Statistics on a sphere

- 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
The “normal” distribution
(Gaussian statistics)
A bunch (1000) of measurements of the length of something 10 cm long

Need to estimate:
most likely true length
and
“spread”
Special case where data fit distribution:

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

mean:

\[ \mu = \frac{1}{N} \sum_{i=1}^{N} z_i \]

variance:

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

[ \sigma \text{ 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.

![Histogram of Means of repeat trials with normal distribution curve and variance bars.]

![Histogram of Variance with bars and normal distribution curve.]

Thursday, April 28, 2011
mostly likely value of true length

1.96\(s\) bounds containing 95% of the measurements

\[ s_m = \frac{s}{\sqrt{N}} \]

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!
\[ P_{dA}(\alpha) \sin \alpha = \frac{\kappa}{2\pi \sinh \kappa} \exp(\kappa \cos \alpha) \sin \alpha \]
$\kappa = 5$  $\kappa = 10$  $\kappa = 50$

$=>$ decreasing scatter $=>$
$=>$ increasing concentration parameter $=>$
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”

$P_{\alpha}(\alpha)$

Angle from true mean ($\alpha$)

$\kappa = 100$

$\kappa = 50$

$\kappa = 10$

$\kappa = 5$
Useful statistics:
Useful statistics

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

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

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

often approximated by: \[ \alpha'_{95} \approx \frac{140}{\sqrt{kN}} \]

but only true for k>25
NB: 5% chance that a95 does not include vertical
Angular variance: \[ S^2 = \frac{1}{N - 1} \sum_{i=1}^{N} \alpha_i^2 \]

Circular standard deviation (CSD): \[ \simeq \frac{81}{\sqrt{k}} \]
How many data points do you need?

• Make a fake directional dataset (drawn from Fisher distribution) with $S=15$
• Take first 4 datapoints, calculate $k$ and CSD
• Add a 5th and repeat
• Keep going until $N=30$
CSD “settles down” with N~7
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!

\[ dm = \alpha_{95} \frac{\cos \lambda}{\cos I}, \quad dp = \frac{1}{2} \alpha_{95} (1 + 3 \sin^2 \lambda) \]
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.
Basic approach

- Scatter is related to R

- We can generate distributions that are uniformly on a sphere (random) [use program uniform.py in PmagPy]

- Generate a bunch (10000) sets of uniform directions with N data points, calculate R and find the 95th percentile of these (95% of the Rs are smaller than that). Call that Ro [This is a “Monte Carlo” type of approach.]

- If R in a given set of directions is > Ro, then your data set is 95% sure not to be random

- Can use shortcut of Watson (1956) in book. (see Chapter 11 and Table C.2)
Comparing directions

- If one is “known”, i.e. has no uncertainty, just see if a95 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.
\( \alpha_{95\%} \) don’t overlap means of other dataset clearly different

\( \alpha_{95\%} \) 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 $V_{ws}$ 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
Take home message

- There are many causes of scatter in paleomagnetic directions
- Using Fisher statistics, we can calculate mean directions and confidence intervals.