

CSE 312

Foundations of Computing II

Lecture 5: Conditional Probability Introduction



Aleks Jovicic

Slide Credit: Based on Stefano Tessaro's slides for 312 19au incorporating ideas from Anna Karlin, Alex Tsun, Rachel Lin, Hunter Schafer & myself 😊

Belonging and CS Tas Research Study

Leah Perlmutter (she/her)

leahperl@uw.edu

tinyurl.com/belonging-study

Review Probability space

Either finite or infinite countable (e.g., integers)

Definition. A (discrete) **probability space** is a pair (Ω, \mathbb{P}) where:

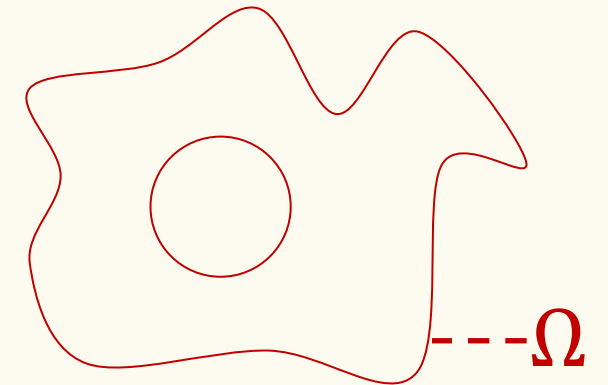
- Ω is a set called the **sample space**.

- \mathbb{P} is the **probability measure**, a function $\mathbb{P}: \Omega \rightarrow [0,1]$ such that:

- $\mathbb{P}(\omega) \geq 0$ for all $\omega \in \Omega$

- $\sum_{\omega \in \Omega} \mathbb{P}(\omega) = 1$

Set of possible elementary outcomes



Specify Likelihood (or probability) of each elementary outcome

Some outcome must show up

The likelihood (or probability) of each outcome is non-negative.

Review Axioms of Probability

Let Ω denote the sample space and $E, F \subseteq \Omega$ be events. Note this is more general to **any** probability space (not just uniform)

Axiom 1 (Non-negativity): $P(E) \geq 0$

Axiom 2 (Normalization): $P(\Omega) = 1$

Axiom 3 (Countable Additivity): If E and F are mutually exclusive, then $P(E \cup F) = P(E) + P(F)$

Corollary 1 (Complementation): $P(E^c) = 1 - P(E)$

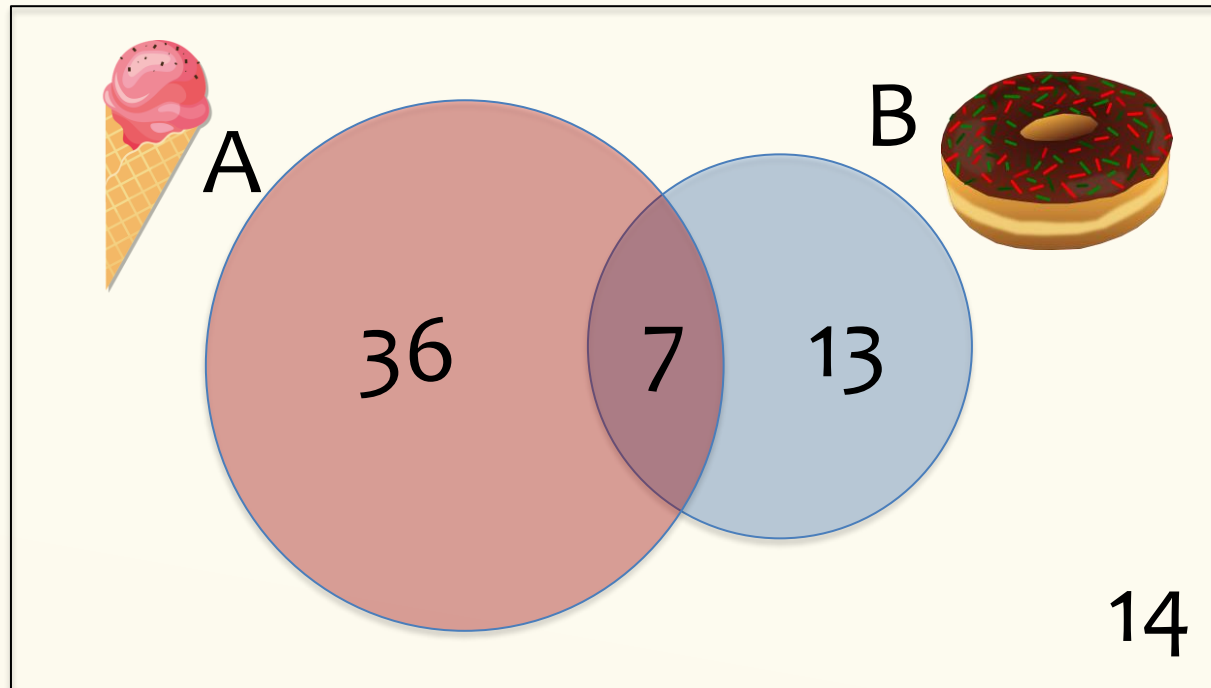
Corollary 2 (Monotonicity): If $E \subseteq F$, $P(E) \leq P(F)$

Corollary 3 (Inclusion-Exclusion): $P(E \cup F) = P(E) + P(F) - P(E \cap F)$

Agenda

- Conditional Probability ◀
- Time Permitting:
 - Bayes Theorem
 - Law of Total Probability
 - Bayes Theorem + Law of Total Probability
 - More Examples

Conditional Probability (Idea)



What's the probability that someone likes ice cream **given** they like donuts?

Conditional Probability

Definition. The **conditional probability** of event A given an event B happened (assuming $P(B) \neq 0$) is

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

An equivalent and useful formula is

$$P(A \cap B) = P(A|B)P(B)$$

Reversing Conditional Probability

Question: Does $P(A|B) = P(B|A)$?

No! The following is purely for intuition and makes no sense in terms of probability

- Let A be the event you are wet
- Let B be the event you are swimming

$$P(A|B) = 1$$

$$P(B|A) \neq 1$$

Example with Conditional Probability

Toss a red die and a blue die (both 6 sided and all outcomes equally likely). What is $P(B)$? What is $P(B|A)$?

	$P(B)$	$P(B A)$
a)	$1/6$	$1/6$
b)	$1/6$	$1/3$
c)	$1/6$	$3/36$
d)	$1/9$	$1/3$

B = red die is 1

A = sum is 4

Example with Conditional Probability

Toss a red die and a blue die (both 6 sided and all outcomes equally likely). What is $P(B)$? What is $P(B|A)$?

	$P(B)$	$P(B A)$
a)	1/6	1/6
b)	1/6	1/3
c)	1/6	3/36
d)	1/9	1/3

B = red die is 1

A = sum is 4

$$\mathbb{P}(B) = \frac{1}{6}$$
$$\mathbb{P}(B|A) = \frac{1}{3}$$

Gambler's fallacy

Assume we toss 51 fair coins.

Assume we have seen 50 coins, and they are all “tails”.

What are the odds the 51st coin is “heads”?

\mathcal{A} = first 50 coins are “tails”

\mathcal{B} = 51st coin is “heads”

$$\mathbb{P}(\mathcal{B}|\mathcal{A}) =$$

Gambler's fallacy

Assume we toss 51 fair coins.

Assume we have seen 50 coins, and they are all “tails”.

What are the odds the 51st coin is “heads”?

\mathcal{A} = first 50 coins are “tails”


\mathcal{B} = 51st coin is “heads”

$$\mathbb{P}(\mathcal{B}|\mathcal{A}) = \frac{\mathbb{P}(\mathcal{A} \cap \mathcal{B})}{\mathbb{P}(\mathcal{A})} = \frac{1/2^{51}}{2/2^{51}} = \frac{1}{2}$$

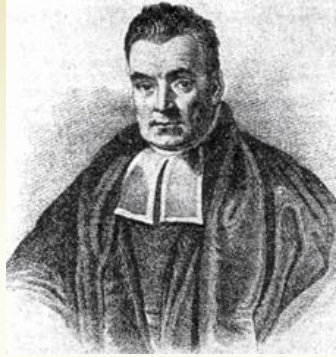
51st coin is independent of
outcomes of first 50 tosses!

Gambler's fallacy = Feels like it's time for “heads”!?

Agenda

- Conditional Probability
- Time Permitting:
 - Bayes Theorem 
 - Law of Total Probability
 - Bayes Theorem + Law of Total Probability
 - More Examples

Bayes Theorem



A formula to let us “reverse” the conditional.

Theorem. (Bayes Rule) For events A and B , where $P(A), P(B) > 0$,

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

$P(A)$ is called the **prior** (our belief without knowing anything)

$P(A|B)$ is called the **posterior** (our belief after learning B)

Bayes Theorem Proof

By definition of conditional probability

$$P(A \cap B) = P(A|B)P(B)$$

Swapping A, B gives

$$P(B \cap A) = P(B|A)P(A)$$

But $P(A \cap B) = P(B \cap A)$, so

$$P(A|B)P(B) = P(B|A)P(A)$$

Dividing both sides by $P(B)$ gives

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

Our First Machine Learning Task: Spam Filtering

Subject: “FREE \$\$\$ CLICK HERE”

What is the probability this email is spam, given the subject contains “FREE”?

Some useful stats:

- 10% of ham (i.e., not spam) emails contain the word “FREE” in the subject.
- 70% of spam emails contain the word “FREE” in the subject.
- 80% of emails you receive are spam.

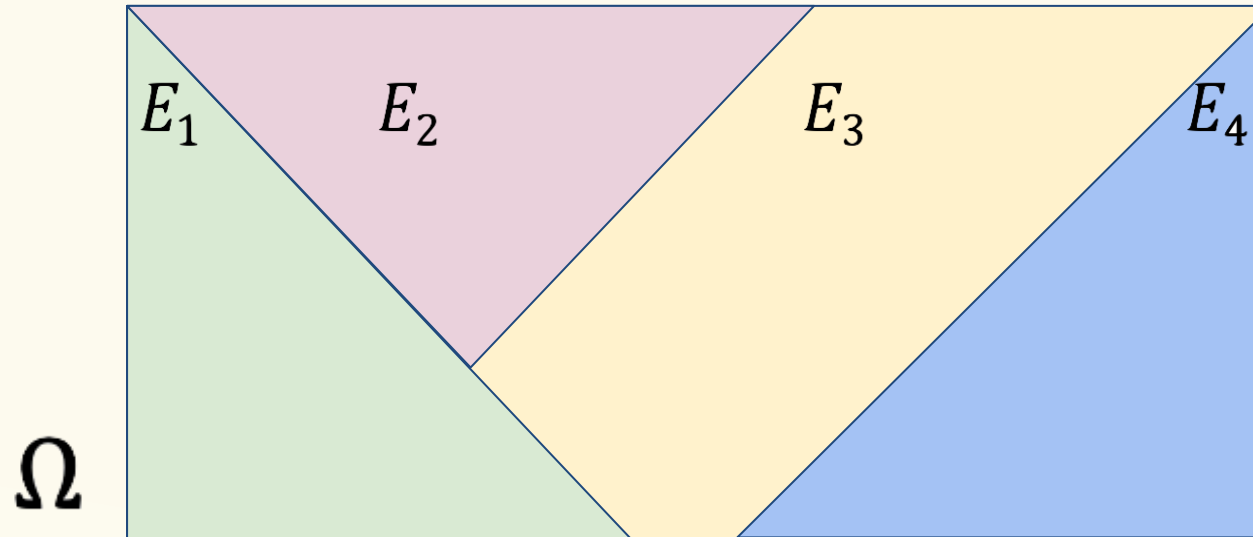
Agenda

- Conditional Probability
- Time Permitting:
 - Bayes Theorem
 - Law of Total Probability ◀
 - Bayes Theorem + Law of Total Probability
 - More Examples

Partitions (Idea)

These events **partition** the sample space

1. They “cover” the whole space
2. They don’t overlap



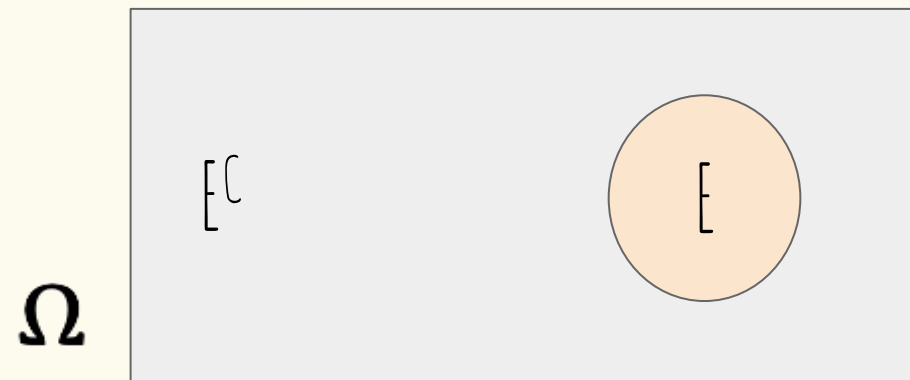
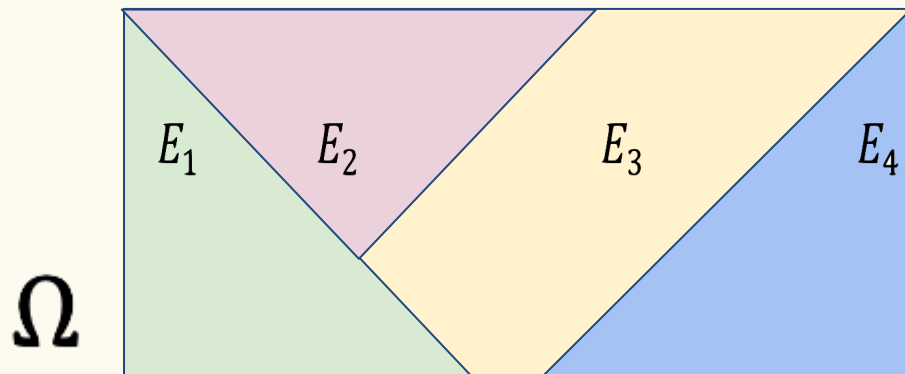
Partition

Definition. Non-empty events E_1, E_2, \dots, E_n **partition** the sample space Ω if
(Exhaustive)

$$E_1 \cup E_2 \cup \dots \cup E_n = \bigcup_{i=1}^n E_i = \Omega$$

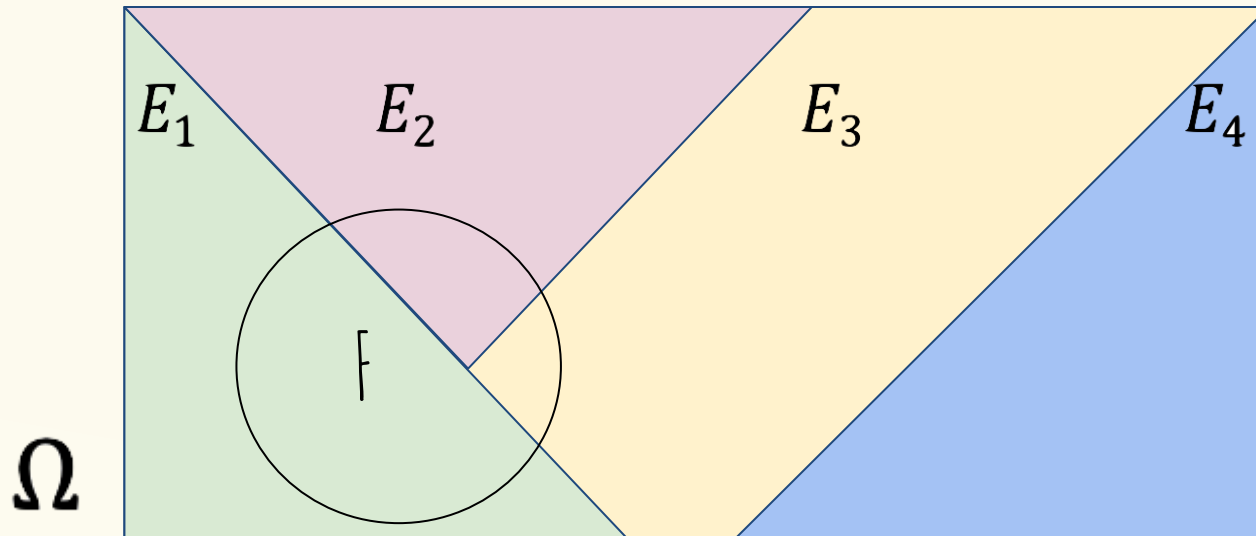
(Pairwise Mutually Exclusive)

$$\forall_i \forall_{i \neq j} E_i \cap E_j = \emptyset$$



Law of Total Probability (Idea)

If we know E_1, E_2, \dots, E_n partition Ω , what can we say about $P(F)$



Law of Total Probability (LTP)

Definition. If events E_1, E_2, \dots, E_n partition the sample space Ω , then for any event F

$$P(F) = P(F \cap E_1) + \dots + P(F \cap E_n) = \sum_{i=1}^n P(F \cap E_i)$$

Using the definition of conditional probability $P(F \cap E) = P(F|E)P(E)$

We can get the alternate form of this that show

$$P(F) = \sum_{i=1}^n P(F|E_i)P(E_i)$$

Another Contrived Example

Alice has two pockets:

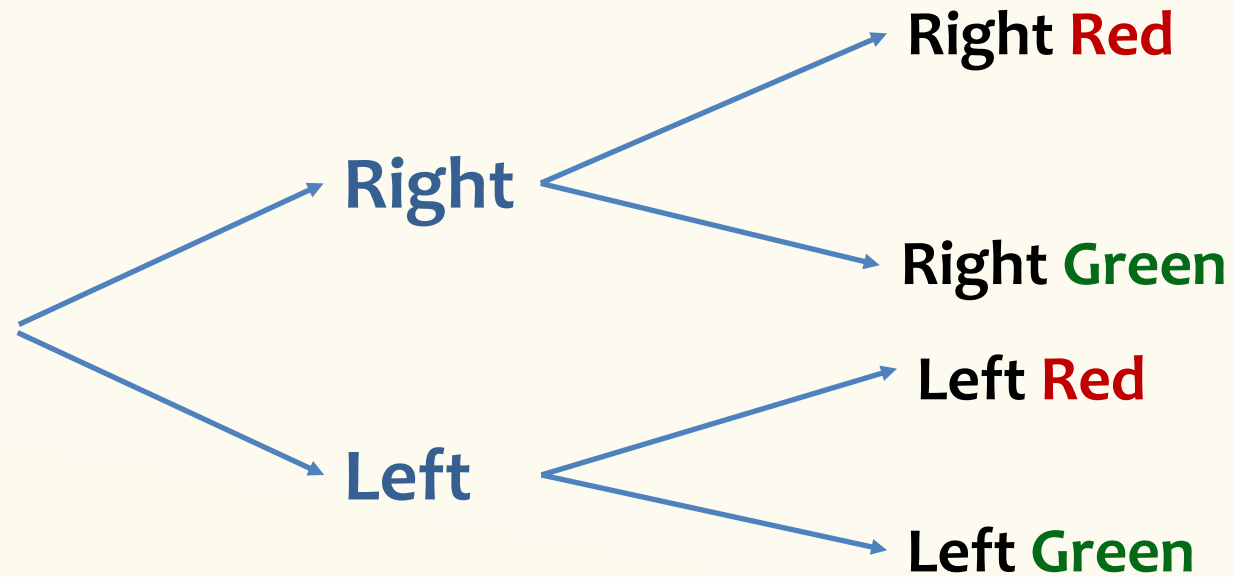
- **Left pocket:** Two red balls, two green balls
- **Right pocket:** One red ball, two green balls.

Alice picks a random ball from a random pocket.

[Both pockets equally likely, each ball equally likely.]

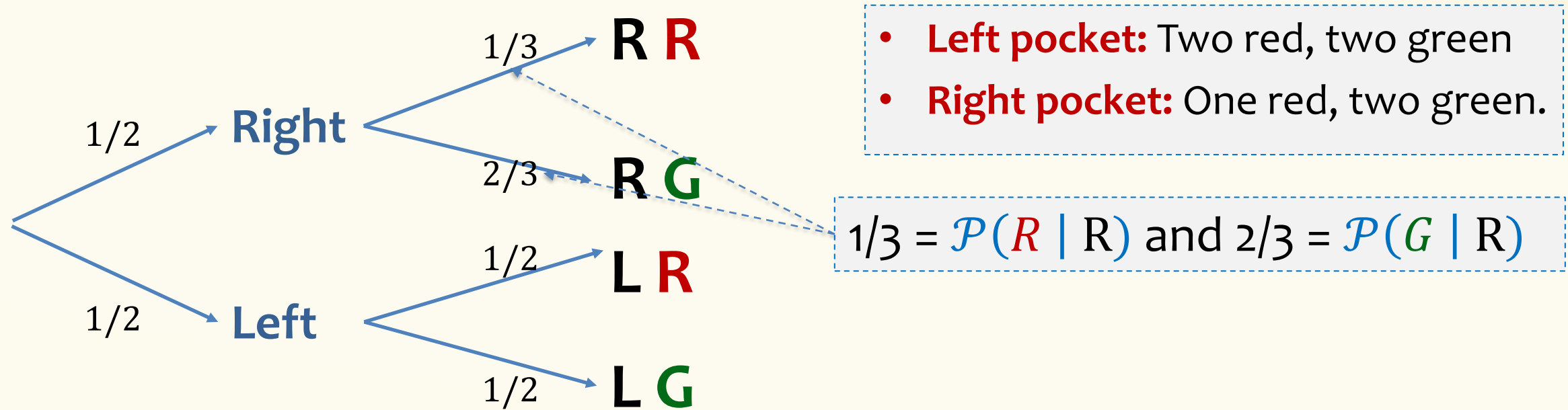
What is $\mathbb{P}(\mathbf{R})$?

Sequential Process – Non-Uniform Case



- **Left pocket:** Two red, two green
- **Right pocket:** One red, two green.
- Alice picks a random ball from a random pocket

Sequential Process – Non-Uniform Case




$$\mathbb{P}(R) = \mathbb{P}(R \cap \text{Left}) + \mathbb{P}(R \cap \text{Right}) \quad (\text{Law of total probability})$$

$$= \mathbb{P}(\text{Left}) \times \mathbb{P}(R|\text{Left}) + \mathbb{P}(\text{Right}) \times \mathbb{P}(R|\text{Right})$$

$$= \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{3} = \frac{1}{4} + \frac{1}{6} = \frac{5}{12}$$

Agenda

- Conditional Probability
- Time Permitting:
 - Bayes Theorem
 - Law of Total Probability
 - Bayes Theorem + Law of Total Probability 
 - More Examples

Our First Machine Learning Task: Spam Filtering

Subject: “FREE \$\$\$ CLICK HERE”

What is the probability this email is spam, given the subject contains “FREE”?

Some useful stats:

- 10% of ham (i.e., not spam) emails contain the word “FREE” in the subject.
- 70% of spam emails contain the word “FREE” in the subject.
- 80% of emails you receive are spam.

Bayes Theorem with Law of Total Probability


Bayes Theorem with LTP: Let E_1, E_2, \dots, E_n be a partition of the sample space, and F and event. Then,

$$P(E_1|F) = \frac{P(F|E_1)P(E_1)}{P(F)} = \frac{P(F|E_1)P(E_1)}{\sum_{i=1}^n P(F|E_i)P(E_i)}$$

Simple Partition: In particular, if E is an event with non-zero probability, then

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

Agenda

- Conditional Probability
- Time Permitting:
 - Bayes Theorem
 - Law of Total Probability
 - Bayes Theorem + Law of Total Probability
 - More Examples 

Example – Zika Testing

Zika fever

OVERVIEW SYMPTOMS SPECIALISTS

Fever
Rash
Joint pain
Red eyes



Spread through mosquito bites *Source*

A disease caused by Zika virus that's spread through mosquito bites.

The image shows a woman with a red rash on her neck and shoulder. A circular inset shows a mosquito biting her skin. The text 'Spread through mosquito bites' and 'Source' is written below the inset.

Usually no or mild symptoms (rash); sometimes severe symptoms (paralysis).

During pregnancy: may cause birth defects.

Suppose you took a Zika test, and it returns “positive”, what is the likelihood that you actually have the disease?

- Tests for diseases are rarely 100% accurate.

Example – Zika Testing

Suppose we know the following Zika stats

- A test is 98% effective at detecting Zika (“true positive”)
- However, the test may yield a “false positive” 1% of the time
- 0.5% of the US population has Zika.

What is the probability you have Zika (event Z) if you test positive (event T).

- A) Less than 0.25
- B) Between 0.25 and 0.5
- C) Between 0.5 and 0.75
- D) Between 0.75 and 1

Example – Zika Testing

Suppose we know the following Zika stats

- A test is 98% effective at detecting Zika (“true positive”)
- However, the test may yield a “false positive” 1% of the time
- 0.5% of the US population has Zika.

What is the probability you have Zika (event Z) if you test positive (event T).

Example – Zika Testing

Have zika blue, don't pink

Suppose we know the following Zika stats

- A test is 98% effective at detecting Zika (“true positive”)
- However, the test may yield a “false positive” 1% of the time
- 0.5% of the US population has Zika. 5% have it.

What is the probability you have Zika (event Z) if you test positive (event T).



Suppose we had 1000 people:

- 5 have Zika and test positive
- 985 do not have Zika and test negative
- 10 do not have Zika and test positive

$$\frac{5}{5 + 10} = \frac{1}{3} \approx 0.33$$

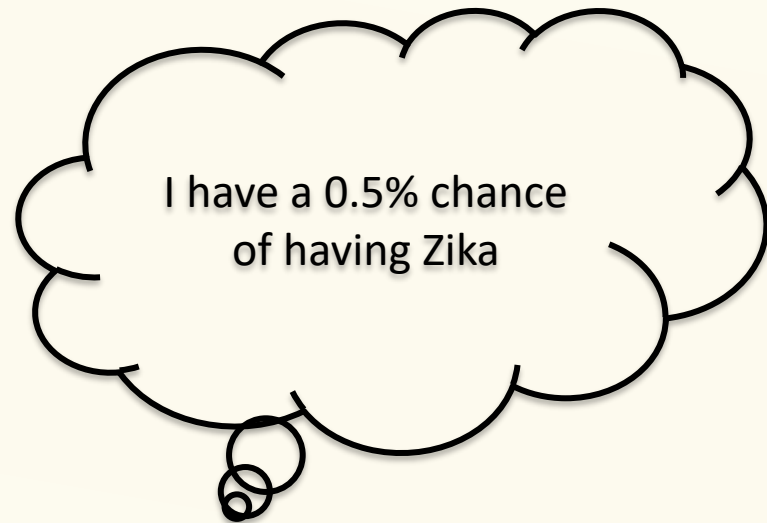
[Demo](#)

Philosophy – Updating Beliefs

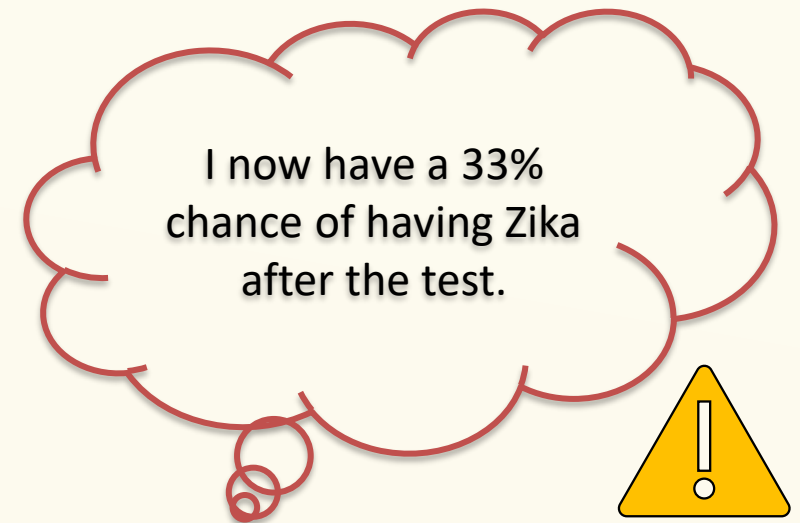
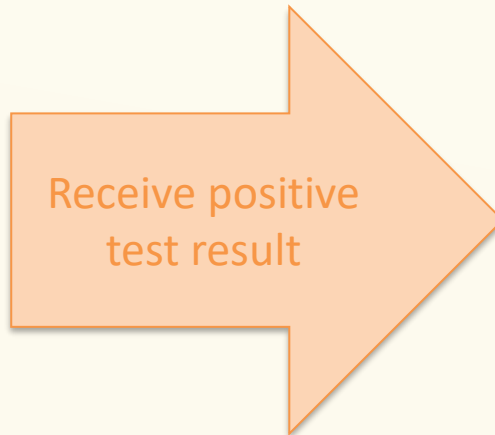
While it's not 98% that you have the disease, your beliefs changed **drastically**

Z = you have Zika

T = you test positive for Zika



Prior: $P(Z)$



Posterior: $P(Z|T)$

Example – Zika Testing

Suppose we know the following Zika stats

- A test is 98% effective at detecting Zika (“true positive”)
- However, the test may yield a “false positive” 1% of the time
- 0.5% of the US population has Zika.

What is the probability you test negative (event \bar{T}) if you have Zika (event Z)?

Conditional Probability Define a Probability Space

The probability conditioned on A follows the same properties as (unconditional) probability.


Example. $\mathbb{P}(\mathcal{B}^c|\mathcal{A}) = 1 - \mathbb{P}(\mathcal{B}|\mathcal{A})$

Conditional Probability Define a Probability Space

The probability conditioned on A follows the same properties as (unconditional) probability.

Example. $\mathbb{P}(\mathcal{B}^c | \mathcal{A}) = 1 - \mathbb{P}(\mathcal{B} | \mathcal{A})$

Formally. (Ω, \mathbb{P}) is a probability space + $\mathbb{P}(\mathcal{A}) > 0$

 $(\mathcal{A}, \mathbb{P}(\cdot | \mathcal{A}))$ is a probability space