

# Working With Circular Data: A Tutorial for Cognitive and Behavioral Research

Stella Wernicke<sup>1</sup>, Julie de Falco<sup>1</sup>, Zahara Gironés<sup>1,2,3</sup>, Chris Jungerius<sup>1</sup>, Ivan Tomić<sup>4</sup>, Adam Triabhall<sup>1</sup>, Xiaolu Wang<sup>1,5</sup>, and Paul M. Bays<sup>1</sup>

<sup>1</sup> Department of Psychology, University of Cambridge

<sup>2</sup> Department of Engineering, University of Cambridge

<sup>3</sup> Department of Clinical Neuroscience, University of Cambridge

<sup>4</sup> Department of Psychology, Faculty of Humanities and Social Sciences, University of Zagreb

<sup>5</sup> Department of Psychology, Sun Yat-sen University

## Abstract

Introductory statistics courses and conventional methods in behavioral and cognitive science are typically restricted to Euclidean space without explicitly stating this, leaving many researchers unaware of how to handle data with circular or periodic properties (e.g., time of day or orientation on a screen). This tutorial provides an accessible yet rigorous introduction, designed to bridge the gap between statistical theory and practical application. After discussing different ways of representing circular data mathematically and fundamental principles for working with them, we explain angular and vector representations and operations, including key considerations for implementing them in computer code. We then survey methods for visualizing circular data, including linear and polar plots and presentation of summary statistics. Next, we introduce circular measures of central tendency, dispersion, and shape: how they relate to their Euclidean counterparts and why they differ. Finally, we turn to more advanced topics of the kind needed to construct models of circular data. We introduce the most common circular distribution families, including circular analogs of the Gaussian distribution and their properties, as well as flexible distribution families that can describe skewed and heavy-tailed data. We survey common applications in the [supplemental materials](#). The tutorial takes a method-oriented approach, not tied to specific software, but highlighting factors relevant to experimental design, data analysis, and statistical modeling. A consistent theme is the need to rigorously account for the periodicity of circular data. We also provide a consistent notation for circular statistics and parameters, addressing the lack of established conventions in the field.

Yunxiao Chen served as action editor.

Stella Wernicke  <https://orcid.org/0009-0001-3542-0238>

This research was funded in whole or in part by the UKRI (includes Economic and Social Research Council; Grant BBSRC DTP).

The authors declare no conflicts of interest. Stella Wernicke received a PhD scholarship from the Foundation of German Businesses, financed by the German Federal Ministry of Research, Technology, and Space (BMFTR). Ivan Tomić was supported by the Croatian Science Foundation (UIP-2025-02-6353). Adam Triabhall was funded by the UK Research and Innovation (UKRI) Biotechnology and Biological Sciences Research Council (BBSRC) through a Doctoral Training Program. We thank Puria Radmard and Julie Lehnert for their helpful discussions and comments during the preparation of the article and Andrey Chetverikov for his valuable feedback on the preprint. A version of this article has been posted on a preprint server (Wernicke et al., 2026).

Open Access funding provided by University of Cambridge: This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0; <https://creativecommons.org/licenses/by/4.0>). This license permits copying and redistributing the work in any medium or format, as well as adapting the material for any purpose, even commercially.

Stella Wernicke served as lead for conceptualization, formal analysis, methodology, project administration, software, validation, visualization, and writing—original draft and contributed equally to writing—review and editing. Julie de Falco contributed equally to methodology and writing—original draft and served in a supporting role for visualization and writing—review and editing. Zahara Gironés contributed equally to methodology and writing—original draft and served in a supporting role for software and visualization. Chris Jungerius served in a supporting role for writing—original draft. Ivan Tomić contributed equally to methodology, software, visualization, and writing—original draft and served in a supporting role for formal analysis and writing—review and editing. Adam Triabhall contributed equally to methodology and writing—original draft. Xiaolu Wang contributed equally to methodology, software, visualization, and writing—original draft and served in a supporting role for writing—review and editing. Paul M. Bays served as lead for conceptualization, methodology, supervision, validation, and writing—review and editing, contributed equally to project administration, and served in a supporting role for visualization.

Correspondence concerning this article should be addressed to Stella Wernicke, Department of Psychology, University of Cambridge, Downing Pl, Cambridge CB2 3EB, United Kingdom. Email: [saiw2@cam.ac.uk](mailto:saiw2@cam.ac.uk)

### Translational Abstract

Many areas of behavioral and cognitive research involve circular data, such as periodic temporal data or orientations in space. Yet researchers often apply “conventional” Euclidean statistical methods that can distort conclusions. This tutorial is intended for researchers who are familiar with statistical measures and distributions in Euclidean space and want to understand their circular equivalents. It is also suitable for readers who want to refresh their knowledge of circular statistics or deepen their understanding of more advanced topics related to circular data. We start by introducing different ways of representing circular data and fundamental principles for operating with them, such as arithmetic operations. We then discuss options for visualizing circular data, including the benefits and limitations of different representations. Afterward, we introduce descriptive circular measures of central tendency, spread, and shape of empirical distributions, drawing parallels to the Euclidean analogs where helpful. The topics become progressively more advanced, concluding with an introduction to a range of circular statistical models. In this section, we also discuss circular analogs of the Gaussian (normal) distribution. In the [supplemental materials](#), we survey common applications of circular statistics, including spatial, temporal, and neural data that are naturally periodic or can be reconceptualized as such. In contrast to many previous methodological publications on circular statistics that emphasize statistical tests, this concise and hands-on tutorial focuses on handling circular data in the context of behavioral and cognitive research, experimental design, and statistical modeling. While being software-agnostic, the tutorial highlights common pitfalls to watch out for during practical implementations.

**Keywords:** circular statistics, periodicity, behavior, cognition, experimental design

**Supplemental materials:** <https://doi.org/10.1037/met0000861.supp>

Mathematical training for biological and psychological scientists typically focuses on the analysis of data in Euclidean spaces, particularly measurements that can be described as points on the real number line (referred to as  $\mathbb{R}^1$ ; [Figure 1A](#)), or as coordinates in a plane ( $\mathbb{R}^2$ ) or in three-dimensional space ( $\mathbb{R}^3$ ). However, many biological and behavioral variables have a periodic or cyclical nature that requires different methods of analysis. Measurements in a circular space (denoted  $\mathbb{S}^1$ ) can be described by a single number (as on the real line), but moving far enough through the value range in either direction returns you to the point where you started ([Figure 1B](#)). Such a space is topologically equivalent to a closed loop or circle and fundamentally different from the real line (which stretches to infinity in either direction) or a bounded interval (a section of the real line with two ends). A simple example of a circular space is the directions of a compass: If you describe directions in terms of angles (e.g., measured clockwise from North), then  $0^\circ$ ,  $360^\circ$ , and  $720^\circ$  all refer to the same direction, and the distance between  $0^\circ$  and  $359^\circ$  is the same as the distance between  $0^\circ$  and  $1^\circ$ . The topic of circular statistics describes methods for analyzing random variables in circular spaces. It is a subset of directional

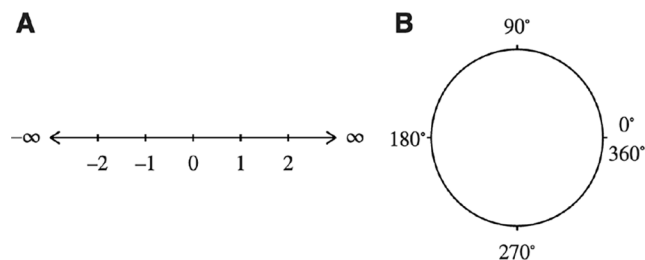
statistics, which extends to variables that can be represented as points on a sphere ( $\mathbb{S}^2$ , which includes directions in three-dimensional space), and analogs in higher dimensions. This tutorial focuses on methods for circular data, as they are widely applicable to behavioral research and also provide a foundation for learning about directional statistics more generally. For an introduction to higher-dimensional directional statistics, refer to ([Mardia & Jupp, 2009](#)) and [Pewsey and García-Portugués \(2021\)](#).

This tutorial targets researchers who are familiar with statistical measures and distributions in Euclidean space and want to understand the corresponding circular equivalents. It develops progressively, from the basic properties and operations needed to present and analyze circular data to more advanced concepts, including the foundations for formal generative models that output circular variables.

Conventions for mathematical notation in circular statistics are less standardized than in Euclidean statistics. In cases where there is a strong convention, we have followed it, but for cases where there is not, we have developed a notation ([Table 1](#) and [Appendix A](#)) that aims to be clear and self-consistent, although it should not be

**Figure 1**

*Value Ranges in One-Dimensional Euclidean and Circular Space*



*Note.* Comparison of the infinite line of real numbers in one-dimensional Euclidean space (A) with the closed set of angles in circular space (B).

**Table 1**

*List of Symbols for Circular Sample and Population Parameters*

| Parameter                              | Sample                 | Population          |
|----------------------------------------|------------------------|---------------------|
| Circular mean                          | $\bar{\theta}$         | $\mu$               |
| Circular median                        | $\theta_{\text{med}}$  | $\mu_{\text{med}}$  |
| Circular mode                          | $\theta_{\text{mode}}$ | $\mu_{\text{mode}}$ |
| Circular variance                      | $\hat{V}_{\theta}$     | $V_{\theta}$        |
| Circular standard deviation            | $s_{\theta}$           | $\sigma_{\theta}$   |
| Circular skewness                      | $\xi_{1\theta}$        | $\gamma_{1\theta}$  |
| Circular kurtosis                      | $\xi_{2\theta}$        | $\gamma_{2\theta}$  |
| Resultant vector                       | $\mathbf{R}$           |                     |
| $p$ th circular moment of a population | $\hat{m}_p$            | $m_p$               |
| Mean resultant vector                  | $\hat{m}_1$            | $m_1$               |
| Resultant length                       | $R$                    |                     |
| Mean resultant length                  | $\bar{R}$              | $\rho$              |

regarded as canonical. Our goal is to give a concise introduction to circular statistics that allows readers to get started with employing them in research involving circular data. For a thorough theoretical introduction to the topic including mathematical derivations and proofs, we recommend [Mardia and Jupp \(2009\)](#). This tutorial will not cover significance tests for circular variables, as there are numerous other sources for these (e.g., [Cremers et al., 2021](#); [Pewsey & García-Portugués, 2021](#), Chapter 5). If readers want to get an overview of the wide range of applications of circular statistics in behavioral and cognitive research before beginning the tutorial, we recommend reading the [supplemental materials on applications](#) first (placed there to meet publication length restrictions). The publications cited in this [supplemental material](#) may also help orient the reader to how circular statistics can be applied to specific topics and types of data later on.

We will begin the tutorial by introducing key characteristics of circular data, different ways of conceptualizing circular values and describing them numerically, and elementary operations on circular variables (Working With Circular Data section). The Visualizations of Circular Data section will provide a brief overview on visually presenting circular data. For an extended discussion of approaches to visualizing circular data, the methodology behind them, and practical guidance for implementing them, see the [supplemental materials](#).

The Descriptive Statistics of Circular Data section will cover the most important statistical measures for circular data, including analogs of mean and variance, and why they differ from their better-known Euclidean counterparts. We will then explore circular statistical models (Probability Theory and Models for Circular Variables section), including the vexed question of circular equivalents to the normal distribution, and concepts of importance to constructing ideal observer and Bayesian models. We will conclude with an overview of further topics in circular statistics that fall outside the scope of this tutorial, providing references to the most useful sources (Beyond This Tutorial section).

## Working With Circular Data

In this section, we describe the fundamental characteristics of circular data to provide the foundation for subsequent sections. We explain the advantages of describing circular values as vectors of unit length in a two-dimensional plane and three standard forms of this vector representation: in polar coordinates, Cartesian coordinates, and on the complex plane. There is no universally correct choice for representing a given circular data set; the optimal representation depends on the context and purpose of the analysis. For instance, angular measurements are often more intuitive for data visualization. If the methods described here are correctly applied, conclusions drawn from circular data will remain valid regardless of the chosen representation.

## Identifying the Periodicity of Data

Finding the appropriate cycle length requires careful consideration, as an unsuitable choice might bias results. One cycle needs to contain every distinct point in the space of possible data values exactly once. Points that are separated by a whole number of cycles must be identical or at least equivalent for the purpose of the analysis. Selecting too narrow a period risks obscuring

important aspects of the data by treating data points as equivalent when they are not: For instance, if temporal data are analyzed on a daily level, but they actually display weekly rhythms. On the other hand, selecting too wide a period will mean the periodic nature of the data is not properly accounted for. [Liu and West \(2016\)](#) discuss a method for identifying an appropriate cycle length for circular analysis of temporal data.

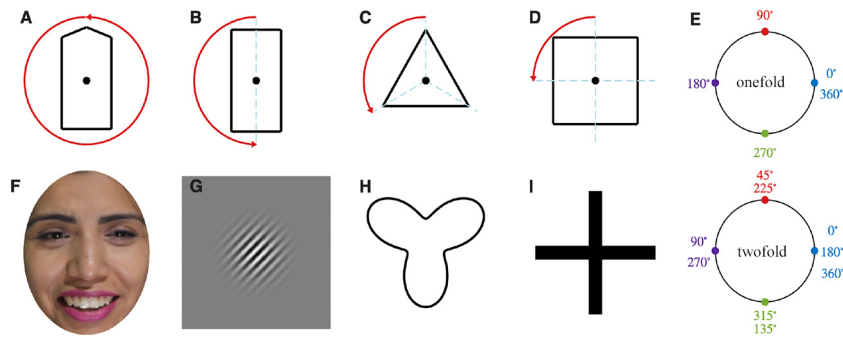
Particular care is needed when working with data that in their raw form are angular or directional, because the appropriate period for circular analysis may not correspond to  $360^\circ$  in the units of the data. If we rotate the black shape in [Figure 2A](#) around the black dot in its center, we will need to rotate  $360^\circ$  before it looks the same again. This stimulus therefore has no (or onefold) *rotational symmetry*, and the circular space of its orientations has a period of  $360^\circ$ . The shape in [Figure 2B](#), in contrast, will first look identical after being rotated through  $180^\circ$  and again every subsequent  $180^\circ$  of rotation. This shape therefore displays twofold rotational symmetry. Analogously, [Figures 2C and 2D](#) show shapes with three- and fourfold symmetry. For shapes with two-fold or higher rotational symmetry, the full space of possible orientations covers less than  $360^\circ$ . It is this smaller period that must be mapped to the full circle for the purposes of circular analysis, and as this will typically be measured in angular units, it is important to keep clear the distinction between the angles of physical orientation and the angles in the circular space of possible values. To illustrate, [Figure 2E](#) shows the orientations corresponding to the cardinal points in the circular space for an object with onefold rotational symmetry (top) and twofold rotational symmetry (bottom). Further care is needed in reporting the results of circular analysis of such data to make clear whether they are in the units of the circular space or of the original data. [Figures 2F–2I](#) show examples of common experimental stimuli used in vision research with one-, two-, three-, and fourfold symmetry, respectively.

## Representations of Circular Data

Having identified the periodicity of our circular data, we can map each observation to its corresponding angle or *phase*, indicating where it falls within one full period or cycle. This requires us to select a data value that will correspond to an angle of 0 and a direction in the data space that will correspond to the positive direction of increasing angle. In some cases, these choices are obvious (e.g., for temporal data with a period of 1 day, one would usually assign 12 o'clock midnight to the zero angle and increasing positive angular values to later times in the day), but in others, they may be arbitrary, or the researcher may have to choose between several viable alternatives.

Angles are typically measured in degrees ( $^\circ$ ), in which case one cycle covers a range of  $360^\circ$ , or radians (rad), in which case (for the reasons explained in the following) one cycle covers  $2\pi$  ( $\approx 6.28$ ) rad. As discussed previously, for data consisting of directions or orientations in space, the angular units of the circular space may differ from those of the raw data. The units in which angular values are reported are a matter of choice, subject to conventions of the field, but some mathematical operations are simplified by using radians. To convert an angle in degrees to radians, multiply by the ratio  $\frac{\pi}{180}$  and divide by the same ratio to convert the other way. Lowercase Greek letters, typically  $\theta$  (theta),  $\phi$

**Figure 2**  
*Rotational Symmetries in Circular Space*



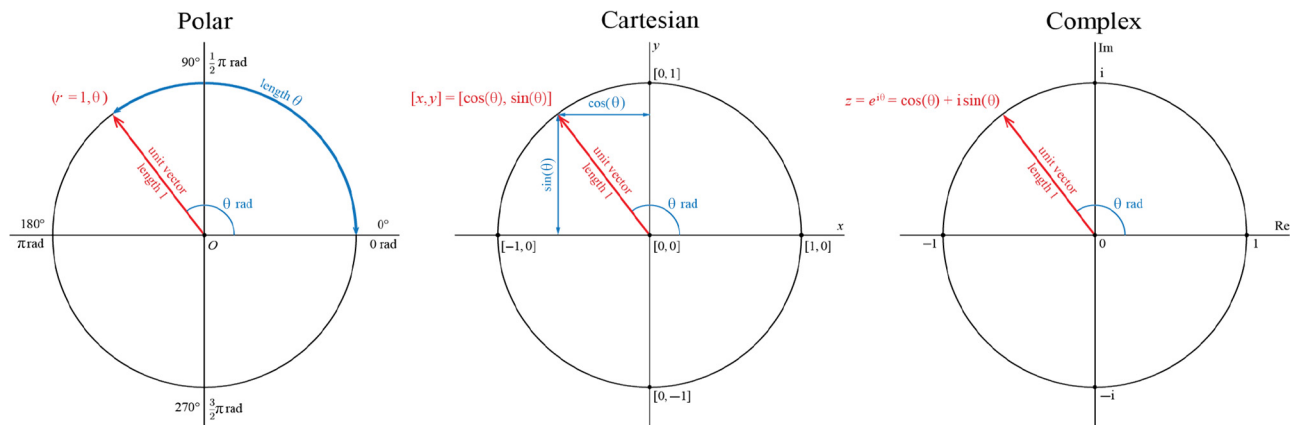
*Note.* (A–D) Abstract stimuli with no or onefold (A), twofold (B), threefold (C), and fourfold (D) rotational symmetry. Black dots show the points around which the stimuli can be rotated; symmetry axes are perpendicular to the page passing through these points. Red arrows indicate one complete rotation in the space of possibilities. (E) Cardinal angles in the circular space of possible orientations marked with corresponding physical rotations in degrees for shapes with onefold (top) and twofold (bottom) rotational symmetry. (F–I) Sample stimuli used in perceptual and memory research. (F) Oval-cropped face with onefold (i.e., no) rotational symmetry. Image from [Mitre-Ortiz and Mitre-Hernandez \(2019\)](#). (G) Gabor patch with twofold symmetry. (H) Radial frequency pattern with a frequency of three cycles per 360° (threefold symmetry) and an amplitude of 50%. Stimuli with different amplitudes can be used to manipulate perceptual and mnemonic discriminability. (I) Fixation cross used for gaze stabilization, with fourfold rotational symmetry. See the online article for the color version of this figure.

(phi), or  $\psi$  (psi), are often used for angular variables in a similar way to  $x$  or  $y$  for Euclidean variables. Similarly, uppercase Greek letters— $\Theta$  (Theta),  $\Phi$  (Phi), or  $\Psi$  (Psi)—are often used for sets or collections of elements analogous to  $X$  or  $Y$  in Euclidean space.

Representing circular data as angles provides a common framework in which to describe and work with different kinds of periodic

data. However, for many of the calculations we will want to perform on circular data, a vector representation (i.e., one combining angle and distance) turns out to be even more powerful. There are three mathematically equivalent ways to represent such a vector, summarized in [Figure 3](#), and each offers distinct advantages depending on the context of application.

**Figure 3**  
*Representations of Circular Data on the Unit Circle*



*Note.* Three mathematically equivalent vector representations of the same circular value  $\theta$ : the polar vector representation (left), the Cartesian vector representation (middle), and the complex vector representation (right). Representations are shown on the unit circle with radius 1. For the polar representation, cardinal angles are annotated in radians (rad) and degrees ( $^\circ$ ). Note that it is a mathematical convention to plot circular data with the zero reference point on the right side of the horizontal axis and a counterclockwise sense of rotation. The angle of  $\theta$  rad corresponds to an arc length  $\theta$  on the circumference of the circle. The circumference of the circle is  $2\pi$ , and a full rotation corresponds to  $2\pi$  rad and  $360^\circ$ . In both Cartesian and complex space, a unit vector (or point on the unit circle) at angle  $\theta$  has horizontal and vertical components  $\cos(\theta)$  and  $\sin(\theta)$ , respectively. See the online article for the color version of this figure.

### Polar Coordinates

Representing a circular value in polar coordinates simply involves supplementing the angular representation of data,  $\theta$ , with a distance or *norm*,  $r$ , resulting in the polar coordinate pair  $(r, \theta)$ . By convention, the norm is set to 1, giving  $(1, \theta)$ . Now each observation is identified not just with a direction but also with a specific point in the two-dimensional plane, and the space of all possible observations is described by a circle of unit radius. Technically, this is an *embedding* of the circular space ( $\mathbb{S}^1$ ) in the real plane ( $\mathbb{R}^2$ ). We can now appreciate the basis for radian units: an angle of  $X$  rad corresponds to a distance  $X$  measured on the circumference of the circle, with a full period corresponding to the full circumference of  $2\pi$ . The polar vector representation is illustrated in the left panel of Figure 3.

### Cartesian Coordinates

We can also describe circular observations as vectors in a Cartesian coordinate system, with origin  $O$  and two perpendicular axes,  $X$  and  $Y$ , passing through  $O$ . By convention, the  $0^\circ$  direction is positive on the  $x$ -axis. In this reference frame, any point on the plane can be represented as a pair of rectangular coordinates  $[x, y]$ , with unit vectors constrained to have  $x^2 + y^2 = 1$ . Compared with a polar representation, this simplifies vector operations (which are used to define statistics like the circular mean; see the Descriptive Statistics of Circular Data section). For example, addition of two vectors simply involves pairwise addition of their Cartesian coordinates. On the other hand, it makes angular operations, such as adding two angles together, more difficult. The Cartesian vector representation is the most easily extended to higher-dimensional directional statistics. For example, spherical data can be represented by unit vectors in three dimensions, described in  $[x, y, z]$  coordinates.

To convert from polar coordinates to Cartesian coordinates, we use the trigonometric functions sine and cosine. For a vector with polar representation  $(r, \theta)$ , the corresponding Cartesian coordinates are

$$[x, y] = [r \cos(\theta), r \sin(\theta)], \quad (1)$$

and for the inverse,

$$(r, \theta) = \left( \sqrt{x^2 + y^2}, \text{atan2}(y, x) \right), \quad (2)$$

where  $\text{atan2}$  is the two-argument arctangent function. These trigonometric functions are available in most programming languages and typically default to using radians as angular units, with  $\text{atan2}$  usually returning angles in the range  $-\pi < \theta \leq \pi$  rad. The Cartesian vector representation is illustrated in the middle panel of Figure 3.

### Complex Coordinates

Finally, each data point can be represented by a unit complex number,  $z$ , interpreted as a vector in the complex plane ( $\mathbb{C}^1$ ). The angle  $\theta$ , measured in radians, and the unit complex number  $z$  are related by Euler's formula:

$$z = e^{i\theta} = \cos(\theta) + i \sin(\theta), \quad (3)$$

where  $i = \sqrt{-1}$  is the imaginary unit. Note that we write the imaginary unit as  $i$  denoted in upright Roman to distinguish it

from the indexing variables  $i$  denoted in italics within the later equations (e.g., summing from  $i = 1$  to  $n$ ). While it may seem esoteric, the complex vector representation of circular observations allows for particularly compact descriptions of both angular and vector operations, as will become clear in subsequent sections.

The complex plane is usually depicted with the horizontal axis as the real component, whereas the vertical axis represents the imaginary component. The unit complex number  $e^{i0}$  evaluates to 1 at  $\theta = 0$ , then moves into the upper half-plane as  $\theta$  grows, tracing the unit circle counterclockwise. This is the basis of the mathematical convention of plotting circular data with the zero reference point on the right side of the horizontal axis and a counterclockwise sense of rotation (see Figure 3), a convention we follow in this tutorial. Nonetheless, many fields deviate from this standard, adopting domain-specific conventions to facilitate data interpretation (e.g., in meteorology, with North =  $0^\circ$  plotted up and clockwise rotation). A Cartesian coordinate  $[x, y]$  corresponds to a complex coordinate  $z = x + iy$ . A polar coordinate  $(r, \theta)$  can be converted to a complex coordinate by either  $z = re^{i\theta}$  or equivalently  $z = r(\cos(\theta) + i \sin(\theta))$ . The complex vector representation is illustrated in the right panel of Figure 3.

### Periodicity, Wrapping, and the Modulo Operation

As described previously, the defining characteristic of circular data is its inherent periodicity, meaning that measurements that differ by a whole number of cycles are considered equivalent. This property fundamentally distinguishes circular from Euclidean space, where proximity and ordering follow conventional numerical relationships. For example, in Euclidean space, 1 is always closer to 3 than 10. In circular spaces, values that may appear numerically distant can in fact be close, for example,  $1^\circ$  and  $359^\circ$  differ by only  $2^\circ$  because  $1^\circ$  and  $361^\circ$  are equivalent angles.

While conceptually any observation in a circular space can be represented by infinitely many angular values, differing by whole numbers of cycles, for practical purposes, it is usually constrained to one unique value within a specific periodic interval. According to the mathematical convention, angular values are constrained to a positive interval  $[0, 2\pi)$  rad. Note that the square bracket indicates the inclusion of the lower bound and the round bracket the exclusion of the upper bound. This specification ensures that the same point is not included twice, which would occur if both bounds were closed, that is, with this definition, an observation can be assigned to 0 but not  $2\pi$ .

The term “wrapping” is often used for the process of mapping circular observations to their representative values within a specified interval (imagine wrapping the real number line around the unit circle, like winding thread onto a bobbin). The process of wrapping angle  $\theta$  to the interval  $[0, 2\pi)$  can be described by a modulo operation,

$$\theta_{\text{wrapped}} = \theta \bmod 2\pi, \quad (4)$$

where  $a \bmod b$  is the *least nonnegative residue of a modulo b*, that is, the smallest number  $\geq 0$  that can be obtained by subtracting an integer multiple of  $b$  from  $a$ .

An alternative convention is to constrain angular values to the symmetric signed interval  $[-\pi, \pi)$ , centered on 0. For this, the wrapping operation is

$$\theta_{\text{wrapped}} = ((\theta + \pi) \bmod 2\pi) - \pi. \quad (5)$$

The same operations can be used to convert angles constrained to one of these conventional intervals into the other. Importantly, these operations preserve the circular topology so that individual observations are not shifted relative to one another; only their numerical representation is changed.

Using the signed interval  $[-\pi, \pi)$  can facilitate data interpretation and visualization when investigating signed deviations or errors relative to a reference angle (see, e.g., [Figure S1D in the supplemental materials on applications](#)) or other observations that tend to cluster around 0. It is also the convention for certain sources of angular data. For example, in sound localization tasks, azimuthal (horizontal) directions are represented with 0 corresponding to the listener's frontal midline and positive values indicating deviations to the listener's right-hand side.

Most programming languages have a built-in modulo function, but, importantly, the implementation for negative inputs frequently differs from the least nonnegative residue assumed previously. As of the software versions available at the time of writing (MATLAB R2026a, R-4.6.1, Python 3.14.7, NumPy 2.5.2, ES2026, C++23), MATLAB's `mod(a, b)`, R's `a%%b`, and Python's `a%b` or `numpy.mod(a, b)` all return a result with the same sign as the divisor  $b$ . This is acceptable for wrapping operations, because the divisor is the length of the desired interval (e.g.,  $2\pi$ ) and so positive. In contrast, JavaScript's `a%b`, C++'s `a%b` and `std::fmod(a, b)`, and Python's `math.fmod(a, b)` return an output equivalent to  $-((-a) \bmod b)$  when  $a$  is negative. One solution is to define your own `mod` function based on the `floor` function, for example, in JavaScript:

```
function mod(a,b) { return(a - b*Math.floor(a/b)) };
```

Alternatively, you can use the following sign-safe versions of [Equations 4](#) and [5](#), which produce acceptable results across programming languages:

$$\theta_{\text{wrapped}} = ((\theta \bmod 2\pi) + 2\pi) \bmod 2\pi \Rightarrow \theta_{\text{wrapped}} \in [0, 2\pi), \quad (6)$$

$$\theta_{\text{wrapped}} = ((\theta \bmod 2\pi) + 3\pi) \bmod 2\pi - \pi \Rightarrow \theta_{\text{wrapped}} \in [-\pi, \pi). \quad (7)$$

## Operations on Angles

When working with Euclidean variables, there exist many operations that map real numbers to other real numbers. By contrast, for circular variables, the need to respect periodic equivalence means there are very few operations that map angles to other angles: Essentially, these are limited to addition, subtraction, and multiplication by integer quantities (equivalent to repeated addition or subtraction). While it is possible to perform other operations on numerical representations of angles, the results will depend on the particular interval you have chosen for representation. For instance,  $0^\circ$  and  $360^\circ$  refer to the same angle, but if you attempt to divide them by 2, you will get  $0^\circ$  and  $180^\circ$ , which are different angles.

Having specified a range of angles in which to represent circular data, we will want to represent the results of angular addition and subtraction within the same interval. Practically, this just means applying a “wrapping” or modulo operation, as described previously,

after the usual arithmetic operation of addition or subtraction. A familiar example of modular arithmetic is the hour hand on a clock dial. If we subtract 2 hours from 1 o'clock, we get 11 o'clock, not  $-1$  o'clock. Similarly, if we add 5 hours to 10 o'clock, we get 3 o'clock and not 15 o'clock.

The modular addition of two angles  $\theta$  and  $\phi$  in the standard interval  $[0, 2\pi)$  is given by

$$(\theta + \phi) \bmod 2\pi \quad (8)$$

and modular subtraction by

$$(\theta - \phi) \bmod 2\pi. \quad (9)$$

The modulo operation ensures that the result always lies in  $[0, 2\pi)$ , regardless of the values of  $\theta$  and  $\phi$ . Modular subtraction can be used to recenter data on a particular reference direction or compute the angular deviation or error of an estimate  $\theta$  from a true value  $\phi$ . In these cases, it may be preferable to wrap the result to the signed interval  $[-\pi, \pi)$  using [Equation 5](#).

Importantly, because a modulo operation changes only the numerical representation of a circular value, it can be applied at every step of an arithmetic calculation or only to the final output without changing the result, for example,

$$((\theta + \phi) \bmod 2\pi) + \psi \bmod 2\pi = (\theta + \phi + \psi) \bmod 2\pi. \quad (10)$$

The symbols  $\oplus$  and  $\ominus$  are sometimes used for modular addition and subtraction, but this is not an established standard, so their intended meaning should always be explicitly stated, including the interval in which their outputs are to be represented.

## Vector Equivalents of Angular Arithmetic

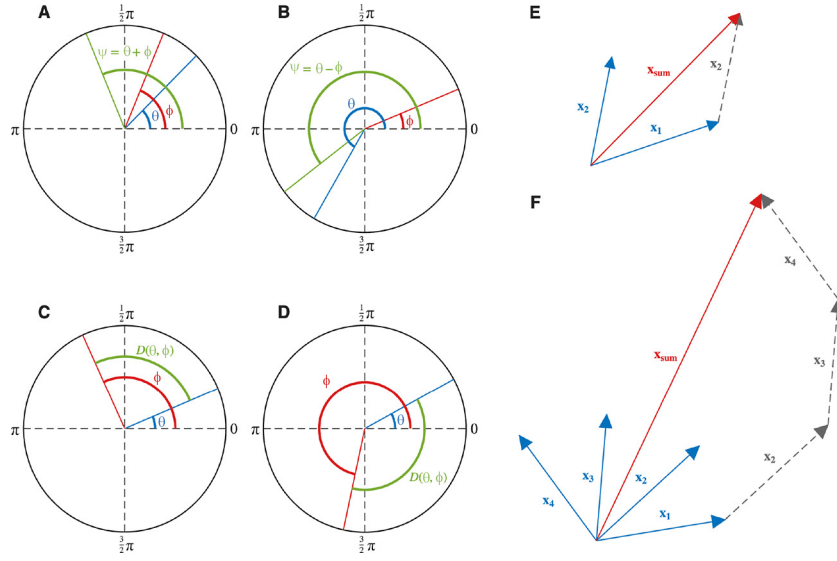
The modular arithmetic formulation is useful for working directly with angles. It is also possible, and often preferable, to implement angular addition and subtraction in a vector representation of circular data. As shown in [Figure 4A](#), in the standard polar coordinate system, adding an angle  $\phi$  to an angle  $\theta$  corresponds to a counterclockwise (positive) rotation of the unit vector corresponding to  $\theta$ . Conversely, subtracting an angle  $\phi$  from  $\theta$  corresponds to a clockwise (negative) rotation ([Figure 4B](#)).

In Cartesian coordinates, rotation of vectors is achieved by multiplying by a *rotation matrix*. A rotation of a vector with coordinates  $[x, y]$  through an angle  $\phi$  about the origin  $O$  can be expressed in matrix form as

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}. \quad (11)$$

The rotation is counterclockwise in standard coordinates if  $\phi$  is positive and clockwise if it is negative. If  $[x, y]$  is a unit vector corresponding to angle  $\theta$ , then  $[x', y']$  is another unit vector corresponding to angle  $\theta + \phi$ . Rotation matrices are a relatively cumbersome way of implementing angular arithmetic, but they find considerable application in computer graphics, including 3D graphics where rotations in three dimensions are expressed analogously as  $3 \times 3$  matrices. Importantly, in computer graphics software, pixel locations on a two-dimensional display are usually represented in a nonstandard “left-handed” Cartesian system, with  $y$  increasing downward instead of upward,  $x$  increasing to the

**Figure 4**  
*Angular Arithmetic and Vector Summation*



*Note.* (A) Modular addition (as defined in Equation 8) of angles  $\theta$  (blue) and  $\phi$  (red) yielding sum  $\psi$  (green). (B) Modular subtraction (as defined in Equation 9) of angle  $\phi$  (red) from angle  $\theta$  (blue) yielding difference  $\psi$  (green). (C, D) Angular distance  $D(\theta, \phi)$  (green) between angles  $\theta$  (blue) and  $\phi$  (red). Although any pair of two angles can be connected by rotation in either direction around the circle, counterclockwise and clockwise rotation lead to the smaller angle in (C) and (D), respectively. (E, F) Vector addition (as defined in Equation 16) of two (E) or multiple (F) vectors. Blue and red arrows show the original vectors  $\mathbf{x}_i$  and summed vectors  $\mathbf{x}_{\text{sum}}$ , respectively. Gray dashed arrows show the tip-to-tail visualization where summing vectors can be viewed as connecting individual vectors sequentially. See the online article for the color version of this figure.

right, and the origin at the top-left corner of the display. In this system, all coordinates lie in the first quadrant where  $x \geq 0$  and  $y \geq 0$ . The inversion of the y-axis reverses the sense of rotation, so for screen coordinates  $[x, y]$ , the rotation described by Equation 11 is clockwise for positive  $\theta$ .

For circular values represented as unit vectors in the complex plane, the addition and subtraction of angles correspond to multiplication and division, respectively. That is, if  $z_1$  and  $z_2$  are two unit complex numbers corresponding to the angles  $\theta$  and  $\phi$ , then their product is

$$z_1 \cdot z_2 = e^{i\theta} \cdot e^{i\phi} = e^{i(\theta+\phi)}, \quad (12)$$

yielding a new unit vector representing the angle  $\theta + \phi$ . Similarly, the subtraction  $\theta - \phi$  corresponds to dividing the vector representation of  $\theta$  by the vector representation of  $\phi$ :

$$\frac{z_1}{z_2} = \frac{e^{i\theta}}{e^{i\phi}} = e^{i(\theta-\phi)}. \quad (13)$$

The complex representation therefore provides a simple and mathematically elegant implementation of angular arithmetic, in which the “wrapping around” of large angle changes is taken care of automatically by the geometry of the complex plane. Many programming languages have support for complex numbers built-in (e.g., in both MATLAB and Python, you can create complex numbers by typing  $1j$  for the imaginary unit  $i$ , and they

will work with all the usual math operators and functions), whereas other languages have well-established libraries that implement them (e.g., `math.js` for JavaScript). Note, however, that the complex representation of circular variables does not have convenient extensions to spherical or higher-dimensional directional variables: The polar and Cartesian representations provide better starting points for working with those kinds of data.

### Measuring Distance

The periodic equivalence of circular data means careful consideration must be given to the measurement of distances. The Euclidean distance between two points on the real number line is just the absolute value of their difference, calculated equivalently as either  $|x - y|$  or  $|y - x|$ . By contrast, given two angles  $\theta$  and  $\phi$  corresponding to two points  $A$  and  $B$  on the unit circle, their angular separation can be measured either clockwise or counterclockwise, yielding two different distances that add up to one complete cycle. The *angular distance* is typically taken as the smaller of the two, equivalent to the shortest arc length between  $A$  and  $B$  (Figures 4C and 4D). This can be calculated as

$$\begin{aligned} D(\theta, \phi) &= \min(|\theta - \phi|, 2\pi - |\theta - \phi|) \\ &= \pi - |\pi - |\theta - \phi||, \end{aligned} \quad (14)$$

where the circular variables  $\theta$  and  $\phi$  must be represented in the same interval of range  $2\pi$ , for example,  $[0, 2\pi)$  or  $[-\pi, \pi)$ . The resulting distance always lies in  $[0, \pi]$ , and it captures the magnitude of the angular separation, without regard to its direction. Note that the angular distance is not itself a circular variable.

An alternative measure of separation between circular variables, and one that is in a sense more “natively” circular, is the *cosine dissimilarity* (also *cosine error*), defined as

$$D_{\cos}(\theta, \phi) = 1 - \cos(\theta - \phi) = 1 - \cos(\phi - \theta). \quad (15)$$

The cosine dissimilarity lies in the range  $[0, 2]$ , with 0 and 2 corresponding to the minimum and maximum possible separation of 0 and  $\pi$  rad, respectively. This measure of distance can be conceptualized as a circular analogue of the squared Euclidean distance and plays a similarly important role in circular statistics, including in the definition of circular variance, as we will discuss in the Circular Variance section. Indeed, with respect to the vector representation of angles, the cosine dissimilarity is equal to half the squared straight-line distance between points on the unit circle, that is,  $\frac{1}{2} \|\mathbf{A} - \mathbf{B}\|^2$  in the Cartesian representation or  $\frac{1}{2} |z_1 - z_2|^2$  in the complex representation. For unit vectors with small differences in angle  $\delta$ , the straight-line distance and the angular distance are approximately equal, so the cosine dissimilarity  $1 - \cos(\delta)$  converges to  $\frac{\delta^2}{2}$ . Note that, like angular distance, the cosine dissimilarity is not a circular variable.

In study design, circular distance becomes relevant when a minimum or maximum separation between angular values must be ensured. Such constraints allow continuous rather than discrete variables while keeping points neither too close nor too far apart. For instance, experimental stimuli could be sampled with a minimum circular distance, so they remain perceptually distinct, or ambulatory-assessment surveys scheduled at random times subject to a maximum gap between them. A method for sampling from a uniform distribution in circular space while ensuring a minimum separation between values is provided in [Appendix B](#).

## Vector Summation

We described previously how addition and subtraction of angles can be represented as rotation of unit vectors in the real or complex plane. In subsequent sections, we will see that vector representations provide a general foundation for the analysis of circular data. This includes the definition of circular analogues of mean and variance, which are based on vector summation.

The vector sum of two unit vectors,  $\mathbf{x}_1$  and  $\mathbf{x}_2$ , is illustrated geometrically in [Figure 4E](#), with vectors represented by arrows. If we shift one vector arrow, for instance,  $\mathbf{x}_2$ , without rotating it, from the origin  $O$  to a point where its tail coincides with the tip of the other vector arrow  $\mathbf{x}_1$ , then we can draw a vector sum, also called the *resultant vector*, as an arrow connecting the origin  $O$  with the tip of the shifted vector. The three vectors  $\mathbf{x}_1$ ,  $\mathbf{x}_2$ , and  $\mathbf{x}_1 + \mathbf{x}_2$  form the so-called vector triangle.

To sum vectors defined in Cartesian coordinates, add up  $x$  and  $y$  coordinates separately:

$$\mathbf{x}_1 + \mathbf{x}_2 = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_1 + x_2 \\ y_1 + y_2 \end{bmatrix}. \quad (16)$$

For a complex vector representation, real and imaginary parts sum separately:

$$z_1 + z_2 = (a + ib) + (c + id) = a + c + i(b + d). \quad (17)$$

When summing more than two vectors, the order of addition does not affect the result. For example, we have  $(\mathbf{x}_1 + \mathbf{x}_2) + \mathbf{x}_3 = \mathbf{x}_1 + (\mathbf{x}_2 + \mathbf{x}_3)$ . Geometrically, each vector can be placed tip-to-tail in sequence. The resulting chain of vectors forms a vector polygon, and the resultant vector is the one that connects the origin to the tip of the final vector in the sequence ([Figure 4F](#)).

## Visualizations of Circular Data

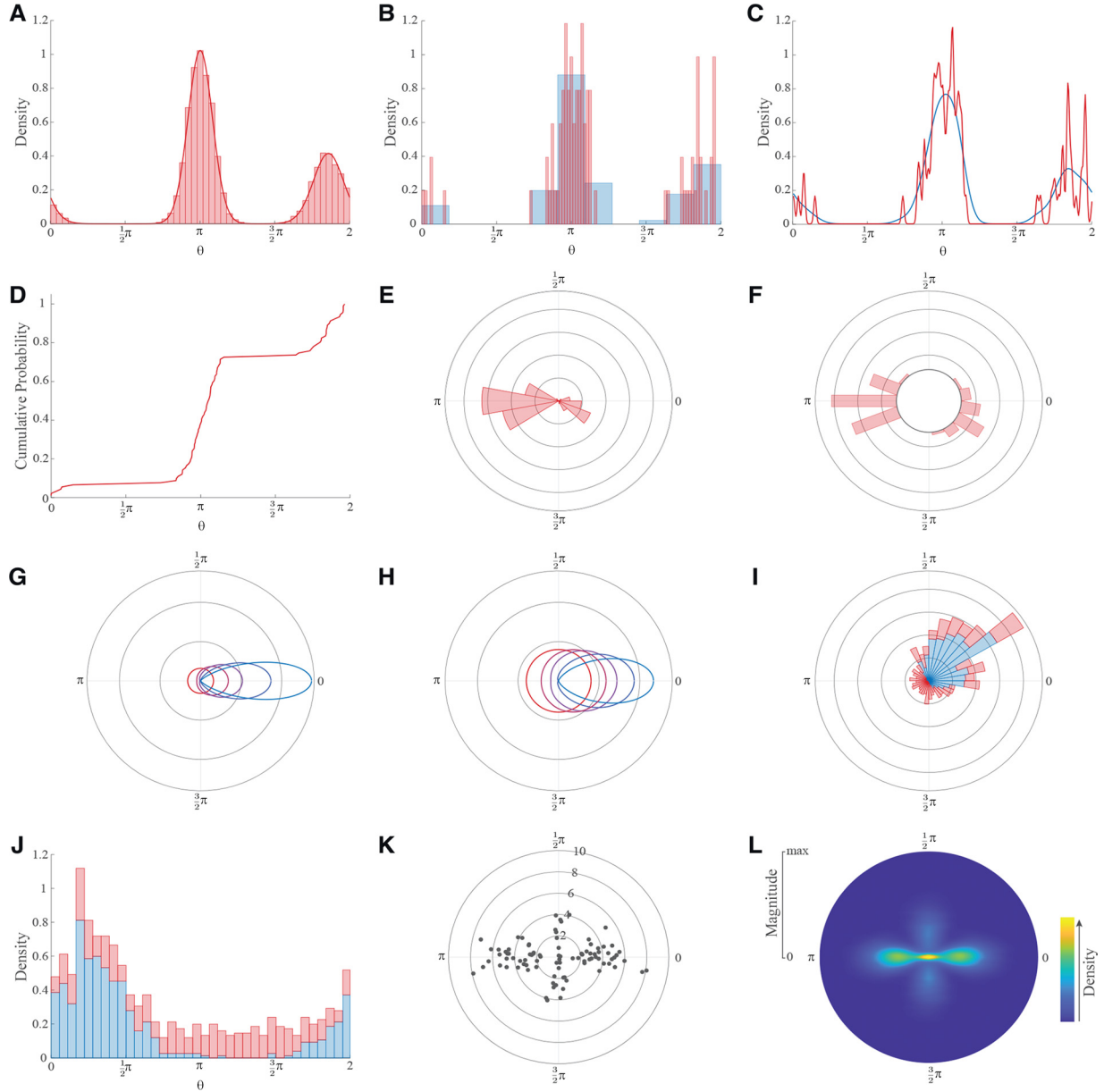
In this section, we briefly discuss approaches for visualizing circular data. Due to publication length restrictions, an extended version of this section, including practical guidance on producing these visualizations in commonly used programming languages, is provided in the [supplemental materials](#).

Visualizing circular data requires approaches that respect its inherent periodicity and directional nature. A key decision to make is whether data are unwrapped onto a line and represented in a Euclidean coordinate system ([Figures 5A–5D](#), [5J](#), and [6D](#)) or represented in a two-dimensional polar coordinate system ([Figures 5E–5L](#), [5K–5L](#), and [6A–6C](#))—where angle is mapped to direction ( $\theta$ ) and magnitude to radius ( $r$ ). In some cases, linear and polar visualizations convey similar information, whereas in others, one format provides a much clearer or more intuitive representation of the data. When creating a linear plot of circular data, a suitable interval must be chosen to prevent the density of a mode positioned close to the boundary from spilling over to the opposite side of the plot ([Figure 5A](#)). In addition, the same considerations for choosing bins must be made as for linear plots of Euclidean data, for example, deciding on a suitable bin width ([Figure 5B](#)). For many linear plots likely familiar to the reader, there are several options for polar analogs. In polar histograms, for instance, bins can be represented by wedges radiating from the center of a circle ([Figure 5E](#)) or by rectangular bars arranged around a reference circle ([Figure 5F](#)). Polar representations can also be suitable for visualizing cylindrical data that combine a circular and a Euclidean variable ([Figures 5K and 5L](#)) and for displaying individual data points alongside summary statistics ([Figures 6A and 6B](#)). Like other approaches to representing circular data, box plots can use both polar ([Figure 6C](#)) and linear ([Figure 6D](#)) coordinates.

## Descriptive Statistics of Circular Data

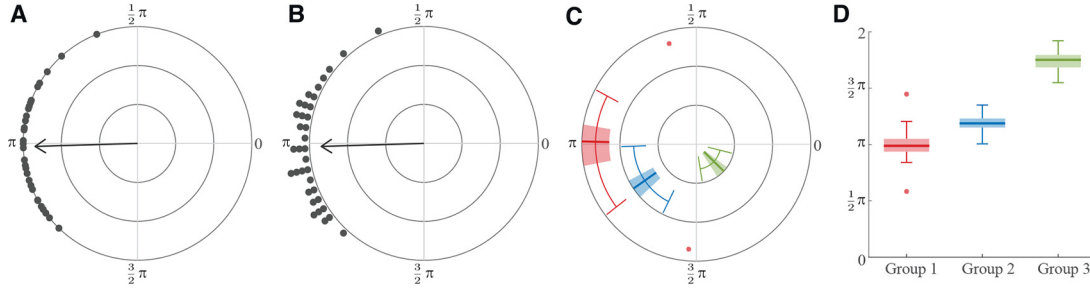
In this section, we will review the descriptive statistics most commonly used to summarize circular data sets. For each statistic, we will give its mathematical definition, the intuition behind it, and its relationship to the corresponding Euclidean statistic. Where applicable, we will also discuss the circumstances under which Euclidean approximations of circular measures provide acceptable accuracy. As mentioned in the introduction, mathematical notation in circular statistics is less standardized than in Euclidean contexts. Some of the symbols introduced in this section represent our own suggested notation (for an overview, see the middle column in [Table 1](#)). Due to publication length constraints, we discuss outlier classification and exclusion in the [supplemental materials](#). For a discussion of the corresponding population parameters and their

**Figure 5**  
*Visualization of Circular Data in Linear and Polar Spaces*



*Note.* (A) Linear histogram of a bimodal circular distribution. Bars show the empirical density estimated from a random sample ( $N = 9,000$ ), and the line shows the corresponding theoretical distribution. (B–F) All plots show the same sample data ( $N = 90$ ) drawn from the theoretical distribution shown as the line in (A). (B) Effect of bin widths on the representation of circular data. Both histograms show the same data but with coarse (blue, 11 bins) and fine (red, 111 bins) binning. (C) Effect of smoothing strength on circular density estimation. Both curves represent the same data with stronger (blue) and weaker (red) smoothing. (D) Empirical cumulative distribution function. (E, F) Polar histograms showing density as a function of angle: polar histogram with wedges (E) and wrapped bar histogram with rectangular bars (F). (G) Polar plot of von Mises distributions (circular normal distributions; see the Probability Theory and Models for Circular Variables section) with probability density plotted in the radial dimension. Note that dense regions appear exaggerated, whereas sparse regions are understated because the area scales with the square of the radius. (H) Polar plot of the same von Mises distributions as in (G) with the square root of probability density on the radial axis. This area-proportional mapping ensures that enclosed areas are linearly proportional to density. (I, J) Stacked histograms of the same data with two subgroups (blue and red) layered one on top of another in polar (I) and linear circular space (J). Note that the red subgroup is approximately uniformly distributed. However, as the area grows with the square of the radius in rose plots, bins of the red subgroup appear larger for bins where density in the blue subgroup is higher in the polar representation (I). (K, L) Representations of cylindrical data with one circular and one Euclidean dimension. (K) Polar scatter plot showing cylindrical data with the circular and Euclidean dimensions represented by angle and distance from the center, respectively. (L) Polar heat map showing a large number of samples drawn from the same distribution as data in (K). Density is represented smoothly by a change in hue across both angular and radial dimensions. See the online article for the color version of this figure.

**Figure 6**  
*Visualization of Summary Statistics*



*Note.* (A, B) Rayleigh plots. The arrows indicate the mean resultant vector (see Circular Mean section). (A) Rayleigh plot showing 35 individual observations (dark gray circles). Note that it is not possible to see whether multiple observations share the same value. (B) Stacked Rayleigh plot of the same data shown in (A) but now binned into  $5^\circ$  intervals, with observations stacked radially within each bin. (C, D) Box plots summarizing three groups of circular observations with circular median (bold line), interquartile range from the 25th to the 75th percentile (opaque box), range of data not including outliers (whiskers), and outlier values (dots). Note that data dispersion is similar for Group 1 (red) and Group 3 (green), which is apparent in the linear representation (D). In the polar representation (C), the radial distance between the two groups gives the impression that Group 1 has a higher dispersion. See the online article for the color version of this figure.

estimation, see the Probability Theory and Models for Circular Variables section.

## Central Tendency

### Circular Mean

A basic requirement of any statistic calculated from circular data is that it respects the equivalence of angles separated by a whole number of cycles, for example, that it is unchanged if we describe a data point as  $360^\circ$  instead of  $0^\circ$ . The usual definition of the arithmetic mean  $\frac{1}{n} \sum_{i=1}^n x_i$  does not fulfill this requirement. However, the problem of identifying a circular counterpart to the Euclidean mean turns out to be even more fundamental in nature. Consider a data sample consisting of the four cardinal directions ( $0^\circ$ ,  $90^\circ$ ,  $180^\circ$ , and  $270^\circ$ ). Intuitively, no single direction can be identified as its average. It is clear that a different conception of mean value is required for circular data: the centroid, or center of mass, fulfills the requirements.

In Euclidean space, the arithmetic mean is equivalent to the centroid of the set of points  $x_i$  on the real number line. This idea can be illustrated using the analogy of a seesaw: Placing equally weighted balls at each location  $x_i$  on a beam, the mean corresponds to the center of mass or the position of a fulcrum that balances the system. We can extend this concept to circular data by treating each angle  $\theta_i$  as a point on the unit circle (the embedded representation introduced in the Polar Coordinates section). In this case, the centroid lies inside the circle rather than on its circumference (purple dots in Figure 7). The analogy is to a disk with equally weighted balls placed along its edge at angles  $\theta_i$ . Placing a small plinth underneath the disk to keep it balanced would require positioning it closer to the center and not on the edge. The position of the plinth reflects the location of the center of mass.

The direction of the center of mass from the origin defines the circular mean—a generalization of the linear mean that respects the periodic and directional nature of circular data. It is the same as the direction of the resultant vector (the sum of the  $n$  unit vectors corresponding to the data; see the Vector Summation section).

The two-dimensional location of the centroid can be obtained by scaling the resultant vector by  $\frac{1}{n}$ . The length of this *mean resultant vector* contains information about the spread of the data, as we will see in the Measures of Dispersion section. If the mean resultant vector has length 0, the circular mean is undefined, as this indicates complete directional cancellation, as in a uniform distribution.

As described in the Complex Coordinates section, a two-dimensional vector can be conveniently represented by a complex number. When represented in this way, we will denote the mean resultant vector by  $\hat{m}_1$  (the origin of this notation will become clear in the Circular Moments section). It is defined as

$$\hat{m}_1 := \frac{1}{n} \sum_{i=1}^n e^{i\theta_i} = \bar{R} e^{i\bar{\theta}}, \quad (18)$$

where  $\bar{R} = |\hat{m}_1|$  is the modulus or length of the mean resultant vector and  $\bar{\theta} = \arg(\hat{m}_1)$  is its argument or direction, which defines the circular mean.

The mean resultant vector can alternatively be represented in polar coordinates, as  $(\bar{\theta}, \bar{R})$ , or in Cartesian coordinates as

$$\frac{1}{n} \left[ \sum_{i=1}^n \cos(\theta_i), \quad \sum_{i=1}^n \sin(\theta_i) \right]. \quad (19)$$

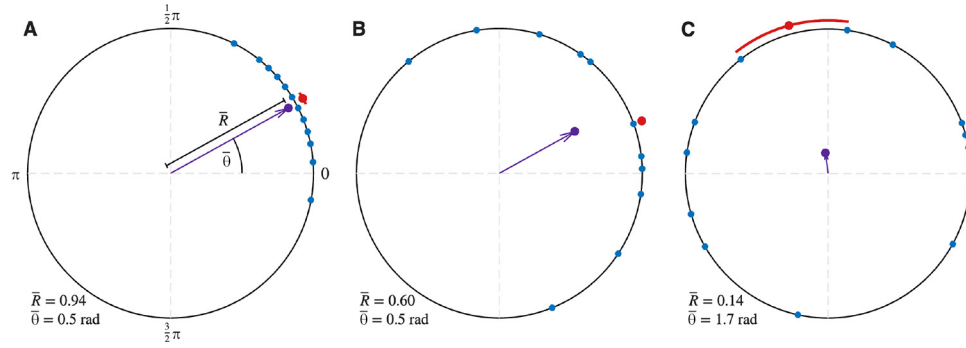
$\bar{R}$  can also be calculated using the Pythagorean formula with the  $x$ - and  $y$ -components of the Cartesian resultant vector,

$$\bar{R} = \sqrt{\left( \frac{1}{n} \sum_{i=1}^n \cos(\theta_i) \right)^2 + \left( \frac{1}{n} \sum_{i=1}^n \sin(\theta_i) \right)^2}, \quad (20)$$

and the circular mean  $\bar{\theta}$  can be obtained by calculating the angle between the positive  $x$ -axis and the resultant vector,

$$\bar{\theta} = \text{atan2} \left( \frac{1}{n} \sum_{i=1}^n \sin(\theta_i), \frac{1}{n} \sum_{i=1}^n \cos(\theta_i) \right), \quad (21)$$

**Figure 7**  
Center of Mass and Central Tendency of Circular Data



*Note.* The panels display samples of 12 (A, C) or 11 (B) data points on the unit circle. Each data point  $\theta_i$  corresponds to a complex number  $e^{i\theta_i}$  with unit length. The purple dots and arrows show the center of mass and the mean resultant vector for each sample. The circular mean,  $\bar{\theta}$ , is the angle of this vector, and the mean resultant vector,  $\bar{R}$ , represents its length. The sample in (B) has the same circular mean as the one in (A), but the center of mass is closer to the center of the circle as the sample displays a higher spread. (C) displays the most dispersed sample, with the center of mass close to the center of the circle. The red points in (A–C) show possible values of the median of the different samples; the red arcs in (A) and (C) mark the range within which each median could fall. Note that the median of the odd number of data points in (B) is unique. Medians are presented outside of the circle to avoid overlapping with data points. See the online article for the color version of this figure.

recalling from the Representations of Circular Data section that `atan2` first takes the vertical component and then the horizontal component.

The circular mean  $\bar{\theta}$  is the angle that maximizes the sum of cosines of angular distances between each data point and that angle:

$$\bar{\theta} = \arg \max_{\phi} \sum_{i=1}^n \cos(\theta_i - \phi). \quad (22)$$

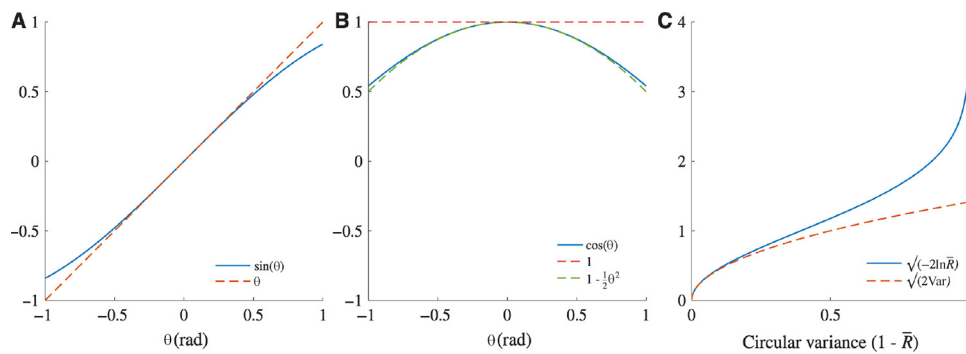
Equivalently, we can say it is the angle with the minimum summed cosine dissimilarity from the data (see the Measuring Distance section). This is analogous to the way the Euclidean mean minimizes the sum of squared deviations from the data.

In the special case that data points are concentrated in a small region of the circular space, some (but not all) circular statistics

can be approximated by their Euclidean counterparts. Specifically, if we choose an interval in which to represent the circular data such that the data are tightly concentrated in their center (e.g.,  $[-\pi, \pi]$  if the data are concentrated around 0), then the “wrapping around” at the bounds of the interval will have a negligible effect. In this case, the Euclidean mean will closely approximate the circular mean.

Formally, we can approximate sine and cosine by  $\sin(\theta_i) \approx \theta_i$  and  $\cos(\theta_i) \approx 1$  (red dashed lines in Figures 8A and 8B). Inserting these approximations into Equation 21 yields the Euclidean mean:  $\bar{\theta} \approx \text{atan2}(\frac{1}{n} \sum_{i=1}^n \theta_i, 1) = \frac{1}{n} \sum_{i=1}^n \theta_i$ . However, this approximation of cosine is only sufficiently precise for very small angles  $\theta_i \approx 0$ . As an alternative that is suitable for concentrated data that span a larger arc, we can use the second-order expansion  $\cos(\theta_i - \phi) \approx 1 - \frac{1}{2}(\theta_i - \phi)^2$  (green dashed line in Figure 8B). If we apply this to

**Figure 8**  
Euclidean Approximations of Circular Measures



*Note.* (A, B) Approximations of sine (A) and cosine (B) for small values. (C) Approximation of circular variance; relationship between circular variance and circular standard deviation. Blue continuous lines represent exact values; dashed lines indicate approximations. See the online article for the color version of this figure.

Equation 22, we obtain  $\bar{\theta} \approx \arg \min_{\phi} \sum_{i=1}^n (\theta_i - \phi)^2$ . The right-hand side coincides with the variational characterization of the Euclidean mean that defines it as the value that minimizes the sum of its squared distances from all data points.

### Circular Median

In Euclidean space, the median of a sample separates the data into an upper half and a lower half. Similarly, the sample *circular median* can be described as splitting the data into a clockwise half, contained in the interval  $(\bar{\theta}_{\text{med}} - \pi, \bar{\theta}_{\text{med}})$ , and a counterclockwise half, in  $(\bar{\theta}_{\text{med}}, \bar{\theta}_{\text{med}} + \pi)$ . However, in circular space, for any direction  $\phi$  that meets this criterion, the diametrically opposite direction  $\phi \pm \pi$  does so as well. To resolve this ambiguity, the circular median is defined as a point on the circle that minimizes the sum of angular distances between each data point and that angle,

$$\bar{\theta}_{\text{med}} := \arg \min_{\phi} \sum_{i=1}^n D(\theta_i, \phi), \quad (23)$$

where  $D(\theta_i, \theta_j)$  denotes the shortest arc length between two angles (see the Measuring Distance section). Of the two opposite directions that split the data in two, this identifies the circular median as the one more strongly in the overall direction of the data. This definition is analogous to the Euclidean median, which can be defined as a point that minimizes the summed absolute distance from the data. If the number of data points is odd, one of them satisfies Equation 23 (red dot in Figure 7B), but if the number is even, the set of solutions is an interval subtended by two neighboring data points (red arcs in Figures 7A and 7C). In this case, it is usual to take the midpoint of the interval as the circular median. This all corresponds closely to the way the median is identified in Euclidean data. Note, however, that for certain precise geometric arrangements of data points, the circular median is still not uniquely defined. As a trivial example, for any odd number of data points arranged evenly around the circle, every data point has an equal claim to be the circular median.

The circular median is less sensitive to extreme observations or *outliers* than the circular mean and can provide a robust measure of central tendency, similarly to the median versus the arithmetic mean in Euclidean space (see the [supplemental materials on outlier classification and exclusion](#)). When most data points are clustered in a small region in the interior of an interval of  $2\pi$ , the circular median is approximately equal to the Euclidean median calculated over that interval.

### Circular Mode

The circular mode of a sample  $\theta_{\text{mode}}$  is the most frequently occurring angle, taking into account the equivalence of angles separated by whole numbers of cycles. As in the Euclidean case, the circular mode is strictly only defined for discrete data, but the term may also be used in a looser sense to describe peaks in a histogram or smoothed density estimates of continuous data. Whereas the global mode is rarely used to estimate central tendency for continuous circular data, local mode estimates are relevant for several applications, including multimodality testing (Ameijeiras-Alonso et al., 2019; Bolón et al., 2025) and mode-based clustering approaches (Chang-Chien et al., 2012; Zhang & Chen, 2021).

## Measures of Dispersion

### Circular Variance

As mentioned previously and displayed in Figure 7, the mean resultant length  $\bar{R}$  conveys information about the spread of a data sample by quantifying how tightly the angles cluster around the mean direction. Specifically,  $\bar{R}$  lies in the range  $0 \leq \bar{R} \leq 1$ , with  $\bar{R} = 1$  indicating perfect alignment of all angles (i.e., maximum concentration) and  $\bar{R} = 0$  implying complete directional cancellation (i.e., maximum dispersion). The sample *circular variance* is defined as

$$\hat{V}_{\theta} := 1 - \bar{R}, \quad 0 \leq \hat{V}_{\theta} \leq 1, \quad (24)$$

reflecting the inverse relationship between  $\bar{R}$  and spread.

The circular variance is equal to the average cosine dissimilarity between the data points and their circular mean,

$$\hat{V}_{\theta} = \frac{1}{n} \sum_{i=1}^n (1 - \cos(\theta_i - \bar{\theta})), \quad (25)$$

in the same way that the Euclidean variance is the average squared deviation from the arithmetic mean. Unlike many other circular statistics, the circular variance of clustered data is not approximately equal to the corresponding Euclidean measure. Instead, for data that are concentrated within a small arc of the circle, circular variance is approximately *half* the value of the Euclidean sample variance (red dashed line in Figure 8C). This is a consequence of the fact that for small  $\delta$ ,  $1 - \cos(\delta) \approx \frac{\delta^2}{2}$ .

### Circular Standard Deviation

In Euclidean statistics, variance is expressed in squared units of the data; to quantify dispersion in the same units as the data themselves, the standard deviation is defined as the square root of the variance. However, applying this transformation to the circular variance, which takes values in  $[0, 1]$ , would not yield a measure of spread in radians or any other natural angular units. Instead, the *circular standard deviation*  $s_{\theta}$  is defined as

$$s_{\theta} := \sqrt{-2 \ln \bar{R}} = \sqrt{-2 \ln (1 - \hat{V}_{\theta})}. \quad (26)$$

This definition might seem arbitrary and a bit peculiar, but it has a logic based on the Gaussian or normal distribution. Specifically, if you wrap a Gaussian probability distribution with standard deviation  $\sigma$  around the unit circle (see the Wrapped Normal Distribution section), the resulting distribution will have a circular standard deviation equal to  $\sigma$  based on this definition. In other words,  $s_{\theta}$  is the standard deviation of a normal distribution wrapped around the circle that has the same mean resultant length as the data.

Importantly, the logarithm in Equation 26 converts the bounded range of  $\bar{R}$  into an unbounded scale. As  $\bar{R}$  approaches 1,  $s_{\theta}$  approaches 0, corresponding to maximum concentration. Conversely, as  $\bar{R}$  approaches 0,  $s_{\theta}$  tends to infinity, that is, maximum spread (blue line in Figure 8C). Thus,  $s_{\theta}$  can in principle take arbitrarily large values, just like the Euclidean standard deviation, even though the data are geometrically confined to a

circle. Provided  $s_0$  is substantially smaller than the maximum angular distance of a point from the circular mean ( $s_0 \ll \pi$ ), it can be interpreted as a standardized measure of dispersion in angular units, directly comparable with  $s$  in Euclidean space. However, once  $s_0$  approaches or exceeds  $\pi$ , although it remains a valid measure of spread, the interpretation in angular units breaks down, because the topology of the circle means angular distances are confined to at most  $\pi$  rad.

For data concentrated within a small arc of the circle,  $s_0$  can be well-approximated by its Euclidean equivalent  $s$ . In this regime,  $\bar{R}$  is close to 1; thus, the circular variance  $\hat{V}_0 = 1 - \bar{R}$  is close to 0. Using the Taylor expansion  $\ln(1+x) \approx x$  for small  $x$ , we obtain  $s_0 = \sqrt{-2\ln(1 - \hat{V}_0)} \approx \sqrt{2\hat{V}_0}$ . As discussed in the Circular Variance section, for concentrated data,  $2\hat{V}_0$  is approximately equal to the Euclidean sample variance, so  $s_0$  is approximately equal to the Euclidean sample standard deviation,  $s$ .

### Circular Moments

To calculate circular mean, variance, and standard deviation, we used the mean resultant vector, which can also be characterized as the *first circular moment* of the data distribution. Circular moments, also called *trigonometric moments*, form an important class of summary statistics. They act as a “fingerprint” of the distribution, allowing complex patterns to be summarized by a few interpretable numbers.

In Euclidean statistics, the  $p$ th moment of the sample distribution is the arithmetic mean of the  $p$ th power of the data values,  $\frac{1}{n} \sum_i x_i^p$ . However, this definition does not transfer directly to circular data, because it ignores periodicity and directionality. To properly capture the geometry of the circle, powers of  $\theta$  are replaced with powers of the complex exponential  $e^{i\theta} = \cos(\theta) + i \sin(\theta)$ , which is periodic with period  $2\pi$ .

The  $p$ th sample circular moment is therefore defined as

$$\hat{m}_p := \frac{1}{n} \sum_{i=1}^n e^{ip\theta_i} = \bar{R}_p e^{i\bar{\theta}_p}. \quad (27)$$

The first circular moment,  $\hat{m}_1$ , is the mean resultant vector (see the Circular Mean section), its modulus  $|\hat{m}_1|$  is the mean resultant length  $\bar{R}$ , and its argument or phase  $\arg(\hat{m}_1)$  is the circular mean  $\bar{\theta}$ . More generally, the modulus and argument of each circular moment  $\hat{m}_p$  yield interesting statistics.

The argument of the  $p$ th circular moment,

$$\bar{\theta}_p := \arg(\hat{m}_p) = \text{atan2}\left(\frac{1}{n} \sum_{i=1}^n \sin(p\theta_i), \frac{1}{n} \sum_{i=1}^n \cos(p\theta_i)\right), \quad (28)$$

corresponds to the mean direction of the data when considered modulo  $\frac{2\pi}{p}$ , that is, when directions separated by  $\frac{2\pi}{p}$  are treated as equivalent. Higher-order moment arguments generalize the circular mean to capture dominant symmetries in circular data. For instance,  $\bar{\theta}_2$  gives the dominant axis when opposite directions are treated as equivalent, whereas  $\bar{\theta}_3$  points to the orientation of a threefold (trimodal) arrangement.

The modulus or length of the  $p$ th moment

$$\begin{aligned} \bar{R}_p &:= |\hat{m}_p| \\ &= \sqrt{\left(\frac{1}{n} \sum_{i=1}^n \cos(p\theta_i)\right)^2 + \left(\frac{1}{n} \sum_{i=1}^n \sin(p\theta_i)\right)^2}, \quad 0 \leq \bar{R}_p \leq 1, \end{aligned} \quad (29)$$

measures how strongly the data cluster when considered modulo  $\frac{2\pi}{p}$ . Across orders, a value of  $\bar{R}_p = 0$  indicates complete uniformity around the circle.  $\bar{R}_p = 1$  corresponds to perfect alignment of the data wrapped onto a circle of length  $\frac{2\pi}{p}$ , which has different implications for different values of  $p$ . For  $p = 1$ , this means all observations fall at the same angle, for  $p = 2$ , they can be split between two opposite angles, and so forth.

While the argument and the modulus of the first circular moment describe central tendency and spread, higher order circular moments allow us to characterize further aspects of a distribution’s shape and symmetry. We turn to these measures in the following sections.

### Circular Skewness

Skewness describes the asymmetry or lopsidedness of a distribution. In Euclidean statistics, a nonzero skewness statistic indicates that the data spread further from the mean in one direction than the other, with its sign corresponding to the direction of the long tail; it is calculated as the third moment of the  $z$ -scored (standardized) data. *Circular skewness* extends this concept to periodic data, quantifying the asymmetry of a distribution around the circular mean. The sample circular skewness is defined by [Mardia \(1975\)](#) as

$$g_{10} := \frac{\bar{R}_2 \sin(\bar{\theta}_2 - 2\bar{\theta})}{(1 - \bar{R})^{\frac{3}{2}}} = \frac{1}{\hat{V}_0^{\frac{3}{2}}} \frac{1}{n} \sum_{i=1}^n \sin(2(\theta_i - \bar{\theta})), \quad (30)$$

where the second expression illustrates that, analogously to  $z$ -scoring in Euclidean space, data are centered on the mean  $\bar{\theta}$  and normalized by the variance  $\hat{V}_0$ . In line with its Euclidean equivalent, a circular skewness of 0 indicates symmetry, whereas larger deviations from 0 reflect increasing asymmetry ([Figure 9A](#)). However, a positive value indicates that the long tail is in the negative (clockwise) direction from the mean (green line in [Figure 9A](#)), whereas a negative value indicates a longer positive (counterclockwise) tail (red line in [Figure 9A](#))—the opposite sense to the Euclidean skewness. For data concentrated in the interior of a  $2\pi$  interval, the circular skewness is not approximated by the Euclidean skewness  $g_1$  directly but rather by

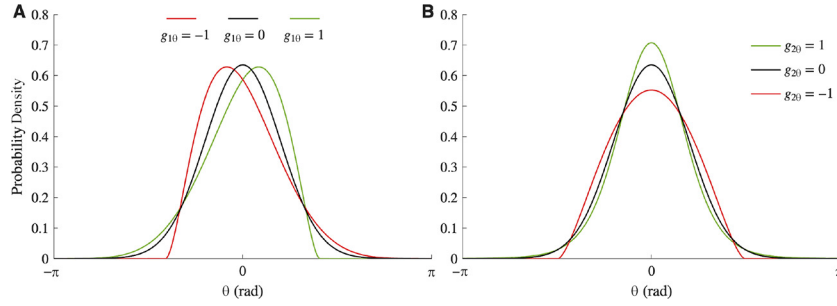
$$g_{10} \approx -2^{\frac{3}{2}} g_1, \quad (31)$$

where the negative sign reflects the reversal of sense described previously and the power-of-two scaling is due to the fact that circular variance approaches half the Euclidean variance for concentrated data (see the Circular Variance section).

### Circular Kurtosis

Kurtosis describes how much probability mass lies in the tails of a distribution. The Euclidean sample kurtosis  $b_2$  is the fourth moment of the  $z$ -scored data. As with skewness,  $z$ -scoring makes the kurtosis invariant to shifts or scaling of the data. A Gaussian distribution has a kurtosis of 3, and as a result, the *excess kurtosis* of a

**Figure 9**  
*Skewness and Kurtosis*



*Note.* Circular distributions with different skewness (A) and kurtosis (B). In each panel, distributions have matching circular means and variances. Circular kurtosis and circular skewness are 0 for all distributions in (A) and (B), respectively. The black curves represent normal distributions wrapped around the circle (see the Wrapping Euclidean Distributions section). Contrary to the Euclidean equivalent, a positive circular skewness (green in Panel A) indicates that the long tail is in the negative (clockwise) direction from the mean, whereas a negative value (red in Panel A) indicates a longer positive (counterclockwise) tail. See the online article for the color version of this figure.

sample in Euclidean space is defined as  $g_2 = b_2 - 3$ . The distribution of a typical unimodal sample with positive excess kurtosis has a narrow peak around the mean and, extending away, heavier tails than the Gaussian distribution. Conversely, the distribution of a sample with negative excess kurtosis is flatter and more spread out around the mean. The sample *circular kurtosis* is defined by [Mardia \(1975\)](#) as

$$g_{20} := \frac{\bar{R}_2 \cos(\bar{\theta}_2 - 2\bar{\theta}) - \bar{R}^4}{(1 - \bar{R})^{\frac{3}{2}}} = \frac{1}{\hat{V}_0^2} \left( \frac{1}{n} \sum_{i=1}^n \cos 2(\theta_i - \bar{\theta}) - \bar{R}^4 \right). \quad (32)$$

It is the circular counterpart of Euclidean excess kurtosis. The reference point for zero circular kurtosis is the wrapping of the Gaussian distribution onto the unit circle (see the Wrapped Normal Distribution section), with positive values indicating heavier tails (green line in [Figure 9B](#)) and negative values lighter tails (red line in [Figure 9B](#)) than the wrapped normal, in agreement with the Euclidean excess kurtosis. For concentrated data, the circular kurtosis can be approximated by twice the Euclidean excess kurtosis,

$$g_{20} \approx 2g_2, \quad (33)$$

with the factor-of-two scaling again attributable to the fact that circular variance approaches half the Euclidean variance for concentrated data (see the Circular Variance section).

### Probability Theory and Models for Circular Variables

In this section, we introduce probabilistic foundations and modeling approaches for random variables in a circular space. The distribution of a circular random variable can be modeled by a probability density function (PDF) in a similar way to a Euclidean variable. The key difference is that a circular PDF must be periodic, such that  $f(\theta) = f(\theta + 2\pi k)$  for any integer  $k$ . While a PDF in  $\mathbb{R}^1$  has integral (“area under the curve”) equal to 1 over the whole real line  $(-\infty, \infty)$ , a valid PDF in circular space must integrate to 1 over any interval of length  $2\pi$ . Circular distributions can also be characterized by their cumulative distribution function

(CDF). Given a chosen reference angle  $\theta_0$ , the CDF is defined as  $F(\theta) = \int_{\theta_0}^{\theta} f(\phi) d\phi$ , so that  $F(\theta_0) = 0$  and  $F(\theta_0 + 2\pi) = 1$ . The CDF can be constructed over any interval of length  $2\pi$ ; thus, the choice of starting point  $\theta_0$  is an essential part of its specification.

In this section, we will outline population counterparts of the sample statistics discussed in the Descriptive Statistics of Circular Data section, address the conditions for sample-to-population convergence in circular space, and describe some common circular population distributions. We will compare two circular analogs of the Gaussian distribution and introduce two more flexible families for fitting heavy-tailed and skewed data. We will examine the properties of circular random variables in more detail before concluding with practical guidelines for choosing between circular distributions and sampling random values from circular distributions. Throughout this section, our focus will be on continuous distributions; discussions of discrete or “lattice” distributions on the circle can be found in [Jammalamadaka and SenGupta \(2001\)](#) and [Mardia and Jupp \(2009\)](#).

### Population Parameters

As in the Euclidean case, each descriptive statistic for circular data has a corresponding population parametric form that can be used to summarize an equivalent aspect of the theoretical distribution of a random variable (for sample and population parameter notations, see [Table 1](#)). The population parameters simply replace normalized sums over the  $n$  data points in a sample with *expectations* or integrals over the circular space. So the  $p$ th circular moment of a random variable with PDF  $f(\theta)$  is

$$m_p := \mathbb{E}_0[e^{ip\theta}] = \int_{\psi}^{\psi+2\pi} e^{ip\theta} f(\theta) d\theta, \quad (34)$$

where the integral can be taken over any interval of length  $2\pi$ . The arguments  $u_p = \arg(m_p)$  and moduli  $\rho_p = |m_p|$  provide definitions for the population circular mean  $\mu \equiv \mu_1$  and population mean resultant length  $\rho \equiv \rho_1$ . Population circular variance  $V_0$ ,

standard deviation  $\sigma_0$ , skewness  $\gamma_{10}$ , and kurtosis  $\gamma_{20}$  all have definitions that follow from those of the corresponding sample statistics (see the Descriptive Statistics of Circular Data section).

The circular median discussed in the Circular Median section is not moment-based. Like the sample version, the population circular median  $\mu_{\text{med}}$  can be defined in two ways: first, as an angle that divides the circle into two semicircles  $(\mu_{\text{med}} - \pi, \mu_{\text{med}})$  and  $(\mu_{\text{med}}, \mu_{\text{med}} + \pi)$  that each contain equal probability mass of  $\frac{1}{2}$ ; second, as an angle that minimizes the expected circular distance from the random variable,

$$\mu_{\text{med}} := \arg \min_{\phi} \mathbb{E}_{\theta} [D(\theta, \phi)] = \arg \min_{\phi} \int_{\psi}^{\psi+2\pi} D(\theta, \phi) f(\theta) d\theta. \quad (35)$$

As for the sample case, the first definition yields at least two possible solutions; therefore, the second one should be preferred. Note, however, that for certain distributions, the circular population median is not uniquely defined, even under the second definition (e.g., for symmetric, bi-, or multimodal distributions).

The properties and approximations for concentrated data described for sample statistics in the preceding section all generalize to their population counterparts.

### Convergence of Sample Statistics to Population Parameters

The trigonometric moments of an independent and identically distributed (i.i.d.) sample, including the mean resultant vector, converge to their population counterparts as the sample size increases (see the Central Limit Theorem section). As a result, sample statistics that are continuous functions of the mean resultant vector (circular mean, variance, and standard deviation) and of higher-order trigonometric moments (circular skewness and kurtosis) inherit convergence via the continuous mapping theorem. The only case where convergence of these parameters is not assured is when the true parameter is not uniquely defined. For example,

although we can calculate the circular mean of a sample drawn from the circular uniform distribution, the population mean resultant vector has length 0, so the mean direction is undefined, and the sample mean direction has no population counterpart to converge to.

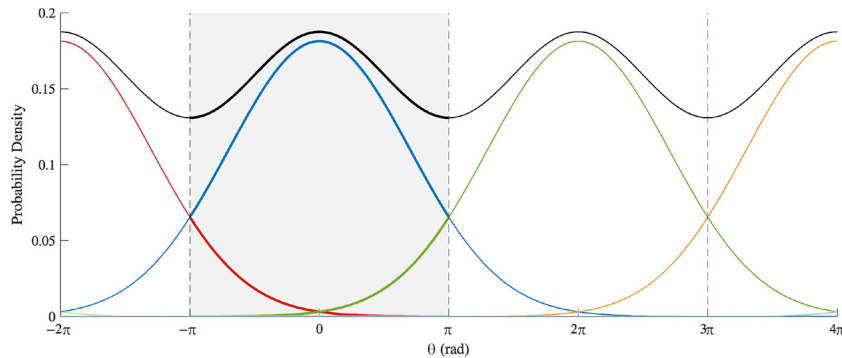
For parameters that are not moment-based, the question of convergence is more complicated. The circular median also converges when the true value is uniquely defined and the density around it is well-behaved, that is, continuous and sufficiently smooth (for a detailed discussion, see [Purkayastha, 1995](#)). Convergence results for local circular mode estimators are given in [Zhang and Chen \(2021\)](#).

### Wrapping Euclidean Distributions

One approach to modeling circular data is to “wrap” a Euclidean distribution around the circle to construct a *wrapped distribution*. For a Euclidean distribution with PDF  $f(x)$ , the result of wrapping is a circular PDF that can be written as the periodic sum  $f_w(\theta) = \sum_{k=-\infty}^{\infty} f(\theta + 2\pi k)$ , where the sum is taken over all integers  $k \in \mathbb{Z}$ . Conveniently, the circular moments of the wrapped distribution are equal to the characteristic function of the unwrapped distribution evaluated at integer arguments, for example, the resultant vector (first circular moment) of a wrapped distribution is given by  $m_1 = \varphi(1)$ , where  $\varphi(t) = \mathbb{E}[e^{itx}]$  is the characteristic function of the unwrapped distribution  $f(x)$ . The characteristic function also provides an alternative formulation of the PDF as a Fourier series, which is an infinite sum of sinusoidal components:  $f_w(\theta) = \frac{1}{2\pi} \sum_{t=-\infty}^{\infty} \varphi(t) e^{-it\theta}$ .

In practice, the infinite Fourier representation of a wrapped distribution cannot be evaluated exactly and must be approximated by truncating the series to a finite number of terms. This means evaluating the characteristic function at a limited set of integer points, which produces a smooth approximation of the true circular PDF. The resulting approximation is a finite Fourier series, and its accuracy improves as more terms are included. For an illustration of wrapping a Gaussian distribution, see [Figure 10](#), where the

**Figure 10**  
Wrapping of a Gaussian Distribution



*Note.* Colored lines show normal (Gaussian) distributions with standard deviation  $\sigma = 2.2$  and means centered on integer multiples of  $2\pi$ , that is,  $\mu = 2\pi k$ ,  $k \in \mathbb{Z}$ . The black line shows the wrapped normal distribution with mean 0 and circular standard deviation  $\sigma_\theta$  equal to  $\sigma$ , obtained by summing over all translations (for every integer  $k$  in the interval  $[-5, 5]$ , including those not shown) of the Gaussian distribution. The wrapped normal distribution is periodic with period  $2\pi$ . The signed interval  $[-\pi, \pi]$  is highlighted in gray, displaying one period of the wrapped normal distribution. Dashed gray lines indicate period boundaries. See the online article for the color version of this figure.

blue line shows a Gaussian distribution in Euclidean space and the black line the same distribution wrapped around the circle. While their familiarity is appealing, the distributions generated by wrapping are often “false friends,” proving difficult to work with analytically and lacking in desirable mathematical properties. More tractable “natively circular” distributions can be defined directly in circular space using trigonometric functions.

### Circular Uniform Distribution

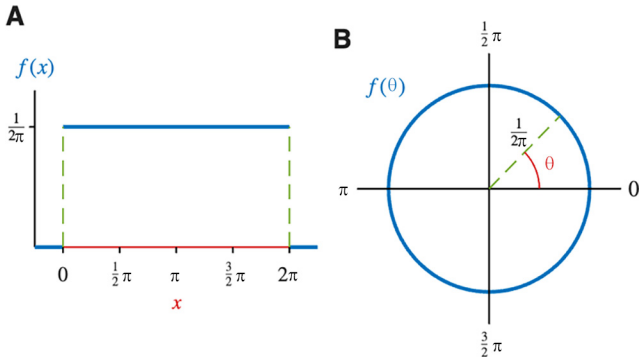
On the real line, a random variable can have a uniform distribution only with respect to a particular bounded interval, with PDF  $f(x) = \frac{1}{b-a}$  for  $x$  in the interval  $[a, b)$  and  $f(x) = 0$  outside it (Figure 11, left). While similar distributions can be defined that have uniform probability over a specified range of circular space, the *circular uniform distribution* describes a random variable that has equal probability everywhere, with PDF  $f_U(\theta) = \frac{1}{2\pi}$  for all  $\theta$  (Figure 11, right). The circular uniform distribution has no well-defined mean direction: Its mean resultant is the zero vector, its circular variance is 1, and its circular standard deviation is infinite. The circular uniform distribution has the maximum entropy of any circular distribution.

### Circular Analogs of the Normal Distribution

The normal or Gaussian distribution family is fundamental in Euclidean space, underpinning many statistical models due to its well-behaved mathematical properties, including closed-form expression, defined moments, and closure under translation, convolution, and scaling. For circular variables, no single family of distributions possesses all the desirable properties of the Gaussian, and some of its properties are not shared by any family of circular distributions. The closest circular equivalents to the Gaussian are the wrapped normal distribution and the von Mises distribution (sometimes referred to as the circular normal).

**Figure 11**

*Uniform Distribution in Euclidean and Circular Space*



*Note.* (A) Uniform distribution in Euclidean space  $f(x)$  over the interval  $[0, 2\pi)$ . Note that density is 0 outside this interval. (B) Uniform distribution in circular space  $f(\theta)$ . Equivalent components are displayed in the same color in (A) and (B). See the online article for the color version of this figure.

### Wrapped Normal Distribution

Wrapping the Gaussian distribution  $\mathcal{N}(\mu, \sigma)$  with mean  $\mu$  and standard deviation  $\sigma$

$$f_N(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) \quad (36)$$

onto a unit circle results in the wrapped normal distribution  $WN(\mu, \sigma)$  with PDF given by

$$f_{WN}(\theta; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \sum_{k=-\infty}^{\infty} \exp\left(-\frac{(\theta - \mu + 2\pi k)^2}{2\sigma^2}\right). \quad (37)$$

The circular mean of the wrapped normal distribution is  $\mu$ , and its circular standard deviation is  $\sigma$  (as discussed in the Circular Standard Deviation section, this motivates the definition of circular standard deviation). Its circular variance is not  $\sigma^2$ , but  $1 - \bar{R} = 1 - \exp\left(-\frac{\sigma^2}{2}\right)$ . Its circular kurtosis and skewness are 0.

For small  $\sigma$ , the probability density is concentrated enough that the circularity of the space becomes irrelevant, and the wrapped normal PDF approximates the Gaussian PDF with mean  $\mu$  and standard deviation  $\sigma$ . As  $\sigma \rightarrow \infty$ , the wrapped normal distribution approaches the circular uniform distribution.

One way in which the wrapped normal distribution can arise is as the result of Brownian diffusion (i.e., accumulation of Gaussian noise) in an angular variable from starting point  $\mu$ , where the distribution of the variable at time  $t$  is wrapped normal with circular standard deviation proportional to  $\sqrt{t}$ . For a visualization of the wrapped normal distribution and its construction, see Figure 10.

### Von Mises Distribution

In practice, the von Mises distribution is often preferred as the circular analog of the normal distribution on the real line. The von Mises distribution  $VM(\mu, \kappa)$  has PDF

$$f_{VM}(\theta; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp(\kappa \cos(\theta - \mu)), \quad (38)$$

where  $\mu$  is the circular mean and  $\kappa$  is called the *concentration* of the distribution. Comparing the von Mises PDF with the Gaussian PDF (Equation 36) reveals a very similar functional form: The concentration parameter  $\kappa$  of the von Mises plays an equivalent role to the inverse variance, or precision, of the Gaussian distribution  $\frac{1}{\sigma^2}$ , and a trigonometric measure of similarity to the mean,  $\cos(\theta - \mu)$ , replaces the term based on squared error  $-\frac{1}{2}(x - \mu)^2$  (see the Measuring Distance section).

The function  $I_0(\cdot)$  refers to the modified Bessel function of the first kind with order 0. Bessel functions are important mathematically as solutions to certain differential equations, but here  $2\pi I_0(\kappa)$  is simply a normalization term that appropriately scales the PDF, so it integrates to 1 over intervals of length  $2\pi$ .

A von Mises distribution with concentration  $\kappa$  has circular variance  $1 - \rho = 1 - A_1(\kappa)$ , with  $A_1(\kappa) = \frac{I_1(\kappa)}{I_0(\kappa)}$ , that is, the ratio of the first- to zeroth-order modified Bessel function of the first kind. Its circular skewness is 0, but its circular kurtosis is greater than 0

for  $\kappa > 0$ , with a peak value of  $\sim .834$  for  $\kappa \approx 2.71$ . As  $\kappa \rightarrow \infty$ , the von Mises PDF approaches the Gaussian PDF with mean  $\mu$  and precision  $\frac{1}{\sigma^2}$  equal to  $\kappa$ . For  $\kappa = 0$ , the von Mises distribution is identical to the circular uniform distribution.

The von Mises is a member of the exponential family of distributions, making it particularly suitable for statistical inference. The maximum likelihood estimator (MLE) of the location parameter  $\mu$  is the circular mean. In fact, Richard von Mises originally derived the distribution named after him as that distribution for which the sample circular mean is the MLE of location, in the same way as the sample mean is the MLE for the location parameter of the Gaussian (von Mises, 1918).

The von Mises distribution can also be derived from a bivariate Gaussian distribution in the real plane  $\mathbb{R}^2$ . The von Mises PDF is equal to the conditional density of polar angle  $\theta$  given unit radial distance  $p(\theta | r = 1)$  for a bivariate normal random variable with mean on the unit circle, at  $[\cos(\mu), \sin(\mu)]$ , equal variances  $\frac{1}{\kappa}$  and zero correlation. This approach can be extended to define analogs of the von Mises in higher dimensions, known as von Mises–Fisher distributions: The von Mises–Fisher PDF in  $S^p$  is proportional to the restriction of an isotropic multivariate Gaussian distribution in  $\mathbb{R}^{p+1}$  to the unit  $p$ -dimensional hypersphere.

Yet another way to obtain the von Mises distribution is as the hitting point on the unit circle of a two-dimensional Brownian diffusion (i.e., an accumulation of isotropic Gaussian noise) with constant drift in direction  $\mu$  with velocity  $\kappa$  (Gordon & Hudson, 1977). Extending this principle to higher dimensions again leads to the von Mises–Fisher distributions.

### Comparison of Wrapped Normal and von Mises Families

Figure 12 gives a comparison of the Gaussian, wrapped normal, and von Mises distributions with mean 0 and matched (circular) standard deviation (Gumbel et al., 1953). Key properties of these distributions are summarized in Table 2. Unlike the wrapped normal PDF, which involves an infinite sum of Gaussian densities, the PDF of the von Mises distribution has an analytic expression, as do some key quantities such as Fisher information and entropy (most of these involve Bessel functions, but most mathematical

software packages have built-in functions for their efficient numerical approximation).

A family of distributions is closed under convolution if the sum of two independent random variables from that family also belongs to the same family (the PDF of the sum is obtained by applying the mathematical operation of convolution to the PDFs of the two variables). The wrapped normal inherits closure under convolution from its Euclidean counterpart, the Gaussian (see the Sampling Distribution of the Circular Mean section). However, a number of other properties that make the Gaussian distribution special are instead possessed by the von Mises distribution in circular space. Like the Gaussian, the von Mises is a *maximum entropy distribution*. Specifically, among all distributions on the circle with a given circular mean and variance (or equivalently, a given mean resultant vector), the von Mises distribution has the maximum entropy. More similarities and differences between these distributions are described in the General Properties of Circular Random Variables section.

### Flexible Distribution Families

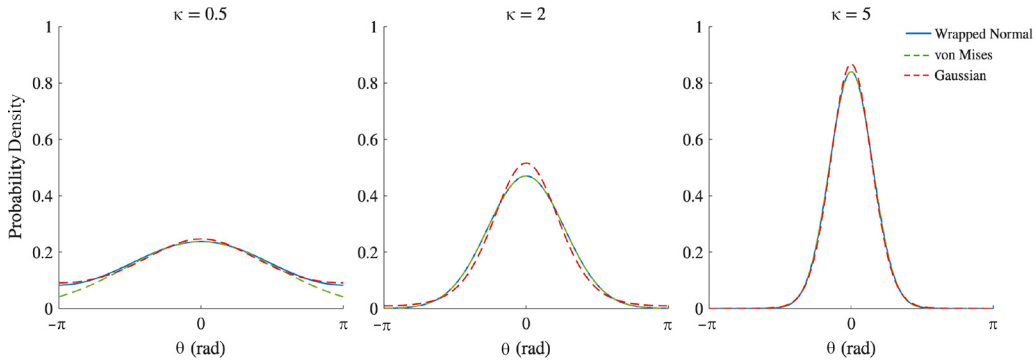
We next introduce two families of distributions that can serve as flexible models for circular data that are asymmetric or heavy-tailed. Both are four-parameter families of distributions varying in tailedness and skewness as well as location and scale. The *wrapped stable family* generalizes the wrapped normal distribution and retains the property of closure under convolution, meaning that sums of random variables belong to the same family. The sine-skewed Jones–Pewsey family generalizes the von Mises distribution, and its PDF has an analytic expression based on trigonometric functions. Particular attention will be given to the wrapped Cauchy distribution, which uniquely belongs to both families.

### Wrapped Stable Family

A distribution is called *stable* if linear combinations of random variables have the same distribution up to location and scale. In Euclidean space, stable distributions are important as they generalize the concept of a central limit distribution (see the Central Limit Theorem section) to include distributions with nonfinite variance.

**Figure 12**

*Comparison of Gaussian, Wrapped Normal, and von Mises Distributions*



*Note.* Gaussian, wrapped normal, and von Mises distributions with concentration parameters  $\kappa = 0.5, 2.0, 5.0$ . Distributions are matched for (circular) standard deviation, using  $\sigma = \sqrt{-2 \ln A_1(\kappa)}$  to obtain  $\sigma$  for the wrapped normal and Gaussian distributions. The wrapped normal PDF is approximated using the sum of the first 10 terms of its Fourier series. PDF = probability density function. See the online article for the color version of this figure.

**Table 2***Properties of the Gaussian, Wrapped Normal, and von Mises Distribution Families*

| Property                                  | Gaussian                                        | Wrapped normal                                   | von Mises                                                                     |
|-------------------------------------------|-------------------------------------------------|--------------------------------------------------|-------------------------------------------------------------------------------|
| Support                                   | $x \in \mathbb{R}^1$                            | $\theta \in \mathbb{S}^1$                        | $\theta \in \mathbb{S}^1$                                                     |
| Parameters                                | $\mu, \sigma^2$                                 | $\mu, \sigma$                                    | $\mu, \kappa$                                                                 |
| Symmetrical                               | ✓                                               | ✓                                                | ✓                                                                             |
| Unimodal                                  | ✓                                               | ✓                                                | ✓                                                                             |
| (Circular) mean/median/mode               | $\mu$                                           | $\mu$                                            | $\mu$                                                                         |
| Circular moments, $m_n$                   | (Not applicable)                                | $e^{in\mu - n^2\sigma^2/2}$                      | $e^{in\mu} A_n(\kappa)^a$                                                     |
| Mean resultant length, $\rho$             | (Not applicable)                                | $e^{-\sigma^2/2}$                                | $A_1(\kappa)^a$                                                               |
| (Circular) variance                       | $\sigma^2$                                      | $1 - e^{-\sigma^2/2}$                            | $1 - A_1(\kappa)^a$                                                           |
| (Circular) standard deviation             | $\sigma$                                        | $\sigma$                                         | $\sqrt{-2 \ln A_1(\kappa)^a}$                                                 |
| (Circular) kurtosis                       | 3                                               | 0                                                | $\frac{A_2(\kappa) - A_1(\kappa)^4}{(1 - A_1(\kappa)^2)^2}$                   |
| Closed under translation <sup>b</sup>     | ✓                                               | ✓                                                | ✓                                                                             |
| Closed under scaling <sup>b</sup>         | ✓                                               | ✗                                                | ✗                                                                             |
| Closed under convolution <sup>b</sup>     | ✓                                               | ✓                                                | ✗                                                                             |
| Fisher information for $\mu$              | $\frac{1}{\sigma^2}$                            | $\sum_{k=-\infty}^{\infty} k^2 e^{-k^2\sigma^2}$ | $\kappa A_1(\kappa)^a$                                                        |
| Entropy (differential)                    | $\frac{1}{2} \ln(2\pi e \sigma^2)$              | (Not analytic)                                   | $\ln(2\pi I_0(\kappa)) - \kappa A_1(\kappa)^a$                                |
| Maximum entropy distribution <sup>c</sup> | ✓                                               | ✗                                                | ✓                                                                             |
| Exponential family                        | ✓                                               | ✗                                                | ✓                                                                             |
| Sufficient statistics                     | $(\bar{x}, s^2)$ for $(\mu, \sigma^2)$          | (Nonfinite)                                      | $(\bar{\theta}, \bar{R})$ for $(\mu, \kappa)$                                 |
| Maximum likelihood estimates              | $\bar{x}$ (for $\mu$ ), $s^2$ (for $\sigma^2$ ) | (Not analytic)                                   | $\bar{\theta}$ (for $\mu$ ), $A_1^{-1}(\bar{R})$ (for $\kappa$ ) <sup>a</sup> |
| Log likelihood                            | Quadratic                                       | (Not analytic)                                   | Cosine                                                                        |
| Product of PDFs                           | Scaled Gaussian                                 | (Not analytic)                                   | Scaled von Mises                                                              |

*Note.* Green check marks (✓) indicate that the distribution has the given property; red crosses (✗) indicate that it does not. PDF = probability density function. See the online article for the color version of this table.

<sup>a</sup>  $A_n(\kappa)$  is shorthand for the ratio of Bessel functions  $\frac{I_n(\kappa)}{I_0(\kappa)}$ . <sup>b</sup> “Closed under X” means that applying operation X to any member of the distribution family produces another member of the family. <sup>c</sup> The Gaussian and von Mises distributions have the maximum entropy for a specified mean and (circular) variance.

In a circular space, all continuous distributions have finite variance, but wrapping the stable family provides a flexible model for skewed and heavy-tailed data that is also closed under convolution. The wrapped stable PDF can be constructed from its moments, which correspond to the characteristic function of a general stable distribution on the real line,

$$m_k = \varphi(k) = \begin{cases} \exp\left(-(k\gamma)^\alpha + ik\delta + i[(k\gamma)^\alpha - k\gamma]\beta \tan\left(\frac{\pi}{2}\alpha\right)\right) & \text{for } \alpha \neq 1, \\ \exp\left(-(k\gamma)^\alpha + ik\delta - ik\gamma \log(k\gamma)\beta \frac{2}{\pi}\right) & \text{for } \alpha = 1, \end{cases} \quad (39)$$

$$f_{\text{WS}}(\theta; \alpha, \beta, \gamma, \delta) = \frac{1}{2\pi} \left(1 + 2 \sum_{k=1}^{\infty} |m_k| \cos(k\theta - \arg m_k)\right). \quad (40)$$

The preceding expressions are referred to as the  $S_0$  parametrization by [Pewsey \(2008\)](#). Note that, despite the piecewise definition, necessary because  $\tan(\frac{\pi}{2})$  is undefined, the wrapped stable PDF and moments change smoothly through  $\alpha = 1$ . The mean resultant length is given by  $\rho = \exp(-\gamma^\alpha)$ .

The stability parameter  $\alpha \in (0, 2]$  controls heaviness of the distribution tails, skewness parameter  $\beta \in [-1, 1]$  controls the asymmetry,  $\gamma$  is a scale (or dispersion) parameter, and  $\delta$  is a location parameter. For  $\alpha = 2$ , the wrapped stable is equivalent to the wrapped normal with  $\mu = \delta$  and  $\sigma = \sqrt{2\gamma}$  ([Figure 13A](#)). As illustrated in [Figures 13B–13E](#), larger  $|\beta|$  increases skewness (for  $\alpha < 2$ ), smaller  $\alpha$  yields heavier tails, and higher  $\gamma$  leads to greater dispersion.

[Figure 13A](#) shows some canonical examples of wrapped stable distributions. The wrapped normal ( $\alpha = 2, \beta = 0$ ), the wrapped Holtsmark ( $\alpha = 1.5, \beta = 0$ ), and the wrapped Cauchy ( $\alpha = 1, \beta = 0$ ) are symmetric cases. The standard wrapped Levy distribution ( $\alpha = 0.5, \beta = 1$ ) is right-skewed and exhibits extremely heavy tails, making it suitable for modeling pulse-like behavior on the circle.

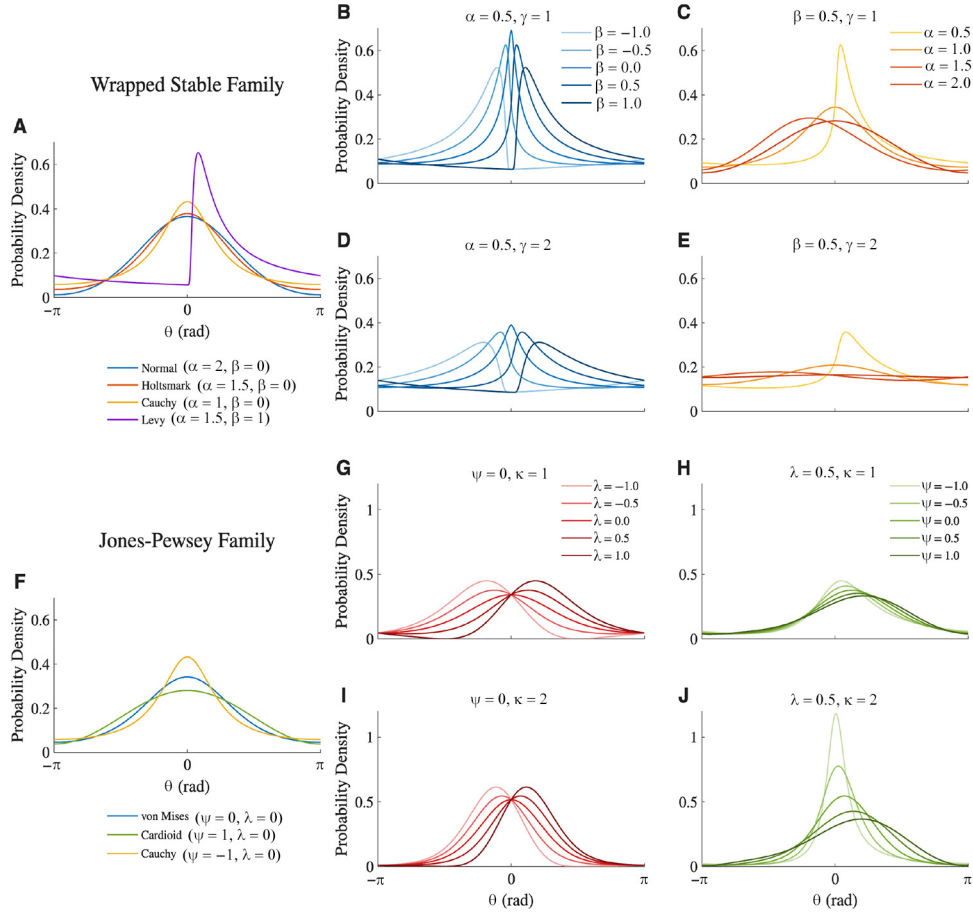
### Sine-Skewed Jones–Pewsey Family

The Jones–Pewsey family is a generalization of the von Mises distribution on the circle ([Jones & Pewsey, 2005](#)), which allows for flexible tailedness and, in its sine-skewed form ([Jones & Pewsey, 2012](#)), flexible skewness as well. The sine-skewed Jones–Pewsey PDF is given by

$$f_{\text{SS-JP}}(\theta; \delta, \kappa, \psi, \lambda) = \begin{cases} \frac{\cosh^{1/\psi}(\kappa\psi)(1 + \tanh(\kappa\psi)\cos(\theta - \delta))^{1/\psi}(1 + \lambda\sin(\theta - \delta))}{2\pi P_{1/\psi}(\cosh(\kappa\psi))} & \text{for } \psi \neq 0, \\ \frac{\exp(\kappa\cos(\theta - \delta))(1 + \lambda\sin(\theta - \delta))}{2\pi I_0(\kappa)} & \text{for } \psi = 0, \end{cases} \quad (41)$$

where  $P_v(z)$  is the associated Legendre function of the first kind of degree  $v$  and order 1, which can also be expressed in terms of the hypergeometric function  ${}_2F_1(-v, v+1; 1; \frac{1-z}{2})$ . Despite the piecewise definition, needed to avoid division by 0, the distribution changes smoothly through  $\psi = 0$ , which

**Figure 13**  
*Examples of Flexible Distribution Families*



*Note.* Example distributions from the wrapped stable family (A–E) and the sine-skewed Jones–Pewsey family (F–J). (A) and (F) show special cases of the wrapped stable and Jones–Pewsey family, respectively. The Cauchy distributions in (A) and (F) are matched in mean resultant length, using  $\gamma = -\ln(\tanh(\frac{\kappa}{2}))$  to convert the concentration parameter of the Jones–Pewsey family,  $\kappa$ , to the scale parameter of the Wrapped stable family,  $\gamma$ . Shapes of the wrapped stable family with changing  $\beta$  (B, D) and  $\alpha$  (C, E). The PDFs of the wrapped stable family are approximated using the first 100 terms of the Fourier series. Shapes of the Jones–Pewsey family with changing  $\lambda$  (G, I) and  $\psi$  (H, J). Location parameters  $\delta$  are 0 in all cases. PDF = probability density function. See the online article for the color version of this figure.

corresponds in this case to the sine-skewed von Mises distribution.

The index  $\psi \in (-\infty, \infty)$  controls the tailedness of the distribution,  $\lambda \in [-1, 1]$  controls skewness,  $\kappa$  is an inverse scale (precision) parameter, and  $\mu$  is a location parameter.

When  $\psi < 0$ , the distribution has heavier tails than the von Mises distribution, whereas  $\psi > 0$  leads to lighter tails (compare the wrapped stable family, which contains only heavier tailed distributions; Figures 13H and 13J).  $\lambda > 0$  leads to right-skew, and conversely for  $\lambda < 0$ ; larger values of  $|\lambda|$  lead to stronger skewness (Figures 13G and 13I).

Figure 13F shows some special cases of the symmetric Jones–Pewsey family ( $\lambda = 0$ ), including the von Mises distribution (with  $\psi = 0$ ; mean resultant length is  $\rho = A_1(\kappa)$ ), the Cardioid

distribution ( $\psi = 1$ ;  $\rho = \tanh(\kappa)$ ), and the wrapped Cauchy distribution ( $\psi = -1$ ;  $\rho = \tanh(\frac{\kappa}{2})$ ).

In general, the analytic form of the Jones–Pewsey family PDF makes it particularly convenient for modeling and statistical inference. As for the von Mises distribution, the normalization term of the PDF involves a nonelementary function, the associated Legendre, but most mathematical software packages provide efficient methods for its computation as a special case of the hypergeometric function. Although analytically convenient, direct numerical evaluation via the hypergeometric function becomes unreliable when  $\psi$  is large and negative. Two practical ways to ensure numerical stability in this regime are computing the Legendre term using more robust algorithms (Gutleb et al., 2022) or bypassing the Legendre representation with a numerical integration approach (see Appendix C).

### Wrapped Cauchy Distribution

The wrapped Cauchy distribution belongs to both the wrapped stable family (with  $\alpha = 1$ ,  $\beta = 0$ ; Figure 13A) and the symmetrical Jones–Pewsey family (with  $\psi = -1$ ,  $\lambda = 0$ ; Figure 13F). It is closed under convolution like the rest of the wrapped stable family, but its PDF also has a closed-form trigonometric expression as part of the Jones–Pewsey family. Specifically, its PDF can be obtained directly as the wrapping of the Cauchy distribution onto the circle,

$$f_{WC}(\theta; \mu, \gamma) = \sum_{k=-\infty}^{\infty} \frac{1}{\pi} \frac{\gamma}{(\gamma^2 + (\theta - \mu + 2\pi k)^2)}, \quad (42)$$

or equivalently in trigonometric terms as

$$f_{WC}(\theta; \mu, \gamma) = \frac{1}{2\pi} \frac{\sinh(\gamma)}{\cosh(\gamma) - \cos(\theta - \mu)}. \quad (43)$$

It is parameterized by a location  $\mu$  (the mean and center of symmetry) and a scale  $\gamma$ . The mean resultant length of the wrapped Cauchy distribution is  $\rho = e^{-\gamma}$ , and its circular standard deviation is  $\sqrt{2\gamma}$ . Its circular moments are  $m_n = e^{in\mu - |n|\gamma}$ . Its differential entropy can be written in closed form as  $\log(2\pi(1 - e^{-2\gamma}))$ . Note that although the Cauchy distribution on the real line is the classic exemplar of a “pathological” distribution, with mean, variance, and higher moments undefined, its circular counterpart retains some of the heavy-tailedness yet has finite circular moments, making it more tractable for modeling circular data with heavy tails.

### General Properties of Circular Random Variables

#### Additivity of Mean and Variance

Models of Euclidean random variables often make use of properties of the mean and variance, such as additivity, that do not apply or apply differently in circular spaces. The Euclidean mean or expectation is additive in the sense that the mean of the sum of any two random variables is equal to the sum of their means,

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y], \quad (44)$$

and Euclidean variance is additive in the sense that the variance of the sum of two random variables is equal to the sum of their variances, *so long as the variables are uncorrelated*,

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] \quad \text{iff } X \perp\!\!\!\perp Y. \quad (45)$$

These properties are key to understanding how variability from different sources combines and how random noise accumulates (e.g., in drift-diffusion models; Bogacz et al., 2006; Ratcliff, 1978). However, caution is required when extrapolating these properties to circular variables (see Smith, 2016, for a diffusion model of circular data).

Circular equivalents to mean and variance are related to the direction and length of the mean resultant vector, which, for the angular sum of two independent random variables, is the *product* of their individual mean resultant vectors. As a result, the circular

mean—although not additive in general—does have the additive property for variables that are uncorrelated,

$$\mu[\theta + \phi] = \mu[\theta] + \mu[\phi] \pmod{2\pi} \quad \text{if } \theta \perp\!\!\!\perp \phi. \quad (46)$$

Circular measures of dispersion are related to the mean resultant length,  $\rho$ , for which

$$\rho[\theta + \phi] = \rho[\theta]\rho[\phi] \quad \text{if } \theta \perp\!\!\!\perp \phi, \quad (47)$$

so  $\rho$  does not have the additive property, and neither does the circular variance, based on its usual definition  $V_\theta = 1 - \rho$ . However, the *logarithm* of the mean resultant length,  $\log \rho$ , is additive for independent circular random variables, and (despite not being a usual circular measure of variance) so is the squared circular standard deviation,

$$\sigma^2[\theta + \phi] = \sigma^2[\theta] + \sigma^2[\phi] \quad \text{if } \theta \perp\!\!\!\perp \phi. \quad (48)$$

A more detailed exposition with derivations can be found in Appendix D.

### Products of Probability Density Functions

Products of PDFs are important in contexts where information from different sources is combined. These include maximum likelihood estimation and Bayesian inference based on independent samples drawn from the same distribution, as well as sensory fusion and cue combination (Trommershauser et al., 2011), where different probabilistic estimates of a parameter are integrated. These computations take a particularly simple form for Gaussian variables, because the product of Gaussian PDFs is proportional to another Gaussian PDF, with precision  $\frac{1}{\sigma^2}$  equal to the sum of the component precisions, and mean equal to the precision-weighted average of the component means.

For circular variables, the von Mises distribution possesses a similar property (Murray & Morgenstern, 2010): The product of von Mises PDFs is proportional to another von Mises PDF with mean and concentration determined by the concentration-weighted resultant vector of the component mean directions (Figure 14A):

$$\prod_{i=1}^n f_{VM}(\theta; \mu_i, \kappa_i) = S f_{VM}(\theta; \arg \bar{r}, |\bar{r}|) \quad (49)$$

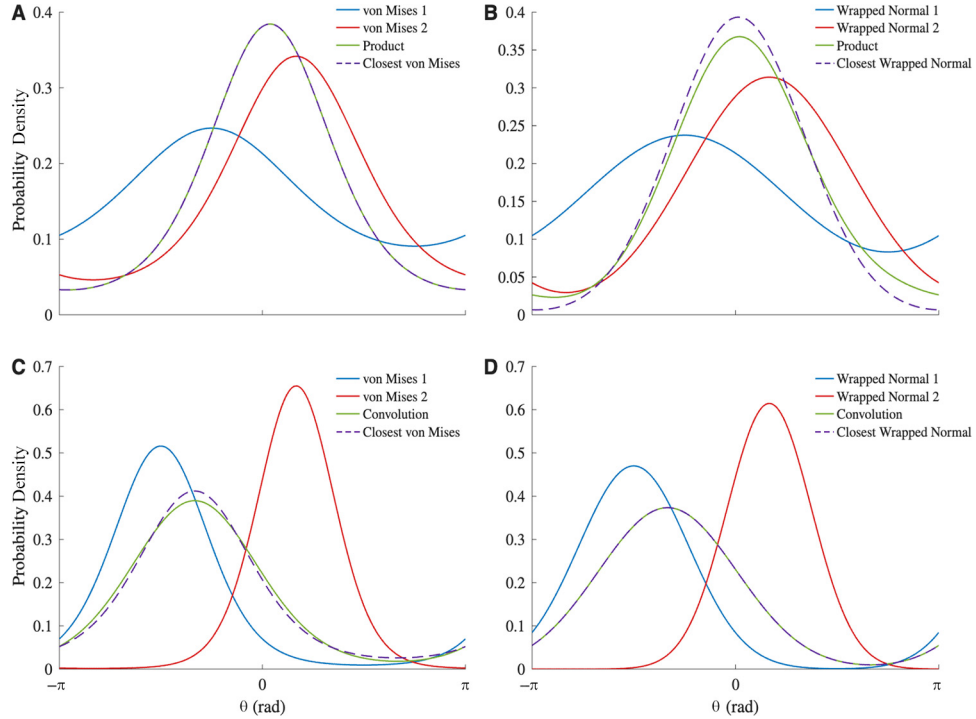
$$\bar{r} = \sum_{i=1}^n \kappa_i e^{i\mu_i}, \quad S = \frac{1}{(2\pi)^{n-1}} \frac{I_0(|\bar{r}|)}{\prod_i I_0(\kappa_i)}.$$

By contrast, the product of two wrapped normal distributions does not yield another wrapped normal distribution, although an approximation based on the Gaussian product formula holds well for small standard deviations, for example,  $\sigma < 1$  (Figure 14B).

### Sums and Differences of Random Variables

The probability distribution of the sum of two independent random variables  $Z = X + Y$  is obtained by convolving their individual PDFs; the distribution of their difference is obtained by reflecting one of the PDFs about 0, then convolving. These operations are the same for circular variables, where the

**Figure 14**  
*Products and Convolutions of von Mises and Wrapped Normal PDFs*



*Note.* (A, B) Products of PDFs of von Mises distributions (A) and wrapped normal distributions (B). Each panel shows two PDFs (red and blue), their normalized product (green), and the closest match to the normalized product by a PDF from the same family (purple, dashed). Note that the product of two von Mises PDFs (A) is proportional to another von Mises PDF, whereas the product of two wrapped normal PDFs (B) does not in general correspond to any wrapped normal PDF. Examples shown are  $VM_1(\mu_1 = -\frac{\pi}{4}, \kappa_1 = 0.5)$  and  $VM_2(\mu_2 = \frac{\pi}{6}, \kappa_2 = 1)$  and wrapped normal PDFs with matching resultant vectors. (C, D) Convolutions of PDFs of von Mises distributions (C) and wrapped normal distributions (D). PDFs shown in green are convolutions of red and blue PDFs, and purple is the closest PDF from the same family. Note that the convolution of the two von Mises distributions (C) is not a von Mises distribution, whereas the convolution of two wrapped normal distributions is another wrapped normal distribution (D). Examples shown are  $VM_1(\mu_1 = -\frac{\pi}{2}, \kappa_1 = 2)$  and  $VM_2(\mu_2 = \frac{\pi}{6}, \kappa_2 = 3)$  and wrapped normal PDFs with matched resultant vectors. Wrapped normal PDFs are approximated using the sum of the first 10 terms of their Fourier series. PDF = probability density function. See the online article for the color version of this figure.

convolution is of their periodic PDFs over  $(-\infty, \infty)$ . For Euclidean variables drawn from Gaussian distributions, the sum is always distributed as another Gaussian: We say the Gaussian family is *closed under convolution*. Because convolving two wrapped distributions is identical to wrapping the convolution of the corresponding unwrapped distributions, the wrapped normal distribution inherits this property directly from the normal distribution (Figure 14D):

$$f_{WN}(\theta; \mu_x, \sigma_x) * f_{WN}(\theta; \mu_y, \sigma_y) = f_{WN}\left(\theta; \mu_x + \mu_y, \sqrt{\sigma_x^2 + \sigma_y^2}\right). \quad (50)$$

The von Mises distribution is not closed under convolution, which means that the sum (or difference) of von Mises distributed variables is not in general a member of any recognized distribution family (Figure 14C). The PDF for the sum can be calculated as

$$f_{VM}(\theta; \mu_x, \kappa_x) * f_{VM}(\theta; \mu_y, \kappa_y) = \frac{I_0\left(\sqrt{\kappa_x^2 + \kappa_y^2 + 2\kappa_x\kappa_y \cos(\theta - (\mu_x + \mu_y))}\right)}{2\pi I_0(\kappa_x) I_0(\kappa_y)}. \quad (51)$$

### Sampling Distribution of the Circular Mean

Because the Gaussian family is closed under both convolution and scaling, the mean of  $n$  i.i.d. Gaussian random variables  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  are distributed as another Gaussian (with the same mean and variance reduced by a factor of  $n$ ). This important property is not shared by any family of circular distributions.

The circular equivalent of the mean is the angle of the resultant vector, the vector sum of the unit vectors corresponding to individual random variables. In general, the distribution of the resultant vector angle cannot be obtained independently of its length. The only exception is the circular uniform distribution (Kent

et al., 1979), for which the distribution of the resultant angle  $\bar{\theta}$  of i.i.d. samples is, trivially, also circular uniform. For two samples from the circular uniform distribution, the distribution of resultant length  $R$  is

$$h_2(R) = \frac{2}{\pi\sqrt{4-R^2}} \text{ for } R \in [0, 2). \quad (52)$$

For three or more samples, the distribution  $h_n$  of resultant length quickly becomes intractable and can only be approximated with numerical methods. An approximation for large  $n$  is given in terms of the chi-square PDF by

$$h_n(R) \approx \frac{4}{n} R f_{\chi^2} \left( \frac{2}{n} R^2; 2 \right). \quad (53)$$

For the von Mises distribution, the circular mean of i.i.d. samples is not von Mises distributed, but it can be described as a mixture of von Mises distributions with different concentrations. Specifically, for  $n$  samples from  $\text{VM}(\mu, \kappa)$ , the conditional distribution of resultant angle  $\bar{\theta}$  for a given resultant length  $R$  is von Mises, with concentration proportional to  $R$ ,

$$p(\bar{\theta} | R) = f_{\text{VM}}(\bar{\theta}; \mu, \kappa R). \quad (54)$$

Informally, when the sample directions are similar (large  $R$ ), their mean  $\bar{\theta}$  tends to be close to the distribution mean,  $\mu$ , whereas when the samples are widely dispersed, the sample mean is too. This contrasts with the Gaussian distribution, for which the sample mean is independent of the sample standard deviation.

The distribution of resultant length is related to the equivalent distribution for the circular uniform by

$$p(R) = \frac{I_0(\kappa R)}{I_0(\kappa)^n} h_n(R). \quad (55)$$

The distribution of the circular mean (the marginal distribution of resultant angle) can then be obtained as the mixture density

$$p(\bar{\theta}) = \int_0^n p(\bar{\theta} | R) p(R) dR. \quad (56)$$

For the wrapped normal, the circular mean of i.i.d. samples does not have any recognized distribution, even when conditioned on resultant length.

### Central Limit Theorem

The central limit theorem states that, under quite general conditions, the sum or mean of a sample of Euclidean random variables converges to a normal (Gaussian) distribution as the number of observations in the sample grows, even if the individual observations do not come from a normal distribution. This fact is often called upon to justify the use of the normal distribution family as a default model for a wide range of behavioral and neural observations and also underlies many parametric statistical tests. Analogous results to the central limit theorem for circular variables are therefore of considerable interest; however, the topic is complex.

In circular spaces, the limiting distribution depends on precisely what operation is performed on the random variables. It should be plain that the angular sum of increasingly many independent circular random variables, regardless of their individual distributions,

will in general grow in dispersion and ultimately converge to the circular uniform distribution. Of more interest is the limiting distribution of the mean resultant vector, the direction of which is the closest circular analog of the Euclidean mean. For i.i.d. circular variables, the mean resultant vector converges to a bivariate Gaussian centered on the mean resultant vector of the distribution they are sampled from, and with covariance that can be calculated from its first two circular moments (Jammalamadaka & SenGupta, 2001, Section 3.4). In particular, if that distribution is symmetrical, the bivariate Gaussian is uncorrelated, so the *tangent* of the resultant direction converges to a Gaussian distribution. If the distribution of resultant direction is narrow enough (which it will be for sufficiently large numbers of samples, as long as they are not uniformly distributed), then it will have approximately the same distribution as its tangent.

Based on this, we can deduce that the limiting distribution of the circular mean of symmetrically distributed circular random variables is approximately Gaussian in shape. Unfortunately, this is not sufficient to identify a specific circular distribution. In particular, both wrapped normal and von Mises PDFs converge to Gaussian as their width approaches 0, so we cannot identify either of these circular analogs of the Gaussian as the unique limiting distribution.

### An Equivalent of the Cramér–Rao Bound

The Cramér–Rao Bound (CRB; Rao, 1945) is an important result in estimation theory; it describes a limit on the precision of a point estimator  $\hat{\theta}(X)$  of an unknown parameter  $\theta \in \mathbb{R}^1$  based on data  $X$  with likelihood function  $p(x | \theta)$ . For an unbiased estimator, that is, one with expectation equal to the true parameter,  $E[\hat{\theta}] = \theta$ , the CRB can be expressed as

$$\text{Var}[\hat{\theta}] \geq \frac{1}{\mathcal{I}(\theta)}, \quad (57)$$

where the Fisher information  $\mathcal{I}(\theta)$  is defined as  $\mathcal{I}(\theta) = \text{Var} \left[ \frac{\partial}{\partial \theta} \log p(x | \theta) \right] = -\mathbb{E} \left[ \frac{\partial^2}{\partial \theta^2} \log p(x | \theta) \right]$ . An unbiased estimator that achieves this bound is said to be efficient (an example is the MLE of the mean based on i.i.d. samples drawn from a normal distribution).

An equivalent of the CRB for an estimator  $\hat{\theta}(X)$  of a circular parameter  $\theta \in \mathbb{S}^1$  that is unbiased in the sense of having a circular mean equal to  $\theta$  (or equivalently,  $\mathbb{E}[\sin(\hat{\theta} - \theta)] = 0$ ) is given by

$$\frac{\text{Var}[\sin(\hat{\theta} - \theta)]}{\rho_0^2} \geq \frac{1}{\mathcal{I}(\theta)}, \quad (58)$$

where  $\rho_0$  is the mean resultant length of  $\hat{\theta}$  (Mardia, 1975). Note that the left-hand side of this equation differs from the definition of circular variance,  $1 - \rho_0$ . However, it can equivalently be written in terms of the first and second circular moments of  $\hat{\theta}$  as

$$\delta := \frac{1 - |m_2|}{2|m_1|^2}, \quad (59)$$

which is an alternative measure of variability for circular variables, defined as the circular dispersion by Mardia and Jupp (2009, Section 3.4.2). The bound is met in the circular case if and only if

$\frac{\partial}{\partial \theta} \log p(x | \theta) \propto \sin(\hat{\theta} - \theta)$ , that is, if the likelihood is von Mises. An example is the MLE of the circular mean based on i.i.d. samples from a von Mises distribution with known concentration  $\kappa > 0$ .

### Choosing an Appropriate Circular Distribution

After being introduced to a range of established circular distribution families and their properties, the reader should now be able to choose suitable distributions when building statistical models of circular data. If the data come from a source with no directional bias, the circular uniform distribution should be used. If not, two criteria can be employed for selecting one of the distribution families discussed previously: first, whether the data can be expected to have an approximately normal distribution; second, whether the purposes of analysis better suit the properties of wrapped distributions or those defined intrinsically on the circle. If a normal distribution can be assumed (based on inspection or because the conditions described in the Central Limit Theorem section are met), then the von Mises or the wrapped normal is appropriate; if not, and assuming the data are nonetheless approximately unimodal, the Jones–Pewsey or wrapped stable should be chosen for their ability to capture asymmetry and tailedness of data. Wrapped distribution families are more appropriate for data obtained by wrapping Euclidean random variables or when the analysis involves addition or subtraction of circular variables (because these families are closed under convolution). Distributions that are defined intrinsically on the circle are appropriate if the reader needs an analytically and computationally tractable model for statistical inference, such as MLE or maximum a posteriori estimation, or if they want a closed-form PDF, entropy, or Fisher information. If theoretical and practical considerations do not identify a single distribution family, we recommend selecting based on how well the different possible distributions fit the data, noting, however, that differences in fit between wrapped and intrinsically circular distributions are typically small.

### Sampling Random Values From Circular Distributions

Sampling from circular distributions is essential for generating synthetic data and experimental stimuli and for use in simulations and model recovery. For wrapped distributions, the usual method is to draw samples from the corresponding distribution in Euclidean space and then wrap the generated random values onto the circle. Distributions that are defined intrinsically on the circle require different sampling methods. For the von Mises distribution, the rejection sampling algorithm introduced by Best and Fisher (1979) has become the standard. This method samples values from an auxiliary distribution on  $[0, 2\pi)$  whose density is proportional to  $\exp(\kappa \cos \theta)$ —a distribution closely resembling the von Mises—and then uses a rejection criterion to ensure that accepted samples exactly follow the von Mises distribution. This algorithm is implemented, for example, in the function `rvm` in the R package `CircStats` (Lund & Agostinelli, 2002) and in `scipy.stats.vonmises.rvs` in the Python library `SciPy` (Virtanen et al., 2020).

More generally, inverse transform sampling provides a universal framework for generating random variables from any continuous distribution with a known CDF: generate a random value  $U$  from a uniform distribution in the interval  $[0, 1]$  and return the

smallest number  $X$  for which  $F(X) \geq U$ , where  $F$  is the CDF of the target distribution. This method can be used for sampling from circular distributions with the CDF defined over any specified interval of length  $2\pi$ .

### Beyond This Tutorial

This tutorial was intended to provide a practical foundation in circular statistics tailored to behavioral and cognitive research. For a comprehensive mathematical introduction to the topic, see the foundational textbooks by Mardia and Jupp (2009) and Jammalamadaka and SenGupta (2001). In this section, we briefly outline topics that we did not cover in the tutorial and point to recommended resources.

This tutorial did not cover circular *inferential statistics*. A wide range of tests assess whether circular data deviate from expected distributions—for an accessible overview, see Landler et al. (2018). For an introduction to circular–circular correlations, refer to Jammalamadaka and SenGupta (2001, Chapter 8) and (Mardia and Jupp (2009, Chapter 3.5.2)). There are circular versions of many of the familiar tests for comparing samples. These include the Watson–Williams Test (a circular equivalent to the two-sample two-sided  $t$  test; Watson & Williams, 1956), analysis of variance (Harrison et al., 1986), and multivariate analysis of variance (Landler et al., 2022). Landler et al. (2021) evaluate different statistical tests comparing two independent samples of circular data in terms of Type I error rate and statistical power. Tasdan and Yeniay (2018) discuss Bonferroni corrections for circular data. For *regression models* of circular variables with circular predictors, there are methodological guidelines for frequentist ( $p$  value) tests (Cremers et al., 2020; Downs, 2002; Jammalamadaka & SenGupta, 2001, Chapter 8) and Bayesian statistics (Jha & Bhuyan, 2021). For a review of methods for principal component analysis of circular and higher-dimensional directional data, we recommend Nodehi et al. (2026).

*Clustering algorithms* offer a way to identify subgroups or outlier candidates based on angular proximity. There are circular versions of many established clustering algorithms, including circular versions of  $k$ -means (Charalampidis, 2005), fuzzy  $c$ -means (Kesemen et al., 2016), and the machine-learning algorithm mean-shift (Zhang & Chen, 2021). Machine learning methods have also been adopted to predict (Lang et al., 2020) and simulate (Wessel et al., 2026) circular variables.

Some specific cases of circular data require special considerations to avoid biases and distorted results. While this tutorial focused on continuous circular variables, treating noncontinuous data as continuous can substantially bias results (Mardia & Sriram, 2023), even with a fine-grained scale such as rounding to the nearest full degree (Landler et al., 2019). Mardia and Sriram (2023) introduce different families of *circular discrete distributions*, including the circular lattice distribution, which is defined over a finite set of equally spaced points on a circle (Mardia & Jupp, 2009, Chapter 3.5.2). For the classification and time-series modeling of ordinal circular data, refer to Gehlot and Laha (2025) and Jha and Biswas (2018), respectively. Adaptations for grouped or rounded data of tests that assess whether circular data deviate from expected distributions are discussed by Jammalamadaka et al. (2024) for uniformity testing and by Landler et al. (2020) for testing against expected distributions in general.

*Incomplete or nonuniform sampling* can also bias results; for instance, if a variable is collected only during a constrained time window of a 24-hour cycle (e.g., during daytime). Alsammani et al. (2024) discuss methods to mitigate these measurement biases. For the analysis of circular longitudinal and time-series data, some special considerations are warranted. Refer to Artes and Toloï (2009), Fisher and Lee (1994), Nuñez-Antonio and Gutiérrez-Peña (2014), and Zhan et al. (2025) for guidelines on time-series modeling, as well as to Abuzaid et al. (2014) for outlier-classification methods in circular time-series models.

Other forms of data closely related to circular data are also relevant for specific topics in behavioral and cognitive research but go beyond the scope of this tutorial. Cylindrical data comprise a circular component and a linear component. Cremers and Klugkist (2018) provide an introduction to cylindrical data in behavioral research. For an introduction to circular-linear correlations, refer to Jammalamadaka and SenGupta (2001, Chapter 8) and Mardia and Jupp (2009, Chapter 3.5.2). For regression models of circular variables on linear predictors, there are methodological guidelines for frequentist statistics (Cremers et al., 2020; Fisher & Lee, 1992; Jammalamadaka & SenGupta, 2001, Chapter 8), as well as Bayesian statistics (Cremers et al., 2018; Mulder & Klugkist, 2017). For perspectives on directional data analysis that include spherical and higher-dimensional data, refer to Fisher et al. (1993), Mardia and Jupp (2009), and Pewsey and García-Portugués (2021).

## References

- Abuzaid, A. H., Mohamed, I. B., & Hussin, A. G. (2014). Procedures for outlier detection in circular time series models. *Environmental and Ecological Statistics*, 21(4), 793–809. <https://doi.org/10.1007/s10651-014-0281-8>
- Alsammani, A., Stacey, W. C., & Gliske, S. V. (2024). Estimation of circular statistics in the presence of measurement bias. *IEEE Journal of Biomedical and Health Informatics*, 28(2), 1089–1100. <https://doi.org/10.1109/jbhi.2023.3334684>
- Ameijeiras-Alonso, J., Benali, A., Crujeiras, R. M., Rodríguez-Casal, A., & Pereira, J. M. C. (2019). Fire seasonality identification with multimodality tests. *Annals of Applied Statistics*, 13(4). <https://doi.org/10.1214/19-aos1273>
- Artes, R., & Toloï, C. M. C. (2009). An autoregressive model for time series of circular data. *Communications in Statistics—Theory and Methods*, 39(1), 186–194. <https://doi.org/10.1080/03610920802650338>
- Best, D. J., & Fisher, N. I. (1979). Efficient simulation of the von Mises distribution. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 28(2), 152–157. <https://doi.org/10.2307/2346732>
- Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, 113(4), 700–765. <https://doi.org/10.1037/0033-295X.113.4.700>
- Bolón, D., Crujeiras, R. M., & Rodríguez-Casal, A. (2025). A likelihood ratio test for circular multimodality. *Environmental and Ecological Statistics*, 32(1), 57–87. <https://doi.org/10.1007/s10651-024-00639-1>
- Chang-Chien, S.-J., Hung, W.-L., & Yang, M.-S. (2012). On mean shift-based clustering for circular data. *Soft Computing*, 16(6), 1043–1060. <https://doi.org/10.1007/s00500-012-0802-z>
- Charalampidis, D. (2005). A modified *k*-means algorithm for circular invariant clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(12), 1856–1865. <https://doi.org/10.1109/TPAMI.2005.230>
- Cremers, J., & Klugkist, I. (2018). One direction? A tutorial for circular data analysis using R with examples in cognitive psychology. *Frontiers in Psychology*, 9, Article 2040. <https://doi.org/10.3389/fpsyg.2018.02040>
- Cremers, J., Mainhard, T., & Klugkist, I. (2018). Assessing a Bayesian embedding approach to circular regression models. *Methodology*, 14(2), 69–81. <https://doi.org/10.1027/1614-2241/a000147>
- Cremers, J., Pennings, H. J. M., & Ley, C. (2020). Regression models for cylindrical data in psychology. *Multivariate Behavioral Research*, 55(6), 910–925. <https://doi.org/10.1080/00273171.2019.1693332>
- Cremers, J., Pennings, H. J. M., Mainhard, T., & Klugkist, I. (2021). Circular modelling of circumplex measurements for interpersonal behavior. *Assessment*, 28(2), 585–600. <https://doi.org/10.1177/1073191119858407>
- Downs, T. D. (2002). Circular regression. *Biometrika*, 89(3), 683–698. <https://doi.org/10.1093/biomet/89.3.683>
- Fisher, N. I., & Lee, A. J. (1992). Regression models for an angular response. *Biometrics*, 48(3), 665–677. <https://doi.org/10.2307/2532334>
- Fisher, N. I., & Lee, A. J. (1994). Time series analysis of circular data. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 56(2), 327–339. <https://doi.org/10.1111/j.2517-6161.1994.tb01981.x>
- Fisher, N. I., Lewis, T., & Embleton, B. J. (1993). *Statistical analysis of spherical data* (New ed.). Cambridge University Press.
- Gehlot, S., & Laha, A. K. (2025). Classification with ordinal circular data. In S. Kumar, B. C. Arnold, K. Shimizu, & A. K. Laha (Eds.), *Directional and multivariate statistics* (pp. 87–102). Springer Nature. [https://doi.org/10.1007/978-981-96-2004-3\\_5](https://doi.org/10.1007/978-981-96-2004-3_5)
- Gordon, L., & Hudson, M. (1977). A characterization of the von Mises distribution. *Annals of Statistics*, 5(4), 813–814. <https://doi.org/10.1214/aos/1176343906>
- Gumbel, E. J., Greenwood, J. A., & Durand, D. (1953). The circular normal distribution: Theory and tables. *Journal of the American Statistical Association*, 48(261), 131–152. <https://doi.org/10.1080/01621459.1953.10483462>
- Gutleb, T., Carrillo, J., & Olver, S. (2022). Computing equilibrium measures with power law kernels. *Mathematics of Computation*, 91(337), 2247–2281. <https://doi.org/10.1090/mcom/3740>
- Harrison, D., Kanji, G. K., & Gadsden, R. J. (1986). Analysis of variance for circular data. *Journal of Applied Statistics*, 13(2), 123–138. <https://doi.org/10.1080/02664768600000021>
- Jammalamadaka, S. R., Ghosh, K., & Akiri, S. (2024). Testing uniformity on the circle using spacings when data are rounded. *Statistics and Probability Letters*, 210, Article 110114. <https://doi.org/10.1016/j.spl.2024.110114>
- Jammalamadaka, S. R., & SenGupta, A. (2001). *Topics in circular statistics*. World Scientific. <https://doi.org/10.1142/4031>
- Jha, J., & Bhuyan, P. (2021). Two-stage circular-circular regression with zero inflation: Application to medical sciences. *Annals of Applied Statistics*, 15(3), 1343–1365. <https://doi.org/10.1214/20-aos1429>
- Jha, J., & Biswas, A. (2018). Distribution and time-series modelling of ordinal circular data. *Environmetrics*, 29(2), Article e2490. <https://doi.org/10.1002/env.2490>
- Jones, M. C., & Pewsey, A. (2005). A family of symmetric distributions on the circle. *Journal of the American Statistical Association*, 100(472), 1422–1428. <https://doi.org/10.1198/016214505000000286>
- Jones, M. C., & Pewsey, A. (2012). Inverse Batschelet distributions for circular data. *Biometrics*, 68(1), 183–193. <https://doi.org/10.1111/j.1541-0420.2011.01651.x>
- Kent, J. T., Mardia, K. V., & Rao, J. S. (1979). A characterization of the uniform distribution on the circle. *Annals of Statistics*, 7(4), 882–889. <https://doi.org/10.1214/aos/1176344737>
- Kesemen, O., Tezel, Ö., & Özkul, E. (2016). Fuzzy *c*-means clustering algorithm for directional data (FCM4DD). *Expert Systems with Applications*, 58, 76–82. <https://doi.org/10.1016/j.eswa.2016.03.034>
- Landler, L., Ruxton, G. D., & Malkemper, E. P. (2018). Circular data in biology: Advice for effectively implementing statistical procedures.

- Behavioral Ecology and Sociobiology*, 72(8), Article 128. <https://doi.org/10.1007/s00265-018-2538-y>
- Landler, L., Ruxton, G. D., & Malkemper, E. P. (2019). Circular statistics meets practical limitations: A simulation-based Rao's spacing test for non-continuous data. *Movement Ecology*, 7(1), Article 15. <https://doi.org/10.1186/s40462-019-0160-x>
- Landler, L., Ruxton, G. D., & Malkemper, E. P. (2020). Grouped circular data in biology: Advice for effectively implementing statistical procedures. *Behavioral Ecology and Sociobiology*, 74(8), Article 100. <https://doi.org/10.1007/s00265-020-02881-6>
- Landler, L., Ruxton, G. D., & Malkemper, E. P. (2021). Advice on comparing two independent samples of circular data in biology. *Scientific Reports*, 11(1), Article 20337. <https://doi.org/10.1038/s41598-021-99299-5>
- Landler, L., Ruxton, G. D., & Malkemper, E. P. (2022). The multivariate analysis of variance as a powerful approach for circular data. *Movement Ecology*, 10(1), Article 21. <https://doi.org/10.1186/s40462-022-00323-8>
- Lang, M. N., Schlosser, L., Hothorn, T., Mayr, G. J., Stauffer, R., & Zeileis, A. (2020). Circular regression trees and forests with an application to probabilistic wind direction forecasting. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 69(5), 1357–1374. <https://doi.org/10.1111/rssc.12437>
- Liu, Y., & West, S. G. (2016). Weekly cycles in daily report data: An overlooked issue. *Journal of Personality*, 84(5), 560–579. <https://doi.org/10.1111/jopy.12182>
- Lund, U., & Agostinelli, C. (2002). *CircStats: Circular statistics* (Version 0.2-7) [Computer software]. <https://doi.org/10.32614/CRAN.package.CircStats>
- Mardia, K. V. (1975). Statistics of directional data. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 37(3), 349–371. <https://doi.org/10.1111/j.2517-6161.1975.tb01550.x>
- Mardia, K. V., & Jupp, P. E. (2009). *Directional statistics*. John Wiley & Sons.
- Mardia, K. V., & Sriram, K. (2023). Families of discrete circular distributions with some novel applications. *Sankhya A*, 85(1), 1–42. <https://doi.org/10.1007/s13171-022-00298-z>
- Mitre-Ortiz, A., & Mitre-Hernandez, H. (2019). Study of spontaneous and acted learn-related emotions through facial expressions and galvanic skin response. *Research in Computing Science*, 148(5), 97–105. <https://doi.org/10.13053/rcs-148-5-11>
- Mulder, K., & Klugkist, I. (2017). Bayesian estimation and hypothesis tests for a circular generalized linear model. *Journal of Mathematical Psychology*, 80, 4–14. <https://doi.org/10.1016/j.jmp.2017.07.001>
- Murray, R. F., & Morgenstern, Y. (2010). Cue combination on the circle and the sphere. *Journal of Vision*, 10(11), Article 15. <https://doi.org/10.1167/10.11.15>
- Nodehi, A., Moghimbeygi, M., & Ley, C. (2026). Recent advances in principal component analysis for directional data. *Journal of Multivariate Analysis*, 212, Article 105528. <https://doi.org/10.1016/j.jmva.2025.105528>
- Núñez-Antonio, G., & Gutiérrez-Peña, E. (2014). A Bayesian model for longitudinal circular data based on the projected normal distribution. *Computational Statistics and Data Analysis*, 71, 506–519. <https://doi.org/10.1016/j.csda.2012.07.025>
- Pewsey, A. (2008). The wrapped stable family of distributions as a flexible model for circular data. *Computational Statistics and Data Analysis*, 52(3), 1516–1523. <https://doi.org/10.1016/j.csda.2007.04.017>
- Pewsey, A., & García-Portugués, E. (2021). Recent advances in directional statistics. *TEST*, 30(1), 1–58. <https://doi.org/10.1007/s11749-021-00759-x>
- Purkayastha, S. (1995). An almost sure representation of sample circular median. *Journal of Statistical Planning and Inference*, 46(1), 77–91. [https://doi.org/10.1016/0378-3758\(94\)00092-A](https://doi.org/10.1016/0378-3758(94)00092-A)
- Rao, C. R. (1945). Information and the accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society*, 37(3), 81–91.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85(2), 59–108. <https://doi.org/10.1037/0033-295X.85.2.59>
- Smith, P. L. (2016). Diffusion theory of decision making in continuous report. *Psychological Review*, 123(4), 425–451. <https://doi.org/10.1037/rev0000023>
- Tasdan, F., & Yeniay, O. (2018). A comparative simulation of multiple testing procedures in circular data problems. *Journal of Applied Statistics*, 45(2), 255–269. <https://doi.org/10.1080/02664763.2016.1273886>
- Trommershauser, J., Kording, K., & Landy, M. S. (2011). *Sensory cue integration*. Oxford University Press.
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., . . . SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in python. *Nature Methods*, 17(3), 261–272. <https://doi.org/10.1038/s41592-019-0686-2>
- von Mises, R. (1918). Über die “Ganzzahligkeit” der Atomgewichte und verwandte Fragen. *Physikalische Zeitschrift*, 19, 490–500.
- Watson, G. S., & Williams, E. J. (1956). On the construction of significance tests on the circle and the sphere. *Biometrika*, 43(3–4), 344–352. <https://doi.org/10.1093/biomet/43.3-4.344>
- Wernicke, S., De Falco, J., Gironés, Z., Jungerius, C., Tomic, I., Triabhall, A., Wang, X., & Bays, P. (2026). *Working with circular data: A tutorial for cognitive and behavioral research*. OSF. <https://osf.io/5n64t>
- Wessel, J. B., Murphy-Barltrop, C. J. R., & Simpson, E. S. (2026). Generative machine learning for multivariate angular simulation. *Extremes*, 29(2), 195–243. <https://doi.org/10.1007/s10687-025-00522-7>
- Zhan, X., Ma, T., Liu, S., & Shimizu, K. (2025). Circular data diagnostics in regression and time series models. In S. Kumar, B. C. Arnold, K. Shimizu, & A. K. Laha (Eds.), *Directional and multivariate statistics* (pp. 3–23). Springer Nature. [https://doi.org/10.1007/978-981-96-2004-3\\_1](https://doi.org/10.1007/978-981-96-2004-3_1)
- Zhang, Y., & Chen, Y.-C. (2021). Kernel smoothing, mean shift, and their learning theory with directional data. *Journal of Machine Learning Research*, 22(154), 1–92. <http://jmlr.org/papers/v22/20-1194.html>

(Appendices follow)

## Appendix A

### List of Symbols

| Symbol                                                | Description                                                                                                                                                                 |
|-------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| $\mathbb{R}^p$                                        | $p$ -dimensional Euclidean space.                                                                                                                                           |
| $\mathbb{S}^p$                                        | $p$ -dimensional directional space.                                                                                                                                         |
| $\mathbb{S}^1$                                        | Circular (one-dimensional directional) space.                                                                                                                               |
| $x, y$ (Set: $X, Y$ )                                 | Random variables.                                                                                                                                                           |
| $\theta$ (Set: $\Theta$ )                             | Polar angle on a circle.                                                                                                                                                    |
| $[x, y]$                                              | Pair of rectangular coordinates. Used to represent circular values in a Cartesian coordinate system.                                                                        |
| $(r, \theta)$                                         | Polar vector representation. $r$ is the radial coordinate, indicating the length of the vector; $\theta$ is the angular coordinate, specifying the direction of the vector. |
| $O$                                                   | The origin, i.e., point where all coordinates are 0, e.g., $[0, 0]$ in $\mathbb{R}^2$ .                                                                                     |
| $i$                                                   | The imaginary unit, $\sqrt{-1}$ .                                                                                                                                           |
| $z$                                                   | A unit complex number.                                                                                                                                                      |
| $\oplus$ and $\ominus$                                | Modular addition and subtraction in a circular space.                                                                                                                       |
| $D(\theta, \phi)$                                     | Circular distance between the angles $\theta$ and $\phi$ .                                                                                                                  |
| $\kappa$                                              | Concentration parameter of a von Mises distribution ( $\frac{1}{\kappa}$ is analogous to $\sigma^2$ ).                                                                      |
| $\sigma^2$                                            | Euclidean population variance.                                                                                                                                              |
| $\hat{m}_p$                                           | $p$ th circular moment of a sample.                                                                                                                                         |
| $\mathbf{R}$                                          | Resultant vector of a sample.                                                                                                                                               |
| $\hat{m}_1$                                           | Sample mean resultant vector, i.e., sample-size-adjusted resultant vector.                                                                                                  |
| $R$                                                   | The resultant length of a sample, i.e., the length of the resultant vector.                                                                                                 |
| $\bar{R}$                                             | Sample mean resultant length, i.e., the length of the mean resultant vector.                                                                                                |
| $\bar{\theta}$                                        | Sample circular mean.                                                                                                                                                       |
| $\theta_{\text{med}}$                                 | Sample circular median.                                                                                                                                                     |
| $\theta_{\text{mode}}$                                | Sample circular mode.                                                                                                                                                       |
| $\hat{V}_{\theta}$                                    | Sample circular variance.                                                                                                                                                   |
| $s_{\theta}$                                          | Sample circular standard deviation.                                                                                                                                         |
| $g_1$                                                 | Sample Euclidean skewness.                                                                                                                                                  |
| $g_{1\theta}$                                         | Sample circular skewness.                                                                                                                                                   |
| $b_2$                                                 | Sample Euclidean kurtosis.                                                                                                                                                  |
| $g_2$                                                 | Sample Euclidean excess kurtosis.                                                                                                                                           |
| $g_{2\theta}$                                         | Sample circular kurtosis.                                                                                                                                                   |
| $m_p$                                                 | $p$ th circular moment of a population.                                                                                                                                     |
| $m_1$                                                 | Population mean resultant vector.                                                                                                                                           |
| $\mu$                                                 | Population circular mean.                                                                                                                                                   |
| $\mu_{\text{med}}$                                    | Population circular median.                                                                                                                                                 |
| $\mu_{\text{mode}}$                                   | Population circular mode.                                                                                                                                                   |
| $V_{\theta}$                                          | Population circular variance.                                                                                                                                               |
| $\sigma_{\theta}$                                     | Population circular standard deviation.                                                                                                                                     |
| $\gamma_{1\theta}$                                    | Population circular skewness.                                                                                                                                               |
| $\gamma_{2\theta}$                                    | Population circular kurtosis.                                                                                                                                               |
| $\rho$                                                | Population mean resultant vector length.                                                                                                                                    |
| $\varphi_k$                                           | Characteristic function of a distribution.                                                                                                                                  |
| $f_N(x; \mu, \sigma)$ or $\mathcal{N}(\mu, \sigma^2)$ | The Gaussian (normal) distribution.                                                                                                                                         |
| $f_U(\theta)$                                         | The circular uniform distribution.                                                                                                                                          |
| $f_{VM}(\theta; \mu, \kappa)$                         | The von Mises distribution (circular normal).                                                                                                                               |
| $f_{WN}(\theta; \mu, \sigma)$                         | The wrapped normal distribution.                                                                                                                                            |
| $f_{WC}(\theta; \mu, \gamma)$                         | The wrapped Cauchy distribution with scale parameter $\gamma$ .                                                                                                             |
| $f_{WS}(\theta; \alpha, \beta, \gamma, \delta)$       | The wrapped stable family with stability parameter $\alpha$ , skewness parameter $\beta$ , scale parameter $\gamma$ , and location parameter $\delta$ .                     |
| $f_{SS-JP}(\theta; \delta, \kappa, \psi, \lambda)$    | The sine-skewed Jones–Pewsey family with location parameter $\delta$ , concentration $\kappa$ , shape parameter $\psi$ , and skewness parameter $\lambda$ .                 |
| $I_n(\kappa)$                                         | The modified Bessel function of the first kind of order $n$ .                                                                                                               |
| $A_n(\kappa)$                                         | The ratio of Bessel functions $\frac{I_n(\kappa)}{I_0(\kappa)}$ .                                                                                                           |
| $\mathbb{E}[X]$                                       | The expectation of random variable $X$ .                                                                                                                                    |
| $\mathbb{V}\text{ar}[X]$                              | The variance of random variable $X$ .                                                                                                                                       |
| $\mathcal{I}(\theta)$                                 | Fisher information of a circular variable $\theta$ .                                                                                                                        |

(Appendices continue)

## Appendix B

### Sampling Circular Values With Minimum Separation

Sampling values from a circular uniform distribution while enforcing a minimum angular separation between samples can be achieved using the following algorithm. The function takes as input the number of samples  $N$ , the minimum angular separation  $\text{minDistance}$ , and the lower and upper bound of the circular space.

---

Algorithm 1: RandMinDistance ( $N$ ,  $\text{minDistance}$ ,  $\text{lower}$ ,  $\text{upper}$ )

**Require:** Number of samples  $N$ , minimum separation  $\text{minDistance}$

```

1: if  $\text{lower}$ ,  $\text{upper}$  is undefined then
2:    $\text{lower} \leftarrow -\pi$ ,  $\text{upper} \leftarrow \pi$                                 ▶ Default circular range
3: end if
4:  $\text{range} \leftarrow \text{upper} - \text{lower}$ 
5:  $l \leftarrow \text{range} - N \times \text{minDistance}$ 
6: if  $l < 0$  then
7:   error: "Range too small to draw  $N$  samples with given minimum distance."
8: end if
9: ▶ Generate  $N$  random values within the remaining free length
10: for  $i = 1$  to  $N$  do
11:    $r[i] \leftarrow \text{Uniform}(0, l)$ 
12: end for
13:  $(s, \text{idx}) \leftarrow \text{Sort}(r)$                                 ▶ Sort and save original indices
14: ▶ Enforce minimum distance
15: for  $i = 1$  to  $N$  do
16:    $s[i] \leftarrow s[i] + (i - 1) \times \text{minDistance}$ 
17: end for
18:  $r \leftarrow \text{Order}(s, \text{idx})$                                 ▶ Restore original order (optional)
19: ▶ Apply global random rotation and wrap onto the circle
20:  $u \leftarrow \text{Uniform}(0, \text{range})$ 
21: for  $i = 1$  to  $N$  do
22:    $r[i] \leftarrow \text{lower} + ((r[i] + u - \text{lower}) \bmod \text{rangeLength})$ 
23: end for
24: return  $r$ 

```

---

First, this algorithm computes the total free length available after enforcing the minimum spacing requirement. It then samples  $N$  random values within this remaining space. After sorting the samples, a minimum separation of at least  $\text{minDistance}$  of adjacent values is ensured by applying a cumulative offset of  $(i - 1) \times \text{minDistance}$ . Optionally, the original order of the values can be restored—for instance, when experimental stimuli are presented sequentially and should not appear in a sorted order. Finally, the list of values is rotated by a random offset and wrapped back onto the circular range, yielding a uniform random arrangement of points while ensuring the specified minimum separation.

The separation constraints and the maximum number of samples are inversely related: for a minimum separation distance  $\Delta_{\min}$ , the number of samples  $n$  is constrained by  $n \leq \left\lfloor \frac{R}{\Delta_{\min}} \right\rfloor$ . Once the separation constraints are specified, angular distance (see the Measuring Distance section) can be computed pairwise to all data points to verify that every pair of circular observations meets the requirements.

(Appendices continue)

## Appendix C

### Jones–Pewsey PDF via Numerical Integration

In Equation 41, the normalizing constant of the Jones–Pewsey distribution is expressed as an associated Legendre function, which becomes numerically unstable when  $\psi$  is large and negative. The algorithm that we introduce in this appendix avoids this by computing the density using numerical integration of the kernel and working in log space to prevent floating-point overflow. When  $\psi = 0$ , the distribution reduces to the von Mises and its analytic normalizing constant is used instead.

---

Algorithm 2: JonesPewseyPDF ( $\theta, \mu, \kappa, \psi$ )

**Require:** Angles  $\theta$ , location parameter  $\mu$ , concentration  $\kappa$ , shape parameter  $\psi$

```

1: if  $|\psi| < 10^{-6}$  then
2:    $f \leftarrow \exp(\kappa \times \cos(\theta - \mu)) / 2 \times \pi \times I_0(\kappa)$  ► Special case: von Mises
3: else
4:    $\text{core} \leftarrow \cosh(\kappa \times \psi) + \sinh(\kappa \times \psi) \times \cos(\theta - \mu)$  ► Compute kernel argument at each angle
5:   if  $\text{core} \leq 0$  for any  $\theta$  then
6:     error: "Non-positive JP kernel; check parameter regime."
7:   end if
8:    $\text{logg} \leftarrow 1/\psi \times \log(\text{core})$  ► Compute log-kernel
9:    $\text{logg} \leftarrow \text{logg} - \max(\text{logg})$  ► Subtract maximum for numerical stability
10:   $g \leftarrow \exp(\text{logg})$  ► Transform to linear scale
11:   $f \leftarrow g / \text{trapezoidalIntegral}(\theta, g)$  ► Normalize numerically
12: end if
13: return  $f$ 

```

---

To calculate the sine-skewed Jones–Pewsey distribution, using the same approach, the following wrapper function can be used:

---

Algorithm 3: SineSkewedJonesPewseyPDF ( $\theta, \mu, \kappa, \psi, \lambda$ )

**Require:** Angles  $\theta$ , location parameter  $\mu$ , concentration  $\kappa$ , shape parameter  $\psi$ , skewness  $\lambda \in [-1, 1]$

```

1:  $f_{\text{sym}} \leftarrow \text{JonesPewseyPDF}(\theta, \mu, \kappa, \psi)$  ► Evaluate Jones-Pewsey base density
2:  $f \leftarrow (1 + \lambda \times \sin(\theta - \mu)) \times f_{\text{sym}}$  ► Apply sine-skewing
3:  $f \leftarrow f / \text{trapezoidalIntegral}(\theta, f)$  ► Renormalize numerically (safeguard)
4: return  $f$ 

```

---

## Appendix D

### Additivity of Mean and Variance

The Euclidean mean or expectation is additive in the sense that the mean of the sum of any two random variables is equal to the sum of their means,

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y], \quad (\text{E1})$$

and Euclidean variance is additive in the sense that the variance of the sum of two random variables is equal to the sum of their variances, *so long as the variables are uncorrelated*,

$$\begin{aligned} \text{Var}[X + Y] &= \text{Var}[X] + \text{Var}[Y] + 2\text{Cov}[X, Y] \\ &= \text{Var}[X] + \text{Var}[Y] \quad \text{iff } X \perp Y. \end{aligned} \quad (\text{E2})$$

Circular equivalents to mean and variance are related to the direction and length of the mean resultant vector, which for the angular sum of two variables  $\theta$  and  $\phi$  is given by

$$\begin{aligned} m_{\theta+\phi} &= \mathbb{E}[e^{i(\theta+\phi)}] = \mathbb{E}[e^{i\theta}e^{i\phi}] \\ &= \mathbb{E}[e^{i\theta}]\mathbb{E}[e^{i\phi}] + \text{Cov}[e^{i\theta}, e^{i\phi}] \\ &= m_\theta m_\phi \quad \text{iff } \theta \perp \phi, \end{aligned} \quad (\text{E3})$$

that is, the mean resultant vector of the sum of two independent random angles is the *product* of their individual mean resultant vectors.

We confine ourselves now to the case of independent  $\theta$  and  $\phi$ . The circular mean is the direction or argument of the mean resultant vector,

$$\begin{aligned} \mu[\theta + \phi] &= \arg m_{\theta+\phi} = \arg(m_\theta m_\phi) \\ &= \arg m_\theta + \arg m_\phi \pmod{2\pi} \\ &= \mu[\theta] + \mu[\phi] \pmod{2\pi}, \end{aligned} \quad (\text{E4})$$

using  $\arg(z_1 z_2) = \arg z_1 + \arg z_2$  for any complex  $z_1, z_2$ . In other words, the circular mean is additive for independent variables. Circular measures of dispersion are related to the mean resultant length,

$$\rho[\theta + \phi] = |m_{\theta+\phi}| = |m_\theta m_\phi| = |m_\theta| |m_\phi| = \rho[\theta] \rho[\phi], \quad (\text{E5})$$

using  $|z_1 z_2| = |z_1| |z_2|$  for any complex  $z_1, z_2$ . In other words, the mean resultant length does not have the additive property, and by extension neither does the circular variance, based on its usual definition  $V_\theta = 1 - \rho$ . However,

$$\log \rho[\theta + \phi] = \log \rho[\theta] + \log \rho[\phi], \quad (\text{E6})$$

that is, the logarithm of mean resultant length is additive for independent circular random variables. Based on the definition of circular standard deviation,

$$\sigma_\theta = \sqrt{-2 \log \rho} \Rightarrow \sigma_\theta^2 = -2 \log \rho, \quad (\text{E7})$$

it follows that squared circular standard deviation is also additive for independent variables,

$$\sigma_\theta^2[\theta + \phi] = \sigma_\theta^2[\theta] + \sigma_\theta^2[\phi] \quad \text{if } \theta \perp \phi. \quad (\text{E8})$$

Received January 12, 2026  
Revision received June 16, 2026  
Accepted June 23, 2026 ■