SWC-GCNU NEUROSTATISTICS 2025:
CIRCULAR STATISTICS

Kira M. Düsterwald
February 17, 2025

Contents

1 What is circular data?

Let's start with some hopefully intuitive examples of circular data: hours in a day; months in a year; wind direction; the directions of different edges in a visual scene; behavioural responses to brief presentations of different orientations; neural tuning to orientation (could be V1 or MT); neural tuning to oscillation (phase); the 3-dimensional direction of flight for a bat; and so many more!

Colloquially, we can think of circular data as data that:

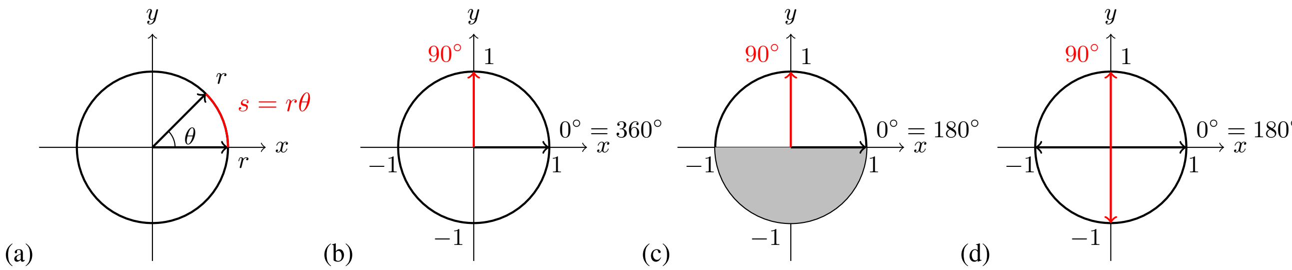

Figure 1: (a) Arc length in radians. (b-d) 90° represented with different frames of reference.

1. Has no true "zero" or (monotonic) ordering
2. Is periodic – the data "wraps around" and repeats
3. Something we can express with an angle, or orientation, or azimuth

In this lecture, we'll discuss a sub-discipline called directional (or circular) statistics: dealing with data in the class of directions or unit vectors on $\mathbb{R}^n$.

1.1 How can we express circular data?

Usually we describe circular data from some reference angle (this is the "zero") and we take the azimuth (angle) from there as a given measurement or data point. Conventionally, the reference is given along the positive x-axis when the circle is embedded in rectangular co-ordinates, and the positive direction is counter-clockwise of that reference (Fig. 1)[1]. When we wrap around the circle and get close to the reference again, we can calculate the reset/wrap using modular arithmetic.

1.1.1 Radians

Preferably, we will use **radians**, for some reasons that you will see later, but most simply because radians easily associate an angle θ with its arc length s: $s = r\theta$, where r is the radius of the circle (Fig. 1 A). If $r = 1$, which as we'll see soon is a reasonable way to represent circular data, then the arc length *is* just the angle in radians!

This simple way of thinking about circular data points out the fundamental difference between it and linear data as we are used to: angles are points on an arc or positions on a circle, and these are *topologically* distinct from a line segment.

To convert from an angle x in degrees to one in radians use:

$$\theta = \frac{2\pi x^\circ}{k}, \tag{1}$$

where k is the scale or normaliser (the periodicity).

As an exercise, calculate in radians what 90° is on a 360° scale, and what 90° is on a 180° scale. What does each case look like graphically (hint: match the cases to Fig. 1 B-D)?

Using angles generally works, but has the annoying property that one needs to use modulus operations to calculate the angle after periodicity, or wrap-around. This will cause some problems with calculating circular qualities later, because it's a pain to convert angles with modular arithmetic all the time, as we will see later.

One way around this is to put everything into trigonometry: trigonometric functions naturally wrap around the circle. So we could constantly use trigonometric formulae and their inverses to map between angles and wraps... Phew!

1.1.2 Complex numbers and Euler's formula

Another way of representing circular data that "automatically" does the wrap is by using complex numbers. This is because we can write *any* complex number z as $z = Re^{i\theta}$ where R is the complex number length, and θ the angle between the complex number and the reference, and complex numbers rotate upon exponential increase and relate neatly to trigonmetry (more on this later!). We can represent the complex formulation graphically using a real and

[1]Although using 180° or π radians as the wrap may actually be more convenient, as it is more symmetric.

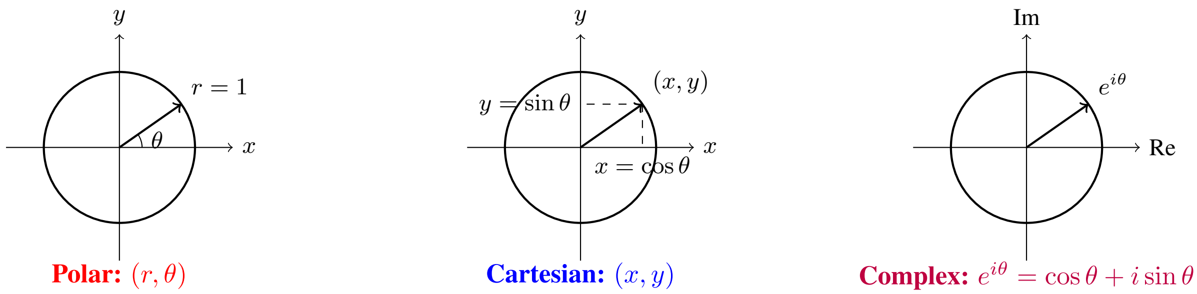

Figure 2: Polar, Cartesian (rectangular), and complex representations of circular data.

imaginary axis (Fig. 2). The last statement about complex numbers rotating on exponential increase follows from Euler's formula:

$$e^{i\theta} = \cos\theta + i\sin\theta \tag{2}$$

If we have time, I'll give some intuition in class as to why this formula is true. For now, just take it as a given, or review 3Blue1Brown's excellent video on YouTube, which explains elementary group theory and how it applies to this formula!

2 The mean of circular data

Given some circular data, we want to be able to say some things about that data. If I were to give you a bunch of linear data, you might try to take some measure of central tendency: (arithmetic) mean, median, mode, etc. We'll discuss different definitions for the mean of circular data.

2.1 Our usual, casual notion of mean is not sufficient

There are several different types of "mean" (cubic, harmonic, ...) but we'll stick to the usual arithmetic (linear) mean for this example. The linear mean of N data points $\{\alpha_1, \alpha_2, ..., \alpha_N\}$ is usually given by $\bar{\alpha} = \frac{\sum \alpha_i}{N}$. Sometimes we might also give it weightings w_i, in which case we get a weighted mean $\bar{\alpha} = \frac{\sum w_i \alpha_i}{\sum w_i}$. This notion can easily be extended to the usual sense of (discrete) expected value when $\sum w_i = 1$.

Let's try to apply this closed form for the arithmetic mean to some data to see why it is problematic for circular data. For example, say our data points are $\{\frac{\pi}{6}, \frac{\pi}{8}, 2\pi - \frac{\pi}{6}, 2\pi - \frac{\pi}{8}\}$. Use the circular reference $(0, 2\pi]$ and do the calculation to show that the linear mean of these data points is π. Does the answer match your expectations for this data? What if we change the reference range to $(-\pi, \pi]$?

Probably the answer does not make much sense to you, because:

- We intuitively expect the mean to be 0 – that's the point around which the data is clustered, anyway – but we do not always get 0.
- The mean changes with different reference ranges, but these really shouldn't change much about the data, so that's also not what we would like.

The main problem here comes about with using a definition on a linear space on our circular data, which is no longer linear! Put with more maths, the usual notion of mean is constrained to simply connected Euclidean spaces and fails in multiply connected spaces. Sure, we can embed our data on the Euclidean plane, but it isn't really based there. Instead, we need a mean defined for our new space.

2.2 Extrinsic or circular mean

The obvious way around this is to forget that the circle is a circle and just embed it in the plane – then we can use our usual notion of mean on the 2-dimensional space, constrained by the equation for the circle, and just convert back to angle afterwards! Sounds good in principle, right? Let's see what happens.

That is, we can use the vector mean:

$$\begin{aligned}\vec{\mu} &= \frac{1}{N}\sum_{j=1}^{N}\vec{\alpha_j} = \frac{1}{N}\left[\sum_{j=1}^{N}x_j, \sum_{j=1}^{N}y_j\right] \\ &= \frac{1}{N}\sum_{j=1}^{N}\alpha\,(\vec{\theta}_j) = \frac{1}{N}\left[\sum_{j=1}^{N}\cos\theta_j, \sum_{j=1}^{N}\sin\theta_j\right]\end{aligned}$$

Let's apply that to the example from earlier: $\{\frac{\pi}{6}, \frac{\pi}{8}, 2\pi - \frac{\pi}{6}, 2\pi - \frac{\pi}{8}\}$. You should get out $\left[\frac{2}{\sqrt{3}}, 0\right]$. How does this look graphically? Why might this be or not be a problem?

The main concern with this way of calculating the mean is that the answer no longer lies on the circle! Rather, we get an answer that is on the *disk* that the circle surrounds. This may not seem like a problem, but from the point of view of defining the space that the circular data sits in, we technically have arrived at an answer that is no longer in that space, and intuitively we might expect that the mean belongs to the same space that our data does.

Of course, we can rescale $\vec{\mu}$ to achieve this: just let $\vec{\nu} = \frac{\vec{\mu}}{|\vec{\mu}|}$, that is, take the unit vector. Alternatively, if we want to map back to angle space, and as you will see commonly, we can take the **argument** of $\vec{\mu}$ by using the inverse tangent function:

$$\bar{\alpha} = \arctan\left(\frac{1}{N}\frac{\sum_j \sin\alpha_j}{\sum_j \cos\alpha_j}\right). \tag{3}$$

2.2.1 Mean resultant vector length

$\vec{\mu}$ might not lie in the same manifold space as the original pure angular data, but what about the original length of $\vec{\mu}$ that we calculated? Does it mean anything?

This quantity, $R =\mid \vec{\mu} \mid$, is called the (sample) mean resultant vector length. It will have a value between 0 and 1 and gives some indication of the "strength" of association of the mean, or the spread of the data. The closer to 1 it is, the more dense the data. Note that R is *not* the same as the variance, since it it does not strictly calculate the deviation from the mean, but sometimes we will use $1 - R$ as a measure of spread of data. It is also useful for frequentist statistical tests.

2.3 Intrinsic mean

To take a stab at another version of circular mean that would be "intrinsic" – lie on the circle implicitly – consider what is desirable from the mean: it should represent the whole dataset, within the space of the data, and be "close" to all points in the data, in some sense.

2.3.1 p-norms: variational approach to quantify central measures of dispersion

One way of quantifying this is via a variational approach with something called the Frechét function: $f(x) = \sum_i d(x_i, x)$, where d is a distance. We would take the x that gives us the minimum of f.

For the mean, we can define the distance so that the Frechét function is $f(x) = \sum_{i,j}(x_i - x)^2$, and the mean is just x that minimises $f(x)$. That is, the mean minimises the sum of squared distances! In full:

$$\bar{x} = \arg\min_{x_0 \in \mathcal{X}} \sum_{j=1}^{N}(x_j - x_0)^2 \tag{4}$$

This is a bit like the reverse direction of defining the variance as a dispersion from the mean first!

In fact, this is a more general version of a way of defining measures of central tendency called "p-norms" which in some sense rely on a measure of dispersion (which should be monotonic) and a metric, d. We'll ignore the former today. In general, we can write:

$$F_p(\{x\}, x_0) = \left(\sum_j d(x_j, x_0)^p\right)^{\frac{1}{p}}, \tag{5}$$

where $p \in \mathbb{N}_{\not{\!\!\!\;}}$. With the Euclidean metric, we can write $d(x_j, x_0) = \mid x_j - x_0 \mid$. Sometimes you will see F written as $|| \cdot ||_p$, and called "$L - p$" (norms).

We can explore the p-norms a bit further. Let's think what $L0$ represents. What about $L1$? We'll get back to this in the circular sense shortly!

In summary, we can define the intrinsic mean with a variational approach as to equation 4. In that case, we are guaranteed to find a unique mean that lies on the unit circle!

2.3.2 Intrinsic mean with complex numbers

So the intrinsic mean is the solution to a minimisation problem but constrains the distances to be on the circle. But the search space is infinite, which means that if we do not have a nice minimisation problem, we will struggle to solve it! So, we will apply it to directional data with complex numbers to try and make the search space smaller.

Let $z = e^{i\theta}$ where $\theta \in (-\pi, \pi]$, as explained in Figure 2. The circle then is the set of points with the same *modulus* (1 in this case) and the position is given by the *argument* or *phase factor*. We can take the distance between two complex numbers using the arc, or simply $\mid \frac{1}{i} \log e^{i(\theta_1 - \theta_2)} \mid = \arg e^{i(\theta_1 - \theta_2)}$, which is often squared. In the variational form this becomes:

$$\bar{\theta} = \arg \min_{\theta_0 \in \mathcal{C}} \sum_j \left(\arg e^{i(\theta_j - \theta_0)} \right)^2 \tag{6}$$

In fact, if you know a bit of complex analysis, there is a closed form for this. All we need are the phase factors (arguments), or parts in the top of the exponents. We can find the optimal argument by using the exponentials directly, replacing sum with product and division with root. That is: $e^{i\bar{\theta}} = \sqrt[N]{\prod_j e^{i\theta_j}}$. In complex analysis, roots are just multiples around the circle (think about square-root to help you out here!), that is:

$$\left(e^{i\bar{\theta}} \right)_k = e^{i\left(\frac{\sum_j \theta_j + k2\pi}{N} \right)}, k \in \{0, 1, 2, ..., N-1\}, \tag{7}$$

or $\bar{\theta}_k = \bar{\theta}_0 + k\frac{2\pi}{N}$.

Note that even with this approach we get *multiple* possible means, but only finitely many. However, it is simple to triage this list: we can use the full distance formula to figure out which of the means has the smallest $L2$ cost and use that as the answer!

As a concrete example, let's work this out for the data $\{\frac{\pi}{6}, -\frac{\pi}{6}\}$.

You should notice that the extrinsic mean projection onto the unit circle and the intrinsic mean coincide when we have two points. However, this isn't the case when there are 3 or more data points, so be wary of assuming this is always true!

3 The median of circular data

Median is a trickier concept. The linear concept of median is something like $\alpha_{med} = \frac{\alpha_{N+1}}{2}$, or colloquially the data point that is closest to the "middle" of all the other data, so that there is an equal amount of data on either side.

The closed form calculation is a bit of a problem for circular data, since there is no ordering! However, we can still think of the median in terms of equal division of data: the median is the angle that lies closes to the diameter of the circle that divides the data into two equally sized groups.

Using what we have learned about the intrinsic circular mean, we can define an equivalent expression for the median:

$$\bar{m} = \arg \min_{x_0 \in \mathcal{X}} \sum d(x_0, x_j)^1 \tag{8}$$

Indeed we can also use complex numbers to write this as $\arg\min_{\theta_0} \sum_j \mid \arg e^{i(\theta_j - \theta_0)} \mid$.

Note that there is no "closed form!"

4 Inferential statistics on a circle

For linear data, when there is lots of it we can invoke the central limit theorem. That is, the distribution of the normalized sample mean converges to a standard normal distribution, and this approximation gets better the more data we have. This is true regardless of the true underlying distribution of the data.

We use this assumption implicitly in a lot of statistical tests on linear data. For circular data, we can get a similar result, but we need a distribution that wraps around the circle. While there are other descriptive probability distributions have this property (e.g. Wrapped Normal or Wrapped Cauchy), the distribution that satisfies the equivalent central limit theorem in directional statistics is called the **von Mises** distribution.

4.0.1 The von Mises distribution

The probability density function (PDF) of the **von Mises distribution** is given by:

$$f(\theta \mid \mu, \kappa) = \frac{e^{\kappa \cos(\theta-\mu)}}{2\pi I_0(\kappa)}, \tag{9}$$

where:

- θ is the angle (in radians),
- μ is the **mean direction** (location parameter),
- κ is the **concentration parameter** ($\kappa \geq 0$),
- $I_0(\kappa)$ is the **modified Bessel function of the first kind** of order zero:

$$I_0(\kappa) = \frac{1}{\pi} \int_0^{\pi} e^{\kappa \cos\phi} d\phi$$

4.0.2 Maximum likelihood estimation

If we have some data, we might try to fit it to a von Mises distribution. For example, one model for a neuron's orientation tuning curve is by using a von Mises distribution. In order to fit the parameters μ, κ to the data, we can use standard maximum likelihood estimation to maximise the likelihood of the data given the parameters. We'll go through this in class in case you are unfamiliar!

4.0.3 How good is the fit? Some (loose) frequentist statistics

To test goodness of fit for directional data, we need a different set of statistical tests. Here I will list a couple of useful ones, which hopefully will be more intuitive with the information above:

- **Rayleigh test for uniformity**: parametric test that assumes a von Mises distribution for the data and tests the null hypothesis that the data is uniformly distributed; will fail when the assumption that the data is von Mises-distributed is poor, e.g. for multi-modal data.
- **Omnibus test for uniformity**: non-parametric test that checks percentage of samples in half-circle with least data-points vs half-circle with most – if these are close to equal, data is uniform – also fails for some multi-modal data
- **Rao's test for uniformity:** percentage test for the mean inter-sample interval – expect $\frac{2\pi}{N}$ for uniform data
- **Watson-Williams test:** something akin to a paired t-test for parametric data, again assuming a von Mises distribution (test if sample - sample means are the same)
- There also exist non-parametric paired tests, something like Kruskal-Wallis tests

5 References

1. Cremers and Klugkist, 2018, *One Direction? A Tutorial for Circular Data Analysis Using R With Examples in Cognitive Psychology*, Front. Psychol.
2. Ken Harris, 2022, *Circular statistics, Maximum likelihood, Local likelihood* , Neuroinformatics – UCL
3. T. Hotz and S. Huckemann, 2011, *Intrinsic Means on the Circle - Uniqueness, Locus and Asymptotics*, https://arxiv.org/abs/1108.2141
4. A. Brun et al., 2007, *Intrinsic and Extrinsic Means on the Circle – A Maximum Likelihood Interpretation*, IEEE Conference