Advanced Review

Circular data

Alan Lee*

The special nature of circular data means that conventional methods suitable for the analysis of linear data do not apply. In this article, we survey a range of methods that have been developed over the last 50 years to handle the special characteristics of data consisting of angular measurements. We discuss summary statistics and graphical methods, methods for the analysis of single and multiple samples of circular data, circular correlation, regression methods, and time series. We discuss the standard probability models on which these analyses are based, and give several examples of the application of these methods. A reasonably comprehensive bibliography is provided. *WIREs Comp Stat* 2010 2 477–486 DOI: 10.1002/wics.98

Keywords: circular data; circular probability models; circular confidence intervals; circular correlation; regression models for a circular response

Observations consisting of directions or angles, while not as common as continuous measurements on a ratio scale, are nevertheless to be found across many areas of science, including ecology, earth sciences, environmental science, and medicine. Examples of such data are the angular movements of an animal relative to a food source or other attractor, wind directions, diurnal measurements of admission times to an intensive care unit, and departure directions of birds after release.

Circular data arise whenever directions are measured, and are usually expressed as angles relative to some fixed reference point, such as Due North. Time data measured on a 24 h clock may also be converted to angular measurements, with 0:00 corresponding to 0° and 24:00 to 360°. Angles can be measured in degrees, or, as we shall do for the most part, in radians.

In this article, we offer a brief survey of the statistical methods that can be used to handle such data. It is immediately apparent that statistical methods that are used for ordinary numerical data are quite inappropriate for circular data. For example, suppose we have angular measurements 1°, 0°, 359°. The arithmetic mean of these three numbers is 120°, but it is clear that a more sensible result is 0°. Over the past 50 years, a reasonably comprehensive suite of special techniques has been developed that respect the fact that the circle (where angular measurements live) is very different topologically from the line (where ordinary numeric measurements live). We now have data-analytic techniques that parallel the standard methods for the analysis of single and multiple samples of angular data, and for handling problems of regression and correlation. Paralleling the usual probability models such as the normal for linear data, we have corresponding models for distributions on the circle and the torus.

*Correspondence to: lee@stat.auckland.ac.nz

Department of Statistics, University of Auckland, Auckland, New Zealand

DOI: 10.1002/wics.98

LOCATION, DISPERSION AND CIRCULAR MOMENTS

Suppose we have a sample $\theta_1, \ldots, \theta_n$ which we can visualize as a set of unit vectors. Sensible measures of angular location and dispersion can be derived by considering the resultant of these n unit vectors. The direction of the resultant, $\overline{\theta}$ say, can be taken as the circular mean direction, analogous to the mean of a sample of linear data. Moreover, the length of the resultant provides a measure of dispersion of the data. At one extreme, if all the angles are identical, the unit vectors are coincident and the length of the resultant is n. At the other, if the data are uniformly dispersed around the circle, the length of the resultant is zero. A convenient measure of angular dispersion is $1 - \overline{R}$, where $\overline{R}$ is the mean resultant length (the length of the resultant divided by n.)

Higher moments can also be defined. The pth sample trigonometric moments are complex numbers defined by

$$m_p = \overline{C}_p + i\overline{S}_p, \tag{1}$$

where

$$\overline{C}_p = \frac{1}{n}\sum_{i=1}^{n}\cos p\theta_i, \quad \overline{S}_p = \frac{1}{n}\sum_{i=1}^{n}\sin p\theta_i.$$

Other summary measures, such as the circular variance $1 - \overline{R}$, analogues of skewness and kurtosis and other measures of location such as the circular median can also be defined. See Fisher [1] pp. 34, 35 for details. The large-sample joint distribution of these statistics is derived in Ref 2.

There are also specialized methods of data display for samples of circular data, such as rose diagrams, circular histograms, circular dotplots and nonparametric density estimates based on kernel smoothing. These data displays are discussed in Refs 1,3,4. Density estimation is considered in Refs 5–8.

CIRCULAR DISTRIBUTIONS

There are several distributions for angular data, corresponding to the usual distributions for linear data. Before describing these, we must define the concept of the circular density and the trigonometric moments of a distribution. Circular distributions are usually described in terms of a circular density, which is a function $f(\theta)$ defined for each angle θ (which need not be confined to $[-\pi, \pi]$) satisfying the conditions

1. $f(\theta) \geq 0$ for all angles θ,
2. f is periodic with period 2π,
3. $\int_{\theta_0}^{\theta_0+2\pi} f(\theta)\mathrm{d}\theta = 1$ for all θ_0.

Any such function describes a probability distribution on the circle. If θ is a circular random variable with density f, and θ_1 and θ_2 are fixed angles with $0 \leq \theta_1 \leq \theta \leq \theta_2 \leq 2\pi$ or $-\pi \leq \theta_1 \leq \theta \leq \theta_2 \leq \pi$, then

$$Pr[\theta_1 \leq \theta \leq \theta_2] = \int_{\theta_1}^{\theta_2} f(\theta)\,\mathrm{d}\theta. \tag{2}$$

The trigonometric moments of the distribution are the complex numbers

$$\mu_p = E(\mathrm{e}^{ip\theta}) = \alpha_p + i\beta_p, \quad p = 0, \pm 1, \pm 2, \ldots, \tag{3}$$

where $\alpha_p = \int_0^{2\pi} \cos p\theta f(\theta)\mathrm{d}\theta$ and $\beta_p = \int_0^{2\pi} \sin p\theta f(\theta)\,\mathrm{d}\theta$. The first trigonometric moment, expressed in polar coordinates, is

$$\mu_1 = E(\mathrm{e}^{i\theta}) = \rho \mathrm{e}^{i\mu}, \tag{4}$$

where ρ is the resultant length and μ the mean direction. The circular variance is defined as $1 - \rho$, and the circular dispersion as $\delta^2 = (1 - \rho_2)/2\rho^2$ where ρ_2 is the modulus of the complex number μ_2. For a sample of size n, the quantity $\delta/\sqrt{n}$ is called the circular standard error of the mean direction, and a sample version of this plays a role in inference for a single sample of data, as will be seen in the following text.

We can also interpret the trigonometric moments in terms of characteristic functions: The periodicity of the circular density (which implies that θ and $\theta + 2\pi$ have the same distribution) means that the characteristic function $\phi(t) = E(\mathrm{e}^{it\theta})$ only makes sense for the integer values of t. As for distributions on the line the characteristic function determines the distribution. It follows that circular distributions are determined by their trigonometric moments.

Univariate Distributions

Two Standard Distributions

Standard circular distributions include the uniform, with density $f(\theta) = \frac{1}{2\pi}$, which has no mean direction, and the von Mises, with density

$$f(\theta) = \frac{1}{2\pi I_0(\kappa)} \mathrm{e}^{\kappa \cos(\theta - \mu)}, \tag{5}$$

where μ is the mean direction, and I_0 is a modified Bessel function of order zero. The parameter κ is a concentration parameter: for $\kappa = 0$, the distribution reduces to the uniform, and for $\kappa \to \infty$, the distribution converges to a point mass at μ. The resultant length is given by $\rho = \frac{I_1(\kappa)}{I_0(\kappa)}$ where $I_1(\kappa)$ is a modified Bessel function of order one. The von Mises is often regarded as the circular analogue of the normal distribution on the line. Several properties of the normal have natural analogues for the von Mises. For example, as we shall see, the maximum likelihood estimate (MLE) of the mean direction is the sample mean direction, and the von Mises is the circular distribution having maximum entropy for a fixed mean direction and circular variance.

The von Mises distribution is unimodal. To model multimodal data, mixtures of von Mises distributions have been widely used. See Refs 9–15 for details.

Wrapped Distributions

If X is a linear random variable, with density g, then we may define the corresponding 'wrapped' circular version of X by

$$\theta = X(\mathrm{mod}\ 2\pi). \tag{6}$$

The variable θ has density

$$f(\theta) = \sum_{k=-\infty}^{\infty} g(\theta + 2\pi k), \quad (7)$$

obtained by 'wrapping' g around the circle. Two of the most commonly used wrapped distributions are the wrapped normal and the wrapped Cauchy. Using the theory of theta functions, the density of the wrapped normal can be expressed as:

$$f(\theta) = \frac{1}{2\pi}\left(1 + 2\sum_{p=1}^{\infty} \rho^{p^2} \cos(\theta - \mu)\right), \quad (8)$$

where ρ is the resultant length, given by $\rho = e^{-\sigma^2/2}$ where σ^2 is the variance of the wrapping normal. The mean direction is μ.

Suppose Y is a (linear) Cauchy random variable, with median μ and scale parameter σ, and density

$$g(y) = \frac{1}{2\pi}\frac{\sigma}{\sigma^2 + (y-\mu)^2}. \quad (9)$$

The density of the corresponding wrapped Cauchy can be expressed as:

$$f(\theta) = \frac{1}{2\pi}\frac{1-\rho^2}{1+\rho^2 - 2\rho\cos(\theta-\mu)} \quad (10)$$

where again μ and ρ are the mean direction and resultant length, with $\rho = e^{-\sigma}$.

Another elegant derivation of the wrapped Cauchy, given in Ref 16, can be made using Mobius transformations in the complex plane. Define a complex random variable Z by

$$Z = \frac{1+iY}{1-iY}. \quad (11)$$

This transformation is a special case of the general Mobius transformation, and as it maps onto the unit circle, Z can be regarded as a circular random variable $\theta = \arg(Z)$. Its density is given by Eq. (10), and μ and ρ are again the mean direction and resultant length. However, the relationship between ρ and the parameters of the (linear) Cauchy is different: let $z_0 = \mu + i\sigma$ and $\psi = (1 + iz_0)/(1 - iz_0)$. Then $\rho = |\psi|$ if $|\psi| < 1$ and $1/|\psi|$ otherwise.

The wrapped Cauchy has some very nice invariance properties which make it a useful distribution for regression modeling, as we explain in the section on Regression Models below. For example, write $Z \sim C^*(z_0)$ if Z is derived from Y through the Mobius transformation as described above. Then if w is a fixed complex number with $|w| = 1$, $wZ \sim C^*(wz_0)$. Moreover, if $Z \sim C^*(z_0)$, and $W = (Z + w)/(1 + \overline{w}Z)$, then $W \sim C^*\left((w + z_0)/(1 + \overline{w}z_0)\right)$, so that the wrapped Cauchy is invariant under these types of Mobius transformations. See Ref 16 for details.

Other wrapped distributions are the wrapped t,[17] the wrapped skew-normal,[18–19] and the wrapped stable distributions.[20] See also Ref 21 for the description of a family of distributions that includes the wrapped Cauchy, as well as the von Mises and the cardioid.

Projection Methods

Another general method for constructing angular distributions is to project univariate or bivariate linear variables onto the circle. For example, the image of a linear variable under a stereographic projection is a circular variable. This amounts to defining $\theta = 2\arctan(X)$, where X is linear. Another common method is to project a bivariate linear random vector $Y = (Y_1, Y_2)$ onto the unit circle. This results in a circular variable θ, where $\cos\theta = Y_1/\sqrt{(Y_1^2 + Y_2^2)}$, $\sin\theta = Y_2/\sqrt{(Y_1^2 + Y_2^2)}$. If Y is bivariate normal, this results in the offset normal distribution (see Ref 3, p. 46).

Bivariate Distributions

A simple way of creating bivariate circular distributions on the torus is by wrapping bivariate distributions on the plane: If $g(x, y)$ is a bivariate density on the plane, the wrapped version of g is

$$f(\theta, \psi) = \sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} g(\theta + 2k\pi, \psi + 2l\pi). \quad (12)$$

A simple calculation shows that the marginal densities of f are the wrapped versions of the marginal densities of g.

Another general method for constructing bivariate densities was given in Ref 22. If f_1, f_2 and f_3 are densities on the circle, then

$$f(\theta, \psi) = 2\pi f_3(2\pi\{F_1(\theta) + F_2(\psi)\})f_1(\theta)f_2(\psi) \quad (13)$$

and

$$f(\theta, \psi) = 2\pi f_3(2\pi\{F_1(\theta) - F_2(\psi)\})f_1(\theta)f_2(\psi) \quad (14)$$

are both bivariate distributions on the torus with marginal distributions f_1 and f_2. The functions F_1 and

F_2 are given by

$$F_1(\theta) = \int_0^\theta f_1(\theta')\,\mathrm{d}\theta' \quad \text{and} \quad F_2(\psi) = \int_0^\psi f_2(\psi')\,d\psi', \tag{15}$$

and are the circular distribution functions corresponding to f_1 and f_2.

The use of projection methods to construct bivariate distributions having specified marginal distributions is described in Ref 23. For example, if U and V are bivariate random vectors such that the elements of U are uncorrelated standard normals and similarly for V, but (U, V) has a multivariate normal distribution, then the projections of U and V onto the unit circle generate two-correlated uniform distributions.

Finally, a bivariate analogue of the von Mises distribution has been considered by several authors. Consider a distribution on the torus having density

$$f(\theta, \psi) = C \exp(\kappa_1 a_1^T U + \kappa_2 a_2^T V + U^T A V) \tag{16}$$

where $U = (\cos\theta, \sin\theta)$, $V(\cos\psi, \sin\psi)$, and A is a 2×2 matrix. In general, this distribution does not have von Mises margins. Special cases have been studied in Refs 24–26.

INFERENCE FOR CIRCULAR DATA

Inference for a Single Sample

Suppose we have a sample $\theta_1 \ldots, \theta_n$ of angular data, assumed to be generated from some circular distribution. How can we be sure our distributional assumption is reasonable? If it is, how can we estimate the model parameters? How can we estimate the underlying mean direction, if indeed it exists, (the underlying model might be the uniform distribution). How can we construct tests and confidence intervals for the mean direction?

Turning first to questions of goodness of fit, we note that the classical methods used to test uniformity of linear data can be used (with suitable modifications to make them invariant to the choice of reference direction) for testing uniformity of circular data. Note that the uniformity of the underlying distribution may be of interest in its own right, e.g., when astronomers wish to test if stars are uniformly distributed on the celestial horizon. Tests for uniformity of linear data are usually based on either the χ^2 test, tests based on the empirical distribution or tests based on spacings between observations. Circular versions of these tests have been developed in Refs 27–29 in the case of empirical distribution function (EDF)-based tests, and Refs 30,31 in the case of tests based on arc lengths.

The linear tests based on EDFs are not very powerful against multimodal alternatives, and this is true for the circular versions as well. Smooth tests of uniformity which have better power against this type of alternative are discussed in Ref 32. Parametric tests of uniformity can be performed by, e.g., testing if $\kappa = 0$ in the von Mises distribution. As $\kappa = 0$ (or equivalently $\rho = 0$) implies uniformity for the von Mises, a natural test is based on the sample mean resultant length $\bar{R}$. This is the Raleigh test.

Having satisfied ourselves that the mean direction exists, we can proceed to estimate it with a confidence interval. For large samples, we can use the interval

$$\bar{\theta} \pm \arcsin(z(\alpha/2)\hat{\sigma}) \tag{17}$$

where $z(\alpha/2)$ is the usual normal quantile and $\hat{\sigma}$ is the estimated circular standard error defined in the discussion above of circular distributions. This interval, introduced in Ref 33, is based on a series of asymptotic approximations and is recommended when the sample size exceeds 25. A bootstrap approach when $n \leq 25$ is recommended in Ref 1. Bootstrap approaches to setting confidence intervals based on asymptotically pivotal statistics are discussed in Refs 34,35, with the solution proposed in the second paper being superior.

If we are prepared to make distributional assumptions, we can employ conventional likelihood methods. For the von Mises, inference is straightforward: the MLE of the mean direction is the sample mean direction, and the MLE of κ is obtained by solving the equation $I_1(\kappa)/I_0(\kappa) = R$, where R is the sample mean resultant length. See Ref 3, p. 85 for details. An approximate confidence interval is described in Ref 36. Tests for the mean direction based on large-sample theory are discussed in Refs 1, p. 93 and 3, p. 122. A goodness of fit test for the von Mises is provided in Ref 37.

For likelihood inference for the wrapped distributions, see the references above in the discussion of wrapped distributions.

Inference for Several Samples

An approximate test for comparing several von Mises mean directions was introduced in Ref 38. Suppose n_i and $R_i, i = 1, 2, \ldots, k$ represent the sample size and resultant length of k samples, from each of k different von Mises populations, assumed to have the same concentration parameter, and let n and R be the size and resultant length of the combined sample. The quantities $n_i - R_i$ represent the concentration of the individual samples, and $n - R$ represents the concentration of the combined sample. If the mean directions coincide, the identity

$$2\kappa(n-R) = 2\kappa\sum_i (n_i - R_i) + 2K\left(\sum_i R_i - R\right) \quad (18)$$

has an approximate χ^2 decomposition

$$\chi^2_{n-1} = \chi^2_{p-1} + \chi^2_{n-p} \quad (19)$$

from which an approximate analysis of variance can be constructed. The approximation improves as the concentration parameter increases, but correction factors are available; see Ref 39. If the assumption of equal dispersions is untenable, see Ref 40. A test based on the likelihood ratio test for this problem is given by Ref 41. For a bootstrap approach, see Ref 42.

More Complicated Designs

Several authors (see Refs 41,43,44) have attempted to extend these results for the several sample case to cover multifactorial designs, but the lack of a suitable notion of interaction has hampered these efforts, with no compelling breakdown of tables of mean directions into main effects and interactions being available.

REGRESSION AND CORRELATION

Circular Correlation

Measuring association between two circular variables (or a circular and a linear variable) is less straightforward than between two linear variables, largely because there is no natural relationship between them. In the linear–linear case, a linear relationship provides a natural standard against which to measure association. No such natural relationship exists where circular variables are involved. Consequently, many different correlation coefficients for angular data have been proposed, corresponding to different types of relationship.

A measure of association between two variables should have the following properties:

1. It should have a particular value (by convention zero) when the variables are independent.
2. It should lie between two particular values (by convention ± 1).
3. It should take these particular values when the variables have some specified relationship, such as a linear relationship for linear variables.

We treat the cases of circular–linear and circular–circular association separately, because different types of relationship are appropriate in these two situations.

Circular–linear Association

The joint distribution of a circular and a linear variable is supported by an infinite cylinder, and a possible form of relationship between a linear variable y and a circular variable θ is

$$y = a + f(\theta - \mu) \quad (20)$$

where f is a continuous function with $f(0) = 0$. If f is a bounded periodic function with period 2π, then f defines a closed curve around the cylinder. In the special case when f is a multiple of the sine function, the relationship becomes

$$y = a + b\sin\theta + c\cos\theta \quad (21)$$

and an obvious measure of correlation in this case, proposed in Ref 45, is the ordinary multiple correlation between y and the variables $(\cos\theta, \sin\theta)$. However, this has the disadvantage that it is not signed. An alternative based on the canonical correlation between y and $(\cos\theta, \sin\theta)$ is considered in Ref 24. In Ref 46, more general forms of the relationship (requiring only that g have one maximum and one minimum in $[0, 2\pi]$) are discussed. A correlation coefficient to measure this type of association is introduced, as are distribution-free tests of independence.

Another form of f is an unbounded periodic function that is monotone on $(-\pi, \pi)$, such as the function $\tan(\frac{\theta}{2})$. Although in Ref 47 a regression model based on this type of relationship is introduced, no corresponding correlation coefficient seems to have been proposed. This regression model is discussed further in the following text.

Circular–circular Association

Consider a one-to-one mapping taking a circular variable θ onto another circular variable ψ, with the property that as θ moves around the circle in a clockwise direction, so does ψ. In Ref 48, this is termed positive toroidal association; correspondingly if ψ moves in an anticlockwise direction as θ moves clockwise, we have negative toroidal association. Two signed correlation coefficients to measure this type of association, together with the associated tests of independence are also presented. A generalization of one of these tests to higher dimensions is given in Ref 49.

A much simpler form of relationship between angular variables can be expressed in terms of reflections and rotations: in Ref 50, relationships of the form

$$\psi = \theta + \alpha (\mathrm{mod}\ 2\pi) \tag{22}$$

and

$$\psi = -\theta + \alpha (\mathrm{mod}\ 2\pi) \tag{23}$$

are considered, together with a signed correlation coefficient

$$\rho = \frac{E[\sin(\theta_1 - \theta_2)\sin(\psi_1 - \psi_2)]}{\{E[\sin^2(\theta_1 - \theta_2)]E[\sin^2(\psi_1 - \psi_2)]\}^{1/2}} \tag{24}$$

where θ_1 and θ_2 are independently distributed as θ, and ψ_1 and ψ_2 are independently distributed as ψ. The correlation ρ takes the value zero when θ and ψ are independent, the value 1 if and only if Eq. (22) holds, and the value 1 if and only if Eq. (23) holds. There is also an associated large-sample test of independence.

Other authors have obtained correlations by applying multivariate methods to the unit vectors $U = (\cos\theta, \sin\theta)$ and $V = (\cos\psi, \sin\psi)$. In Ref 24, an alternative correlation is proposed based on the maximum canonical correlation between the vectors U and V, i.e., the largest singular value of the matrix $\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$, where

$$\mathrm{Cov}\begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}. \tag{25}$$

This correlation is not signed, but takes the value 1 if and only if the relationships (22) or (23) hold. A correlation based on on the trace of $\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ is studied in Ref 51. Using the maximum singular value of the matrix Σ_{12} is suggested in Ref 25.

Tests for Independence

If we are prepared to assume a particular form for the bivariate distribution of (θ, ψ) which includes independence as a special case, we can test the independence of the variables by testing the parameters of the distribution assume the appropriate values corresponding to independence. For example, in the wrapped bivariate normal, if the correlation of the underlying bivariate normal is zero, then the wrapped bivariate normal has independent margins.

If we do not wish to assume a parametric form the joint distribution, we can use a nonparametric test based on circular ranks.[52] Other nonparametric tests include those in Ref 53 based on EDFs, and Ref 54, based on weighted degenerate U-statistics.

Regression Models for a Circular Response

Regression models have been proposed for the three cases

1. Linear response, angular covariates;
2. Angular response, linear covariates;
3. Angular response, angular covariate.

Case 1 is straightforward with the angular covariates entering as trigonometric polynomials, and will not be discussed further. Case 2 has been considered in Ref 47, where a model of the form

$$\theta = \beta_0 + g(\beta^T x) + \epsilon\ (\mathrm{mod}\ 2\pi), \tag{26}$$

is proposed. As discussed earlier, g is a monotone link function mapping the real line to $(-\pi, \pi)$, with $g(0) = 0$, and ϵ has a von Mises distribution with mean direction 0. In this reference, a procedure for estimating the parameters using maximum likelihood is proposed, but the likelihood is difficult to work with for more than a few covariates.

An alternative approach is to use the offset normal distribution, allowing the mean vector (μ_1, μ_2) of the bivariate normal to be a linear function of the covariates. See Ref 55 for parameter estimation using maximum likelihood. This method seems to have better numerical properties than the method using monotone link functions.

Case 3 is studied in Ref 56, using the relationship

$$\psi = \mu_\psi + g(\omega g^{-1}(\theta - \mu_\theta)) + \epsilon\ (\mathrm{mod}\ 2\pi) \tag{27}$$

with $g(x) = 2\arctan x$, where ϵ has a von Mises distribution. Parameter estimation and certain degenerate cases are discussed. Note that the regression relationship $\psi = \mu_\psi + g(\omega g^{-1}(\theta - \mu_\theta))$ is a special case of a Mobius transformation.

A regression model also involving Mobius transformations, but based on the circular Cauchy distribution, has recently been proposed in Ref 57. Consider a circular variate U (regarded as a complex number with modulus 1) having a $C^*(\phi)$ distribution, where $0 < \phi < 1$. Their regression model, for an angular covariate θ, is given by

$$Z = \beta_0 \frac{e^{i\theta} + \beta}{\overline{\beta} e^{i\theta} + 1} U, \qquad (28)$$

where β_0 and β are complex numbers with $|\beta_0| = 1$ and $|\beta| \neq 1$. By the properties of the circular Cauchy distribution, the angular variable $\psi = \arg Z$ has a circular Cauchy distribution, with parameter $\beta_0 \frac{e^{i\theta}+\beta}{\overline{\beta} e^{i\theta}+1}\phi$. The parameter ϕ represents the strength of the regression relationship. As $\phi \to 1$, the variables are related without error, whereas $\phi \to 0$, the response is uniformly distributed independently of the covariates. A bivariate distribution for θ, ψ in which the conditional distribution of ψ given θ is the regression function above is also described in Ref 57. Estimation via maximum likelihood is straightforward. A test of independence corresponding to this bivariate distribution can be obtained by testing $\beta = 1$.

Finally, a tree-based regression model, with a data-driven choice of regression function, is considered in Ref 58.

Circular Time Series

The regression models above can be adapted to analyze time series of angles. For example, in Ref 59 an autoregressive model is obtained by wrapping a linear AR(1) model around the circle, and applied it to the analysis of wind directions. Four types of time series model are studied in Ref 60. The first is obtained by wrapping standard linear ARMA processes, and the second by projecting a bivariate linear process onto the unit circle. Two further models are based on the regression models discussed above. One is a linked process of the form

$$\theta_t = \mu + g(X_t) \qquad (29)$$

where X_t is a linear process, and the second one of the form

$$\theta_t = \mu + g(\omega_1 g^{-1}(\theta_{t-1} - \mu)) + \cdots + \omega_p g^{-1}(\theta_{t-p} - \mu)) + \epsilon_t \qquad (30)$$

where the ϵ_t's are independent von Mises variables. Model identification and fitting issues are also discussed. Other recent approaches include the one based on generalized estimating equations,[61] and another on hidden Markov models.[62]

APPLICATIONS

The techniques described above have been applied to several fields of scientific endeavor, most notably in meteorology, ecology, and medicine. In meteorology, the von Mises distribution has been used as a model for the analysis of wind directions. See Ref 59 for time series of wind directions, see Ref 63 for the estimation of the mean direction of winds, and the calculation of correlations between wind direction and wave heights. Mixtures of von Mises distributions are used to model wind speeds and wind directions in Refs 64,65. Angular-linear correlation is used to examine the relationship between wind directions and pollution concentrations in Ref 66.

In ecology, angular measurements arise when the directional movement of animals in response to stimuli is considered. Projected normal regression is used to study the orientation of sandhoppers in Ref 67. Circular regression is applied to the same problem in Ref 68. Circular time series methods are applied to the tracks of bark beetles in Ref 69. A multifactorial analysis of intertidal snail movements is presented in Refs 43 and 44. Circular regression is used to study the influence of a landscape feature on the direction of animal movement in Ref 70.

Diurnal or seasonal data can also be analyzed by these methods, by identifying the position in the cycle with an angle. Thus, in Ref 71, von Mises distributions are fitted and density estimation is used to study crime incident statistics, identifying time of day with an angle. Circular regression is used to study the relationship between the seasonality of disease onset and other covariates in Ref 72. Mixtures of von Mises distributions are applied to model the seasonal pattern of sudden infant death syndrome in Ref 73.

OTHER TOPICS

Bayesian Approaches to Circular Data

Several of the techniques described above have Bayesian counterparts. For the one-sample problem, a Bayesian analysis based on the von Mises distribution is described in Ref 74. In Ref 75, the same ground is covered from an empirical Bayes perspective. One and two sample problems are treated in Ref 76. A hierarchical Bayesian treatment of the circular–linear regression model is given in Ref 77.

Robustness and Outliers

For directional data, the problem of outliers is not as acute as is the case for linear data, as angles are confined to the interval $[0, 2\pi]$. However, individual points can still exert an undue influence, particularly on measures of concentration. See Ref 78, Ch11, for a discussion. Several tests to detect outliers have been proposed. In Ref 79, a test based on the effect on the resultant length of a sample when a single observation is deleted is presented. Another approach is to fit a mixture model, and test for the homogeneity of the mixture. See Ref 4, Ch 10 for a discussion. Finally, Collett[79] and SenGupta[80] approach the outlier problem though slippage tests, where all observations but one have a common distribution, but the distribution of the remaining observation has a different mean direction.

Software for Circular Data Analysis

The book by Jammalamadaka and SenGupta[4] contains examples illustrating the use of an R package `CircStats`, written by Ulrick Lund, that will perform many simple calculations (such as summary statistics and some simple tests) on circular data. It will also generate random angles from the standard distributions, plot circular data and fit the standard distributions to data. The package `circular` by Claudio Agostinelli and Ulrick Lund is an updated version with more functionality.

CONCLUSION

In this article, we have given a brief survey of statistical methods for the analysis of circular data. As we have pointed out, the standard methods used for linear data are not appropriate for the analysis of angles. However, over the last 50 years, a comprehensive set of methods for circular data analysis have been developed, covering summary statistics, probability distributions, one- and multi-sample problems, correlation, regression, and time series. There is an active and expanding literature, and statistical software is available for the implementation of the methods, which are finding application in many fields of scientific endeavor.

The reader wanting to know more about circular data should consult one of the excellent texts available. The first comprehensive treatment was published by Mardia,[81] followed by Batschelet,[82] Fisher,[1] Mardia and Jupp,[3], and Jammalamadaka and SenGupta.[4] The texts by Batschelet and Fisher are more applied books oriented toward the working scientist wanting to apply the methods, whereas the others are more theoretical. The book by Mardia and Jupp is an updated version of Ref 81.

REFERENCES

1. Fisher NI. *Statistical Analysis of Circular Data*. Cambridge: Cambridge University Press; 1993.
2. Pewsey A. The large-sample joint distribution of key circular statistics. *Metrika* 2004, 60:25–32.
3. Mardia KV, Jupp PE. *Directional Statistics*. New York: John Wiley & Sons; 2000.
4. Jammalamadaka SR, SenGupta A. *Topics in Circular Statistics*. Singapore: World Scientific; 2001.
5. Hall P, Watson GS, Cabrera J. Kernel density estimation with spherical data. *Biometrika* 1987, 74:751–762.
6. Fisher NI. Smoothing a sample of circular data. *J Struct Geol* 1989, 11:775–778.
7. Klemela J. Estimation of densities and derivatives of densities with directional data. *J Multivariate Anal* 2000, 73:18–40.
8. Taylor CC. Automatic bandwidth selection for circular density estimation. *Comput Stat Data Anal* 2008, 52:3493–3500.
9. Fraser YS, Hsu MD, Walker JJ. Identifiability of finite mixtures of von Mises distributions. *Ann Econ Stat* 1981, 9:1130–1131.
10. Bartels R. Estimation in a bidirectional mixture of von Mises distributions. *Biometrics* 1994, 40:777–784.
11. Kim PT, Koo JY. Directional mixture models and optimal estimation of the mixing density. *Can J Stat* 2000, 28:383–398.
12. Grimshaw SD, Whiting DG, Morris TM. Likelihood ratio tests for a mixture of two von Mises distributions. *Biometrics* 2001, 57:260–265.
13. Holzmann A, Munk H, Startman B. Identifiability of finite mixtures - with applications to circular distributions. *Sankhya* 2004, 66:440–450.
14. Fu Y, Chen J, Li P. Modified likelihood ratio test for homogeneity in a mixture of von Mises distributions. *J Stat Plan Inference* 2008, 138:667–681.
15. Chen J, Li P, Fu Y. Testing homogeneity in a mixture of von Mises distributions with a structural parameter. *Can J Stat* 2008, 36:129–142.
16. McCullaugh P. Mobius transformations and Cauchy parameter estimation. *Ann Econ Stat* 1996, 24: 787–808.
17. Pewsey A, Lewis T, Jones MC. The wrapped *t* family of circular distributions. *Aust N Z J Stat* 2007, 49:79–91.

18. Pewsey A. The wrapped skew-normal distribution on the circle. *Commun Stat Theory Methods* 2000, 29:2459–2472.
19. Pewsey A. Modelling asymmetrically distributed circular data using the wrapped skew-normal distribution. *Environ Ecol Stat* 2006, 13:257–269.
20. Pewsey A. The wrapped stable family of distributions as a flexible model for circular data. *Comput Stat Data Anal* 2008, 52:1516–1523.
21. Jones MC, Pewsey A. A family of symmetric distributions on the circle. *J Am Stat Assoc* 2005, 100:1422–1428.
22. Wehrly TE, Johnson RA. Bivariate models for independence of angular observations and a related Markov process. *Biometrika* 1980, 67:255–256.
23. Saw JG. Dependent unit vectors. *Biometrika* 1983, 70:665–671.
24. Johnson RA, Wehrly TE. Measures and models for angular correlation and angular-linear correlation. *J R Stat Soc., Ser A* 1977, 39:222–229.
25. Rivest L-P. A distribution for dependent unit vectors. *Commun Stat Theory Methods* 1988, 17:661–483.
26. Singh H, Hnizdo V, Demchuk E. Probabilistic model for two dependent circular variables. *Biometrika* 2002, 89:719–723.
27. Kuiper NH. Tests concerning random points on the circle. *Proc K Ned Akad Wet C* 1960, 63:38–47.
28. Watson GS. Goodness-of-fit tests on the circle. *Biometrika* 1961, 48:109–114.
29. Ajne B. A simple test for uniformity of a circular distribution. *Biometrika* 1968, 55:343–354.
30. Sethuraman J, Rao JS. Pitman efficiencies of tests based on spacings. *Nonparametric Techniques in Statistical Inference*. Cambridge: Cambridge University Press; 1970,, 405–415.
31. Rao JS. Some tests based on arc lengths for the circle. *Sankhya* 1976, 38:329–338.
32. Bogdan K, Bogdan M, Futschik S. A data driven smooth test for circular uniformity. *Ann Inst Stat Math* 2002, 54:29–43.
33. Fisher NI, Lewis T. Estimating the mean direction of several circular or spherical distributions with differing dispersions. *Biometrika* 1983, 70:333–341.
34. Fisher NI, Hall P. Bootstrap confidence regions for directional data. *J Am Stat Assoc*, 1989, 84 996–1002 (correction 1990), 85:608.
35. Fisher NI, Hall P, Jing B-Y, Wood ATA. Improved pivotal methods for constructing confidence intervals with directional data. *J Am Stat Assoc* 1996, 91:1062–1070.
36. Upton GTG. Approximate confidence intervals for the mean direction of the von Mises distribution. *Biometrika* 1986, 73:525–527.
37. Lund U, Jammalamadaka SR. An entropy-based test for goodness of fit of the von Mises distribution. *J Stat Comput Simul* 2000, 67:319–332.
38. Watson GS, Williams E. On the construction of significance tests on the circle and the sphere. *Biometrika* 1956, 43:344–352.
39. Stephens MA. Tests for the von Mises distribution. *Biometrika* 1969, 56:149–160.
40. Watson GS. Large sample theory of the Langevin distribution. *J Stat Plan Inference* 1983, 8:245–256.
41. Anderson CFJ, Wu CM. Measuring location effects from factorial experiments with a directional response. *Int Stat Rev* 1995, 63:345–363.
42. Amaral GJA, Dryden IL, Wood ATA. Pivotal bootstrap methods for *k*-sample problems in directional statistics and shape analysis. *J Am Stat Assoc* 2007, 102:695–707.
43. Underwood AJ, Chapman MG. Multifactorial analyses of directions of movement of animals. *J Exp Mar Biol Ecol* 1985, 91:17–43.
44. Underwood AJ, Chapman MG. Experimental analyses of the influences of topography and of the substratum on the movements and density of an intertidal snail, Littorina unifasciata. *J Exp Mar Biol Ecol* 1989, 134:175–196.
45. Mardia KV. Linear-circular correlation coefficients and rythmometry. *Biometrika* 1976, 63:403–405.
46. Fisher NI, Lee AJ. Nonparametric measures of angular-linear association. *Biometrika* 1981, 68:629–636.
47. Fisher NI, Lee AJ. Regression models for an angular response. *Biometrics* 1992, 48:665–677.
48. Fisher NI, Lee AJ. Nonparametric measures of angular-angular association. *Biometrika* 1982, 69:315–321.
49. Jupp PE. A nonparametric correlation coefficient and a two-sample test for random vectors or directions. *Biometrika* 1987, 74:887–890.
50. Fisher NI, Lee AJ. A correlation coefficient for circular data. *Biometrika* 1983, 70:327–332.
51. Mardia KV, Jupp PE. A general correlation coefficient for directional data and related regression problems. *Biometrika* 1980, 67:163–173.
52. Rothman ED. Tests of coordinate independence for a bivariate sample on a torus. *Ann Math Stat* 1971, 42:1962–1969.
53. Rao JS, Puri ML. Problems of association for bivariate circular data and a new test of independence. In: Krishnaiah PR, ed. *Multivariate Analysis*, vol. 4. Amsterdam: North Holland; 1977,, 515–223.
54. Shieh GS, Johnson RA, Frees EW. Testing independence of bivariate circular data and weighted degenerate u-statistics. *Stat Sin* 1994, 4:729–747.
55. Presnell B, Morrison SP, Littell RC. Projected multivariate linear models for directional data. *J Am Stat Assoc* 1998, 93:1068–1077.

56. Mardia KV, Downs TD. Circular regression. *Biometrika* 2002, 89:683–697.

57. Kato K, Shimizu S, Shieh G. A circular-circular regression model. *Stat Sin* 2008, 18:633–645.

58. Lund UJ. Tree-based regression for a circular response. *Commun Stat Theory Methods* 2002, 31:1549–1560.

59. Breckling J. *The Analysis of Directional Time Series: Application to Wind Speed and Direction*. New York: Springer Verlag; 1989.

60. Fisher NI, Lee AJ. Time series analysis of circular data. *J R Stat Soc., Ser A* 1994, 56:327–339.

61. Artes R, Paula GA, Ranvaud R. Analysis of circular longitudinal data based on generalised estimating equations. *Aust N Z J Stat* 2000, 42:347–358.

62. Holzmann H, Munk A, Suster M, Zucchini W. Hidden Markov models for circular and linear-circular time series. *Environ Ecol Stat* 2006, 13:325–347.

63. Bowers ID, Morton JA, Mould GI. Directional statistics of the wind and waves. *Appl Ocean Res* 2000, 22:13–30.

64. Martin LV, Cremades A, Santabarbara JM. Analysis and modelling of time series of surface wind speed and direction. *Int J Climatol* 1999, 19:197–209.

65. Carta C, Bueno JA, Ramirez P. Statistical modelling of directional wind speeds using mixtures of von Mises distributions. *Energy Conserv Manage* 2008, 49:897–907.

66. Somerville MC, Mukerjee S, Fox DL, Stevens RK. Statistical approaches to wind sector analyses for assessing local source impacts. *Atmos Environ* 1994, 28:3483–3493.

67. Marchetti GM, Scapini F. Use of multiple regression models in the study of sandhopper orientation under natural conditions. *Estuar Coast Shelf Sci* 2003, 58S:207–215.

68. D'Elia A, Borgioli C, Scapini F. Orientation of sandhoppers under natural conditions in repeated trials: an analysis using longitudinal directional data. *Estuar Coast Shelf Sci* 2001, 53:839–847.

69. Preisler HK, Akers RP. Autoregressive-type models for the analysis of bark beetle tracks. *Biometrics* 1995, 51:259–267.

70. Tracey JA, Zu J, Crooks K. A set of nonlinear regression models for animal movement in response to a single landscape feature. *J Agric Biol Environ Stat* 2005, 10:1–18.

71. Brunsdon C, Corcoran J. Using circular statistics to analyse time patterns of crime incidence. *Comput Environ Urban Syst* 2006, 30:300–319.

72. Gao K-S, Chia F, Krantz I, Nordin P, Machin D. On the application of the von Mises distribution and angular regression methods to investigate the seasonality of disease onset. *Stat Med* 2006, 25:1593–1618.

73. Mooney JA, Helms PJ, Joliffe IT. Fitting mixtures of von Mises distributions: a case study involving sudden infant death syndrome. *Comput Stat Data Anal* 2003, 41:505–513.

74. Nunez-Antonio G, Gutierrez-Pena E. A Bayesian analysis of directional data using the von Mises-Fisher distribution. *Commun Stat Simul Comp* 2005, 34:989–999.

75. Rodrigues J, Leite JG, Milan LA. An empirical Bayes inference for the von Mises distribution. *Aust N Z J Stat* 2000, 42:433–440.

76. Ghosh SR, Jammalamadaka K, Tiwari RC. Semiparametric Bayesian techniques for problems in circular data. *J Appl Stat* 2003, 30:145–161.

77. George BJ, Ghosh K. A semiparametric Bayesian model for circular-linear regression. *Commun Stat Simul Comp* 2006, 35:911–923.

78. Barnett V, Lewis T. *Outliers in Statistical Data*, 2nd ed. London: John Wiley & Sons; 1984.

79. Collett D. Outliers in circular data. *Appl Stat* 1980, 29:50–57.

80. SenGupta A, Laha AK. The slippage problem for the circular normal distribution. *Aust N Z J Stat* 2001, 43:461–471.

81. Mardia KV. *Statistics of Directional Data*. New York: Academic Press; 1972.

82. Batschelet E. *Circular Statistics in Biology*. London: Academic Press; 1981.

