Lecture 20 – Law of Total Probability and Bayes' Theorem

DSC 40A, Winter 2024

Announcements

- Homework 6 is due **Tonight at 11:59pm**.
  - Problem 3 needs knowledge from today’s lecture

Agenda

- Partitions and the Law of Total Probability.
- Bayes’ Theorem.
- Frequentist vs. Bayesian

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?

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

$P(\text{trolley} \cap \text{late})$

↑ probability

	Late	Not Late
Trolley	0.06	0.24
Bike	0.03	0.07
Drive	0.36	0.24
	0.45	

Discussion Question

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

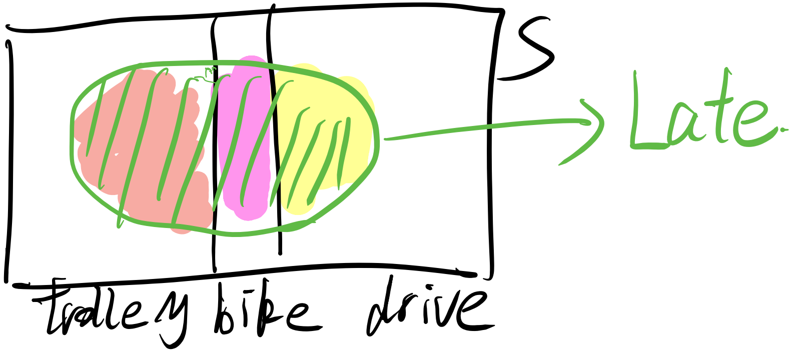
a) 0.24
b) 0.30
[✓ c) 0.45]
d) 0.50
e) None of the above

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

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



$$\frac{.06}{.06+.24} = \frac{.06}{.3} = 1/5$$

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

Discussion Question

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

a) 0.06
[✓ b) 0.2]
c) 0.25
d) 0.45
e) None of the above

$$P(\text{late} \mid \text{trolley}) = \frac{P(\text{late} \cap \text{trolley})}{P(\text{trolley})}$$

$$= \frac{P(\text{late} \cap \text{trolley})}{P(\text{trolley} \cap \text{late}) + P(\text{trolley} \cap \text{not late})}$$

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

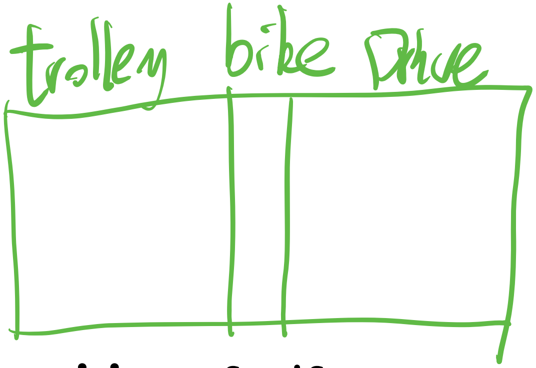
$$P(A \cap B) = P(A) \cdot P(B|A)$$

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

- Another way of expressing the same thing:

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

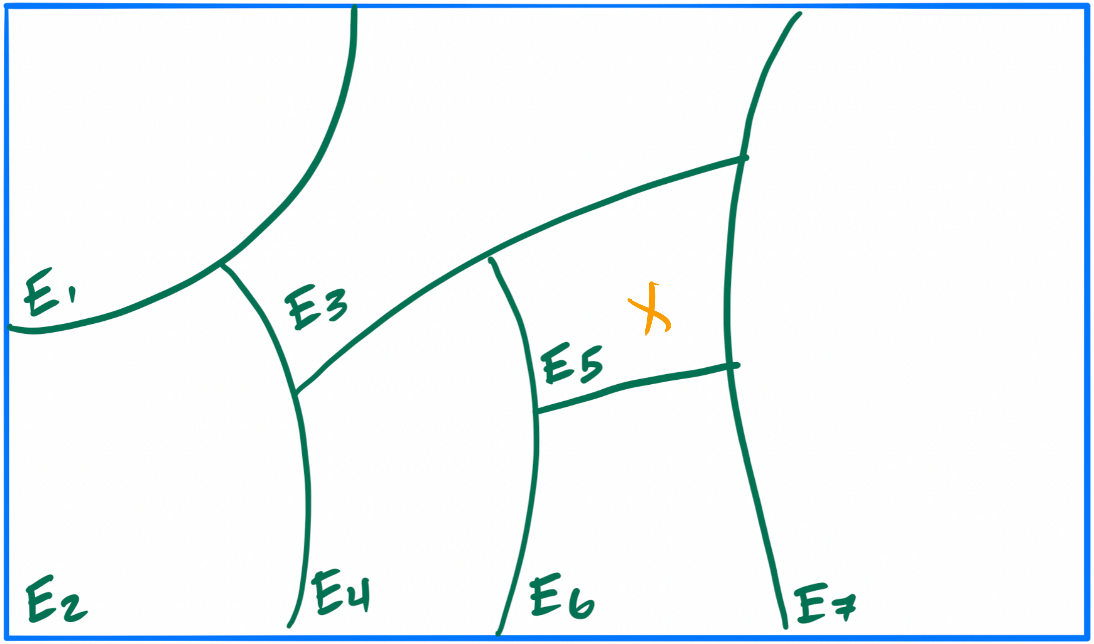
Partitions



- A set of events $E_1, E_2, \dots, E_k$ is a **partition** of S if
  - $P(E_i \cap E_j) = 0$ for all pairs $i \neq j$.

  - $P(E_1 \cup E_2 \cup \dots \cup E_k) = 1$. 1 (or $E_1 \cup E_2 \cup E_3 \cup \dots E_k = S$)
    - Equivalently, $P(E_1) + P(E_2) + \dots + P(E_k) = 1$.
- In other words, $E_1, E_2, \dots, E_k$ is a partition of S if every outcome s in S is in **exactly** one event E_i.

Partitions, visualized



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.
- 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.

Example partitions

- In getting to school, the events Trolley, Bike, and Drive.
- In getting to school, the events $\overset{A}{\text{Late}}$ and Not $\overset{\bar{A}}{\text{Late}}$.
- In selecting an undergraduate student at random, the events Freshman, Sophomore, Junior, and Senior.
- 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 $\bar{A}$.

The Law of Total Probability

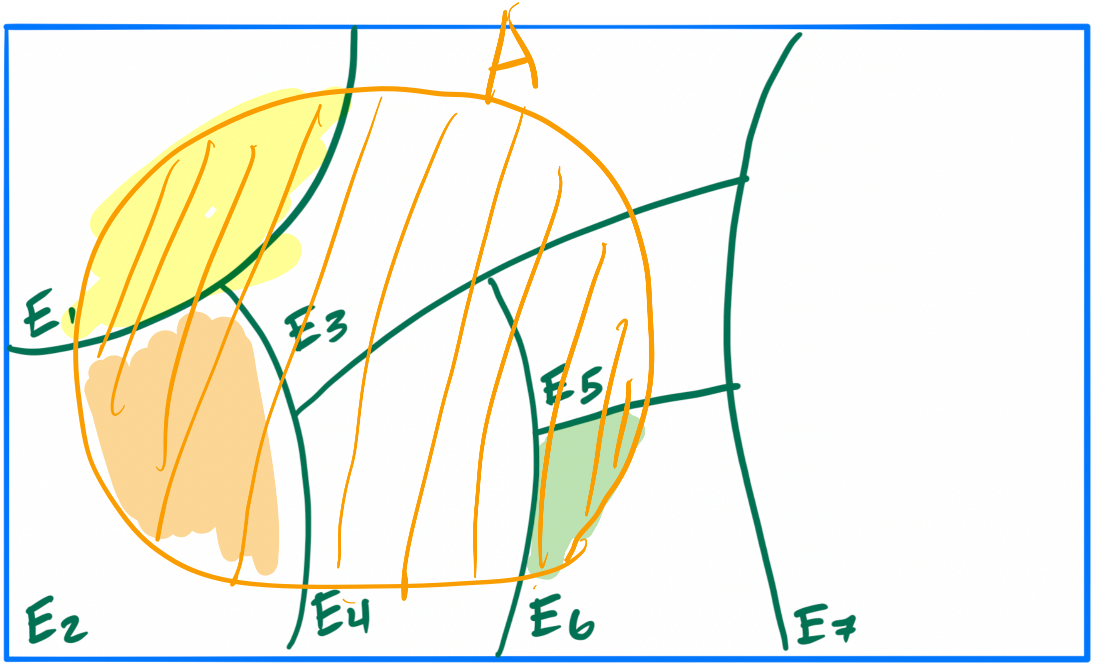
- If A is an event and $E_1, E_2, \ldots, E_k$ is a **partition** of S, then

$$P(A) = P(A \cap E_1) + P(A \cap E_2) + \ldots + P(A \cap E_k)$$

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

$$P(\text{late}) = P(\text{late} \cap \text{trolley}) + P(\text{late} \cap \text{bike}) + P(\text{late} \cap \text{drive})$$

The Law of Total Probability, visualized



$$P(A) = P(A \cap E_1) + P(A \cap E_2) + \ldots + P(A \cap E_6) + \underbrace{P(A \cap E_7)}_{40}$$

The Law of Total Probability

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

Use

$\cap$ "and"

$$P(A) = P(A \cap E_1) + P(A \cap E_2) + \dots + P(A \cap E_k)$$

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

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

use Conditional Probability

$$P(A) = P(E_1) \cdot P(A|E_1) + P(E_2) \cdot P(A|E_2) + \dots + P(E_k) \cdot P(A|E_k)$$

$$= \sum_{i=1}^{k} P(E_i) \cdot P(A|E_i)$$

$\frac{0.06}{.45} = \frac{.12}{.9}$

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

Discussion Question

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

- a) Close to 0.05
- [✓ b) Close to 0.15]
- c) Close to 0.3
- d) Close to 0.4

$$P(\text{trolley} \mid \text{late}) = \frac{P(\text{trolley} \cap \text{late})}{P(\text{late})}$$

$$= \frac{0.06}{P(\text{late} \cap \text{drive}) + P(\text{late} \cap \text{trolley}) + P(\text{late} \cap \text{bike})}$$

Bayes’ Theorem

Example: getting to school

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

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

- Can you still find $P(\text{Trolley}|\text{Late})$?

$$P(\text{trolley} \mid \text{late}) = \frac{P(\text{trolley} \cap \text{late})}{P(\text{late})} = \frac{P(\text{trolley}) \cdot P(\text{late}|\text{trolley})}{P(\text{late})}$$

$$= \frac{0.3 \cdot 0.2}{0.45} = \frac{0.06}{0.45} = \frac{0.12}{0.9}$$

Bayes’ Theorem

- Recall that the multiplication rule states that

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

$$\|$$

- It also states that

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

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

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

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

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

Bayes' Theorem and the Law of Total Probability

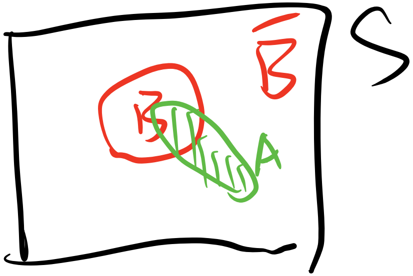
- Bayes' Theorem:

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

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

law of total probability

$$P(A) = P(A\cap B) + P(A\cap \bar{B}) = \underbrace{P(B)\cdot P(A|B) + P(\bar{B})\cdot P(A|\bar{B})}_{\text{sub this into the denom. of Baye's Theorem}}$$



Bayes' Theorem and the Law of Total Probability

- Bayes' Theorem:

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

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

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

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

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

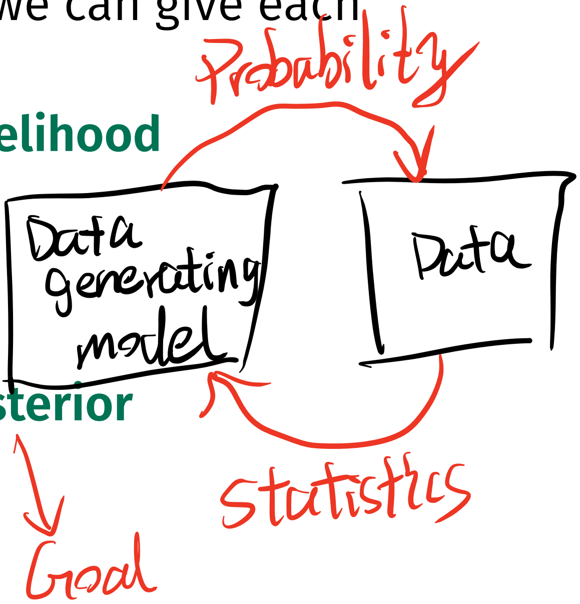
Frequentist vs. Bayesian

Bayes' Theorem Reformulated

- Bayes' Theorem:

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

- Let's assume B is a hypothesis (i.e. some prediction by a model) and A is observed data, then we can give each term a name:
  - $P(A|B) = P(Data|Hypothesis)$ **Likelihood**
  - $P(B) = P(Hypothesis)$ **Prior**
  - $P(B|A) = P(Hypothesis|Data)$ **Posterior**
  - $P(A) = P(Data)$ **Bayes Evidence**



Example: How Long is One Day?

One day is defined to be 24 hours:

- When you measure the length of one day with the super old clock in your grandfather garagae, you get 23.5±2.2 hrs
- When you measure the length of one day with a mechnical watch, you get 23.99±0.05 hrs
- When you measure the length of one day with your apple watch, you get 23.999±0.002 hrs
- When you measure the length of one day the best atomic clock in a lab, you get 23.999999999±0.00000007 hrs

How long is one day?

Frequentist vs. Bayesian

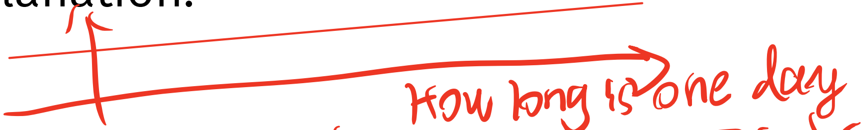
Bayesian assigns probability to hypothesis

Frequentist don't.

Frequentist's explanation:

- A day is always 24 hours, that does not change, but my measurement of "how long is one day" comes with uncertainties.

Bayesian explanation:



$$\text{Posterior} = \frac{\text{Prior} \cdot \text{Likelihood}}{\text{Evidence}}$$

P(How long is one day | observed measurement of 1 day)

P(observed lenth of a day | my hypothesis about How long is one day)

P(observed data)

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

$P(A|B)$

Example: drug test

A manufacturer claims that its drug test will **detect steroid use 95% of the time**. 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?

$$P(B|A) = \frac{P(B)\cdot P(A|B)}{P(A)} \qquad P\left(\begin{matrix}\text{Use}\\ \text{Steroid}\end{matrix} \,\middle|\, \begin{matrix}\text{Positive}\\ \text{Test}\end{matrix}\right)$$

$$\uparrow \qquad \uparrow$$

$$B \qquad A$$

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

law of total probability

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

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

$$= \frac{0.1\cdot 0.95}{0.1 * 0.95 + 0.9\cdot 0.15}$$

$$\simeq 0.41 < \frac{1}{2}$$

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?

Discussion Question

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

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

a) $P(A)$
b) $1 - P(B)$
c) $P(B)$
d) $P(\bar{B})$
e) 1

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

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

- Bayes' Theorem states that

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

- We often re-write the denominator $P(A)$ in Bayes' Theorem using the Law of Total Probability.