Northeastern University

INFO 6105
Data Sci Eng Tools & Mthds
Lecture 4 Statistics and Data Science

Review Probabilities and Bayesian Statistics

28 January 2019

Starting point of probability theory

- Given the probabilities of *two* events, the probability of *both* at the same time is:

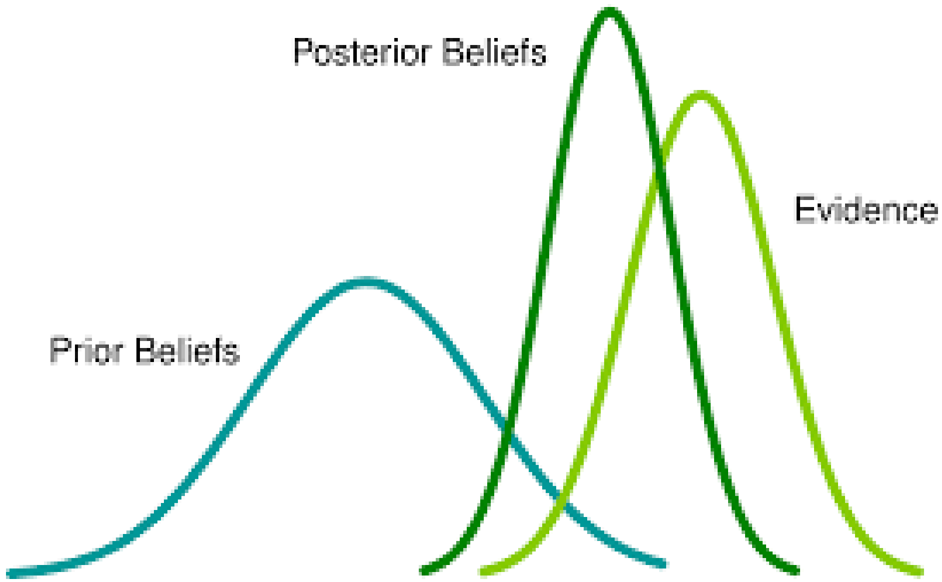
$$P(a, b) = P(a) + P(b) - P(a \cap b)$$

- Given the probabilities of two events, the probability of one event *after* the other is:

$$P(b|a) = P(a) * P(b)$$

Probabilities Exercise

- What is the probability that ***someone*** in this classroom shares ***your*** birthday?
- What is the probability that ***two students*** in this classroom have the same birthday?



DATA SCIENCE & BAYESIAN STATISTICS

Philosophy of Bayesian Inference

- You are a skilled programmer, but bugs still slip into your code. After a particularly difficult implementation of an algorithm, you decide to test your code on a trivial example. It passes. You test the code on a harder problem. It passes once again. And it passes the next, *even more difficult*, test too! You are starting to believe that there may be no bugs in this code...
- **If you think this way, then congratulations, you already are thinking *Bayesian*!**
  - **Bayesian inference is simply *updating your beliefs* after considering *new evidence***
  - **A Bayesian can rarely be certain about a result, but he or she can be very confident**
- We can never be 100% sure that our code is bug-free unless we test it *on every possible problem*
  - Instead, we can test it on a *large* number of problems, and if it succeeds we can feel more *confident* about our code, but still not certain
  - Bayesian inference works identically: We update our beliefs about an outcome based on evidence

Frequentist statistics

- ..is the more *classical, non-Bayesian* version of statistics
- Assumes that probability is the long-run frequency of events
  - For example, the *probability of plane accidents* under a frequentist philosophy is interpreted as *long-term frequency of plane accidents*
  - *The probability of bugs in your code is the number of buggy SLOCs over total number of SLOCs you write <u>in your lifetime</u>*
- This makes logical sense for many probabilities of events, but becomes more difficult to understand when events have no long-term frequency of occurrences
  - We often assign probabilities to outcomes of presidential elections, but the election itself only happens once!
  - Frequentists get around this by invoking *alternative realities* and saying across all these realities, the frequency of occurrences defines the probability. Like Star Trek parallel universes. Yuck!
- Frequentist methods are still useful in some areas

Dilemma at the movies

This person dropped their ticket in the hallway.

Do you call out

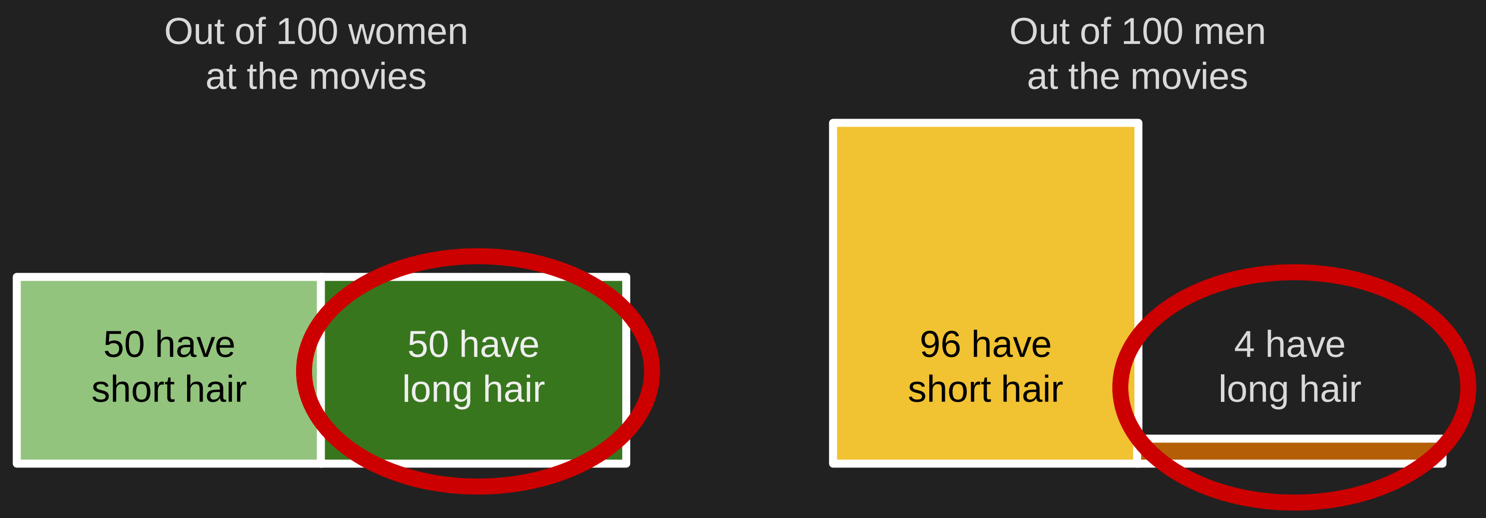
“Excuse me, ma’am!”

or

“Excuse me, sir!”

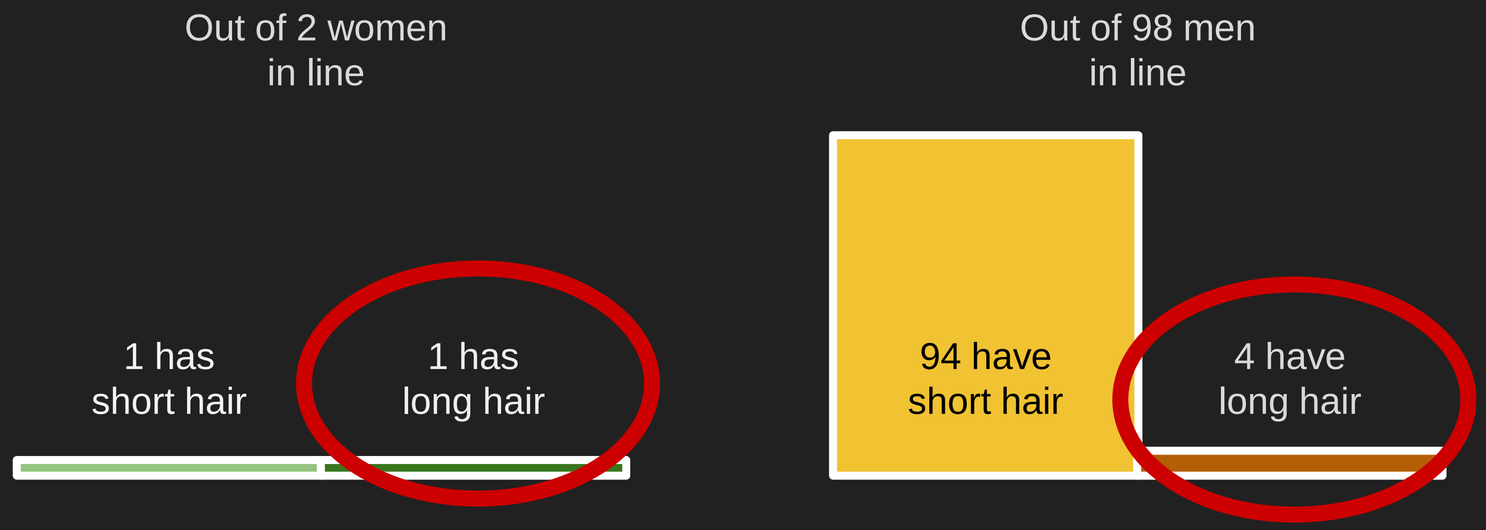
You have to make a guess.

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)

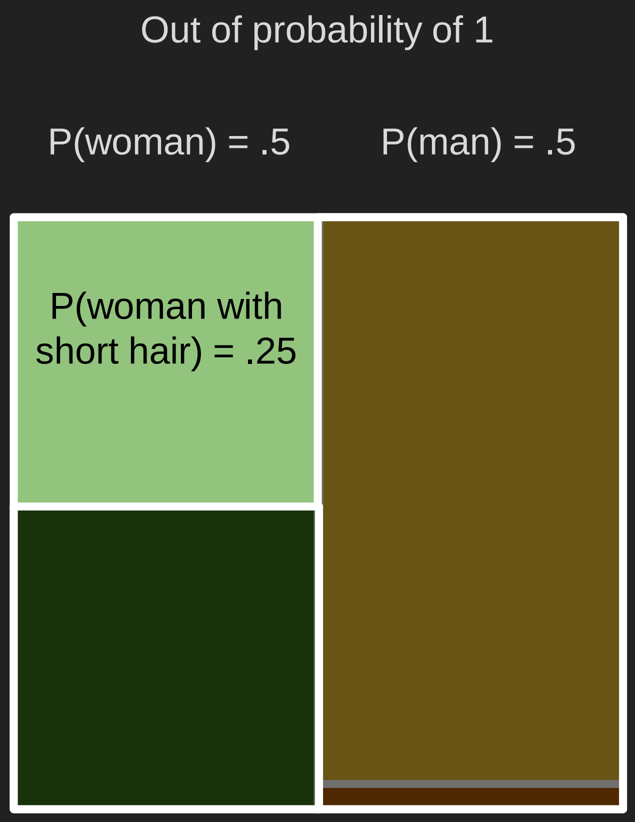
Joint (∩) probabilities

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**

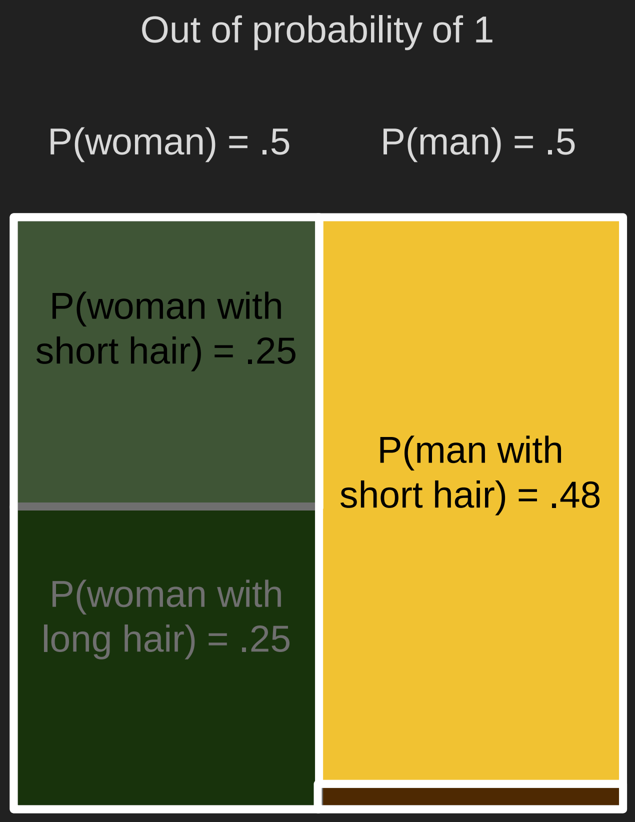


Joint (∩) probabilities

P(man with short hair)

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

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



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)

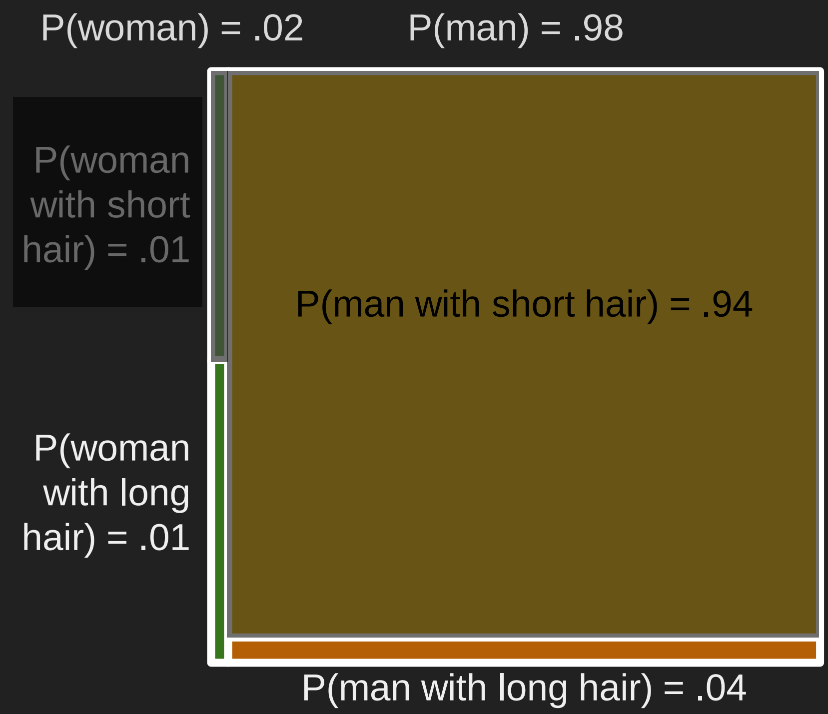
Marginal probabilities

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

P(man with long hair)

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

Out of probability of 1



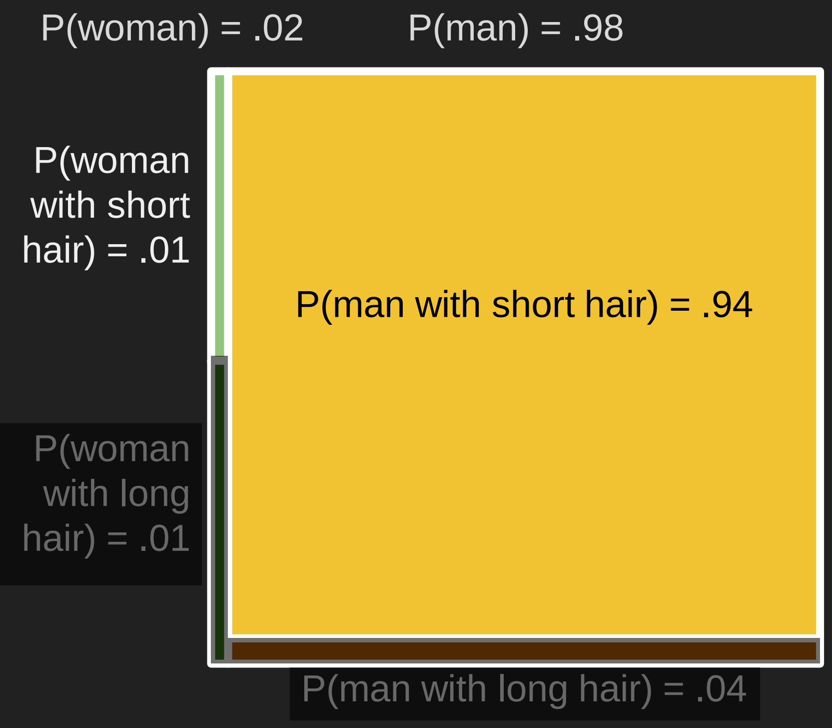
Marginal probabilities

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

P(man with short hair)

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

Out of probability of 1



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)

$\boxed{\text{P(long hair)}}$ * P(man | long hair) = P(man) * P(long hair | man)

P(man | long hair) = P(man) * P(long hair | man) / $\boxed{\text{P(long hair)}}$

Bayes' Theorem

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

Back to the movie theater, this time with Bayes

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

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

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

Back to the bathroom line, this time with Bayes

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

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

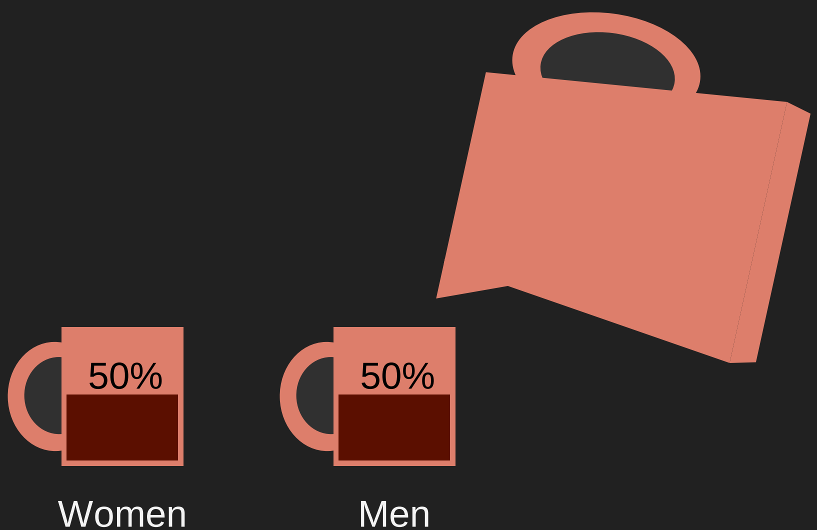
$$P(\text{man} \mid \text{long hair}) = \frac{.98 * .04}{.01 + .04} = .04 / .05 = \boxed{.80}$$

Conclusion

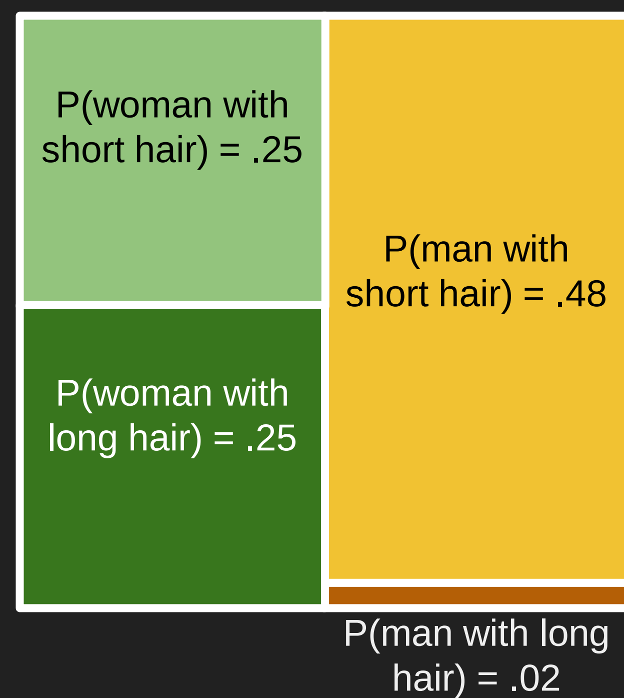
- You update your belief based on ***evidence***

Probability distributions

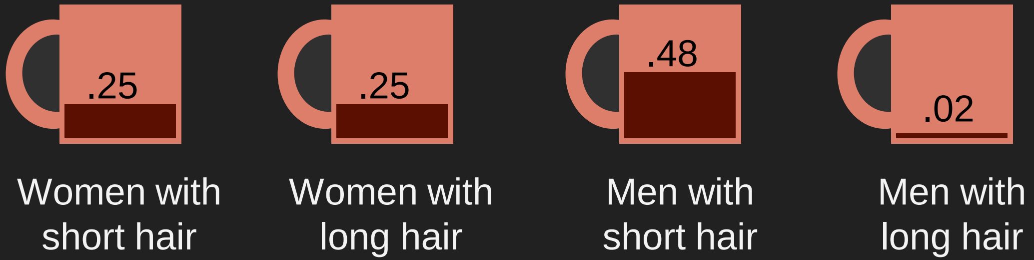
Our people are distributed between two groups, women and men.



Probability distributions

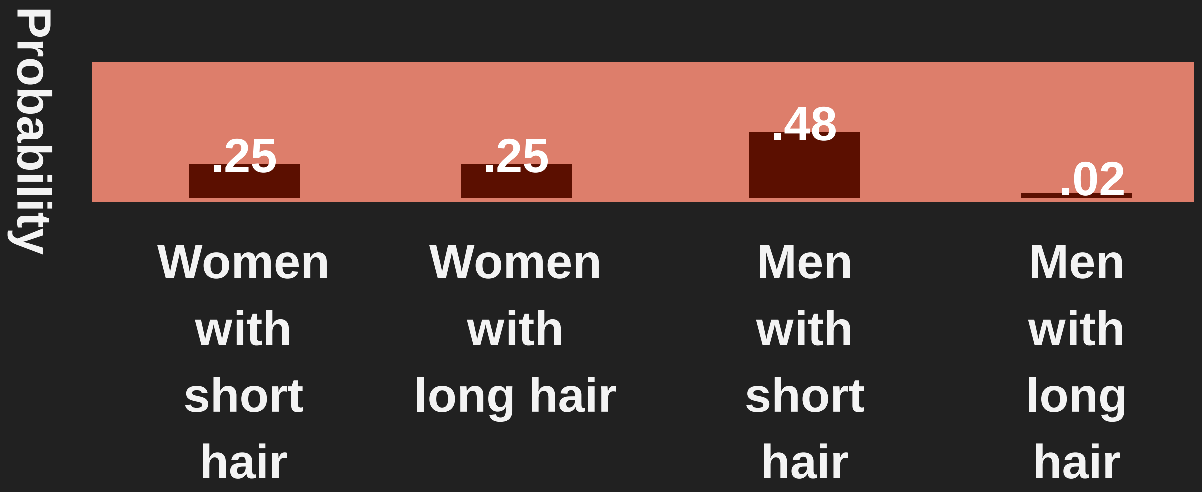


Probability distributions



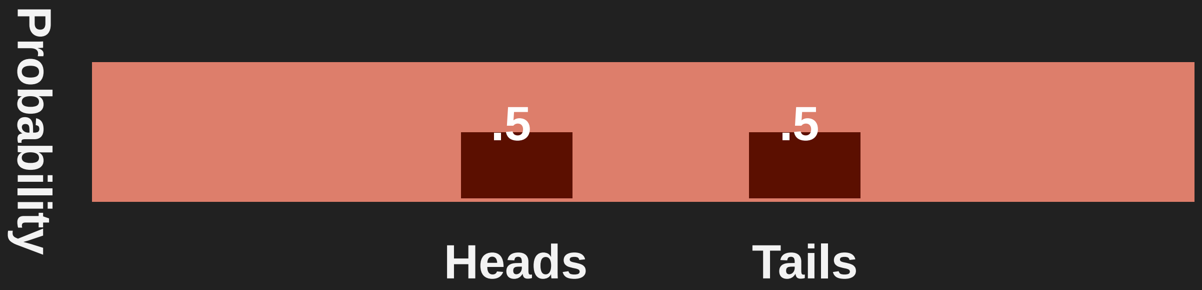
Probability distributions

It’s helpful to think of probabilities as beliefs



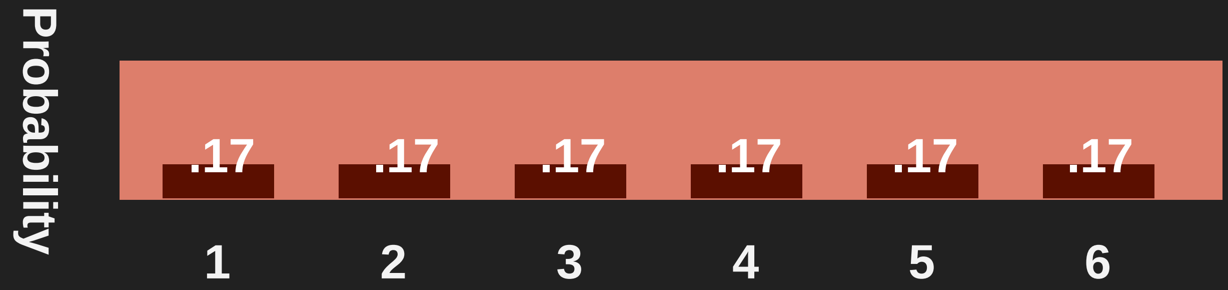
Probability distributions

Flipping a fair coin



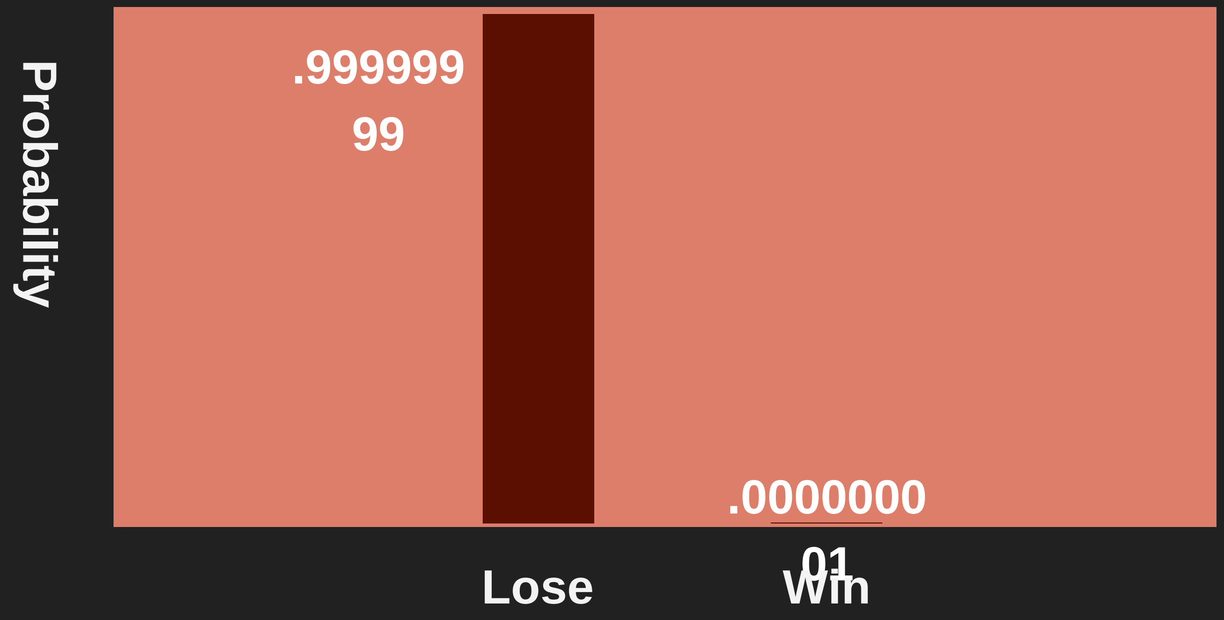
Probability distributions

Rolling a fair die



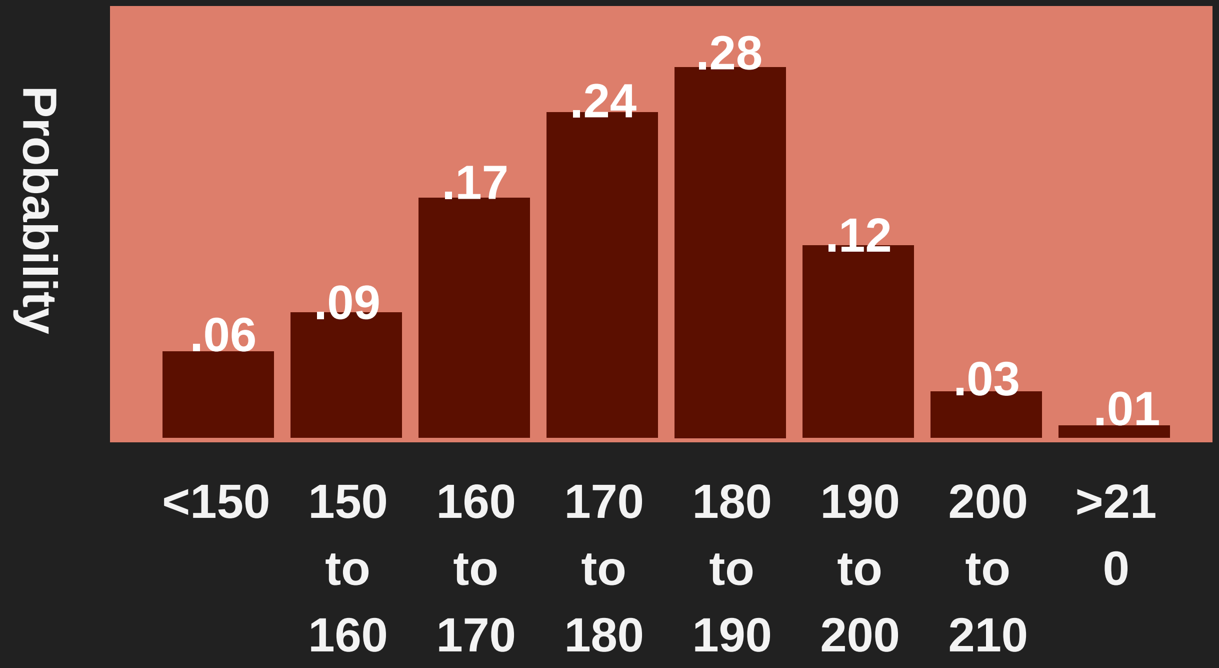
Probability distributions

Playing for the Powerball jackpot



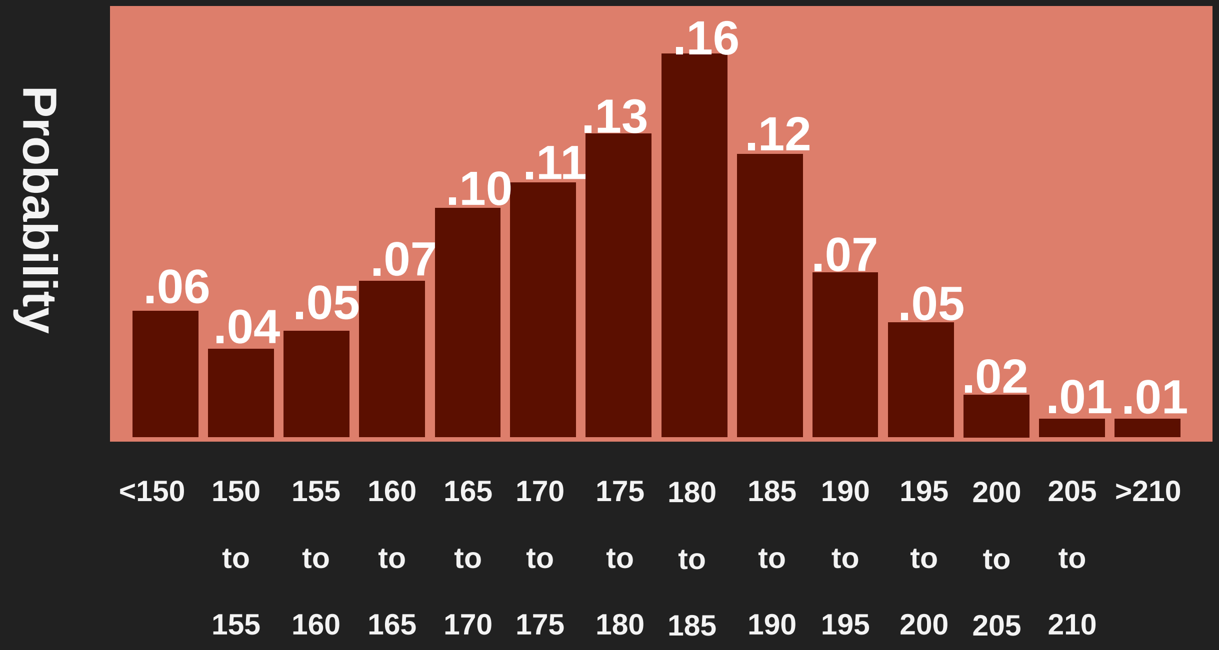
Probability distributions

Height of adults in cm



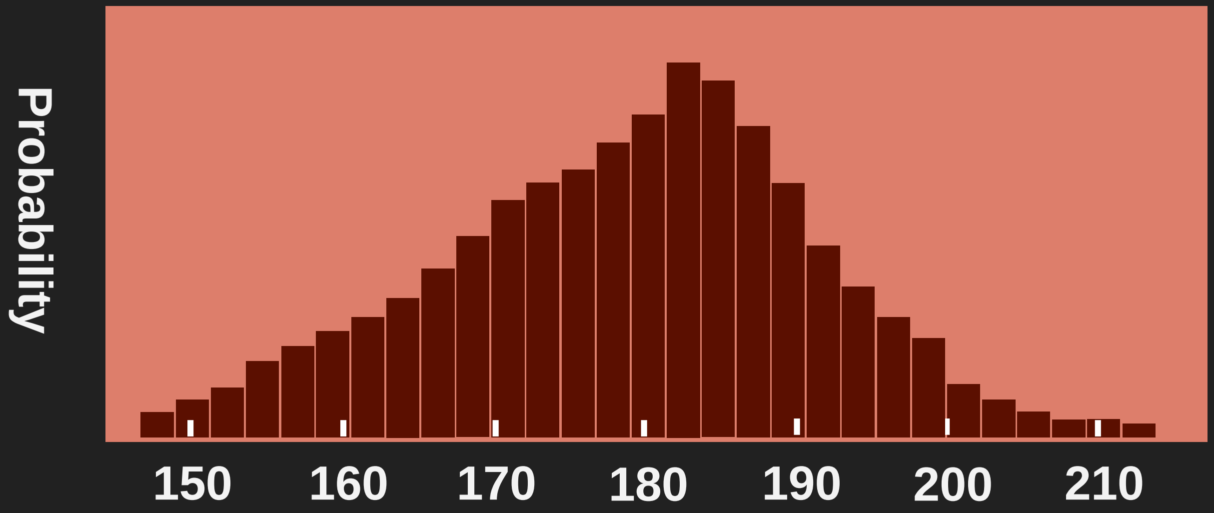
Probability distributions

Height of adults in cm



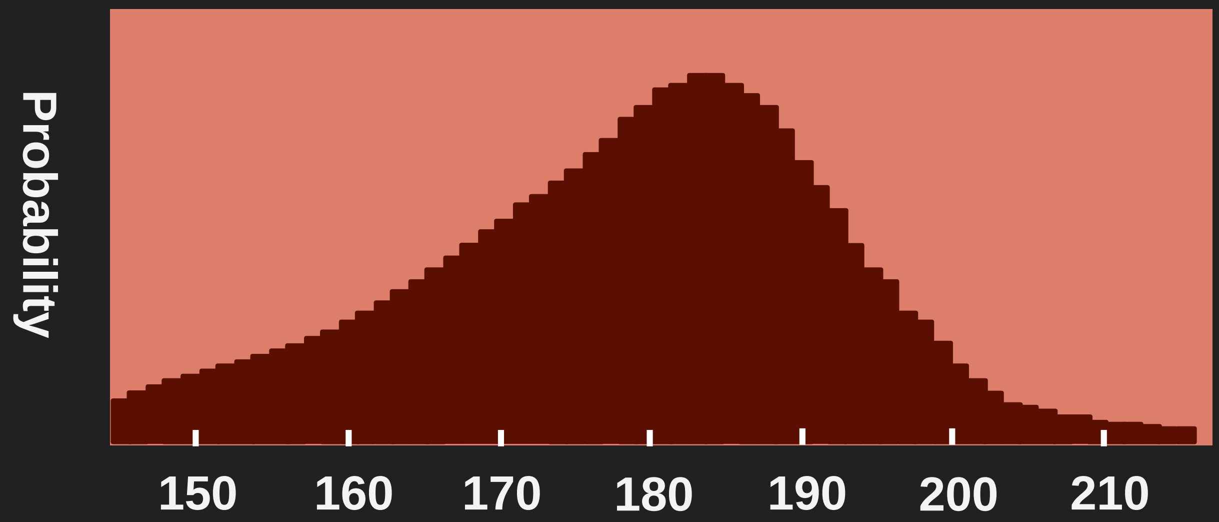
Probability distributions

Height of adults in cm



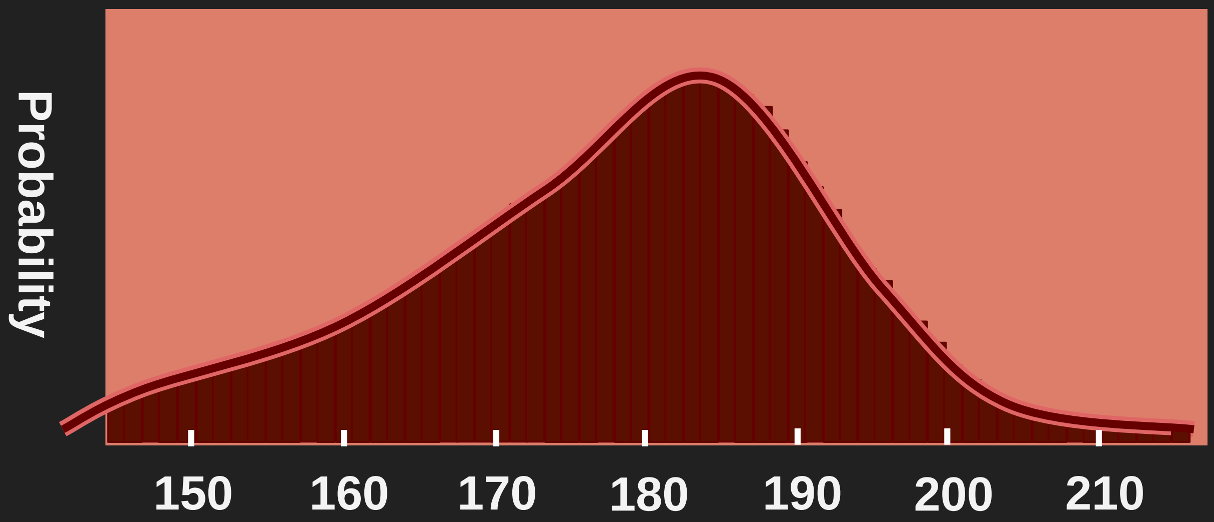
Probability distributions

Height of adults in cm



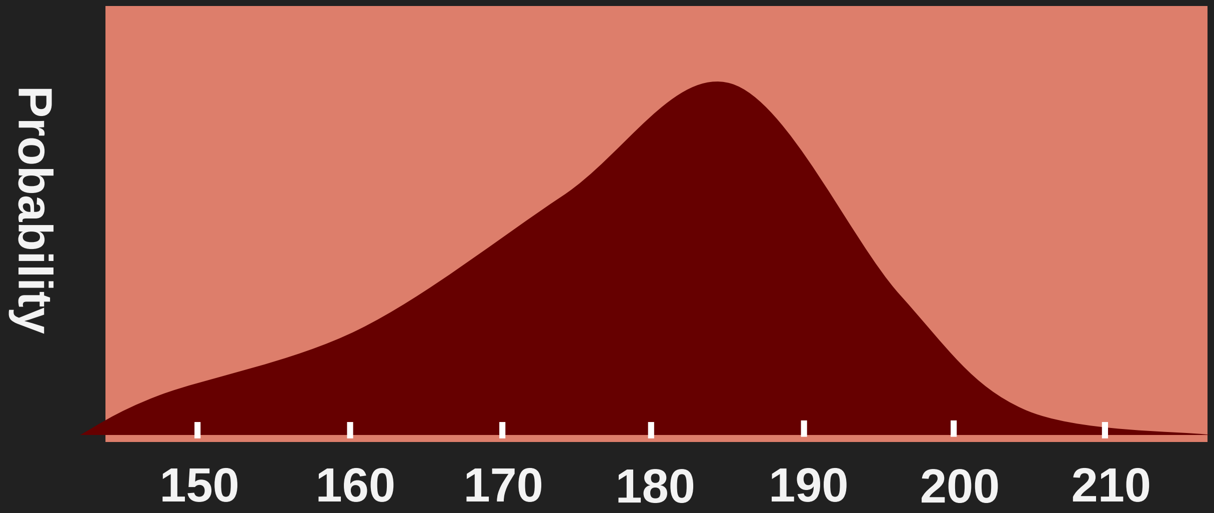
Probability distributions

Height of adults in cm



Probability distributions

Height of adults in cm



Bayes' Theorem

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

Bayes' Theorem

$$P(w \mid m) = \frac{\overbrace{\boxed{P(m \mid w)}}^{\textbf{likelihood}}\ P(w)}{P(m)}$$

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

Another example: Formula 1

- Suppose, out of all the 4 championship races (F1) between Lewis Hamilton and Fernando Alonso
  - Lewis won 3 times while Fernando 1
- So, if you were to bet on the winner of next race, who would it be?

Informative `Prior`

- It rained once when Lewis won, and once when Fernando won and it is definite that it will rain on the next date
  - So, who would you bet your money on now ?

Rainy F1

- Suppose, B be the *event of winning for Fernando*
- A be the *event of raining*
- P(A) =1/2, since it rained twice out of four days
- P(B) is 1/4, since Fernando won only one race out of four
- P(A|B)=1, since it rained every time when Fernando won
- Substituting the values in the conditional probability formula, we get:
  - P(B|A) = P(A|B)*P(B) / P(A) = 1 . ¼ . 2 = 1/2
- The probability is 50%, which is almost the double of 25% when rain was not taken into account!
- Further strengthened our belief of Fernando Alonso winning in the light of new *evidence* i.e rain
- Pretty amazing..

Why do we care about Bayesian ML?

- Because when it snows, that changes the equations of self-driving cars
  - **A lot of things can occur which we *don't have enough data on***
- Machine learning is a set of methods for creating models that describe or predicting something about the world
  - It does so by learning those models from data
- **Bayesian machine learning allows us to encode our prior beliefs about what those models should look like, independent of what the data tells us**
  - **This is especially useful when we don't have a ton of data to confidently learn our model**
- **Bayesian machine learning also yields errors on guesses, and we want to know *how sure* (any why) a machine is about something**

Bayesian ML

1. **Have many models or distributions**
2. **Specify the prior belief we have about parameters**
3. **Observe some data (evidence)**
4. **Compute posterior P(θ/D) - Probability Distribution of the model obtained after reviewing the evidence**
5. **Do this for multiple models and see which model fits best subsequent data**
6. **Keep model, throw away data**