1.7 Bayesian Statistics

September 25, 2025

qBio I

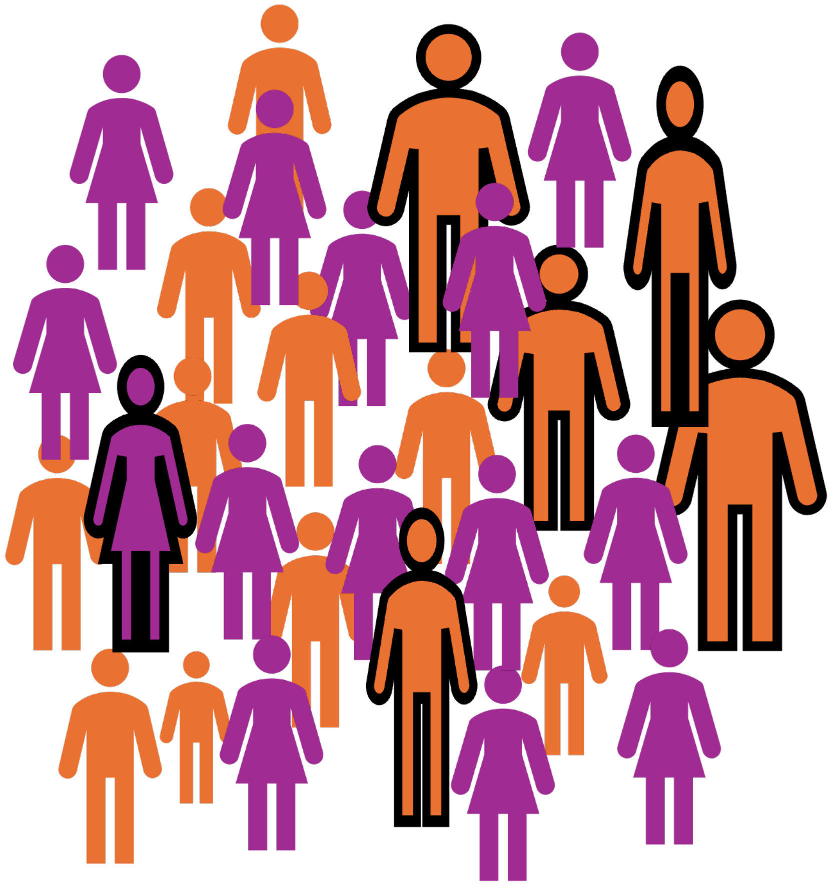
Slide Compilation by Aria Dandawate

Based on Lecture notes by Jason Banfelder

Bayesian Framework

- **Prior**: *what we believe about the system before seeing new data*
- **Likelihood**: *probability of observing the data given our hypothesis*
- **Posterior**: *updated belief about the hypothesis after considering the data*

Problem Setup



Class of Students

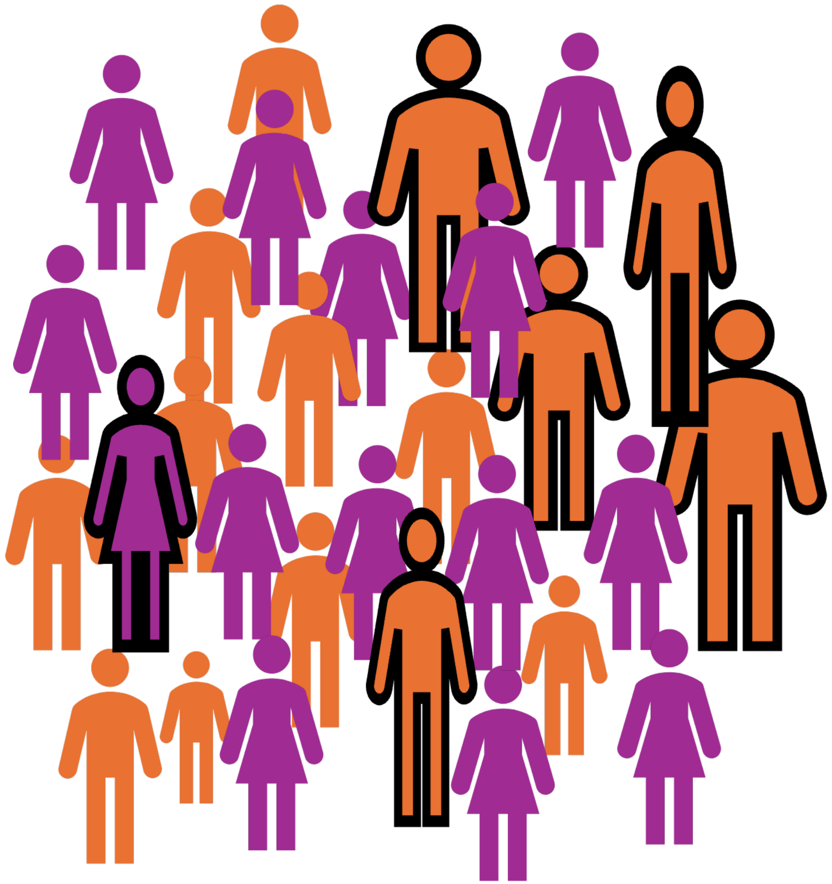
Let's pick a student at random.

P(T) = probability that student is ≥ 6 ft tall

P(F) = probability that student is female

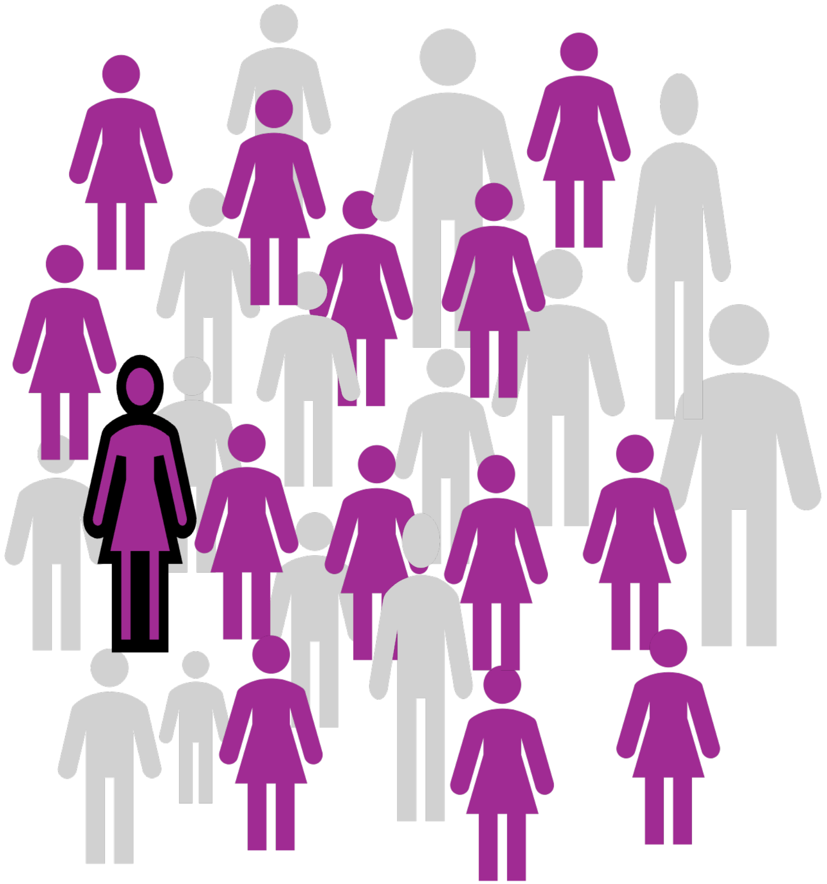
What is the probability that the chosen student is ≥ 6 *ft tall* ***and*** female?

Probability Table



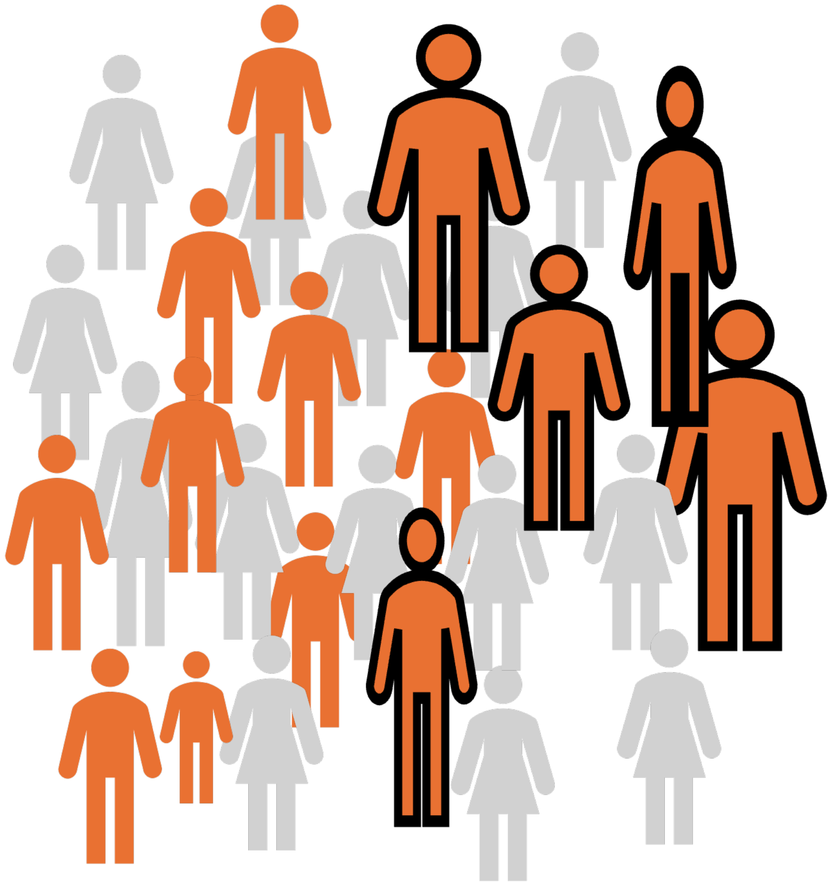
	Female	Male	
Tall			
Not Tall			

Probability Table



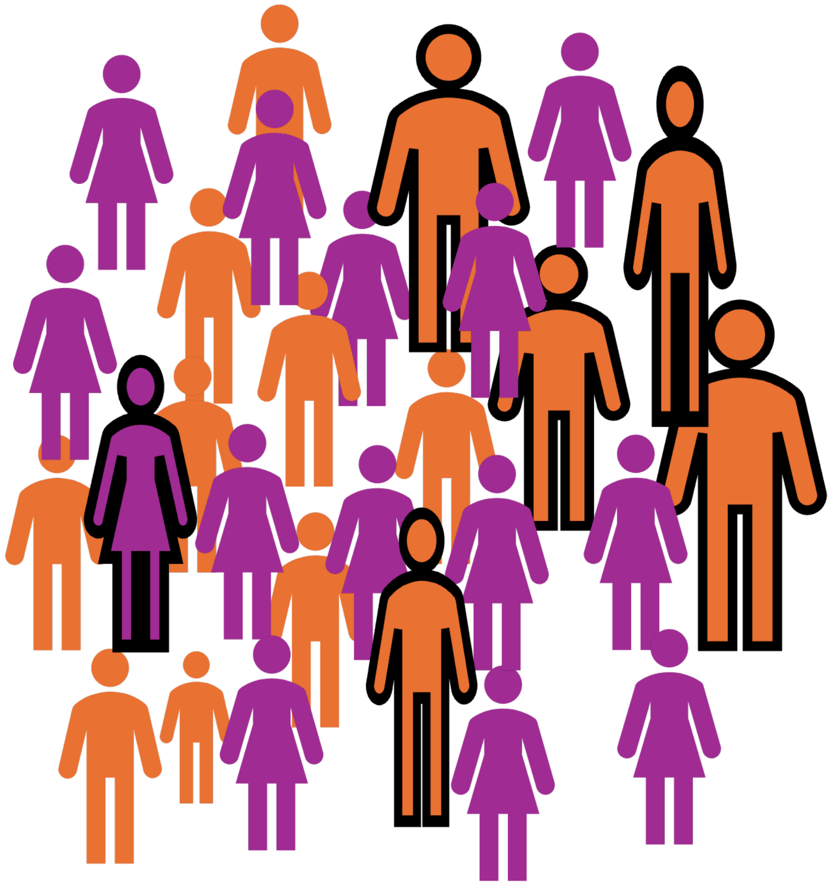
	Female	Male	
Tall	1		
Not Tall	13		
	14		

Probability Table



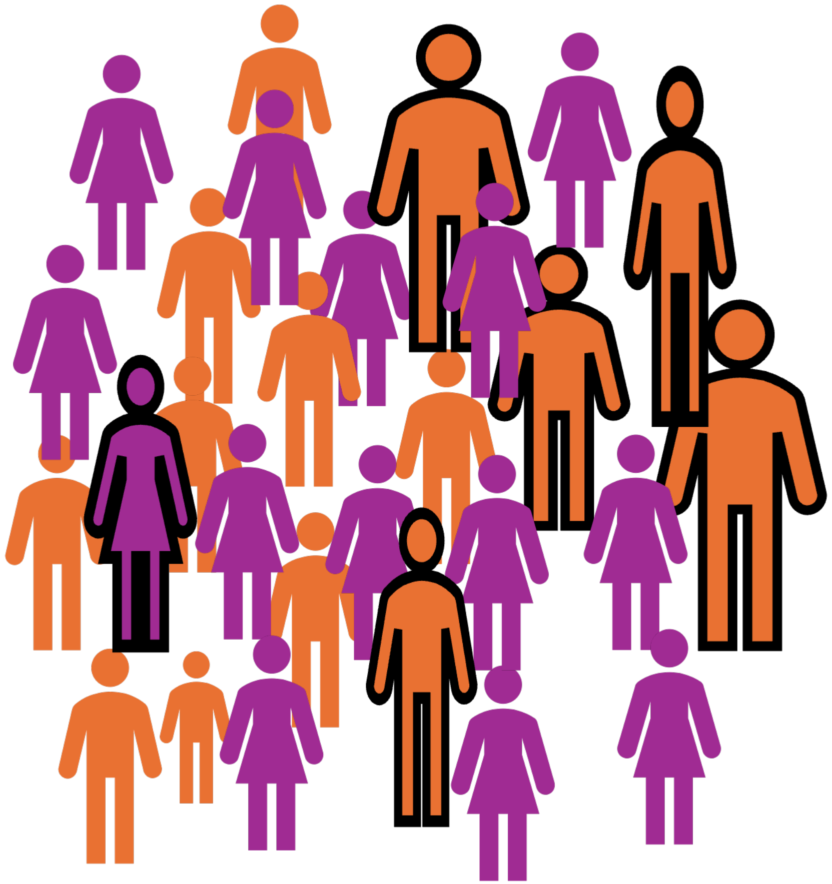
	Female	Male	
Tall	1	5	
Not Tall	13	9	
	14	14	

Probability Table



	Female	Male	
Tall	1	5	6
Not Tall	13	9	22
	14	14	28

Probability Table



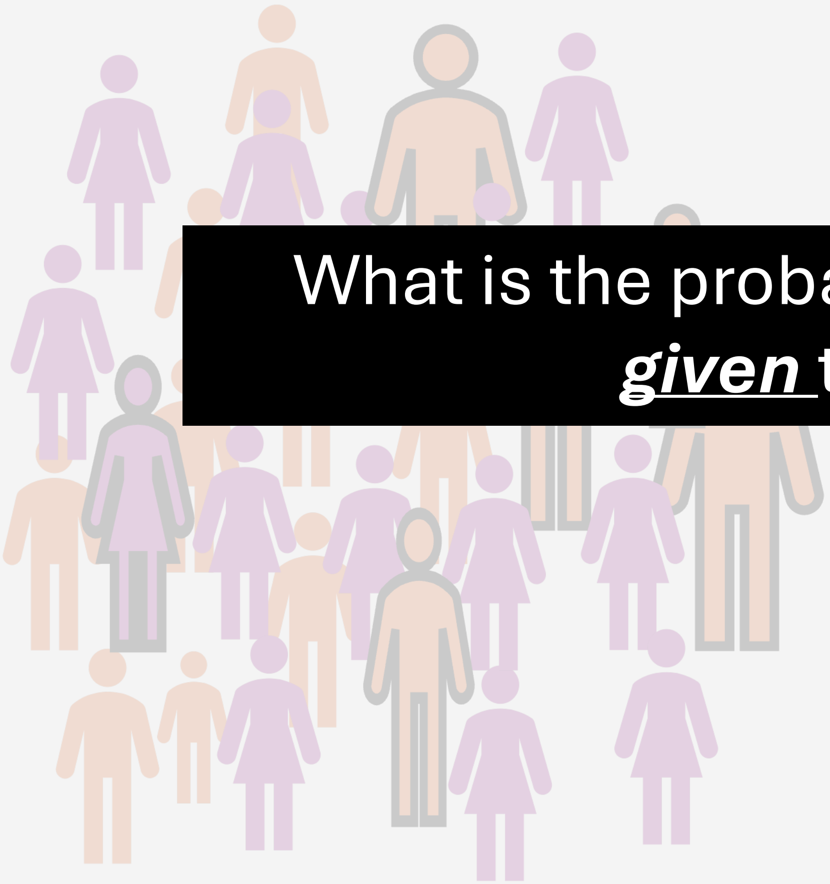
	Female	Male	
Tall	1	5	6
Not Tall	13	9	22
	14	14	28

Let's compute some probabilities from looking at this table:

$$P(T) = \frac{6}{28}, P(F) = \frac{14}{28}$$

$$P(T \cap F) = P(F \cap T) = \mathrm{P(T, F)} = \mathrm{P(F,T)} = \frac{1}{28}$$

Probability Table



What is the probability that a student is ≥ 6ft tall ***given*** that they are female?

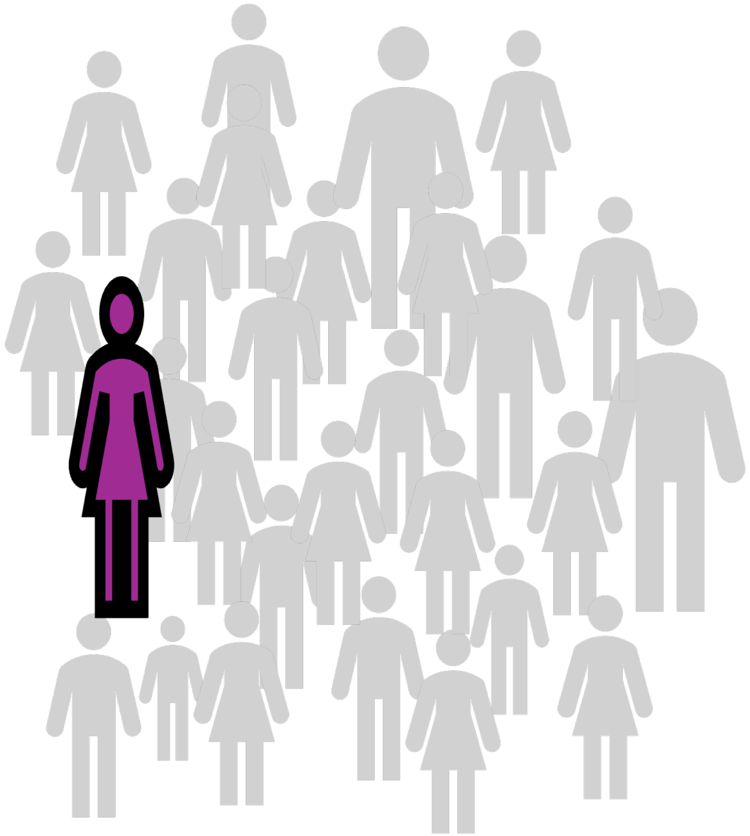
		6
		22
14	14	28

Let's compute some probabilities from looking at this table:

$$P(T) = \frac{6}{28}, P(F) = \frac{14}{28}$$

$$P(T \cap F) = P(F \cap T) = \mathrm{P(T, F)} = \mathrm{P(F,T)} = \frac{1}{28}$$

Joint Probability using Conditional Probability

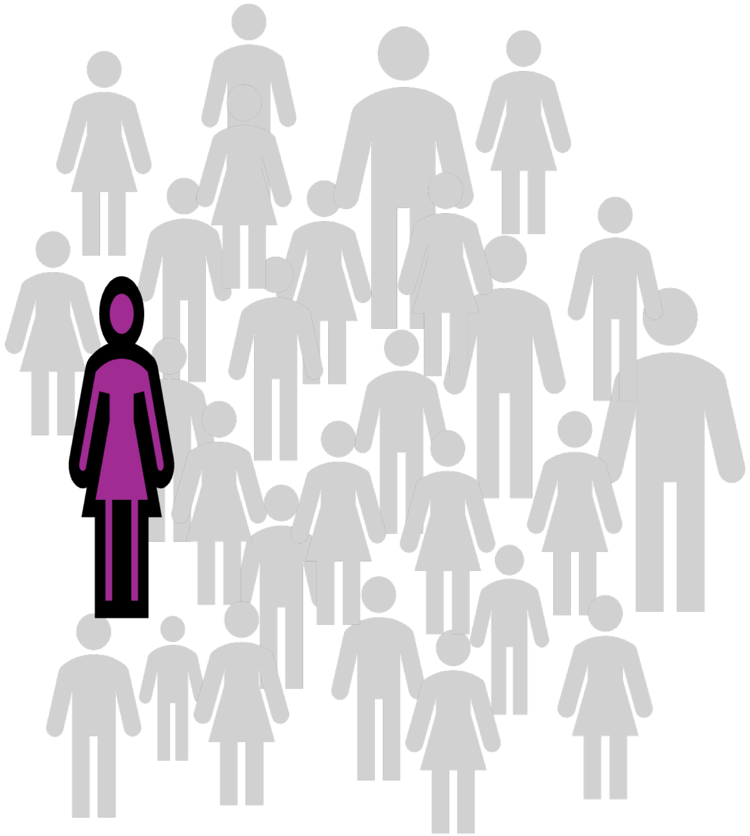


	Female	**Male**	
Tall	1	5	6
Not Tall	13	9	22
	14	14	28

Conditional Probability:

$P(T \mid F)$

Joint Probability using Conditional Probability

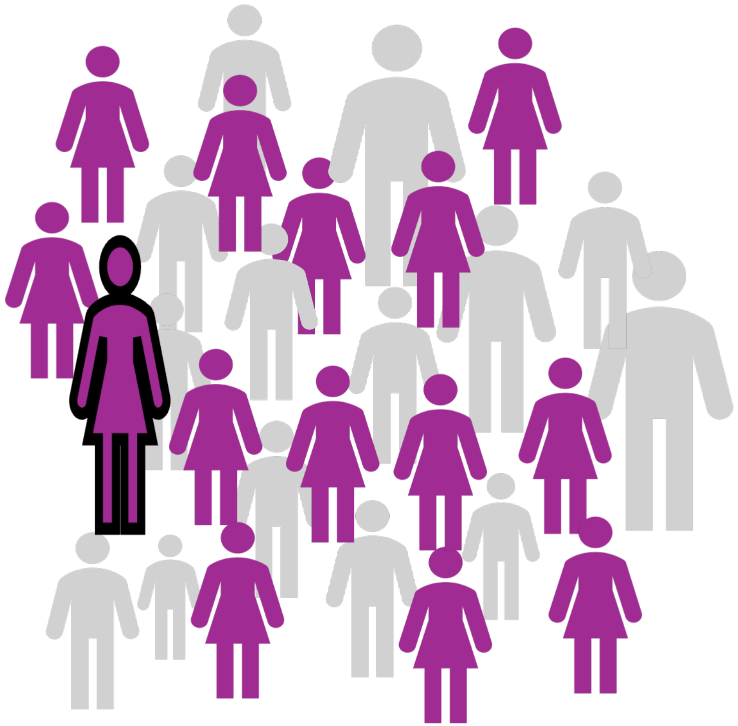


	Female	Male	
Tall	**1**	5	6
Not Tall	13	9	22
	14	14	28

Conditional Probability:

$$P(T \mid F) = \frac{1}{14}$$

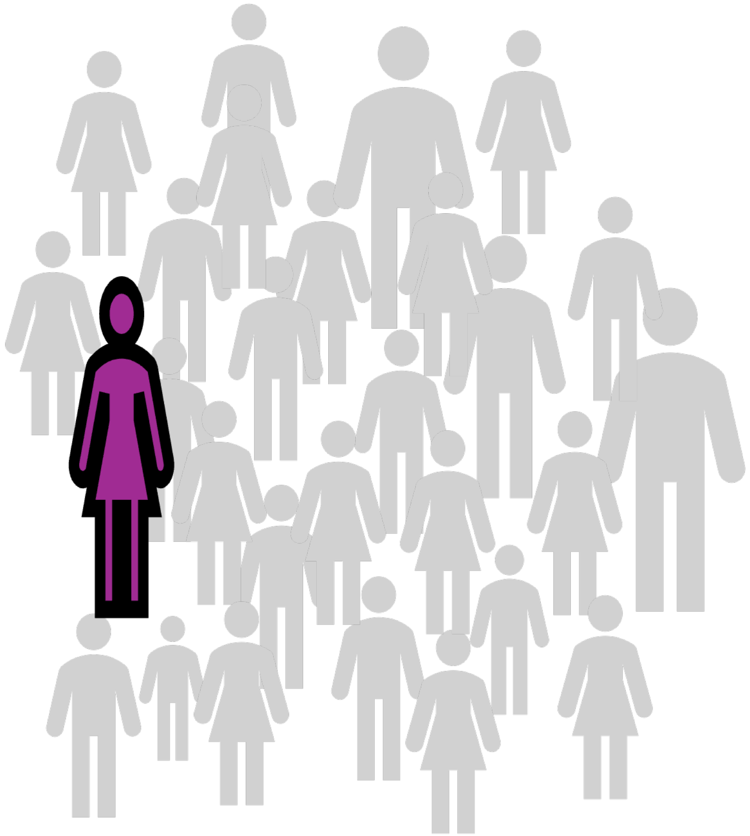
Joint Probability using Conditional Probability



	Female	Male	
Tall	**1**	5	6
Not Tall	13	9	22
	14	14	**28**

$$P(T, F) = P(F) * P(T \mid F)$$

Joint Probability using Conditional Probability

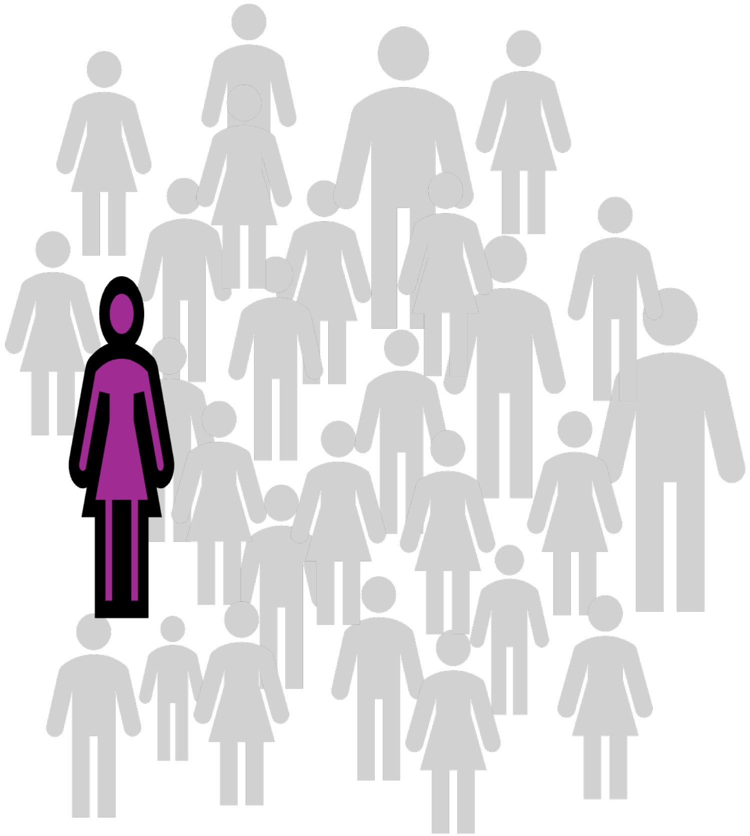


	Female	Male	
Tall	1	5	6
Not Tall	13	9	22
	14	14	28

$$P(T, F) = P(F) * P(T \mid F)$$

$$\frac{14}{28} * \frac{1}{14} = 0.5 * 0.07 = 0.036$$

Joint Probability using Conditional Probability



	Female	Male	
Tall	1	5	6
Not Tall	13	9	22
	14	14	28

P(T, F) = P(T) * P(F | T)

$$\frac{6}{28} * \frac{1}{6} = 0.21 * 0.17 = 0.036$$

Joint Probability (dependence case)

$$\boxed{\begin{matrix} P(T,F) \\ P(T,F) \end{matrix}} = \boxed{\begin{matrix} P(F) \\ P(T) \end{matrix}} * \boxed{\begin{matrix} P(T \mid F) \\ P(F \mid T) \end{matrix}}$$

Joint Prior Conditional

Deriving Bayes' rule from the Joint probability

$$P(T, F) = P(F) * P(T \mid F)$$
$$P(T, F) = P(T) * P(F \mid T)$$

Set them equal, and solve for P(T | F).

$$P(F) * P(T \mid F) = P(T) * P(F \mid T)$$

$$P(T \mid F) = \frac{P(T) * P(F \mid T)}{P(F)}$$

Bayes' Rule

	Female	Male	
Tall			
Not Tall			

"Margins" of the probability table

Generalizable formula:

Prior probability

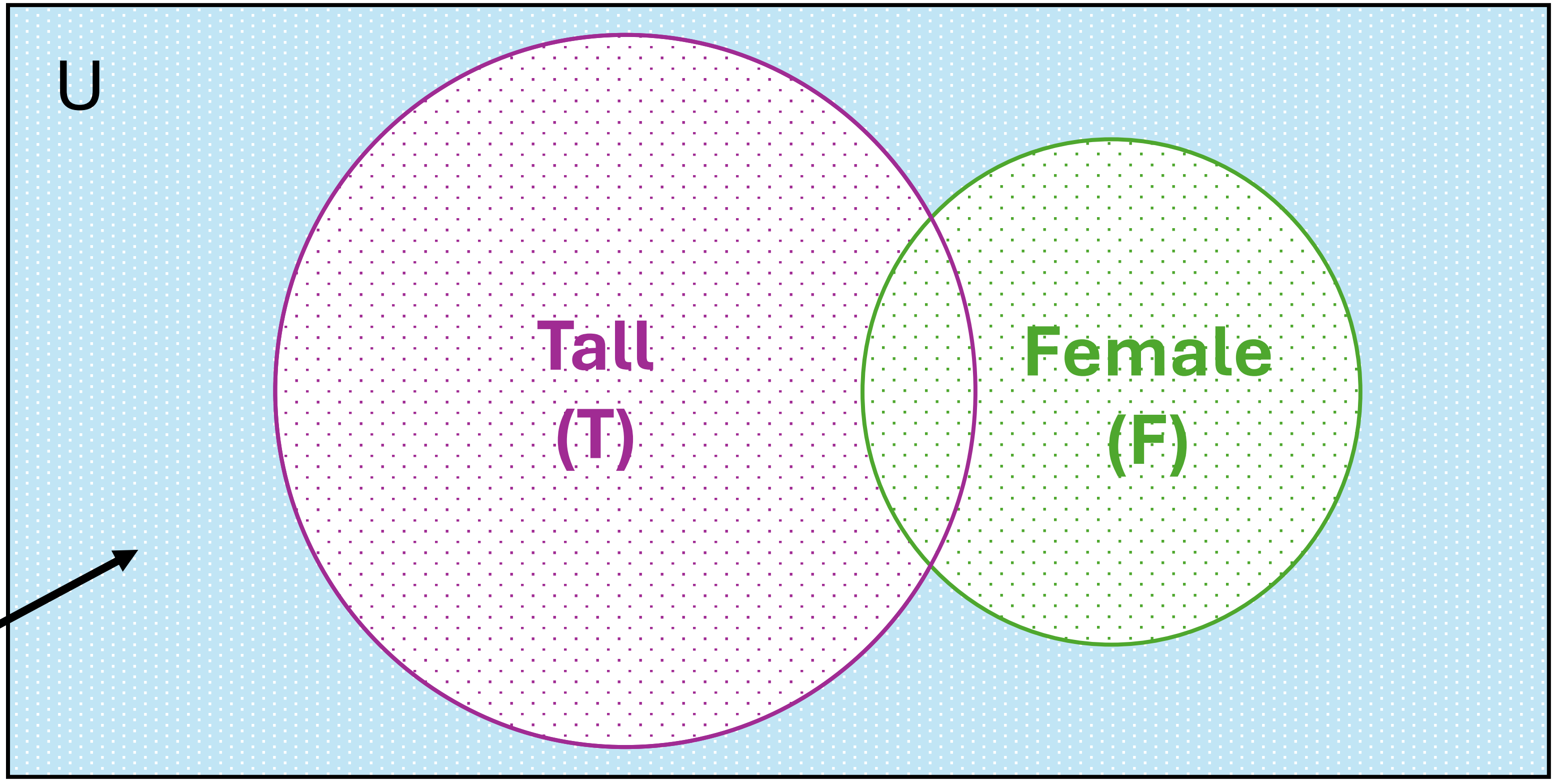
$$\boxed{P(A/B)} = \frac{\boxed{P(A)} * \boxed{P(B|A)}}{\boxed{P(B)}}$$

Likelihood

Posterior probability

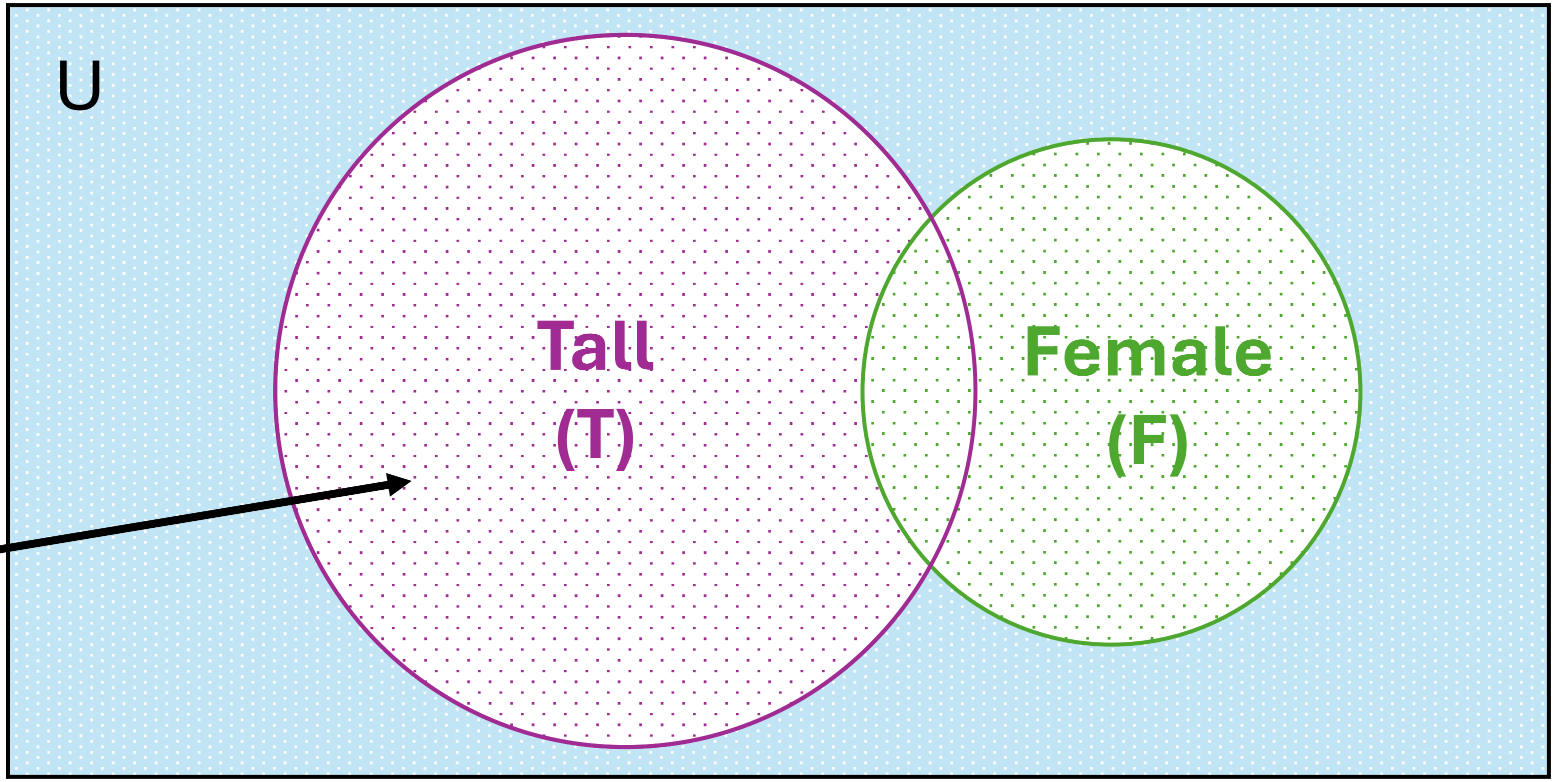
Marginal

Set Theory intuition of Bayes' Rule



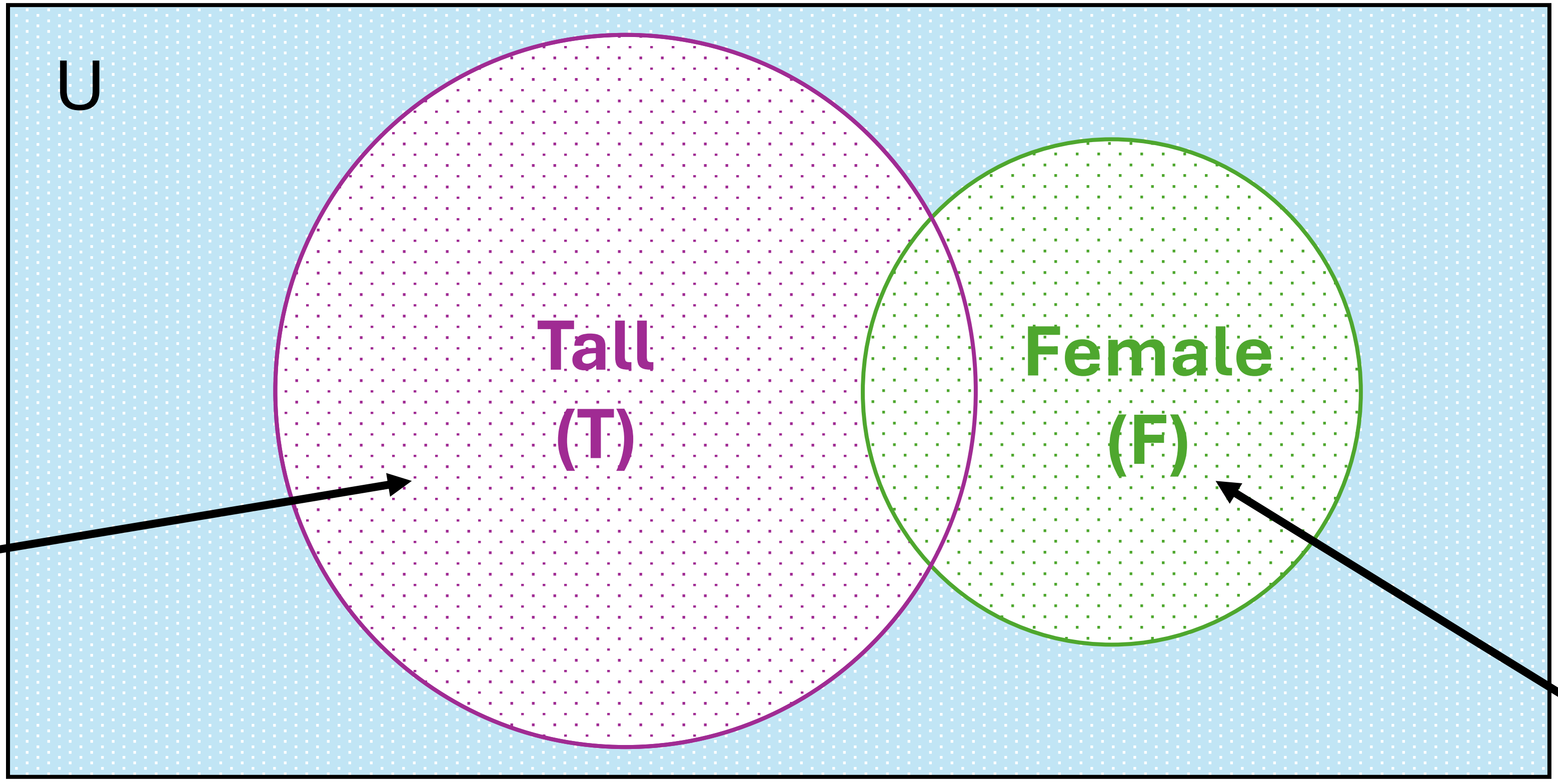
U = all events in universe

Set Theory intuition of Bayes' Rule



$$P(T) = \frac{T}{U}$$

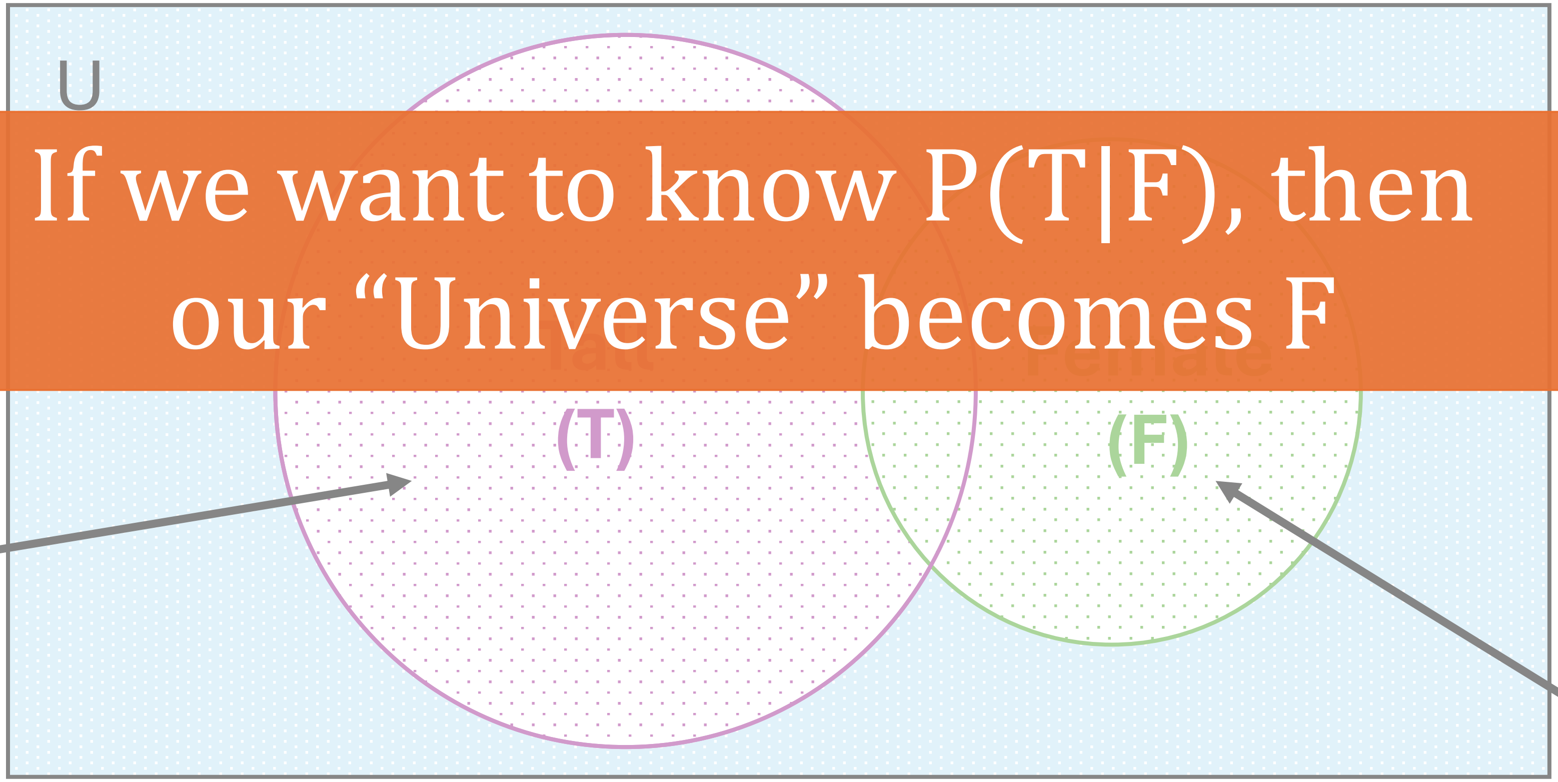
Set Theory intuition of Bayes' Rule



$$P(T) = \frac{T}{U}$$

$$P(F) = \frac{F}{U}$$

Set Theory intuition of Bayes' Rule

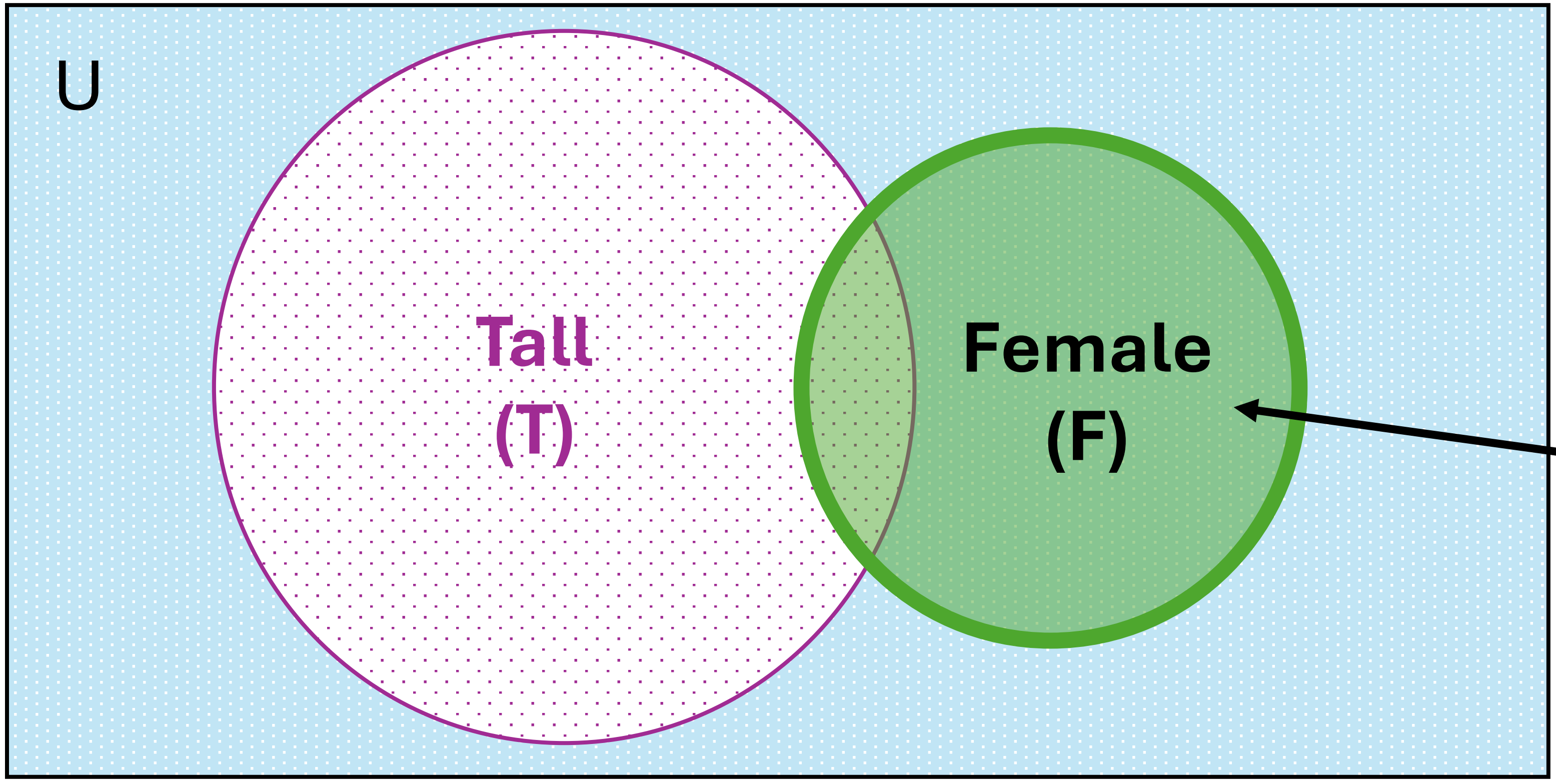


If we want to know P(T|F), then our "Universe" becomes F

$$P(T) = \frac{T}{U}$$

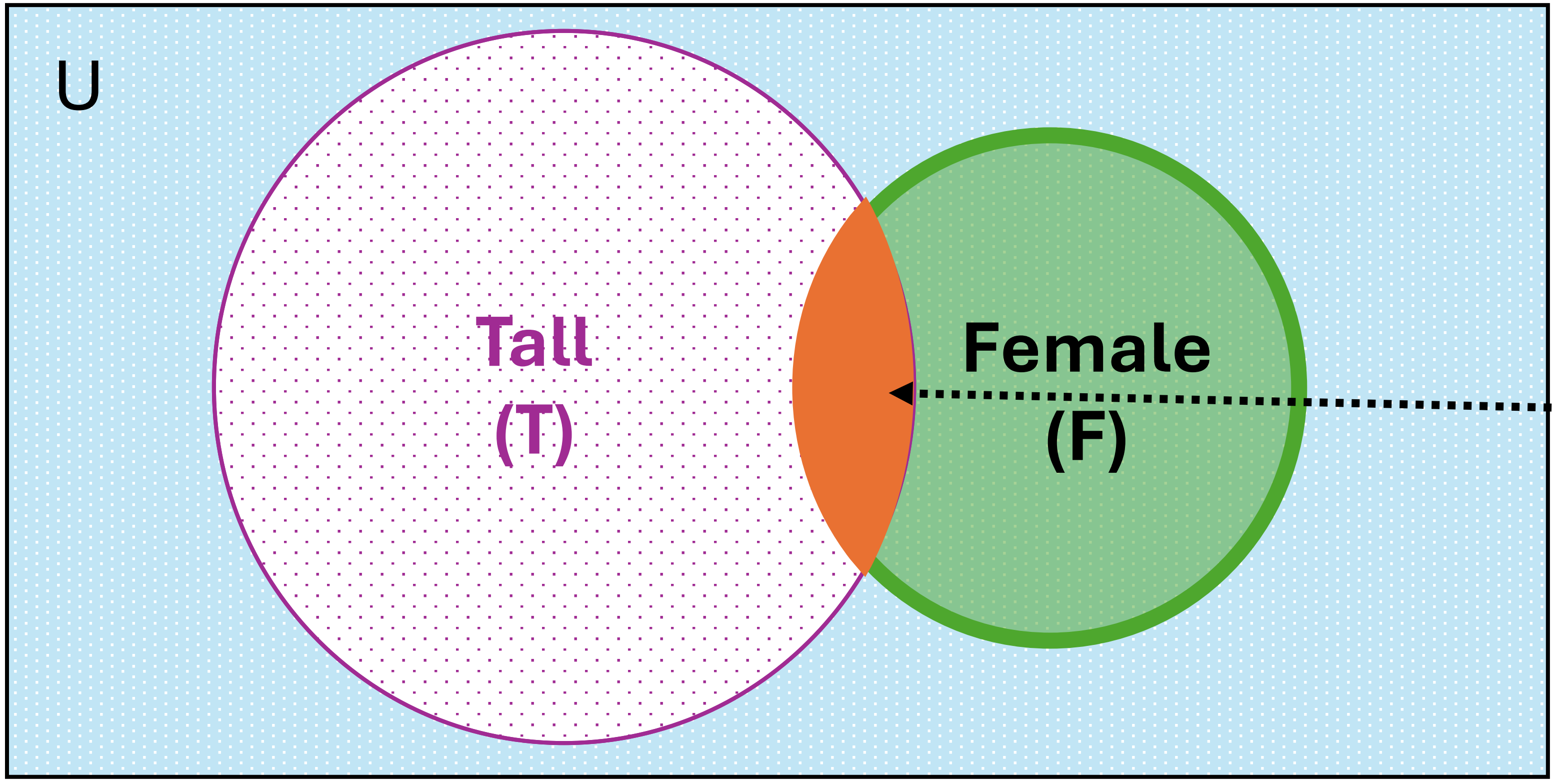
$$P(F) = \frac{F}{U}$$

Set Theory intuition of Bayes' Rule



$$\frac{?}{F}$$

Set Theory intuition of Bayes' Rule



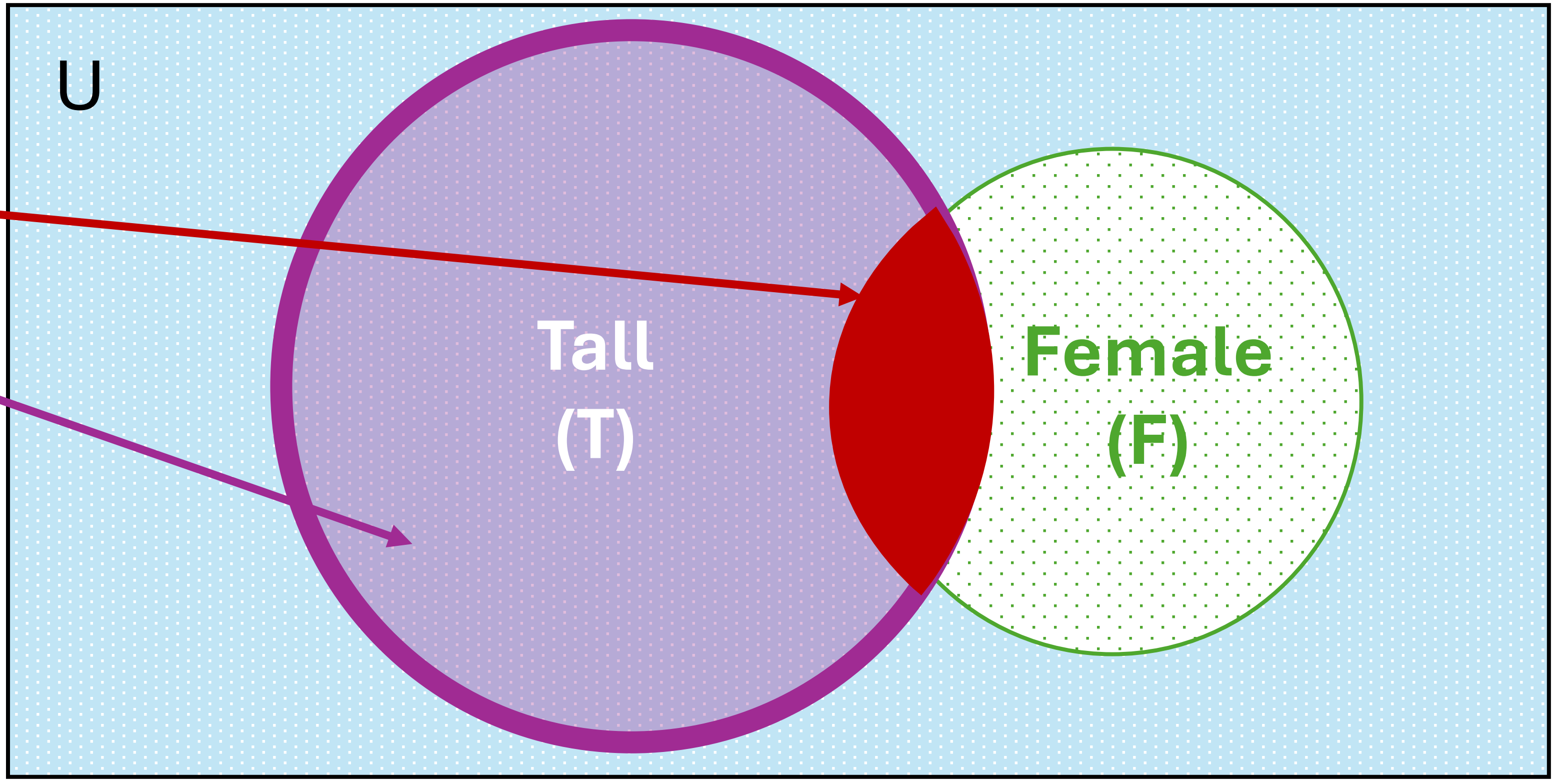
$$\frac{T \cap F}{F}$$

$$\frac{P(T \cap F)}{P(F)}$$

Set Theory intuition of Bayes' Rule

$$P(T \cap F) =$$

$$P(T) * \mathbf{P(F|T)}$$



Applying Bayes' Rule

- 1% of women in a given population have breast cancer.
- If a woman has breast cancer, there is a 90% chance that a particular diagnostic test will return a positive result. (90% TP)
- If a woman does not have breast cancer, there is a 10% chance that this diagnostic test will return a positive result. (10% FP)

What is the probability that a patient actually has cancer given a positive test result?

Write our events in terms of Bayes' rule

- C = having cancer
- H = not having cancer
- + = positive result
- - = negative result

What is P(C|+)?

$$P(C/+) = \frac{P(C) * P(+|C)}{P(+)}$$

Applying Bayes' Rule

- 1% of women in a given population have breast cancer.
- If a woman has breast cancer, there is a 90% chance that a particular diagnostic test will return a positive result. (90% TP)
- If a woman does not have breast cancer, there is a 10% chance that this diagnostic test will return a positive result. (10% FP)

$$P(C|+) = \frac{P(C) * P(+|C)}{P(+)}$$

Which one of these statements gives each of our probabilities in Bayes' rule?

Applying Bayes' Rule

- 1% of women in a given population have breast cancer.

- If a woman has breast cancer, there is a 90% chance that a particular diagnostic test will return a positive result. (90% TP)

- If a woman does not have breast cancer, there is a 10% chance that this diagnostic test will return a positive result. (10% FP)

$$P(C|+) = \frac{P(C) * P(+|C)}{P(+)}$$

When do we get a positive result??

Computing P(+)

Given:

$P(C) = 0.01$
$P(+ \mid C) = 0.9$
$P(+ \mid H) = 0.1$

Then:
$P(H) = 1 - P(C) = 0.99$

$$P(+) = \underbrace{P(\text{true positives})}_{\substack{P(C,\,+) \\ P(C) * P(+ \mid C)}} + \underbrace{P(\text{false positives})}_{\substack{P(H,\,+) \\ P(H) * P(+ \mid H)}}$$

$$\begin{aligned} P(+) &= P(C) * P(+ \mid C) + P(H) * P(+ \mid H) \\ &= 0.01 * 0.9 + 0.99 * 0.1 \\ &= 0.009 + 0.099 \\ &= 0.108 \end{aligned}$$

Almost 10% of our positives are false positives !

Plug into Bayes' Rule

$$\mathrm{P}(\mathrm{C}|+) = \frac{\mathrm{P}(+|\mathrm{C}) \cdot \mathrm{P}(\mathrm{C})}{P(+)}$$

$$= \frac{0.9 \cdot 0.01}{0.108}$$

$$= 8.3\%$$

Model optimization using Bayesian Methods

Estimating the Bias of Coin

- **Problem Setup:**
  - Flip a coin **20 times**
  - Observe **13 heads**
  - Question: *What is the true probability of landing heads in the long-run (i.e. bias of this coin)?*

Use a Binomial test!

```r
binom.test(13, 20)
```

```
##
##  Exact binomial test
##
## data:  13 and 20
## number of successes = 13, number of trials = 20, p-value =
## 0.2632
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.4078115 0.8460908
## sample estimates:
## probability of success
##                   0.65
```

Bayesian Approach: Estimating Bias of a Coin

- Bayes' rules relates to two types of events:
  - Model (M) = coin's bias
  - Data (D) = observed flips

Bayesian thinking: *What is the probability of each model (M) given the observed Data (D)?*

```r
dbinom(13, size = 20, prob = 0.5)

## [1] 0.07392883
```

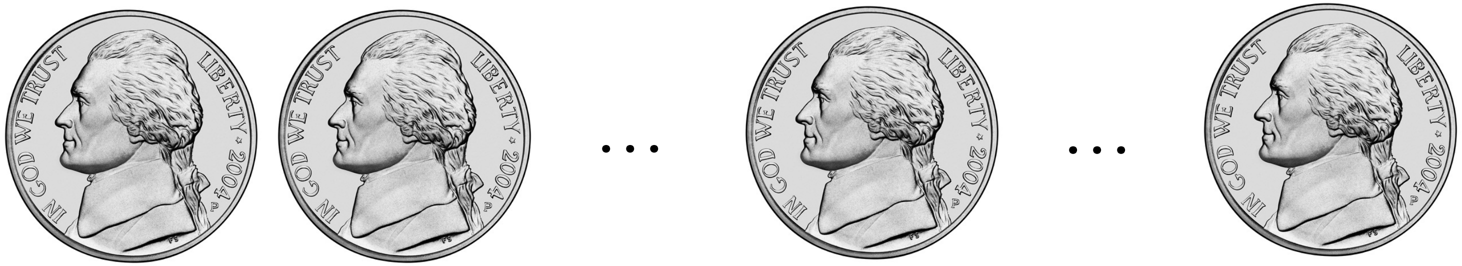
"likelihood of observing 13 heads given that we have a fair coin"

```r
dbinom(13, size = 20, prob = 0.25)

## [1] 0.0001541923
```

"likelihood of observing 13 heads given that we have a coin with bias 0.25"

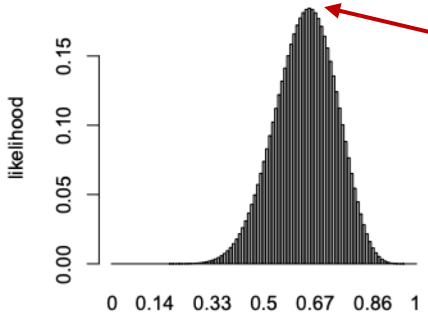
Setup:

1) Take 101 coins with different bias:



$M_0 = 0.0$ $M_1 = 0.01$ … $M_{50} = 0.50$ … $M_{100} = 1.0$

```r
coin.bias <- seq(from = 0, to = 1, by = 0.01)
likelihood <- dbinom(13, 20, prob = coin.bias)
barplot(likelihood, names.arg = coin.bias, ylab = "likelihood")
```

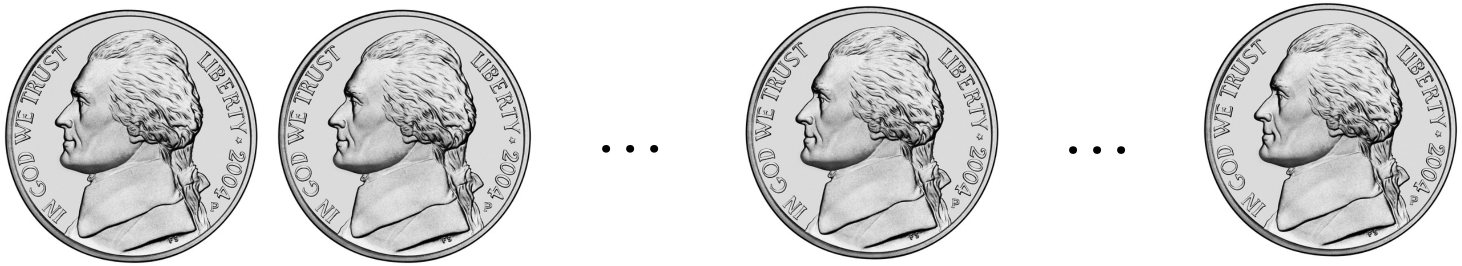


Maximum likelihood is ~0.65

This is not yet a probability distribution

Setup:

1) Take 101 coins with different bias:



$M_0 = 0.0$ $M_1 = 0.01$ $M_{50} = 0.50$ $M_{100} = 1.0$

2) Choose a coin ("model") with bias b at random:

$$P(M_b) = 1/101 = 0.0099$$

3) Flip the coin 20 times and observe 13 heads:

$$\textbf{What is } P(M_{0.5} \mid D_{13})\textbf{?}$$

What is the probability that we have a fair coin given we observed 13 heads?

Bayes' Rule for Coin Bias

$$\mathbf{P(M_{0.5} \mid D_{13})} = \frac{P(M_{0.5})\, P(D_{13}|M_{0.5})}{P(D_{13})}$$

Recall, we computed $P(D_{13}|M_{0.5})$ already:

```
dbinom(13, size = 20, prob = 0.5)

## [1] 0.07392883
```

And we already said $P(M_{0.5})$ = 0.0099

How do we calculate the denominator?

The Denominator

- Denominator = probability of **observing 13 heads across all models**:

$$P(D_{13}) = \sum_{b=0.0}^{1} \underbrace{P(D_{13} \mid M_b) \cdot P(M_b)}_{\text{Joint probability of observing 13 heads and coin with bias } b}$$

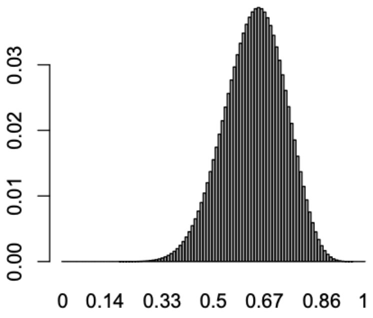
```r
(p.d13 <- sum(dbinom(13, 20, coin.bias) * (1 / 101)))
## [1] 0.04714757
```

Plug and Chug!

$$P(M_{0.5} \mid D_{13}) = \frac{P(M_{0.5})\, P(D_{13}|M_{0.5})}{P(D_{13})} = \frac{0.0739 \cdot 0.0099}{0.04715} = 0.0155$$

Therefore we can calculate $P(M_b \mid D_{13})$ for all possible bias' we are considering....

```
posterior.probability <- dbinom(13, 20, coin.bias) * (1 / 101) / p.d13
sum(posterior.probability)

## [1] 1

barplot(posterior.probability, names.arg = coin.bias)
```



This is a probability distribution!

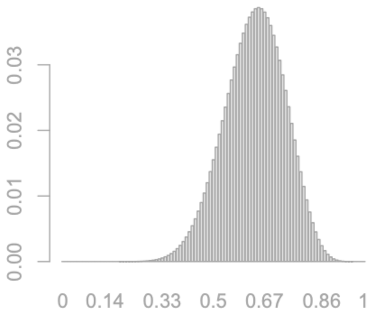
Plug and Chug!

$$P(M_{0.5} \mid D_{13}) = \frac{P(M_{0.5})\,P(D_{13}|M_{0.5})}{P(D_{13})} = \frac{0.0739 \cdot 0.0099}{0.04715} = 0.0155$$

There

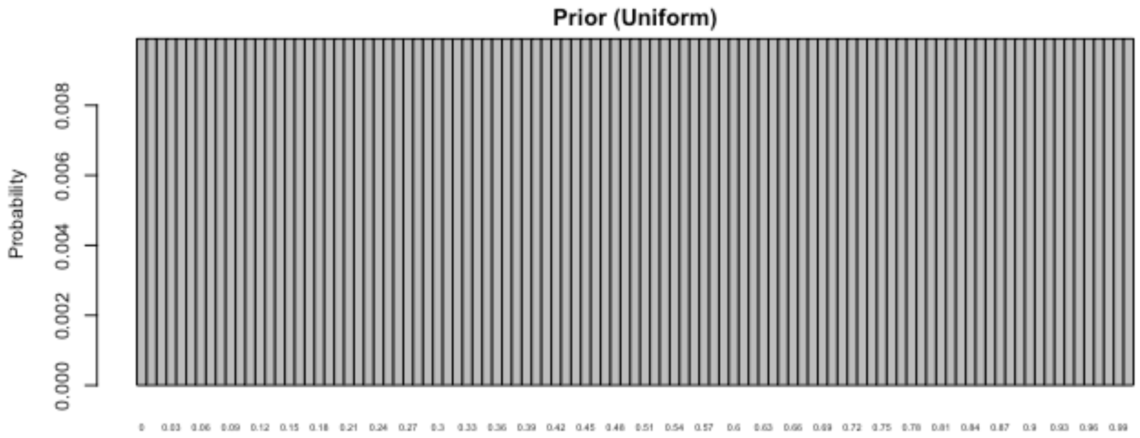
How is this different from running dbinom?

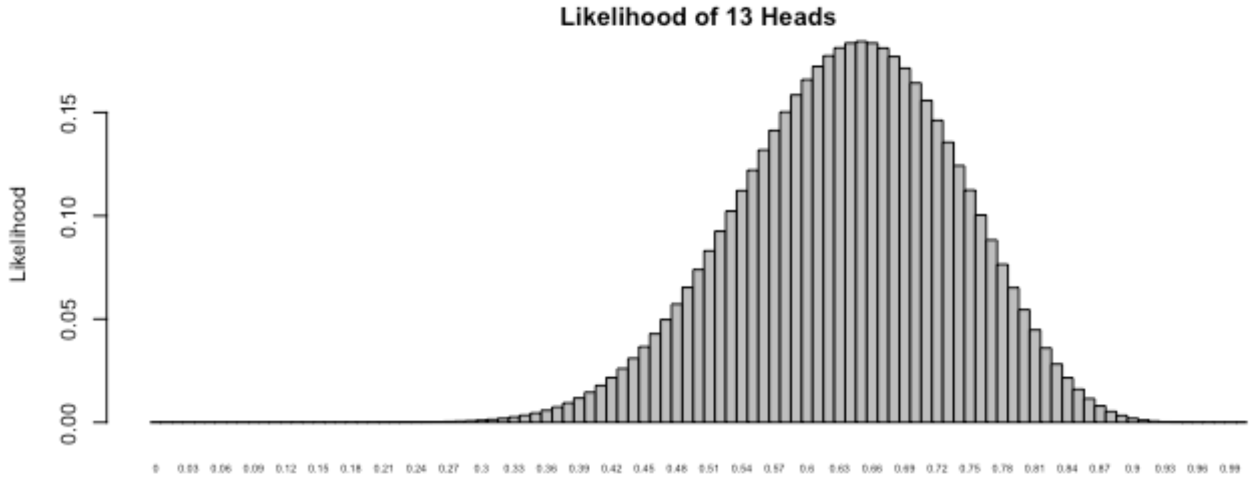
```
pos
sum(posterior.probability)

## [1] 1

barplot(posterior.probability, names.arg = coin.bias)
```

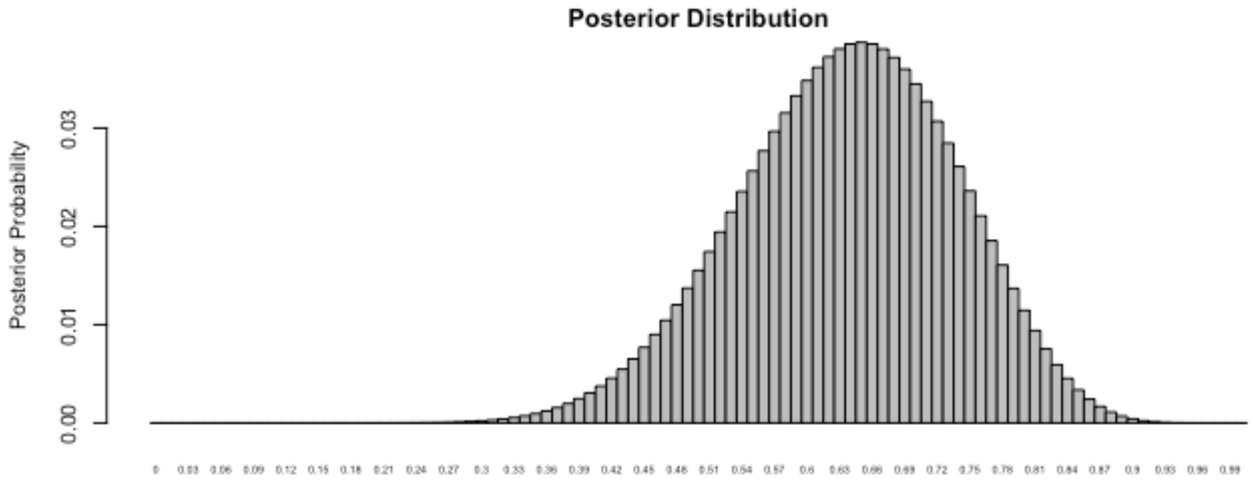


This is a probability distribution!

Defining Priors: Uniform Prior



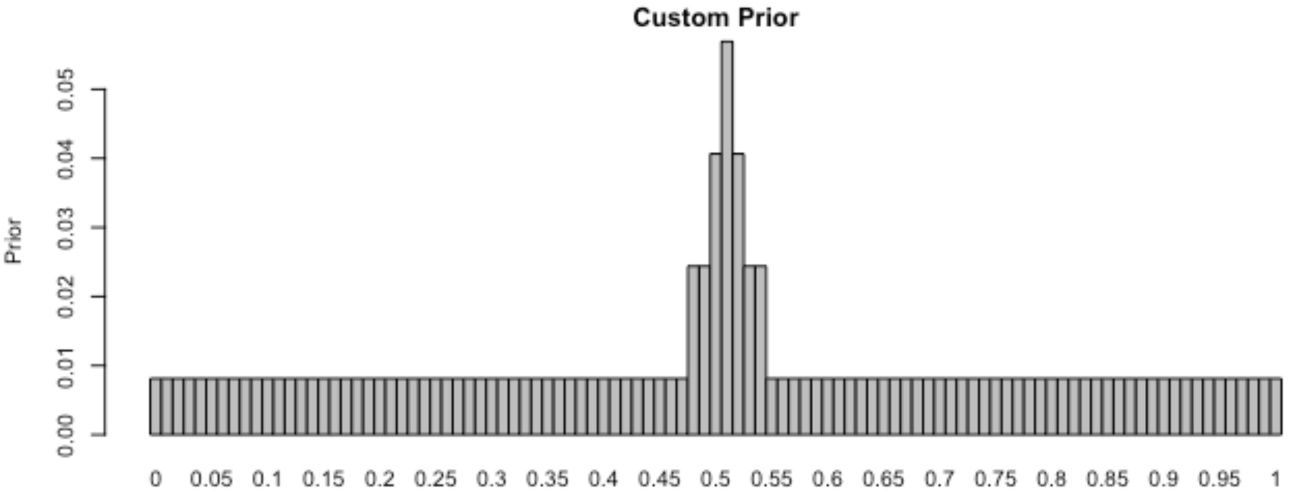


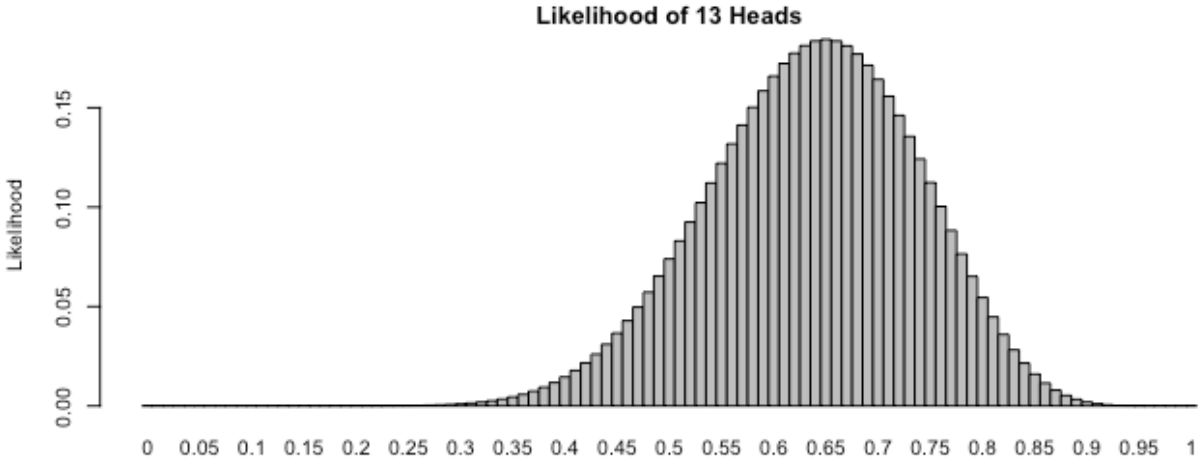


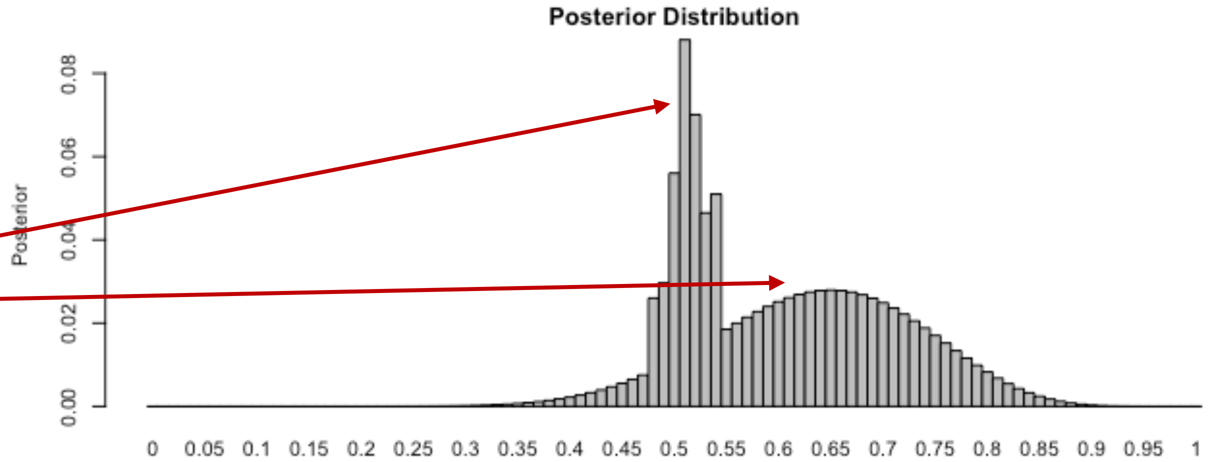
- **Uniform Prior:** *equal probability of choosing all coin biases*
- ***the Likelihood looks like the Posterior!***

But what if we know that we are most likely to choose a "fair" coin?

Defining Priors: Custom Prior







Bimodal posterior = interplay between prior and data

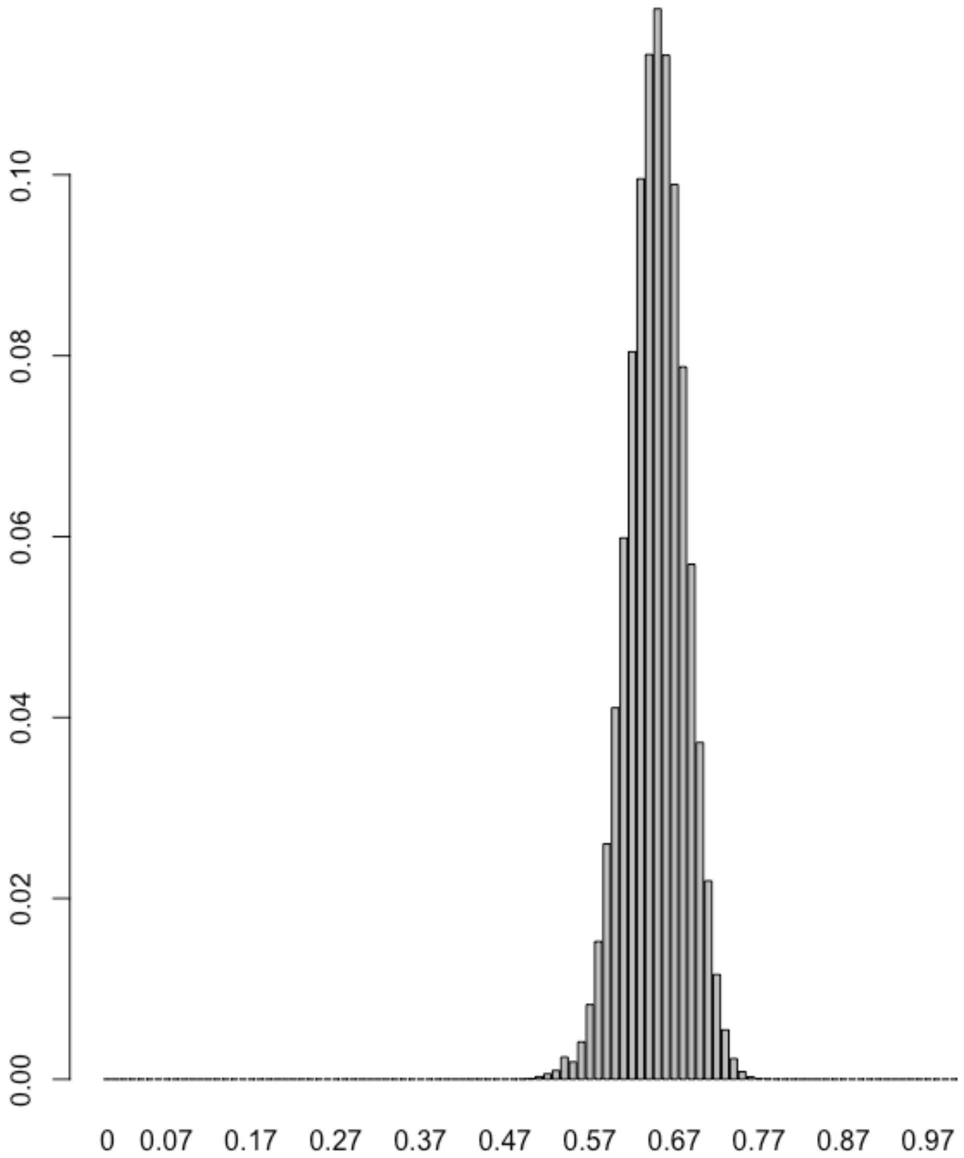
Gather more evidence

- Now we observe:

$$\frac{130\ heads}{200\ coin\ flips}$$

- Evidence now overwhelms the prior!

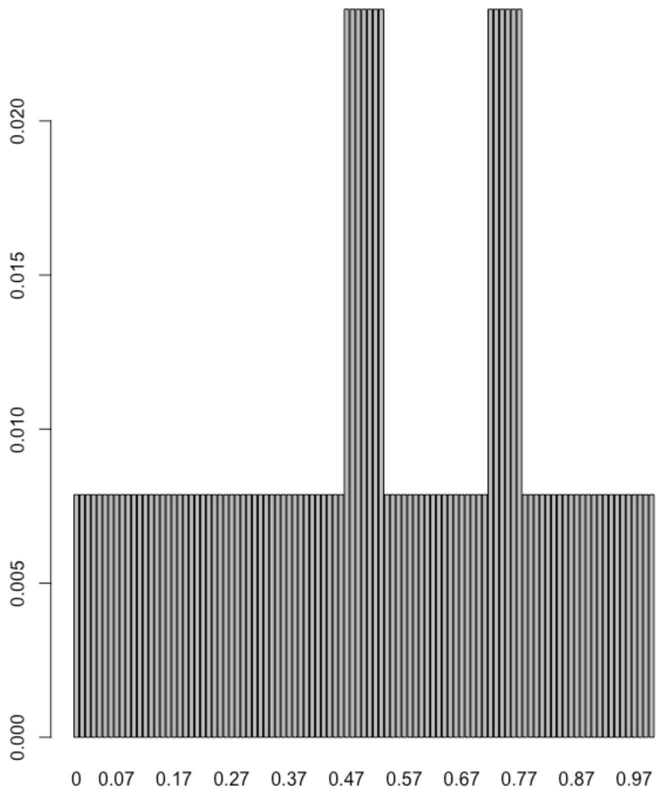
- Even though we had a prior belief, the evidence is overwhelming



What if our coin is from a magic shop?

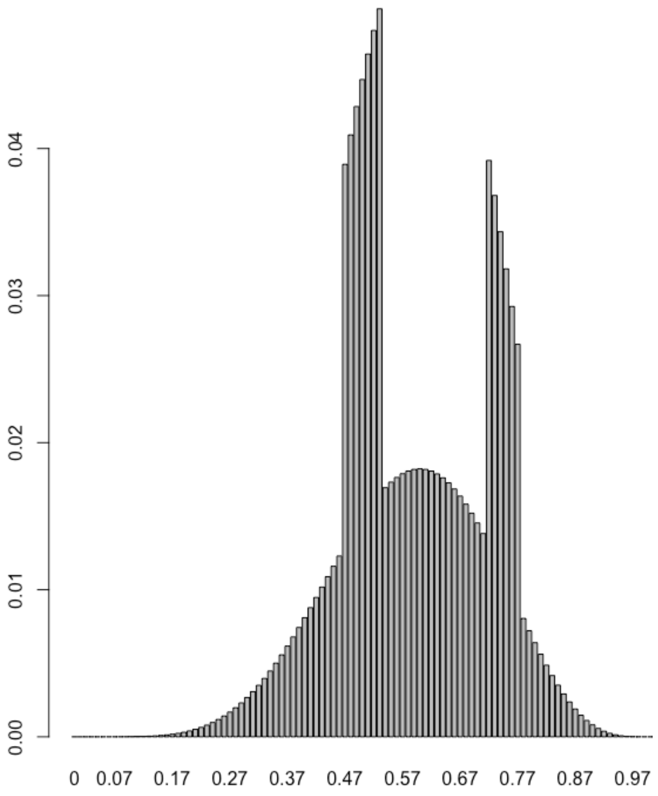
- We know that the shop sells coins with 0.75 bias

Our Prior



We observe 6 heads out of 10 coin flips

Our Posterior



Choosing a prior

- “Strong” priors
  - Need strong evidence (previous experiments, expert knowledge, or well-established theory)
  - Can overwhelm the posterior

Recap

- **Bayesian Approaches:**
  - Allow us to incorporate our prior beliefs into our analysis

  - **BUT** we should be cautious that our prior does not influence our analysis
    - Don’t pick the prior based on what you ***want*** the result to look like

"*Extraordinary claims require extraordinary evidence*"
- Carl Sagan