Circular statistics:

Introduction & background:

The main issue in circular statistics is that quantities of interest (e.g., angles, time, date) "wrap" around and come back to the beginning. For example, here's a circle outlined in degrees:

Figure on board

Suppose now we observe the direction that a particular type of bird flies off to after we release it. We let N = 0º, and find the bird flies at 5º.

We repeat this with 20 birds and get the following results (all in degrees):

3 344 1 24 11 350 16 337 343 15 355 349 19 6 354 5 22 357 12 355

Now let's calculate the the "average" direction our birds take:

$$\bar{y} = 163.9$$

Notice that most of our birds flew in a northerly direction, but the "average" directions is close to south. Obviously the usual way of calculating means and such will not work with circular statistics as they give us garbage.

We need a different way of thinking about averages and such. But first, let's note the following.

1) In circular statistics 0º are often considered to be vertical. This comes from navigation where 0º is considered north, 90º is east (so we move clockwise), and so on. As such, some authors (including your text) consider the *x*-axis to be the vertical axis when doing circular statistics.

2) Mathematicians consider 0º to lie exactly on the (usual) *x*-axis, and the we proceed counterclockwise such that 90º is straight up along the *y*-axis, etc.

Figure on board

Whichever system you use, it all works, but you need to be consistent or you can easily get yourself confused. We'll stick with (1) since a lot of things biologists are interested in move clockwise from vertical (degrees, hours, etc.).

3) About units. Mathematicians (**and R!**) generally use radians when talking about angles and such. Most other folks use degrees. To convert use the following formulas:

1 radian = 57.2958 degrees

1 degree = 0.0174533 radians

So, for example, if you want to calculate the sin(30º) in R, you would do:

```
sin(30*0.0174533)
```

And you would get back 0.5.

4) About units part II. Circular statistics doesn't just deal with directions. You might be interested in the hours of a day or dates of the year. All of these can be considered to be on a circle as you "return" to the starting point after going around the "circle" once.

Figure on board

For example (using 24 hour time, which most of the rest of the world and the U.S. military uses), after 23:59 we get to 24:00, which is the same as 0:00 and time starts over (the a.m./p.m. system is more confusing than anything else here).

Since some of the same methods apply, we need to convert circular units which are not measured in degrees to degrees (so, for example, 12:00 would be 180 degrees). Here's how you convert, with an example (see also your text on pp. 606 - 607):

$$a = \frac{(360)(X)}{k}$$

a = angle you want

X = unit of the other scale (e.g., hours, day, etc.)

k = time (or other) units to make a full circle (e.g., 24 for hours, or 365 for days)

Example: convert 07:37 to degrees

First notice that 37 minutes = 0.617 hours in decimal notation.

Then do:

$a = 360^{\circ}\ (07.617)/24 = 114.255^{\circ}$

So 7:37 would be equivalent to 114.255° on a circle

See the other examples in your text on p. 607.

Averages and a (very!) brief review of trigonometry:

As pointed out above, we do need to be able to calculate averages for angles and such. If we can't just take degrees and do this, we need to use a different approach.

What we do is to take all our degrees and convert them into points on the plane (using a unit circle). So, for instance, if we have an angle of 30 degrees, we look on our unit circle and get the coordinates of the point that lies on the circle at 30 degrees.

Figure on board

To do this, we need to remember our trigonometry (don't worry, we won't get very in-depth). We'll need to remember three functions: sine, cosine, and arctangent (your book makes this a little too complicated by trying to avoid arctangent).

sine = length of opposite side / length of hypotenuse

cosine = length of adjacent side / length of hypotenuse

arctangent = inverse of the tangent function

(where **tangent** = length of opposite side / length of adjacent side)

The sine gives us the *Y*-coordinate on the unit circle, and cosine gives us the *X*-coordinate on the unit circle (remember comments 1 & 2 above, as the *X* and *Y* axes may be reversed).

We'll hold off on explaining why we need the arctangent for a moment. So how do we proceed? With our bird example we convert all our angles into sines and cosines (rounded a bit below):

angles:	3	344	1	24	11	350	16	337	343	15	355	349
	19	6	354	5	22	357	12	355				
sines:	0.052	-0.276	0.017	0.407	0.191	-0.174	0.276	-0.391	-0.292	0.259	-0.087	-0.191
	0.326	0.105	-0.105	0.087	0.375	-0.052	0.208	-0.087				
cosines:	0.999	0.961	1.000	0.914	0.982	0.985	0.961	0.921	0.956	0.966	0.996	0.982
	0.946	0.995	0.995	0.996	0.927	0.999	0.978	0.996				

So now each angle has a *X*-coordinate given by the cosine, and a *Y*-coordinate given by the sine.

We now add all the sines (sum up all the sines) and cosines:

$$\sum_{i=1}^{n} \sin(a) = 0.6472117 \qquad \sum_{i=1}^{n} \cos(a) = 1.905612$$

We can now proceed one of two ways; the text uses the following method (see comments below for the other method):

1) Get the average sine and cosine and call these Y and X. To do this we just divide the above sums by $n = 20$:

Y = average sine = 0.03236059 X = average cosine = 0.9726228

2) Get the distance of this point from the origin:

$$r = \sqrt{X^2 + Y^2} = \sqrt{0.9726228 + 0.03236059} = 0.9731609$$

3) This gives us the following quantities:

$$\sin(\bar{a}) = \frac{Y}{r} = \frac{0.03236059}{0.9731609} = 0.03325307$$

$$\cos(\bar{a}) = \frac{X}{r} = \frac{0.9726228}{0.9731609} = 0.999447$$

4) Now the book gets a little vague. It says we get the angle with the above sine and cosine.

To do this we remember the tangent:

$$\tan(a) = \frac{\sin(a)}{\cos(a)}$$

So we have:

$$\tan(a) = \frac{0.03325307}{0.999447} = 0.03327147$$

Finally, we remember that to invert the tangent, we have the arctangent:

$a = \arctan(\tan(a))$

So if we want the angle, we just ask for the arctangent of 0.03327147:

arctan(0.03327147) = 1.906

And we can finally say that the average direction of our birds was 1.906º.

Some comments on this method:

Notice that technically you can skip steps 2 and 3 and just use the average sine and average cosine to calculate the tangent in step 4 (you're dividing both the quantities in step 3 by r, which just cancels out in step 4).

You can even skip step 1 and just use the sums directly in the tangent function and use the following shortcut method:

$$\tan(a) = \frac{0.6472117}{19.45246} = 0.03327147$$ which is the same as above.

So why go through all the extra steps?

As it turns out, the quantity r is important as it gives us the "length" of the mean angle. If most birds fly in the same direction, r will be bigger than if they fly in many different directions. It is possible for $r = 0$, in which case there is no mean angle.

In other words, r is an important quantity, and since we should calculate it in any case, the above steps make as much sense as anything else.

About the arctangent.....(**and the real final answer**)

The arctangent function can not distinguish which "quadrant" the angle is in. Notice, for example, that if our sine and cosine had both been negative, we would have gotten the same answer for arctan(0.03327147).

In this case, 1.906º would obviously be incorrect - we would have needed to add 180º and get 181.906º.

So now we need the following rules to adjust the answer from step 4:

If both sine + and cosine +	angle computation is correct as is (don't adjust)
If sine + and cosine -	Angle = 180 - average angle you calculated
If sin - and cosine -	Angle = 180 + average angle you calculated
If sin - and cosine +	Angle = 360 - average angle you calculated

If you use R to do this there is a built in function (arctan2) which does all of this automatically (see the last bit of the code below).

Here is the R code I used for this example:

(Remember R uses radians, not degrees, so make sure you do the conversions correctly).

```
# read in the data
deg <- scan(nlines = 1)
3 344   1  24  11 350  16 337 343  15 355 349  19   6 354   5  22 357  12 355

# convert to radians, then get sines
sindeg <- sin(deg*0.0174533)

# if you want to see the output
round(sindeg,3)

# convert to radians, then get cosines
cosdeg <- cos(deg*0.0174533)

# if you want to see the output
round(cosdeg,3)

# get the sum of the sines and cosines:
sum(sindeg)
sum(cosdeg)
```

```
# get the averages of the sines and cosines
ms = mean(sindeg)
mc = mean(cosdeg)


# calculate the mean vector length (radius):
r2 = sqrt(mean(sindeg)^2 + mean(cosdeg)^2)


# get the tangen (again, you could just do this with the sums (see notes):
tandeg <- (ms/r2)/(mc/r2)

# finally, get the arctangent and convert to degrees
atan(tandeg)*57.2958

# and don't forget to check what quadrant your in and adjust your degrees as needed
(see notes)


# OR ALL IN ONE STEP ONCE YOU HAVE YOUR SINES AND COSINES:
atan2(sum(sindeg),sum(cosdeg))*57.2958
```

As should be obvious, directional statistics are not straight forward. Before we end, we should make a few comments about other topics in directional statistics:

There are hypothesis tests available to test:

> That there is no mean direction (this relies on the magnitude of *r*)
>
> That the directions of the angles have a uniform distribution (i.e., there isn't a “favored” direction).
>
> That two sets of data have directions that are not different
>
> > (This is similar to the MWU test).
>
> As well as others.

If you really need this, there is a package (and book) available for R that go into a lot more detail on directional statistics.