Northeastern University

INFO 6105
Data Sci Eng Tools & Mthds
Lecture 3 Bayes Formula

18 September 2019

Bayes' Formula

- We have a prior belief in event A, beliefs formed by previous information, e.g., our prior belief about bugs being in our code before performing tests
- Secondly, we observe our evidence. if our code passes X tests, we want to update our belief to incorporate this. We call this new belief the posterior probability
- Updating our belief is done via the following equation, known as Bayes' Theorem, after its discoverer Thomas Bayes:

$$P(A|X) = \frac{P(X|A)P(A)}{P(X)}$$

The formula is not unique to Bayesian inference: it is a mathematical fact with uses outside Bayesian inference

- Bayesian inference merely uses it to connect prior probabilities P(A) with updated posterior probabilities P(A|X)

More useful Formulation

- **When multiple events A_i form an exhaustive set with another event B**
- B can be written as $B = \sum_{i=1}^{n} B \cap A_i$
- So, probability of B can be written as $P(B) = \sum_{i=1}^{n} P(B \cap A_i)$
- Since $P(B \cap A_i) = P(B|A_i) \times P(A_i)$
- Replacing P(B) in the equation of conditional probability:

$$P(A_i|B) = (P(B|A_i) \times P(A_i))/(\sum_{i=1}^{n}(P(B|A_i) \times P(A_i)))$$

Bayes' Rule

- As Bayesians, we start with a belief, called a **prior**
- Then we obtain some data and use it to update our **belief**
- The outcome is called a **posterior**
- Should we obtain even more data, the ***old*** **posterior** becomes a ***new*** **prior** and the cycle repeats
- **All components are probability distributions**
- This process employs the **Bayes rule**:
  - **P(A | B) = P(B | A) * P(A) / P(B)**
- In Bayesian machine learning we use the Bayes rule to infer model parameters (theta) from data (D):
  - **P(theta | D) = P(D | theta) * P(theta) / P(data)**
  - ***This is how we can ask machines what parameters they use to make decisions***

Bayesian ML: P(theta | D) = P(D | theta) * P(theta) / P(data)

- **P(data) is something we generally cannot compute**
  - Since it's just a normalizing constant, it doesn't matter much. When comparing models, we're mainly interested in expressions containing theta, because P(data) stays the same for each model
- **P(theta) is *a prior*, our belief of what the model parameters *might* be**
  - Most often our opinion in this is vague and if we have enough data, we simply don't care
  - Inference converges to probable theta as long as it's not 0 in the prior
  - One specifies a prior in terms of a parametrized distribution
- **P(D | theta) is called *likelihood of data given model parameters***
  - The formula for likelihood is model-specific. People often use likelihood for evaluation of models: a model that gives higher likelihood to real data is better
- **Finally, P(theta | D), *a posterior*, is what we're after**
  - It's a probability distribution over model parameters obtained from prior beliefs and data

Does this make sense?

- Let's look at our intuition
  - (in black)

Intuition: What does “Bayesian inference” mean?

Inference = *Educated guessing*

Thomas Bayes = A nonconformist Presbyterian minister in London back when the United States were still The Colonies.

He also wrote an essay on probability. His friend Richard Price edited and published it after he died.

Bayesian inference = Guessing in the style of Bayes

Dilemma at the movies

This person dropped their ticket in the hallway.

Do you call out

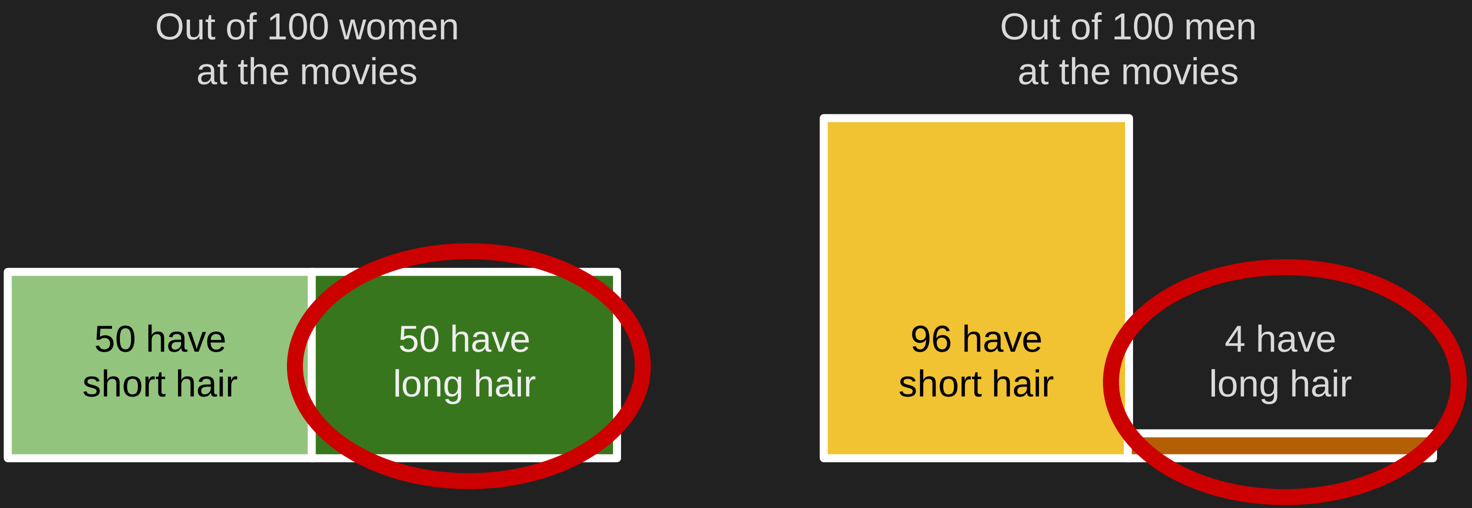
> “Excuse me, ma’am!”

or

> “Excuse me, sir!”

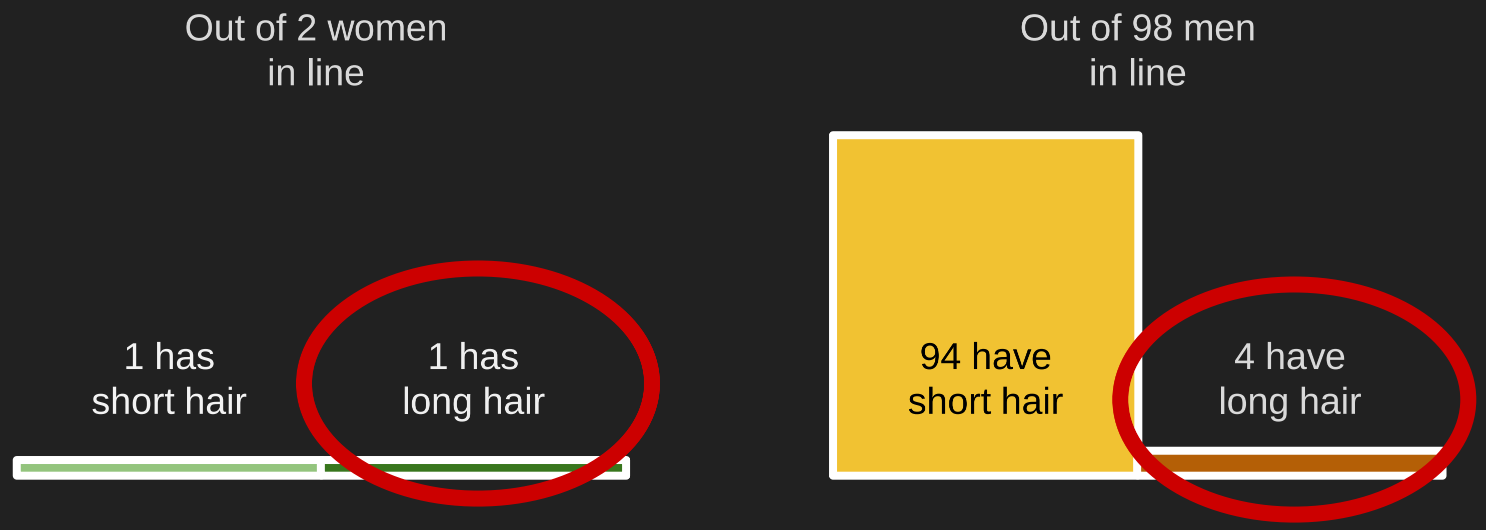
You have to make a guess.

Put numbers to our dilemma



About 12 times more women have long hair than men.

Put numbers to our dilemma



In the line, 4 times more men have long hair than women.

Conditional probabilities

P(long hair | woman)

If I know that a person is a woman, what is the probability that person has long hair?

P(long hair | woman)

= # women with long hair / # women

= 25 / 50 = **.5**

Out of 100 people
at the movies

50 are women

Conditional probabilities

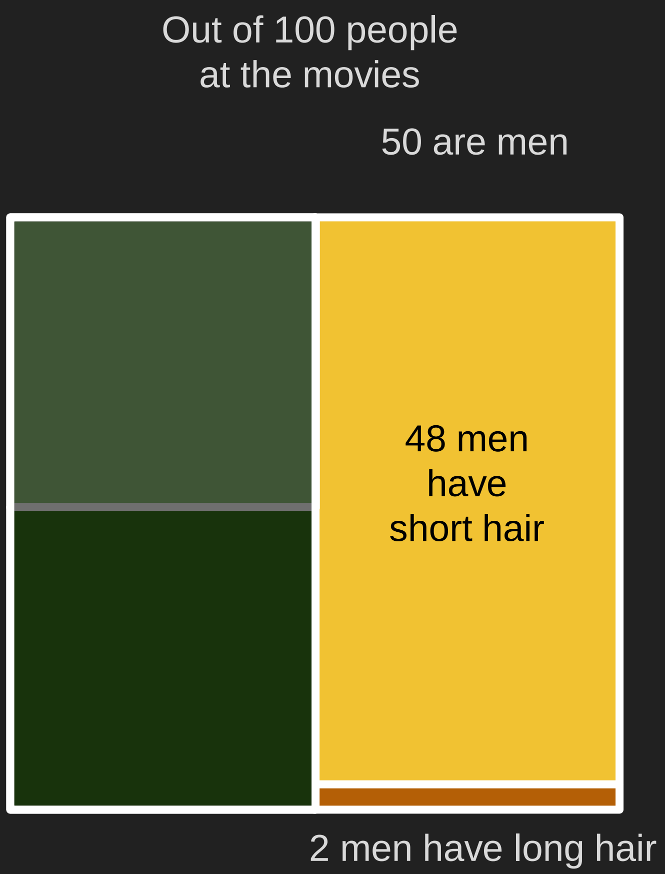
If I know that a person is a man, what is the probability that person has long hair?

P(long hair | man)

= # men with long hair / # men

= 2 / 50 = **.04**

Whether in line or not.



Joint probabilities (∩)

P(A and B) is the probability that both A and B are the case.

Also written P(A, B) or P(A ∩ B)

P(A + B) = P(A) * P(B)

P(A and B) is the same as P(B and A)

The probability that I am having a jelly donut with my milk is the same as the probability that I am having milk with my jelly donut.

P(donut and milk) = P(milk and donut)

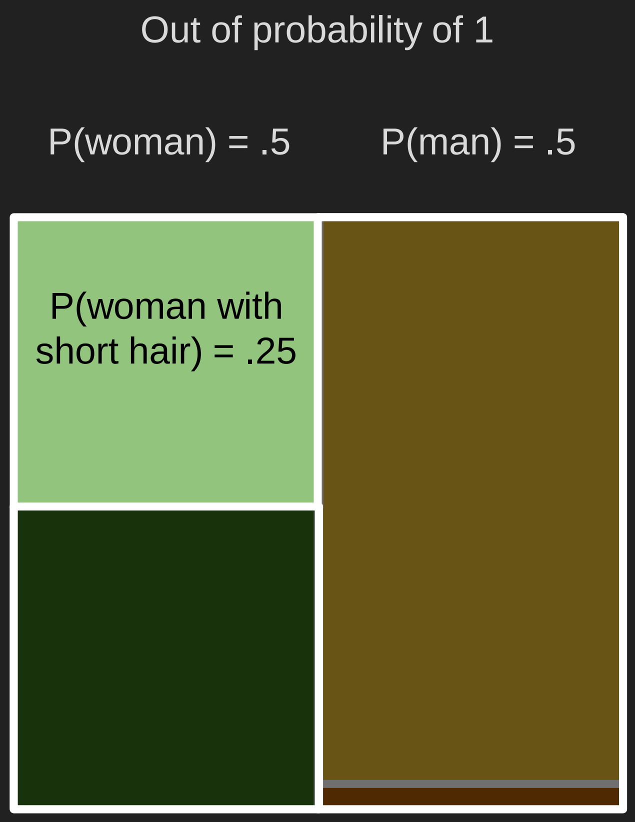
Joint (∩) probabilities

What is the probability that a person is both a woman and has short hair?

P(woman with short hair)

= P(woman) * P(short hair | woman)

= .5 * .5 = **.25**



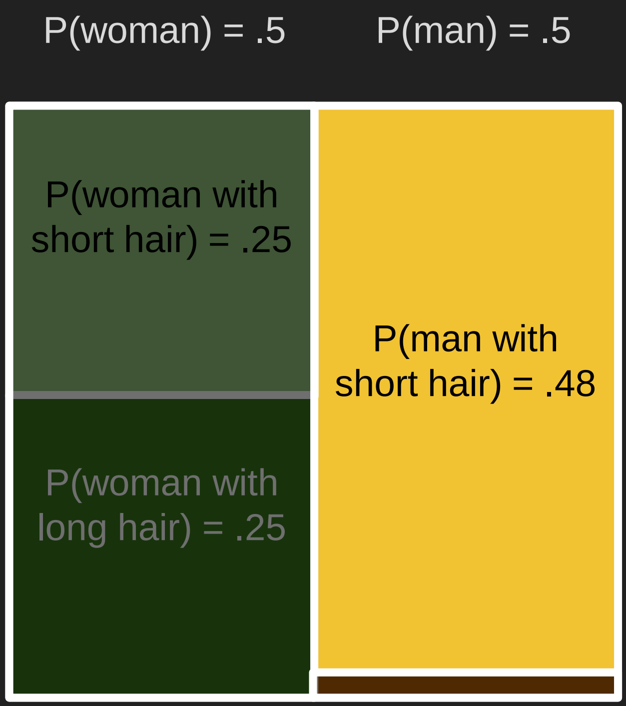
Joint (∩) probabilities

P(man with short hair)

= P(man) * P(short hair | man)

= .5 * .96 = **.48**

Out of probability of 1



Marginal probabilities (∪)

P(A or B) is the probability that either A or B is the case

Also written P(A | B) or P(A ∪ B)

P(A | B) = P(A) + P(B) – P(a∩b)

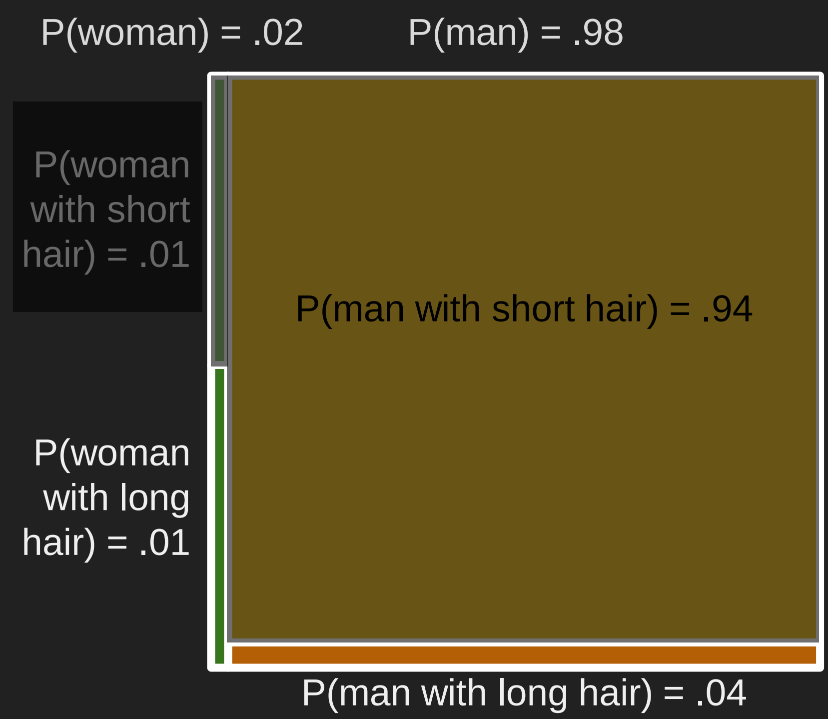
Marginal probabilities

P(long hair) = P(woman with long hair) +

P(man with long hair)

= .01 + .04 = **.05**

Out of probability of 1



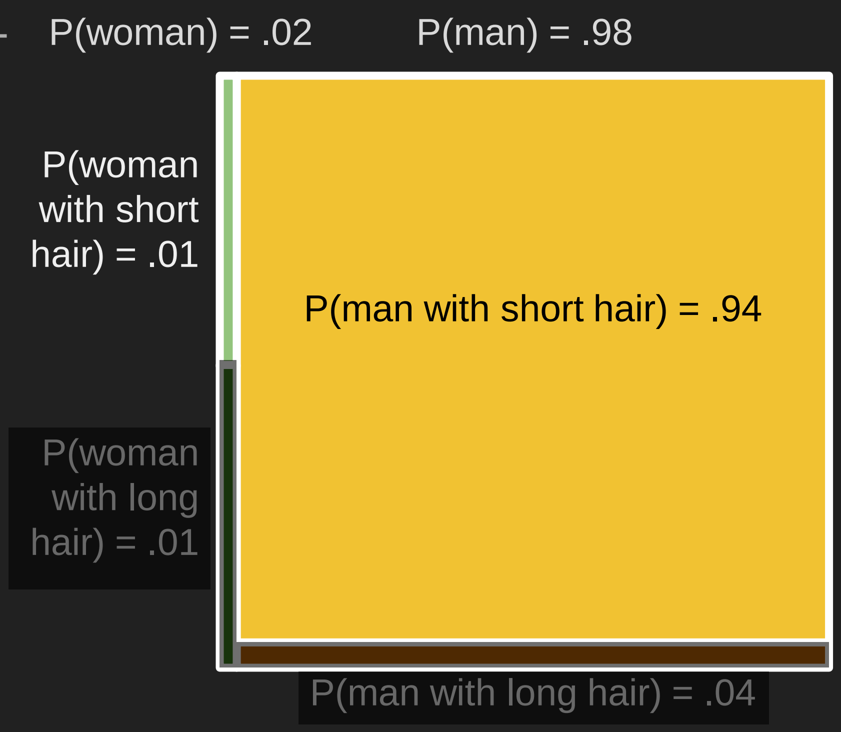
Marginal probabilities

Out of probability of 1

P(short hair) = P(woman with short hair) +

P(man with short hair)

= .01 + .94 = **.95**



What we really care about

We know the person has long hair.
Are they a **man** or a **woman**?

P(man | long hair)

We don't know this answer yet.

Thomas Bayes noticed something cool

P(man with long hair) = P(long hair) * P(man | long hair)

P(long hair and man) = P(man) * P(long hair | man)

Because P(man and long hair) = P(long hair and man)

P(long hair) * P(man | long hair) = P(man) * P(long hair | man)

Thomas Bayes noticed something cool

P(man with long hair) = P(long hair) * P(man | long hair)

P(long hair and man) = P(man) * P(long hair | man)

Because P(man and long hair) = P(long hair and man)

P(long hair) * P(man | long hair) = P(man) * P(long hair | man)

P(man | long hair) = P(man) * P(long hair | man) / P(long hair)

Thomas Bayes noticed something cool

P(man with long hair) = P(long hair) * P(man | long hair)

P(long hair and man) = P(man) * P(long hair | man)

Because P(man and long hair) = P(long hair and man)

P(long hair) * P(man | long hair) = P(man) * P(long hair | man)

P(man | long hair) = P(man) * P(long hair | man) / P(long hair)

P(A | B) = P(B | A) * P(A) / P(B)

Bayes' Theorem

$$P(A \mid B) = \frac{P(B \mid A)\ P(A)}{P(B)}$$

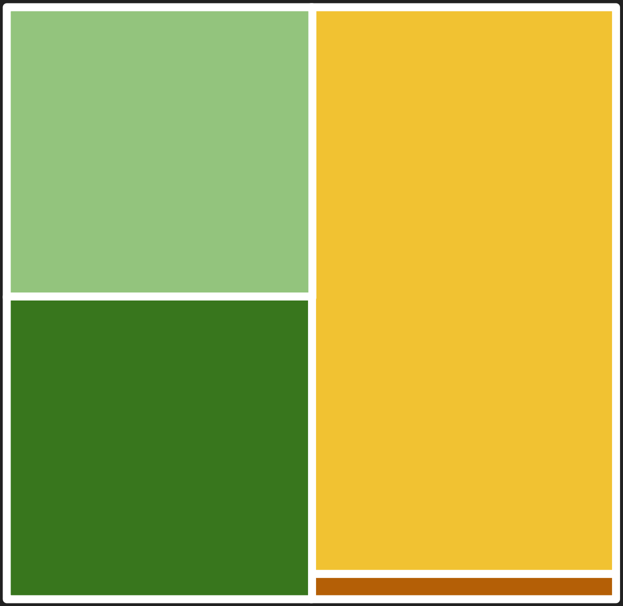
Back to the movie theater, this time with Bayes

$$P(\text{man} \mid \text{long hair}) = \frac{P(\text{man}) * P(\text{long hair} \mid \text{man})}{P(\text{long hair})}$$

$$= \frac{P(\text{man}) * P(\text{long hair} \mid \text{man})}{P(\text{woman with long hair}) + P(\text{man with long hair})}$$

$$P(\text{man} \mid \text{long hair}) = \frac{.5 * .04}{.25 + .02} = .02 / .27 = \boxed{.07}$$

P(woman) = .5 P(man) = .5



P(long hair | man) = .04
P(long hair | woman) = .5

Back to the bathroom line, this time with Bayes

$$P(\text{man} \mid \text{long hair}) = \frac{P(\text{man}) * P(\text{long hair} \mid \text{man})}{P(\text{long hair})}$$

$$= \frac{P(\text{man}) * P(\text{long hair} \mid \text{man})}{P(\text{woman with long hair}) + P(\text{man with long hair})}$$

$$P(\text{man} \mid \text{long hair}) = \frac{.98 * .04}{.01 + .04} = .04 / .05 = \boxed{.80}$$

P(woman) = .02 P(man) = .98

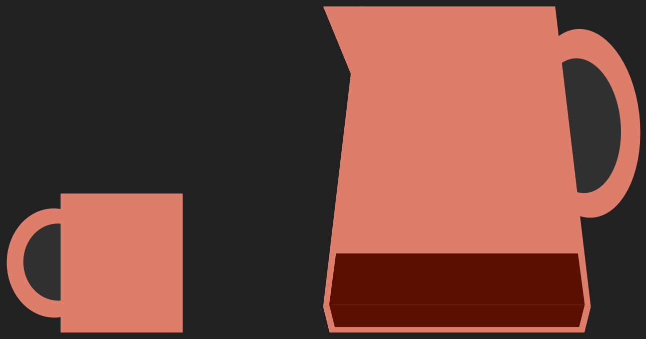
P(long hair | man) = .04
P(long hair | woman) = .5

Conclusion

- You update your belief based on ***evidence***

Probability distributions

Probability is like a pot with just one cup of coffee left in it.

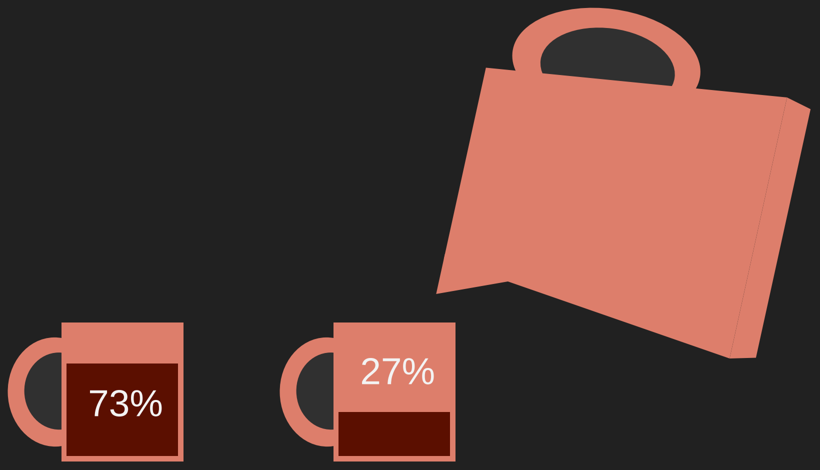


Probability distributions

If you only have one cup, you can fill it completely.

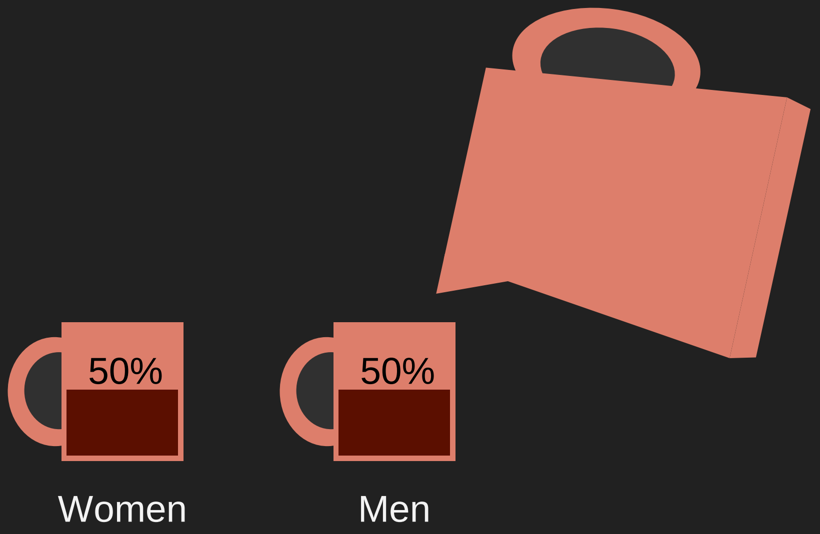
Probability distributions

If you have two cups, you have to decide how to share (distribute) it.



Probability distributions

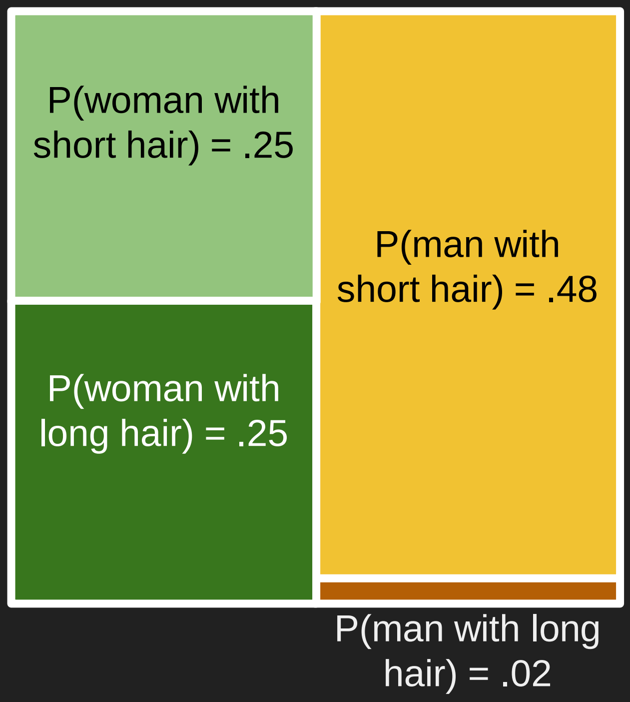
Our people are distributed between two groups, women and men.



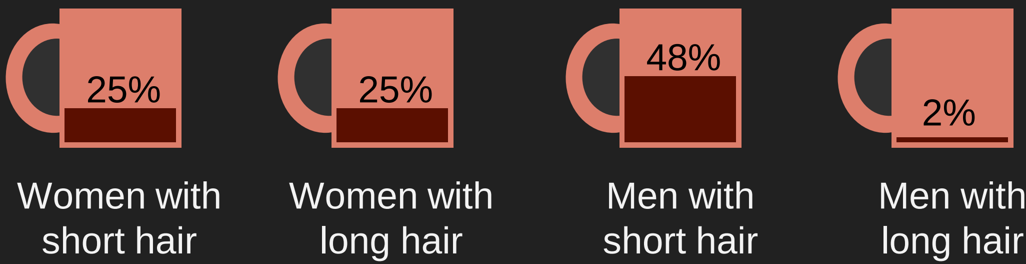
Probability distributions

We can distribute them more.

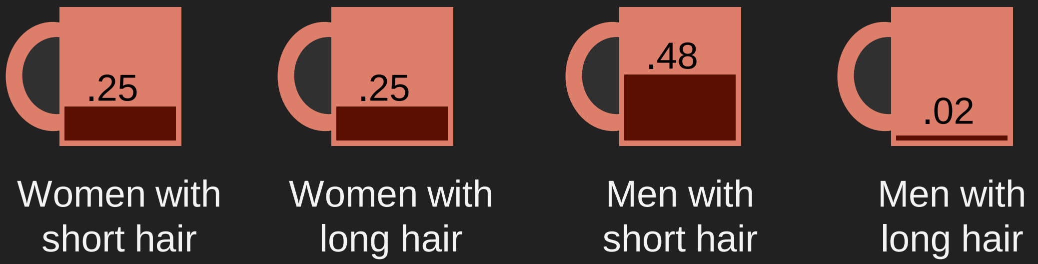
Probability distributions



Probability distributions

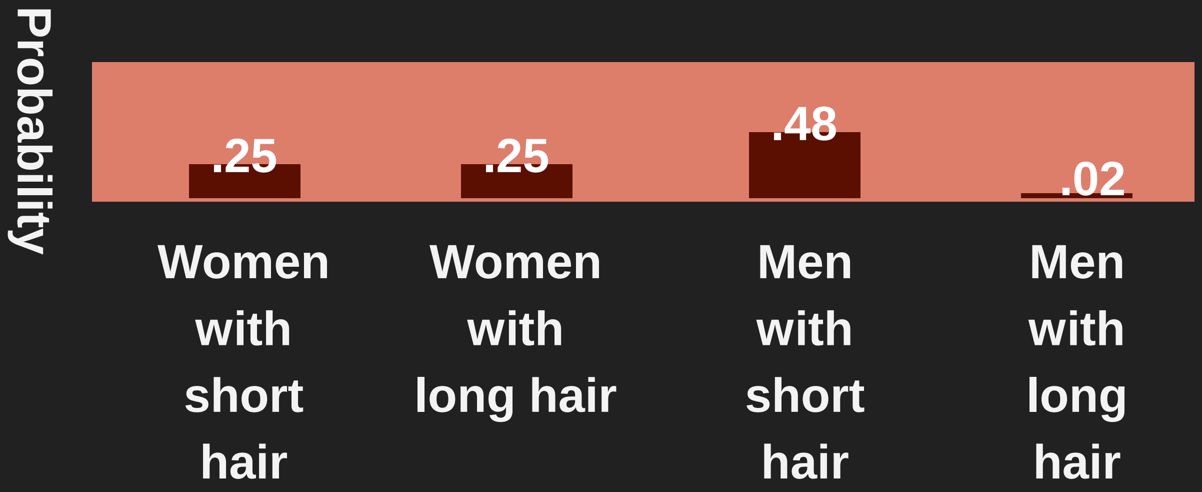


Probability distributions



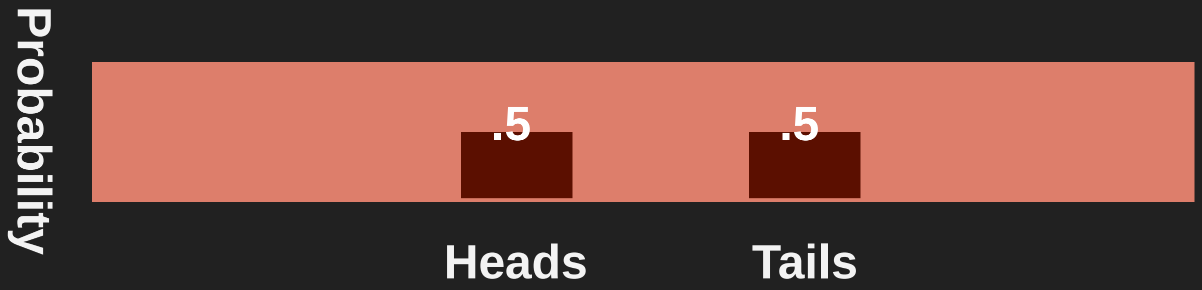
Probability distributions

It’s helpful to think of probabilities as beliefs



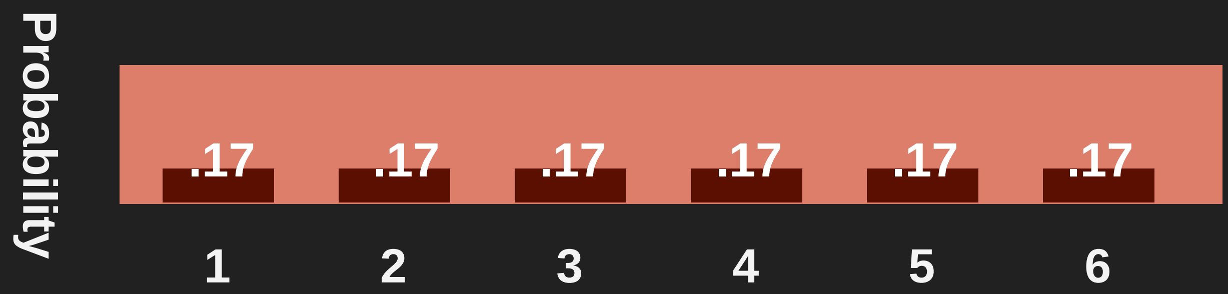
Probability distributions

Flipping a fair coin



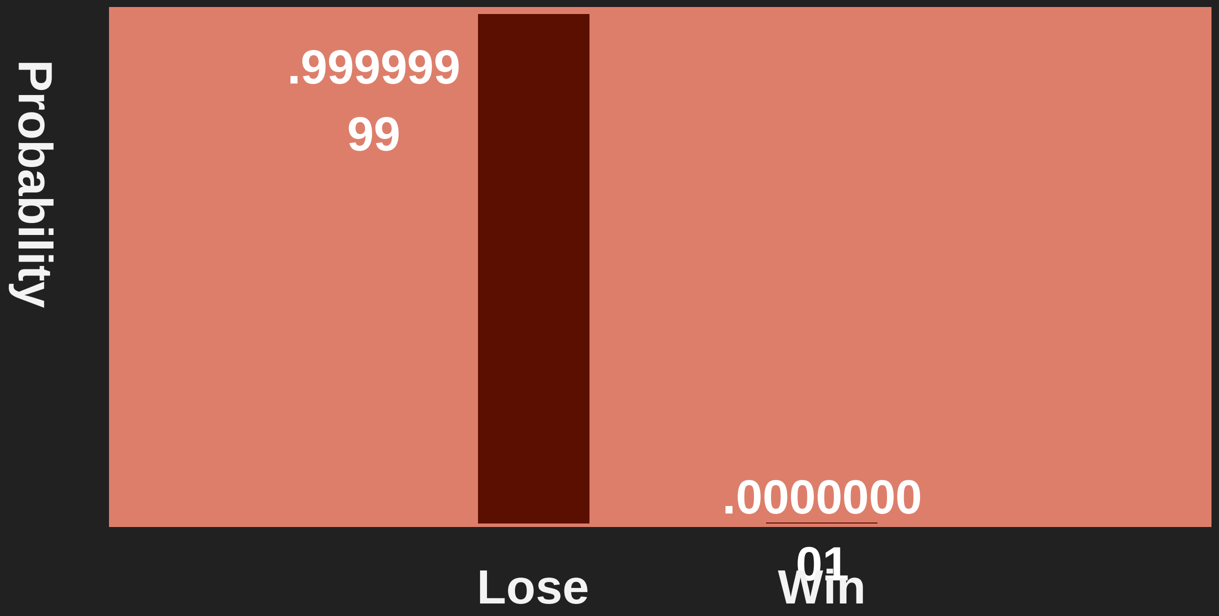
Probability distributions

Rolling a fair die



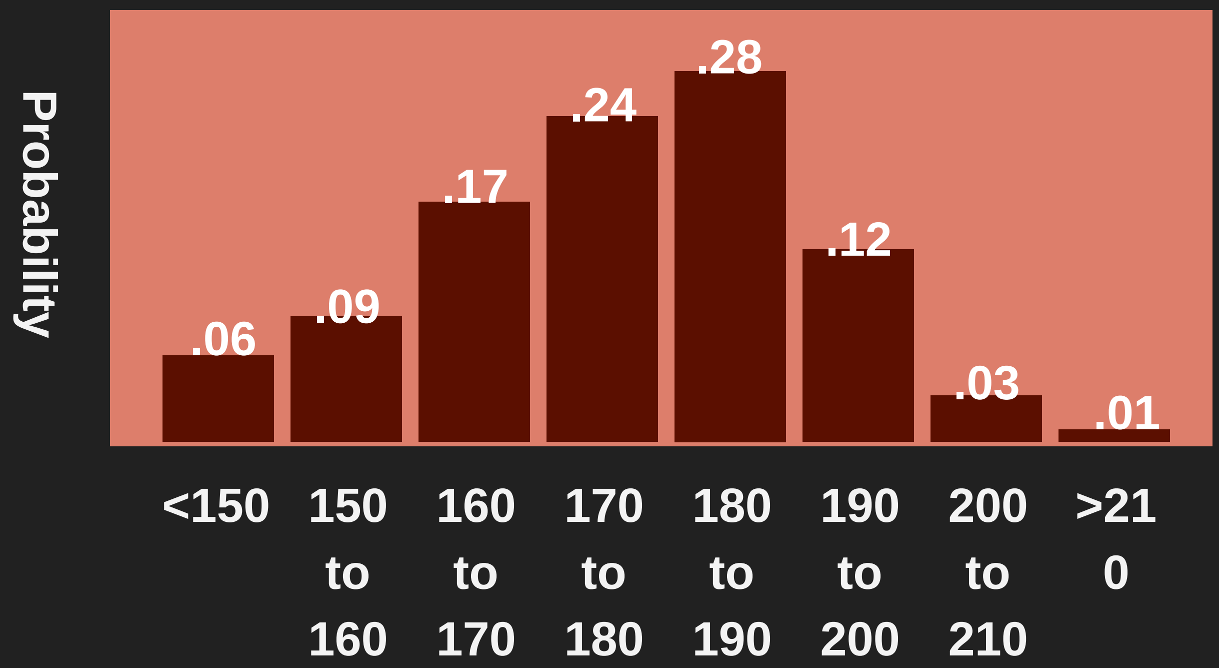
Probability distributions

Playing for the Powerball jackpot



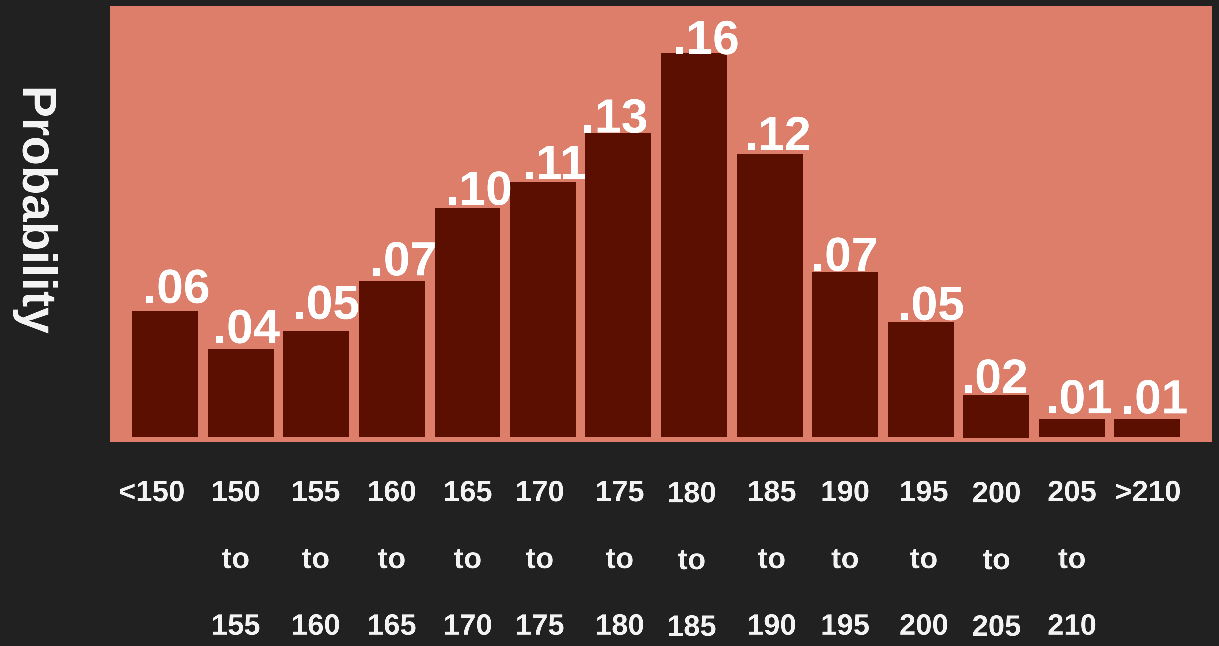
Probability distributions

Height of adults in cm



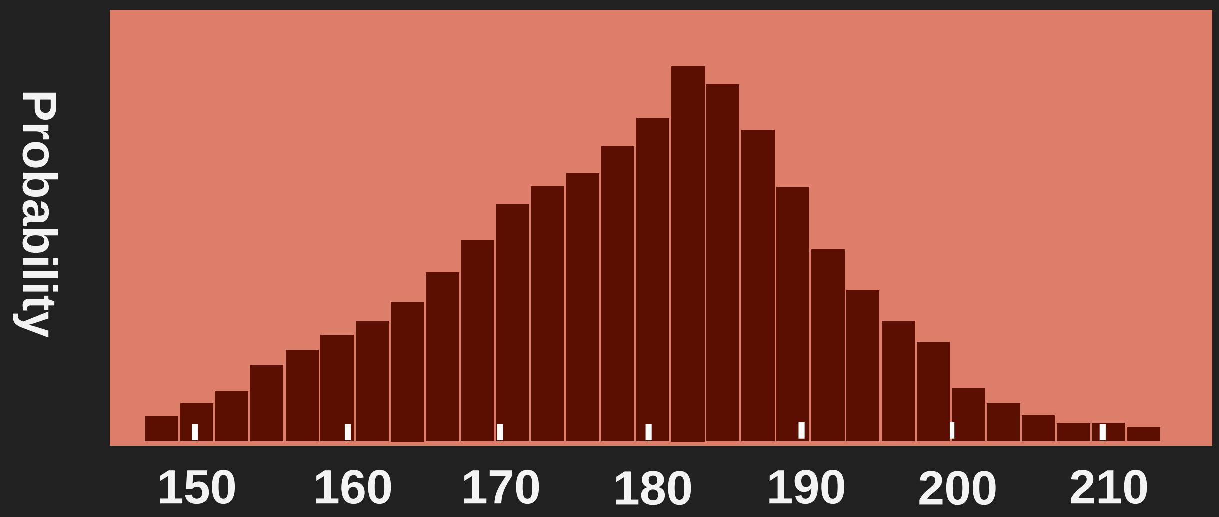
Probability distributions

Height of adults in cm



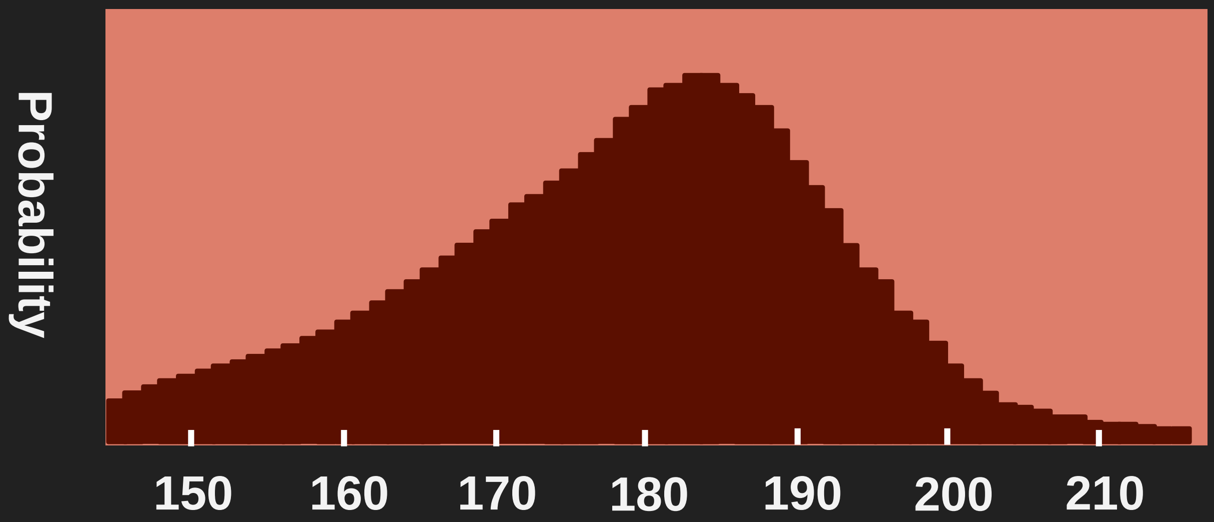
Probability distributions

Height of adults in cm



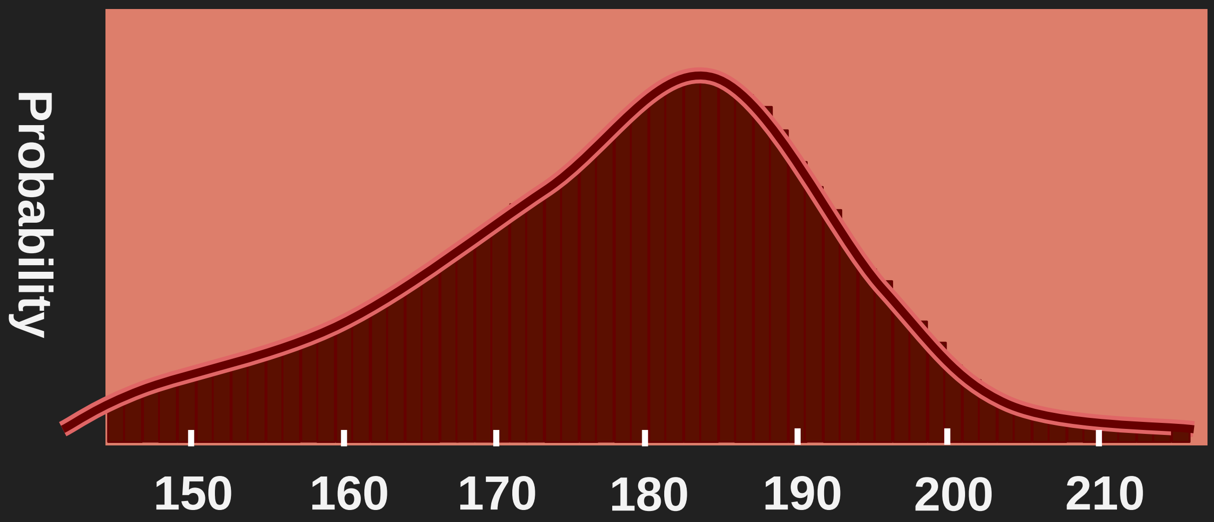
Probability distributions

Height of adults in cm



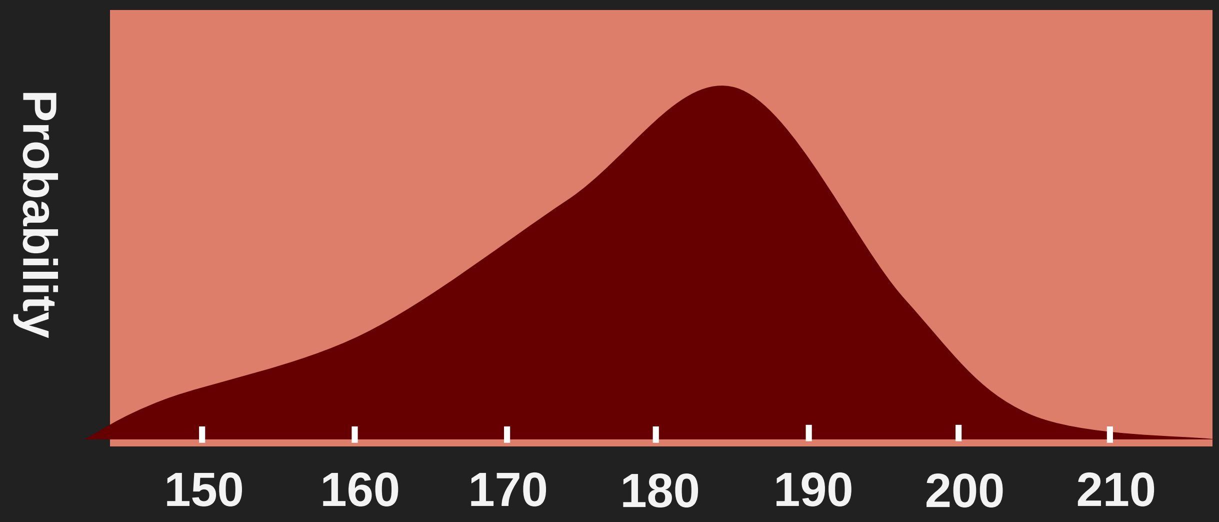
Probability distributions

Height of adults in cm



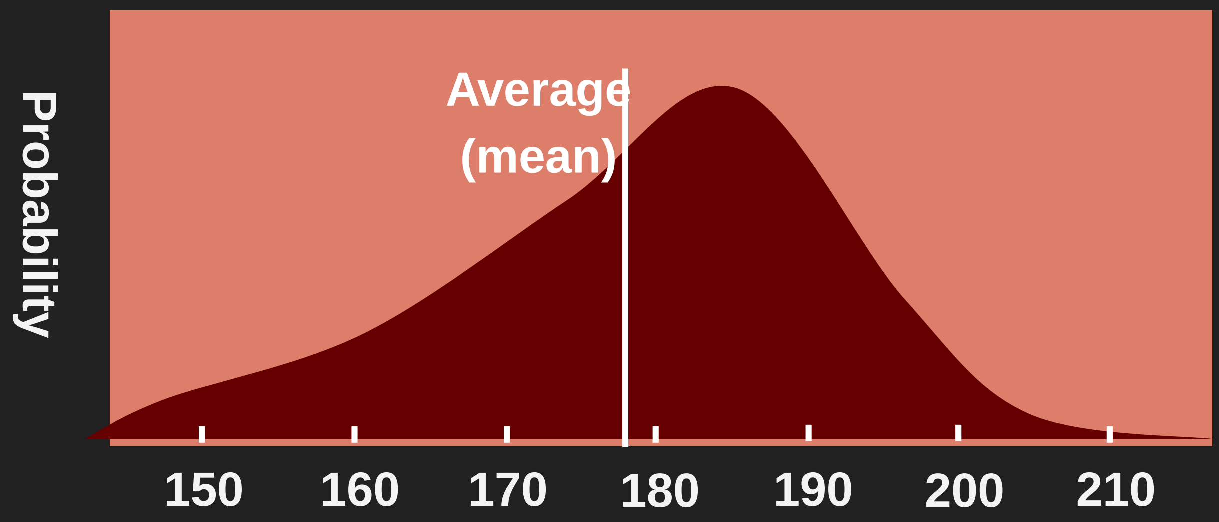
Probability distributions

Height of adults in cm



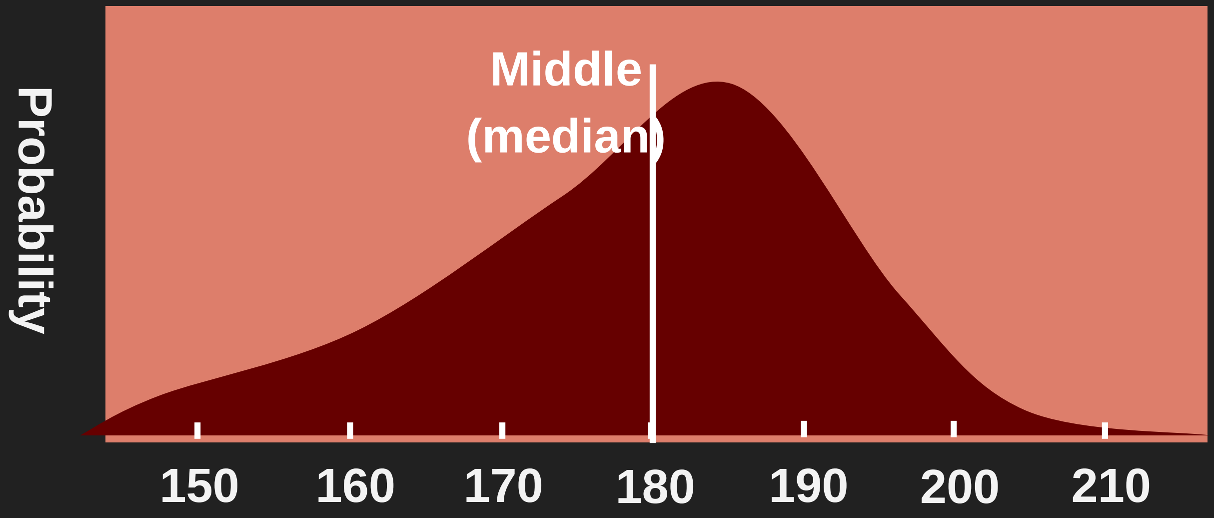
Probability distributions

Height of adults in cm



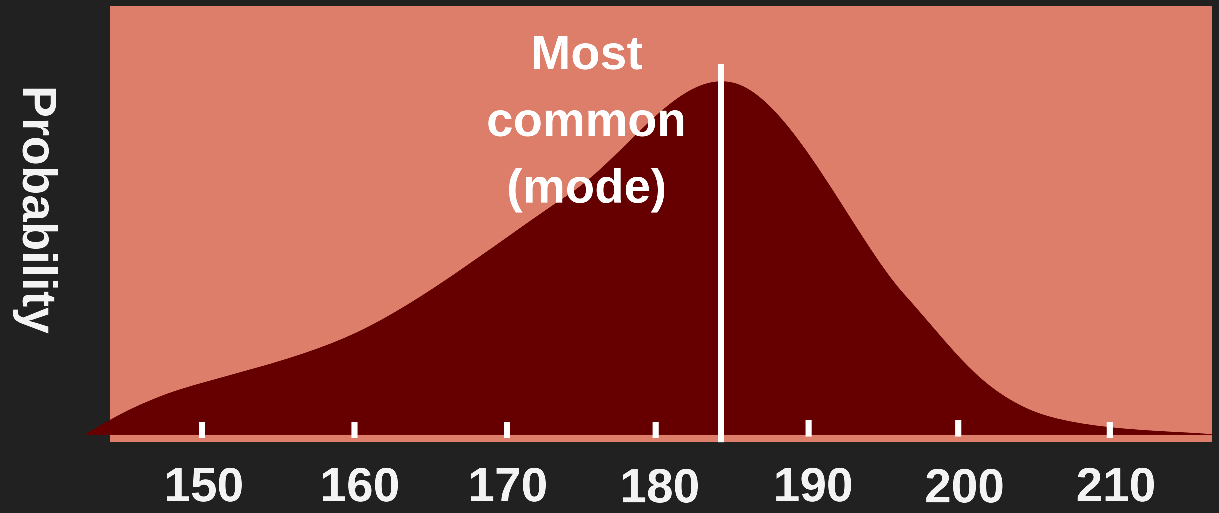
Probability distributions

Height of adults in cm



Probability distributions

Height of adults in cm



Bayes' Theorem

$$P(w \mid m) = \frac{P(m \mid w)\ P(w)}{P(m)}$$

Bayes' Theorem

prior

$$P(w \mid m) = \frac{P(m \mid w)\ \boxed{P(w)}}{P(m)}$$

Bayes’ Theorem

likelihood

$$P(w \mid m) = \frac{\boxed{P(m \mid w)}\ P(w)}{P(m)}$$

Bayes' Theorem

posterior

$$\boxed{P(w \mid m)} = \frac{P(m \mid w)\ P(w)}{P(m)}$$

Bayes' Theorem

$$P(w \mid m) = \frac{P(m \mid w)\ P(w)}{\boxed{P(m)}}$$

marginal likelihood

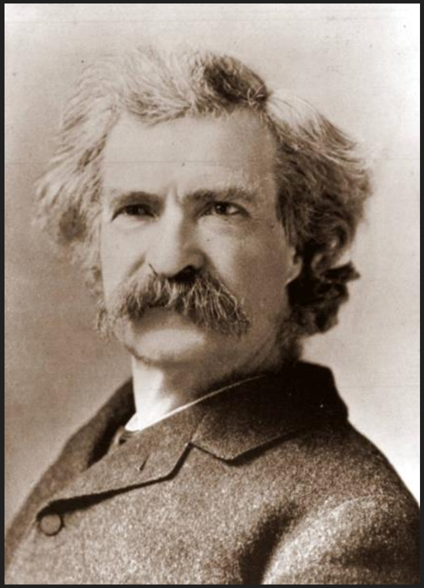
Why Bayesian inference makes us nervous

We’re not always aware of what we believe.

Putting what we believe into a distribution correctly is tricky.

We want to be able to be surprised by our data.
Inaccurate beliefs can make it hard or impossible to learn.

> “It ain't what you don't know that gets you into trouble.
> It's what you know for sure that just ain't so.”
>
> \- Mark Twain



Believe the impossible, at least a little bit

Leave room for believing the unlikely. Leave a non-zero probability unless you are absolutely certain.

> "When you have excluded the impossible, whatever remains, however improbable, must be the truth"
>
> - Sherlock Holmes (Sir Arthur Conan Doyle)

Believe the impossible, at least a little bit

"Alice laughed: "There's no use trying," she said; "one can't believe impossible things."

"I daresay you haven't had much practice," said the Queen. "When I was younger, I always did it for half an hour a day. Why, sometimes I've believed as many as six impossible things before breakfast."

- Lewis Carroll (Alice’s Adventures in Wonderland)


CLASS
DISMISSED
iStockphoto