Lecture 15

Bayes' Theorem and Independence

DSC 40A, Spring 2024

Announcements

- Homework 6 is due **tonight**.
  - Combinatorics is hard! Come to office hours.
- No discussion or office hours on Monday.
  - We'll still post a groupwork worksheet and its solutions, just for extra practice.
- There will be no live lecture on Tuesday. Instead, the lecture video will be pre-recorded and posted on the course website by Tuesday morning.
  - There's also a lecture note I wrote for Tuesday's lecture that you should read.
- The final exam is in two weeks from Saturday: start practicing at practice.dsc40a.com!
  - (Even) more probability problems coming soon.

Agenda

- Law of Total Probability.
- Bayes' Theorem.
- Independence.

Remember, we've posted **many** probability resources on the resources tab of the course website. These will come in handy! Specific resources you should look at:

- The DSC 40A probability roadmap, written by Janine Tiefenbruck.
- The textbook Theory Meets Data, which explains many of the same ideas and contains more practice problems.

For combinatorics specifically, there are two supplementary videos I created that you should watch. Both are linked in this playlist, which is also linked at dsc40a.com.

Question 🤔

Answer at q.dsc40a.com

Remember, you can always ask questions at q.dsc40a.com!

If the direct link doesn't work, click the "🤔 Lecture Questions" link in the top right corner of dsc40a.com.

Law of Total Probability

Example: Getting to school

You conduct a survey where you ask students two questions:

1. How did you get to campus today – trolley, bike, or drive? (Assume these are the only options.)

2. Were you late?

all sum to 1!

	Late	Not Late
Trolley	0.06	0.24
Bike	0.03	0.07
Drive	0.36	0.24

$P(\text{Trolley} \cap \text{Late})$

$P(\text{Bike} \cap \text{Not Late})$

Question 🤔

Answer at q.dsc40a.com

$\mathbb{P}(\text{Trolley} \cap \text{Late})$

$\mathbb{P}(\text{Trolley})$

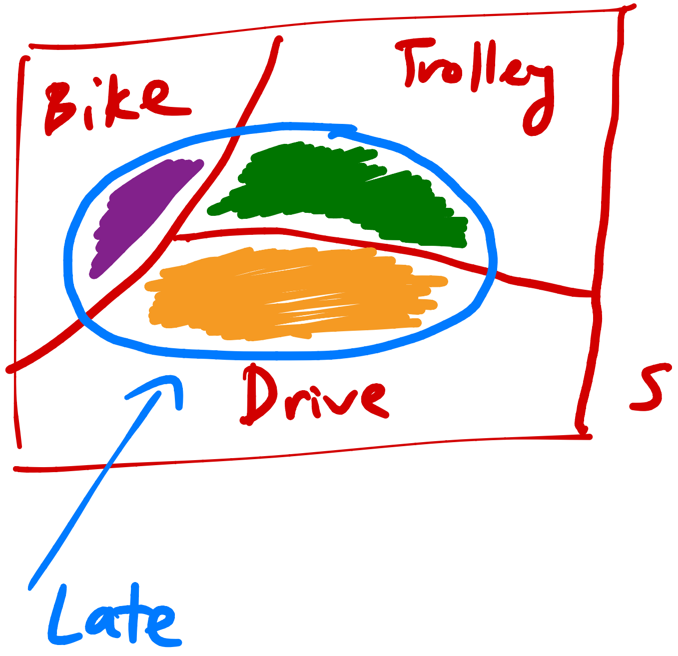
	Late	Not Late	
Trolley	0.06	0.24	0.3
Bike	0.03	0.07	0.1
Drive	0.36	0.24	0.6
	0.45	0.55	

What's the probability that a randomly selected person was late?

A. 0.24 B. 0.30 C. 0.45 D. 0.50 E. None of the above.

Example: Getting to school

	Late	Not Late
Trolley	0.06	0.24
Bike	0.03	0.07
Drive	0.36	0.24



- Since everyone either takes the trolley, bikes, or drives to school, we have:

$$\underbrace{\mathbb{P}(\text{Late})}_{0.45} = \underbrace{\mathbb{P}(\text{Late} \cap \text{Trolley})}_{0.06} + \underbrace{\mathbb{P}(\text{Late} \cap \text{Bike})}_{0.03} + \underbrace{\mathbb{P}(\text{Late} \cap \text{Drive})}_{0.36}$$

Question 🤔

Answer at q.dsc40a.com

$$\mathbb{P}(\text{Late} \mid \text{Trolley}) = \frac{\mathbb{P}(\text{Late} \cap \text{Trolley})}{\mathbb{P}(\text{Trolley})}$$

$$= \frac{\mathbb{P}(\text{Late} \cap \text{Trolley})}{\mathbb{P}(\text{Late} \cap \text{Trolley}) + \mathbb{P}(\text{Not Late} \cap \text{Trolley})}$$

$$= \frac{0.06}{0.06 + 0.24}$$

$$= \frac{6}{30} = \frac{1}{5} = \boxed{0.2}$$

	Late	Not Late
Trolley	0.06	0.24
Bike	0.03	0.07
Drive	0.36	0.24

Avi took the trolley to school. What is the probability that he was late?

A. 0.06 B. 0.20 C. 0.25 D. 0.45 E. None of the above.

Example: Getting to school

	Late	Not Late
Trolley	0.06	0.24
Bike	0.03	0.07
Drive	0.36	0.24

- Since everyone either takes the trolley, bikes, or drives to school, we have:

$$\mathbb{P}(\text{Late}) = \mathbb{P}(\text{Late} \underbrace{\cap}_{} \text{Trolley}) + \mathbb{P}(\text{Late} \cap \text{Bike}) + \mathbb{P}(\text{Late} \cap \text{Drive})$$

"and"

same, from multiplication rule!

- Another way of expressing the same thing:

$$\mathbb{P}(\text{Late}) = \mathbb{P}(\text{Trolley})\ \mathbb{P}(\text{Late}|\text{Trolley}) + \mathbb{P}(\text{Bike})\ \mathbb{P}(\text{Late}|\text{Bike}) + \mathbb{P}(\text{Drive})\ \mathbb{P}(\text{Late}|\text{Drive})$$

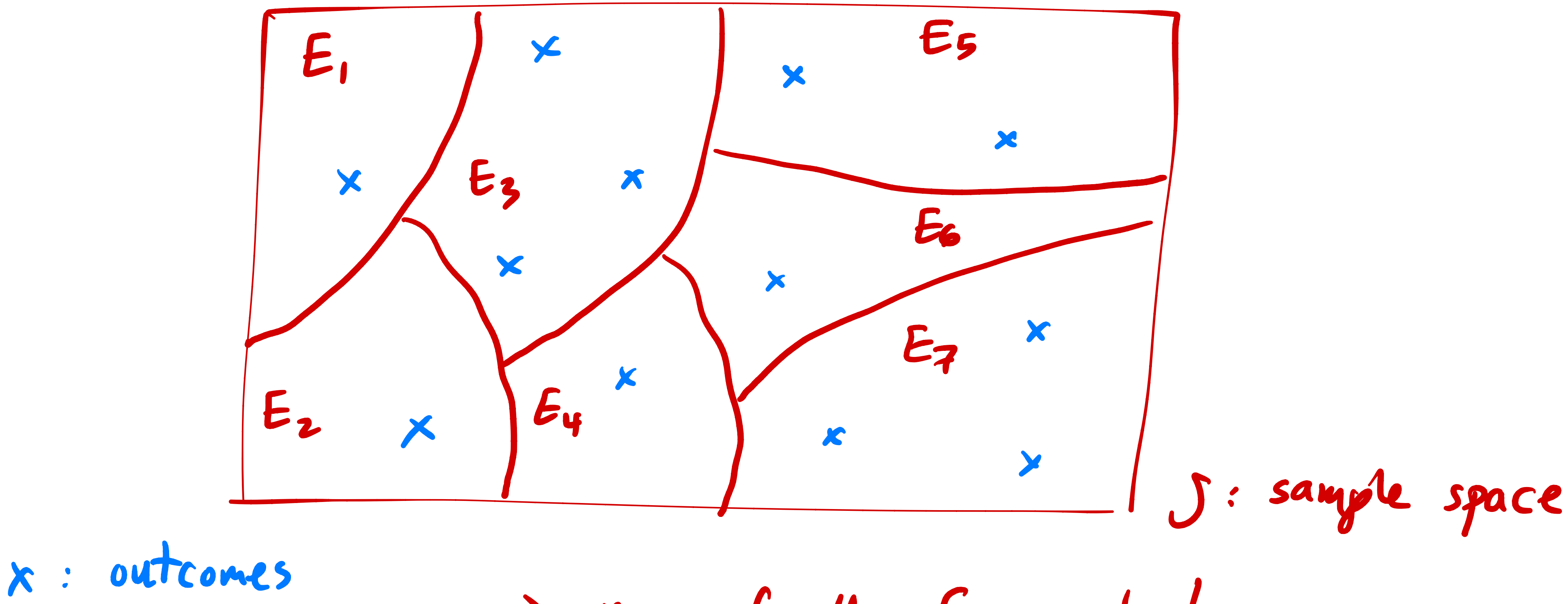
0.2, from previous slide

Partitions

sample space

- A set of events $E_1, E_2, \ldots, E_k$ is a **partition** of S if:
  - $\mathbb{P}(E_i \cap E_j) = 0$ for all pairs $i \neq j$. no overlap!
  - $\mathbb{P}(E_1 \cup E_2 \cup \ldots \cup E_k) = 1.$
    - Equivalently, $\mathbb{P}(E_1) + \mathbb{P}(E_2) + \ldots + \mathbb{P}(E_k) = 1.$
- In other words, $E_1, E_2, \ldots, E_k$ is a partition of S if every outcome $s \in S$ is in **exactly** one event E_i.

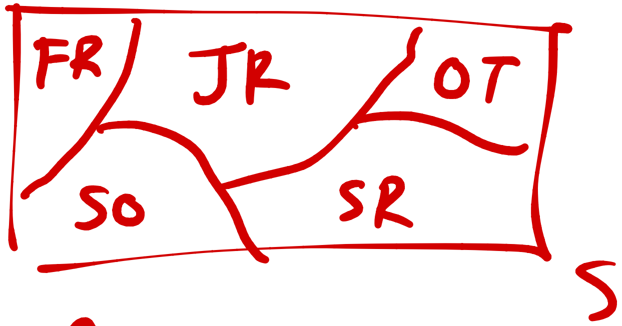
Example



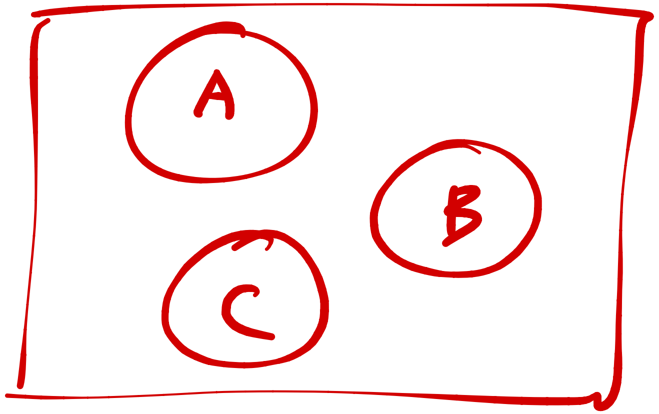
$\Rightarrow$ none of the Es overlap!

$\Rightarrow$ every outcome x is in <u>exactly</u> one E_i!

Example partitions



- In getting to school, the events Trolley, Bike, and Drive.
- In getting to school, the events Late and Not Late.
- In selecting an undergraduate student at random, the events Freshman, Sophomore, Junior, and Senior, and Other.
- In rolling a die, the events Even and Odd.
- In drawing a card from a standard deck of cards, the events Spades, Clubs, Hearts, and Diamonds.
- Special case: Any event A and its complement $\bar{A}$.



Here, A, B, and C are all pairwise mutually exclusive, but are $\boxed{\text{not a partion}}$, because $\mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) \neq 1$.

The Law of Total Probability

- If A is an event and $E_1, E_2, \ldots, E_k$ is a **partition** of S, then:

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(A \cap E_1) + \mathbb{P}(A \cap E_2) + \ldots + \mathbb{P}(A \cap E_k) \\ &= \sum_{i=1}^{k} \mathbb{P}(A \cap E_i)\end{aligned}$$

We've already seen:

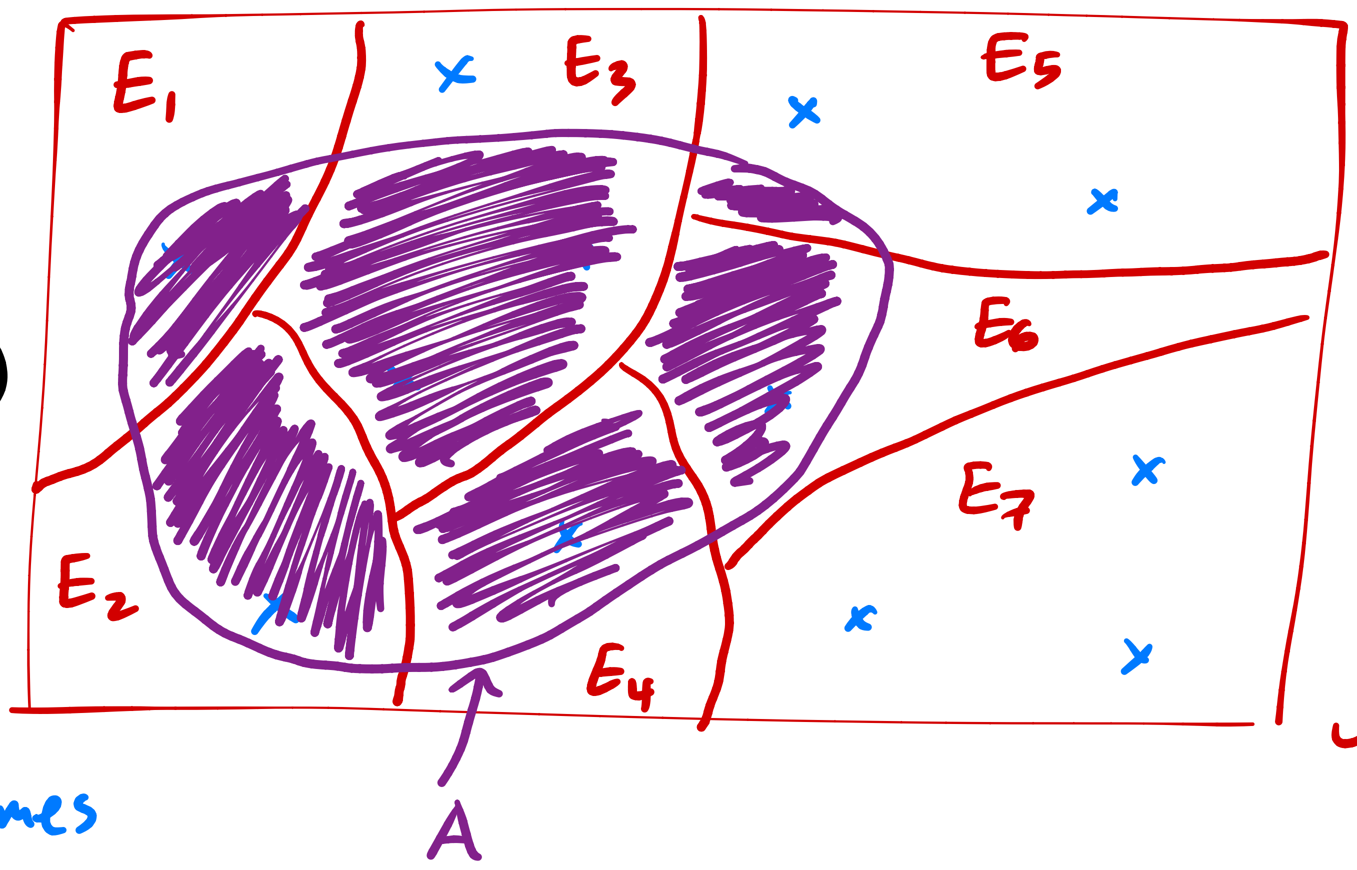
$\mathbb{P}(\text{Late}) = \mathbb{P}(\text{Late} \cap \text{Trolley}) + \mathbb{P}(\text{Late} \cap \text{Drive}) + \mathbb{P}(\text{Late} \cap \text{Bike})$

A : Late

E_1 : Trolley, E_2 : Drive, E_3 : Bike

Aside:

$$\mathbb{P}(A \cap E_1) = \mathbb{P}(A)\cdot \mathbb{P}(E_1 \mid A) = \mathbb{P}(E_1)\cdot \mathbb{P}(A \mid E_1)$$



S : sample space

x : outcomes

0

$$\mathbb{P}(A) = \mathbb{P}(A\cap E_1) + \mathbb{P}(A\cap E_2) + \mathbb{P}(A\cap E_3) + \mathbb{P}(A\cap E_4) + \mathbb{P}(A\cap E_5) + \mathbb{P}(A\cap E_6) + \mathbb{P}(A\cap E_7)$$

$$\mathbb{P}(A) = \sum_{i=1}^{7} \mathbb{P}(A\cap E_i)$$

The Law of Total Probability

- If A is an event and $E_1, E_2, \ldots, E_k$ is a **partition** of S, then:

uses "and"

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(A \cap E_1) + \mathbb{P}(A \cap E_2) + \ldots + \mathbb{P}(A \cap E_k) \\ &= \sum_{i=1}^{k} \mathbb{P}(A \cap E_i)\end{aligned}$$

- Since $\mathbb{P}(A \cap E_i) = \mathbb{P}(E_i) \cdot \mathbb{P}(A|E_i)$ by the multiplication rule, an equivalent formulation is:

uses conditional prob.

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(E_1) \cdot \mathbb{P}(A|E_1) + \mathbb{P}(E_2) \cdot \mathbb{P}(A|E_2) + \ldots + \mathbb{P}(E_k) \cdot \mathbb{P}(A|E_k) \\ &= \sum_{i=1}^{k} \mathbb{P}(E_i) \cdot \mathbb{P}(A|E_i)\end{aligned}$$

Question 🤔

Answer at q.dsc40a.com

$$\mathbb{P}(\text{Trolley} \mid \text{Late}) = \frac{\mathbb{P}(\text{Trolley} \cap \text{Late})}{\mathbb{P}(\text{Late})}$$

$$= \frac{\mathbb{P}(\text{Trolley} \cap \text{Late})}{\mathbb{P}(\text{Trolley} \cap \text{Late}) + \mathbb{P}(\text{Bike} \cap \text{Late}) + \mathbb{P}(\text{Drive} \cap \text{Late})}$$

	Late	Not Late
Trolley	0.06	0.24
Bike	0.03	0.07
Drive	0.36	0.24

$$= \frac{0.06}{0.06 + 0.03 + 0.36}$$

$$= \frac{0.06}{0.45} = \frac{6}{45} = \frac{12}{90} \approx \frac{15}{100}$$

Lauren is late to school. What is the probability that she took the trolley? Choose the best answer.

A. About 0.05 B. About 0.15 C. About 0.30 D. About 0.40

Bayes' Theorem

Example: Getting to school

- Now, suppose we don't have that entire table. Instead, all you know is:
  - $\mathbb{P}(\text{Late}) = 0.45$.
  - $\mathbb{P}(\text{Trolley}) = 0.3$.
  - $\mathbb{P}(\text{Late} | \text{Trolley}) = 0.2$.
- Can we still find $\mathbb{P}(\text{Trolley} | \text{Late})$?

given $\mathbb{P}(A|B)$

multiplication rule, applied carefully!

$$\mathbb{P}(\text{Trolley} | \text{Late}) = \frac{\mathbb{P}(\text{Trolley} \cap \text{Late})}{\mathbb{P}(\text{Late})} = \frac{\mathbb{P}(\text{Trolley}) \cdot \mathbb{P}(\text{Late} | \text{Trolley})}{\mathbb{P}(\text{Late})}$$

asked for $\mathbb{P}(B|A)$

$$= \frac{0.3 \cdot 0.2}{0.45} = \frac{0.06}{0.45} = \boxed{\frac{6}{45}}$$

Bayes' Theorem "Bayes' Rule"

- Recall that the multiplication rule states that:

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

- It also states that:

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

- But since $A \cap B = B \cap A$, we have that:

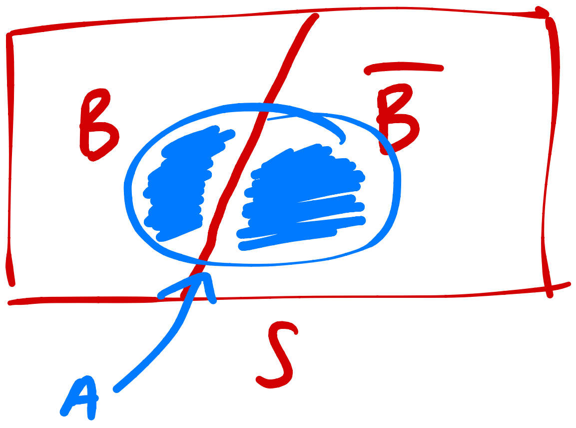
$$\mathbb{P}(A) \cdot \underbrace{\mathbb{P}(B|A)}_{\text{isolate}} = \mathbb{P}(B) \cdot \mathbb{P}(A|B)$$

- Re-arranging yields **Bayes' Theorem**:

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

"reverse" a conditional probability

Bayes' Theorem and the Law of Total Probability



- Bayes' Theorem:

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

$$\mathbb{P}(A) = \underbrace{\mathbb{P}(A \cap B)}_{\text{mult.}} + \underbrace{\mathbb{P}(A \cap \bar{B})}_{\text{mult.}}$$

- Recall from earlier, for any sample space S, B and $\bar{B}$ partition S. Using the Law of Total Probability, we can re-write $\mathbb{P}(A)$ as:

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap \bar{B}) = \mathbb{P}(B) \cdot \mathbb{P}(A|B) + \mathbb{P}(\bar{B}) \cdot \mathbb{P}(A|\bar{B})$$

- This means that we can re-write Bayes' Theorem as:

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(B) \cdot \mathbb{P}(A|B)}{\mathbb{P}(B) \cdot \mathbb{P}(A|B) + \mathbb{P}(\bar{B}) \cdot \mathbb{P}(A|\bar{B})}$$

95% of the time!

Example: Drug test

A manufacturer claims that its drug test will **detect steroid use**. What the company does not tell you is that 15% of all steroid-free individuals also test positive (the "false positive rate"). Suppose 10% of the Tour de France bike racers use steroids and your favorite cyclist just tested positive. What's the probability that they used steroids?

A: tests positive

B: actually uses steroids

$\mathbb{P}(A \mid B) = 0.95$

$\mathbb{P}(A \mid \bar{B}) = 0.15$

$\mathbb{P}(B) = 0.1$

$\boxed{\text{Want } \mathbb{P}(B \mid A)}$ goal!

A: tests positive

B: actually uses steroids

$\mathbb{P}(A \mid B) = 0.95$

$\mathbb{P}(A \mid \bar{B}) = 0.15$

$\mathbb{P}(B) = 0.1$

$\boxed{\text{Want } \mathbb{P}(B \mid A)}$ goal!

$$\mathbb{P}(B \mid A) \underset{\text{Bayes' Theorem}}{=} \frac{\mathbb{P}(B)\,\mathbb{P}(A \mid B)}{\mathbb{P}(A)} \underset{\text{Total Probability}}{=} \frac{\mathbb{P}(B)\,\mathbb{P}(A \mid B)}{\mathbb{P}(A \cap B) + \mathbb{P}(A \cap \bar{B})}$$

$$= \frac{\mathbb{P}(B)\,\mathbb{P}(A \mid B)}{\mathbb{P}(B)\,\mathbb{P}(A \mid B) + \mathbb{P}(\bar{B})\,\mathbb{P}(A \mid \bar{B})}$$

$$= \frac{\underbrace{0.1 \cdot 0.95}}{\underbrace{0.1 \cdot 0.95} + (1 - 0.1) \cdot 0.15} \approx 0.41 < \frac{1}{2}$$

tip: numerator should appear in the denominator!

Example: Taste test

- Your friend claims to be able to correctly guess what restaurant a burger came from, after just one bite.
- The probability that she correctly identifies an In-n-Out Burger is 0.55, a Shake Shack burger is 0.75, and a Five Guys burger is 0.6.
- You buy 5 In-n-Out burgers, 4 Shake Shack burgers, and 1 Five Guys burger, choose one of the burgers randomly, and give it to her.
- **Question**: Given that she guessed it correctly, what's the probability she ate a Shake Shack burger?

I : In-n-Out, S : shake shack, F : Five Guys, C : correct guess

$\mathbb{P}(I) = \frac{5}{10} = 0.5, \quad \mathbb{P}(S) = 0.4, \quad \mathbb{P}(F) = 0.1$

$\mathbb{P}(C \mid I) = 0.55, \quad \mathbb{P}(C \mid S) = 0.75, \quad \mathbb{P}(C \mid F) = 0.6$

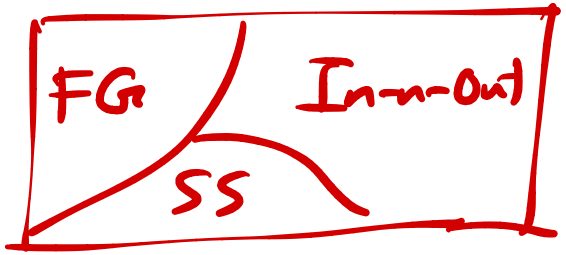
Want $\boxed{\mathbb{P}(S \mid C)}$

I : In-n-Out, S : Shake Shack, F: Five Guys, C : correct guess

$\mathbb{P}(I) = \frac{5}{10} = 0.5$, $\mathbb{P}(S) = 0.4$, $\mathbb{P}(F) = 0.1$

$\mathbb{P}(C|I) = 0.55$, $\mathbb{P}(C|S) = 0.75$, $\mathbb{P}(C|F) = 0.6$

Want $\boxed{\mathbb{P}(S|C)}$



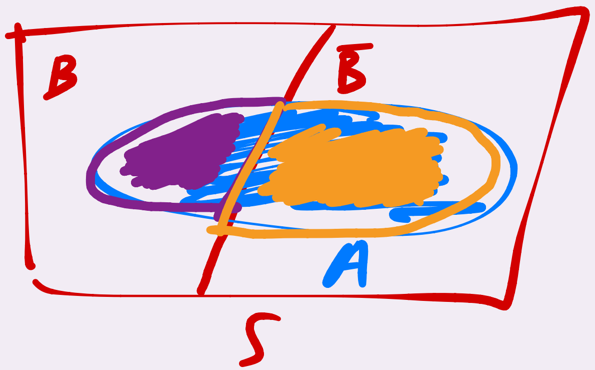
$$\mathbb{P}(S|C) \underset{\text{Bayes Theorem}}{=} \frac{\mathbb{P}(S)\cdot\mathbb{P}(C|S)}{\mathbb{P}(C)} \underset{\text{Total Probability}}{=} \frac{\mathbb{P}(S)\cdot\mathbb{P}(C|S)}{\mathbb{P}(C\cap I)+\mathbb{P}(C\cap S)+\mathbb{P}(C\cap F)}$$

$$= \frac{\mathbb{P}(S)\cdot\mathbb{P}(C|S)}{\mathbb{P}(I)\mathbb{P}(C|I)+\mathbb{P}(S)\mathbb{P}(C|S)+\mathbb{P}(F)\mathbb{P}(C|F)}$$

$$= \frac{0.4\cdot 0.75}{\underbrace{0.5}\cdot 0.55+\underbrace{0.4}\cdot 0.75+\underbrace{0.1}\cdot 0.6} \approx \boxed{0.47}$$

add to 1!

Question 🤔



Answer at q.dsc40a.com

Consider any two events A and B. Choose the expression that's equivalent to:

$$\mathbb{P}(B|A) + \mathbb{P}(\bar{B}|A)$$

- A. $\mathbb{P}(A)$
- B. $1 - \mathbb{P}(B)$
- C. $\mathbb{P}(B)$
- D. $\mathbb{P}(\bar{B})$
- E. 1

$$= \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)} + \frac{\mathbb{P}(\bar{B} \cap A)}{\mathbb{P}(A)}$$

$$= \frac{\boxed{\mathbb{P}(B \cap A)} + \boxed{\mathbb{P}(\bar{B} \cap A)}}{\mathbb{P}(A)}$$

$$= \frac{\mathbb{P}(A)}{\mathbb{P}(A)} = \boxed{1}$$

Example: Prosecutor's fallacy

A bank was robbed yesterday by one person. Consider the following facts about the crime:

- The person who robbed the bank wore Nikes.
- Of the 10,000 other people who came to the bank yesterday, only 10 of them wore Nikes.

11 wore Nikes!

The prosecutor finds the prime suspect, and states that "given this evidence, the chance that the prime suspect was not at the crime scene is $\boxed{\text{1 in 1,000".}}$

they lied! they reversed the conditional probability :(

1. What is wrong with this statement?
2. Find the probability that the prime suspect is guilty given only the evidence in the exercise.

Example from http://stat88.org/textbook/content/Chapter_02/06_Exercises.html

	guilty	innocent
Nikes	1	10
no Nikes	0	9990

11 total Nike wearers!

10,001 total people!

$$\frac{1}{1000} = \frac{10/10001}{10000/10001} = \frac{\mathbb{P}(\text{innocent} \cap \text{Nike})}{\mathbb{P}(\text{innocent})} = \underbrace{\mathbb{P}(\text{Nike} \mid \text{innocent})}$$

they lied! they said they gave us $\mathbb{P}(\text{innocent} \mid \text{Nike})$

$$\mathbb{P}(\text{innocent} \mid \text{Nike}) = \frac{\mathbb{P}(\text{innocent} \cap \text{Nike})}{\mathbb{P}(\text{Nike})} = \frac{10/10001}{11/10001} = \boxed{\frac{10}{11}} \approx 91\%$$

Independence

the remaining slides
will be covered
in Tuesday's lecture

Updating probabilities

- Bayes' Theorem describes how to update the probability of one event, given that another event has occurred.

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

  - $\mathbb{P}(B)$ can be thought of as the "prior" probability of B occurring, before knowing anything about A.
  - $\mathbb{P}(B|A)$ is sometimes called the "posterior" probability of B occurring, given that A occurred.

- What if knowing that A occurred doesn't change the probability that B occurs? In other words, what if:

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

Independent events

- A and B are **independent events** if one event occurring does not affect the chance of the other event occurring.

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

- Otherwise, A and B are **dependent events**.
- Using Bayes' theorem, we can show that if one of the above statements is true, then so is the other.

Independent events

- **Equivalent definition**: A and B are independent events if:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

- To check if A and B are independent, use whichever is easiest:
  - $\mathbb{P}(B|A) = \mathbb{P}(B)$.
  - $\mathbb{P}(A|B) = \mathbb{P}(A)$.
  - $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$.

Mutual exclusivity and independence

Suppose A and B are two events with non-zero probabilities. Is it possible for A and B to be both mutually exclusive and independent?

- A. Yes.
- B. No.

Example: Venn diagrams

For three events A, B, and C, we know that:

- A and C are independent,
- B and C are independent,
- A and B are mutually exclusive,
- $\mathbb{P}(A \cup C) = \frac{2}{3}$, $\mathbb{P}(B \cup C) = \frac{3}{4}$, $\mathbb{P}(A \cup B \cup C) = \frac{11}{12}$.

Find $\mathbb{P}(A)$, $\mathbb{P}(B)$, and $\mathbb{P}(C)$.

Summary

Summary

- A set of events $E_1, E_2, \ldots, E_k$ is a **partition** of S if each outcome in S is in exactly one E_i.
- The Law of Total Probability states that if A is an event and $E_1, E_2, \ldots, E_k$ is a partition of S, then:

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(E_1) \cdot \mathbb{P}(A|E_1) + \mathbb{P}(E_2) \cdot \mathbb{P}(A|E_2) + \ldots + \mathbb{P}(E_k) \cdot \mathbb{P}(A|E_k) \\ &= \sum_{i=1}^{k} \mathbb{P}(E_i) \cdot \mathbb{P}(A|E_i)\end{aligned}$$

- Bayes' Theorem states that $\mathbb{P}(B|A) = \dfrac{\mathbb{P}(B) \cdot \mathbb{P}(A|B)}{\mathbb{P}(A)}$.
- We often re-write the denominator $\mathbb{P}(A)$ in Bayes' Theorem using the Law of Total Probability.

Summary

- Two events A and B are **independent** when knowledge of one event does not change the probability of the other event.
- There are there equivalent definitions of independence:
  - $\mathbb{P}(B|A) = \mathbb{P}(B)$
  - $\mathbb{P}(A|B) = \mathbb{P}(A)$
  - $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$