How Bayesian inference Works

Brandon Rohrer

Bayesian inference is not magic

Bayesian inference is not incomprehensible

$$P(x_1, x_2, \cdots, x_n \mid \mu, \sigma^2) \propto \frac{1}{\sigma^n} \exp\left(-\frac{1}{2\sigma^2}\sum(x_i-\mu)^2\right) \tag{1}$$

$$P(\mu \mid \mu_0, \sigma_0^2) \propto \frac{1}{\sigma_0} \exp\left(-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\right) \tag{2}$$

$$E(z_1 \mid z_2) = E(z_1) + \frac{\cdots}{Var(z_2)}(z_2 - E(z_2)) \tag{3}$$

$$Var(z_1 \mid z_2) = Var(z_1) - \frac{Cov^2(z_1, z_2)}{Var(z_2)} \tag{4}$$

$$x = \mu + \sigma\varepsilon \qquad \varepsilon \sim \mathcal{N}(0,1)$$
$$\mu = \mu_0 + \sigma_0\delta \qquad \delta \sim \mathcal{N}(0,1)$$

$$E(x) = \mu_0 \tag{5}$$
$$\mathrm{Var}(x) = E(\mathrm{Var}(x \mid \mu)) + \mathrm{Var}(E(x \mid \mu)) = \sigma^2 + \sigma_0^2 \tag{6}$$
$$\mathrm{Cov}(x, \mu) = E(x-\mu_0)(\mu-\mu_0) = \sigma_0^2 \tag{7}$$

$$E(\mu \mid x) = \mu_0 + \frac{\sigma_0^2}{\sigma^2+\sigma_0^2}(x-\mu_0) = \frac{\sigma_0^2}{\sigma^2+\sigma_0^2}\underbrace{x}_{\text{MLE}} + \frac{\sigma^2}{\sigma^2+\sigma_0^2}\underbrace{\mu_0}_{\text{prior mean}} \tag{8}$$

$$\mathrm{Var}(\mu \mid x) = \frac{\sigma^2\sigma_0^2}{\sigma^2+\sigma_0^2} = \frac{1}{\frac{1}{\sigma_0^2}+\frac{1}{\sigma^2}} = (\tau_{\text{prior}} + \tau_{\text{data}})^{-1} \tag{9}$$

Definition 4. $1 / \sigma^2$ is usually called the *precision* and is denoted by τ

The posterior mean is usually a convex combination of the prior mean and the MLE.

The posterior precision is, in this case, the sum of the prior precision and the data precision

$$\tau_{\text{post}} = \tau_{\text{prior}} + \tau_{\text{data}}$$

We summarize our results so far:

Lemma 5. *Assume $x \mid \mu \sim \mathcal{N}(\mu, \sigma^2)$ and $\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$. Then:*

$$\mu \mid x \sim \mathcal{N}\left(\frac{\sigma_0^2}{\sigma^2+\sigma_0^2}\, x + \frac{\sigma^2}{\sigma^2+\sigma_0^2}\, \mu_0\,,\ \left(\frac{1}{\sigma_0^2}+\frac{1}{\sigma^2}\right)^{-1}\right)$$

$$\begin{aligned}
P(x_{\text{new}} \mid x, \mu, \alpha, \beta) &= \int P(x_{\text{new}} \mid x, \mu, \tau, \alpha, \beta) P(\tau \mid x, \alpha, \beta) d\tau \\
&= \int P(x_{\text{new}} \mid \mu, \tau) P(\tau \mid x, \alpha, \beta) d\tau \\
&= \int P(x_{\text{new}} \mid \mu, \tau) P(\tau \mid \alpha_{\text{post}}, \beta_{\text{post}}) d\tau
\end{aligned} \tag{17}$$

$$\tau \mid \alpha, \beta \sim \mathrm{Ga}(\alpha, \beta)$$
$$x \mid \tau, \mu \sim \mathcal{N}(\mu, \tau)$$

$$x_i \mid \mu \sim \mathcal{N}(\mu, \sigma^2) \text{ i.i.d.} \quad \Rightarrow \quad \bar{x} \mid \mu \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \tag{10}$$

$$P(x_1, x_2, \cdots, x_n \mid \mu) \propto_\mu \frac{1}{\sigma} \exp\left(-\frac{1}{2\sigma^2}\sum(x_i-\mu)^2\right)$$

Then for the posterior probability, we get

$$\begin{aligned}
P(\mu \mid x_1, x_2, \cdots, x_n) &\propto P(x_1, x_2, \cdots, x_n \mid \mu) P(\mu) \propto P(\bar{x} \mid \mu) P(\mu) \\
&\propto P(\mu \mid \bar{x})
\end{aligned} \tag{12}$$

We can now plug $\bar{x}$ into our previous result and we get:

Lemma 6. *Assume $x_i \mid \mu \sim \mathcal{N}(\mu, \sigma^2)$ i.i.d.and $\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$. Then:*

$$\mu \mid x_1, x_2, \cdots, x_n \sim \mathcal{N}\left(\frac{\sigma_0^2}{\frac{\sigma^2}{n}+\sigma_0^2}\bar{x} + \frac{\sigma^2}{\frac{\sigma^2}{n}+\sigma_0^2}\mu_0,\ \left(\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}\right)^{-1}\right)$$

Assuming μ is fixed, then the conjugate prior for σ^2 is an inverse Gamma distribution:

$$z \mid \alpha, \beta \sim \mathrm{IG}(\alpha, \beta) \qquad P(z \mid \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} z^{-\alpha-1} \exp\left(-\frac{\beta}{z}\right) \tag{13}$$

For the posterior we get another inverse Gamma:

$$\begin{aligned}
P(\sigma^2 \mid \alpha, \beta) &\propto (\sigma^2)^{-(\alpha+\frac{n}{2})-1} \exp\left(-\frac{\beta + \frac{1}{2}\sum(x_i-\mu)}{\sigma^2}\right) \\
&\propto (\sigma^2)^{-\alpha_{\text{post}}-1} \exp\left(-\frac{\beta_{\text{post}}}{\sigma^2}\right)
\end{aligned} \tag{14}$$

Lemma 7. *If $x_i \mid \mu, \sigma^2 \sim \mathcal{N}(\mu, \sigma^2)$ i.i.d.and $\sigma^2 \sim \mathrm{IG}(\alpha, \beta)$. Then:*

$$\sigma^2 \mid x_1, x_2, \cdots, x_n \sim \mathrm{IG}\left(\alpha + \frac{n}{2}, \beta + \frac{1}{2}\sum(x_i-\mu)\right)$$

If we re-parametrize in terms of precisions, the conjugate prior is a Gamma distribution.

$$\tau \mid \alpha, \beta \sim \mathrm{Ga}(\alpha, \beta) \qquad P(\tau \mid \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \tau^{\alpha-1} \exp(-\tau\beta) \tag{15}$$

And the posterior is:

$$P(\tau \mid \alpha, \beta) \propto \tau^{(\alpha+\frac{n}{2})-1} \exp\left(-\tau\left(\beta + \frac{1}{2}\sum(x_i-\mu)\right)\right) \tag{16}$$

Lemma 8. *If $x_i \mid \mu, \tau \sim \mathcal{N}(\mu, \tau)$ i.i.d.and $\tau \sim \mathrm{Ga}(\alpha, \beta)$. Then:*

$$\tau \mid x_1, x_2, \cdots, x_n \sim \mathrm{Ga}\left(\alpha + \frac{n}{2}, \beta + \frac{1}{2}\sum(x_i-\mu)\right)$$

What does “Bayesian inference” even mean?

Inference = Educated guessing

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

He wrote two books. One was about theology, and one was about probability.

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.

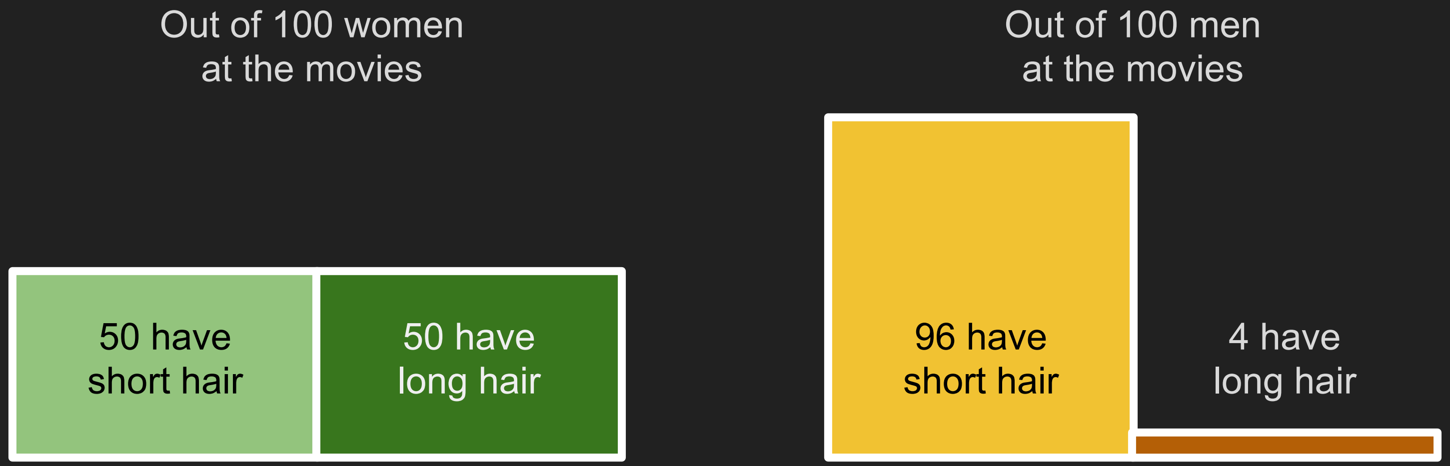
Dilemma at the movies

What if they’re standing in line for the men’s restroom?

Bayesian inference is a way to capture common sense.

It helps you use what you know to make better guesses.

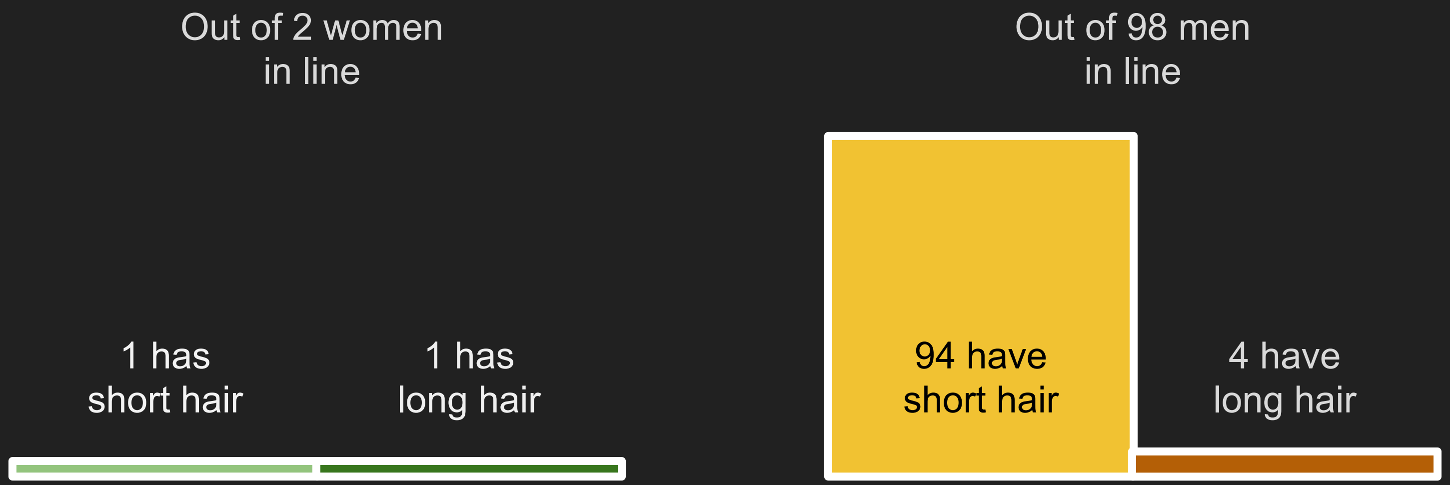
Put numbers to our dilemma



Put numbers to our dilemma

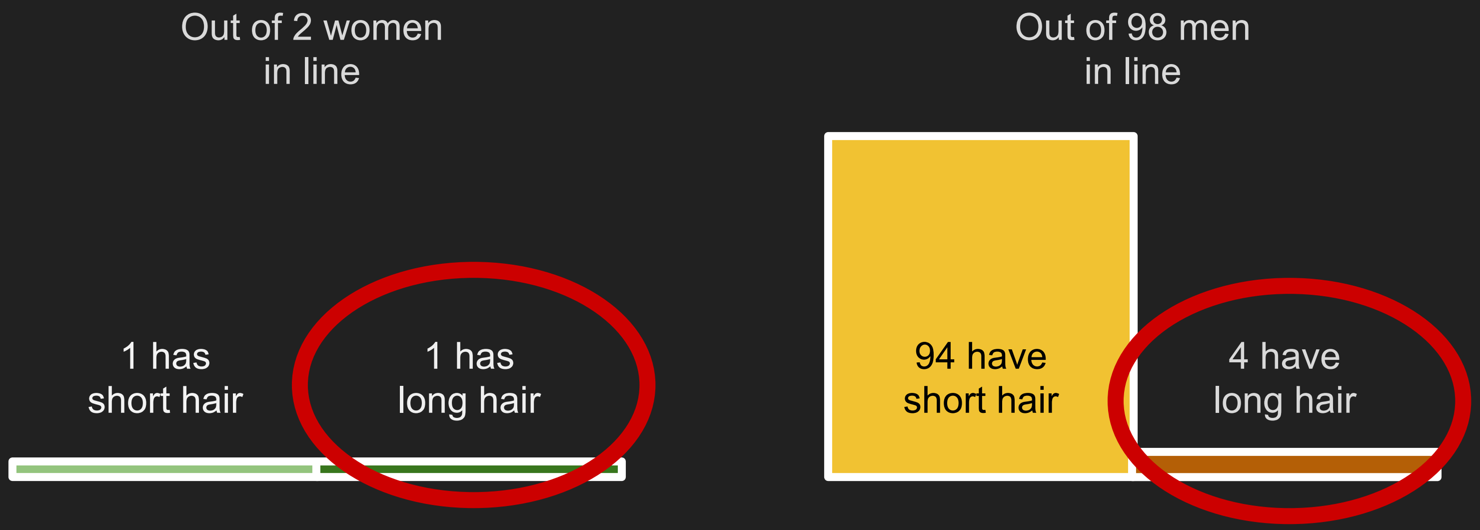
About 12 times more women have long hair than men.

Put numbers to our dilemma



But there are 98 men and 2 women in line for the men’s restroom.

Put numbers to our dilemma



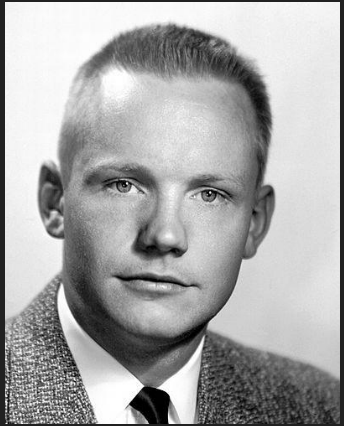
In the line, 4 times more men have long hair than women.

Out of 100 people
at the movies

50 are women

50 are men

2 men have long hair



Out of 100 people
In line for the men's
restroom

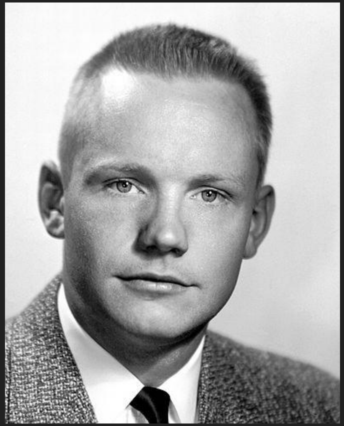
2 are women

98 are men

One woman has short hair

One woman has long hair

4 men have long hair



Translate to math

P(something) = # something / # everything

P(woman) = Probability that a person is a woman

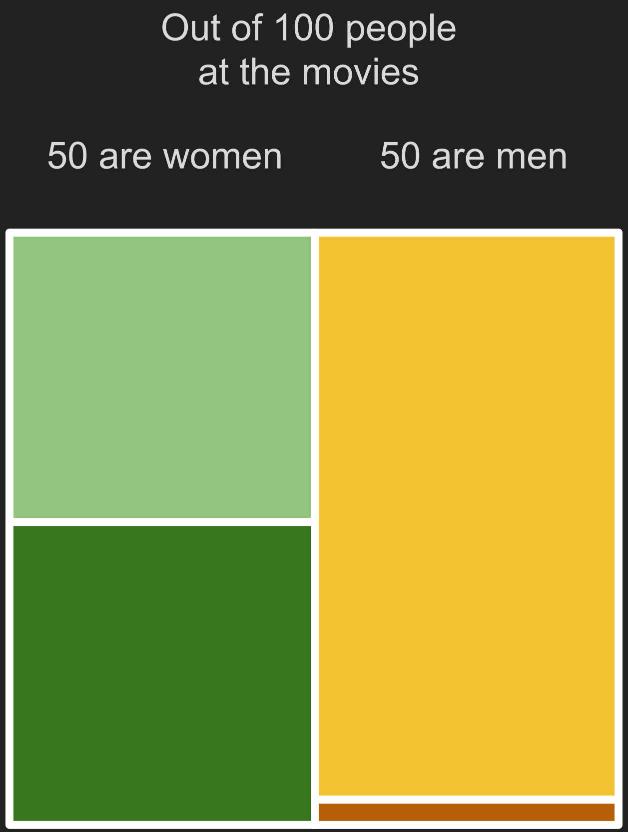
= # women / # people

= 50 / 100 = **.5**

P(man) = Probability that a person is a man

= # men / # people

= 50 / 100 = **.5**



Translate to math

P(something) = # something / # everything

P(woman) = Probability that a person is a woman

= # women / # people

= 2 / 100 = **.02**

P(man) = Probability that a person is a man

= # men / # people

= 98 / 100 = **.98**

Conditional probabilities

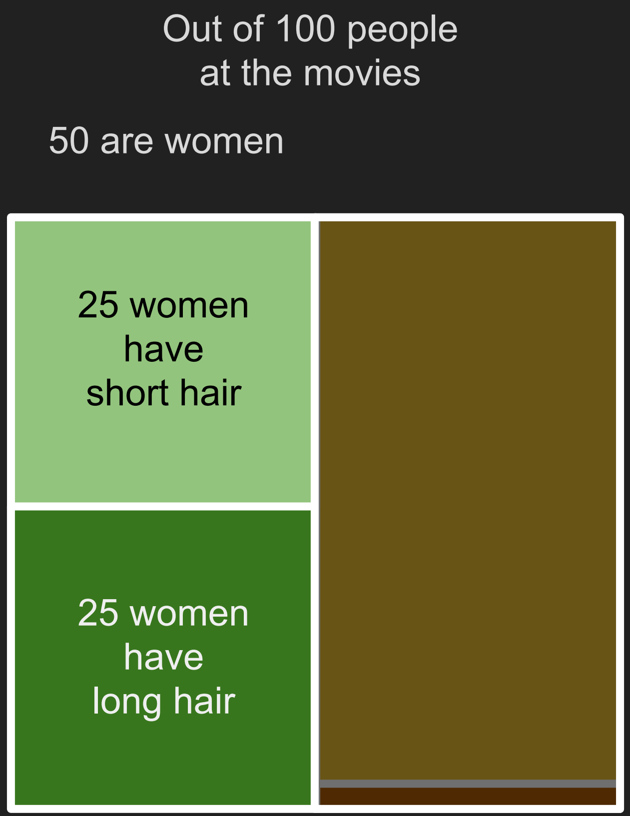
P(long hair | woman)

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



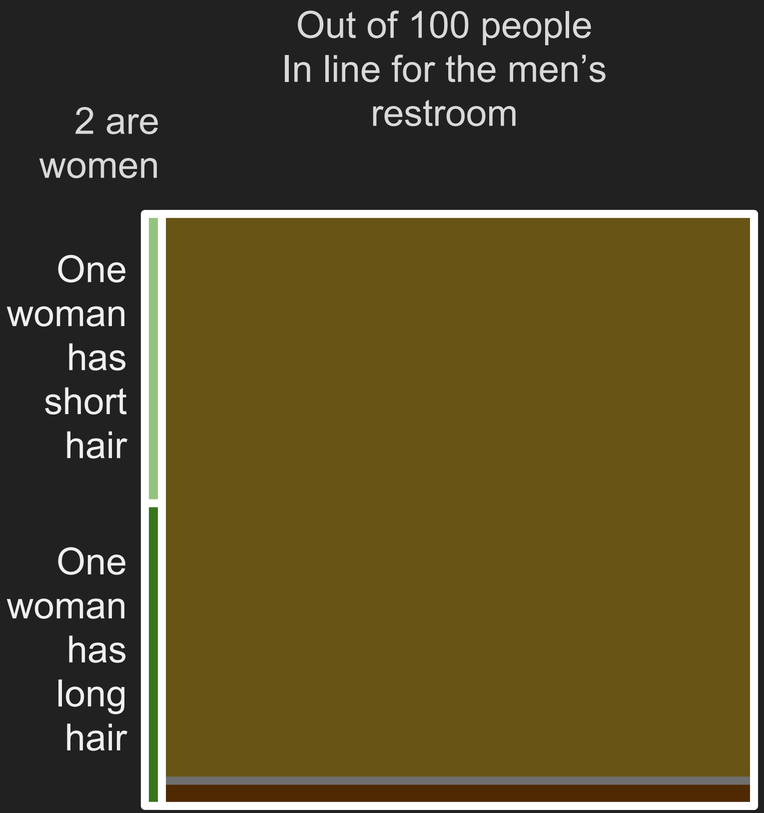
Conditional probabilities

This doesn't change when we consider people in line.

P(long hair | woman)

= # women with long hair / # women

= 1 / 2 = **.5**



Conditional probabilities

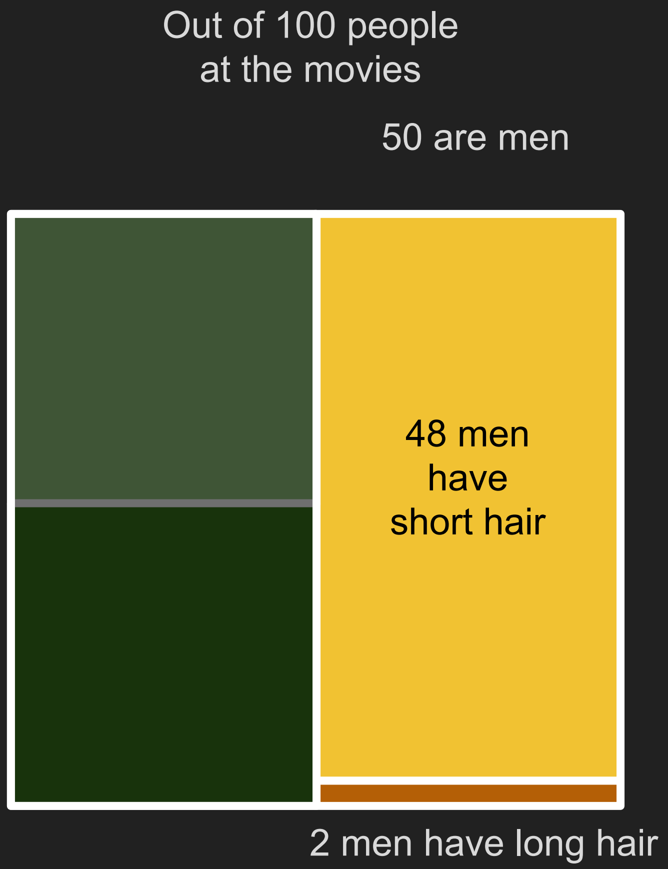
If I know that a person is a man, what is the probability that person has long hair?

P(long hair | man)

= # men with long hair / # men

= 2 / 50 = **.04**

Whether in line or not.



Conditional probabilities

P(A | B) is the probability of A, given B.

“If I know B is the case, what is the probability that A is also the case?”

P(A | B) is not the same as P(B | A).

P(cute | puppy) is not the same as P(puppy | cute)

If I know the thing I’m holding is a puppy, what is the probability that it is cute?

If I know the the thing I’m holding is cute, what is the probability that it is a puppy?

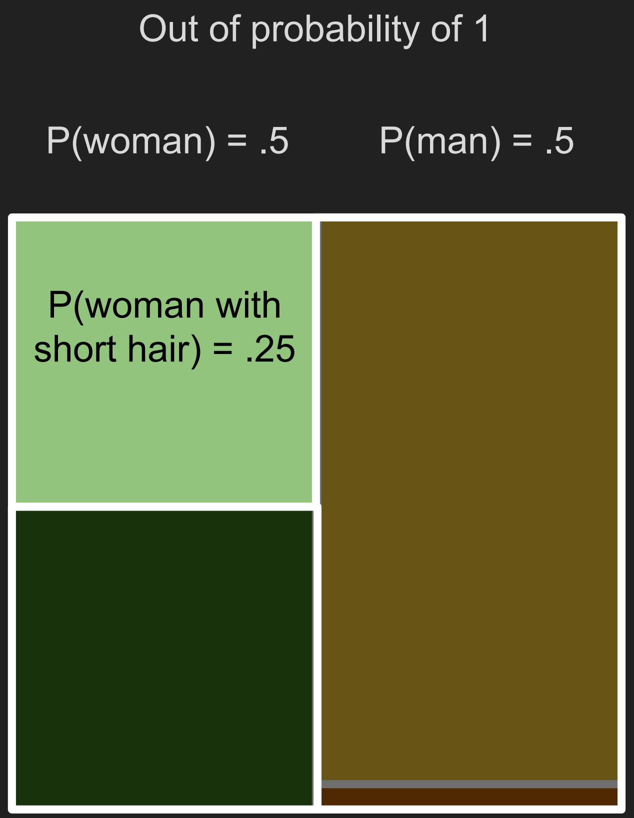
Joint probabilities

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**

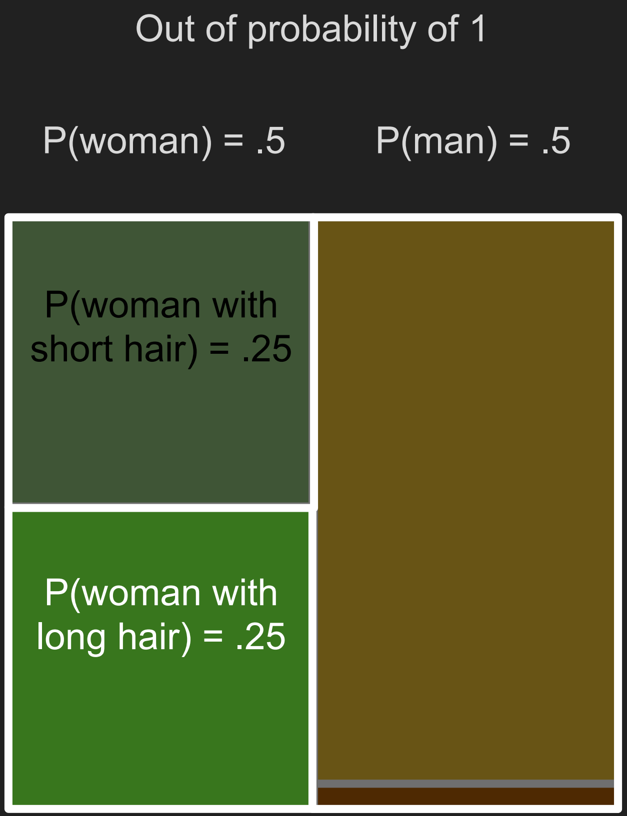


Joint probabilities

P(woman with long hair)

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

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

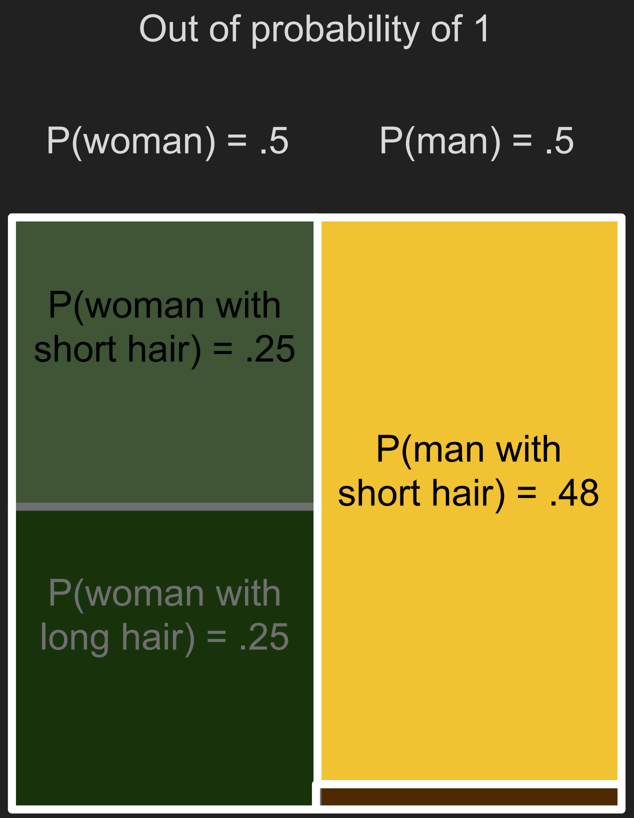


Joint probabilities

P(man with short hair)

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

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



Joint probabilities

P(man with long hair)

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

= .5 * .04 = **.02**

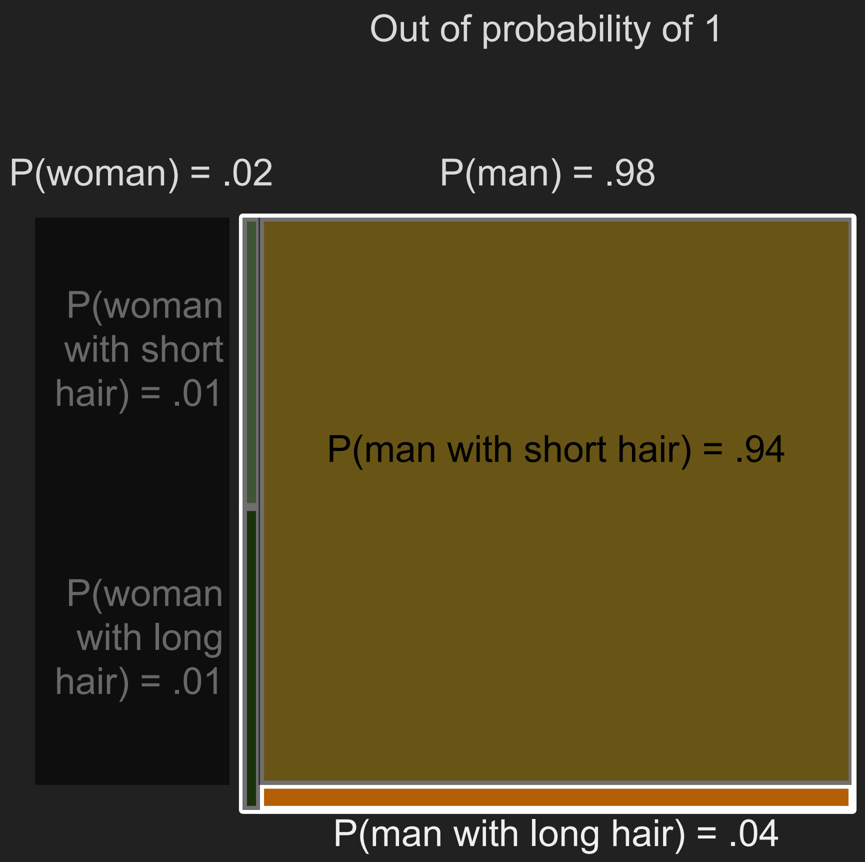
Joint probabilities

If P(man) = .98 and P(woman) = .02, then the answers change.

P(man with long hair)

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

= .98 * .04 = **.04**



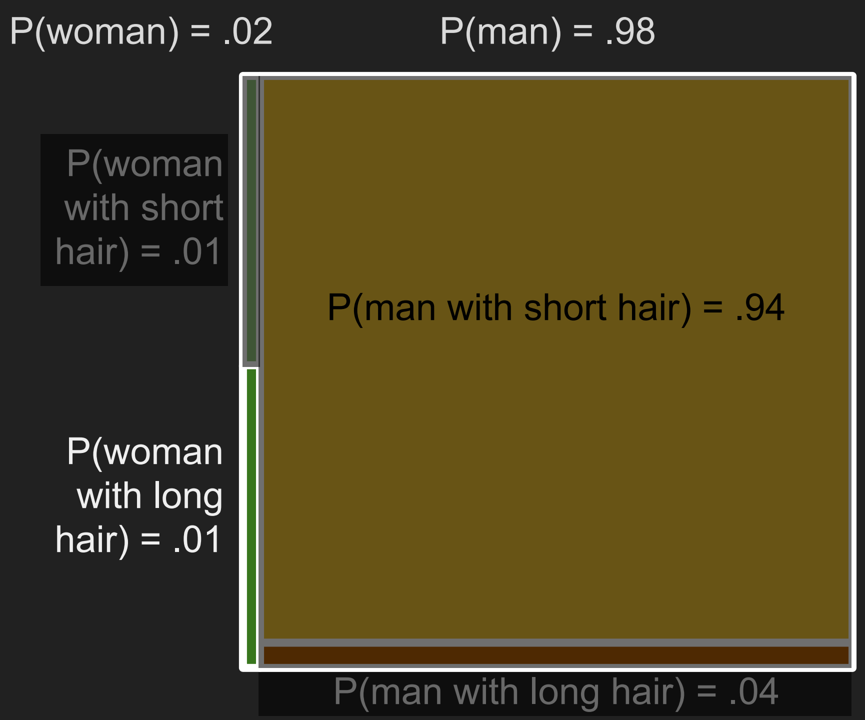
Joint probabilities

Out of probability of 1

P(woman with long hair)

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

= .02 * .5 = **.01**



Joint probabilities

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

Also written P(A, B) or $P(A \cap B)$

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.

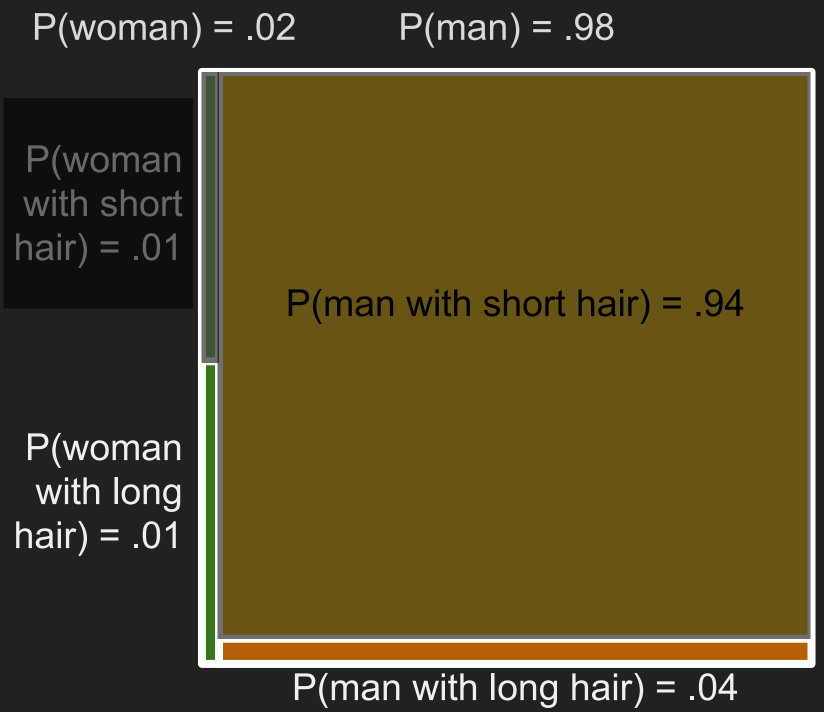
P(donut and milk) = P(milk and donut)

Marginal probabilities

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

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

Out of probability of 1

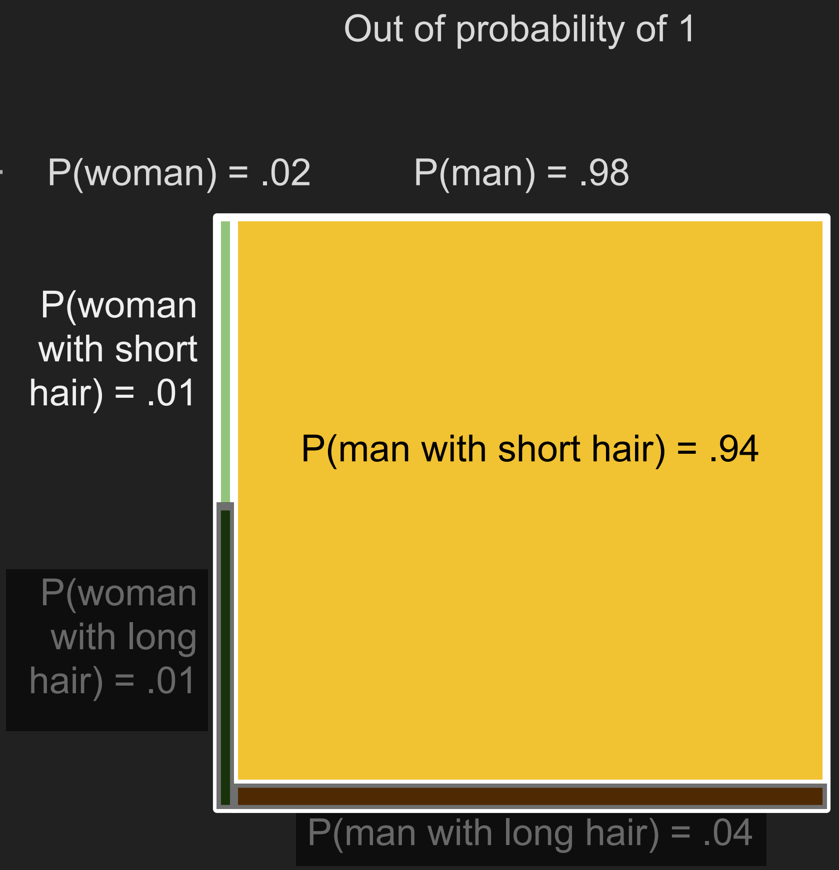


Marginal probabilities

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

P(man with short hair)

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



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)

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)

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)

Thomas Bayes noticed something cool

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

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

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

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

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

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

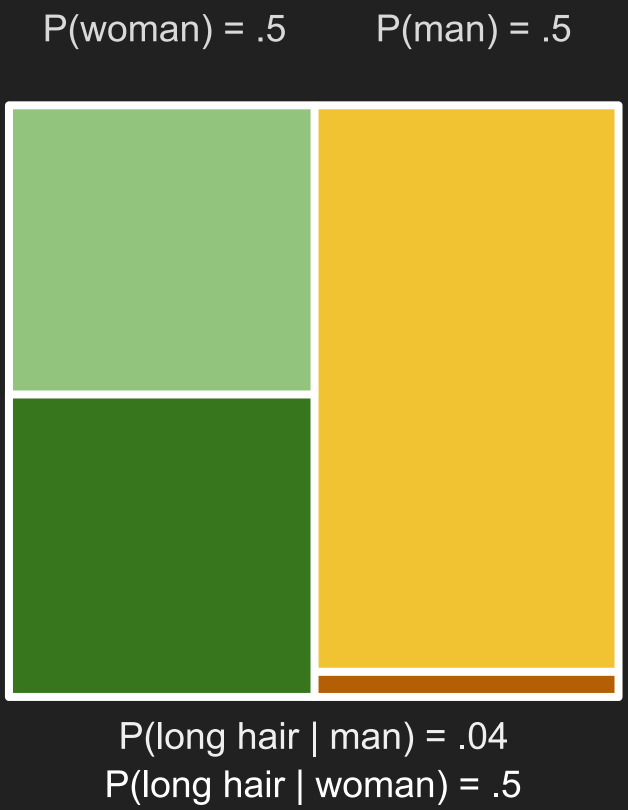
Back to the movie theater, this time with Bayes

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

Back to the movie theater, this time with Bayes

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

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



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

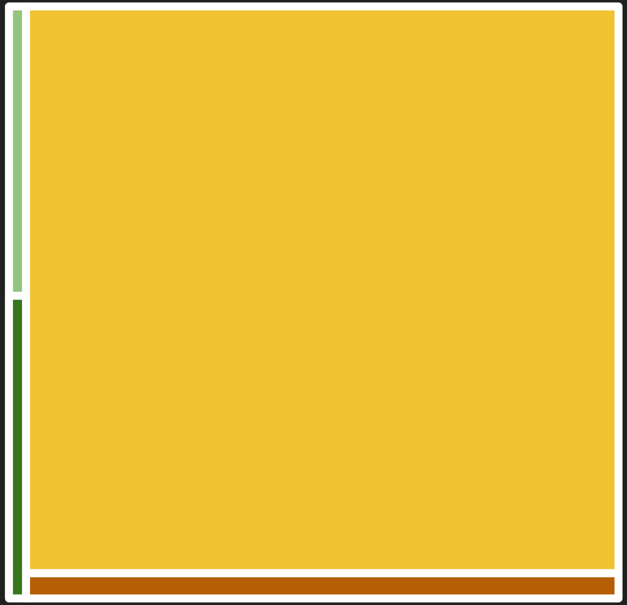
P(long hair | man) = .04
P(long hair | woman) = .5

Back to the 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(woman) = .02 P(man) = .98



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

Back to the movie theater, this time with Bayes

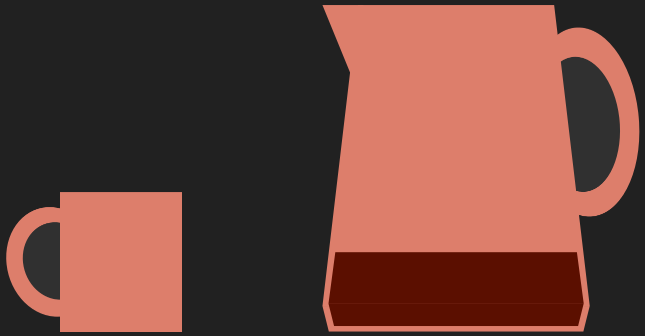
$$P(\text{man} \mid \text{long hair}) = \frac{P(\text{man}) * P(\text{long hair} \mid \text{man})}{P(\text{long hair})}$$

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

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

Probability distributions

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

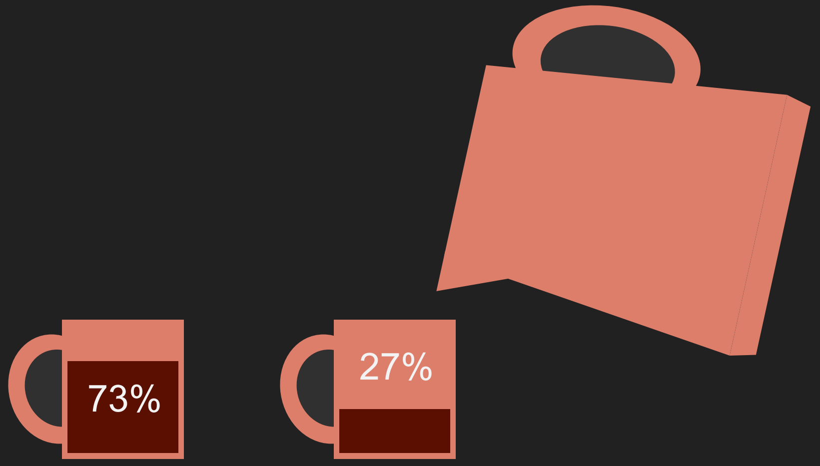


Probability distributions

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

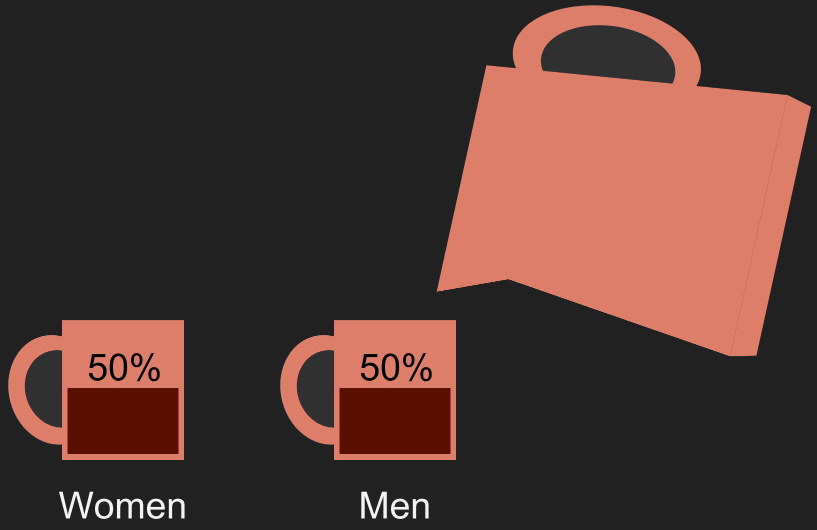
Probability distributions

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



Probability distributions

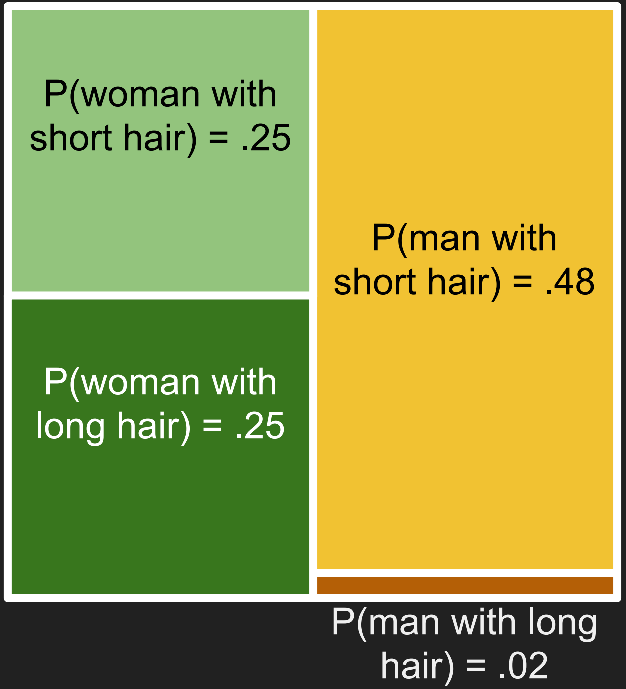
Our people are distributed between two groups, women and men.



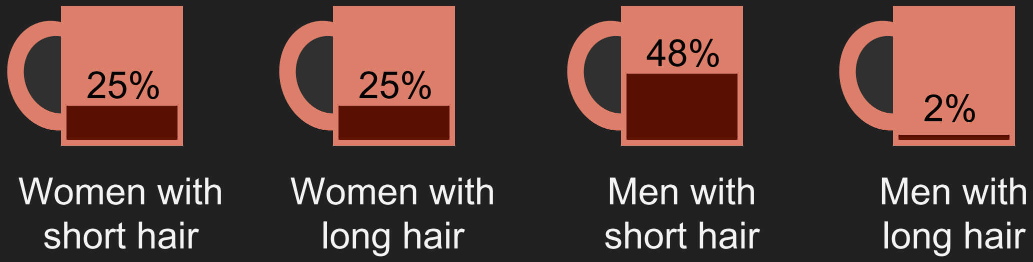
Probability distributions

We can distribute them more.

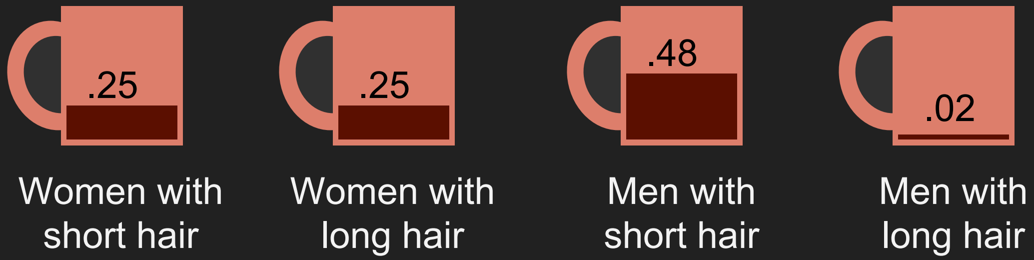
Probability distributions



Probability distributions

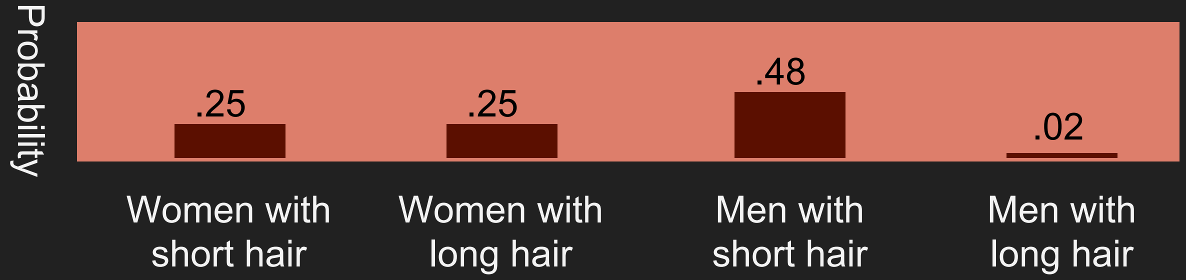


Probability distributions



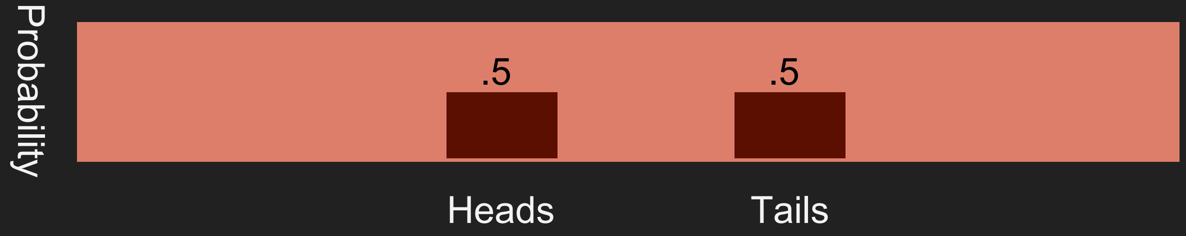
Probability distributions

It’s helpful to think of probabilities as beliefs



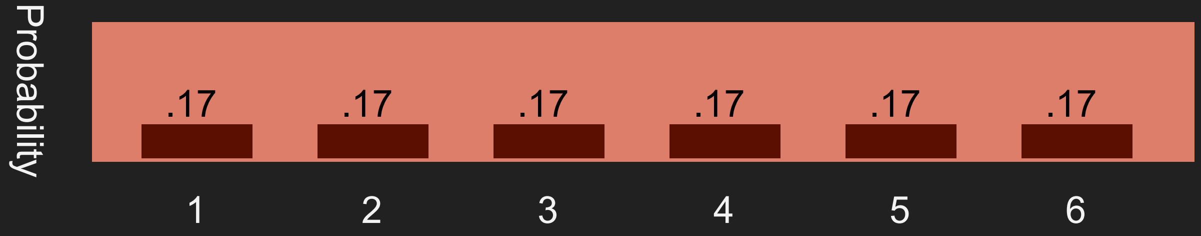
Probability distributions

Flipping a fair coin



Probability distributions

Rolling a fair die

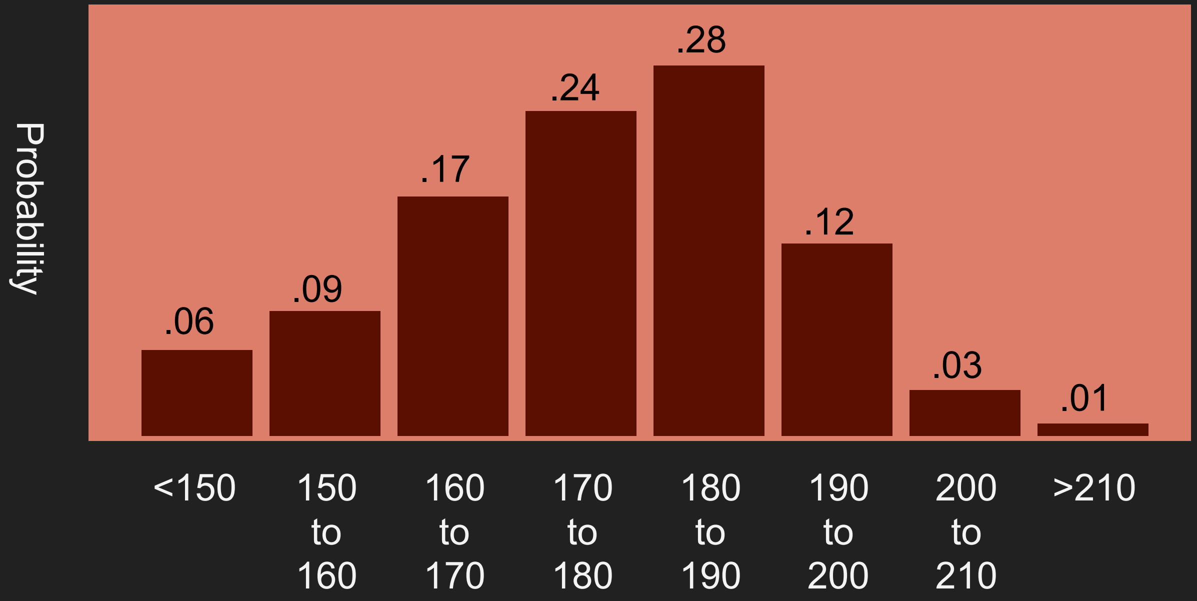


Probability distributions

Playing for the Powerball jackpot

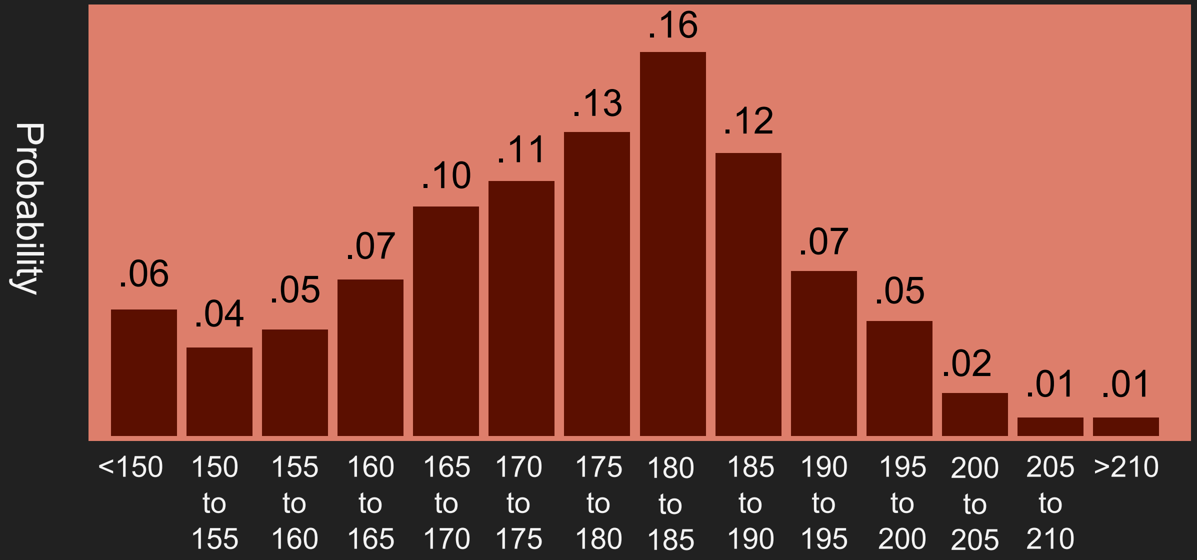
Probability distributions

Height of adults in cm



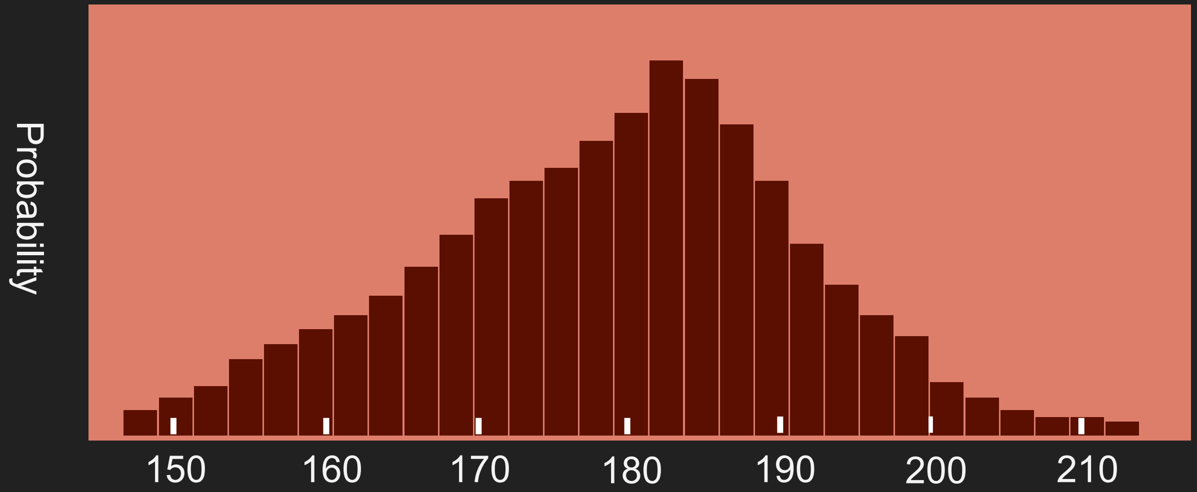
Probability distributions

Height of adults in cm



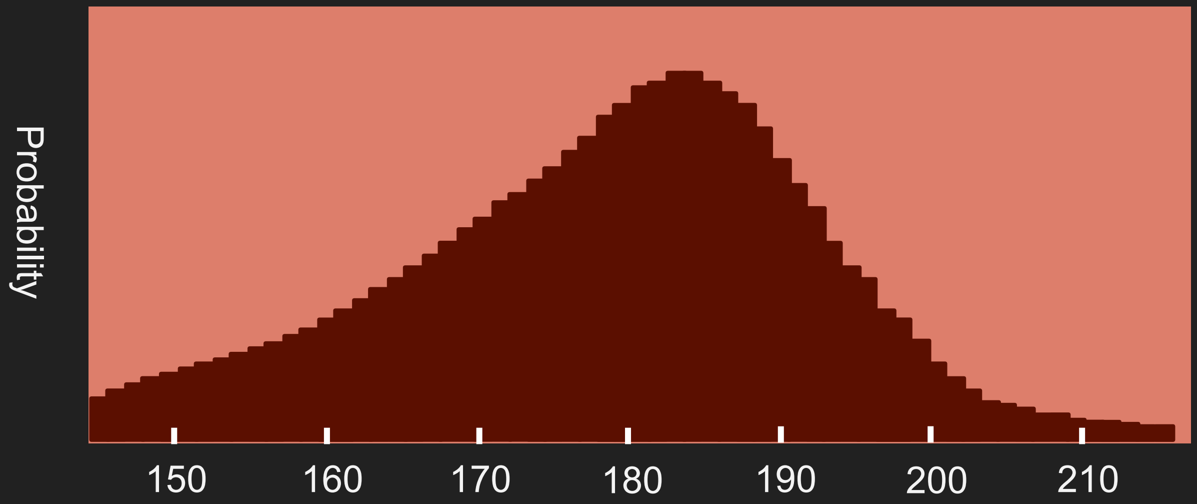
Probability distributions

Height of adults in cm



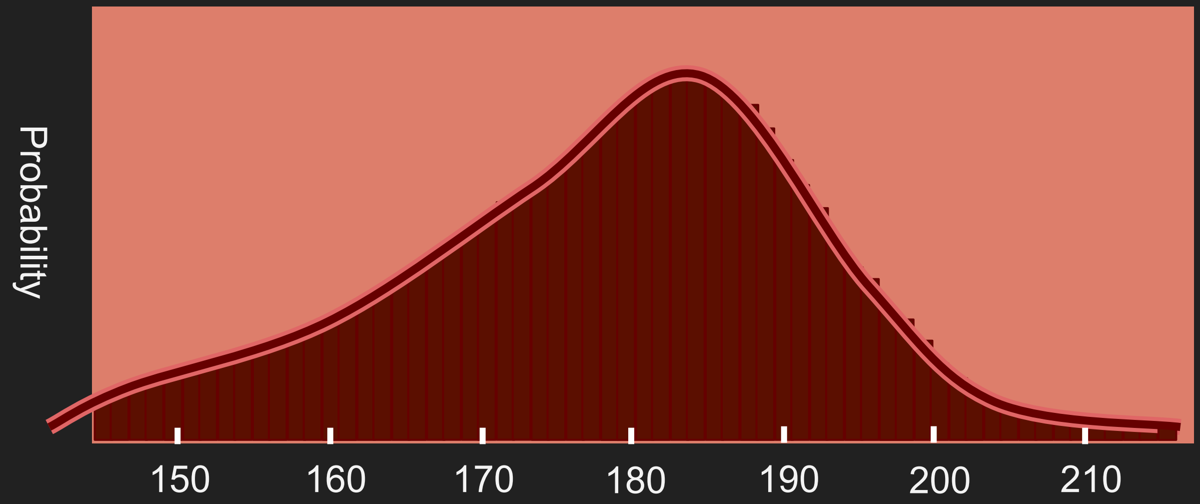
Probability distributions

Height of adults in cm



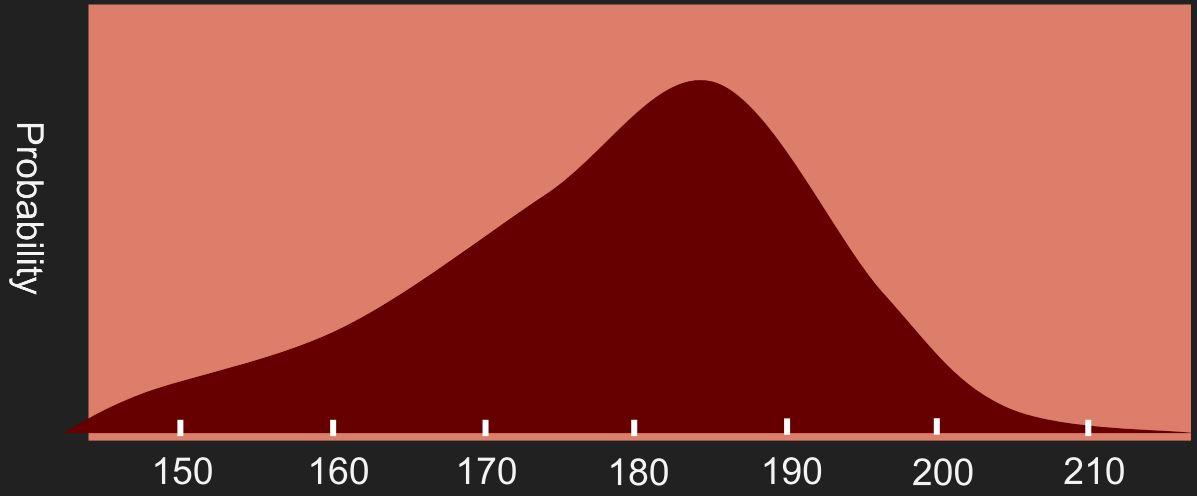
Probability distributions

Height of adults in cm



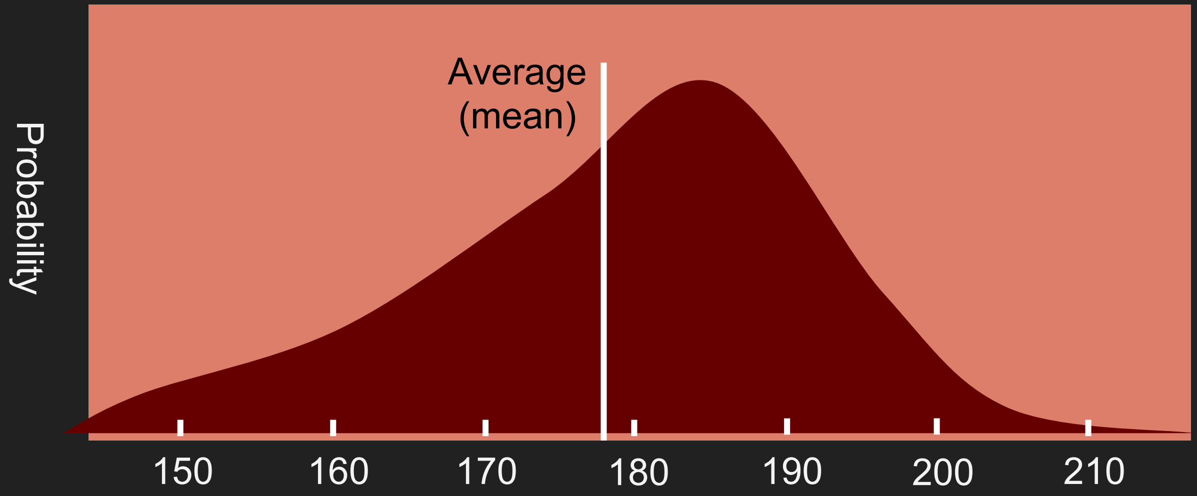
Probability distributions

Height of adults in cm



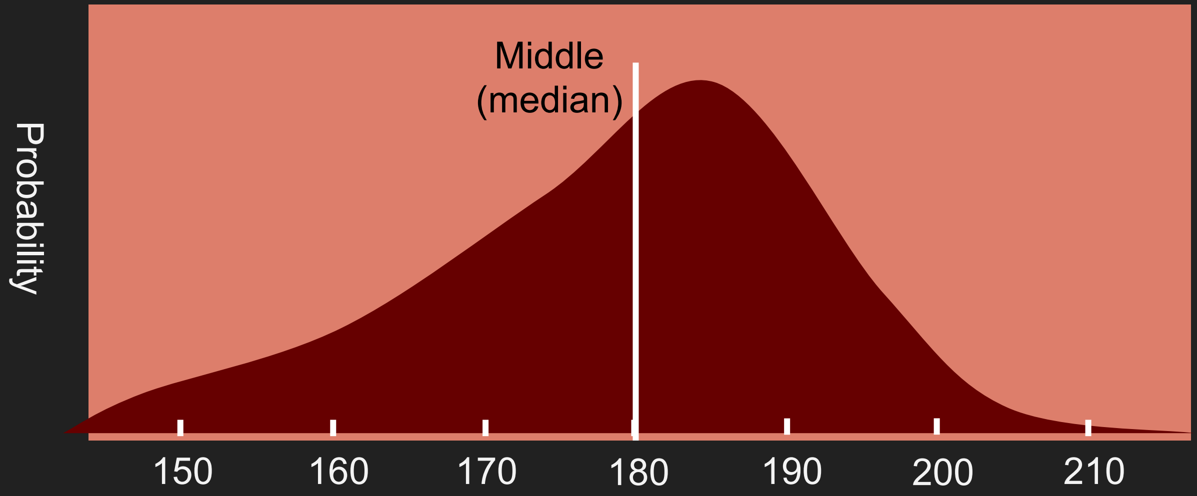
Probability distributions

Height of adults in cm



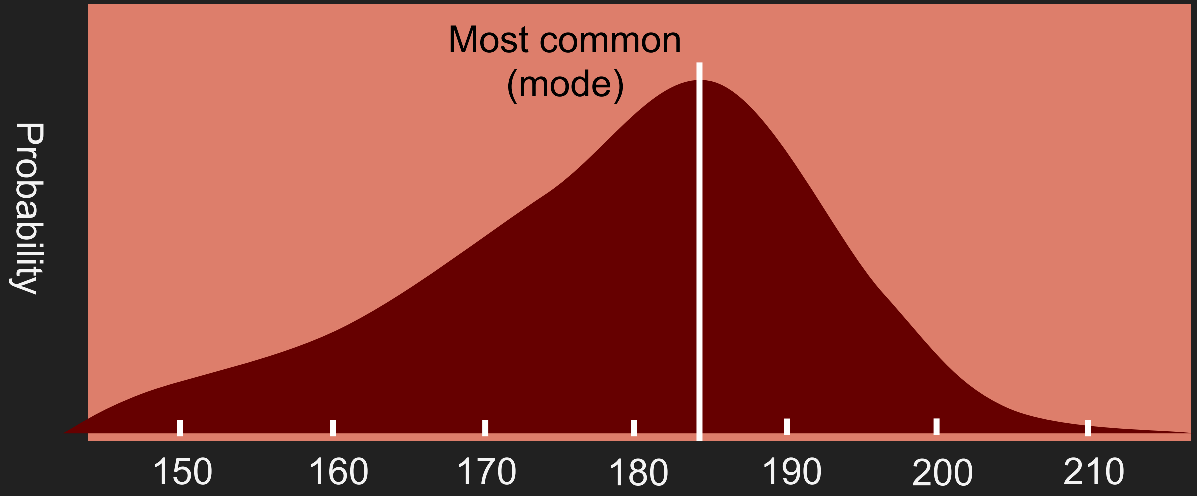
Probability distributions

Height of adults in cm



Probability distributions

Height of adults in cm



Weighing my dog

My dog is named Reign of Terror.

When we go to the veterinarian, Reign squirms on the scale. Each time we get a different weight measurement.

13.9 lb
17.5 lb
14.1 lb

How much does she weigh?

Take the average.

mean = (13.9 + 17.5 + 14.1) / 3 = 15.2 lb

Calculate the standard deviation.

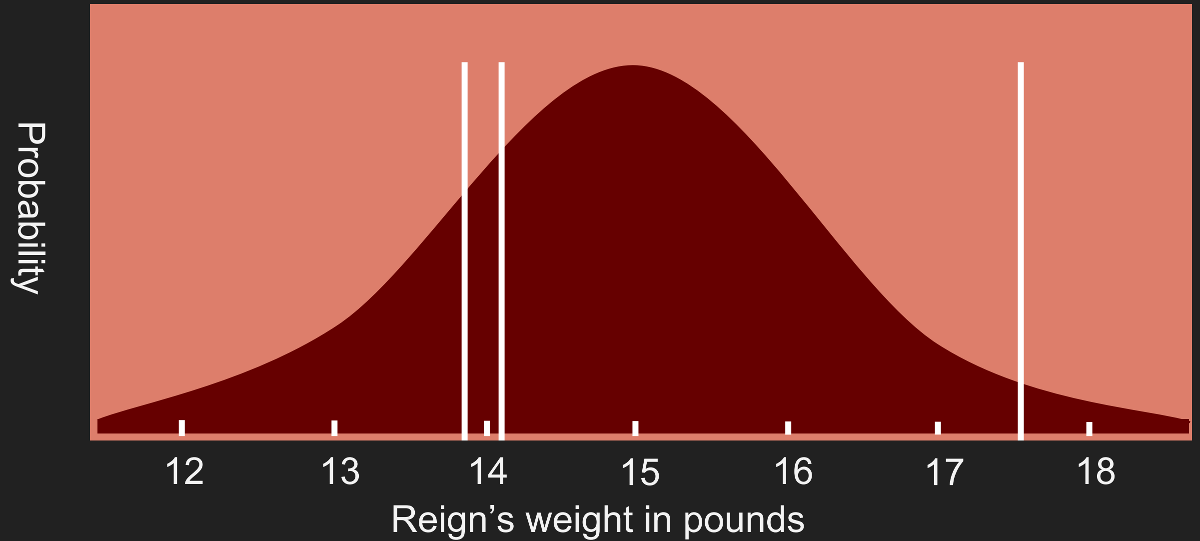
std dev = sqrt((13.9 - 15.1)^2 + (17.5 - 15.1)^2 + (14.1 - 15.1)^2) / 2 = 2.0 lb

Calculate the standard error.

std err = std dev / sqrt(3) = 1.16

How much does she weigh?

The estimate of the mean is a Normal distribution with a mean of 15.2 lb and a standard deviation of 1.2 lb.



Bayes' Theorem

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

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

$$P(w \mid m) = \frac{\overset{\text{likelihood}}{\boxed{P(m \mid w)}}\, P(w)}{P(m)}$$

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

Bayes' Theorem

Start with

$$P(w) = \text{uniform}$$

Bayes' Theorem

$$P(w \mid m) = \frac{P(m \mid w)\ C_1}{C_2}$$

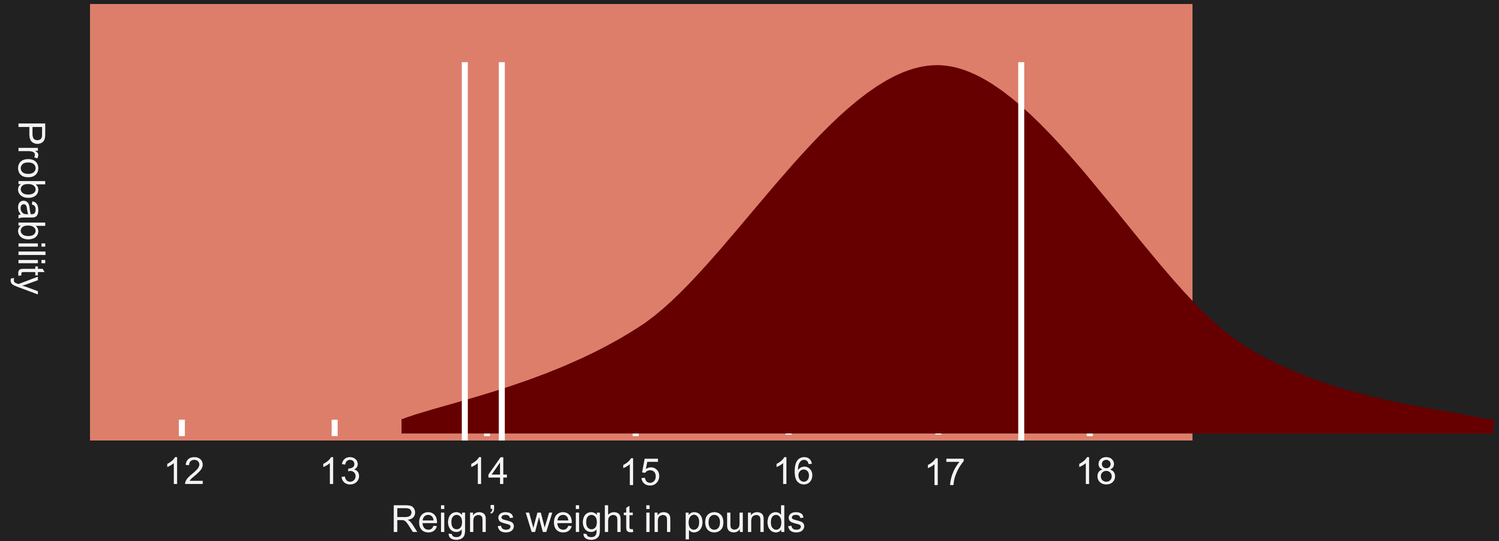
Bayes' Theorem

Assumes that the mean of the weight is equally likely to be anything.

$$P(w \mid m) = P(m \mid w)$$

$$P(w \mid m) = P(m \mid w)$$

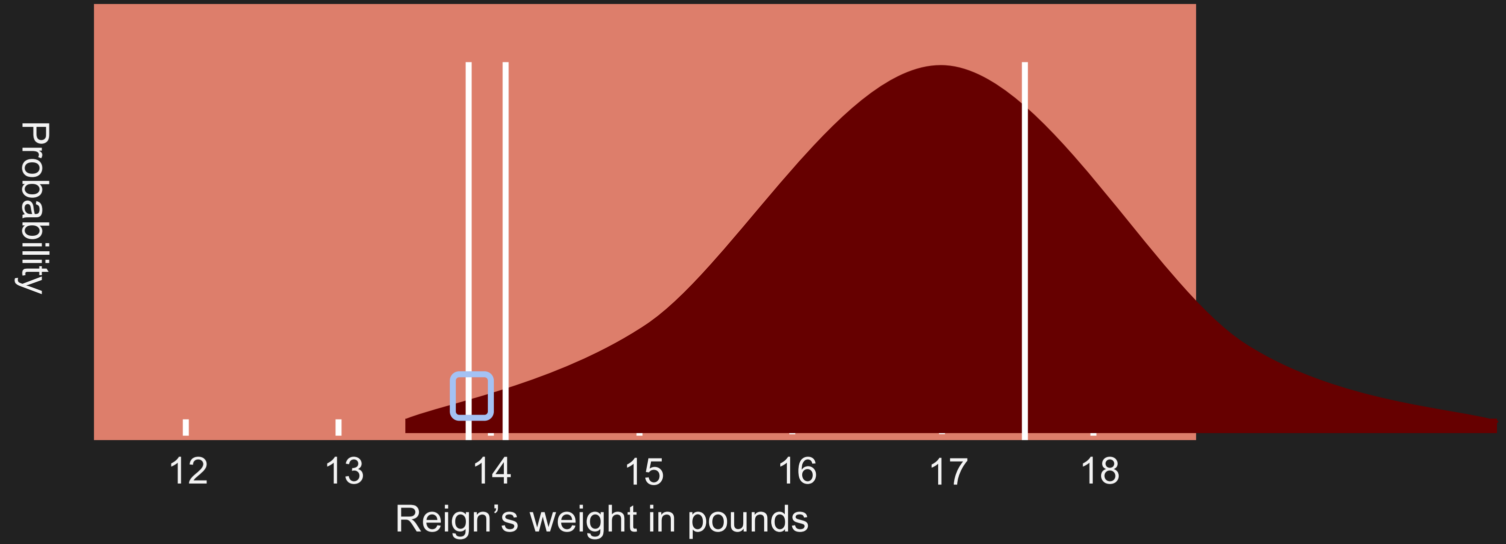
$$P(w = 17 \mid m = [13.9, 14.1, 17.5]) \quad = P(m = [13.9, 14.1, 17.5] \mid w = 17)$$



$$P(w \mid m) = P(m \mid w)$$

$$P(w = 17 \mid m = [13.9, 14.1, 17.5]) = P(m = [13.9, 14.1, 17.5] \mid w = 17)$$

$$= \boxed{P(m = 13.9 \mid w = 17)} * P(m = 14.1 \mid w = 17) * P(m = 17.5 \mid w = 17)$$



$$P(w \mid m) = P(m \mid w)$$

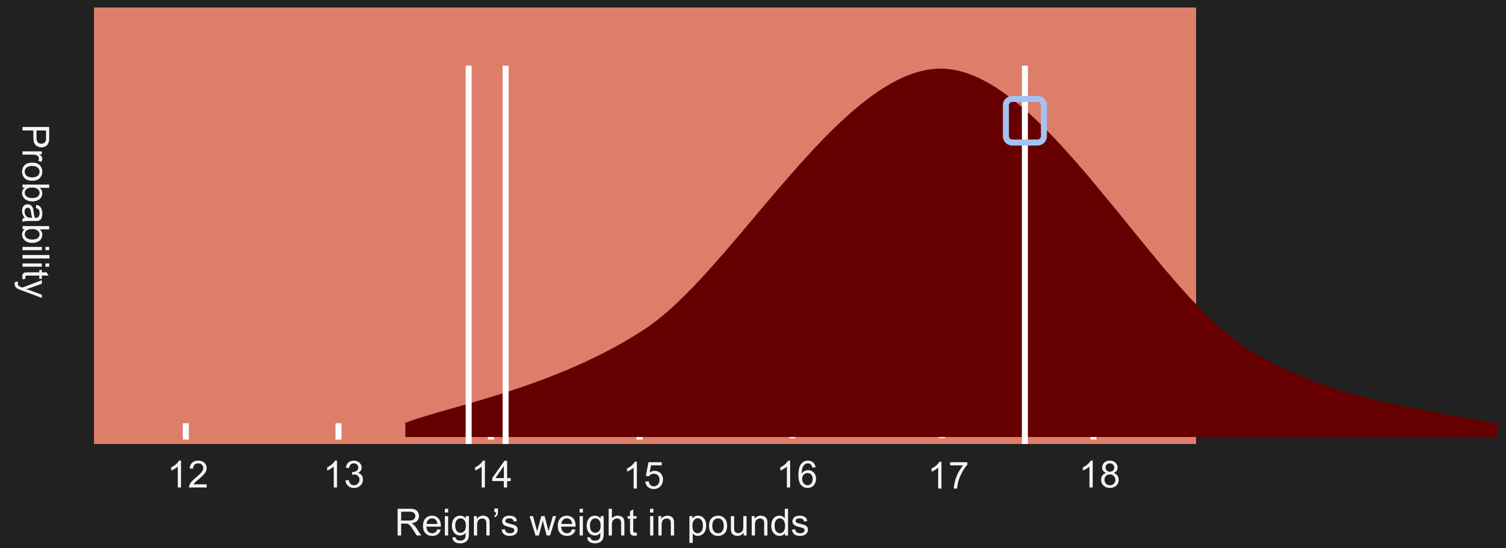
$$P(w = 17 \mid m = [13.9, 14.1, 17.5]) = P(m = [13.9, 14.1, 17.5] \mid w = 17)$$

$$= P(m = 13.9 \mid w = 17) * \boxed{P(m = 14.1 \mid w = 17)} * P(m = 17.5 \mid w = 17)$$

$$P(w \mid m) = P(m \mid w)$$

$$P(w = 17 \mid m = [13.9, 14.1, 17.5]) = P(m = [13.9, 14.1, 17.5] \mid w = 17)$$

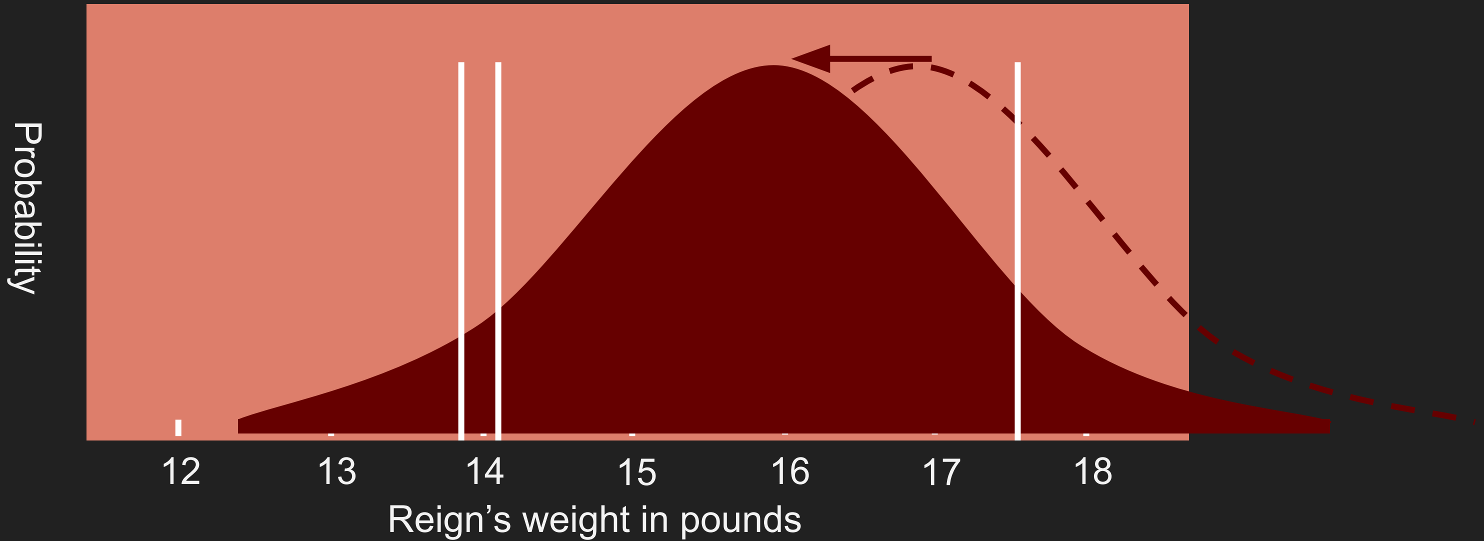
$$= P(m = 13.9 \mid w = 17) * P(m = 14.1 \mid w = 17) * \boxed{P(m = 17.5 \mid w = 17)}$$



$$P(w \mid m) = P(m \mid w)$$

$$P(w = 16 \mid m = [13.9, 14.1, 17.5]) \quad = P(m = [13.9, 14.1, 17.5] \mid w = 16)$$

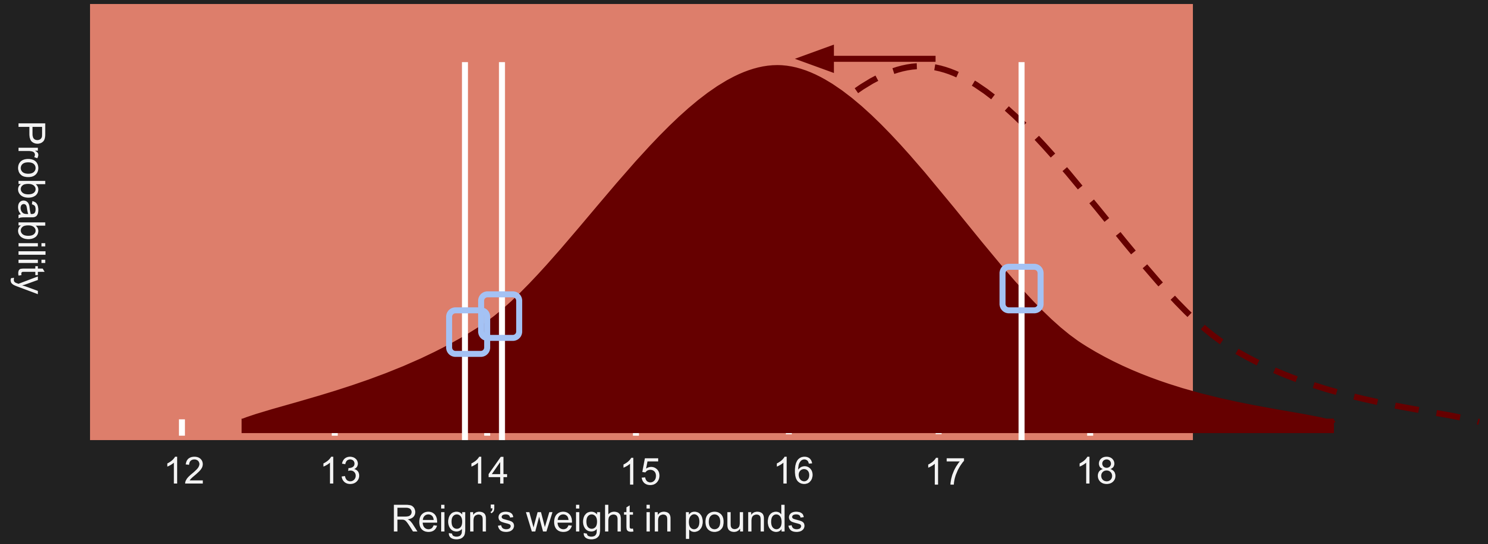
$$= P(m = 13.9 \mid w = 16) * P(m = 14.1 \mid w = 16) * P(m = 17.5 \mid w = 16)$$



$$P(w \mid m) = P(m \mid w)$$

$$P(w = 16 \mid m = [13.9, 14.1, 17.5]) \quad = P(m = [13.9, 14.1, 17.5] \mid w = 16)$$

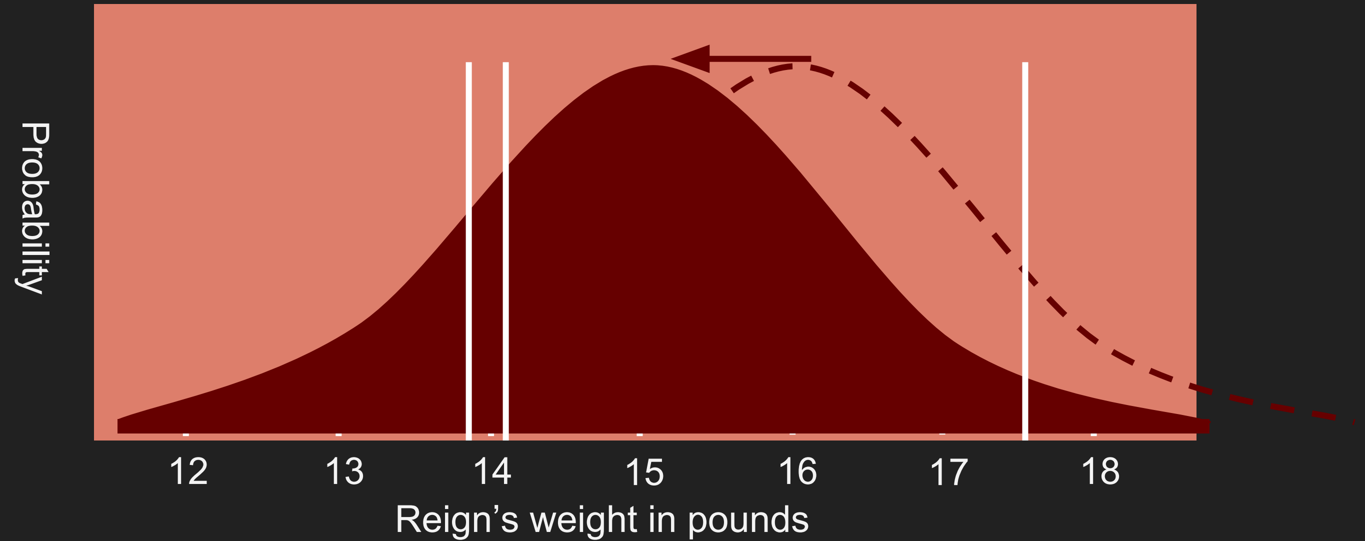
$$= P(m = 13.9 \mid w = 16) * P(m = 14.1 \mid w = 16) * P(m = 17.5 \mid w = 16)$$



$$P(w \mid m) = P(m \mid w)$$

$$P(w = 15 \mid m = [13.9, 14.1, 17.5]) \quad = P(m = [13.9, 14.1, 17.5] \mid w = 15)$$

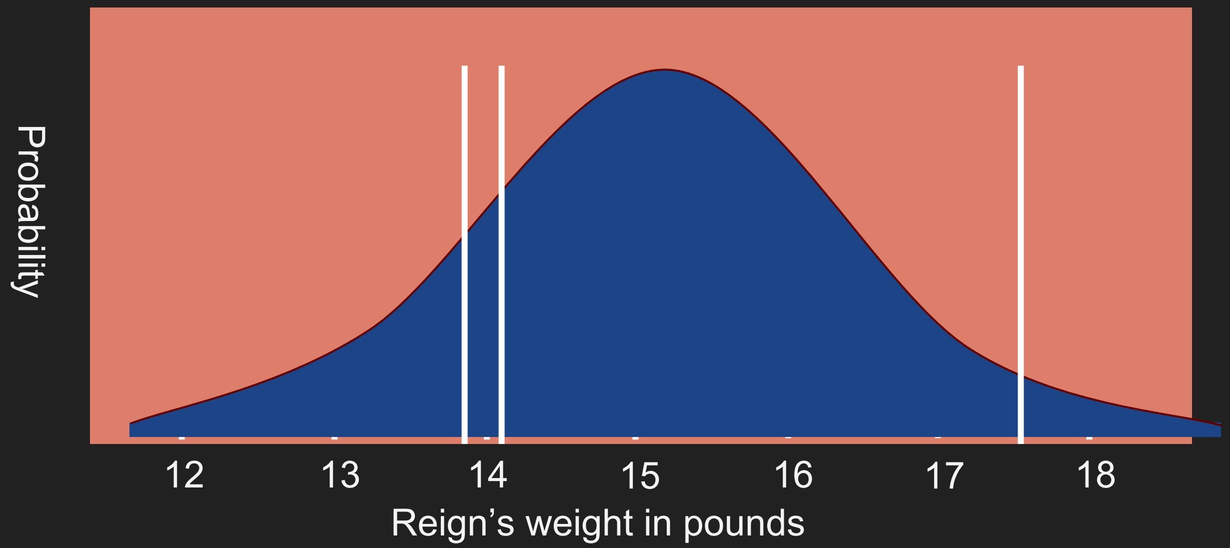
$$= P(m = 13.9 \mid w = 15) * P(m = 14.1 \mid w = 15) * P(m = 17.5 \mid w = 15)$$



$$P(w \mid m) = P(m \mid w)$$

mean = 15.2 lb

Also known as a Maximum Likelihood Estimate (MLE)



What would Thomas do?

Start with what we know.

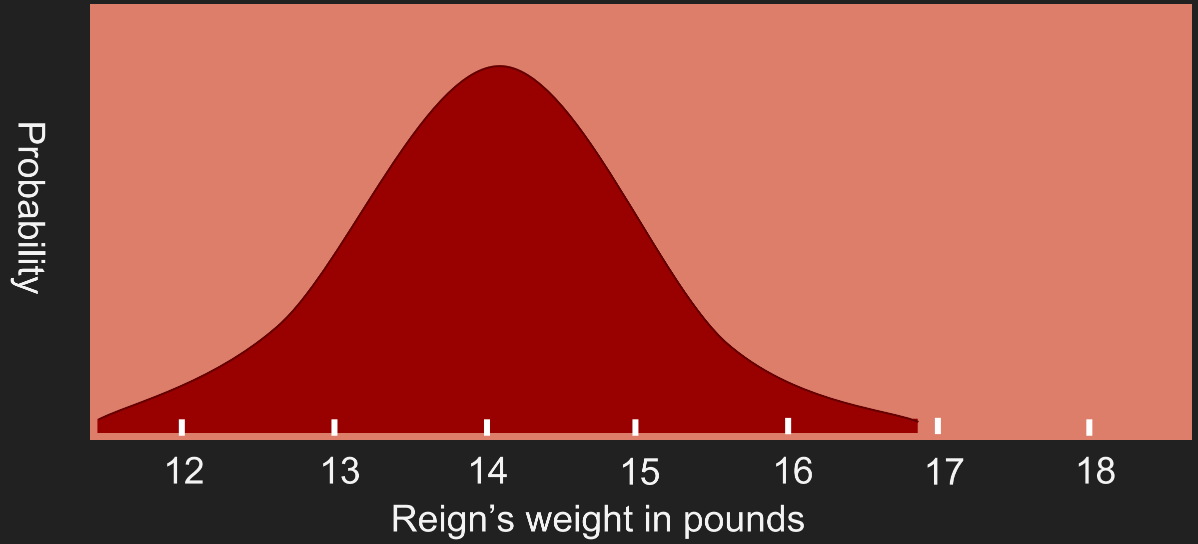
Reign was 14.2 lb the last time we came in.

She doesn't seem noticeably more heavy to me.

I can start with a prior belief--a bias toward what I think the answer will be.

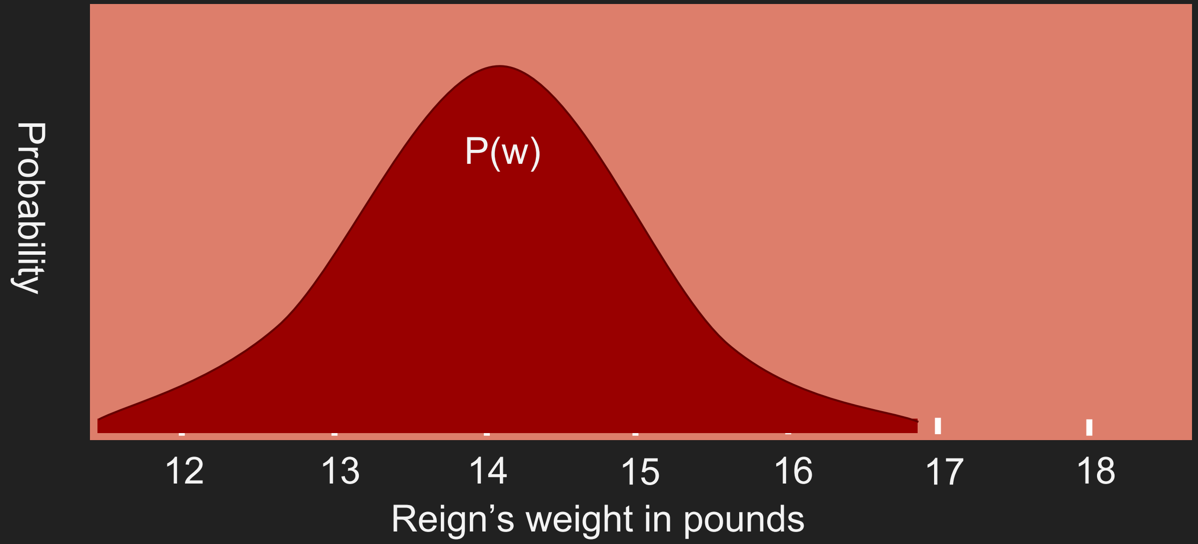
How much does she weigh?

My prior is a normal curve with a mean at 14.2 lb and a standard error of .5 lb.



How much does she weigh?

My prior is a normal curve with a mean at 14.2 lb and a standard error of .5 lb.



Bayes' Theorem

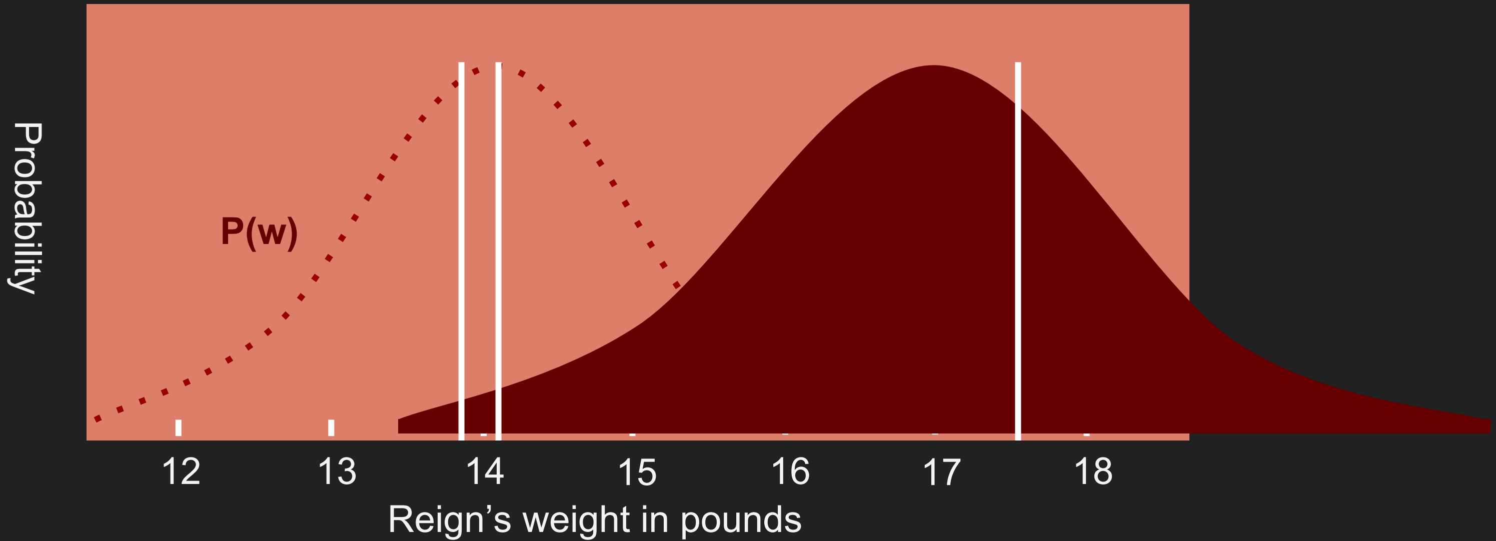
This time I don't neglect the P(w) term. I still assume the P(m) term is constant.

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

$$P(w \mid m) = P(m \mid w) * P(w)$$

$$P(w = 17 \mid m = [13.9, 14.1, 17.5]) = P(m = [13.9, 14.1, 17.5] \mid w = 17)$$

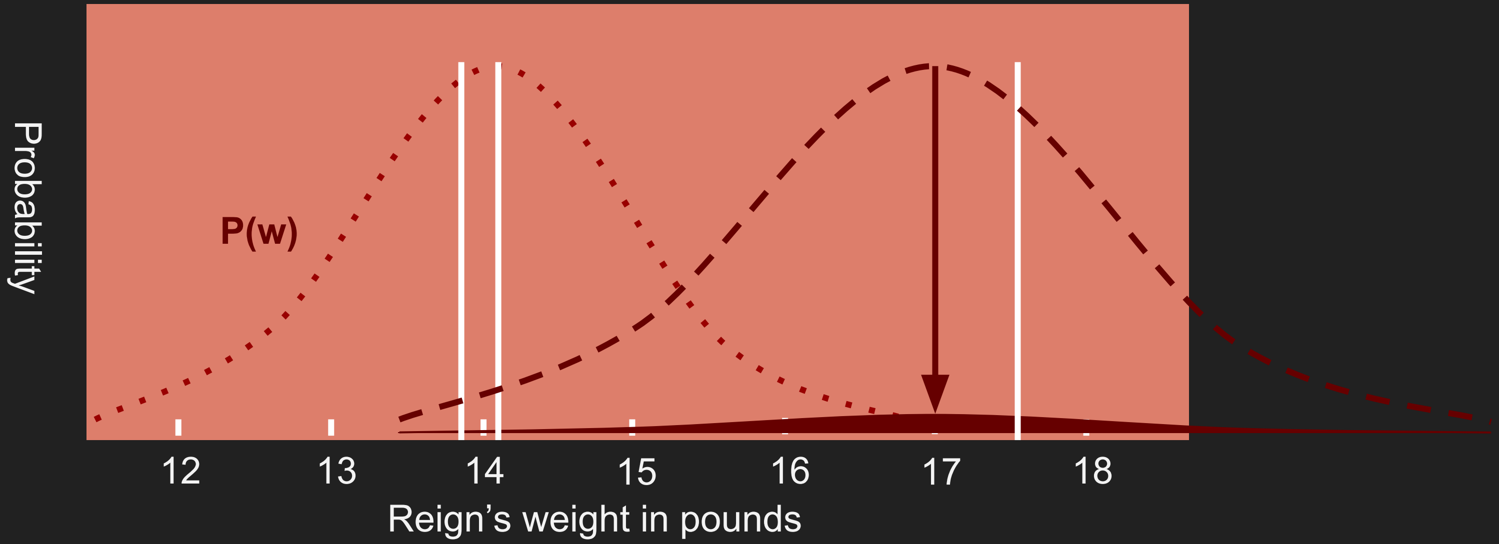
$$= P(m = 13.9 \mid w = 17) * P(m = 14.1 \mid w = 17) * P(m = 17.5 \mid w = 17)$$



$$P(w \mid m) = P(m \mid w) * P(w)$$

$$P(w = 17 \mid m = [13.9, 14.1, 17.5]) = P(m = [13.9, 14.1, 17.5] \mid w = 17) * P(w=17)$$

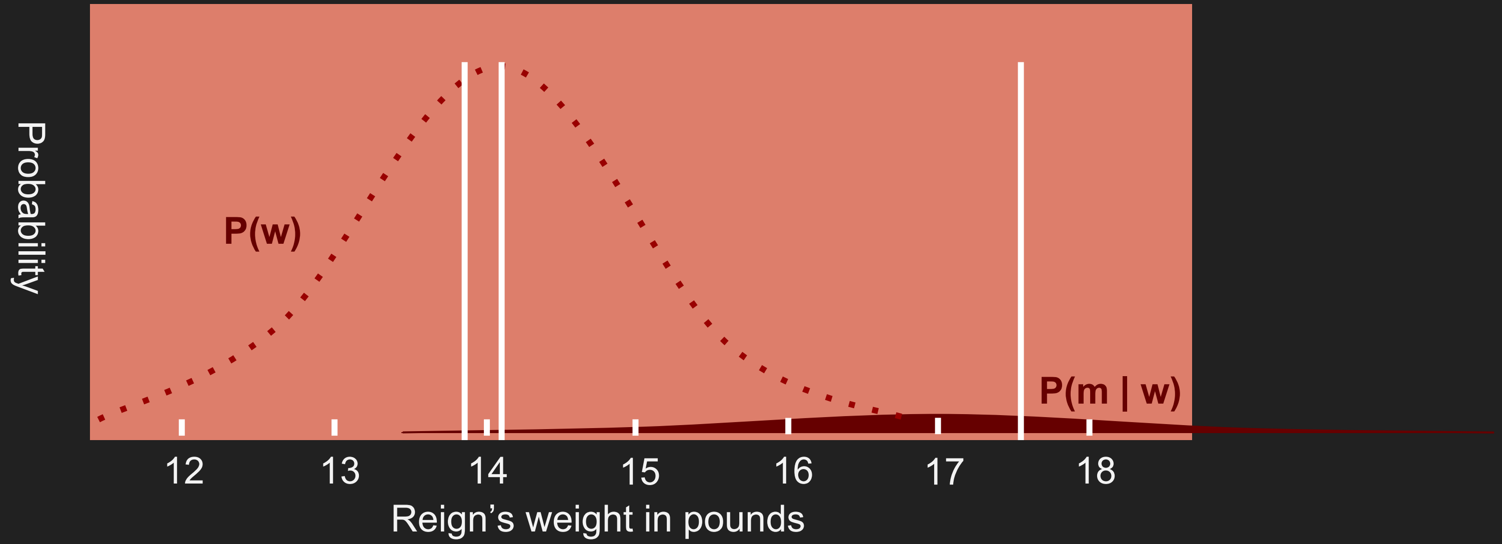
$$= P(m = 13.9 \mid w = 17) * P(w=17) * P(m = 14.1 \mid w = 17) * P(w=17) * P(m = 17.5 \mid w = 17) * P(w=17)$$



$$P(w \mid m) = P(m \mid w) * P(w)$$

$$P(w = 17 \mid m = [13.9, 14.1, 17.5]) = P(m = [13.9, 14.1, 17.5] \mid w = 17) * P(w=17)$$

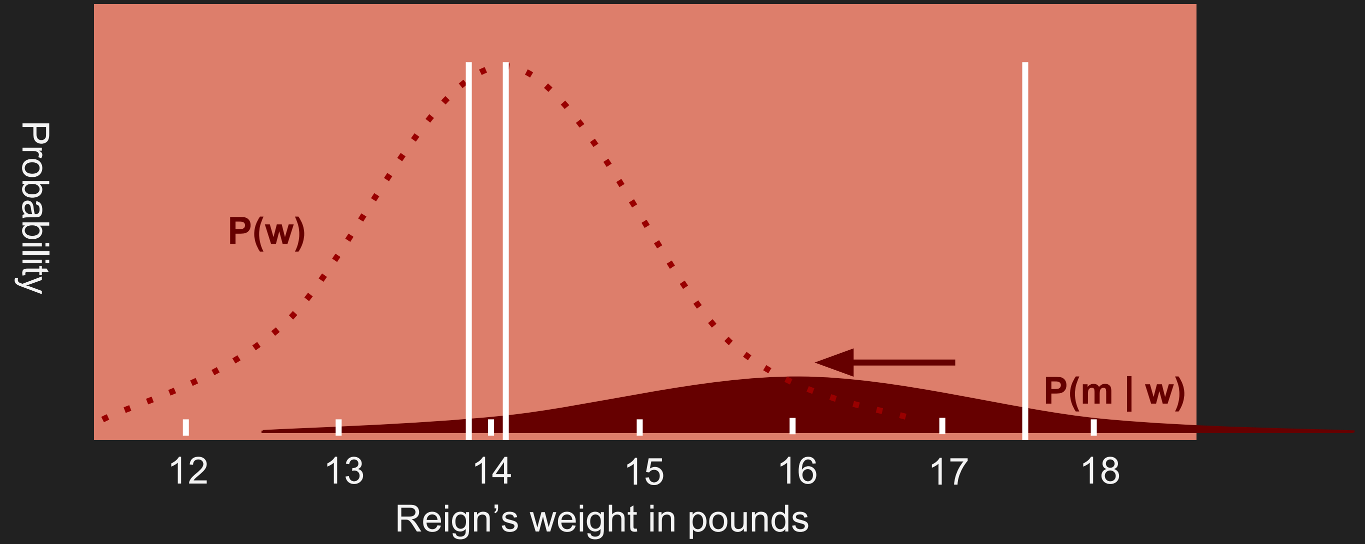
$$= P(m = 13.9 \mid w = 17) * P(w=17) * P(m = 14.1 \mid w = 17) * P(w=17) * P(m = 17.5 \mid w = 17) * P(w=17)$$



$$P(w \mid m) = P(m \mid w) * P(w)$$

$$P(w = 16 \mid m = [13.9, 14.1, 17.5]) = P(m = [13.9, 14.1, 17.5] \mid w = 16) * P(w{=}16)$$

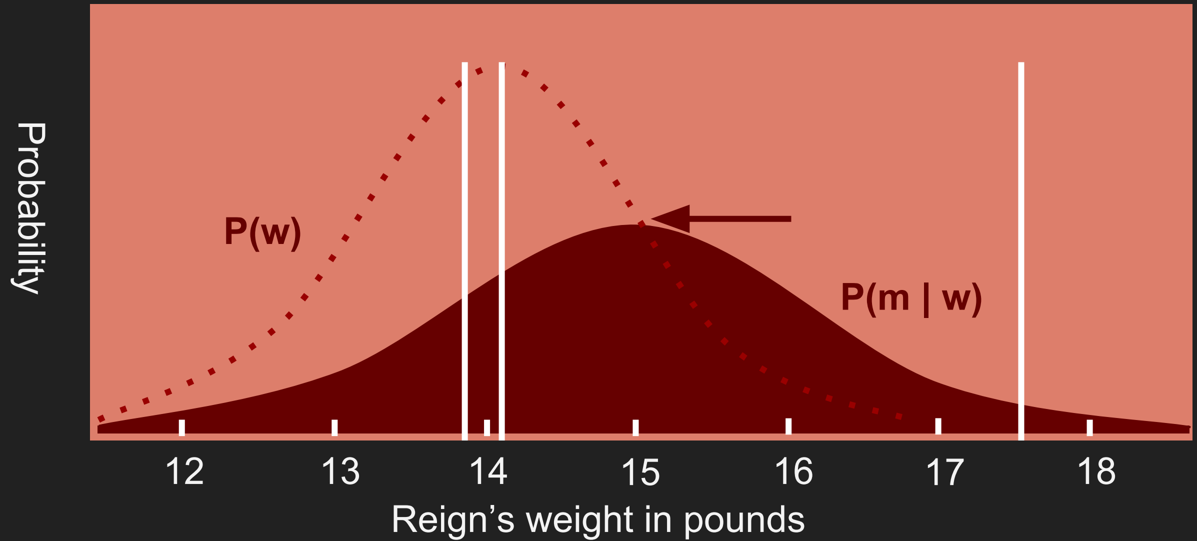
$$= P(m = 13.9 \mid w = 16) * P(w{=}16) * P(m = 14.1 \mid w = 16) * P(w{=}16) * P(m = 17.5 \mid w = 16) * P(w{=}16)$$



$$P(w \mid m) = P(m \mid w) * P(w)$$

$$P(w = 15 \mid m = [13.9, 14.1, 17.5]) = P(m = [13.9, 14.1, 17.5] \mid w = 15) * P(w=15)$$

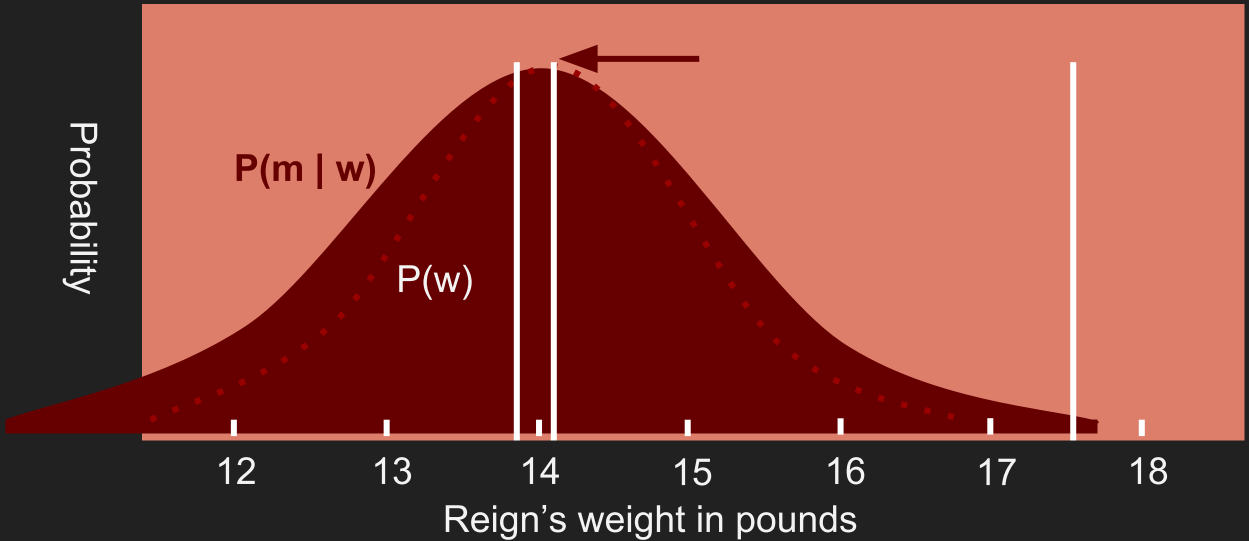
$$= P(m = 13.9 \mid w = 15) * P(w=15) * P(m = 14.1 \mid w = 15) * P(w=15) * P(m = 17.5 \mid w = 15) * P(w=15)$$



$$P(w \mid m) = P(m \mid w) * P(w)$$

$$P(w = 14 \mid m = [13.9, 14.1, 17.5]) = P(m = [13.9, 14.1, 17.5] \mid w = 14) * P(w=14)$$

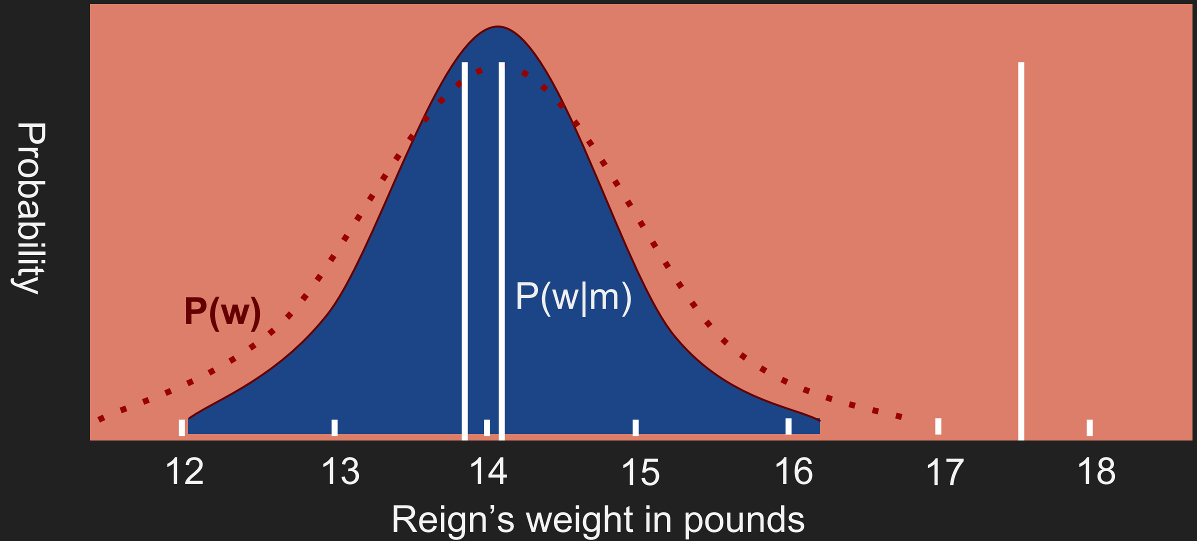
$$= P(m = 13.9 \mid w = 14) * P(w=14) * P(m = 14.1 \mid w = 14) * P(w=14) * P(m = 17.5 \mid w = 14) * P(w=14)$$



$P(w \mid m) = P(m \mid w) * P(w)$

Our new estimate of Reign's weight, $P(w \mid m)$, is a normal distribution with a mean at about 14.1 lb and a std err of .4 lb.

Also known as Maximum A Posteriori (MAP).

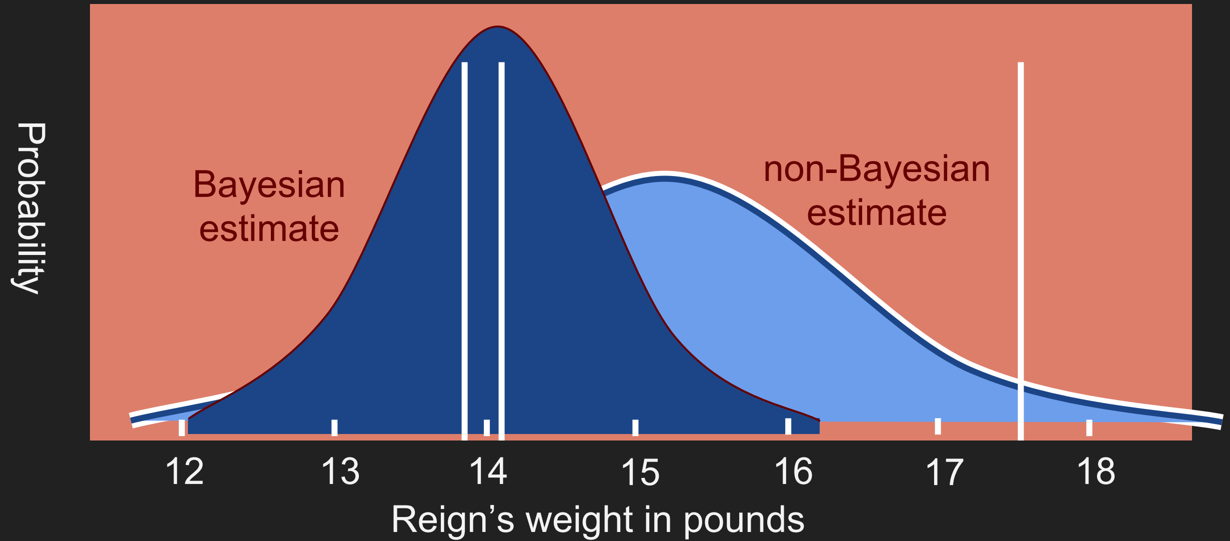


Bayesian vs. not

The Bayesian estimate ignores 17.5 lb like an outlier.

The distribution is narrower. Confidence is greater.

The answer is probably much closer to correct.



Why we might want to use Bayesian inference

We know things

Age is more than zero

Temperature is greater than -273 Celsius

Height is probably less than 2.4 meters (8 feet)

Starting with a belief helps us get to a confident answer with fewer data points.

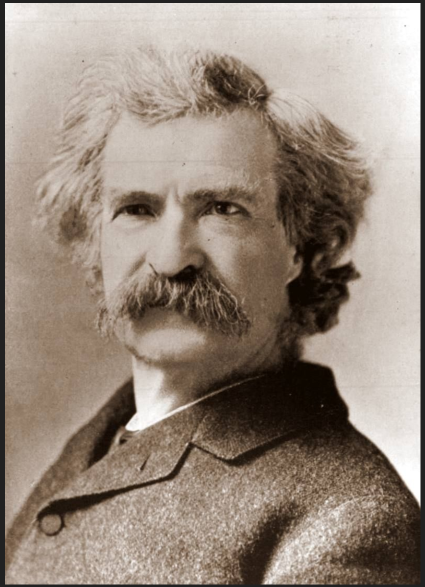
Why Bayesian inference makes us nervous

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)

Questions?

Here’s how you can get in touch with me.

brohrer@gmail.com

Blog (Data Science and Robots)
LinkedIn
Twitter
YouTube
Facebook
GitHub

Link to these slides:
https://docs.google.com/presentation/d/1325yenZP_VdHoVj-tU0AnbQUxFwb8Fl1VdyAAUxEzfg/edit?usp=sharing

Acknowledgements

Businessman Consulting Glowing Crystal Ball, InfoWire.dk, Public domain

Equation images from The Conjugate Prior for the Normal Distribution, Michael I. Jordan, Teodor Mihai Moldovan, February 8 th, 2010

Portrait of an unknown 19th-century Presbyterian clergyman. Identified as Thomas Bayes, Public domain

Woman blond hair, face long hair laleyla5, pixabay, Public Domain

hairfreaky long hair, Creative Commons 2.0

Amber Rose at the "PEOPLE Magazine Awards", Creative Commons Attribution-Share Alike 2.0 Generic

Neil Armstrong, Public Domain

A blond long-haired young lady-woman, "Mike" Michael L. Baird, Creative Commons Attribution 2.0 Generic

White Long Coated Puppy, Creative Commons Zero

Long Beach Comic & Horror Con 2011, Creative Commons 2.0

Donuts and milk, Creative Commons Zero

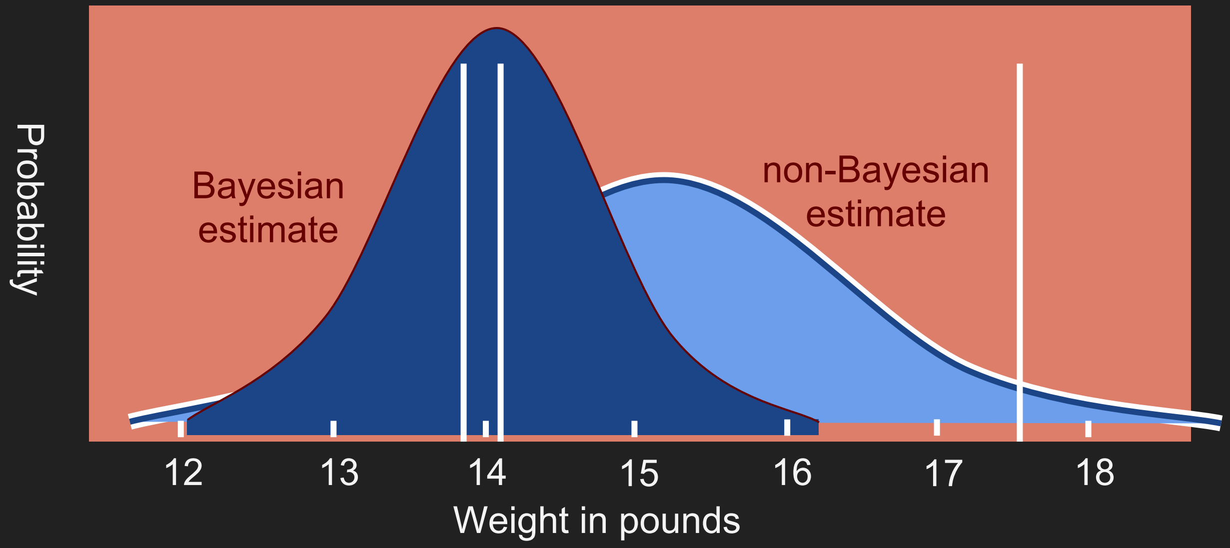
Photo of Reign of Terror, Diane Rohrer under Creative Commons Zero

Mark Twain, Public domain

Sherlock Holmes, Public domain

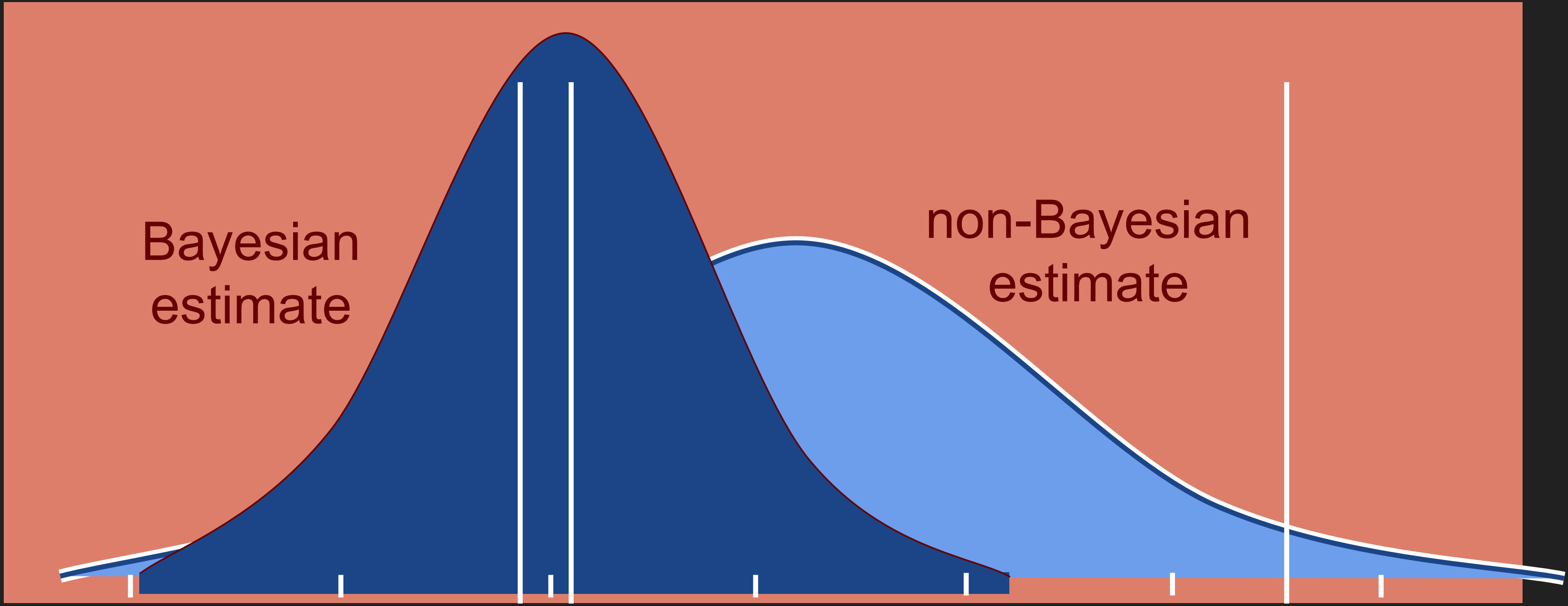
Red Queen with Alice, Public domain

How Bayesian Inference Works



by Brandon Rohrer

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



Bayesian
estimate
non-Bayesian
estimate