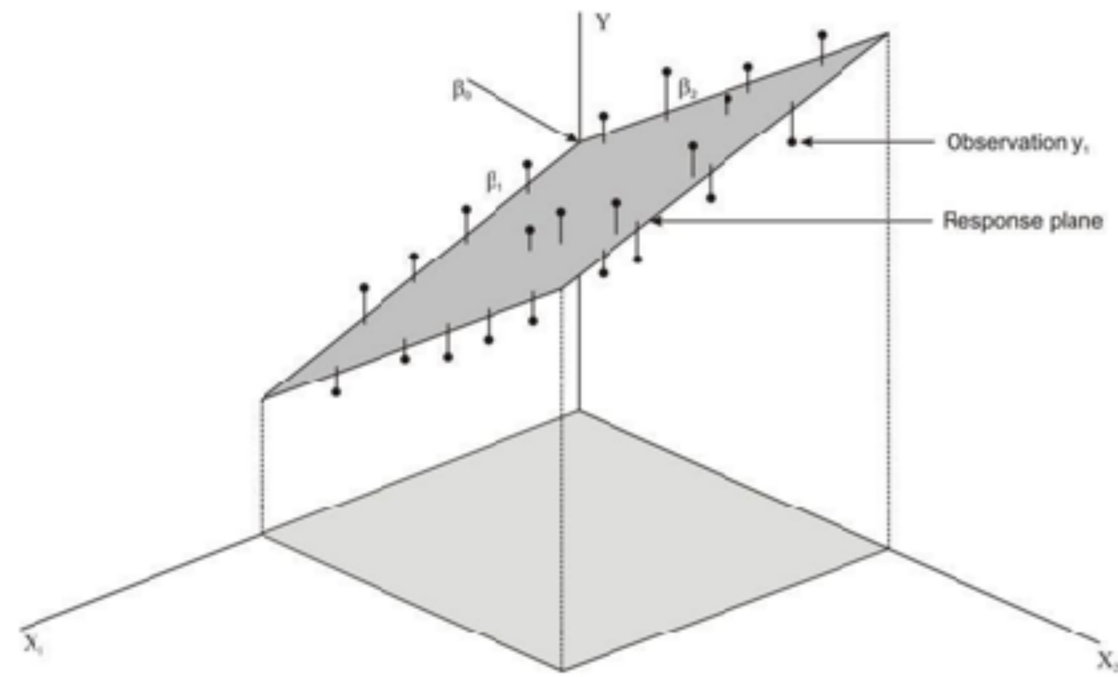
Introduction to Probability

Stephen Keeley
SML 201
Tuesday, March 26th, 2019



- When doing predictive modeling before break, we assumed our data were *noisy* or *random*

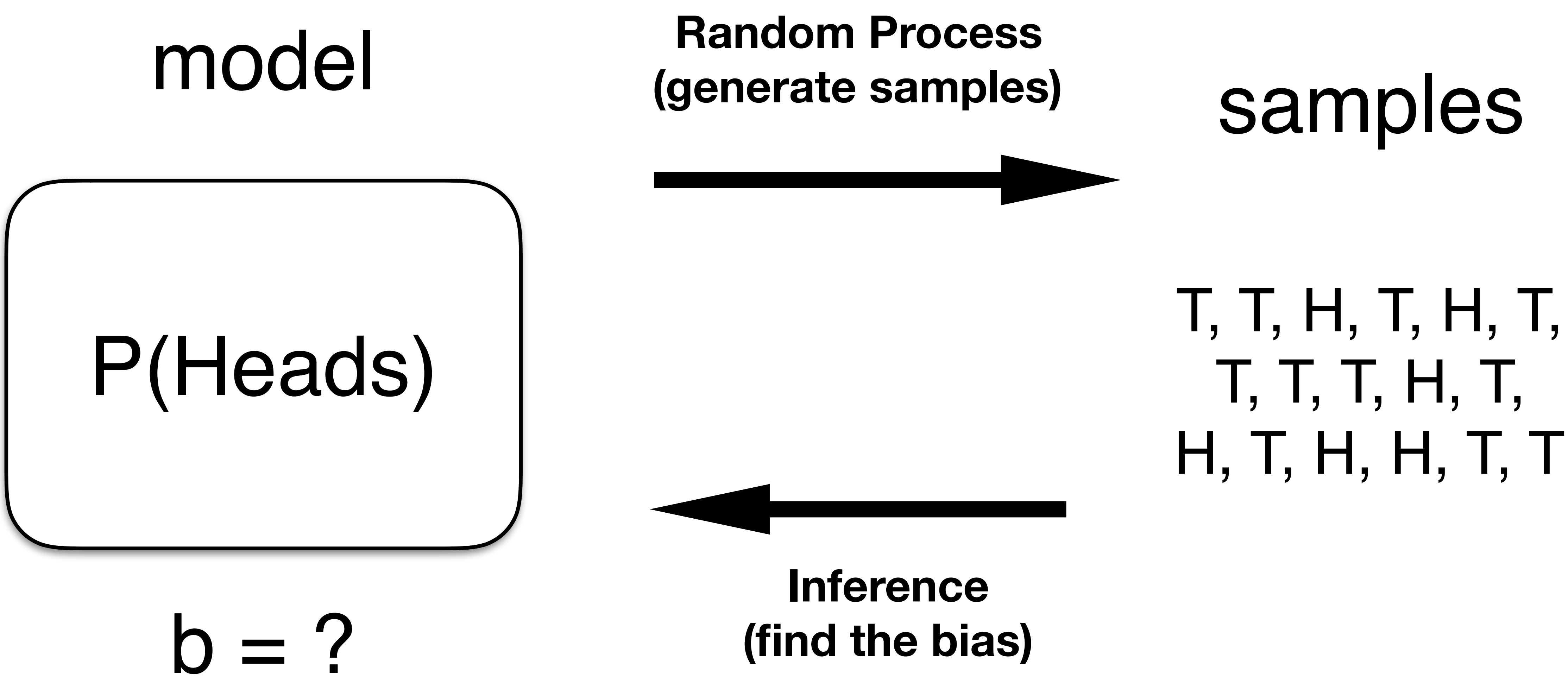
- i.e. the data never *exactly* corresponded to our model

- Instead, there was some random *error* we took into account

How do we describe randomness?

- We frame probability in terms of a **random process** giving rise to an **outcome**
  - **Roll a die → 1, 2, 3, 4, 5, or 6**
  - **Flip a coin → H or T**
- The **probability** of an outcome is the proportion of times the outcome would occur if we observed the random process an infinite number of times

- Simple example : coin-tossing
  - Try to determine the ‘bias’ of the coin (b)
- parameter estimation / Inference

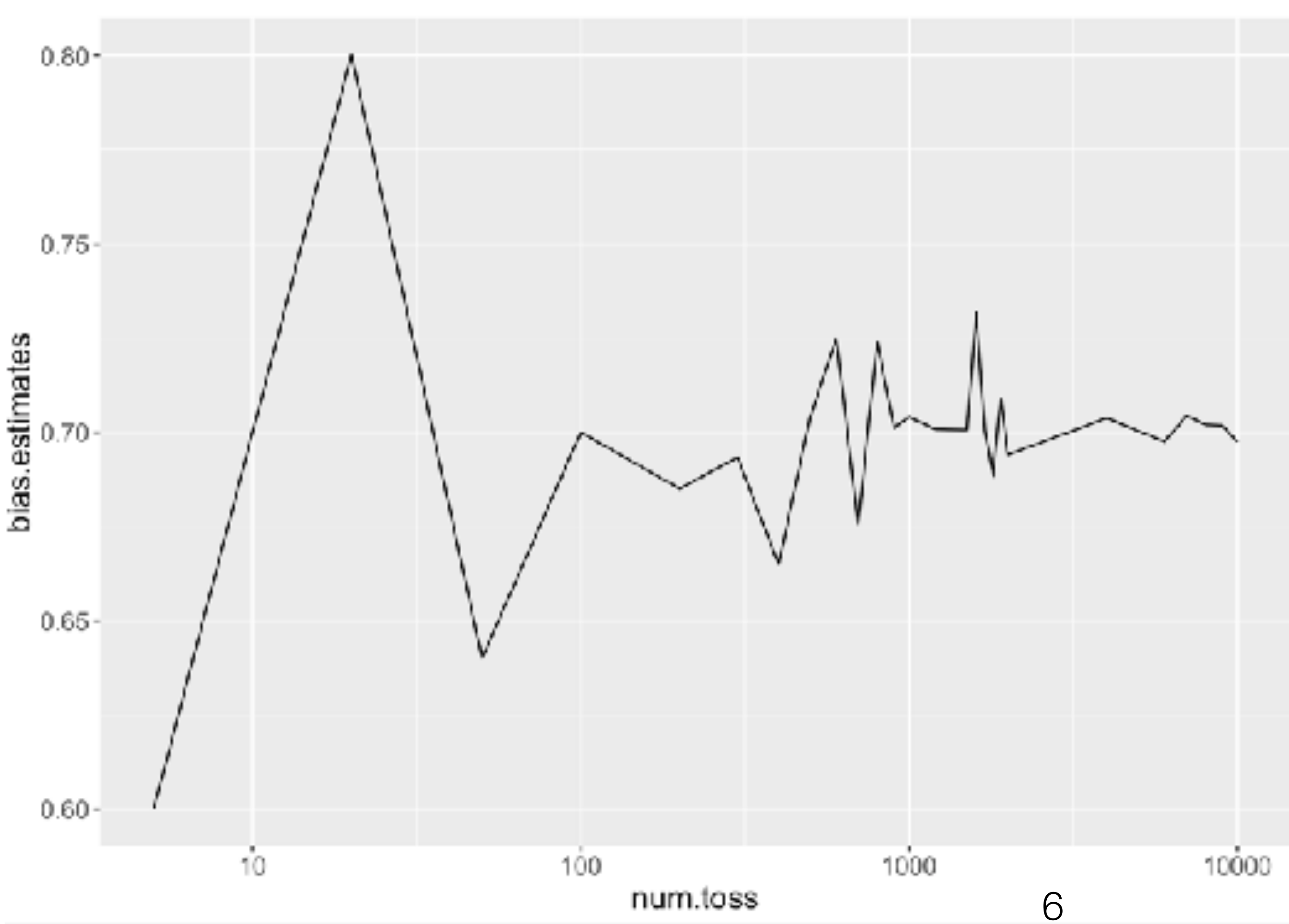


R Code

- An estimator of the **probability** (bias.estimates) can be calculated for any number of **outcomes**

- For estimating the probability of Heads, this is the total number of heads as a proportion of the total number of tosses

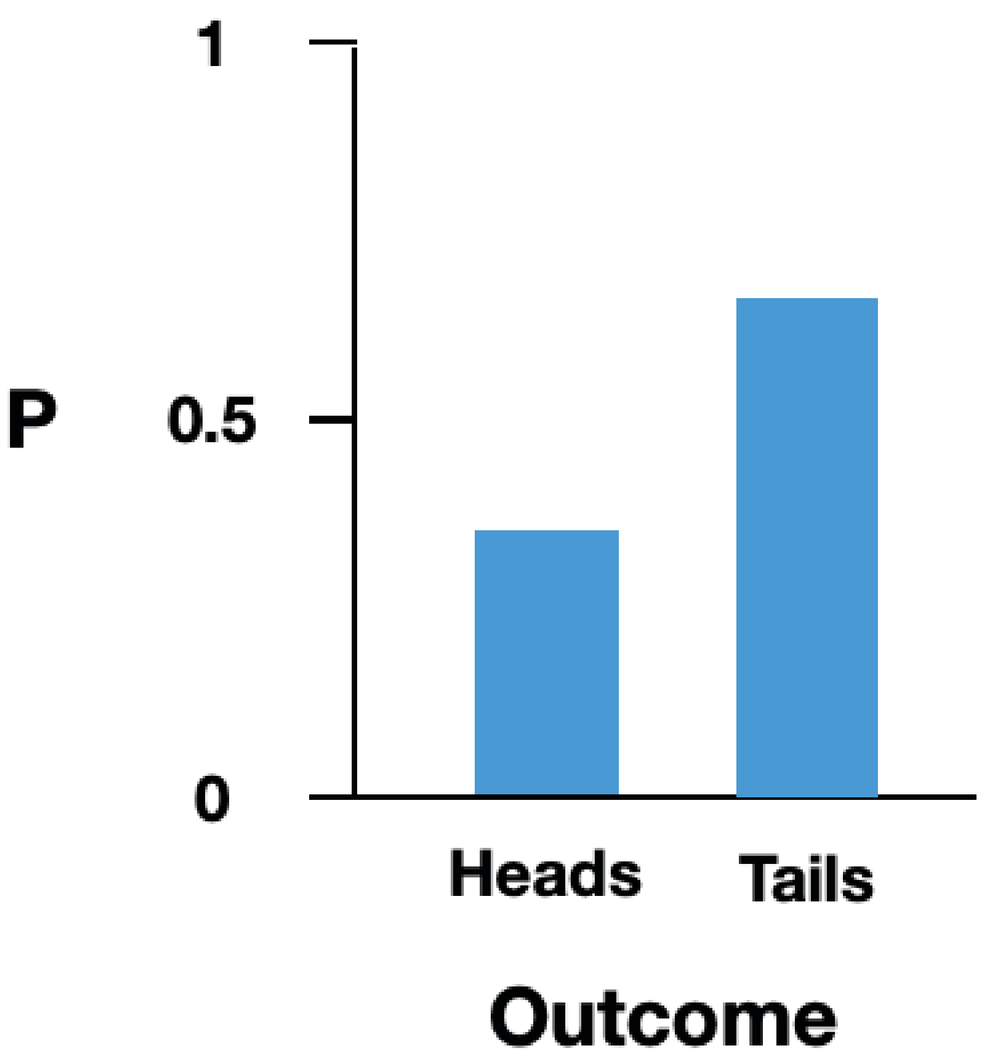
bias.est = # of Heads/# of tosses



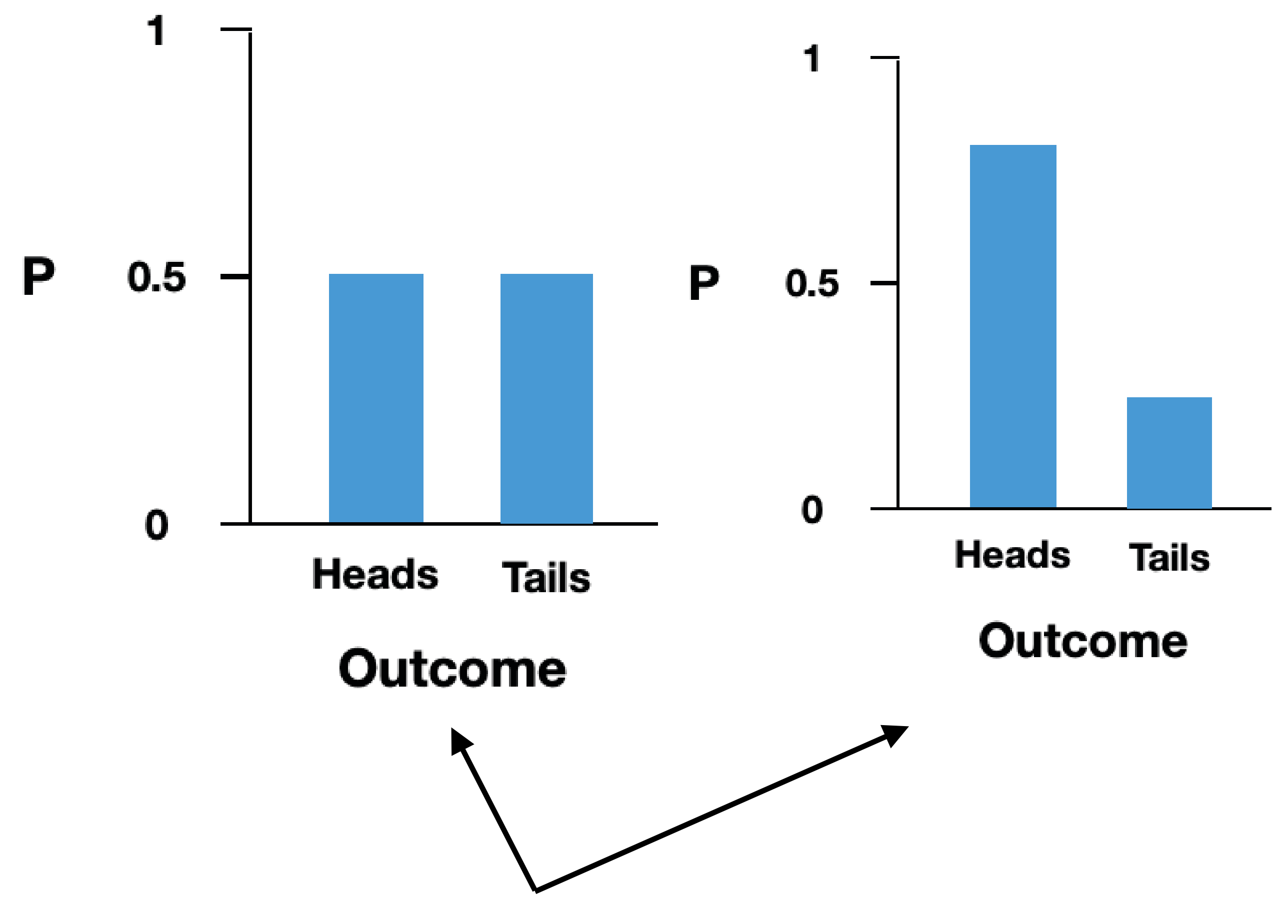
As the number of tosses increases, the estimate of the bias, b, is more accurate

Coin tossing (probability mass func)

Probability Mass Function (PMF)



- 0<=P<=1
- Probability values sum to 1



Coins with a different bias

Bernoulli Random variable PMF

(discrete probability distribution)

$$f(k;p) = p^k(1-p)^{1-k} \quad \text{for } k \in \{0,1\}$$

- probability of heads is **p ,** probability of tails is **1-p**
- 1 or 0 corresponds to ‘Heads’ or ‘Tails’
- When p =.5 —> fair coin toss
- when p = anything else —> weighted coin

R Code

Binomial Random Variable

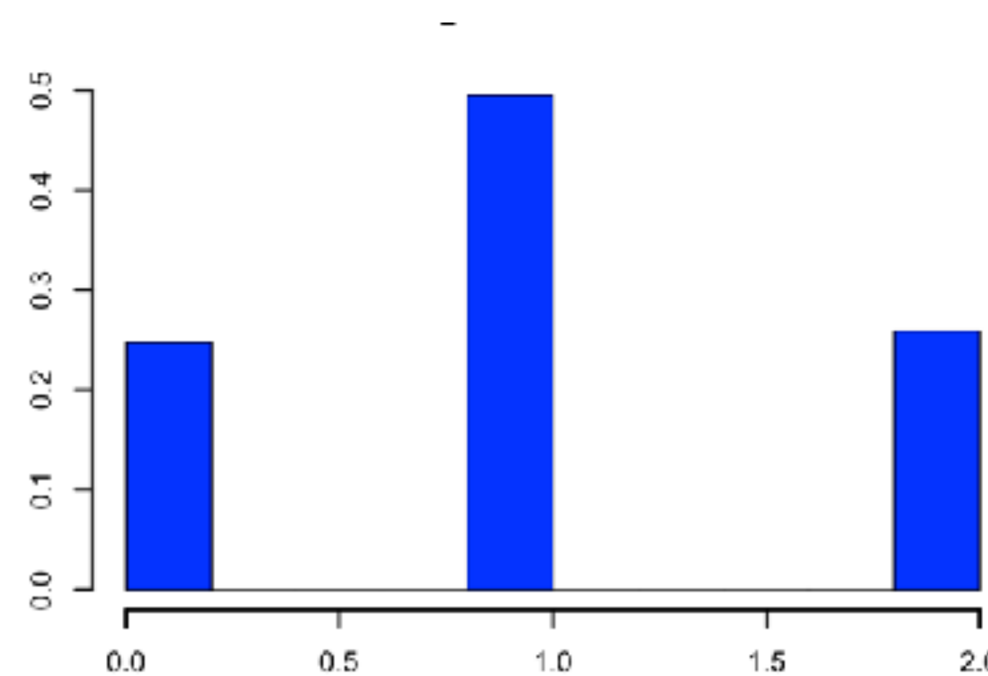
- Let's consider N *fair* coin tosses. What is the probability of getting M "Heads" outcomes.

PMF

N = 2

Number of Heads	Probability
0	1/4
1	1/2
2	1/4

P

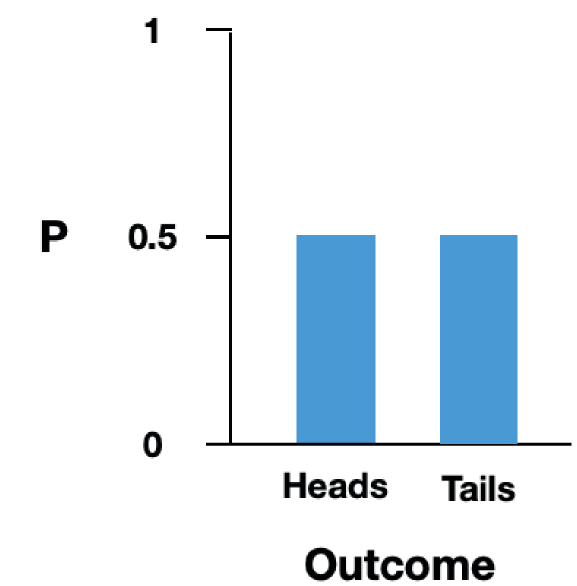


Number of Heads

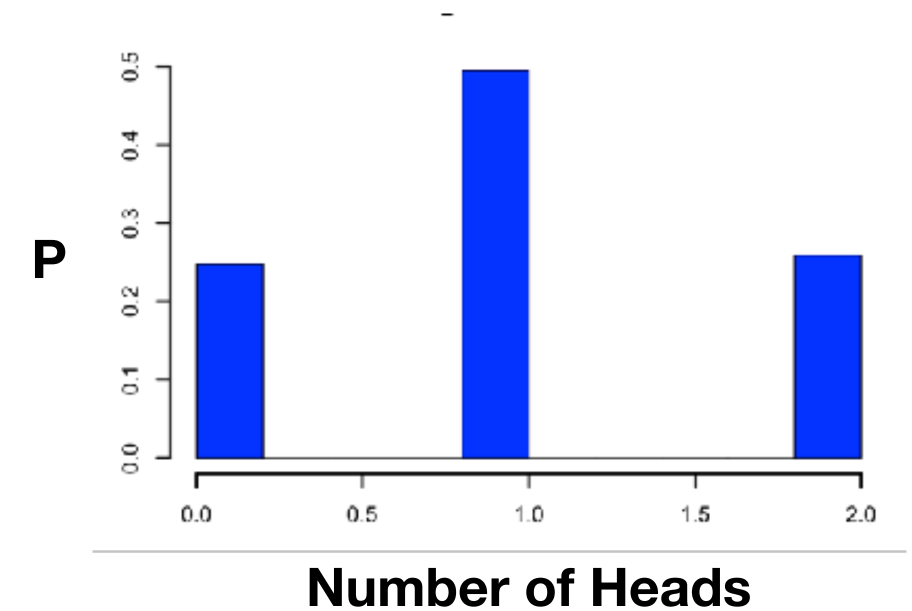
- Let's simulate this! If we simulate two coin tosses many times, the outcomes should follow the PMF above.

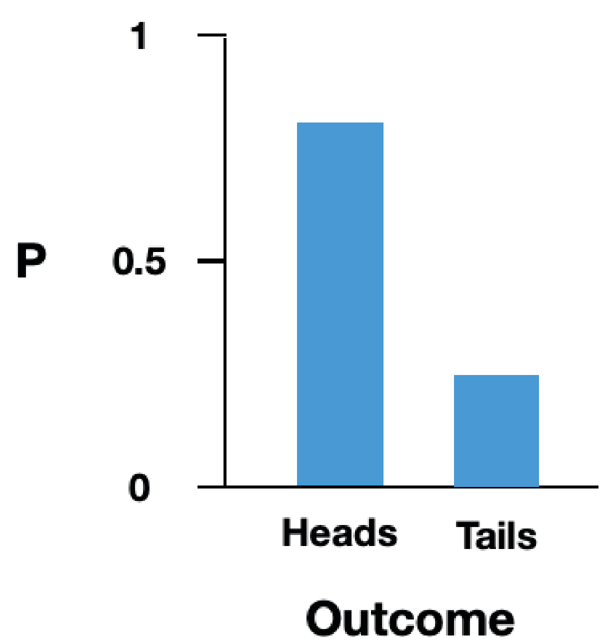
Probability mass functions

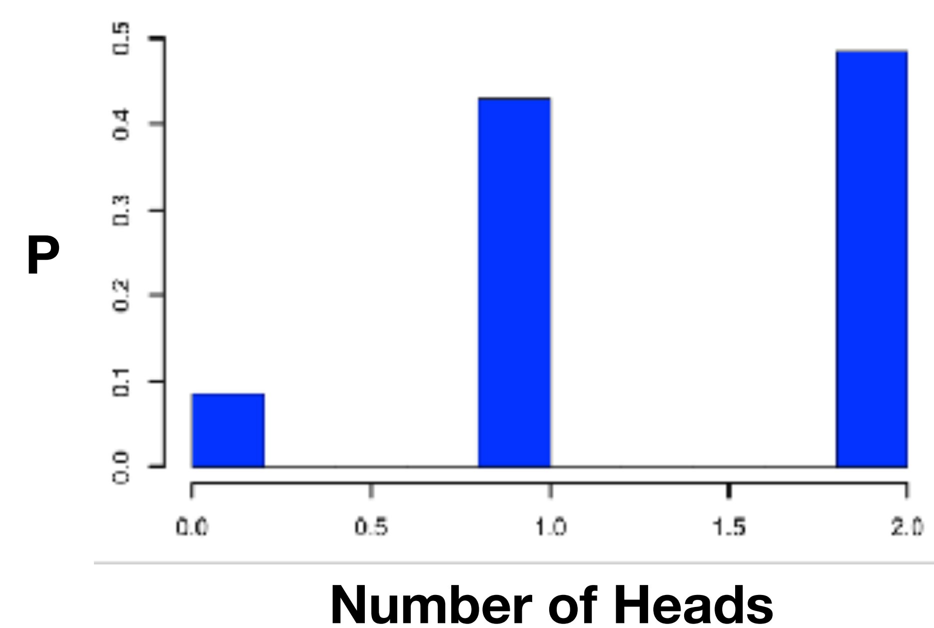
Bernoulli



Binomial



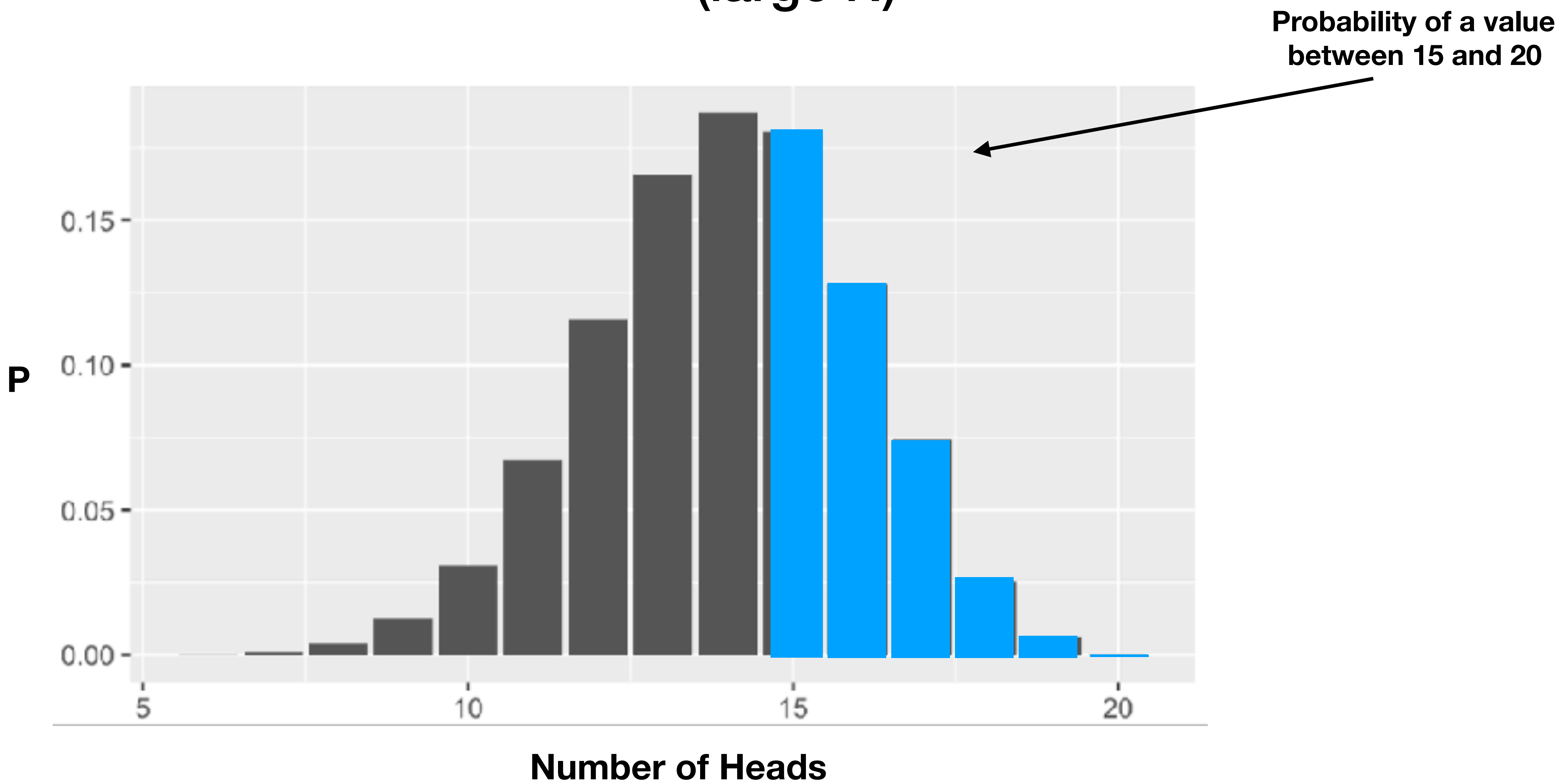




- All sum to 1
- **Outcomes** may be different
- Shapes (parameters) may be different

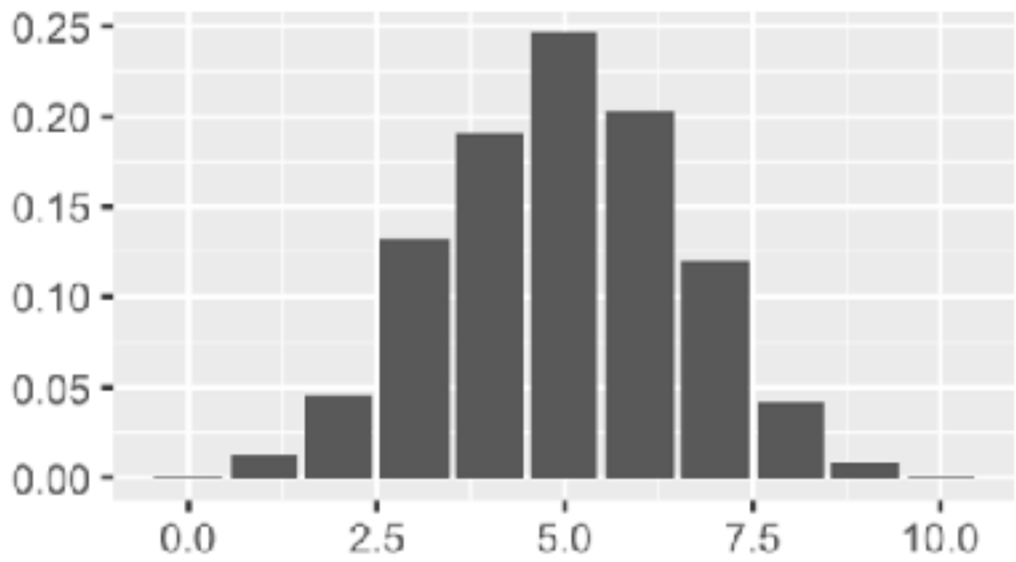
Binomial probability mass function

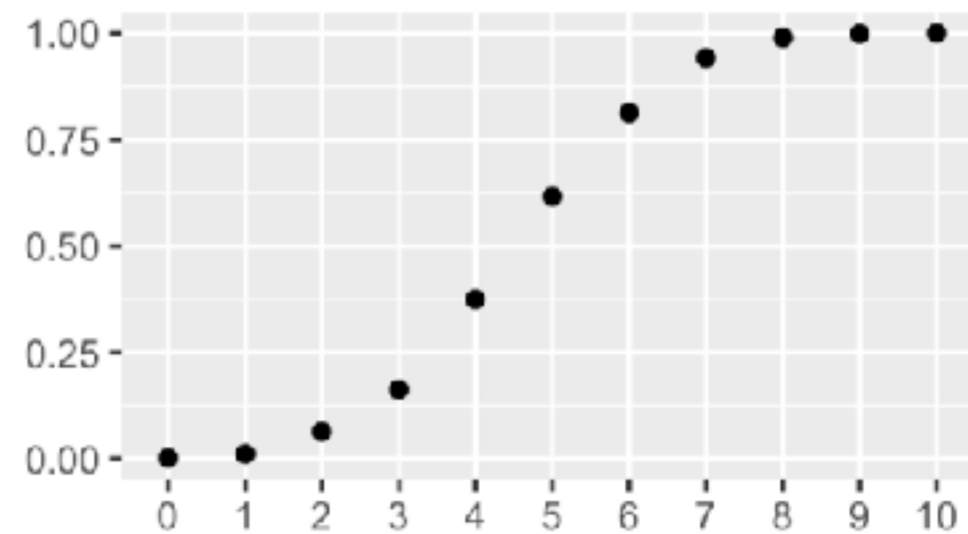
(large N)



- Probabilities of ranges of outcome possibilities can be determined by summing across outcomes
- Sum is still between 0 and 1

Cumulative Mass Function



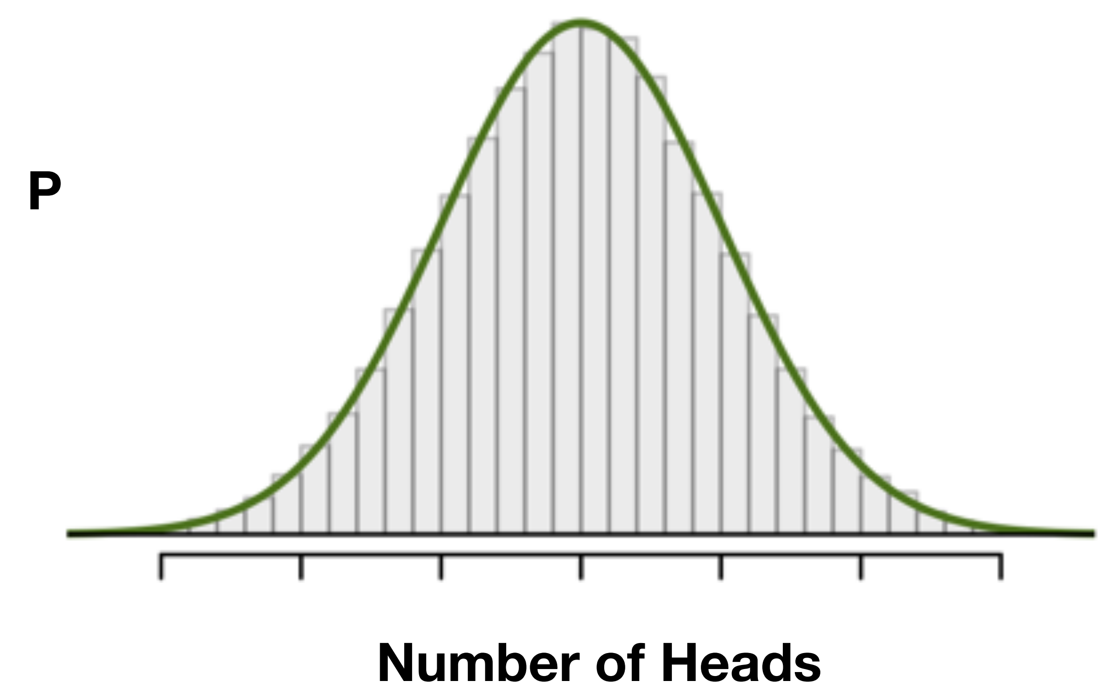


- Consider the probability that the variable is *equal to or less than* some outcome
- Plot this for all possible outcomes
- This is called a Cumulative Mass function

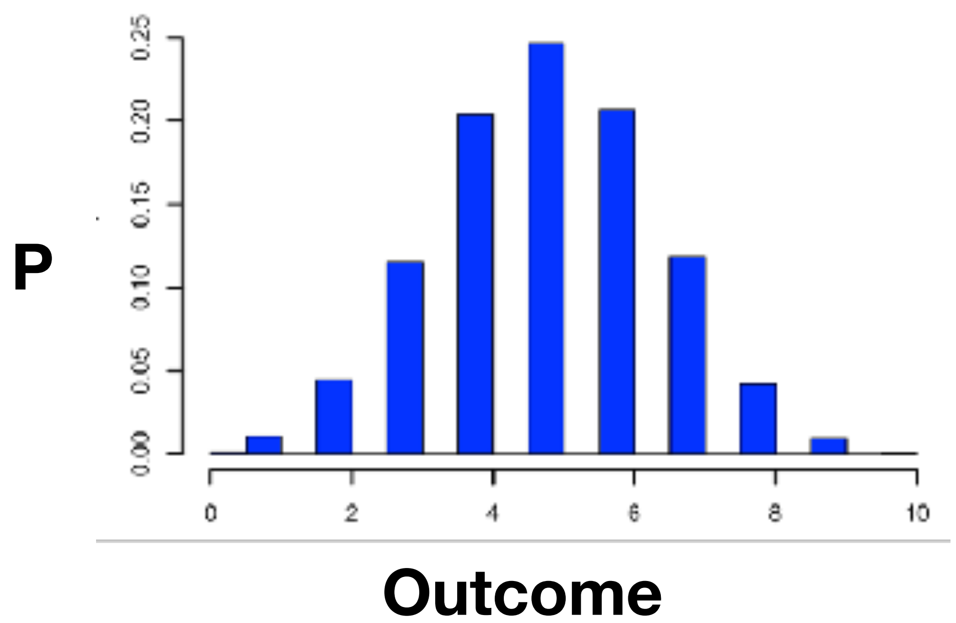
- The rightmost value always is equal to one
- Each outcome's value represents the fraction of the data that achieve *that outcome or a smaller value*

What about if number of coin tosses is *really* big?

- As N grows, the number of possible outcomes grows
- The PMF then becomes approximately **smooth**
- This can be described with a *normal curve* or *approximated as a Gaussian distribution*

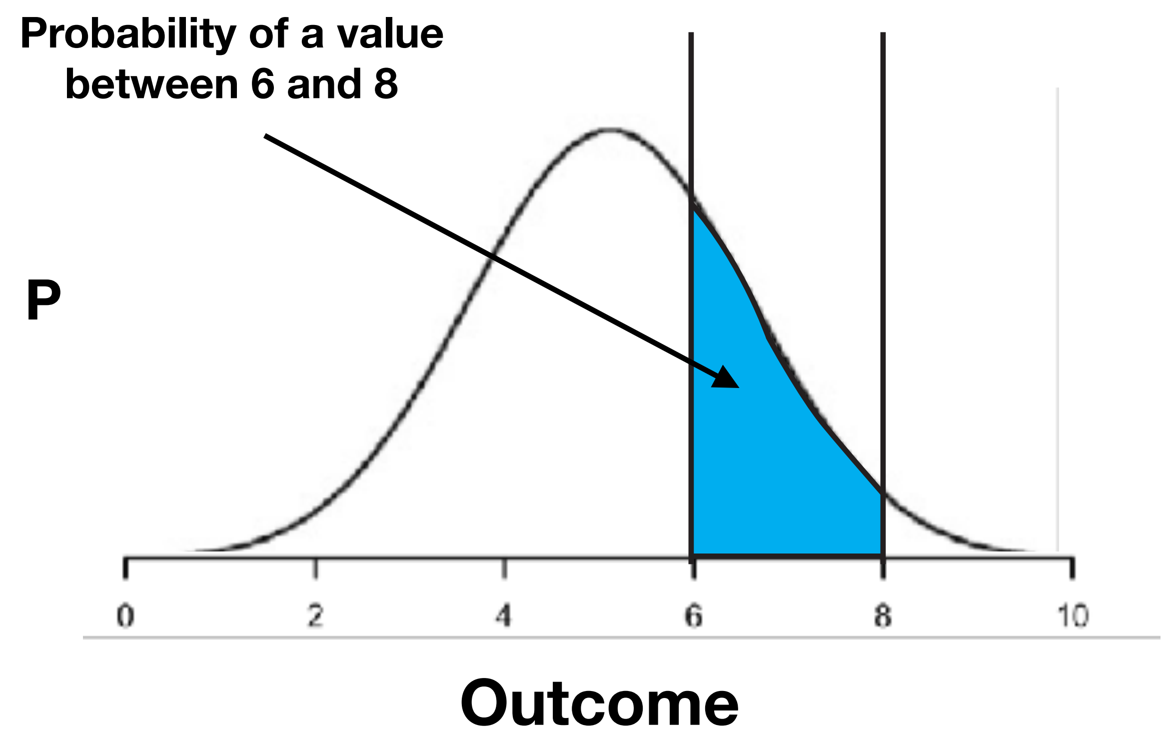


PMF



- Discrete probability distribution
- probabilities are associated with particular values/outcomes
- probabilities sum to 1

PDF



- Continuous probability distribution
- the probability of any exact value is 0, the probability of a range of values has some finite value
- probabilities integrate to 1