

# DSC 40A

*Theoretical Foundations of Data Science I*

# Announcements

- Homework 6 due Monday
- Homework 7 will be released 11/27 and due 12/6.

# Question

Answer at [q.dsc40a.com](http://q.dsc40a.com)

Remember, you can always ask questions at  
[q.dsc40a.com](http://q.dsc40a.com)!

If the direct link doesn't work, click the "Lecture Questions" link in the top right corner of [dsc40a.com](http://dsc40a.com).

# Last Time

- We defined Bayes' Theorem:

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

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

# Today

- What does it mean for one event not to influence the probability of another?
- Independence and conditional independence.

The background features abstract, overlapping green geometric shapes, primarily triangles and polygons, in various shades of green, ranging from light lime to dark forest green. These shapes are concentrated on the right side of the image, with some extending towards the left. The overall effect is a modern, minimalist design.

# Independence

# Updating Probabilities

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

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

# Updating Probabilities

- $P(A)$  is our prior belief that A happens.
- $P(A | B)$  is our updated belief that A happens, now that we know B happens.
- Sometimes knowing that B happens doesn't change anything.
- Sometimes,  $P(B|A) = P(B)$ . Knowing that A occurs doesn't change anything.

## Example

We flip a fair coin twice.

- $P(\text{Second Flip} = \text{Heads}) =$
- $P(\text{Second Flip} = \text{Heads} \mid \text{First Flip} = \text{Heads}) =$



# Independent Events

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

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

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

- Otherwise, A and B are **dependent events**.

If one of the above is true, must the other be true?

- A. yes
- B. not necessarily

# Independent Events

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

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

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

- Using Bayes' Theorem, if one is true, then so is the other.

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

# Independent Events

- A and B are independent events if

$$P(A \text{ and } B) = P(A) * P(B)$$

- This more general definition allows for the probability of A or B to be zero.

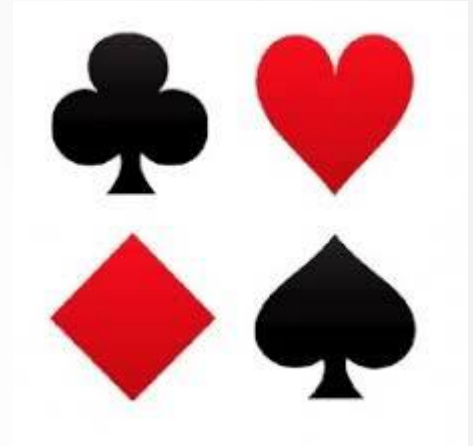
# Example

You draw two cards, one at a time, **with replacement**.

- A is the event that the first card is a heart.
- B is the event that the second card is a club.

You draw two cards, one at a time, **without replacement**.

- A is the event that the first card is a heart.
- B is the event that the second card is a club.



Are A and B independent?

- A. yes in both cases
- B. yes with replacement, no without replacement
- C. no with replacement, yes without replacement
- D. no in both cases

# Example

You draw one card from a deck of 52.

- A is the event that the card is a heart.
- B is the event that the card is a face card (J, Q, K).

Are A and B  
independent?

- A. yes  
B. no

♥: 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K, A

♦: 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K, A

♣: 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K, A

♠: 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K, A

# Assuming Independence

- Sometimes we assume that events are independent to make calculations easier.
- Real-world events are almost never exactly independent, but may be close.
- Example: A is event that a student is a data science major, B is the event that they bike to campus.

# Example

1% of UCSD students are data science majors. 25% of UCSD students eat avocado toast for breakfast. Assuming that being a DSC major and eating avocado toast for breakfast are independent:

- a) What percentage of DSC majors eat avocado toast for breakfast?
  
- a) What percentage of UCSD students are DSC majors who eat avocado toast for breakfast?

# Conditional Independence



# Conditional Independence

- Sometimes, events that are dependent *become* independent upon learning some new information.
- Or independent events can become dependent, given additional information.

# Example

Oops, you lost the King of Clubs! Draw one card from a deck of 51.

- A is the event that the card is a heart.
- B is the event that the card is a face card (J, Q, K).

♥: 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K, A

♦: 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K, A

♣: 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, A

♠: 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K, A

Are A and B independent?

# Example

Oops, you lost the King of Clubs! Draw one card from a deck of 51.

- A is the event that the card is a heart.
- B is the event that the card is a face card (J, Q, K).

♥: 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K, A

♦: 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K, A

♣: 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q,    A

♠: 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K, A

Now suppose you learn that the card is red. Are A and B independent given this new information?

# Conditional Independence

- Recall that A and B are independent if

$$P(A \text{ and } B) = P(A) * P(B)$$

- A and B are conditionally independent given C if

$$P((A \text{ and } B)|C) = P(A|C) * P(B|C)$$

- Given that C occurs, this says that A and B are independent of one another.

# Assuming Conditional Independence

- Sometimes we assume that events are conditionally independent to make calculations easier.
- Real-world events are almost never conditionally independent, but may be close.

# Example

Suppose that 80% of UCSD students use TikTok and 25% of UCSD students eat avocado toast for breakfast. What is the probability that a random UCSD student uses TikTok and eats avocado toast for breakfast, assuming that these events are conditionally independent given that a person is a UCSD student?

# Independence vs. Conditional Independence

- Is it reasonable to assume conditional independence of
  - using TikTok
  - eating avocado toast for breakfastgiven that a person is a UCSD student?
- Is it reasonable to assume independence of these events in general, among all people?

Which assumptions do you think are reasonable?

- A. both
- B. conditional independence only
- C. independence (in general) only
- D. neither

# Independence vs. Conditional Independence

- In general, there is **no relationship** between independence and conditional independence.



# Summary

- Events are independent when knowledge of one event does not change the probability of the other event.
- Events that are not independent *can* be conditionally independent given new information (and the opposite is true).
- **Next time:** Solving the classification problem using Bayes' Theorem and an assumption of conditional independence.