.E. (2000), “Directional Statistics”, Wiley. Chichester.
[17]. Rao, J.S. (1984), “Nonparametric methods in directional data”, Chapter 31 in Handbook of Statistics, Vol. 4. Nonparametric statistics, editors P.R. Krishnaiah and P.K. Sen. 755-770. Elsevier Science Publishers. Amsterdam.
[18]. SenGupta, A. (1991), “A review of optimality of multivariate tests”, *Statistics & Probability Letters*, Special issue - Multivariate Optimality and Related Topics, S.R. Jammalamadaka and A. SenGupta (Guest eds.), 12, 527 - 535.
[19]. SenGupta, A. (1998) DDSTAP – “Directional Data Statistical Analysis Package”, *Indian Statistical Institute*, Kolkata.
[20]. SenGupta, A. (2004), “On the constructions of probability distributions for directional data”, *Bulletin of the Calcutta Mathematical Society*, 96, 139-154.
[21]. SenGupta, A. and Chang, H-J. (2005), “Optimal tests for the mean direction of a von Mises distribution”, In Felicitation volume in honor of B.K. Kale, B.C. Arnold, S. Bendre and U. Gather (eds). (To appear)
[22]. SenGupta, A. and Jammalamadaka, S.R. (2001), “Geometry of the von Mises distribution and of tests for its mean direction”, *Stochastic Modelling and Applications*, 4, 16-37.
[23]. SenGupta, A. and Pal, C. (2001), “On optimal tests for isotropy against the symmetric wrapped stable - circular uniform mixture family”, *Journal of Applied Statistics*, 28, 129-143.
[24]. SenGupta, A. and Vermeire, L. (1986), “On Locally Optimal Tests for Multiparameter Hypotheses”, *Journal of the American Statistical Association*, 81, 819-825.
[25]. Stephens, M.A. (1963), “Random walk on a circle”, *Biometrika*,Vol. 50, 385-390.
[26]. Upton, G.J.G., and Fingleton, B. (1989), “Spatial Data Analysis by Example” Volume 2. *Categorical and Directional Data*, Wiley. NY.
[27]. Wintner, A. (1933), “On the stable distribution law”, *American Journal of Mathematics*, 55, 335-339.
[28]. Zar, J.H. (1999), “Biostatistical Analysis”, 4th Edition. Prentice-Hall, Inc.
[29]. Zernike, F. (1928), “Wahrscheinlickkeitsrechnung und Mathematische Statistik. In Handbuch der Physik”, 3, 419-492. Springer Berlin.

[30]Girija S.V.S,“Construction of New Circular Models”,*VDM Publishing*, 2010.

[31] Rita Ferraz De Oliveira et al., “Late information pick-up is preferred in basketball jump shooting”, *Journal of Sports Sciences*, September 2006; 24(9): 933-940.