Lecture 15 – Independence

DSC 40A, Fall 2021 @ UC San Diego

Suraj Rampure, with help from **many others**

Announcements

worth 5% !

- Fill out Survey 6 if you haven't already!
- **My office hours today will be both in-person (SDSC 2E) AND remote.**
- Groupwork 7 is due **tonight at 11:59pm**.
- Homework 7 is due **Monday at 11:59pm**.
- Homework 5 grades are out.
- **Great source of practice problems for this week's content:** stat88.org/textbook. → look at resources page!
- **Last plug:** consider signing up for my History of Data Science seminar (DSC 90)! → won't actually be 3 hours

Agenda

- Independence.
- Conditional independence.

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.

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 1 in 1,000".

1. What is wrong with this statement?

2. Find the probability that the prime suspect is guilty given only the evidence in the exercise.

	Guilty	Innocent
Nike	1	10
No Nike	0	9990

$$P(\text{Innocent} \mid \text{Nike}) = \frac{P(\text{Innocent} \cap \text{Nike})}{P(\text{Nike})}$$

$$= \frac{10}{11}$$

$$P(\text{Nike} \mid \text{Innocent}) = \frac{P(\text{Innocent} \cap \text{Nike})}{P(\text{Innocent})} = \frac{10}{10000} = \frac{1}{1000}$$

Independence

Updating probabilities

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

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

ratio

  - $P(B)$ can be thought of as the "prior" probability of B occurring, before knowing anything about A.

  - $P(B|A)$ 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

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

Independent events

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

$$P(B|A) = P(B) \quad \rightleftarrows \quad P(A|B) = P(A)$$

equivalent statements

- 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.

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

always true

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

Independent events

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

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

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

multiplication rule in general:

only if independent!!

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

Mutual exclusivity and independence

Discussion Question

Suppose A and B are two events with non-zero probability.

Is it possible for A and B to be both mutually exclusive and independent?

A) Yes

B) No

C) It depends on A and B

To answer, go to `menti.com` **and enter 5938 8210.**

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

$P(A \cap B) = 0$

$P(A)P(B) = 0$

Example: Venn diagrams

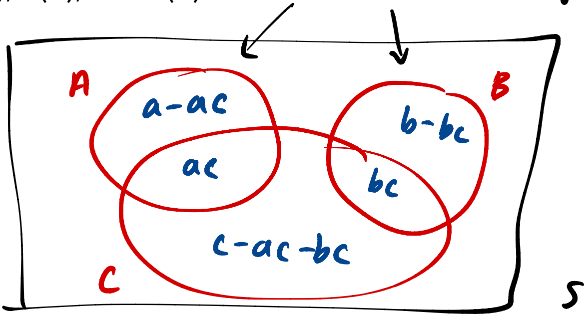
For three events A, B, and C, we know that

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

Find $P(A)$, $P(B)$, and $P(C)$.

For simplicity, let $a = P(A)$, $b = P(B)$, $c = P(C)$.

A and B don't overlap



Venn Diagram not necessary, but helpful.

3 equations:

① $P(A \cup C) = a + c - ac = \frac{2}{3}$

② $P(B \cup C) = b + c - bc = \frac{3}{4}$

③ $P(A \cup B \cup C) = a+b+c - ac - bc = \frac{11}{12}$

$\rightarrow$ ① + ② $\quad a + b + 2c - ac - bc = \frac{2}{3} + \frac{3}{4}$

$\rightarrow$ ① + ② − ③ $\quad c = \frac{2}{3} + \frac{3}{4} - \frac{11}{12} = \frac{8}{12} + \frac{9}{12} - \frac{11}{12} = \boxed{\frac{1}{2}}$

$\Rightarrow$ into ①: $a + \frac{1}{2} - \frac{1}{2}a = \frac{2}{3} \Rightarrow \boxed{a = \frac{1}{3}}$

$\Rightarrow$ into ②: $b + \frac{1}{2} - \frac{1}{2}b = \frac{3}{4} \Rightarrow \boxed{b = \frac{1}{2}}$

$\Rightarrow \therefore \boxed{P(A) = \frac{1}{3},\ P(B) = \frac{1}{2},\ P(C) = \frac{1}{2}}$

Example: cards

♥: 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

- Suppose you draw two cards, one at a time.
  - A is the event that the first card is a heart.
  - B is the event that the second card is a club.

- If you draw the cards **with** replacement, are A and B independent? Yes!

$P(B) = \frac{13}{52}$

- If you draw the cards **without** replacement, are A and B independent?

No!

$P(B|A) = \frac{13}{51} > \frac{13}{52}$

Example: cards

prop of A taken up by $B = \frac{3}{13}$

prop of S taken up by $B = \frac{12}{52} = \frac{3}{13}$

♥: 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

- Suppose 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? Yes!

$P(A) = \frac{13}{52} = \frac{1}{4}$, $P(B) = \frac{12}{52} = \frac{3}{13}$

$P(A \cap B) = \frac{3}{52} = P(A) \cdot P(B) = \frac{1}{4} \cdot \frac{3}{13} = \frac{3}{52}$

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

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

$P(B|A)$

"the proportion of A taken up by B" = "the proportion of S taken up by B"

independent!

Assuming independence

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

Example: breakfast

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

1. What percentage of DSC majors eat avocado toast for breakfast?

$$P(\text{avo toast} \mid DSC) = P(\text{avo toast}) = 25\%$$

2. What percentage of UCSD students are DSC majors who eat avocado toast for breakfast?

$$P(\text{avo toast} \cap DSC) = P(DSC)\,P(\text{avo toast} \mid DSC)$$

$$= P(DSC)\,P(\text{avo toast}) = 1\% \text{ of } 25\%$$

$$= 0.25\%$$

Conditional independence

Conditional independence

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

Example: cards

♥: 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

- Your dog ate the King of Clubs. Suppose you 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).

- Are A and B independent?

$P(A) = \frac{13}{51}$ $P(A \cap B) = \frac{3}{51}$

$P(B) = \frac{11}{51}$ $P(A)P(B) = \frac{13}{51} \cdot \frac{11}{51} \neq \frac{3}{51}$

Example: cards

B

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, A~~

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

- Your dog ate the King of Clubs. Suppose you 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).

event C

- Suppose you learn that the card is red. Are *A* and *B* independent given this new information?

within red cards (i.e. within C):

prop of A taken up by B $= \frac{3}{13} =$ prop of C taken up by B $= \frac{6}{26} = \frac{3}{13}$

Conditional independence

- Recall that A and B are independent if

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

- A and B are **conditionally independent** given C if

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

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

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

Assuming conditional independence

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

Example: Harry Potter and TikTok

Suppose that 50% of UCSD students like Harry Potter and 80% of UCSD students use TikTok. What is the probability that a random UCSD student likes Harry Potter and uses TikTok, assuming that these events are conditionally independent given that a person is a UCSD student?

$$P\left(\left(\text{like HP} \cap \text{use TikTok}\right) \mid \text{UCSD}\right)$$

$$= P\left(\text{like HP} \mid \text{UCSD}\right) \cdot P\left(\text{use TikTok} \mid \text{UCSD}\right)$$

$$= 50\% \text{ of } 80\% = \boxed{40\%}$$

Independence vs. conditional independence

- Is it reasonable to assume conditional independence of
  - liking Harry Potter
  - using TikTok

  given that a person is a UCSD student?
- Is it reasonable to assume independence of these events in general, among all people?

Discussion Question

Which assumptions do you think are reasonable?

A) ~~Both~~
B) Conditional independence only
C) ~~Independence (in general) only~~
D) Neither

To answer, go to `menti.com` **and enter 5938 8210.**

Independence vs. conditional independence

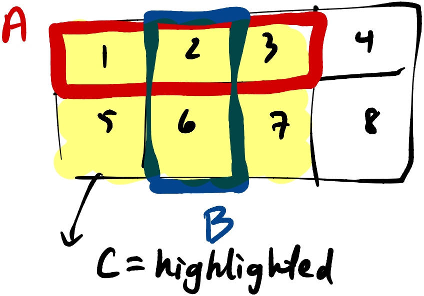
In general, there is **no relationship** between independence and conditional independence. All of these are possibilities, given three events A, B, and C.

- A and B are independent, and are conditionally independent given C.
- A and B are independent, and are conditionally dependent given C.
- A and B are dependent, and are conditionally independent given C.
- A and B are dependent, and are conditionally dependent given C.

Example: constructing events

- Consider a sample space $S = \{1, 2, 3, 4, 5, 6, 7, 8\}$ where all outcomes are equally likely.
- For each scenario, specify events A, B, and C that satisfy the given conditions. (e.g. $A = \{2, 5, 6\}$)
- Choose events that are neither impossible nor certain, i.e. $0 < P(A), P(B), P(C) < 1$.

Scenario 1: A and B **are not** independent. A and B **are** conditionally independent given C.



$P(A) = \frac{3}{8}$, $P(B) = \frac{1}{4}$,

$P(A \cap B) = \frac{1}{8}$, $P(A)P(B) = \frac{3}{32} \neq \frac{1}{8}$,

$\therefore$ A, B not ind.

$P(A|C) = \frac{1}{2}$, $P(B|C) = \frac{1}{3}$,

$P((A \cap B)|C) = \frac{1}{6} \cdot \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$ ✓

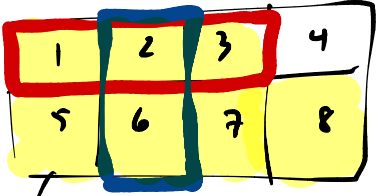
$\therefore$, A and B are cond. ind. given C

Example: constructing events

- Consider a sample space $S = \{1, 2, 3, 4, 5, 6, 7, 8\}$ where all outcomes are equally likely.
- For each scenario, specify events A, B, and C that satisfy the given conditions. (e.g. $A = \{2, 5, 6\}$)
- Choose events that are neither impossible nor certain, i.e. $0 < P(A), P(B), P(C) < 1$.

Scenario 2: A and B **are not** independent. A and B **are not** conditionally independent given C.

A



B

C = highlighted

A and B are same as last slide: not independent.

$P(A|C) = \frac{3}{7}$, $P(B|C) = \frac{2}{7}$,

$P((A \cap B)|C) = \frac{1}{7}$

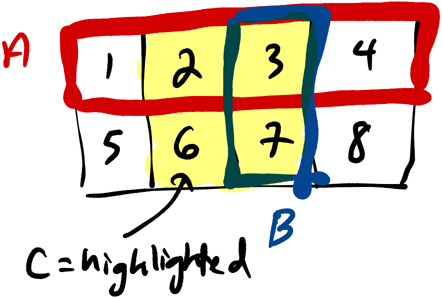
$\Rightarrow \frac{3}{7} \cdot \frac{2}{7} = \frac{6}{49} \neq \frac{1}{7}$,

$\therefore$ A, B not cond ind given C

Example: constructing events

- Consider a sample space $S = \{1, 2, 3, 4, 5, 6\}$ $7, 8$ where all outcomes are equally likely.
- For each scenario, specify events A, B, and C that satisfy the given conditions. (e.g. $A = \{2, 5, 6\}$)
- Choose events that are neither impossible nor certain, i.e. $0 < P(A), P(B), P(C) < 1$. $A = \{1, 2, 3, 4\}$ $B = \{3, 7\}$ $C = \{2, 3, 6, 7\}$

Scenario 3: A and B **are** independent. A and B **are** conditionally independent given C.



$P(A) = \frac{1}{2}$ $P(B) = \frac{1}{4}$

$P(A \cap B) = \frac{1}{8}$ ✓

within C:

$P(A|C) = \frac{1}{2}$, $P(B|C) = \frac{1}{2}$

$P((A \cap B)|C) = \frac{1}{4}$ ✓

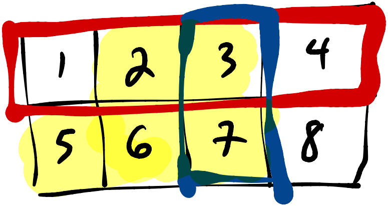
Example: constructing events

7, 8

- Consider a sample space $S = \{1, 2, 3, 4, 5, 6\}$ where all outcomes are equally likely.
- For each scenario, specify events A, B, and C that satisfy the given conditions. (e.g. $A = \{2, 5, 6\}$)
- Choose events that are neither impossible nor certain, i.e. $0 < P(A), P(B), P(C) < 1$.

Scenario 4: A and B **are** independent. A and B **are not** conditionally independent given C.

A, B ind from last slide



$P(A|C) = \frac{2}{5}$

$P(B|C) = \frac{2}{5}$

$P((A \cap B)|C) = \frac{1}{5}$

$\leftarrow \frac{2}{5} \cdot \frac{2}{5} = \frac{4}{25} \neq \frac{1}{5}$

$\therefore$ A and B not cond. ind., given C.

Summary

Summary

- Two events A and B are independent when knowledge of one event does not change the probability of the other event.
  - Equivalent conditions: $P(B|A) = P(B)$, $P(A|B) = P(A)$, $P(A \cap B) = P(A) \cdot P(B)$.

- Two events A and B are conditionally independent if they are independent given knowledge of a third event, C.
  - Condition: $P((A \cap B)|C) = P(A|C) \cdot P(B|C)$.

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

- **Next time:** Using Bayes' theorem and conditional independence to solve the **classification problem** in machine learning.