

MIMM4750G

Probability and likelihood



REMINDER: A 50% INCREASE
IN A TINY RISK IS **STILL TINY.**

Probability

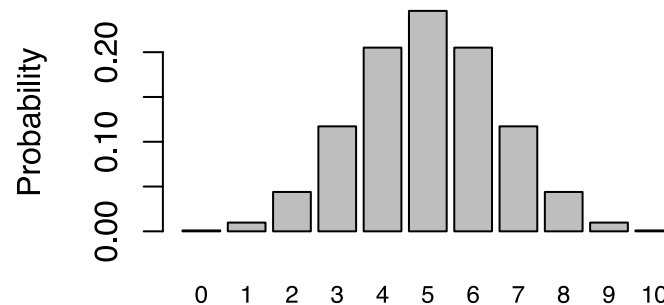
- What is the probability of getting 5 heads in 10 coin tosses?
- Let the probability of heads is p – this is a *parameter* (hypothesis).
- Let the number of heads ($y = 5$) and tosses ($N = 10$) be our *data*.
- As usual, let's make some assumptions:
 1. My coin tossing is truly random.
 2. The coin tosses are independent - one result does not influence another.
 3. p is constant - the coin and my tossing action do not change.

Binomial probability

- The probability of the data (N,y) given the hypothesis (p) can be described by the *binomial model*:

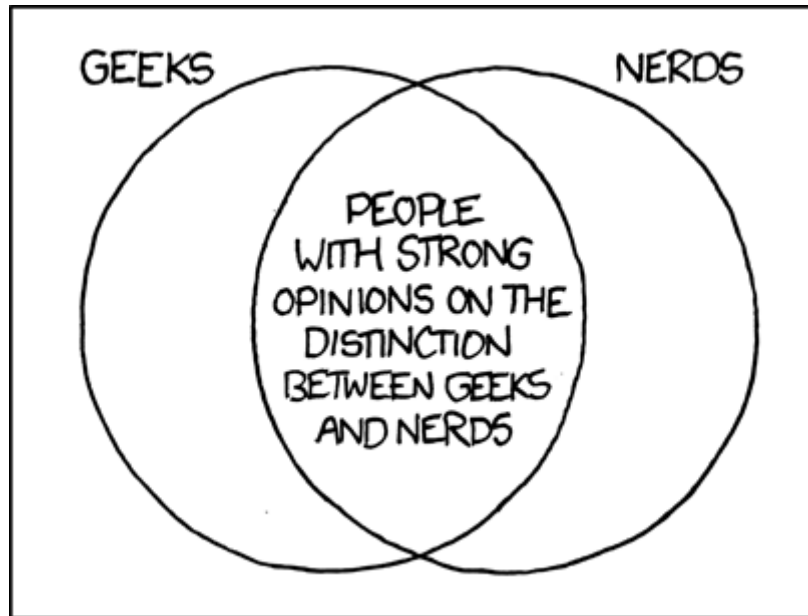
$$P(N, y \mid p) = \binom{N}{y} p^y (1 - p)^{(N-y)}$$

- ("Binomial" indicates *two* and only two outcomes.)
- The probability distribution looks like this:



Properties of a probability distribution

- A probability distribution function (PDF) can be *discrete* or *continuous*.
- The binomial PDF is discrete. There can be either 1 or 2 heads but never 1.5.
- A PDF *must* sum to 1 (100%) across all possible data outcomes, *for a given hypothesis*.



Probability and inference

- When our model of some biological or epidemiological process is simple enough, we may be able to calculate the probability distribution.
- This is the probability of the *data* given the *hypothesis*.
- **But we already know the data!** What we want to learn about is the hypothesis!

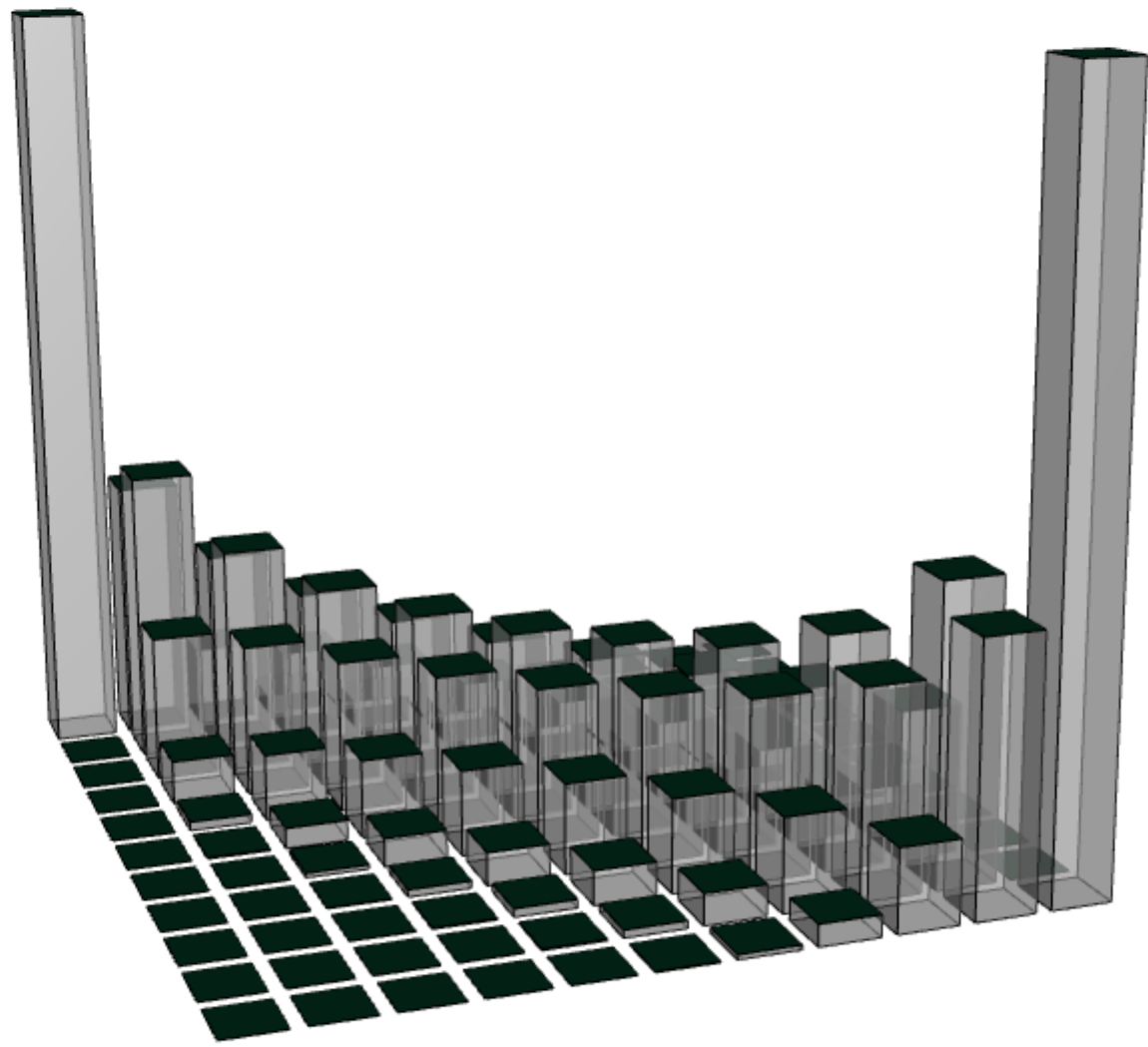
Likelihood

- This probability distribution

$$P(N, y \mid p) = \binom{N}{y} p^y (1 - p)^{(N-y)}$$

has two sets of variables:

1. Parameters that define the hypothesis (p).
2. Variables that comprise the data (N, y).



You must enable Javascript to view this page properly.

Drag mouse to rotate model. Use mouse wheel or middle button to zoom it.

Likelihood

- The term *likelihood* is a signal that we using the PDF to measure the probability of the **hypothesis** given the **data**.
- Since summing across parameters does not sum to 1, all that matters is the *relative* difference in likelihood.
- Might as well drop factors independent of parameters, e.g., $\binom{N}{y}$:

$$P(N, y \mid p) \propto p^y (1 - p)^{(N-y)}$$

- Sometimes we replace the P with \mathcal{L} , e.g., $\mathcal{L}(p \mid N, y)$

Maximum likelihood

- Trying to learn about the world by determining parameter values that maximize the likelihood of the model, given the data.
- These parameter values are the *maximum likelihood estimates* (MLE).
- They are only as useful as the model is an accurate representation of reality.

Finding the MLE

- To estimate the parameter that maximizes likelihood, we can use first-year calculus (ha!)

$$\mathcal{L}(p|N, y) \propto p^y (1 - p)^{N-y}$$

$$\log(\mathcal{L}) \propto y \log(p) + (N - y) \log(1 - p)$$

$$\frac{d \log(\mathcal{L})}{dp} \propto \frac{y}{p} + (-1) \frac{N - y}{1 - p}$$

- If we set the left side to 0, some algebra gives us $\hat{p} = y/N$.

Log-likelihood

- As we add more data, the likelihood gets smaller:

$$P(\text{everything}) = P(\text{first thing}) \times P(\text{another thing}) \\ \times \dots P(\text{last thing})$$

- Taking the log of a very small value gives you a very negative number:
 $\log(3.45 \times 10^{-73}) = -166.8503$
- The log likelihood should never be 0 or greater.
- The MLE is the log-likelihood that is the least negative.

ML tree reconstruction

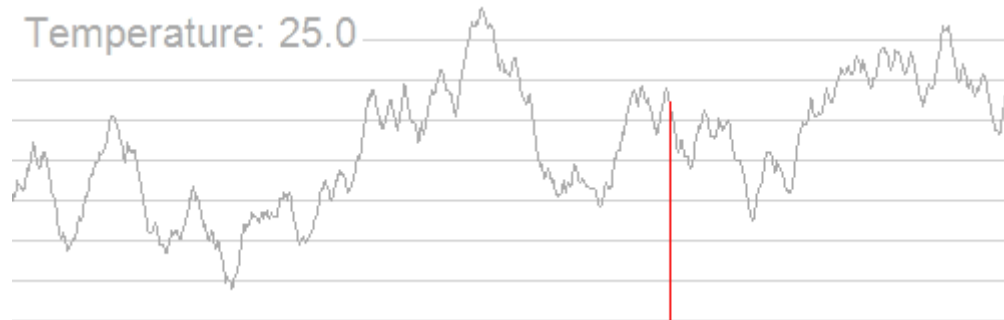
- For reconstructing trees, the parameters include:
 - the distribution of branches relating the sequences
 - the lengths of those branches
 - parameters of the model of evolution along those branches
- (We'll get more into models of evolution later.)
- Obviously, it is not possible to calculate an exact likelihood for this model!

Maximum likelihood heuristics

- As long as we can calculate the likelihood for a specific set of parameters, we can search parameter space for the MLE (*optimization*).
- Heuristic search methods (again!)
- For phylogenies, this requires a way to move through the space of trees.
- First proposed by [Joe Felsenstein](#).

Optimization

- A major challenge of optimization heuristics is that there can be multiple "peaks" (locally optimal solutions)
- A "greedy" heuristic that searches for the MLE by always tuning parameters to increase likelihood.
- A more robust heuristic should have some strategy for climbing down local peaks.



Other strategies

- Random restarts - begin searches at different parameter combinations to sample different peaks.
- Starting with a pretty good guess: many ML-based tree reconstruction programs will initiate the search at a distance-based tree (such as NJ).

Tree rearrangements

- Nearest-neighbor interchange: swap subtrees that are connected by a single branch.
- Subtree pruning and regrafting: move a subtree to an entirely different location in the tree.
- Small moves are easier to make, but can be slow to traverse tree space.

Software

- RAxML
- PhyML
- FastTree2
- GARLI
- MEGA