@circular

A Matlab Class for Circular Statistics

Bart Krekelberg, The Salk Institute, Vision Center Laboratory
bart@salk.edu

If you've ever tried to determine an average angle, or even the distance between two angles, you will have noticed that circular data should be treated differently from normal linear data. This is true when you determine averages, but even more so when more advanced statistical techniques are used to analyse the data. The normal ANOVAs, T-tests, etc. do not apply to data that vary along a circular dimension.

An easy example shows this. Suppose you had one sample of directions with the values [1 2 1 2 4 2] degrees and another [359 355 359 358 355]. If you'd do a T-test to determine whether the means of the samples are different, the answer would be yes. However, when you realize that 359 degrees equals –1 degree, it is clear that these samples are not different. The various statistical tests implemented in this Matlab class take the circular nature or periodicity of the variables into account.

The @circular class stores, analyses and graphs circular data. The statistical tests are all based on (Batschelet, 1981). This excellent book explains the concepts behind the tests and gives clear examples. Sadly, it is currently out of print (August, 2001). An alternative book is (Fisher, 1996).

I implemented those statistical analyses from the book that seemed most useful to me, if you have specific demands that are not covered in this Matlab class, you may want to go back to the source and implement them yourself. I could include them in future releases. Currently, there are some descriptive statistics, tests for directedness, two- and k-sample tests and bivariate tests. See Functions for a complete list. Whenever possible, I tested implementations of tests with the examples Batschelet gives in his book. You can run these tests yourself by running the 'test' script in the docs directory of this release.

Katherine H. kindly entered the tables given in the appendices of Batschelet's book. Each appendix has its own m-file in the private directory of the class. Not all details of every table have been entered. For instance, p-levels above 0.1 have generally been set to 1 and p-levels below 0.001 to 0.001.

Even though I have tested all analyses on a few problems and I have checked that the tables of statistics agree with those in Batschelet's book, errors or bugs may have crept in. You use this toolbox at your own risk.

Installation

Copy the @circular directory to a folder that is on your Matlab search path. Do *not* add @circular itself or any of its subdirectories to the search path.

To get familiar with the functions in this toolbox, type
'what circular'
on the Matlab command line to get a list of methods that are associated with this class.
To get help on a particular function, use the Matlab help function. For instance,
'help mstd'
will give you a short description of the function that determines mean and standard deviation of circular data. For easy reference, the complete collection of these help-files is given as the 'reference.pdf ' portable document format file.

Some of the statistical tests rely on cumulative probability distributions that are available in the Matlab statistics toolbox : The F- and Chi-squared distribution. Alos, the ranksum.m script from the toolbox is used. The 'fminbnd' Matlab 6 function (is it?) is used by some tests as well. You can find out if you have access to the required functions by going into the/@circular/docs directory and running the test script. If this runs without errors, all necessary functions are there.

Getting Started

Let's assume you determined the preferred direction of 10 MT neurons.
data = [15 20 10 40 130 120 270 230 100 300];
To analyse these data, first create a circular data object:
c= circular(data,'deg')
The second argument of the circular() call specifies that the angles in the data vector are specified in degrees, not radians. If you forgot this, the values would be interpreted as radians.
The first thing you can do is to look at the distribution of your data points
plot(c)
will show a polar plot of the data in the circular data object, more interestingly
plot(c,'mean')
will also show the mean direction in this set. In fact, this plot also shows the result of a statistical analysis of directedness: the p value near the tip of the red arrow shows the significance level at which you can claim that there is a directional bias in this data set. This significance level is based on the Rayleigh test and could have been obtained by typing:
[phi,r,meanC,conf,means,p] = mstd(c)
which also gives you access to the mean angle, standard deviation and confidence intervals.

Functions

This section gives a broad overview of the functions that are available. For more detailed information on how to use these functions, see the reference manual or type 'help functionname' on the Matlab command prompt. For a more detailed description of the reason for choosing one test over another, see (Batschelet, 1981).

Setting up the Object

To construct an object you use the circular function. For univariate data (in which only the angle is relevant, not the length of the vector), use:

c = circular(data)

for a data-vector that contains angles in radians, or:

c =circular(data,’deg’)

for a data-vector with angles in degrees. All angles are automatically warped to 0-2*pi or 0-360 degrees.

If your data points not only have a direction, but also a length (say the firing rate of a neuron when stimulated with motion in a particular direction, or the number of observations in a particular direction), specify direction and length as separate vectors:

C =circular(angles,lengths)

For angles in radians or

C =circular(angles,length,’deg’)

For angles in degrees.

Data that was grouped or binned must be treated in a slightly different way than directions that were recorded continuously. For instance if you tested a neuron with 8 directions of motion, specify:

c.groups = 8;

By default, the number of groups is infinite which means that the analysis assumes that any angle could have been measured in principle (i.e. no binning).

Also, if you are not interested in direction but only in orientation, tests for non-randomness, but also the calculation of the mean orientation should be adapted. To use the correct calculations for such axial data, set:

c.axial = 1

By default the axial property is 0. Note that setting this property changes the way in which some of the statistical tests are performed (Rao for instance). I have not had the opportunity to test this rigorously. Hence, be careful when you need to describe axial data.

Descriptive Statistics

Mean	mstd
Standard Deviation	mstd
Confidence Intervals for the Mean	mstd
Median	median
Skew	skew
Kurtosis	kurtosis
Von Mises Fit (~circular Gaussian)	vonmises

Bivariate Descriptive Statistics

Mean length in a range of angles	meanr

Data Manipulation

Adding data	c1 +c2 where both are circular data objects will combine the data stored in c1 and c2
Getting the raw data	To extract the data from the object, there are 4 functions: deg(c) Returns the data in degrees rad(c) Returns the data in radians x(c) returns the x-components of the data y(c) returns the y-components of the data
Subtracting Values	c1-c2 will vector subtract the elements in c2 from the elements in c1. This will only work when c1 and c2 have the same number of elements or c2 is a single vector in which case this vector is subtracted from each c1 element. Note that – is not the complement of +.
Distance	The distance(c1,c2) function returns the angular distance between the elements of c1 and those of c2. Note that this is a value between 0 and 180 degrees (or 0 and pi), that is *not* a circular variable itself. You can do normal linear statistics on it.
Sum	The function sum(c) returns the sum or resultant vector of this dataset as a circular data object.
Normalize	Set all vector lengths to 1.

One Sample Tests

Unimodal data only	rayleigh
Unimodal data plus hypothesis about the peak	vtest
Multimodal data	rao
Grouped data, compare to model	chisq
Non-grouped data, compare to model (~Kolmogorov-Smirnov)	kuipers
Small sample size, compare to model, non grouped	Watson

Two Sample Tests

Underlying distribution is vonMises.	watsonwilliams
No grouping, ties are broken randomly.	ranksum
Test whether the two samples have the same dispersion	Ftest

Correlation (All give significance values, but are not [-1,1]!

Correlation Coeffeciient	corrcoef
Rank Order Correlation	rankcorr
Vector Correlation	veccorr

k-Sample Tests

Underlying distribution is vonMises. No Posthoc	Watsonwilliams.

Bivariate Tests
For data vectors that not only have a direction, but also a length (for instance the firing rate of a neuron when stimulated with motion in a particular direction), you may want to test whether the average vector is different from the origin, or whether the average vector in one sample is different from the average in another sample.

Parametric	Hotelling
Non-Parametric, non-grouped data.	Mardia

References

Batschelet E (1981) Circular Statistics in Biology. London: Academic Press.
Fisher NI (1996) Statistical Analysis of Circular Data. Cambridge, UK: Cambridge University Press.

Feedback

This is a first release and will contain bugs. Use this toolbox at your own risk, and please help me to remove the bugs from future releases by sending bug-reports to bart@salk.edu