CSE 312

Foundations of Computing II

Lecture 6: More Conditional Probability

PAUL G. ALLEN SCHOOL
OF COMPUTER SCIENCE & ENGINEERING

Aleks Jovcic

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

Agenda

- Review: Conditional Probability, Bayes
- Law of Total Probability (w/ Bayes)
- Chain Rule
- Independence
- Conditional Independence
- Assumptions and Correlation
- Bonus: Monty Hall Problem

Last Class:

- Conditional Probability
- Bayes Theorem

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

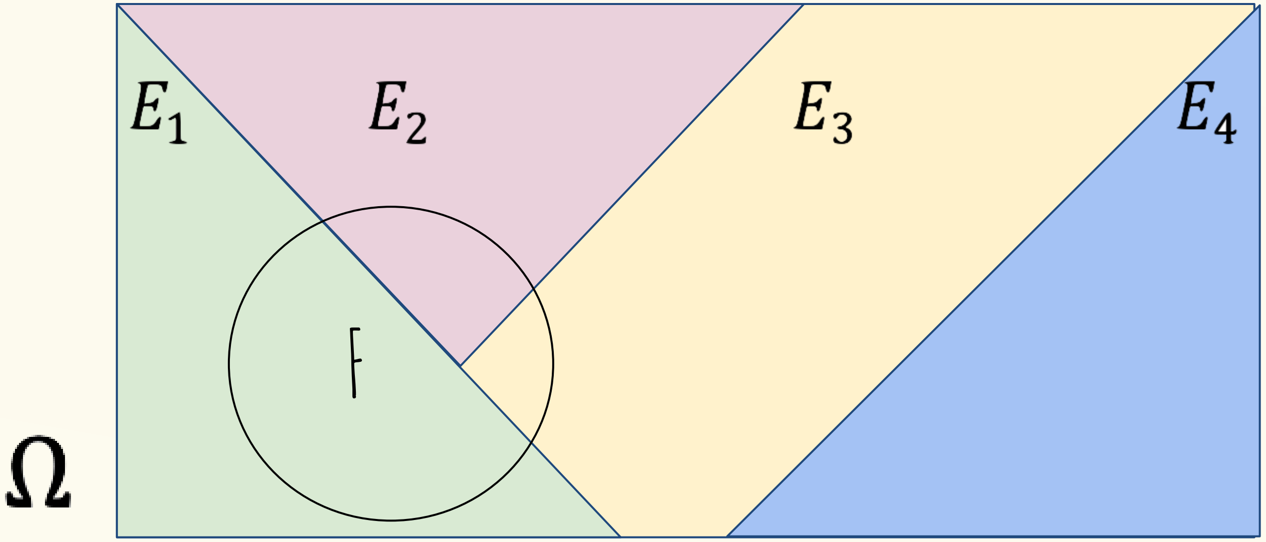
$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}$$

Agenda

- Review: Conditional Probability, Bayes
- **Law of Total Probability (w/ Bayes)**
- Chain Rule
- Independence
- Conditional Independence
- Assumptions and Correlation
- Bonus: Monty Hall Problem

Law of Total Probability (Idea)

If we know $E_1, E_2, \ldots, E_n$ 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

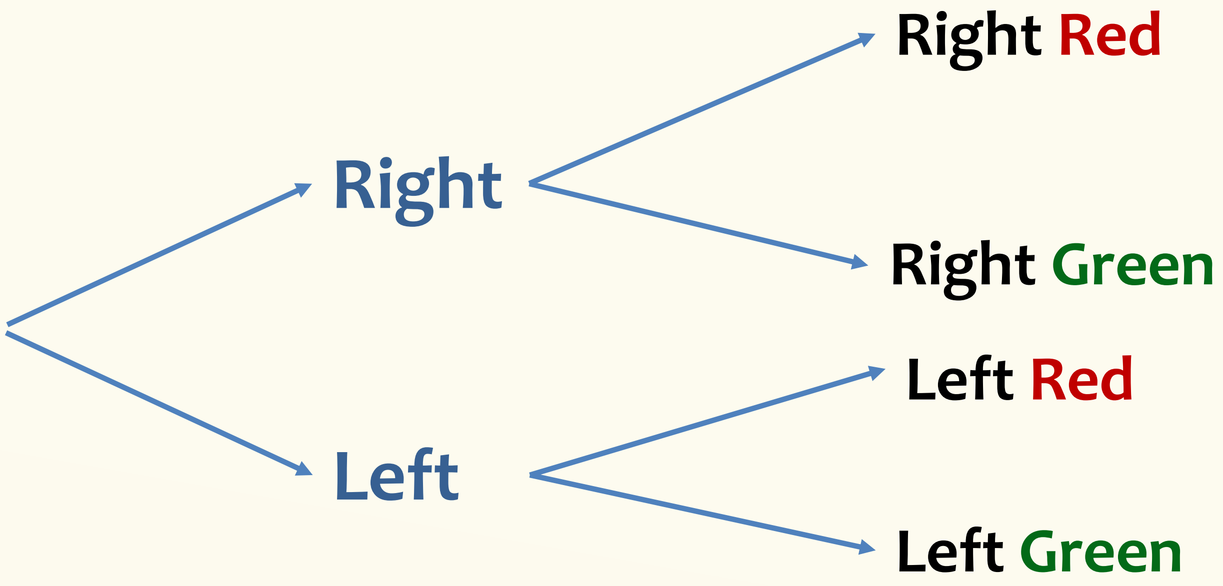
Alice has two pockets:

- **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 $\mathbb{P}(\mathbf{R})$?

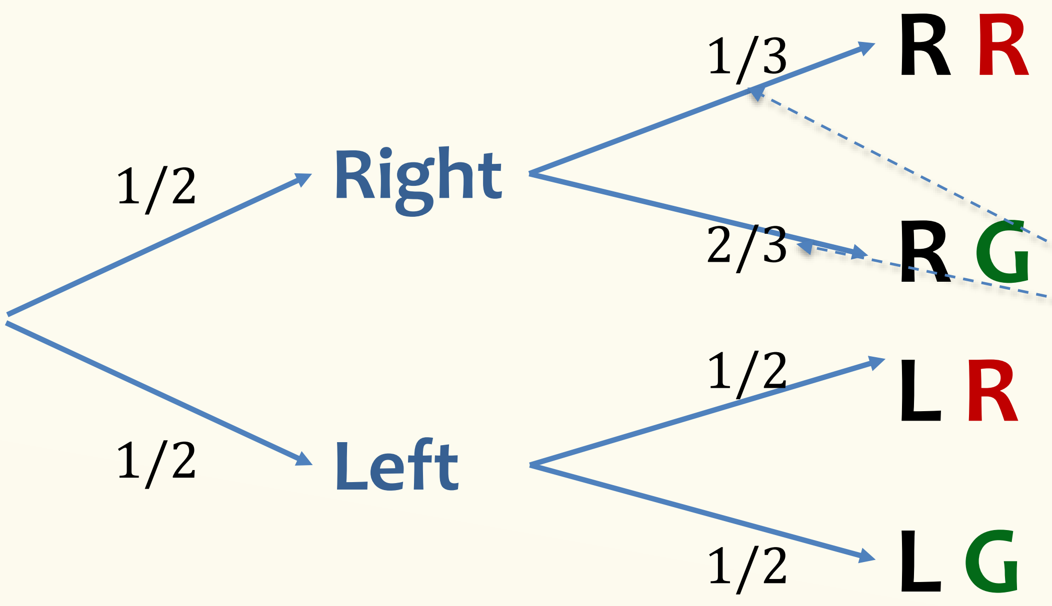
Sequential Process – Non-Uniform Case



- **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



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

1/3 = $\mathcal{P}(R \mid \mathrm{R})$ and 2/3 = $\mathcal{P}(G \mid \mathrm{R})$

$$\mathbb{P}(\mathbf{R}) = \mathbb{P}(\mathbf{R} \cap \mathbf{Left}) + \mathbb{P}(\mathbf{R} \cap \mathbf{Right})$$

(Law of total probability)

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

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

Bayes Theorem with Law of Total Probability

Bayes Theorem with LTP: Let $E_1, E_2, \ldots, E_n$ 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

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

Example – Zika Testing

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.

This example and following slides are from Lisa Yan (Stanford).

Example – Zika Testing

Suppose we know the following Zika stats

- A test is 98% effective at detecting Zika ("true positive")
- However, the test yields 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 yields 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).

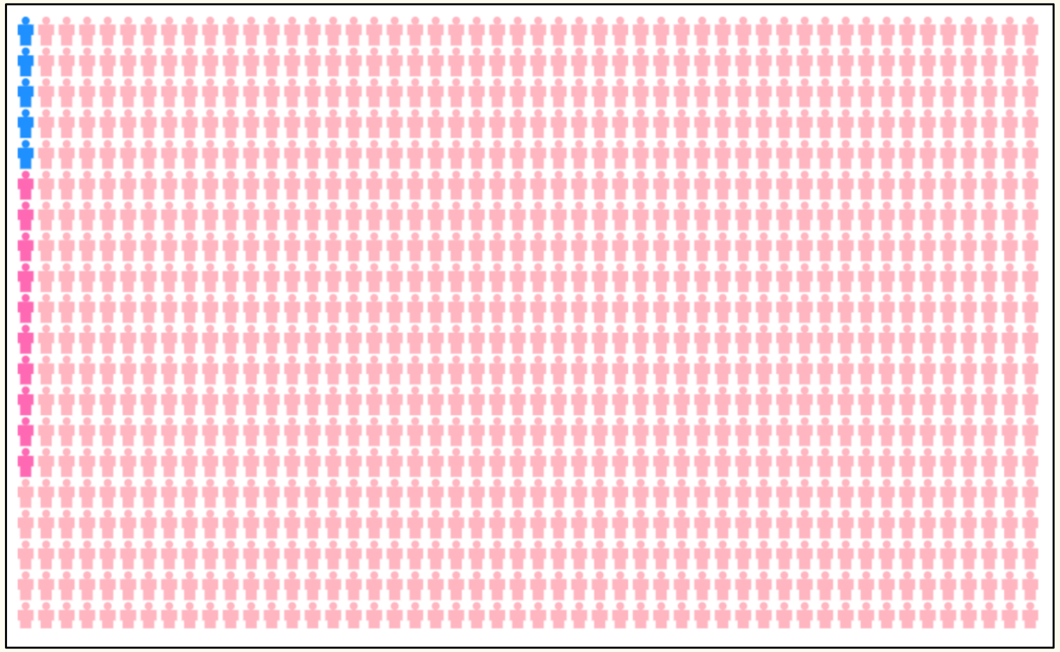
Example – Zika Testing

Have zika blue, don't pink

Suppose we know the following Zika stats

- A test is 98% effective at detecting Zika ("true positive") 100%
- However, the test may yield a "false positive" 1% of the time 10/995 = approximately 1%
- 0.5% of the US population has Zika. 5 people 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**

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

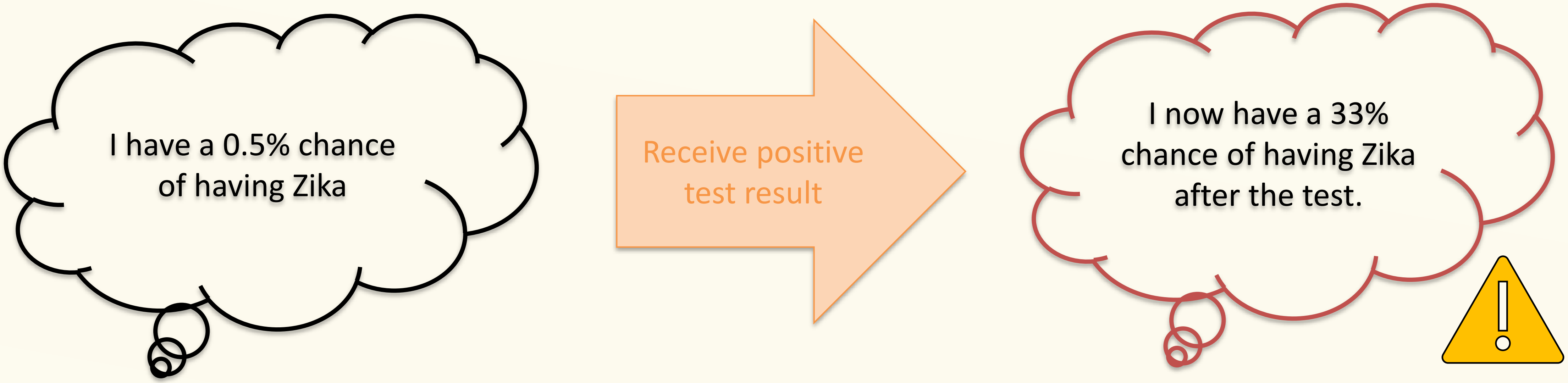
Demo

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 $\mathcal{A}$ follows the same properties as (unconditional) probability.

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

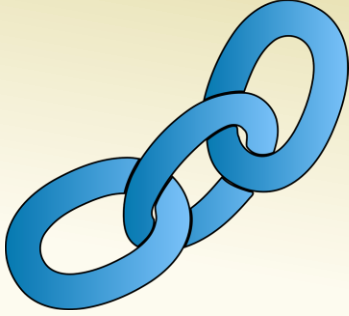
Formally. $(\Omega, \mathbb{P})$ is a probability space + $\mathbb{P}(\mathcal{A}) > 0$

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

Agenda

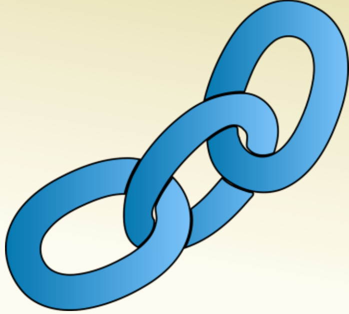
- Review: Conditional Probability, Bayes
- Law of Total Probability (w/ Bayes)
- Chain Rule
- Independence
- Conditional Independence
- Assumptions and Correlation
- Bonus: Monty Hall Problem

Chain Rule



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

Chain Rule



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

Theorem. (Chain Rule) For events $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$,

$$\mathbb{P}(\mathcal{A}_1 \cap \cdots \cap \mathcal{A}_n) = \mathbb{P}(\mathcal{A}_1) \cdot \mathbb{P}(\mathcal{A}_2|\mathcal{A}_1) \cdot \mathbb{P}(\mathcal{A}_3|\mathcal{A}_1 \cap \mathcal{A}_2)$$

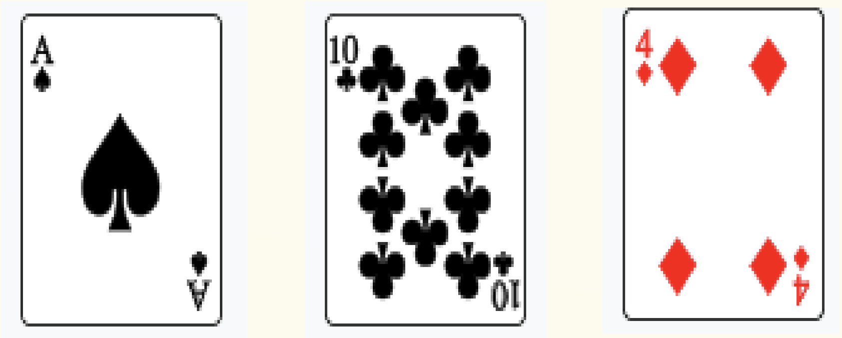
$$\cdots \mathbb{P}(\mathcal{A}_n|\mathcal{A}_1 \cap \mathcal{A}_2 \cap \cdots \cap \mathcal{A}_{n-1})$$

An easy way to remember: We have n tasks and we can do them sequentially, conditioning on the outcome of previous tasks

Chain Rule Example

Have a Standard 52-Card Deck. Shuffle It, and draw the top 3 cards **in order**. (uniform probability space).

What is P (



) = P(**A** ∩ **B** ∩ **C**)?

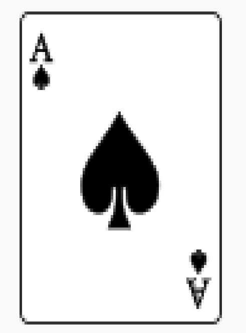
A: Ace of Spades First

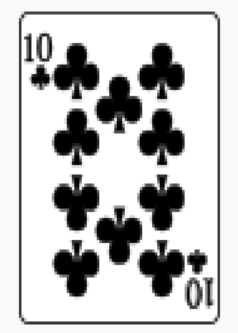
B: 10 of Clubs Second

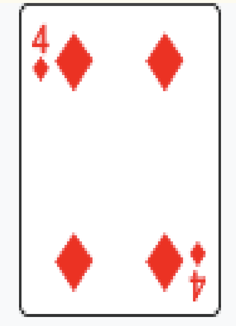
C: 4 of Diamonds Third

Chain Rule Example

Have a Standard 52-Card Deck. Shuffle It, and draw the top 3 cards in order. (uniform probability space).

What is P (





) = P(**A** ∩ **B** ∩ **C**)?

A: Ace of Spades First

B: 10 of Clubs Second

C: 4 of Diamonds Third

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

$$\frac{1}{52} \cdot \frac{1}{51} \cdot \frac{1}{50}$$

Agenda

- Review: Conditional Probability, Bayes
- Law of Total Probability (w/ Bayes)
- Chain Rule
- Independence
- Conditional Independence
- Assumptions and Correlation
- Bonus: Monty Hall Problem

Independence

Definition. If two events $\mathcal{A}$ and $\mathcal{B}$ are **independent** then

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

Alternatively,

- If $\mathbb{P}(\mathcal{A}) \neq 0$, equivalent to $\mathbb{P}(\mathcal{B}|\mathcal{A}) = \mathbb{P}(\mathcal{B})$
- If $\mathbb{P}(\mathcal{B}) \neq 0$, equivalent to $\mathbb{P}(\mathcal{A}|\mathcal{B}) = \mathbb{P}(\mathcal{A})$

"The probability that $\mathcal{B}$ occurs after observing $\mathcal{A}$" -- Posterior
= "The probability that $\mathcal{B}$ occurs" -- Prior

Example -- Independence

Toss a coin 3 times. Each of 8 outcomes equally likely.

- A = {at most one T} = {HHH, HHT, HTH, THH}
- B = {at most 2 Heads}= $\{HHH\}^c$

Independent?

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

<u>Poll:</u>

A. Yes, independent

B. No

Often probability space $(\Omega, \mathbb{P})$ is **defined** using independence

Events generated independently ➔ *their probabilities satisfy independence*

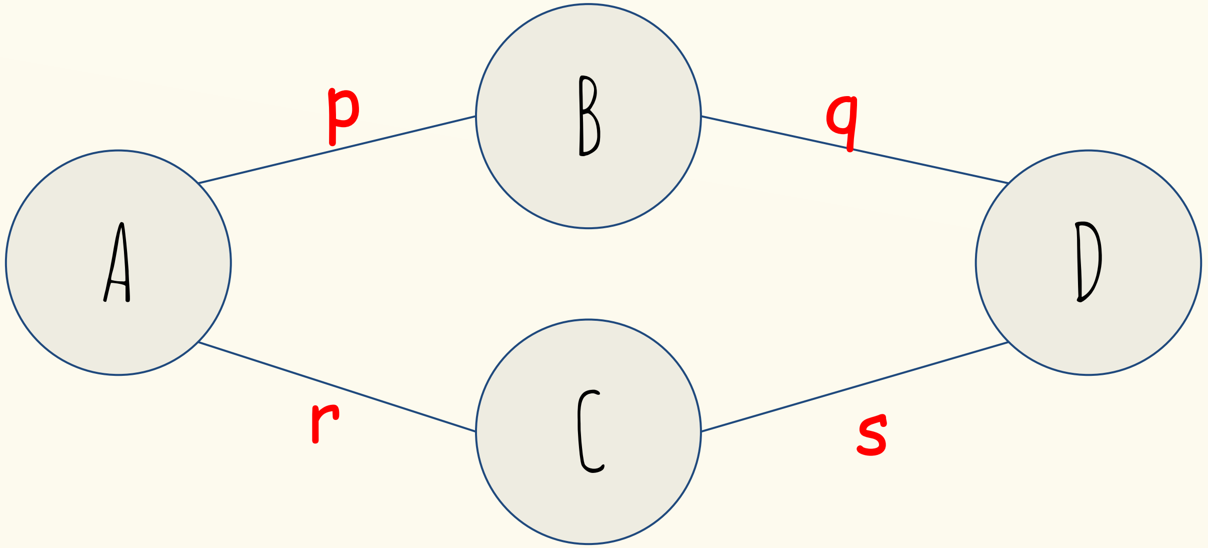
~~⇐~~ *Not necessarily*

This can be counterintuitive!

Example – Network Communication

Each link works with the probability given, **independently**.
What's the probability A and D can communicate?

$\mathbb{P}(AD) = ?$



Example – Network Communication

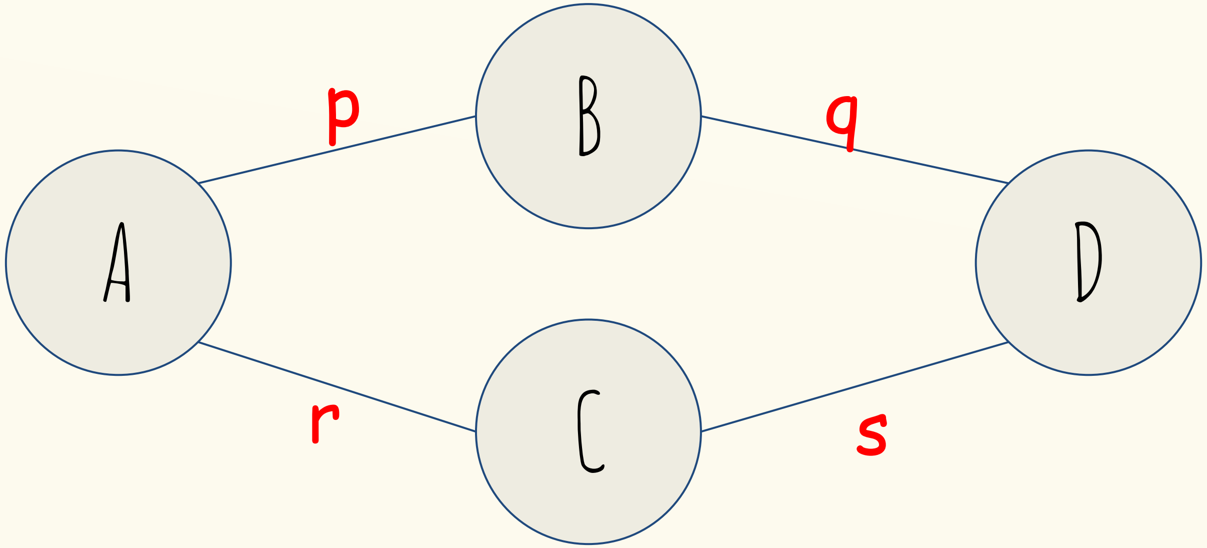
Each link works with the probability given, **independently**.
What's the probability A and D can communicate?

$$\mathbb{P}(AD) = \mathbb{P}(AB \cap BD \text{ or } AC \cap CD)$$

$$= \mathbb{P}(AB \cap BD) + \mathbb{P}\,(AC \cap CD) - \mathbb{P}(AB \cap BD \cap AC \cap CD)$$

$$\mathbb{P}(AB \cap BD) = \mathbb{P}(AB) \cdot \mathbb{P}(BD) = pq$$

$$\mathbb{P}(AC \cap CD) = \mathbb{P}(AC) \cdot \mathbb{P}(CD) = rs$$



$$\mathbb{P}(AB \cap BD \cap AC \cap CD) = \mathbb{P}(AB) \cdot \mathbb{P}(BD) \cdot \mathbb{P}(AC) \cdot \mathbb{P}(CD) = pqrs$$

Example – Biased coin

We have a biased coin comes up Heads with probability 2/3; Each flip is independent of all other flips. Suppose it is tossed 3 times.

$$\mathbb{P}(HHH) =$$

$$\mathbb{P}(TTT) =$$

$$\mathbb{P}(HTT) =$$

Example – Biased coin

We have a biased coin comes up Heads with probability 2/3, independently of other flips. Suppose it is tossed 3 times.

$\mathbb{P}(2\ heads\ in\ 3\ tosses) =$

A) $(2/3)^2\ 1/3$
B) 2/3
C) $3\ (2/3)^2\ 1/3$
D) $(1/3)^2$

Agenda

- Review: Conditional Probability, Bayes
- Law of Total Probability (w/ Bayes)
- Chain Rule
- Independence
- Conditional Independence
- Assumptions and Correlation
- Bonus: Monty Hall Problem

Conditional Independence

Definition. Two events $\mathcal{A}$ and $\mathcal{B}$ are **independent** conditioned on C if

$$\mathbb{P}(C) \neq 0 \text{ and } \mathbb{P}(\mathcal{A} \cap \mathcal{B} \mid C) = \mathbb{P}(\mathcal{A} \mid C) \cdot \mathbb{P}(\mathcal{B} \mid C).$$

Plain Independence. Two events $\mathcal{A}$ and $\mathcal{B}$ are **independent** if

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

Equivalence:

- If $\mathbb{P}(\mathcal{A}) \neq 0$, equivalent to $\mathbb{P}(\mathcal{B}|\mathcal{A}) = \mathbb{P}(B)$
- If $\mathbb{P}(\mathcal{B}) \neq 0$, equivalent to $\mathbb{P}(\mathcal{A}|\mathcal{B}) = \mathbb{P}(\mathcal{A})$

Conditional Independence

Definition. Two events $\mathcal{A}$ and $\mathcal{B}$ are **independent** conditioned on C if

$$\mathbb{P}(C) \neq 0 \text{ and } \mathbb{P}(\mathcal{A} \cap \mathcal{B} \mid C) = \mathbb{P}(\mathcal{A} \mid C) \cdot \mathbb{P}(\mathcal{B} \mid C).$$

Equivalence:

- If $\mathbb{P}(\mathcal{A} \cap C) \neq 0$, equivalent to $\mathbb{P}(\mathcal{B} | \mathcal{A} \cap C) = \mathbb{P}(B \mid C)$
- If $\mathbb{P}(\mathcal{B} \cap C) \neq 0$, equivalent to $\mathbb{P}(\mathcal{A} | \mathcal{B} \cap C) = \mathbb{P}(\mathcal{A} \mid C)$

Plain Independence. Two events $\mathcal{A}$ and $\mathcal{B}$ are **independent** if

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

Equivalence:

- If $\mathbb{P}(\mathcal{A}) \neq 0$, equivalent to $\mathbb{P}(\mathcal{B} | \mathcal{A}) = \mathbb{P}(B)$
- If $\mathbb{P}(\mathcal{B}) \neq 0$, equivalent to $\mathbb{P}(\mathcal{A} | \mathcal{B}) = \mathbb{P}(\mathcal{A})$

Example – More coin tossing

Suppose there is a coin C1 with Pr(Head) = 0.3 and a coin C2 with Pr(Head) = 0.9. We pick one randomly with equal probability and flip that coin twice independently. What is the probability we get all heads?

$$\Pr(HH) = \Pr(HH \mid C1)\Pr(C1) + \Pr(HH \mid C2)\Pr(C2) \qquad \text{LTP}$$

Example – More coin tossing

Suppose there is a coin C1 with Pr(Head) = 0.3 and a coin C2 with Pr(Head) = 0.9. We pick one randomly with equal probability and flip that coin 2 times independently. What is the probability we get all heads?

$$\Pr(HH) \;=\; \Pr(HH \mid C1)\Pr(C1) \;+\; \Pr(HH \mid C2)\Pr(C2)$$ LTP

$$= \Pr(H \mid C2)^2 \Pr(C1) \;+\; \Pr(H \mid C2)^2 \Pr(C2)$$ Conditional Independence

$$= 0.3^2 \cdot 0.5 \;+ 0.9^2 \cdot 0.5 = 0.45$$

$$\Pr(H) \;=\; \Pr(H \mid C1)\Pr(C1) \;+\; \Pr(H \mid C2)\Pr(C2) = 0.6$$

Agenda

- Review: Conditional Probability, Bayes
- Law of Total Probability (w/ Bayes)
- Chain Rule
- Independence
- Conditional Independence
- Assumptions and Correlation
- Bonus: Monty Hall Problem

Correlation

- Pick a person at random
- A : event that the person has lung cancer
- B : event that the person is a heavy smoker

- Fact: $\mathbb{P}(A|B) = 1.17 \cdot \mathbb{P}(A)$

- Conclusions?

Correlation

- Pick a person at random
- A : event that the person has lung cancer
- B : event that the person is a heavy smoker

- Fact: $\mathbb{P}(A|B) = 1.17 \cdot \mathbb{P}(A)$

- Conclusions?
  - Lung cancer increases the the probability of smoking by 17%.
  - Lung cancer causes smoking.

Causality vs. Correlation

- Events A and B are **positively correlated** if

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

- E.g. smoking and lung cancer.

- But A and B being positively correlated does not mean that A causes B or B causes A.

Causality vs. Correlation

- Events A and B are **positively correlated** if

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

- But A and B being positively correlated does not mean that A causes B or B causes A.

Other examples:

- Tesla owners are more likely to be rich. That does not mean poor people should buy a Tesla to get rich.
- People who go to the opera are more likely to have a good career. That does not mean that going to the opera will improve your career.
- Rabbits eat more carrots and do not wear glasses. Are carrots good for eyesight?

Independence as an assumption

- People often assume it **without justification.**
- Example: A sky diver has two chutes

A : event that the main chute doesn't open $\quad \mathbb{P}(A) = 0.02$

B : event that the backup doesn't open $\quad \mathbb{P}(B) = 0.1$

- What is the chance that at least one opens assuming independence?

Independence as an assumption

- People often assume it **without justification.**
- Example: A sky diver has two chutes

  A : event that the main chute doesn't open $\quad \mathbb{P}(A) = 0.02$

  B : event that the backup doesn't open $\quad \mathbb{P}(B) = 0.1$

- What is the chance that at least one opens assuming independence?

- Assuming independence doesn't justify the assumption! Both chutes could fail because of the same rare event e.g., freezing rain.

Agenda

- Review: Conditional Probability, Bayes
- Law of Total Probability (w/ Bayes)
- Chain Rule
- Independence
- Conditional Independence
- Assumptions and Correlation
- Bonus: Monty Hall Problem

Monty Hall Problem

Suppose you're on a game show, and you're given the choice of three doors. Behind one of the doors is a car, behind the other, goats. You pick a door, say number 1, and the host, who knows what's behind the doors, opens another door, say number 3, which has a goat. He says to you, "Do you want to switch to door number 2?" Is it to your advantage to switch your choice of doors?

Assumptions

- The player is equally likely to pick each of the three doors.
- After the player picks a door, the host must open a different door with a goat behind it and offer the player the choice of staying with the original door or switching.
- If the host has a choice of which door to open, then he is equally likely to select each of them.

Should you switch or stay?