

# Lecture 9:

## Statistics on a sphere

- review of getting a paleomagnetic vector
- A brief review of "Gaussian" statistics
- Sources of scatter in pmag vectors and their distributions
- Fisher statistics: nuts & bolts
  - estimating mean directions
  - estimating concentration parameter
  - estimating confidence intervals
- Applications

We developed these equations for the transformation of coordinates:

$$x'_i = a_{ij}x_j$$

or

$$x'_1 = a_{11}x_1 + a_{12}x_2 + a_{13}x_3,$$

$$x'_2 = a_{21}x_1 + a_{22}x_2 + a_{23}x_3,$$

$$x'_3 = a_{31}x_1 + a_{32}x_2 + a_{33}x_3$$

$a_{ij}$  here is a “tensor”

a 3x3 tensor transforms one vector into another

$$A = a_{ij}$$

There exists vectors called “eigenvectors”,  $\lambda$ , whereby

$$Ax = \lambda x$$

To find these, we use the equation:

$$(\mathbf{A} - \lambda \mathbf{1})\mathbf{p} = \mathbf{0}$$

The only way for this to have a non-trivial solution is for:

$$\det[\mathbf{A} - \lambda \mathbf{1}] = 0$$

$$\det[\mathbf{A} - \lambda \mathbf{1}] = 0$$

Lots of math gives you three vectors that satisfy the equation, see:

<http://www.ce.berkeley.edu/~sanjay/ce23/mse211/eigenv1.pdf>

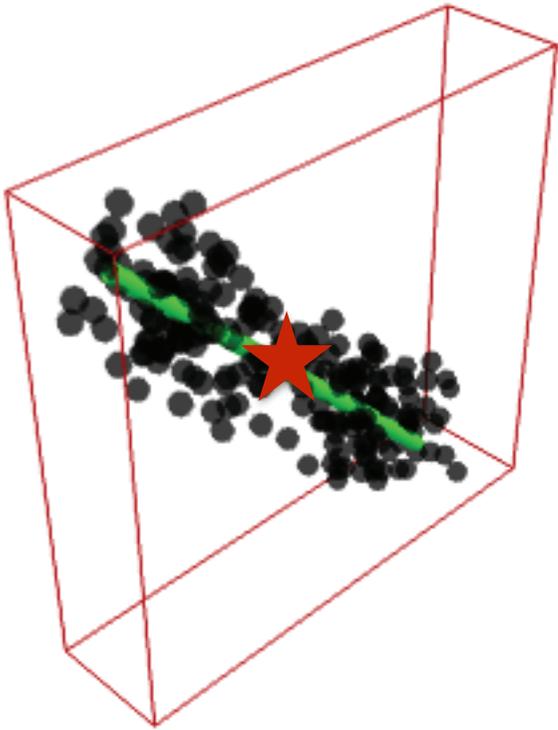
The unit vectors along these three directions are called “eigenvectors” and the lengths are the “eigenvalues”

You can get these using functions in numpy, so you don't have to do it yourself.

# Uses of "eigenparameters" in Earth Science

- Get the 'best-fit' direction through a bunch of 3-D data points (e.g., demagnetization data)
- Best-fit plane, say the fault surface through a bunch of earthquakes
- Orientations of stress or strain tensors
- Preferred orientation of mineral grains or class
- Directions of anisotropy of, say, magnetic susceptibility.....

# Best-fit directions through points in 3D



- Convert D,I,M vector from each step to list of (N,E,D) components (Chapter 2)
- Calculate “center of mass” as average of each component (e.g.,  $\langle x \rangle$  from all selected points ( $X_i$ ))
- Subtract the center of mass from all your points (e.g.,  $x' = X_i - \langle x \rangle$ )

- review of getting a paleomagnetic vector
- A brief review of "Gaussian" statistics
- Sources of scatter in pmag vectors and their distributions
- Fisher statistics: nuts & bolts
  - estimating mean directions
  - estimating concentration parameter
  - estimating confidence intervals
- Applications

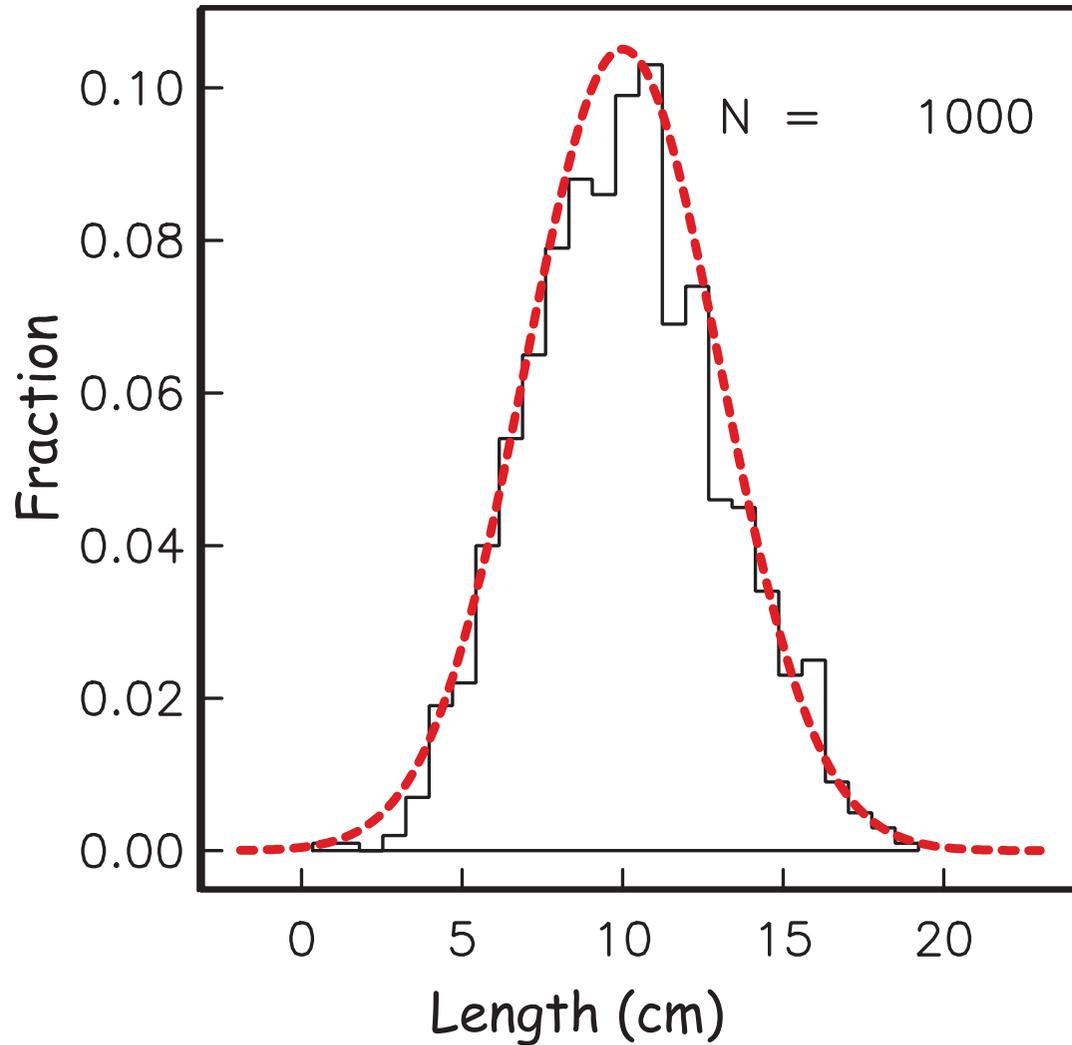
Now that we have a bunch of  
directions....

- How do we calculate a mean direction and an estimate of scatter?
- How do we tell if two directions are the same or different?

“Statistics on a sphere”  
but first, plain old statistics

# The “normal” distribution (Gaussian statistics)

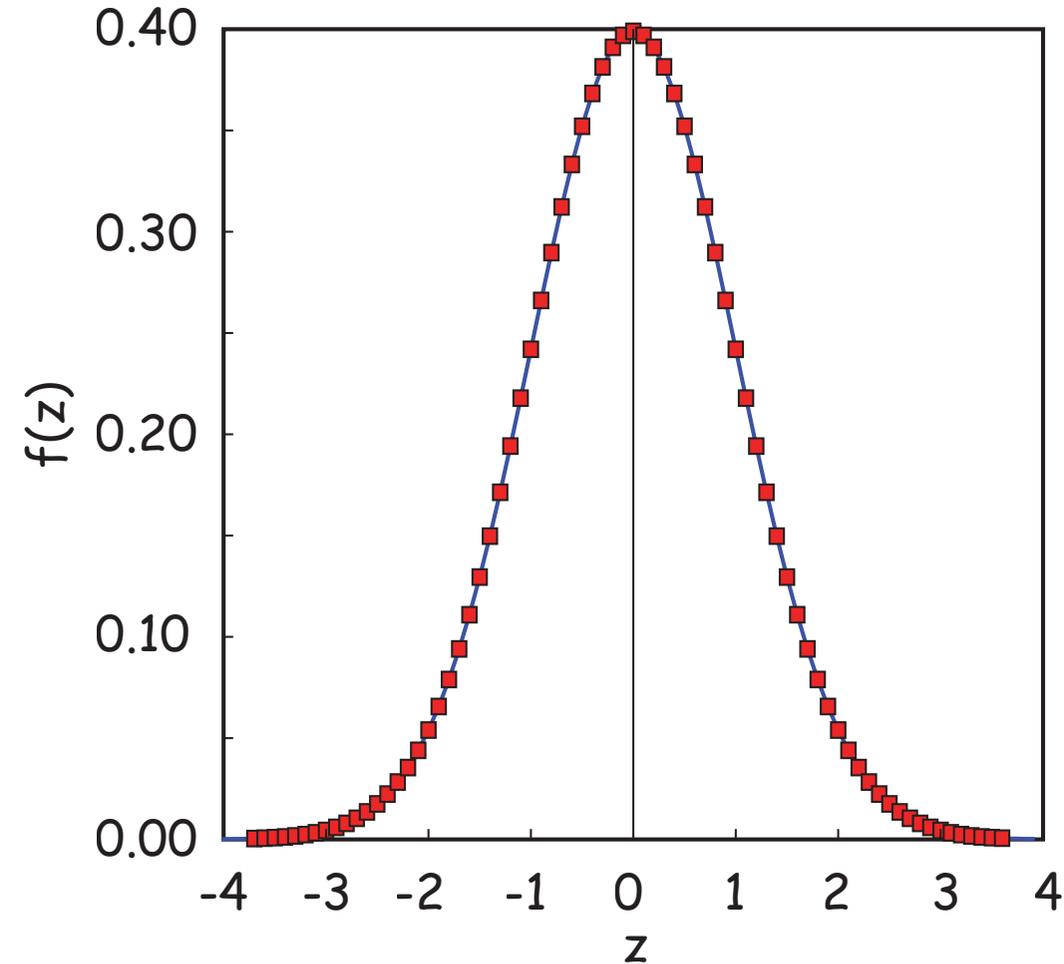
A bunch (1000) of estimates  
of the length of something 10 cm long



Need to estimate:  
most likely true length  
and  
“spread”

Special case where data fit distribution:

$$f(z) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(z - \mu)^2}{2\sigma^2}\right)$$



mean:

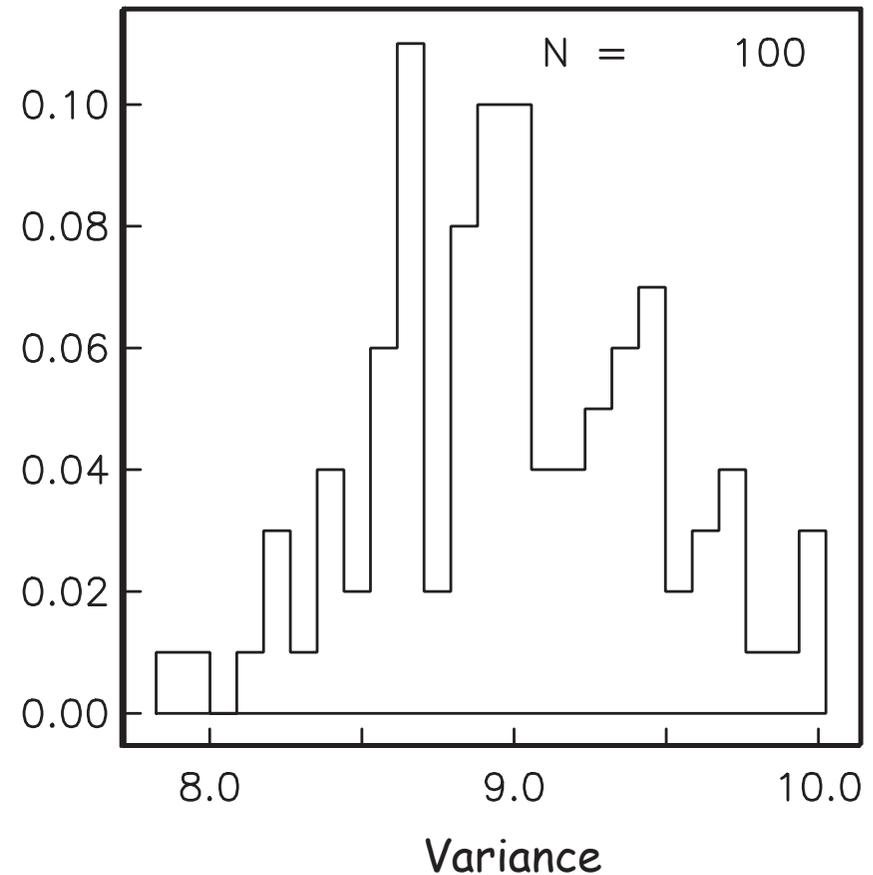
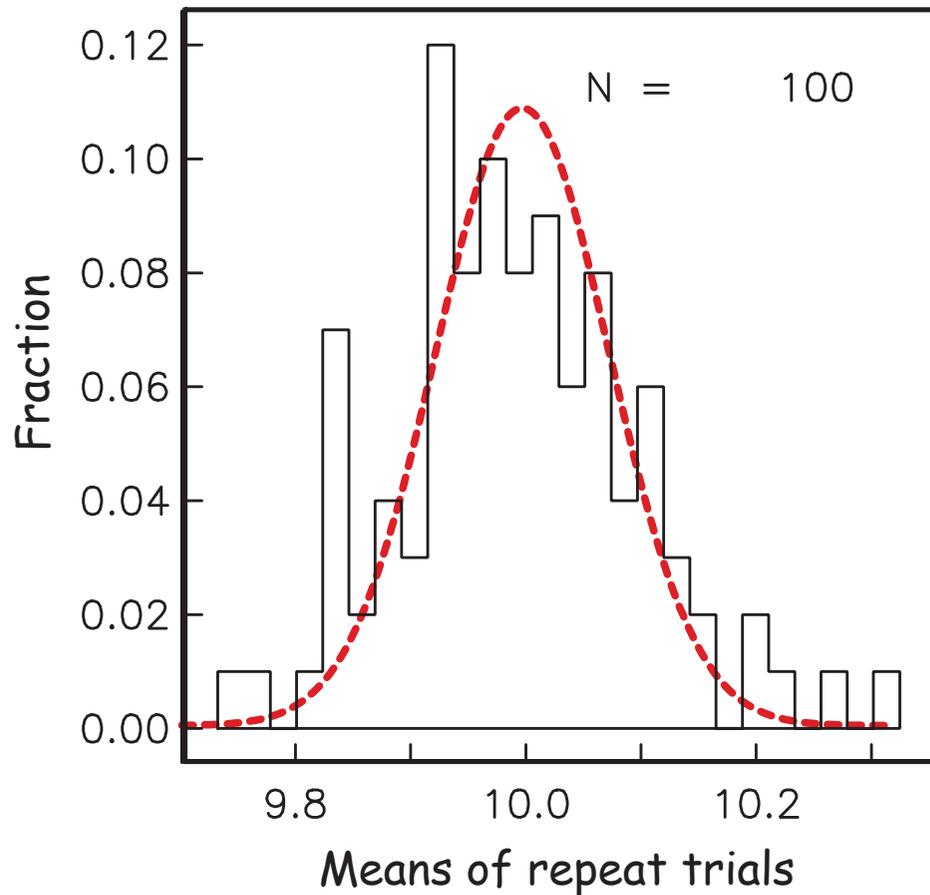
$$\mu \simeq \bar{z} = \frac{1}{N} \sum_{i=1}^N z_i$$

variance:

$$\sigma^2 \simeq s^2 = \frac{1}{N-1} \sum_{i=1}^N (z_i - \bar{z})^2$$

[ $\sigma$  is the “standard deviation”]

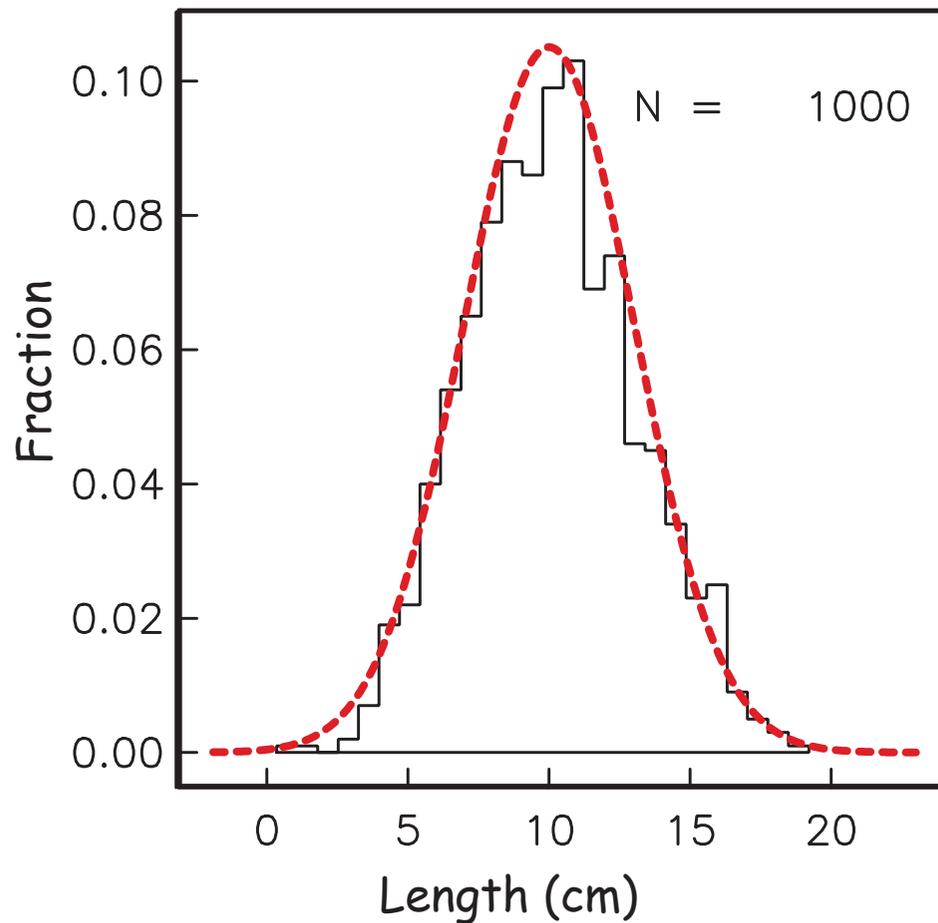
**NB: every time you make a set of 1000 measurements, you will get a slightly different set, hence different means and variances**



$\bar{z}$  mostly likely value of true length

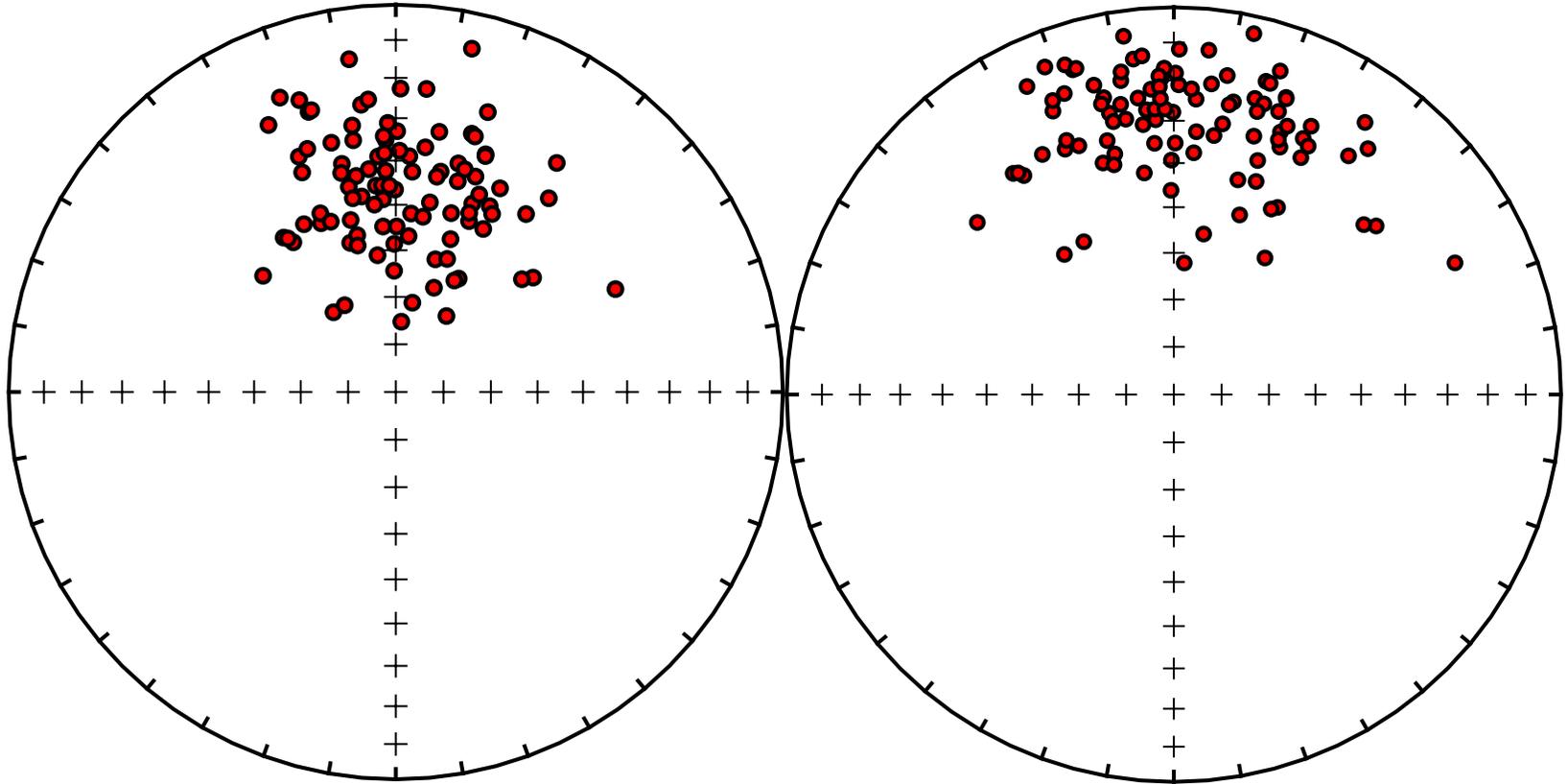
$1.96s$  bounds containing 95% of the measurements

$$s_m = \frac{s}{\sqrt{N}} \quad \text{95\% confidence in the mean}$$



# Statistics on a sphere (Fisher statistics)

# Styles of scatter



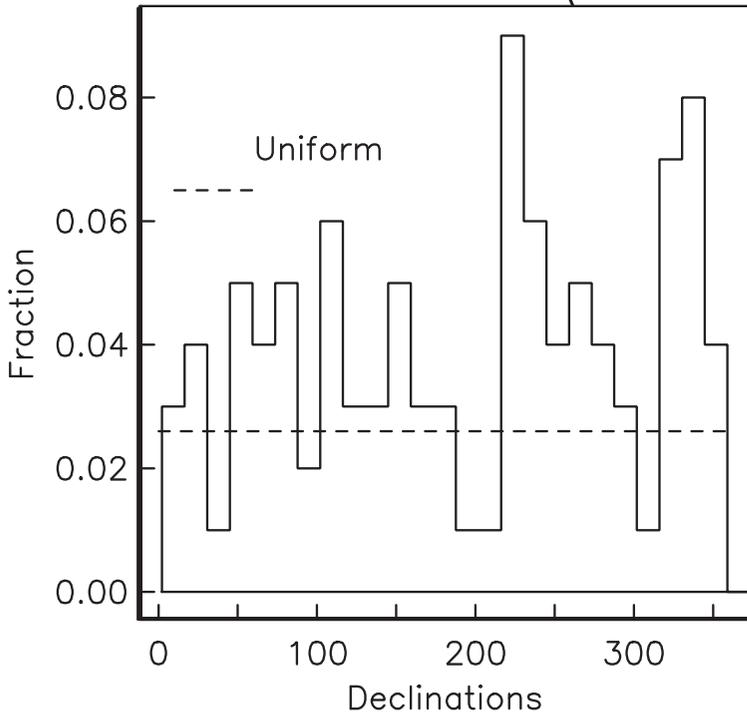
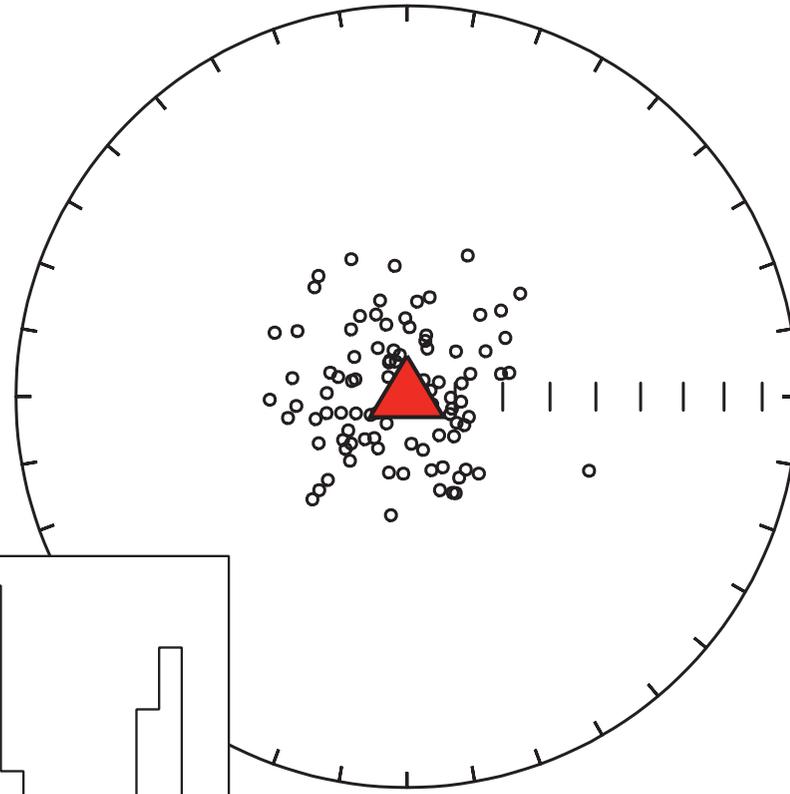
circularly symmetric

streaked

# Sources of scatter

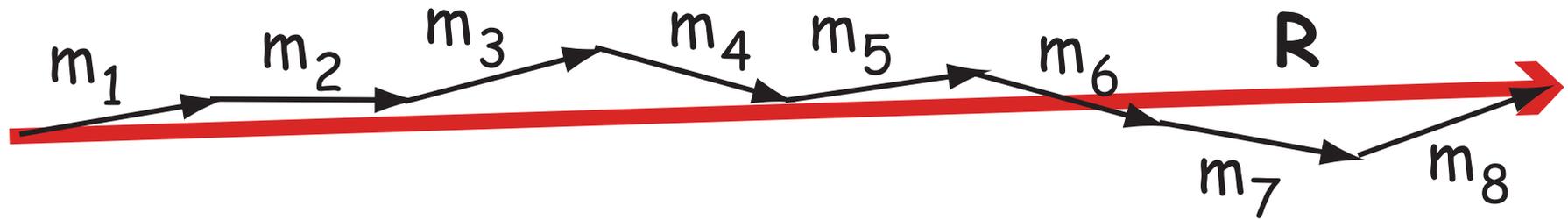
- uncertainty in the measurement caused by instrument noise (symmetric)
- uncertainties in sample orientation (symmetric?)
- uncertainty in the orientation of the sampled rock unit (streaked?)
- variations among samples in the degree of removal of a secondary component (streaked)
- uncertainty caused by the process of magnetization (streaked or symmetric)
- secular variation of the Earth's magnetic field (streaked)
- lightning strikes (streaked)

# Special case: Fisher distribution



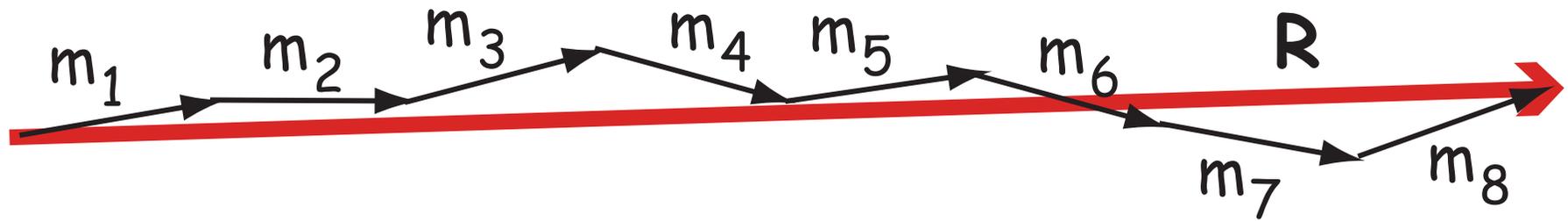
**Circularly symmetric about  
the mean direction**

# Estimating mean direction



- Assume each vector has unit length
- Do vector addition (see Appendix A!)
  - break each vector down to cartesian components ( $x_i$ ); (Chapter 2)
  - sum over each component
  - get resultant vector length R

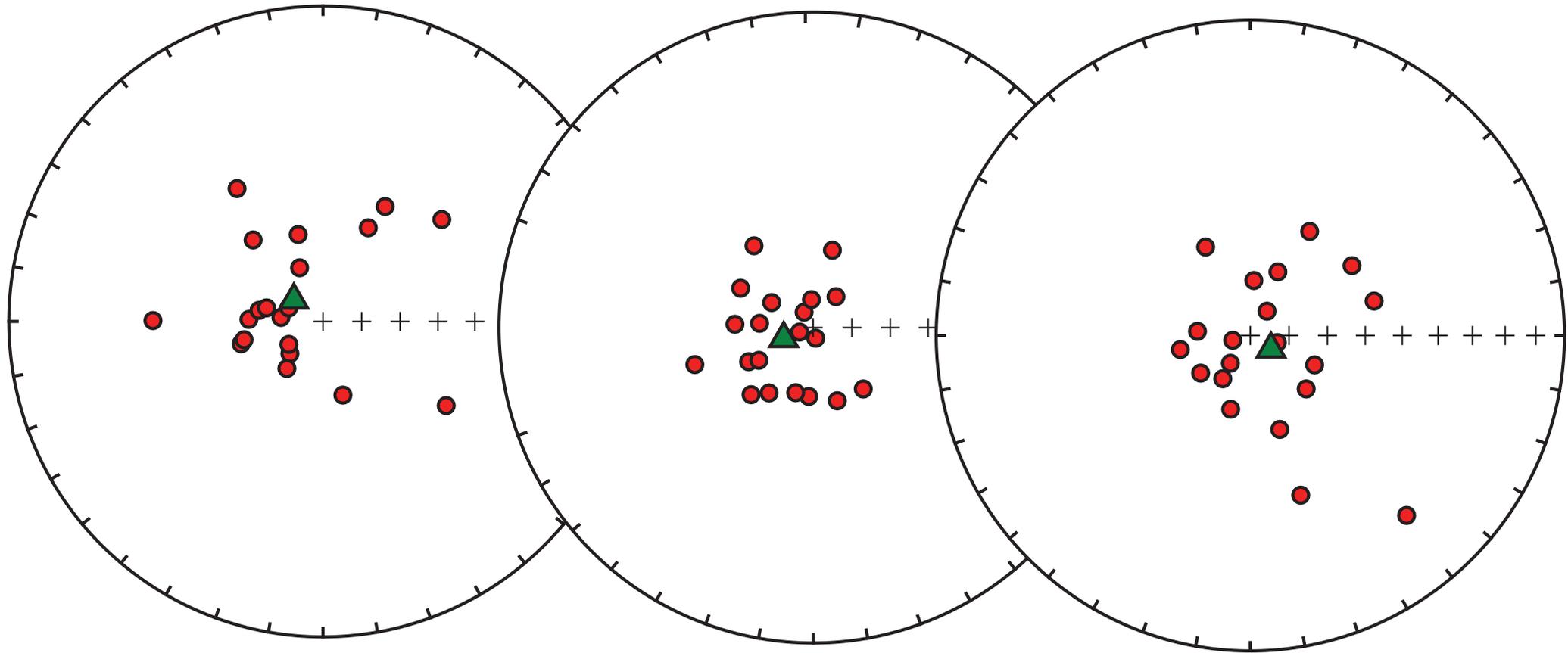
$$R^2 = \left( \sum_i x_{1i} \right)^2 + \left( \sum_i x_{2i} \right)^2 + \left( \sum_i x_{3i} \right)^2$$



- Components of mean direction are:

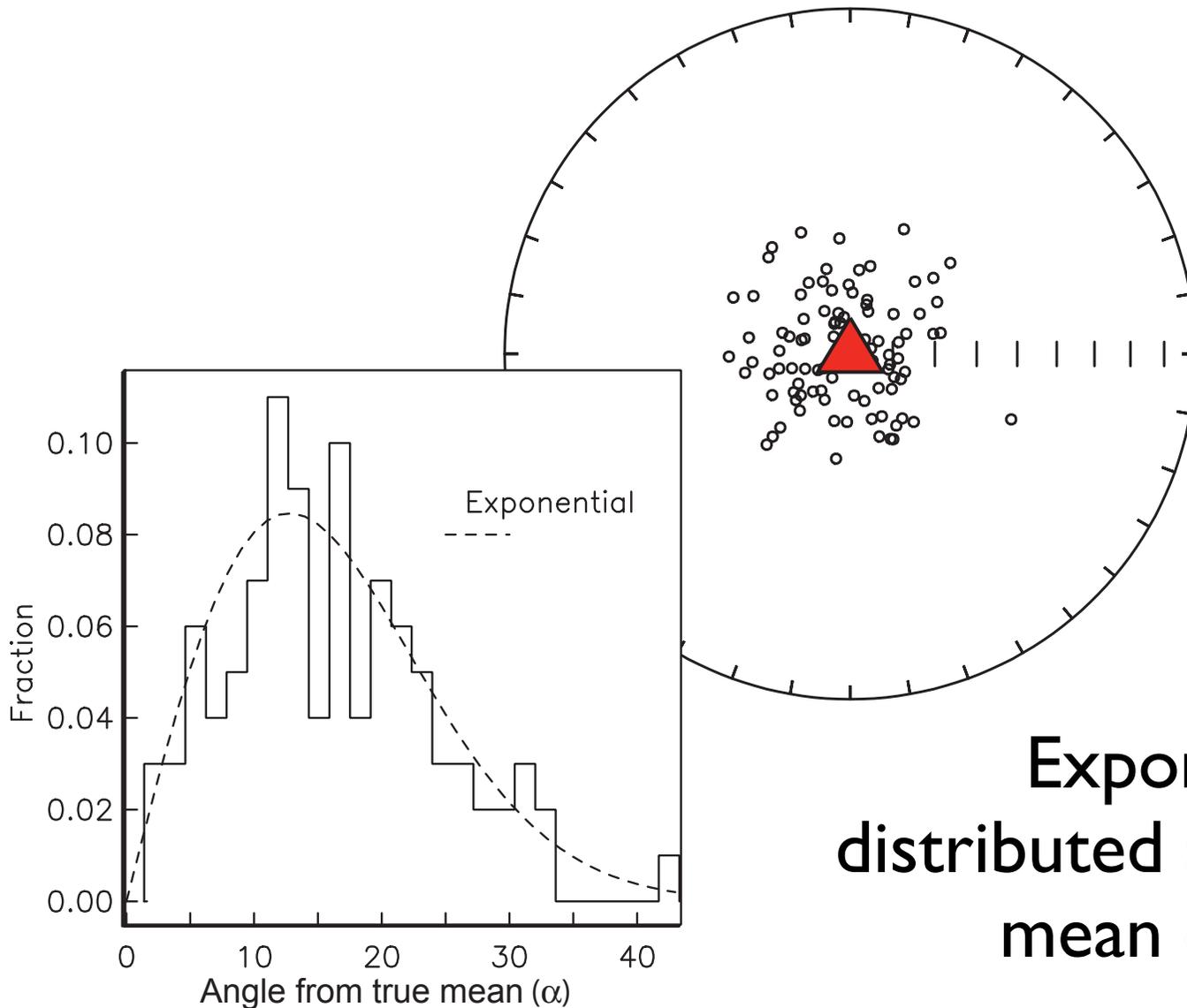
$$\bar{x}_1 = \frac{1}{R} \left( \sum_i x_{1i} \right); \quad \bar{x}_2 = \frac{1}{R} \left( \sum_i x_{2i} \right); \quad \bar{x}_3 = \frac{1}{R} \left( \sum_i x_{3i} \right)$$

- Convert these back to Dec, Inc, Int if desired (see Chapter 2)



same “true” mean (vertical)  
different estimated means!

# Scatter about the mean



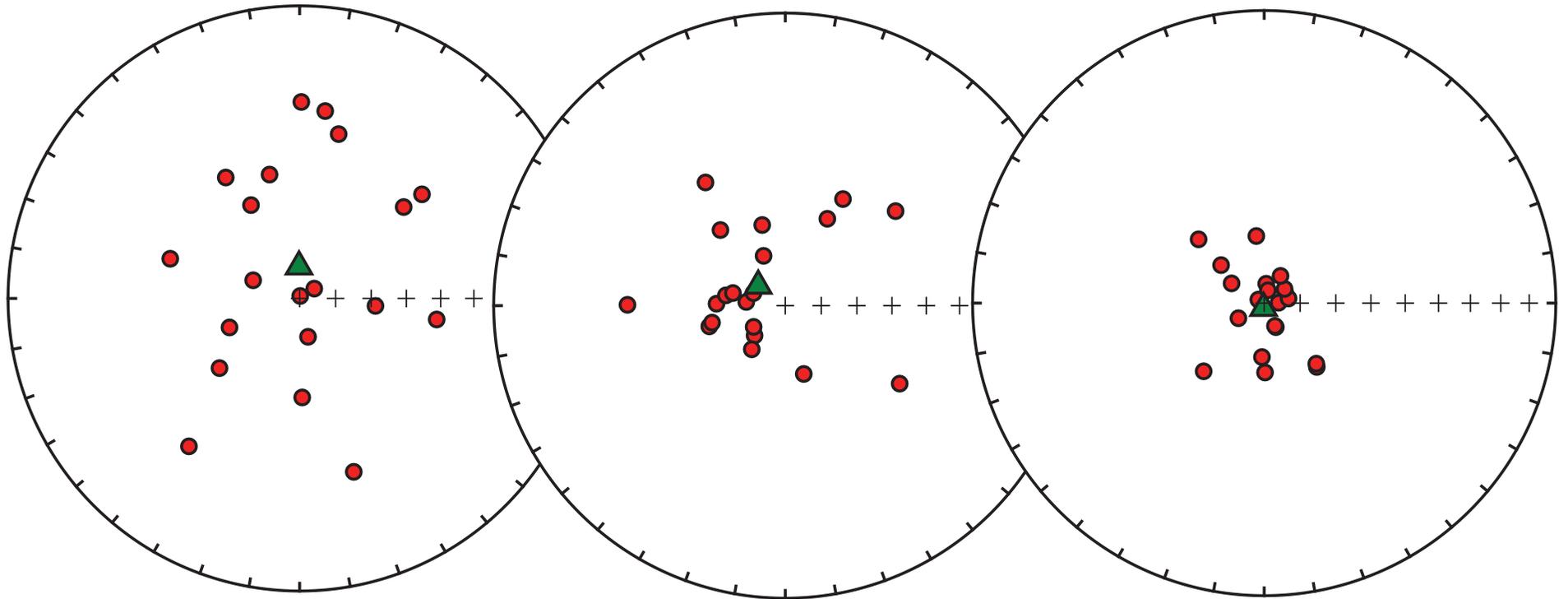
**Exponentially  
distributed away from the  
mean direction**

$$P_{dA}(\alpha) \sin \alpha = \frac{\kappa}{2\pi \sinh \kappa} \exp(\kappa \cos \alpha) \sin \alpha$$

$\kappa = 5$

$\kappa = 10$

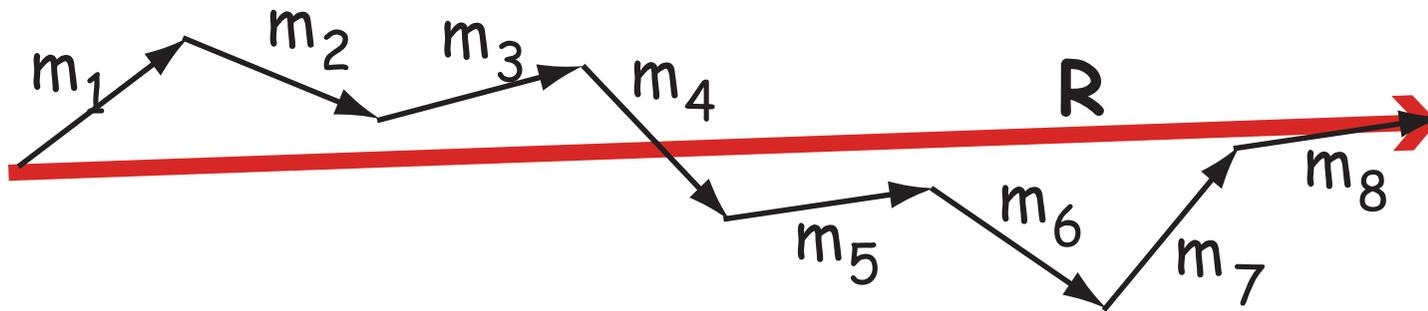
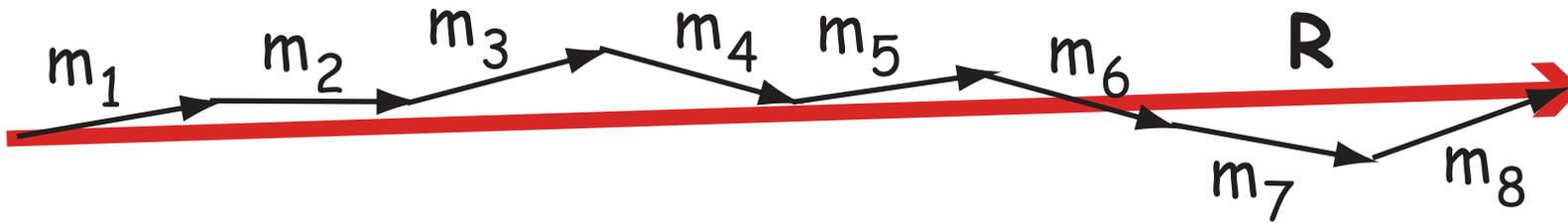
$\kappa = 50$



$\Rightarrow$  decreasing scatter  $\Rightarrow$

$\Rightarrow$  increasing concentration parameter  $\Rightarrow$

Estimating concentration  
parameter,  $\kappa$

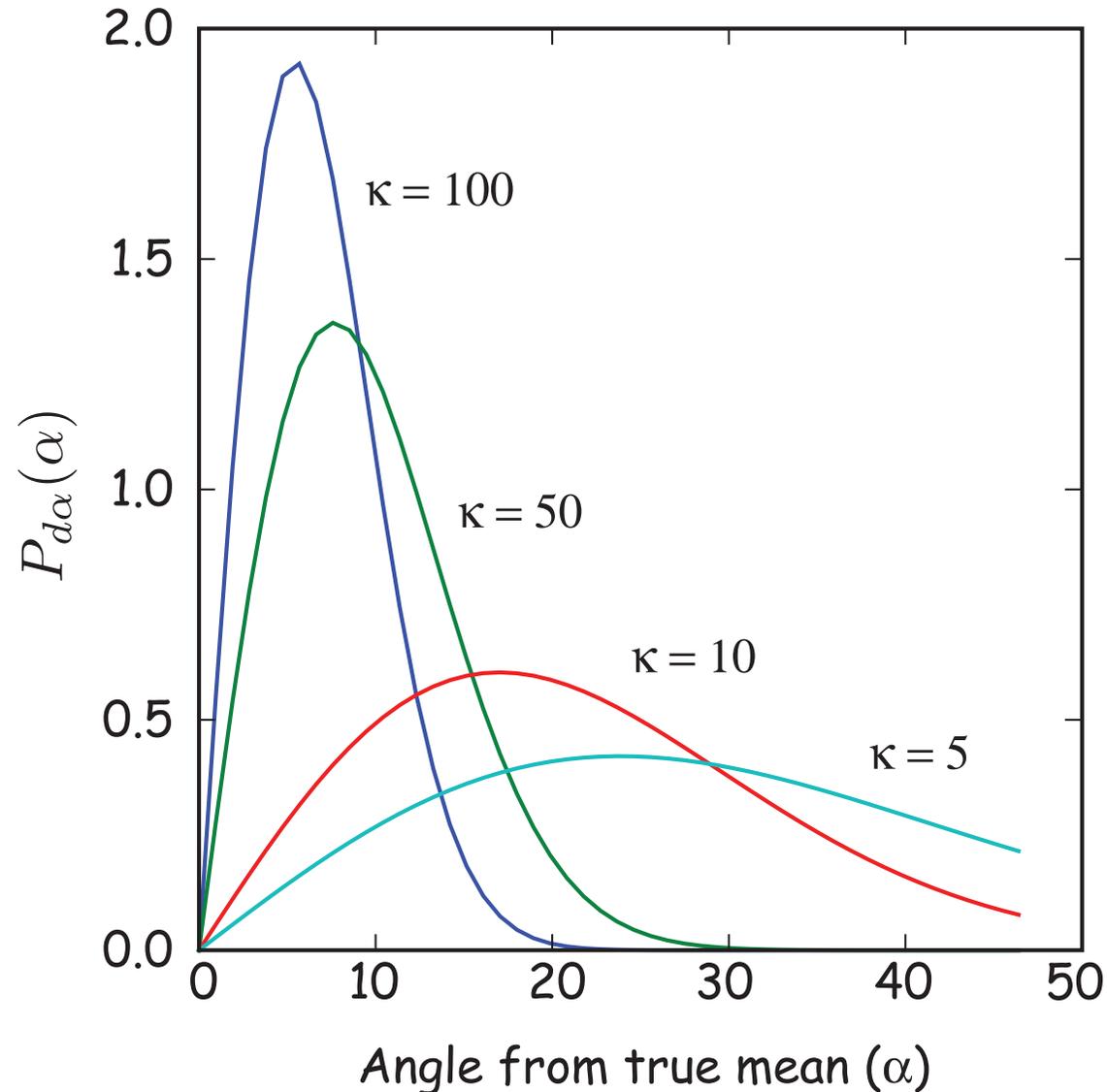


the higher the scatter, the smaller  $R$   
 so... concentration must be inversely related  
 to  $R$ :

$$\kappa \simeq k = \frac{N - 1}{N - R}$$

$k$  is an estimate of  $\kappa$

Probability of finding a direction with a given angle from the true mean: controlled by  $\kappa$  the “concentration parameter”



Useful statistics:

# Useful statistics

- 95% confidence in the mean direction
- Angular variance
- circular standard deviation

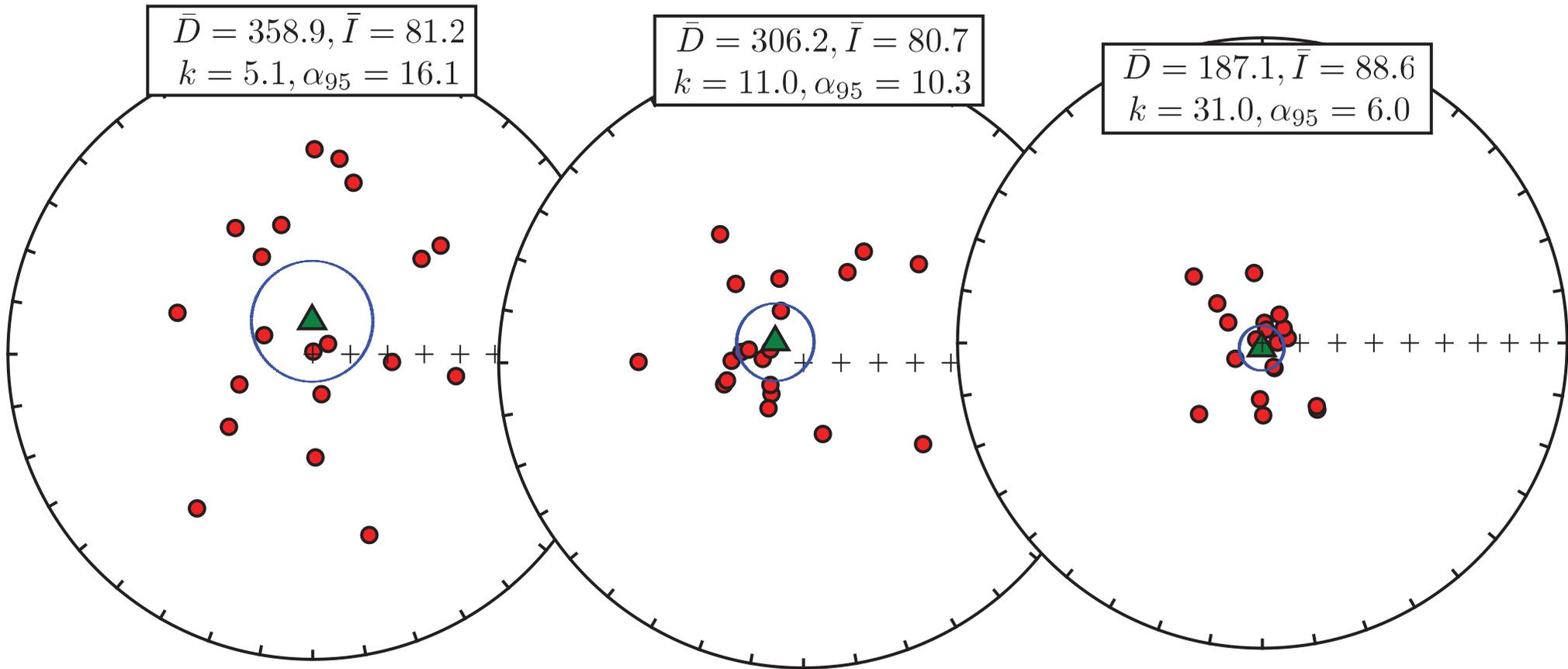
# Confidence in the mean

$$\alpha_{95} = \cos^{-1} \left[ 1 - \frac{N - R}{R} \left[ \left( \frac{1}{p} \right)^{\frac{1}{(N-1)}} - 1 \right] \right]$$

95% confident ( $p=0.05$ ) that true direction lies within this cone

often approximated by:  $\alpha'_{95} \simeq \frac{140}{\sqrt{kN}}$

but only true for  $k > 25$



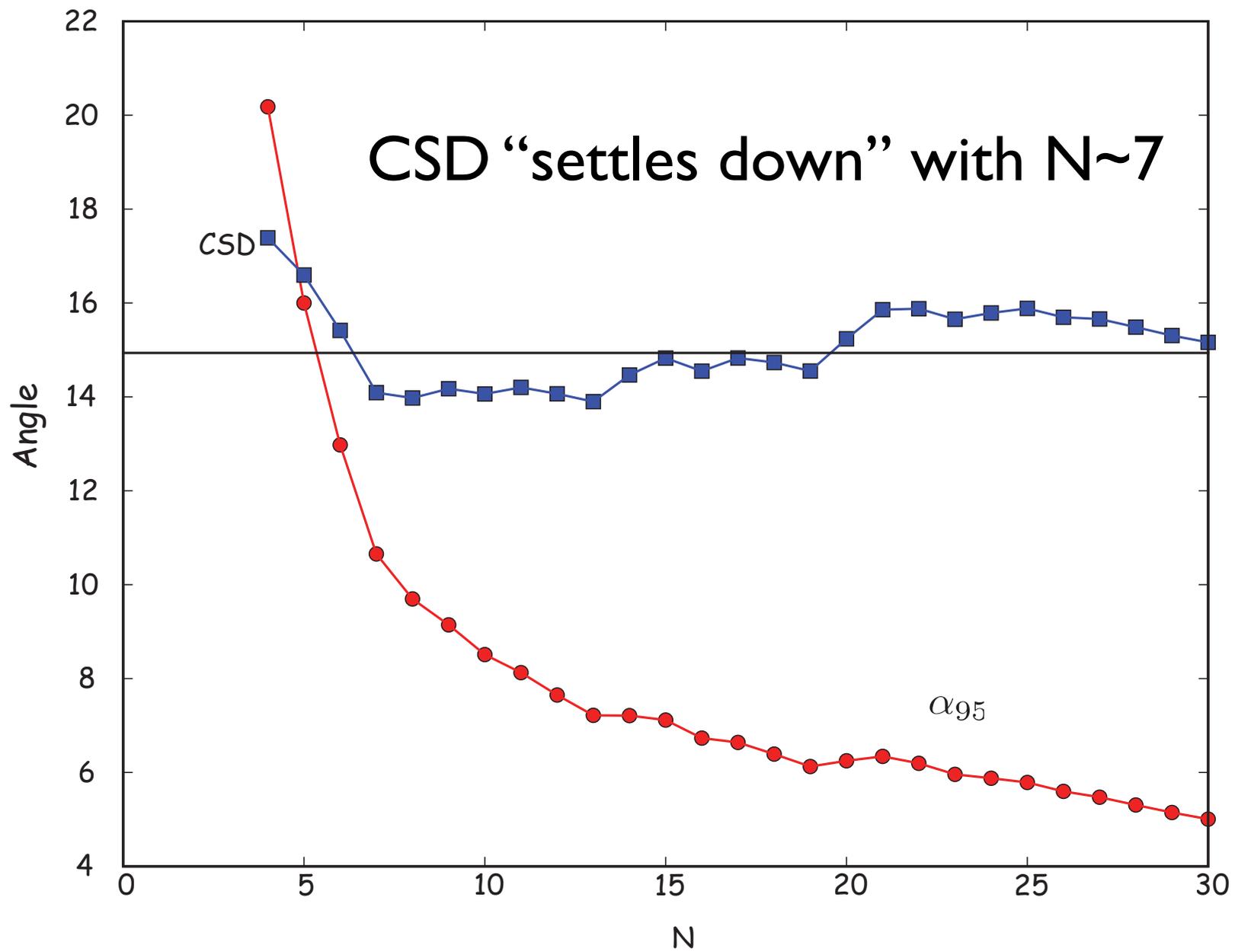
**NB: 5% chance that  $\alpha_{95}$  does not include vertical**

Angular variance:  $S^2 = \frac{1}{N-1} \sum_{i=1}^N \alpha_i^2$

Circular standard deviation (CSD):  $\approx \frac{81}{\sqrt{k}}$

# How many data points do you need?

- Make a fake directional dataset (drawn from Fisher distribution) with  $S=15$
- Take first 4 data points, calculate  $k$  and CSD
- Add a 5th and repeat
- Keep going until  $N=30$

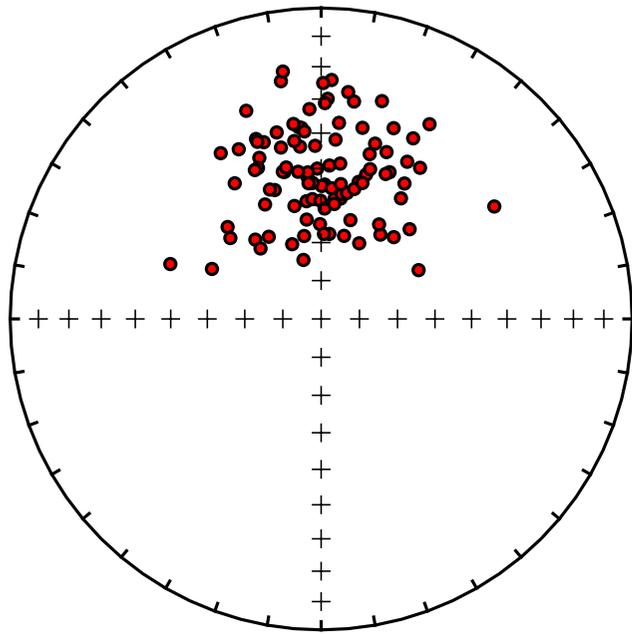


# More useful statistics

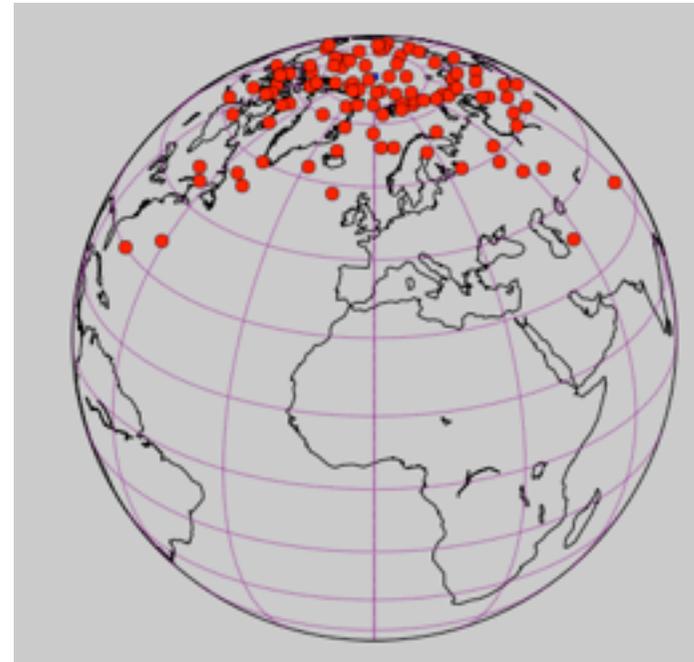
- what about confidence in VGPs?
- Test for randomness
- Are two directions significantly different from each other?
- How to combine best fit lines and planes
- What to do with inclination only data (see book)
- Test for fishiness (see book)

# Mapping of D,I to VGP

- Review Chapter 2 for how to do it



=>



Directions  
measured at  
latitude of 30

Not a circular distribution!

$$dm = \alpha_{95} \frac{\cos \lambda}{\cos \bar{I}}, \quad dp = \frac{1}{2} \alpha_{95} (1 + 3 \sin^2 \lambda)$$

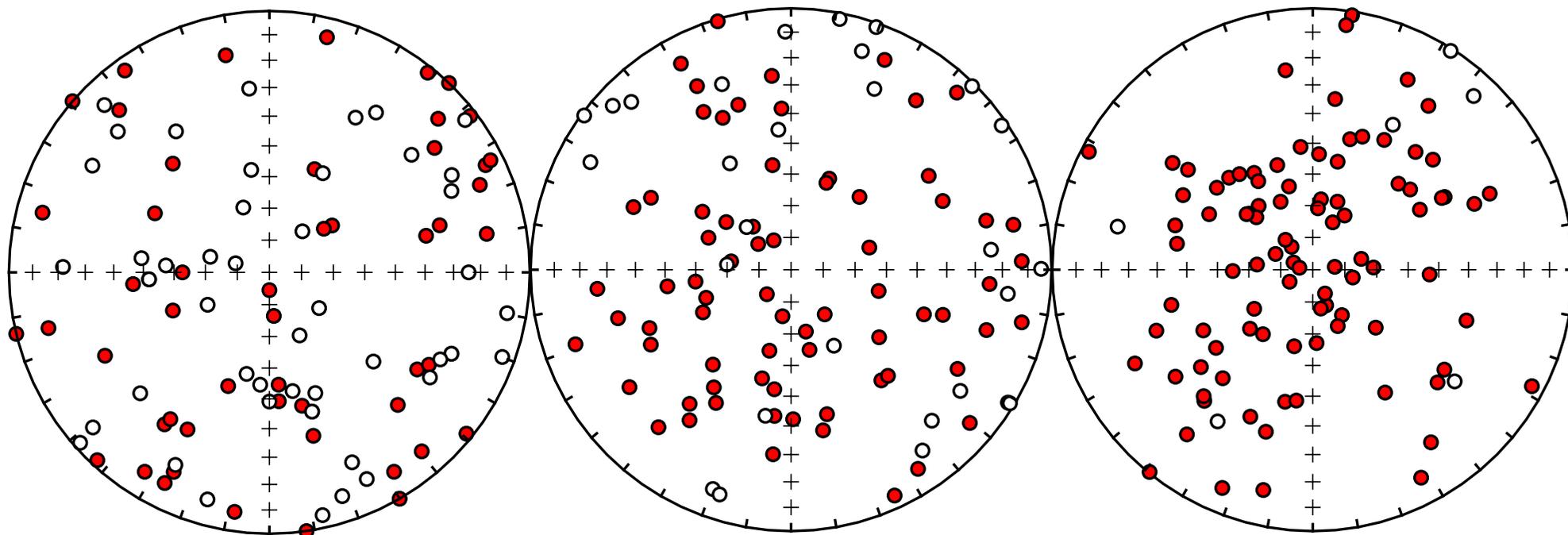
# Randomness: who wants to know?

- The conglomerate test (Chapter 9) relies on a test for randomness - if cobble directions are not random, then they were magnetized AFTER deposition
- If a paleomagnetic site has random directions, then the mean is meaningless

Random

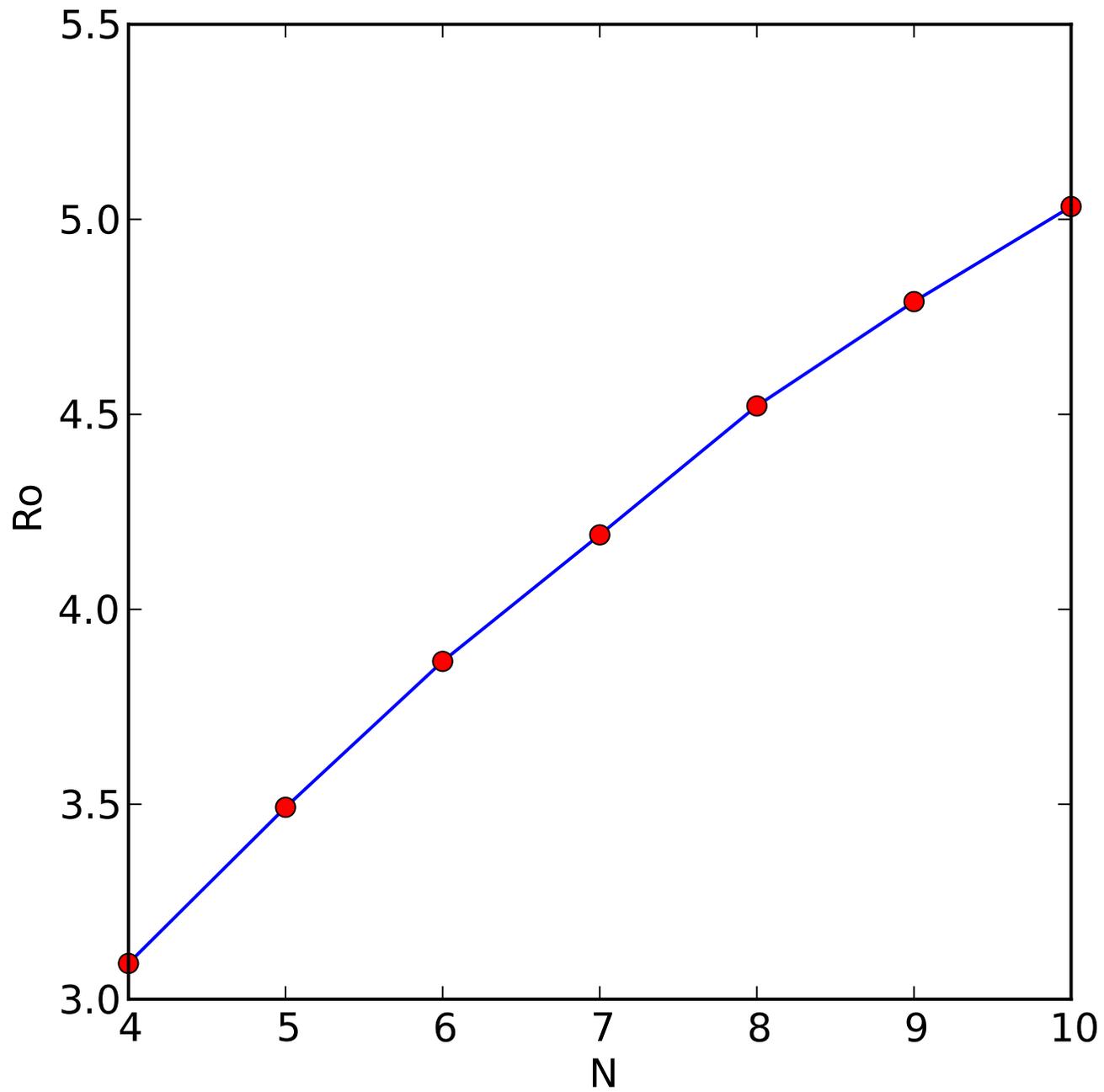
$\kappa = 1$

$\kappa = 3$



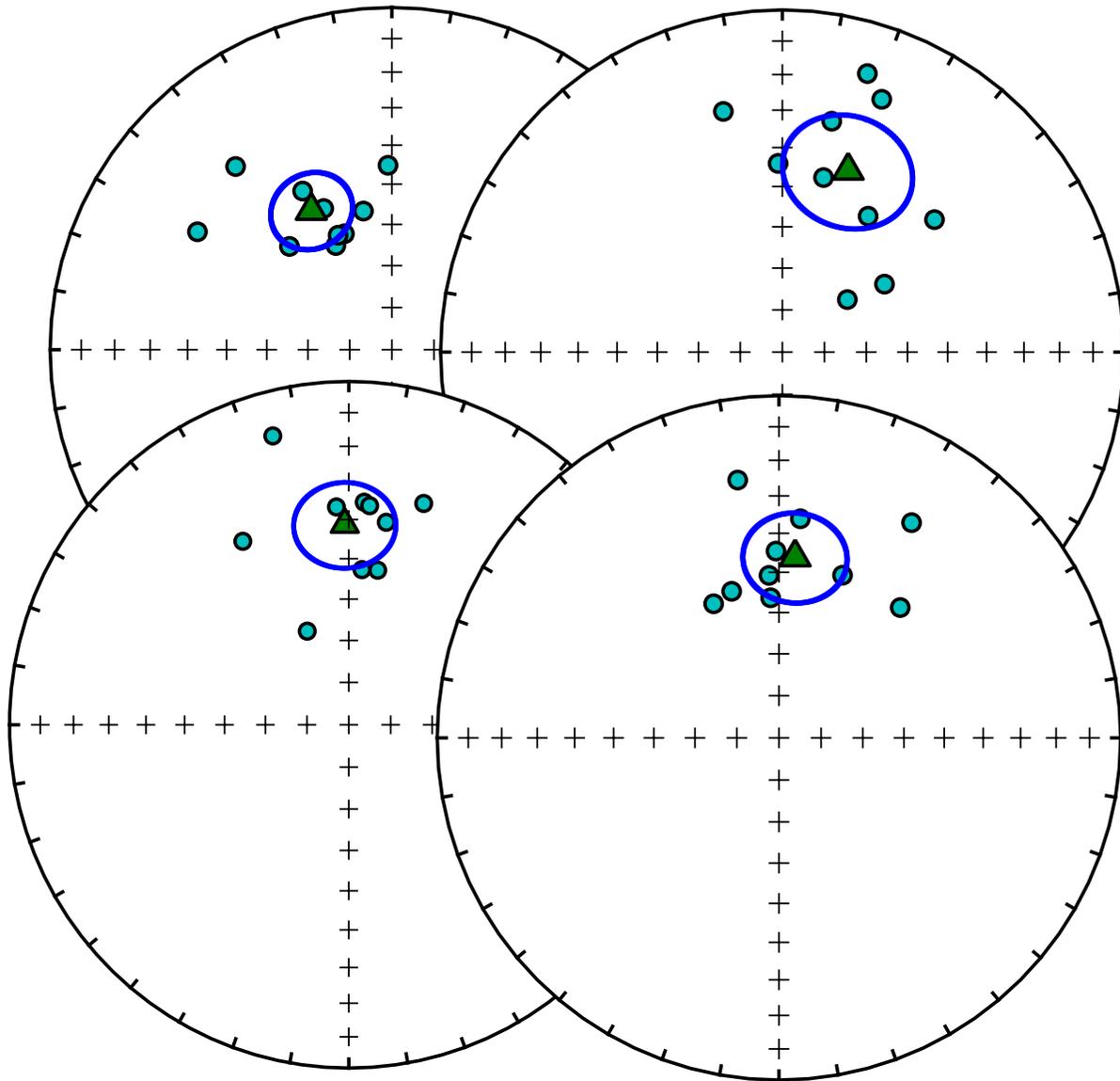
# Basic approach

- Scatter is related to  $R$
- We can generate distributions that are spread uniformly on a sphere (random) [use program `uniform.py` in `PmagPy`]
- Generate a bunch (10,000) of sets of uniform directions with  $N$  data points; calculate  $R$  and find the 95th percentile of these (95% of the  $R$ s are smaller than that). Call that  $R_0$  [This is a "Monte Carlo" type of approach.]
- If  $R$  in a given set of directions is  $> R_0$ , then your data set is 95% sure not to be random
- Can use shortcut of Watson (1956) in book. (see Chapter 11 and Table<sup>38</sup>C.2)



# Comparing directions

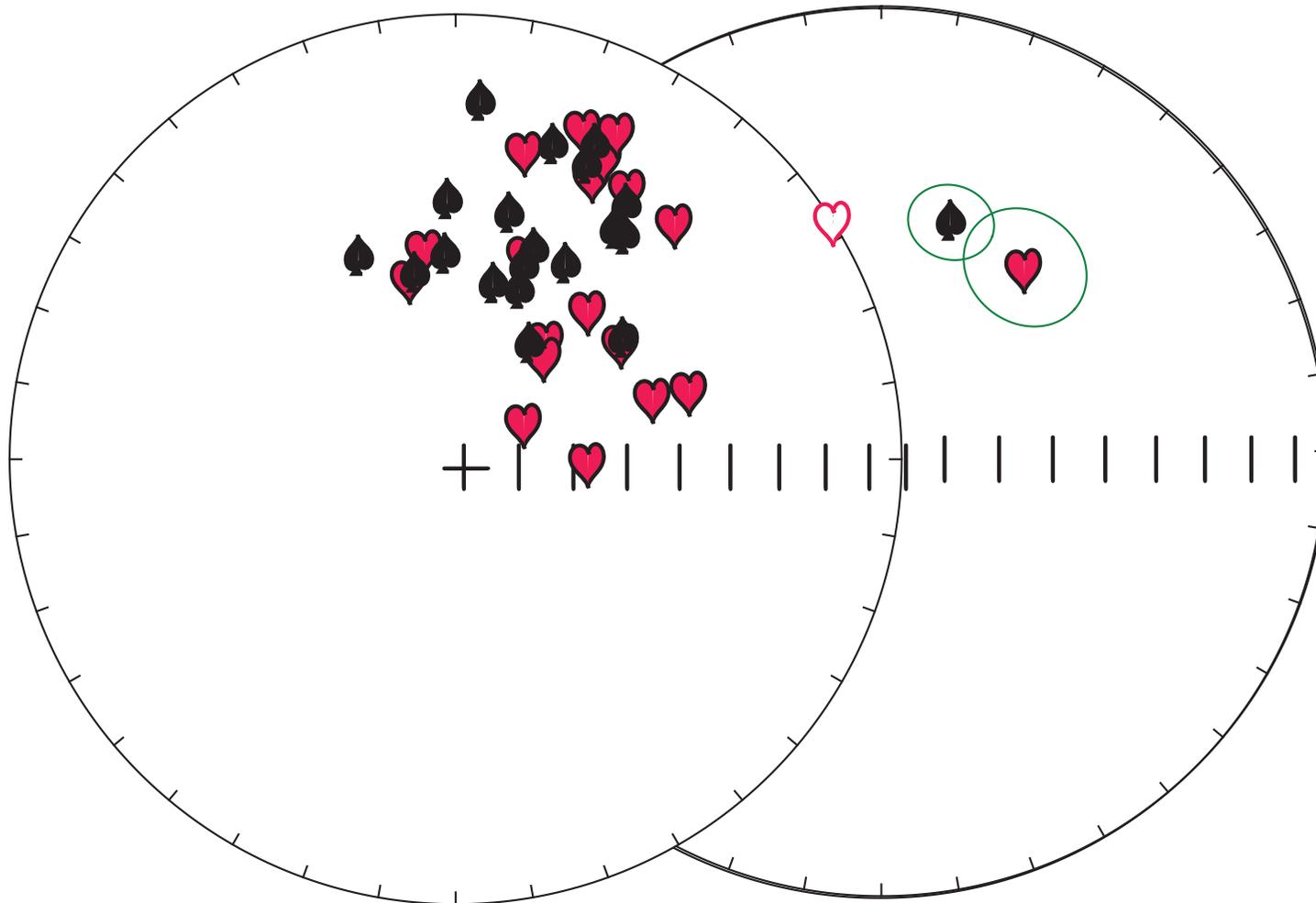
- If one is “known”, i.e. has no uncertainty, just see if  $\alpha_{95}$  of other includes it: Is a given direction vertical? Is a given direction coincident with the IGRF direction at the site?
- If both have some uncertainty (compare two paleomagnetic directions – for example the normal and reverse data from a study), this is a trickier case



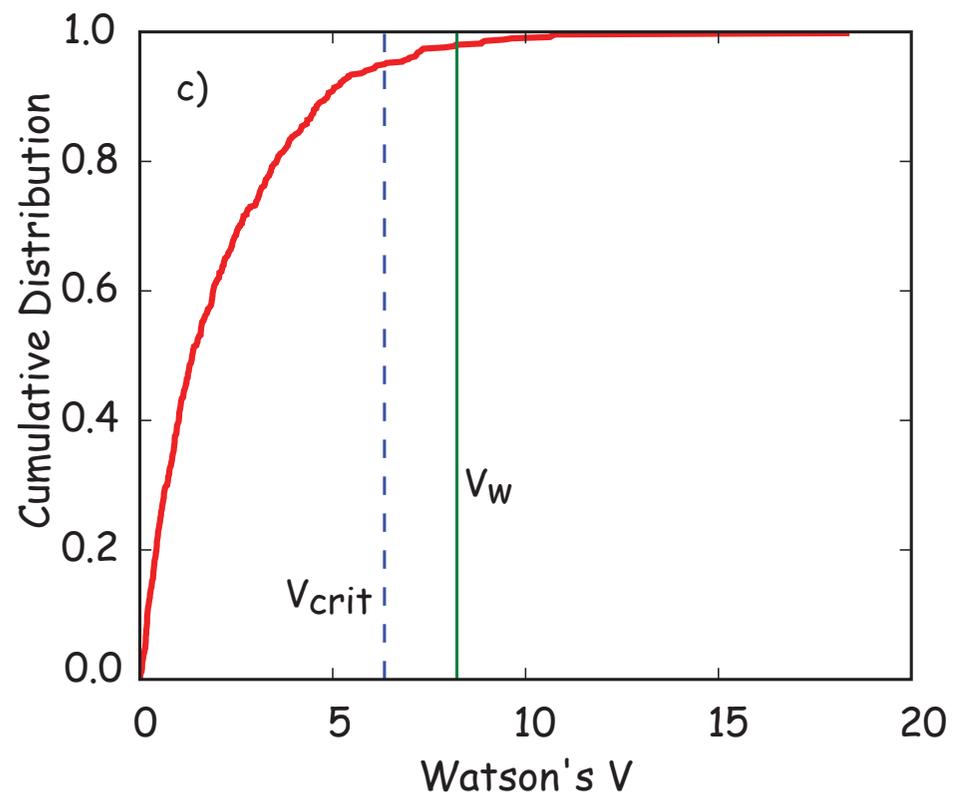
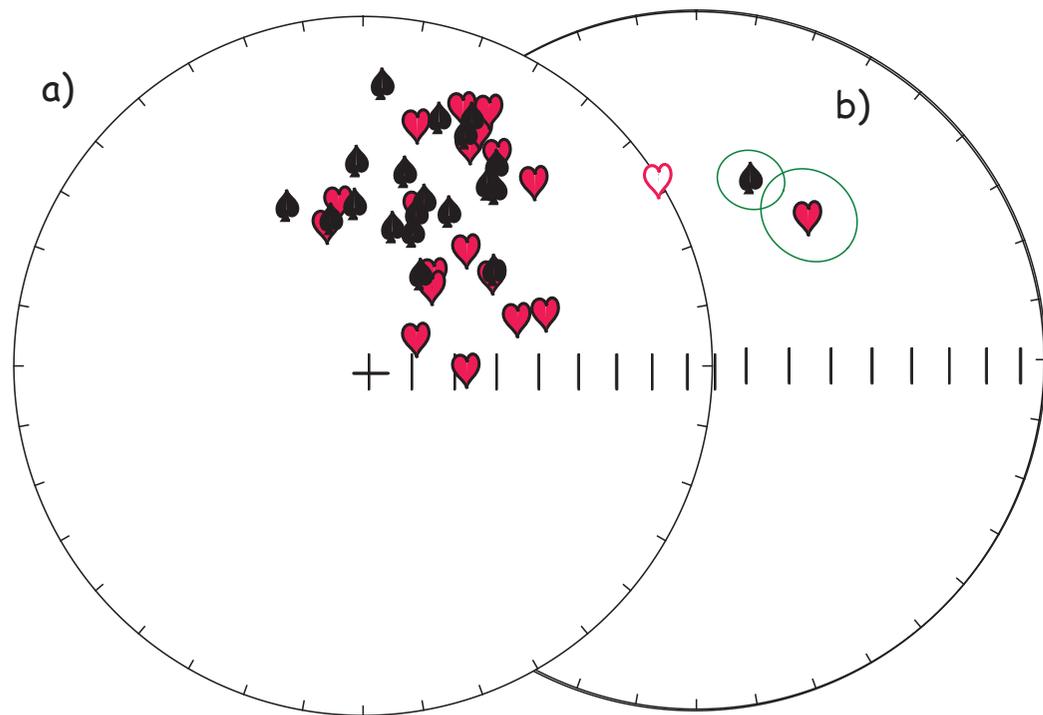
$\alpha_{95S}$  don't overlap  
means of other  
dataset  
clearly different

$\alpha_{95S}$  overlap mean of  
other dataset  
clearly the same

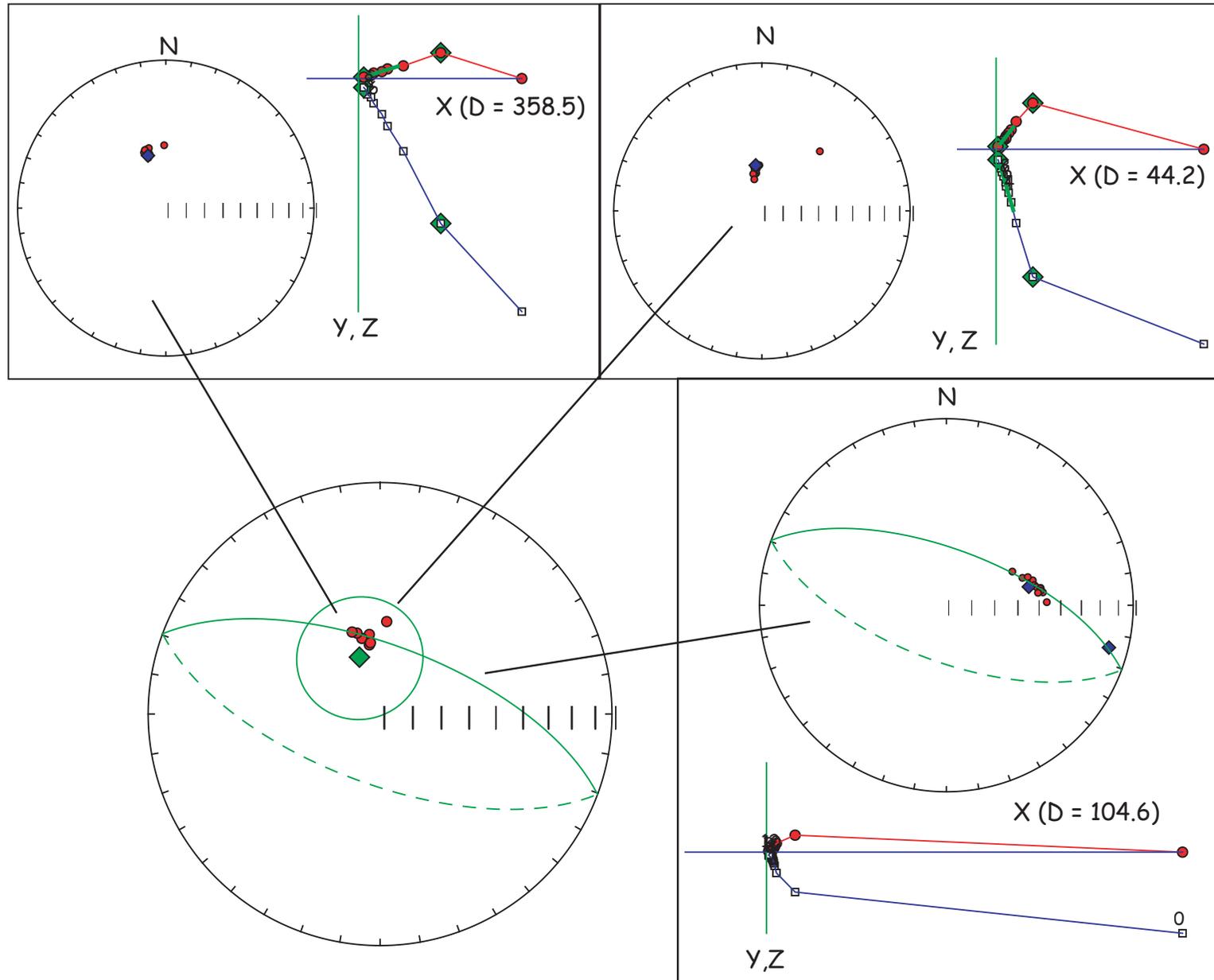
What about this case? Not so clear



- use statistic  $V_w$  – it increases with increasing distance between two data sets (see Chapter 11 and Appendix C.2.1) [check out `watsonsV.py` in `PmagPy`]
- null hypothesis that two datasets share common mean can be rejected if  $V_w$  is bigger than some critical value.
- Use Monte Carlo simulation to determine  $V_{crit}$  by calculating  $V_w$ s for lots of data sets with same  $N_s$  and  $k_s$  that DO share a common mean (e.g., `fishrot.py` in `PmagPy`). Determine 95th percentile for  $V_{crit}$
- If  $V_w > V_{crit}$ , two data sets are different (95% confidence)



# Combining lines and planes:



McFadden & McElhinny (1988) see Chapter 11

# Summary of useful programs in PmagPy

- `uniform.py` – generates sets of uniformly distributed data
- `fishrot.py` – generates sets of Fisher distributed directions
- `gofish.py`
- `watsonsV.py` – calculates  $V_{crit}$  and  $V_w$  for pairs of data sets
- others: `incfish.py`, `fishqq.py`, `eqarea_ell.py`, `watsonsF.py`, `lnp_magic.py`