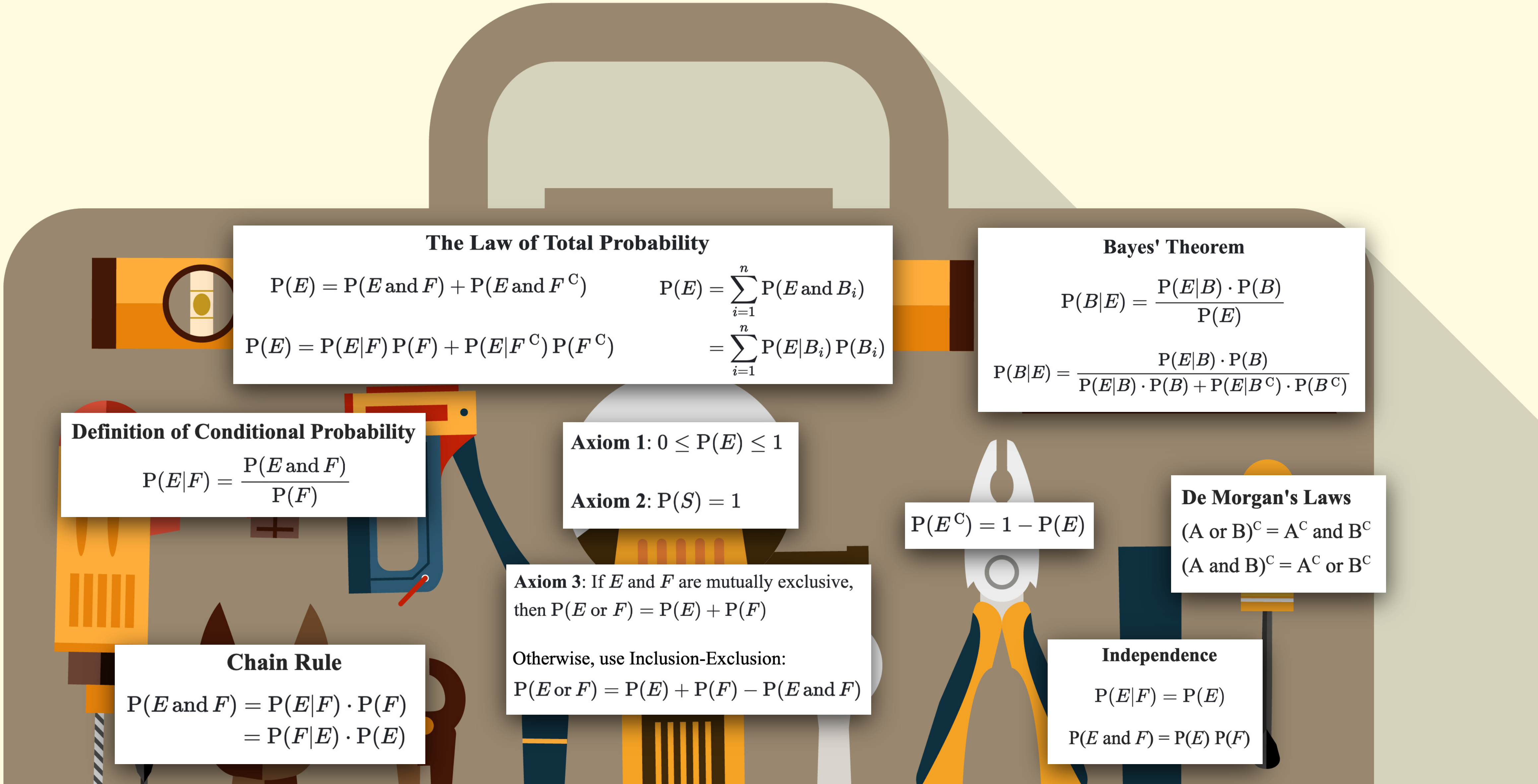
Conditional Probability & Bayes

CS109

The Core Probability Toolkit

The Core Probability Toolkit



The Law of Total Probability

$$P(E) = P(E \text{ and } F) + P(E \text{ and } F^{C})$$

$$P(E) = P(E|F)\,P(F) + P(E|F^{C})\,P(F^{C})$$

$$P(E) = \sum_{i=1}^{n} P(E \text{ and } B_i) = \sum_{i=1}^{n} P(E|B_i)\,P(B_i)$$

Bayes' Theorem

$$P(B|E) = \frac{P(E|B)\cdot P(B)}{P(E)}$$

$$P(B|E) = \frac{P(E|B)\cdot P(B)}{P(E|B)\cdot P(B) + P(E|B^{C})\cdot P(B^{C})}$$

Definition of Conditional Probability

$$P(E|F) = \frac{P(E \text{ and } F)}{P(F)}$$

Axiom 1: $0 \le P(E) \le 1$

Axiom 2: $P(S) = 1$

$$P(E^{C}) = 1 - P(E)$$

De Morgan's Laws

$(A \text{ or } B)^{C} = A^{C} \text{ and } B^{C}$

$(A \text{ and } B)^{C} = A^{C} \text{ or } B^{C}$

Axiom 3: If E and F are mutually exclusive, then $P(E \text{ or } F) = P(E) + P(F)$

Otherwise, use Inclusion-Exclusion:

$$P(E \text{ or } F) = P(E) + P(F) - P(E \text{ and } F)$$

Chain Rule

$$P(E \text{ and } F) = P(E|F)\cdot P(F) = P(F|E)\cdot P(E)$$

Independence

$$P(E|F) = P(E)$$

$$P(E \text{ and } F) = P(E)\,P(F)$$

The Core Probability Toolkit

The Law of Total Probability

$$\mathrm{P}(E) = \mathrm{P}(E \text{ and } F) + \mathrm{P}(E \text{ and } F^{\mathrm{C}})$$

$$\mathrm{P}(E) = \mathrm{P}(E|F)\,\mathrm{P}(F) + \mathrm{P}(E|F^{\mathrm{C}})\,\mathrm{P}(F^{\mathrm{C}})$$

$$\mathrm{P}(E) = \sum_{i=1}^{n} \mathrm{P}(E \text{ and } B_i) = \sum_{i=1}^{n} \mathrm{P}(E|B_i)\,\mathrm{P}(B_i)$$

Bayes' Theorem

$$\mathrm{P}(B|E) = \frac{\mathrm{P}(E|B)\cdot\mathrm{P}(B)}{\mathrm{P}(E)}$$

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

Definition of Conditional Probability

$$\mathrm{P}(E|F) = \frac{\mathrm{P}(E \text{ and } F)}{\mathrm{P}(F)}$$

Axiom 1: $0 \le \mathrm{P}(E) \le 1$

Axiom 2: $\mathrm{P}(S) = 1$

Axiom 3: If E and F are mutually exclusive, then $\mathrm{P}(E \text{ or } F) = \mathrm{P}(E) + \mathrm{P}(F)$

Otherwise, use Inclusion-Exclusion:

$\mathrm{P}(E \text{ or } F) = \mathrm{P}(E) + \mathrm{P}(F) - \mathrm{P}(E \text{ and } F)$

$\mathrm{P}(E^{\mathrm{C}}) = 1 - \mathrm{P}(E)$

De Morgan's Laws

$(\mathrm{A} \text{ or } \mathrm{B})^{\mathrm{C}} = \mathrm{A}^{\mathrm{C}} \text{ and } \mathrm{B}^{\mathrm{C}}$

$(\mathrm{A} \text{ and } \mathrm{B})^{\mathrm{C}} = \mathrm{A}^{\mathrm{C}} \text{ or } \mathrm{B}^{\mathrm{C}}$

Chain Rule

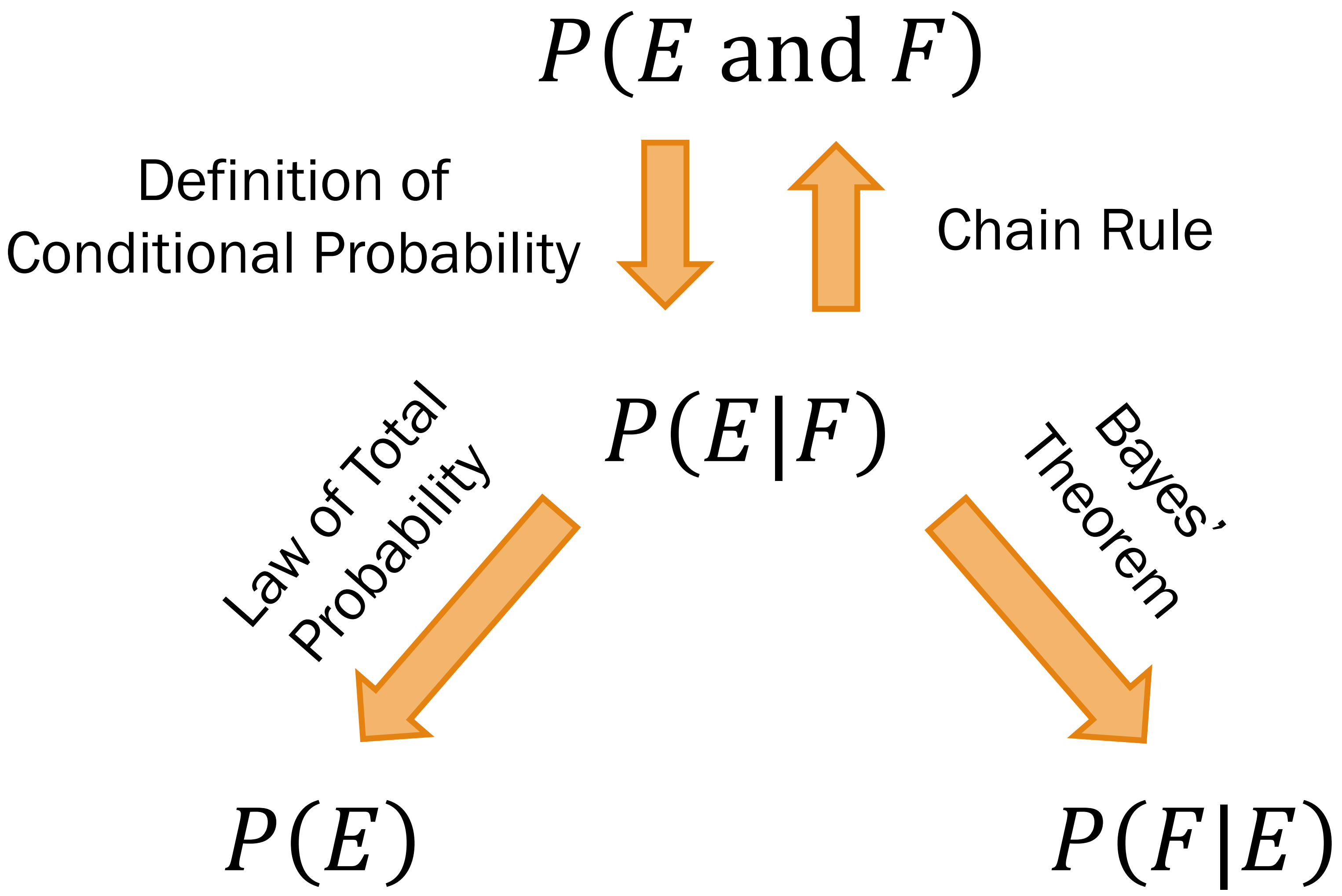
$$\mathrm{P}(E \text{ and } F) = \mathrm{P}(E|F)\cdot\mathrm{P}(F) = \mathrm{P}(F|E)\cdot\mathrm{P}(E)$$

Independence

$$\mathrm{P}(E|F) = \mathrm{P}(E)$$

$$\mathrm{P}(E \text{ and } F) = \mathrm{P}(E)\,\mathrm{P}(F)$$

Learning Goal for Today: Conditional Probability



Learning Goal for Today: Conditional Probability

$$P(E \text{ and } F)$$

Definition of Conditional Probability

Chain Rule

$$P(E|F)$$

Law of Total Probability

Bayes' Theorem

$$P(E)$$

$$P(F|E)$$

Review

The Axioms of Probability

- Axiom 1: $0 \leq P(E) \leq 1$

All probabilities are between 0 and 1

- Axiom 2: $P(S) = 1$

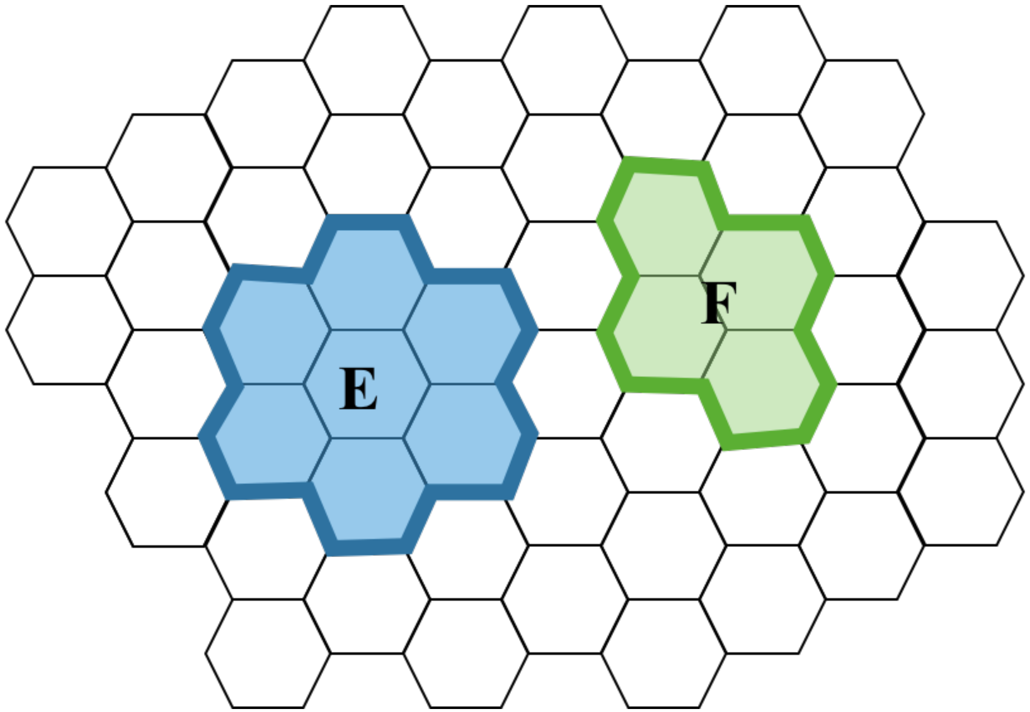
The probability of the sample space is 1

- Axiom 3: If events E and F are mutually exclusive,

$$P(E \cup F) = P(E) + P(F)$$

Probability of event E *or* event F

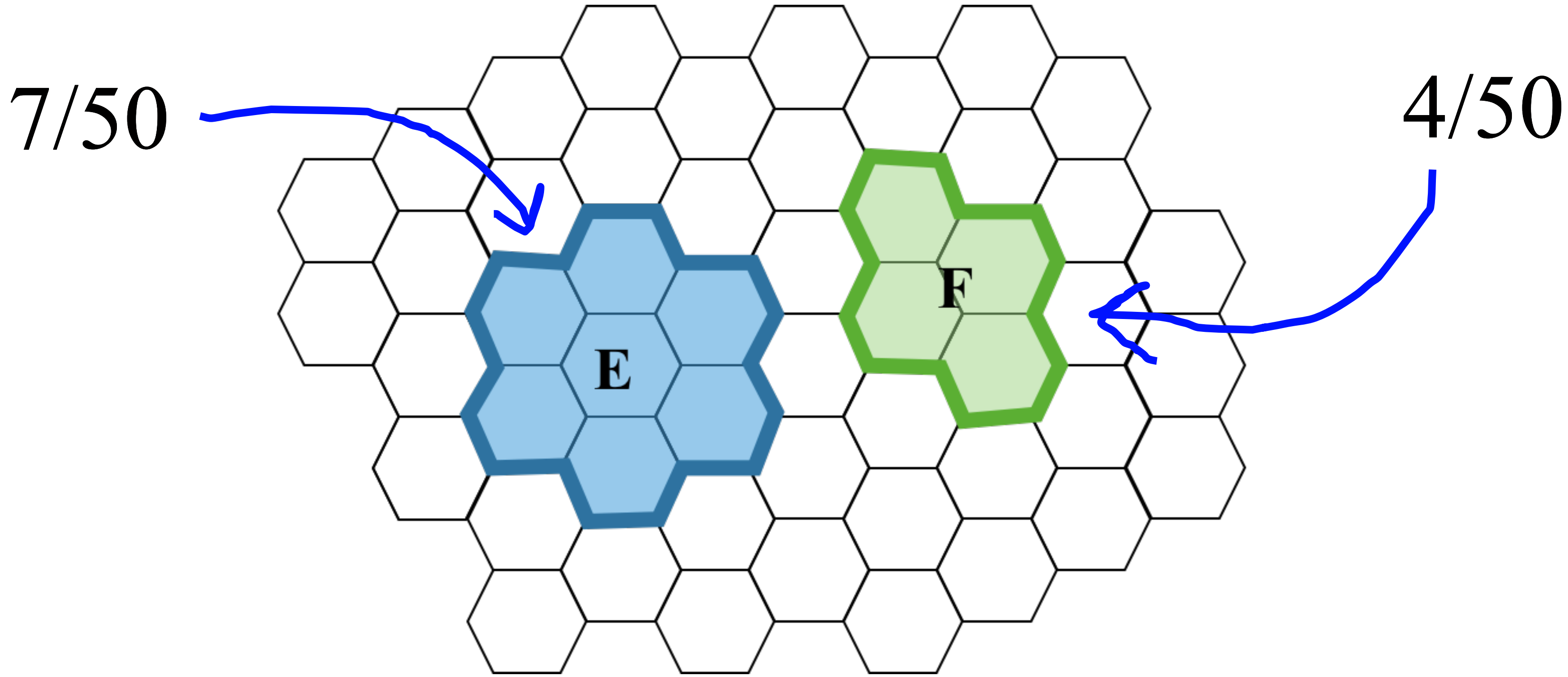
Axiom 3: Probability of Or, With Mutually Exclusive Events



If events have no outcomes in common, probability of OR is simple:

$$P(E \cup F) = P(E) + P(F)$$

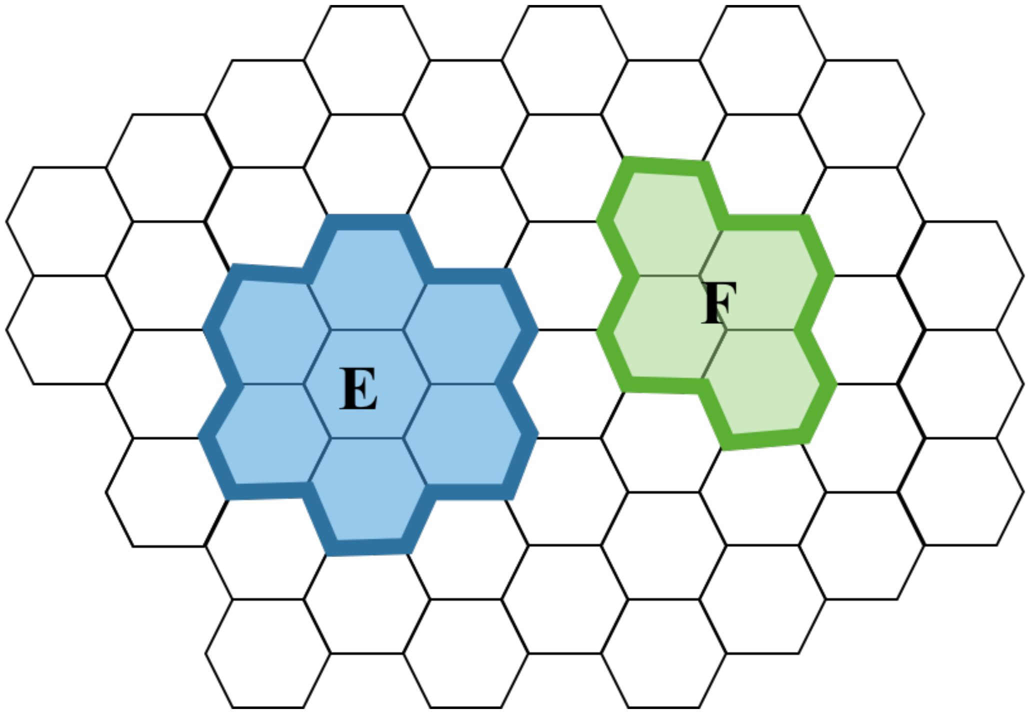
Axiom 3: Probability of Or, With Mutually Exclusive Events



If events have no outcomes in common, probability of OR is simple:

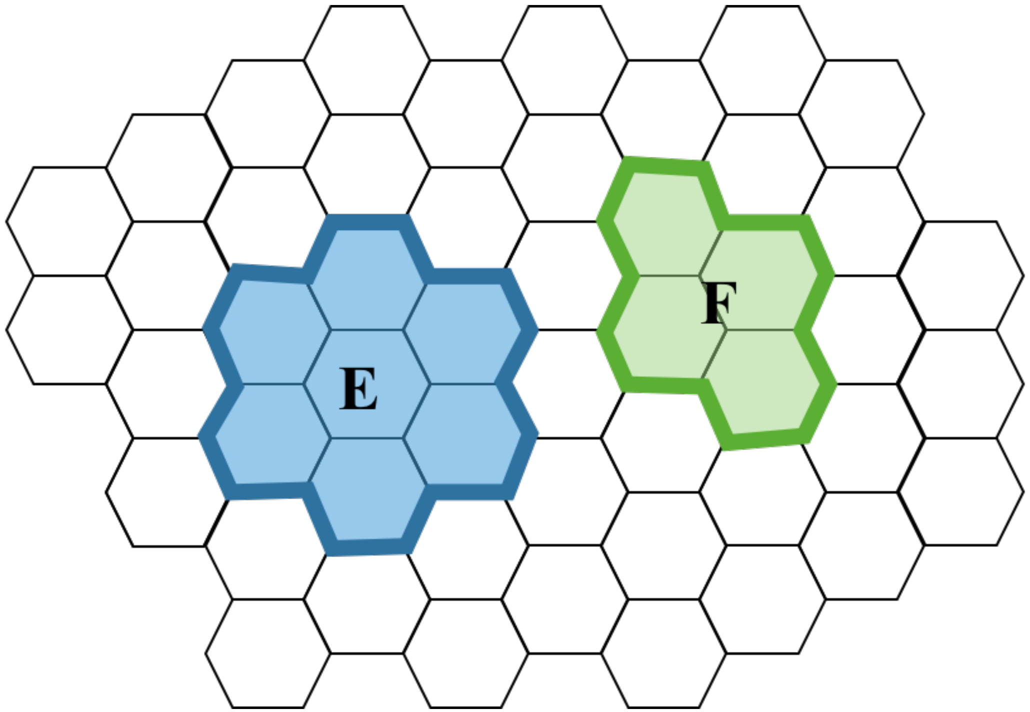
$$P(E \cup F) = \frac{7}{50} + \frac{4}{50} = \frac{11}{50}$$

Axiom 3: Probability of Or, With Mutually Exclusive Events



What is $P(E \cap F)$ in this example?

Axiom 3: Probability of Or, With Mutually Exclusive Events



What is $P(E \cap F)$ in this example?

0! Since E and F are mutually exclusive

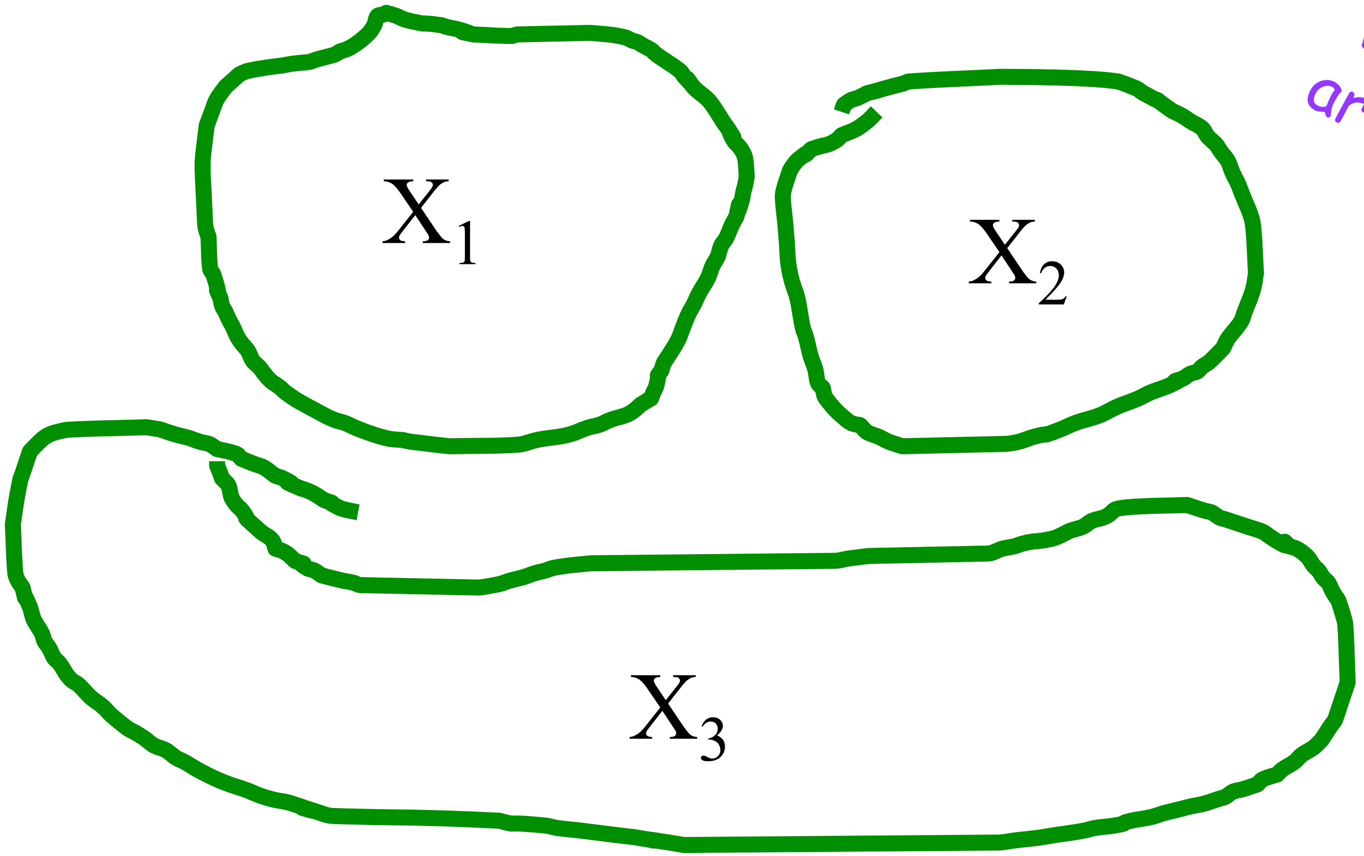
Probability of Or With 3 Or More Events

X_1

X_2

X_3

Probability of Or With 3 Or More Events



Wahoo! All my events are mutually exclusive

$$P(X_1 \cup X_2 \cup \cdots \cup X_n) = \sum_{i=1}^{n} P(X_i)$$

The Axioms of Probability

- Axiom 1: $0 \leq P(E) \leq 1$

All probabilities are between 0 and 1

- Axiom 2: $P(S) = 1$

The probability of the sample space is 1

- Axiom 3: If events E and F are mutually exclusive,

$$P(E \cup F) = P(E) + P(F)$$

Probability of event E *or* event F

- Identity 3*: $P(E^c) = 1 - P(E)$

"not E"

Events either happen…or don't

Where does $P(E^c) = 1 - P(E)$ come from?

Where does $P(E^c) = 1 - P(E)$ come from?

$$P(E \cup E^c) = P(E) + P(E^c)$$

Axiom 3: E and E^c are mutually exclusive

Where does $P(E^c) = 1 - P(E)$ come from?

$$P(E \cup E^c) = P(E) + P(E^c)$$

Axiom 3: E and E^c are mutually exclusive

$$P(S) = P(E) + P(E^c)$$

Since all outcomes are either in E or E^c

Where does $P(E^c) = 1 - P(E)$ come from?

$$P(E \cup E^c) = P(E) + P(E^c)$$

Axiom 3: E and E^c are mutually exclusive

$$P(S) = P(E) + P(E^c)$$

Since all outcomes are either in E or E^c

$$1 = P(E) + P(E^c)$$

Axiom 2: $P(S) = 1$

$$P(E^c) = 1 - P(E)$$

Subtract $P(E)$ from both sides

Chip Defect Detection

Your company has manufactured *n* chips, 1 of which is defective.

k chips are randomly selected from *n* for testing.

What is the probability that the defective chip is in the *k* selected chips?

$$|S| =$$

$$|E| =$$

Chip Defect Detection

Your company has manufactured *n* chips, 1 of which is defective.

k chips are randomly selected from *n* for testing.

What is the probability that the defective chip is in the *k* selected chips?

$$|S| = \binom{n}{k}$$

$$|E| =$$

Chip Defect Detection

Your company has manufactured *n* chips, 1 of which is defective.

k chips are randomly selected from *n* for testing.

What is the probability that the defective chip is in the *k* selected chips?

$$|S| = \binom{n}{k}$$

$$|E| = \binom{1}{1}\binom{n-1}{k-1}$$

Choose the defective chip,
then choose k – 1 other chips

Chip Defect Detection

Your company has manufactured *n* chips, 1 of which is defective.

k chips are randomly selected from *n* for testing.

What is the probability that the defective chip is in the *k* selected chips?

$$|S| = \binom{n}{k}$$

$$|E| = \binom{1}{1}\binom{n-1}{k-1}$$

Choose the defective chip, then choose k – 1 other chips

P(defective chip is in *k* selected chips) $= \dfrac{\binom{1}{1}\binom{n-1}{k-1}}{\binom{n}{k}} = \dfrac{\frac{(n-1)!}{(k-1)!(n-k)!}}{\frac{n!}{k!(n-k)!}} = \dfrac{k}{n}$

Chip Defect Detection

Your company has manufactured n chips, **five** of which are defective.

k chips are randomly selected from n for testing.

What is the probability that **any** defective chips are in the k selected chips?

Chip Defect Detection

Your company has manufactured *n* chips, **five** of which are defective.

k chips are randomly selected from *n* for testing.

What is the probability that **any** defective chips are in the *k* selected chips?

Tip: sometimes, it is easier to calculate $P(E^C)$ than $P(E)$.

Chip Defect Detection

Your company has manufactured *n* chips, **five** of which are defective.

k chips are randomly selected from *n* for testing.

What is the probability that **any** defective chips are in the *k* selected chips?

Let's find *P*(no defective chips chosen) instead:

$$|S| =$$

$$|E| =$$

Chip Defect Detection

Your company has manufactured *n* chips, **five** of which are defective.

k chips are randomly selected from *n* for testing.

What is the probability that **any** defective chips are in the *k* selected chips?

Let's find *P*(no defective chips chosen) instead:

$$|S| = \binom{n}{k}$$

$$|E| = \binom{n-5}{k}$$

Then,

$$P(\text{any defective}) = 1 - P(\text{no defective}) = 1 - \frac{\binom{n-5}{k}}{\binom{n}{k}}$$

End Review

Conditional Probability

Conditional Probability: Why?

So far, we've only learned how to calculate probability in a global, static way.

Day	Outcome
1	Rainy
2	Sunny
3	Rainy
4	Cloudy
5	Rainy
6	Sunny
7	Sunny
8	Sunny
	…
10000	Cloudy

Let E be the event that it is **Sunny** tomorrow. What is $P(E)$?

Conditional Probability: Why?

So far, we've only learned how to calculate probability in a global, static way.

Day	Outcome
1	Rainy
2	Sunny
3	Rainy
4	Cloudy
5	Rainy
6	Sunny
7	Sunny
8	Sunny
	…
10000	Cloudy

Let E be the event that it is **Sunny** tomorrow. What is $P(E)$?

- Weather depends on seasons…shouldn't we only use data from summers to get a probability for tomorrow?

Conditional Probability: Why?

So far, we've only learned how to calculate probability in a global, static way.

Day	Outcome
1	Rainy
2	Sunny
3	Rainy
4	Cloudy
5	Rainy
6	Sunny
7	Sunny
8	Sunny
…	
10000	Cloudy

Let E be the event that it is **Sunny** tomorrow. What is $P(E)$?

- Weather depends on seasons…shouldn't we only use data from summers to get a probability for tomorrow?

- Weather tomorrow depends on weather today…how can we incorporate today's weather into our math?

Conditional Probability: Why?

So far, we've only learned how to calculate probability in a global, static way.

Day	Outcome
1	Rainy
2	Sunny
3	Rainy
4	Cloudy
5	Rainy
6	Sunny
7	Sunny
8	Sunny
	…
10000	Cloudy

Let E be the event that it is **Sunny** tomorrow. What is $P(E)$?

- Weather depends on seasons…shouldn't we only use data from summers to get a probability for tomorrow?
- Weather tomorrow depends on weather today…how can we incorporate today's weather into our math?
- Generally: how do we calculate probabilities of events *under specific conditions*, or *given specific information*?

Conditional Probability: In Words

The **conditional probability** of E given F is the probability that E occurs, *given that F occurs*. This is called "conditioning on F".

Written as: $P(E|F)$

Means: "$P(E, \text{given } F \text{ is observed})$"

Conditional Probability: In Words

The **conditional probability** of E given F is the probability that E occurs, *given that F occurs*. This is called "conditioning on F".

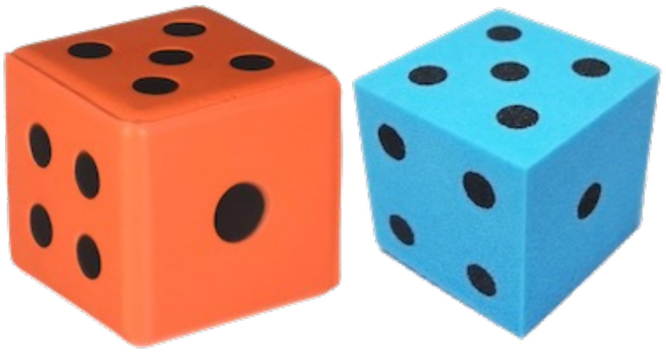
Written as: $P(E|F)$

Means: "$P(E, \text{given } F \text{ is observed})$"

Examples:

- P(sunny | summer) = "Probability it's sunny tomorrow, given that it's summer"
- P(sunny | rain today) = "Probability it's sunny tomorrow, given that it rained today"

Returning to Rolling Two Dice



D_1 D_2

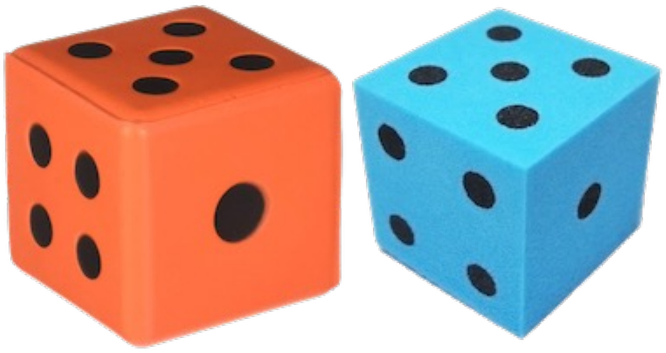
You roll two 6-sided dice.

Let E be the event that the dice sum to 7. What is $P(E)$?

S = {	[1,1]	[1,2]	[1,3]	[1,4]	[1,5]	**[1,6]**
	[2,1]	[2,2]	[2,3]	[2,4]	**[2,5]**	[2,6]
	[3,1]	[3,2]	[3,3]	**[3,4]**	[3,5]	[3,6]
	[4,1]	[4,2]	**[4,3]**	[4,4]	[4,5]	[4,6]
	[5,1]	**[5,2]**	[5,3]	[5,4]	[5,5]	[5,6]
	[6,1]	[6,2]	[6,3]	[6,4]	[6,5]	[6,6] }

$$P(E) = \frac{|E|}{|S|} = \frac{6}{36}$$

Returning to Rolling Two Dice



D_1 D_2

You roll two 6-sided dice.

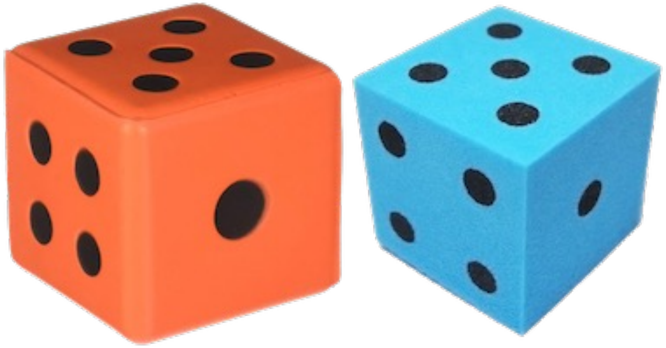
Let E be the event that the dice sum to **4**. What is $P(E)$?

S = {	[1,1]	[1,2]	**[1,3]**	[1,4]	[1,5]	[1,6]
	[2,1]	**[2,2]**	[2,3]	[2,4]	[2,5]	[2,6]
	[3,1]	[3,2]	[3,3]	[3,4]	[3,5]	[3,6]
	[4,1]	[4,2]	[4,3]	[4,4]	[4,5]	[4,6]
	[5,1]	[5,2]	[5,3]	[5,4]	[5,5]	[5,6]
	[6,1]	[6,2]	[6,3]	[6,4]	[6,5]	[6,6] }

$$P(E) = \frac{|E|}{|S|} = \frac{3}{36} = \frac{1}{12}$$

Returning to Rolling Two Dice

You roll two 6-sided dice.



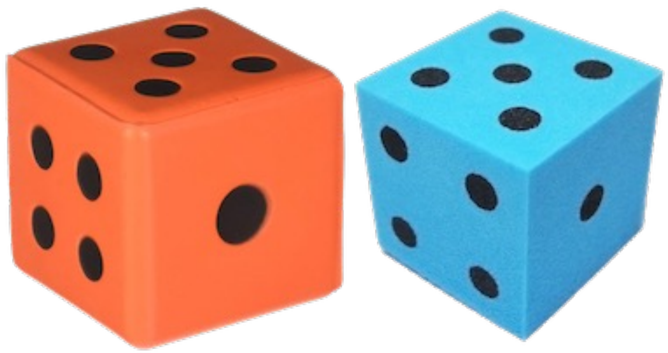
Let E be the event that the dice sum to **4**. What is $P(E)$?

$$
S = \{ \begin{array}{cccccc}
[1,1] & [1,2] & \mathbf{[1,3]} & [1,4] & [1,5] & [1,6] \\
[2,1] & \mathbf{[2,2]} & [2,3] & [2,4] & [2,5] & [2,6] \\
\mathbf{[3,1]} & [3,2] & [3,3] & [3,4] & [3,5] & [3,6] \\
[4,1] & [4,2] & [4,3] & [4,4] & [4,5] & [4,6] \\
[5,1] & [5,2] & [5,3] & [5,4] & [5,5] & [5,6] \\
[6,1] & [6,2] & [6,3] & [6,4] & [6,5] & [6,6]
\end{array} \}
$$

$$P(E) = \frac{|E|}{|S|} = \frac{3}{36} = \frac{1}{12}$$

Note: if we roll a 4, 5, or 6 on the first dice, getting a sum of 4 becomes impossible

Stacking The Odds In Our Favor



D_1 D_2

What outcome for D_1 is most likely to lead to a sum of 4?

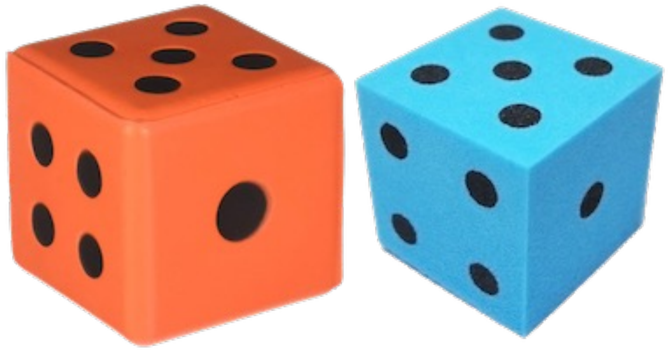
A. 1 and 3 tie for best

B. 1, 2, and 3 tie for best

C. 2 is the best

D. None of the above

Stacking The Odds In Our Favor



D_1 D_2

What outcome for D_1 is most likely to lead to a sum of 4?

A. 1 and 3 tie for best

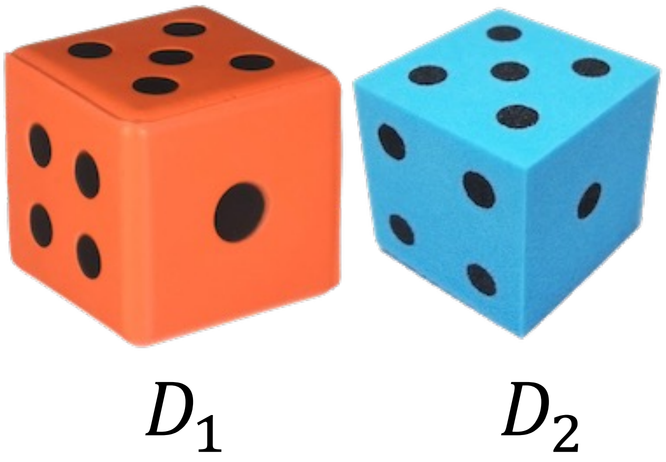
B. 1, 2, and 3 tie for best

C. 2 is the best

D. None of the above

for each, there is one outcome where the sum is 4

Returning to Rolling Two Dice



You roll two 6-sided dice.

Let *F* be the event that $D_1 = 2$.

Let *E* be the event that the dice sum to **4**. What is *P*(*E*, ***given F***)?

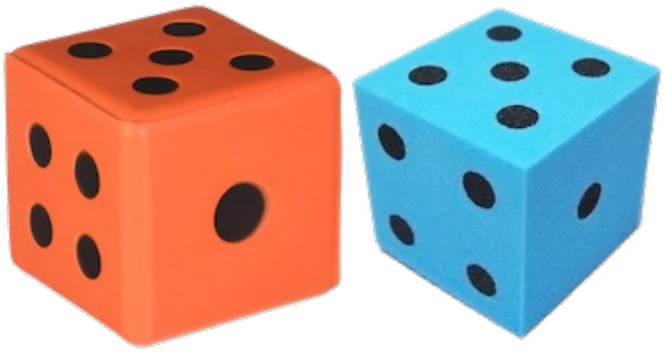
S = { [1,1] [1,2] **[1,3]** [1,4] [1,5] [1,6]
[2,1] **[2,2]** [2,3] [2,4] [2,5] [2,6]
[3,1] [3,2] [3,3] [3,4] [3,5] [3,6]
[4,1] [4,2] [4,3] [4,4] [4,5] [4,6]
[5,1] [5,2] [5,3] [5,4] [5,5] [5,6]
[6,1] [6,2] [6,3] [6,4] [6,5] [6,6] }

Returning to Rolling Two Dice

You roll two 6-sided dice.

Let F be the event that $D_1 = 2$.

Let E be the event that the dice sum to **4**. What is $P(E$, ***given F***$)$?

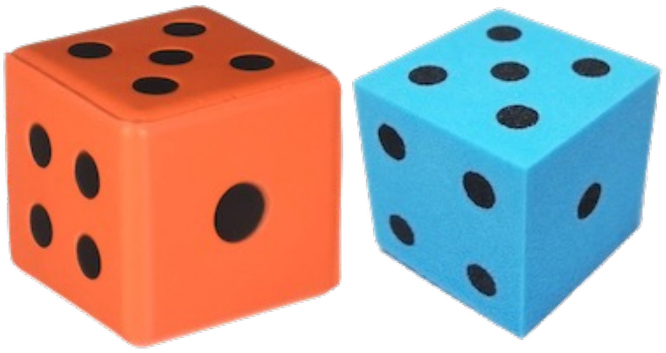


D_1 D_2

S = { [1,1] [1,2] [1,3] [1,4] [1,5] [1,6]
[2,1] [2,2] [2,3] [2,4] [2,5] [2,6]
[3,1] [3,2] [3,3] [3,4] [3,5] [3,6]
[4,1] [4,2] [4,3] [4,4] [4,5] [4,6]
[5,1] [5,2] [5,3] [5,4] [5,5] [5,6]
[6,1] [6,2] [6,3] [6,4] [6,5] [6,6] }

These are the only possible outcomes now…

Returning to Rolling Two Dice



D_1 D_2

You roll two 6-sided dice.

Let F be the event that $\boldsymbol{D_1 = 2}$.

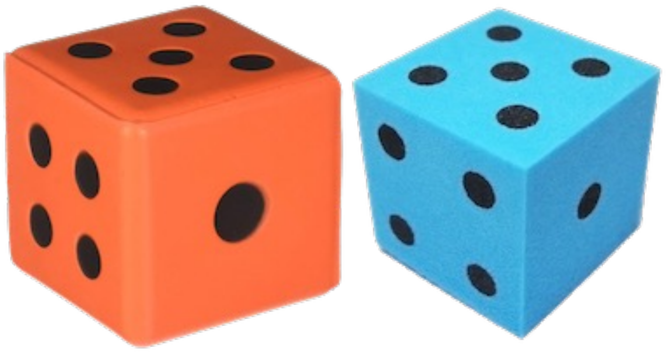
Let E be the event that the dice sum to **4**. What is P(E, ***given F***)?

S = { [2,1] **[2,2]** [2,3] [2,4] [2,5] [2,6] }

These are the only possible outcomes now….

…so our sample space shrinks!

Returning to Rolling Two Dice



You roll two 6-sided dice.

Let F be the event that $\boldsymbol{D_1 = 2}$.

Let E be the event that the dice sum to **4**. What is P(E, ***given F***)?

S = { [2,1]   **[2,2]**   [2,3]   [2,4]   [2,5]   [2,6] }

These are the only possible outcomes now….

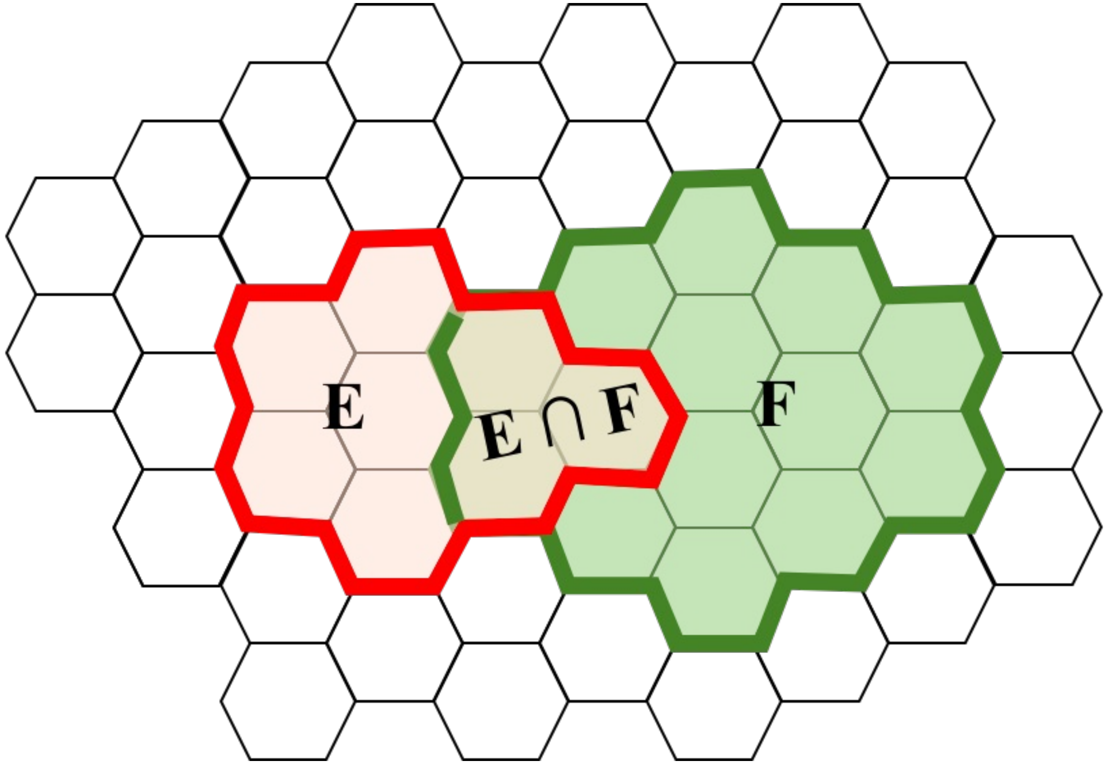
…so our sample space shrinks!

$$P(E|F) = \frac{|E| \text{ (given F)}}{|S| \text{ (given F)}} = \frac{|E \text{ and } F|}{|F|} = \frac{1}{6}$$

"E given F"

Conditional Probability: Visual Intuition

The **conditional probability** of E given F is the probability that E occurs, *given that F occurs*. This is called "conditioning on F".

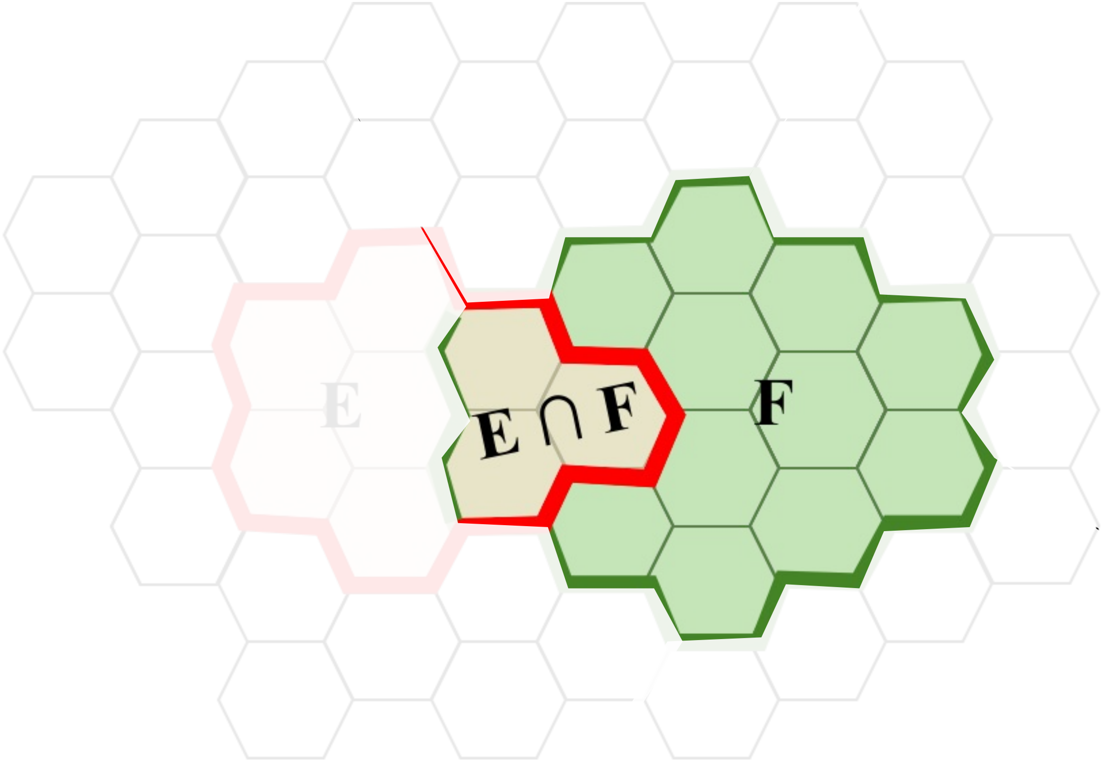


$$P(E) = \frac{8}{50} \approx 0.16$$

$$P(E|F) = \ ?$$

Conditional Probability: Visual Intuition

The **conditional probability** of E given F is the probability that E occurs, *given that F occurs*. This is called "conditioning on F".

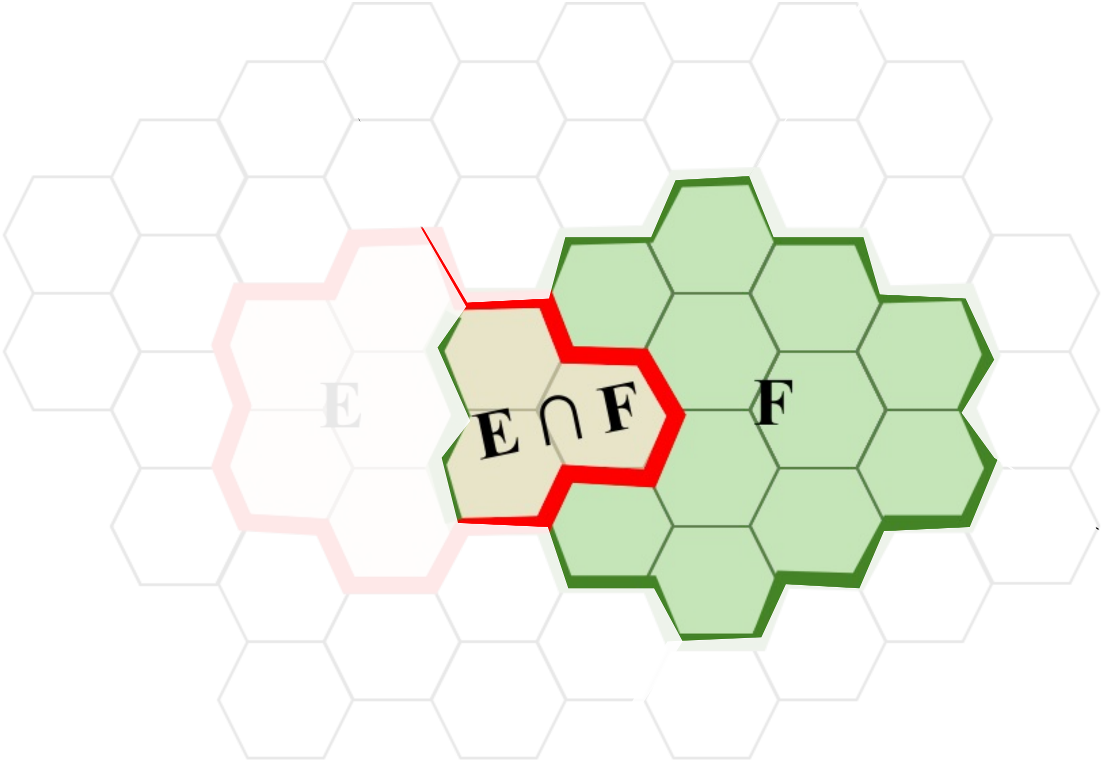


$$P(E) = \frac{8}{50} \approx 0.16$$

$$P(E|F) = \ ?$$

Conditional Probability: Visual Intuition

The **conditional probability** of E given F is the probability that E occurs, *given that F occurs*. This is called "conditioning on F".



$$P(E) = \frac{8}{50} \approx 0.16$$

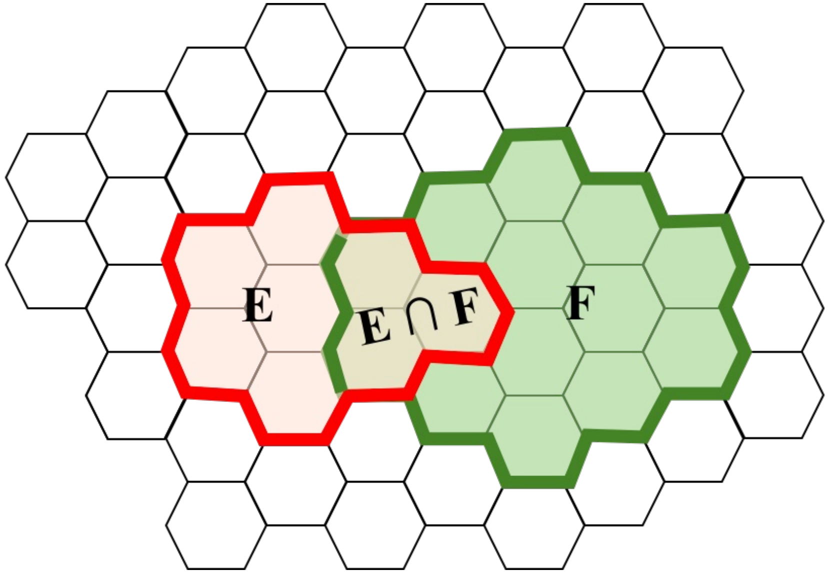
$$P(E|F) = \frac{3}{14} \approx 0.21$$

Conditional Probability: Visual Intuition

The **conditional probability** of E given F is the probability that E occurs, *given that F occurs*. This is called "conditioning on F".

$$\Pr(E|F) = \frac{\#\text{ of outcomes in } E \text{ consistent with } F}{\#\text{ of outcomes in } S \text{ consistent with } F} = \frac{|EF|}{|SF|} = \frac{|EF|}{|F|}$$

Shorthand notation for set intersection (aka set "and")



Definition of Conditional Probability

The **conditional probability** of E given F is the probability that E occurs, *given that F occurs*. This is called “conditioning on F”.

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

- This definition works even when outcomes are not equally likely

Definition of Conditional Probability

The **conditional probability** of E given F is the probability that E occurs, *given that F occurs*. This is called "conditioning on F".

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

- This definition works even when outcomes are not equally likely
- What if $P(F) = 0$??

Definition of Conditional Probability

The **conditional probability** of E given F is the probability that E occurs, *given that F occurs*. This is called "conditioning on F".

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

- This definition works even when outcomes are not equally likely
- What if *P*(*F*) = 0?? Undefined: *you observed the impossible!*

Definition of Conditional Probability

The **conditional probability** of E given F is the probability that E occurs, *given that F occurs*. This is called "conditioning on F".

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

- This definition works even when outcomes are not equally likely
- What if $P(F) = 0$?? Undefined: *you observed the impossible!*

We can multiply both sides by $P(F)$ to get the **chain rule**:

$$P(E \cap F) = P(E|F)P(F)$$


goodbye, land of equally likely outcomes

Practice: Multiple Choice & Probability

Imagine a multiple choice test, where every question has 4 answer choices.

- Students can get questions correct by knowing the answer, or just by guessing.
- How can you know if a student really knows their stuff, or if they're just guessing?

Practice: Multiple Choice & Probability

Imagine a multiple choice test, where every question has 4 answer choices.

- Students can get questions correct by knowing the answer, or just by guessing.
- How can you know if a student really knows their stuff, or if they're just guessing?

Let G be the event that a student guesses an answer to a question. $P(G) = 1/5$.

Let R be the event that the student gets the answer right.

1. What is the probability that a student gets a question right, given that they guess?

Practice: Multiple Choice & Probability

Imagine a multiple choice test, where every question has 4 answer choices.

- Students can get questions correct by knowing the answer, or just by guessing.
- How can you know if a student really knows their stuff, or if they're just guessing?

Let *G* be the event that a student guesses an answer to a question. *P*(*G*) = 1/5.

Let *R* be the event that the student gets the answer right.

$$P(R|G) = \frac{1}{4}$$

1. What is the probability that a student gets a question right, given that they guess?

Practice: Multiple Choice & Probability

Imagine a multiple choice test, where every question has 4 answer choices.

- Students can get questions correct by knowing the answer, or just by guessing.
- How can you know if a student really knows their stuff, or if they're just guessing?

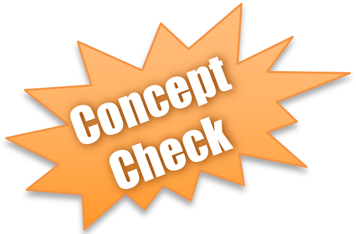
Let G be the event that a student guesses an answer to a question. $P(G)$ = 1/5.

Let R be the event that the student gets the answer right.

$$P(R|G) = \frac{1}{4}$$

1. What is the probability that a student gets a question right, given that they guess?

2. What is the probability that a student guesses on a question and gets it right?



Practice: Multiple Choice & Probability

Imagine a multiple choice test, where every question has 4 answer choices.

- Students can get questions correct by knowing the answer, or just by guessing.
- How can you know if a student really knows their stuff, or if they're just guessing?

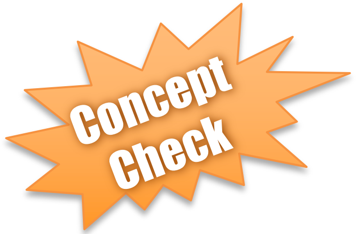
Let *G* be the event that a student guesses an answer to a question. *P*(*G*) = 1/5.

Let *R* be the event that the student gets the answer right.

$$P(R|G) = \frac{1}{4}$$

1. What is the probability that a student gets a question right, given that they guess?

2. What is the probability that a student guesses on a question and gets it right?



$$P(G \cap R) = P(R|G)P(G) = \frac{1}{4} \cdot \frac{1}{5} = \frac{1}{20}$$

Common Confusion With Conditional Probability

$$P(E|F) \neq P(E \cap F)$$

What is the difference between “E, given F”, vs. “E and F”?

Common Confusion With Conditional Probability

$$P(E|F) \neq P(E \cap F)$$

What is the difference between “E, given F”, vs. “E and F”?

- Imagine: E is the event that it is hot outside.

Common Confusion With Conditional Probability

$$P(E|F) \neq P(E \cap F)$$

What is the difference between “E, given F”, vs. “E and F”?

- Imagine: *E* is the event that it is hot outside.
- *F* is the event that the sun explodes into a massive supernova that engulfs the Earth in flames!!!
- If the Earth is engulfed in flames, it’s definitely hot, so $P(E|F)$ is near 1.
- But because *F* is so unlikely, $P(F)$ is almost zero, and $P(E \text{ and } F)$ is near zero too.

Common Confusion With Conditional Probability

$$P(E|F) \neq P(F|E)$$

What is the difference between “E, given F”, vs. “F, given E”?

Common Confusion With Conditional Probability

$$P(E|F) \neq P(F|E)$$

What is the difference between "E, given F", vs. "F, given E"?

- If the Earth is engulfed in flames, it's definitely hot, so $P(E|F)$ is near 1.
- If it's hot outside, do we usually believe that a supernova must have happened? No! There are lots of other reasons it could be hot outside. So $P(F \mid E)$ is near zero too.

Tip: imagining extremes can help you build intuition!

NETFLIX

and Learn

Netflix: Watch Probabilities

What is the probability that a user will watch The Great British Bake-off?

Netflix: Watch Probabilities

What is the probability that a user will watch The Great British Bake-off?

$$P(E) = \lim_{n \to \infty} \frac{n(E)}{n} \approx \frac{\#\text{people who watched movie}}{\#\text{people on Netflix}}$$

$$P(E) \approx 50{,}234{,}231 \,/\, 269{,}613{,}547 \approx 0.18$$

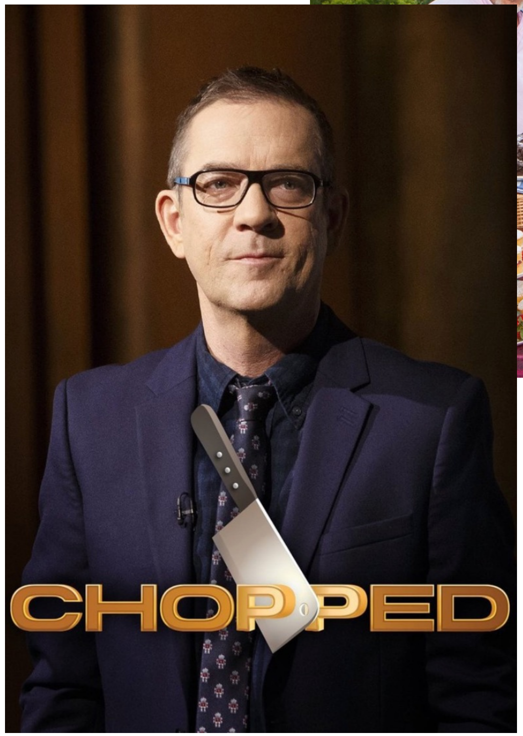
Netflix & Conditional Probability

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$ Definition of Cond. Probability

Let E be the event that a user watches The Great British Bake-off.

Let F be the event that a user watches Chopped.

What is the probability that a user watches
The Great British Bake-off, given they watched Chopped?



Netflix & Conditional Probability

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$ Definition of Cond. Probability

Let E be the event that a user watches The Great British Bake-off.

Let F be the event that a user watches Chopped.

What is the probability that a user watches
The Great British Bake-off, given they watched Chopped?

$$P(E|F) = \frac{P(E \cap F)}{P(F)} \approx \frac{\frac{\text{\# people who have watched both}}{\text{\# people on Netflix}}}{\frac{\text{\# people who have watched Chopped}}{\text{\# people on Netflix}}}$$

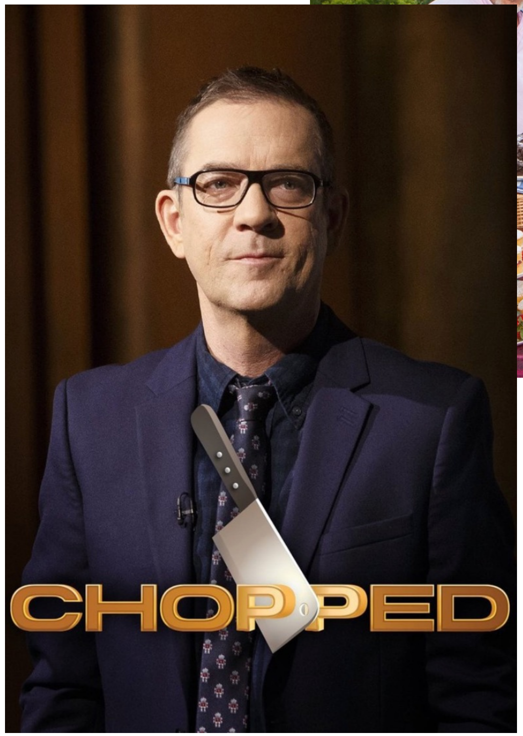
Netflix & Conditional Probability

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$ Definition of Cond. Probability

Let E be the event that a user watches The Great British Bake-off.

Let F be the event that a user watches Chopped.

What is the probability that a user watches
The Great British Bake-off, given they watched Chopped?



$$P(E|F) = \frac{P(E \cap F)}{P(F)} \approx \frac{\dfrac{\text{\# people who have watched both}}{\text{\# people on Netflix}}}{\dfrac{\text{\# people who have watched Chopped}}{\text{\# people on Netflix}}}$$

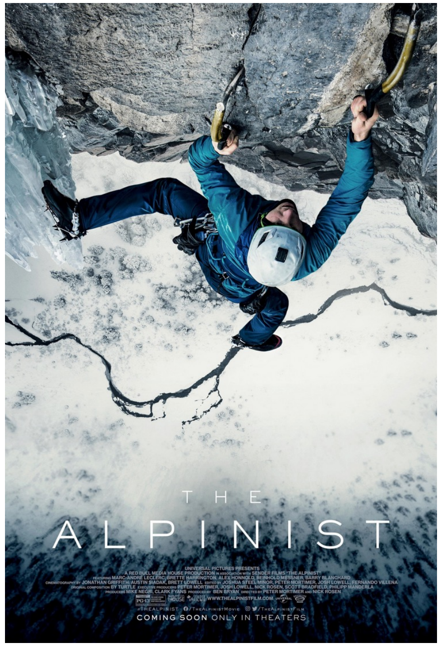
$$\approx \frac{\text{\# people who have watched both}}{\text{\# people who have watched Chopped}} \approx 0.42$$

Netflix & Conditional Probability

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$ Definition of Cond. Probability

Let E be the event that a user watches The Great British Bake-off.

Let F be the event that a user watches each of the shows/movies below.



Netflix & Conditional Probability

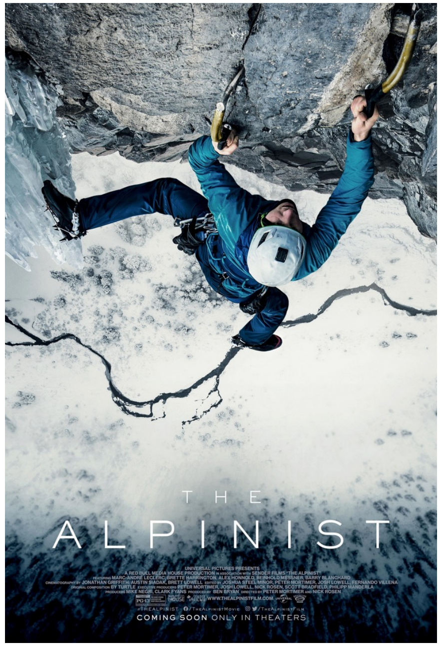
$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$ Definition of Cond. Probability

Let E be the event that a user watches The Great British Bake-off.

Let F be the event that a user watches each of the shows/movies below.

$P(E|F) \approx 0.31$

$P(E|F) \approx 0.20$



$P(E|F) \approx 0.17$

$P(E|F) \approx 0.09$

What does it mean if $P(E \mid F) > P(E)$? How could Netflix use this for recommendations?

A Guide To Notation

And	Or	Given
$P(E \text{ and } F)$	$P(E \text{ or } F)$	$P(E \mid F)$
$P(E \cap F)$	$P(E \cup F)$	$P(E \mid F, G)$
$P(E, F)$		Probability of E given F and G
$P(EF)$		

Generalized Chain Rule

For just two events, the **chain rule** looks like this:

$$P(E \cap F) = P(E|F)P(F)$$

For more than two events:

$$\begin{aligned}\Pr(E_1 \text{ and } E_2 \text{ and } E_3 \text{ and } \ldots E_n) \\ = \Pr(E_1) \cdot \Pr(E_2|E_1) \cdot \Pr(E_3|E_1, E_2) \cdots \Pr(E_n|E_1, E_2 \ldots E_{n-1})\end{aligned}$$

But often, we can simplify this equation a lot…we'll see how next lecture!

Law of Total Probability

Learning Goal for Today: Conditional Probability

$$P(E \text{ and } F)$$

Definition of Conditional Probability

Chain Rule

$$P(E|F)$$

Law of Total Probability

Bayes' Theorem

$$P(E)$$

$$P(F|E)$$

Learning Goal for Today: Conditional Probability

$P(E \text{ and } F)$

Definition of Conditional Probability

Chain Rule

$P(E|F)$

Law of Total Probability

Bayes' Theorem

$P(E)$

$P(F|E)$

Practice: Multiple Choice & Probability

Imagine a multiple choice test, where every question has 4 answer choices.

Let G be the event that a student guesses an answer to a question. $P(G) = 1/5$.

Let R be the event that the student gets the answer right.

Let the probability a student gets the answer right without guessing be $P(R \mid G^C) = 9/10$.

What is the probability that a student gets an answer right?

Law of Total Probability

Let E and F be our two events. The **Law of Total Probability** is:

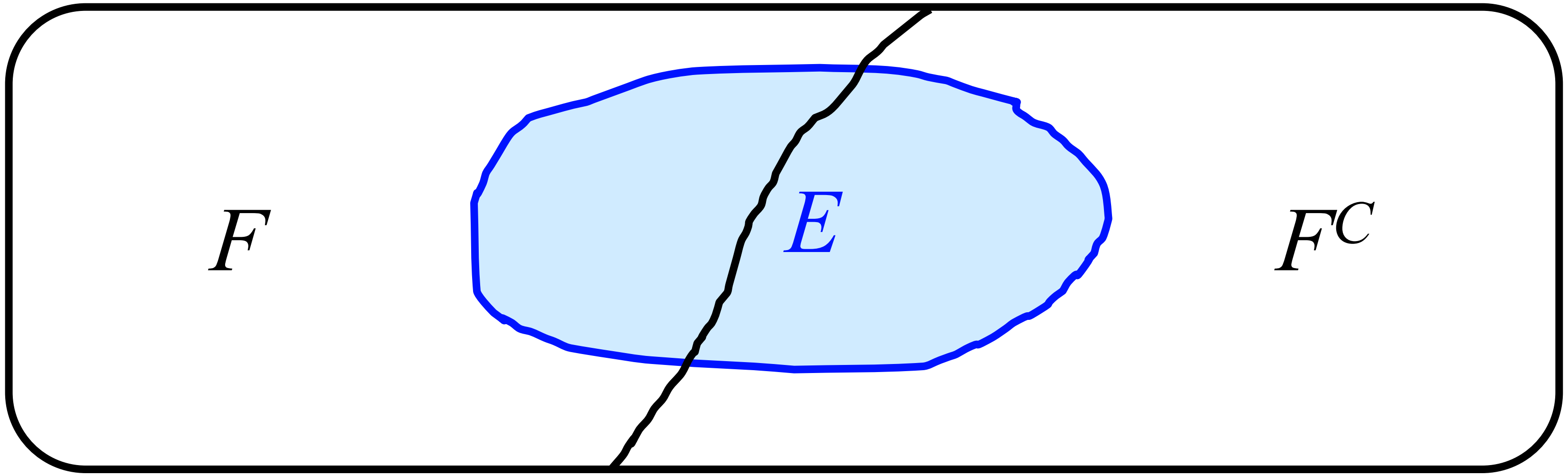
$$P(E) = P(EF) + P(EF^C)$$
$$= P(E|F)P(F) + P(E|F^C)P(F^C)$$

Law of Total Probability

Let E and F be our two events. The **Law of Total Probability** is:

$$P(E) = P(EF) + P(EF^C)$$
$$= P(E|F)P(F) + P(E|F^C)P(F^C)$$

Sample Space

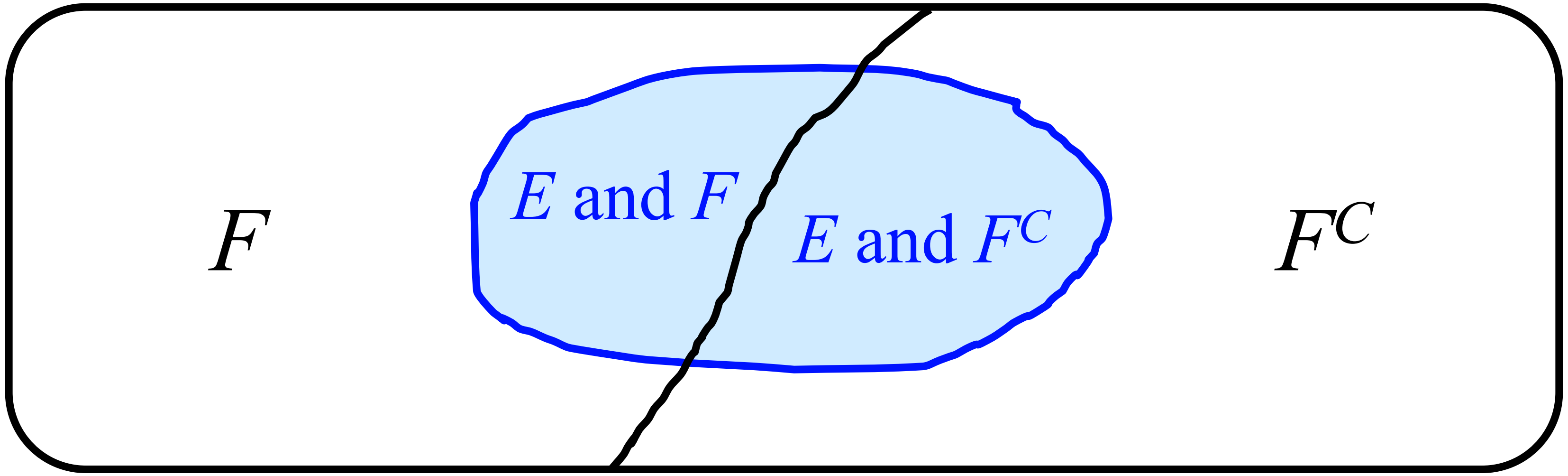


Law of Total Probability

Let E and F be our two events. The **Law of Total Probability** is:

$$P(E) = P(EF) + P(EF^C)$$
$$= P(E|F)P(F) + P(E|F^C)P(F^C)$$

Sample Space

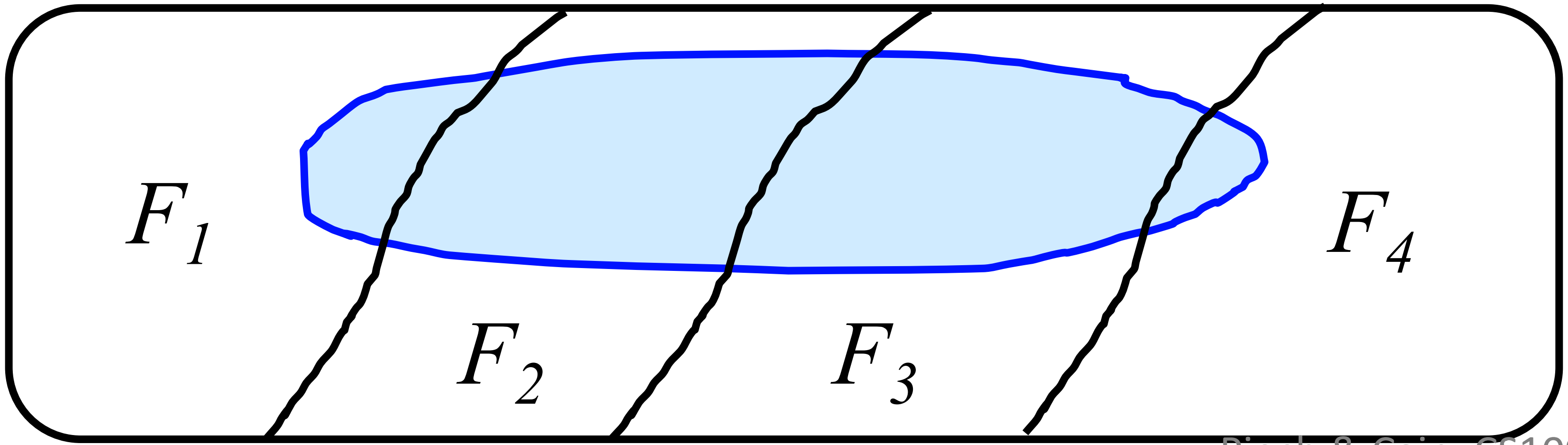


Generalized Law of Total Probability

Let $F_1, F_2, \ldots, F_n$ be mutually exclusive events that cover the sample space.

$$P(E) = \sum_{i=1}^{n} P(E \cap F_i)$$
$$= \sum_{i=1}^{n} P(E|F_i)P(F_i)$$

Sample Space



Practice: Multiple Choice & Probability

Imagine a multiple choice test, where every question has 4 answer choices.

Let G be the event that a student guesses an answer to a question. $P(G) = 1/5$.

Let R be the event that the student gets the answer right.

Let the probability a student gets the answer right without guessing be $P(R \mid G^C) = 9/10$.

What is the probability that a student gets an answer right?

Practice: Multiple Choice & Probability

Imagine a multiple choice test, where every question has 4 answer choices.

Let G be the event that a student guesses an answer to a question. $P(G) = 1/5$.

Let R be the event that the student gets the answer right.

Let the probability a student gets the answer right without guessing be $P(R \mid G^C) = 9/10$.

What is the probability that a student gets an answer right?

$$P(R) = P(R|G)P(G) + P(R|G^C)P(G^C)$$

Practice: Multiple Choice & Probability

Imagine a multiple choice test, where every question has 4 answer choices.

Let G be the event that a student guesses an answer to a question. $P(G) = 1/5$.

Let R be the event that the student gets the answer right.

Let the probability a student gets the answer right without guessing be $P(R \mid G^C) = 9/10$.

What is the probability that a student gets an answer right?

$$P(R) = P(R|G)P(G) + P(R|G^C)P(G^C)$$
$$= \frac{1}{4} \cdot \frac{1}{5} + \frac{9}{10}(1 - \frac{1}{5}) = 0.77$$

Evolution of Bacteria

$P(E) = P(E|F)P(F) + P(E|F^C)P(F^C)$ Law of Total Probability

You have bacteria in your gut which is causing a disease.
10% of bacteria have a mutation making them resistant to antibiotics.
You take half a course of antibiotics…

Probability a bacteria survives, given it has the mutation: 20%
Probability a bacteria survives, given it doesn't have the mutation: 1%
What is the probability that a randomly chosen bacteria survives?

Tip: start problems like this by defining events.

Evolution of Bacteria

$P(E) = P(E|F)P(F) + P(E|F^C)P(F^C)$ Law of Total Probability

You have bacteria in your gut which is causing a disease.
10% of bacteria have a mutation making them resistant to antibiotics.
You take half a course of antibiotics…

Probability a bacteria survives, given it has the mutation: 20%
Probability a bacteria survives, given it doesn't have the mutation: 1%
What is the probability that a randomly chosen bacteria survives?

Let E be the event that a bacteria survives.
Let M be the event that a bacteria has the mutation.

Evolution of Bacteria

$$P(E) = P(E|F)P(F) + P(E|F^C)P(F^C)$$ Law of Total Probability

You have bacteria in your gut which is causing a disease.
10% of bacteria have a mutation making them resistant to antibiotics.
You take half a course of antibiotics…

Probability a bacteria survives, given it has the mutation: 20%
Probability a bacteria survives, given it doesn't have the mutation: 1%
What is the probability that a randomly chosen bacteria survives?

Let E be the event that a bacteria survives.
Let M be the event that a bacteria has the mutation.

$$\begin{aligned}\Pr(E) &= \Pr(E|M)\Pr(M) + \Pr(E|M^C)\Pr(M^C) \\ &= 0.20 \cdot 0.10 + 0.01 \cdot 0.90 \\ &= 0.029\end{aligned}$$

Real question. What is the probability that a surviving bacteria has the mutation?

$$P(\text{ Mutation } | \text{ Survives })$$

$$P(M \mid E)$$

Real Question: P(M | E)?

You have bacteria in your gut which is causing a disease.
10% of bacteria have a mutation making them resistant to antibiotics.
You take half a course of antibiotics…

$P(E \mid M) = 0.20$
$P(E \mid M^C) = 0.01$
What is the probability that a surviving bacteria **has the mutation**?

Learning Goal for Today: Conditional Probability

$$P(E \text{ and } F)$$

Definition of Conditional Probability

Chain Rule

$$P(E|F)$$

Law of Total Probability

Bayes' Theorem

$$P(E)$$

$$P(F|E)$$

Learning Goal for Today: Conditional Probability

$P(E \text{ and } F)$

Definition of Conditional Probability

Chain Rule

$P(E|F)$

Law of Total Probability

Bayes' Theorem

$P(E)$

$P(F|E)$

It’s Time For Bayes’ Theorem

Thomas Bayes

Rev. Thomas Bayes (~1701-1761):
British mathematician and Presbyterian minister

Thomas Bayes

Rev. Thomas Bayes (~1701-1761):
British mathematician and Presbyterian minister

He looked remarkably similar to Sean Astin

Thomas Bayes

$P(E|F) \Rightarrow P(F|E)$

$P(F \mid E)$

I want to calculate
P(State of the world F | Observation E)
It seems so tricky!…

The other way around is easy
P(Observation E | State of the world F)
What options to I have, chief?

$P(E \mid F)$

Thomas Bayes

$P(E|F) \Rightarrow P(F|E)$

$$P(F|E) = \frac{P(EF)}{P(E)}$$

Definition of Conditional Probability

Thomas Bayes

$P(E|F) \Rightarrow P(F|E)$

$$P(F|E) = \frac{P(EF)}{P(E)}$$

Definition of Conditional Probability

A little while later…

$$= \frac{P(E|F)P(F)}{P(E)}$$

The Chain Rule is symmetric

Thomas Bayes

$P(E|F) \Rightarrow P(F|E)$

$$P(F|E) = \frac{P(EF)}{P(E)}$$

Definition of Conditional Probability

A little while later…

$$= \frac{P(E|F)P(F)}{P(E)}$$

The Chain Rule is symmetric

A little while later…

$$= \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^C)P(F^C)}$$

Law of Total Probability

Bayes' Theorem

$P(E|F) \Rightarrow P(F|E)$

For any events E and F where $P(E) > 0$ and $P(F) > 0$,

$$P(F|E) = \frac{P(E|F)P(F)}{P(E)}$$

Expanded form:

$$P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^C)P(F^C)}$$

Bayes' Theorem

$P(E|F) \Rightarrow P(F|E)$

For any events E and F where $P(E) > 0$ and $P(F) > 0$,

$$\underbrace{P(F|E)}_{\text{posterior}} = \frac{\overbrace{P(E|F)}^{\text{likelihood}}\overbrace{P(F)}^{\text{prior}}}{\underbrace{P(E)}_{\text{normalization constant}}}$$

Expanded form:

$$P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^C)P(F^C)}$$

Real Question: P(M | E)?

You have bacteria in your gut which is causing a disease.
10% of bacteria have a mutation making them resistant to antibiotics.
You take half a course of antibiotics…

$P(E \mid M) = 0.20$
$P(E \mid M^C) = 0.01$
What is the probability that a surviving bacteria **has the mutation**?

Real Question: P(M | E)?

You have bacteria in your gut which is causing a disease.
10% of bacteria have a mutation making them resistant to antibiotics.
You take half a course of antibiotics…

$P(E \mid M) = 0.20$
$P(E \mid M^C) = 0.01$
What is the probability that a surviving bacteria **has the mutation**?

$$P(M \mid E) = \frac{P(E \mid M)\, P(M)}{\mathrm{P}(E \mid M)\, \mathrm{P}(M) + \mathrm{P}(E \mid M^c)\, \mathrm{P}(M^c)}$$

Real Question: P(M | E)?

You have bacteria in your gut which is causing a disease.
10% of bacteria have a mutation making them resistant to antibiotics.
You take half a course of antibiotics…

$P(E \mid M) = 0.20$
$P(E \mid M^C) = 0.01$
What is the probability that a surviving bacteria **has the mutation**?

$$P(M \mid E) = \frac{P(E \mid M)\, P(M)}{\mathrm{P}(E \mid M)\, \mathrm{P}(M) + \mathrm{P}(E \mid M^c)\, \mathrm{P}(M^c)}$$

$$P(M \mid E) = \frac{0.2 * 0.1}{0.2 * 0.1 + 0.01 * 0.9} \approx 0.69$$

Zika Virus Testing

$$P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^C)P(F^C)}$$

A test is 98% effective at detecting Zika.

- However, the test has a "false positive" rate of 1%.
- 0.5% of US population has Zika.
- Let E = you test positive for Zika with this test.
- Let F = you actually have Zika.
- What is P(F | E)?

Zika Virus Testing

$$P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^C)P(F^C)}$$

A test is 98% effective at detecting Zika.

- However, the test has a “false positive” rate of 1%.
- 0.5% of US population has Zika.
- Let E = you test positive for Zika with this test.
- Let F = you actually have Zika.
- What is P(F | E)?

$$P(F \mid E) = \frac{P(E \mid F)\, P(F)}{\mathrm{P}(E \mid F)\, \mathrm{P}(F) + \mathrm{P}(E \mid F^c)\, \mathrm{P}(F^c)}$$

Zika Virus Testing

$$P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^C)P(F^C)}$$

A test is 98% effective at detecting Zika.

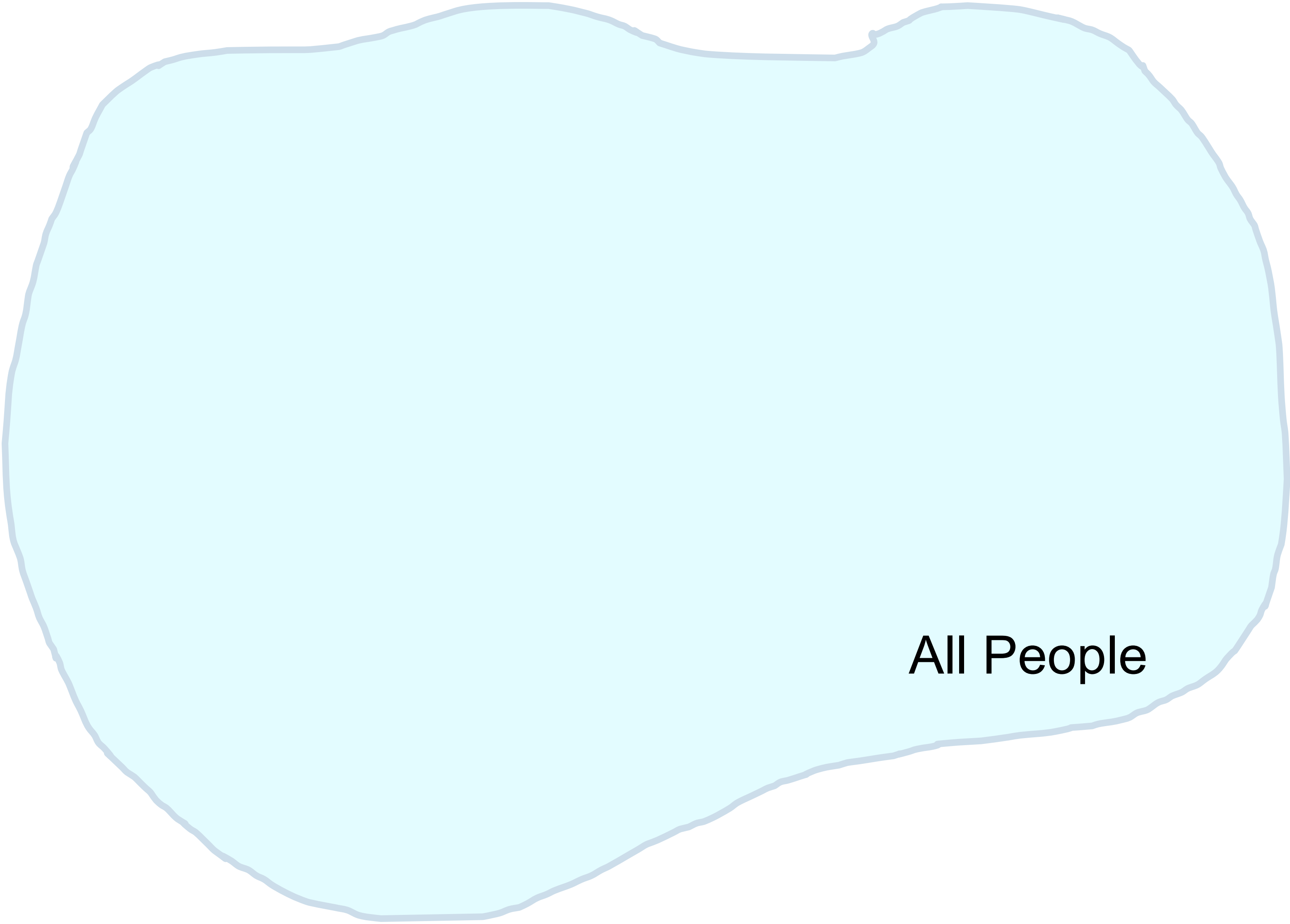
- However, the test has a "false positive" rate of 1%.
- 0.5% of US population has Zika.
- Let E = you test positive for Zika with this test.
- Let F = you actually have Zika.
- What is P(F | E)?

$$P(F \mid E) = \frac{P(E \mid F)\, P(F)}{\mathrm{P}(E \mid F)\, \mathrm{P}(F) + \mathrm{P}(E \mid F^c)\, \mathrm{P}(F^c)}$$

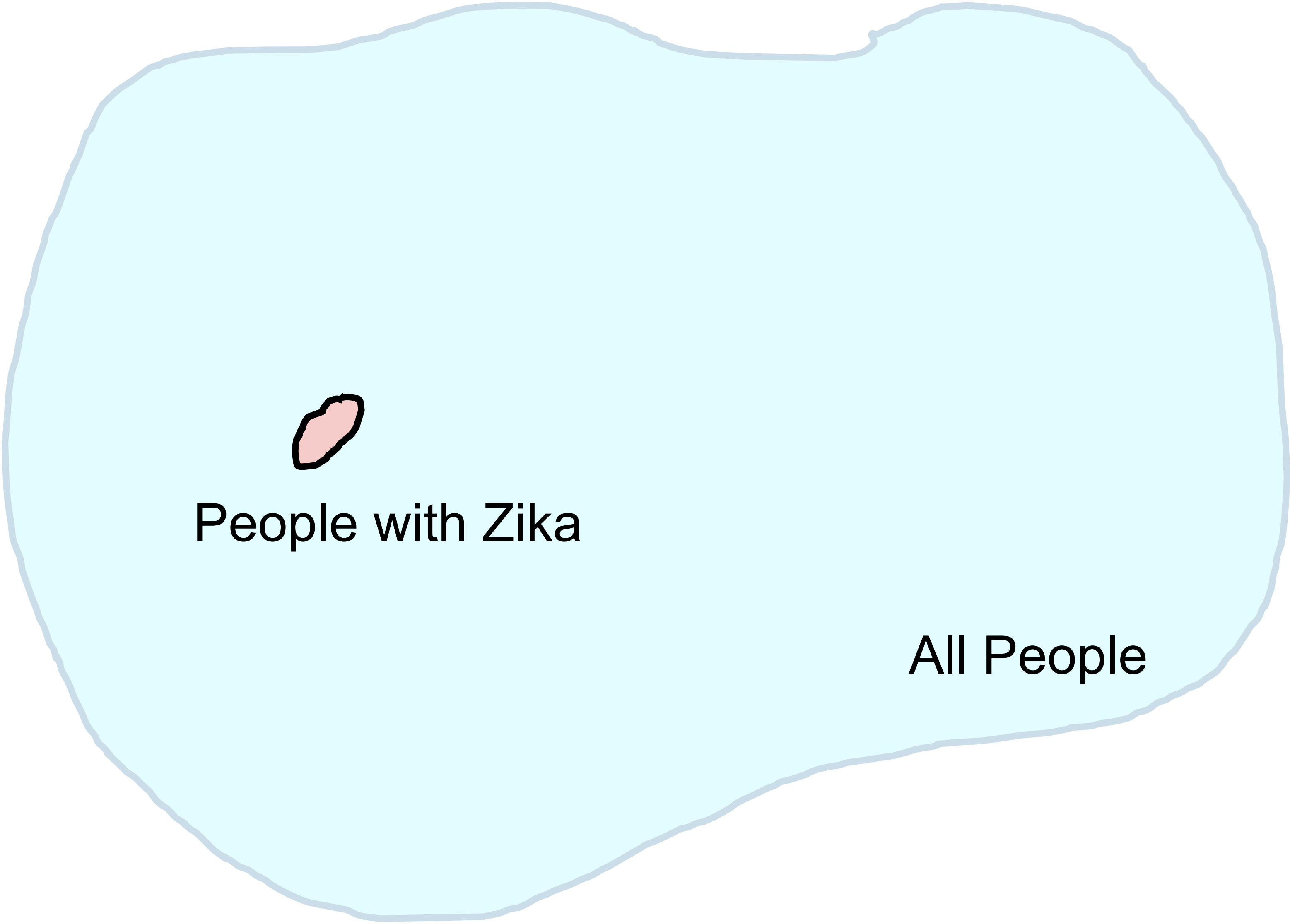
$$P(F \mid E) = \frac{(0.98)(0.005)}{(0.98)(0.005) + (0.01)(1 - 0.005)} \approx 0.33$$

Intuition Time

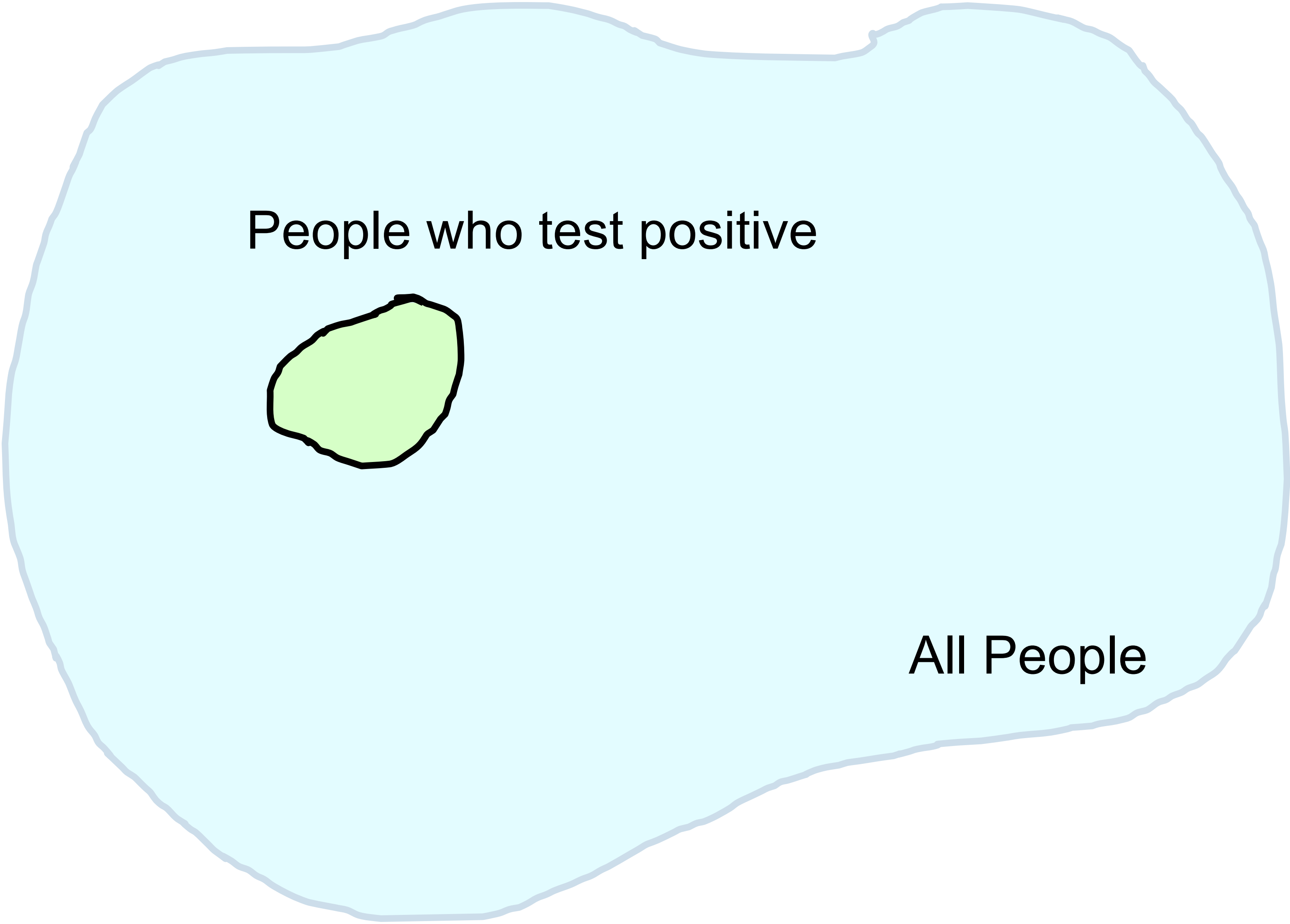
Bayes Thorem Intuition



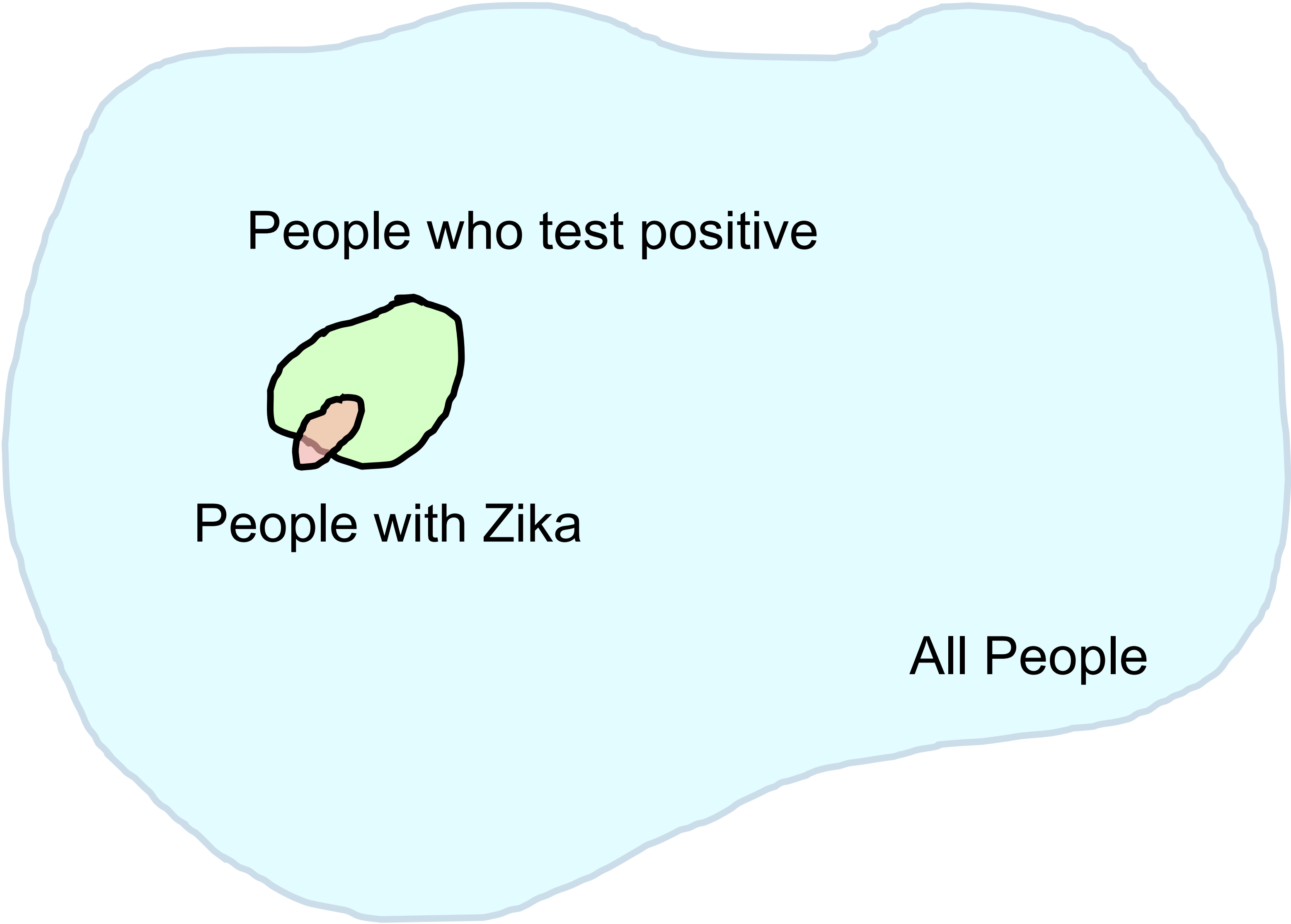
Bayes Thorem Intuition



Bayes Thorem Intuition

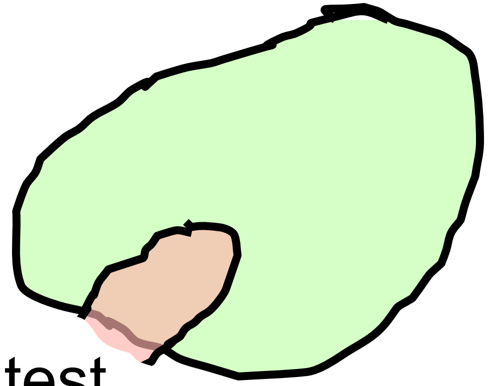


Bayes Thorem Intuition



Bayes Thorem Intuition

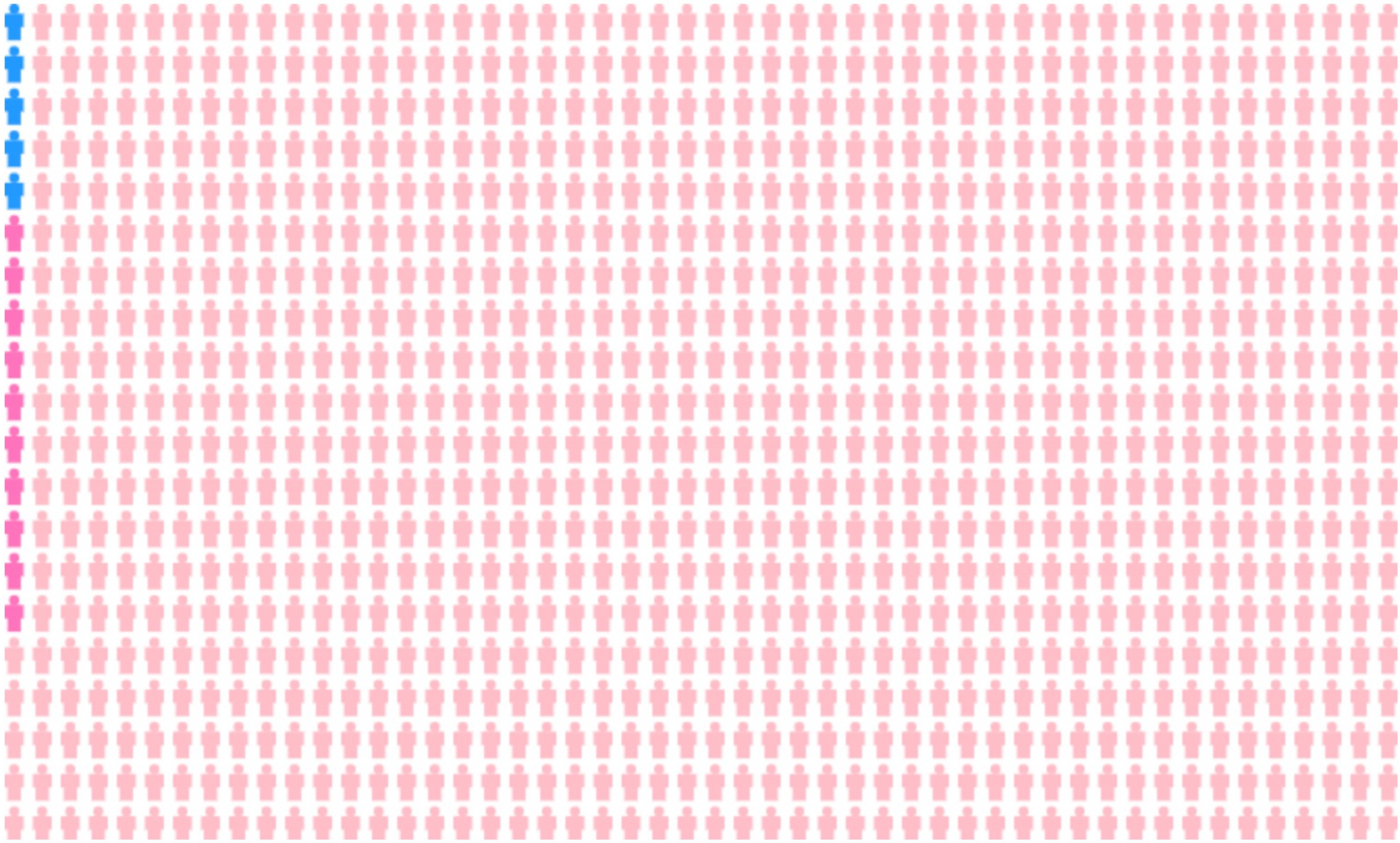
Conditioning on a positive result changes the sample space to this:



≈ 0.330

Bayes Thorem Intuition

Say we have 1000 people:



5 have Zika and test positive, 985 **do not** have Zika and test negative.
10 **do not** have Zika and test positive.

Bayes Thorem Intuition

Conditioned on just those that test positive:

Notice that all the people with Zika are here, but the group is still mainly folks without Zika

5 have Zika and test positive, 985 **do not** have Zika and test negative.
10 **do not** have Zika and test positive.

Why it is still good to get tested

- Let E^c = you test negative for Zika with this test.
- Let F = you actually have Zika.
- What is $P(F \mid E^c)$?

	Zika +	**Zika –**
Test +	$0.98 = P(E \mid F)$	$0.01 = P(E \mid F^c)$
Test –	$0.02 = P(E^c \mid F)$	$0.99 = P(E^c \mid F^c)$

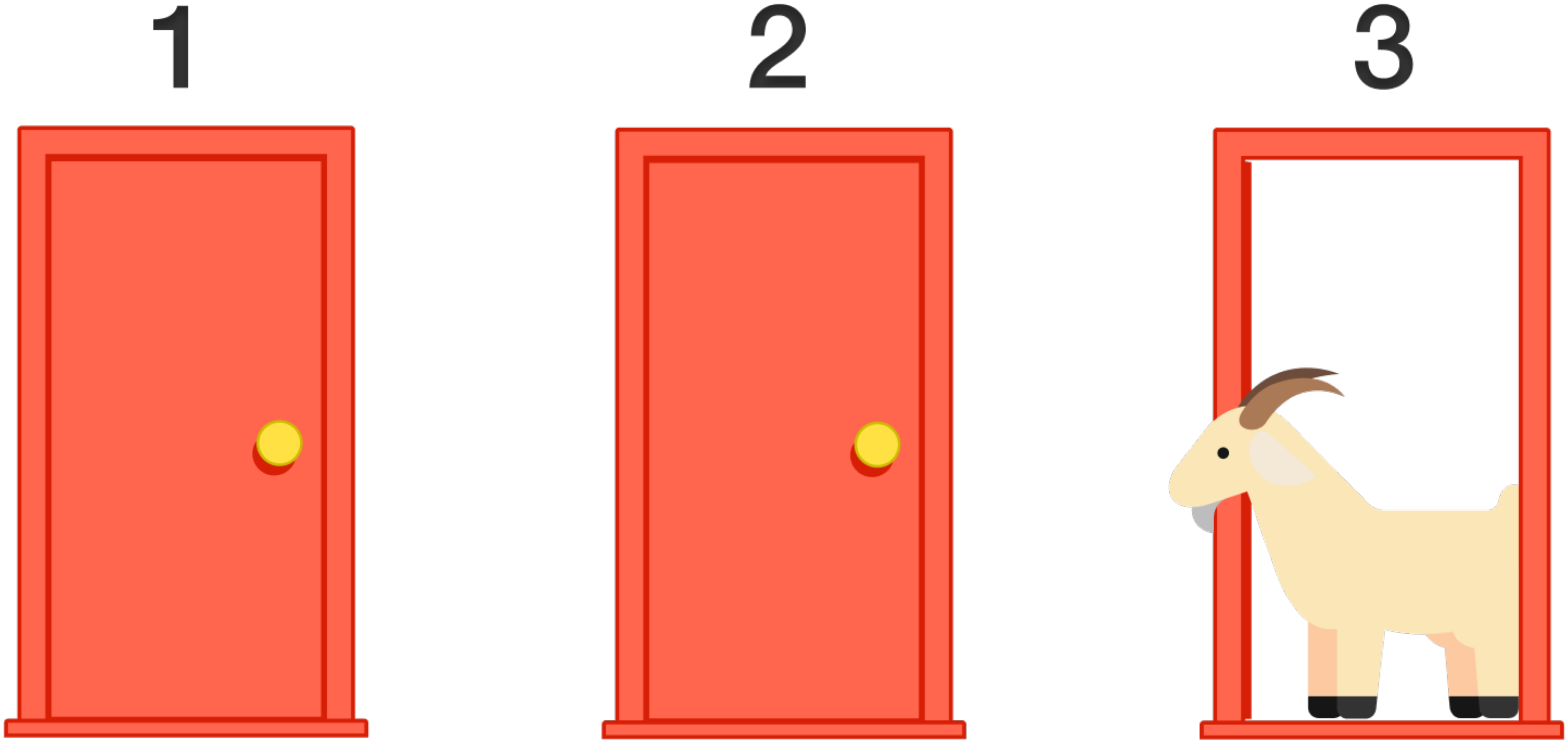
Why it is still good to get tested

- Let E^c = you test <u>negative</u> for Zika with this test.
- Let F = you actually have Zika.
- What is $P(F \mid E^c)$?

	Zika +	Zika –
Test +	$0.98 = P(E \mid F)$	$0.01 = P(E \mid F^c)$
Test –	$0.02 = P(E^c \mid F)$	$0.99 = P(E^c \mid F^c)$

$$P(F \mid E^c) = \frac{P(E^c \mid F)\, P(F)}{P(E^c \mid F)\, P(F) + P(E^c \mid F^c)\, P(F^c)}$$

$$P(F \mid E^c) = \frac{(0.02)(0.005)}{(0.02)(0.005) + (0.99)(1 - 0.005)} \approx 0.0001$$



The Monty Hall Problem

Monty Hall Problem aka Let's Make a Deal

Behind one door is a prize (equally likely to be any door).

Behind the other two doors is nothing

1. We choose a door
2. Host opens 1 of other 2 doors, revealing nothing
3. We are given an option to change to the other door.

Should we switch?

Doors A,B,C

Note: If we don't switch, P(win) = 1/3 (random)

We are comparing P(win) and P(win|switch).

Marilyn Vos Savant: "You Should Always Switch"

If we switch

Without loss of generality, say we pick A (out of Doors A,B,C).

1/3 1/3 1/3

A = prize
- Host opens B or C
- We switch
- We lose

P(win | A prize, picked A, switched) = 0

B = prize
- Host must open C
- We switch to B
- We win

P(win | B prize, picked A, switched) = 1

C = prize
- Host must open B
- We switch to C
- We **win**

P(win | C prize, picked A, switched) = 1

P(win | picked A, switched) = 1/3 * 0 + 1/3 * 1 + 1/3 * 1 = 2/3

You should switch.

Monty Hall, 1000 envelope version

Start with 1000 envelopes (of which 1 is the prize).

1. You choose 1 envelope.

$$\left\{\begin{array}{l} \frac{1}{1000} = P(\text{envelope is prize}) \\ \frac{999}{1000} = P(\text{other 999 envelopes have prize}) \end{array}\right.$$

2. I open 998 of remaining 999 (showing they are empty).

$$\frac{999}{1000} = P(\text{998 empty envelopes had prize}) + P(\text{last other envelope has prize})$$

$$= P(\text{last other envelope has prize})$$

3. Should you switch?

$$\text{No: } P(\text{win without switching}) = \frac{1}{\text{original \# envelopes}}$$

$$\text{Yes: } P(\text{win with new knowledge}) = \frac{\text{original \# envelopes - 1}}{\text{original \# envelopes}}$$

Next time: gambling!