CSE 312

Foundations of Computing II

Lecture 7: Chain Rule and Independence

PAUL G. ALLEN SCHOOL
OF COMPUTER SCIENCE & ENGINEERING

Anna R. Karlin

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

Announcements

- No concept check today!
- Section tomorrow is **important** with new content that you will need on pset 3, problem 7. Bring your laptops.
- I have to be out of town (and will be largely unreachable) Thursday-Saturday – Aleks will give Friday's lecture!
- Quiz 1 out later next week. Will cover material from the first two problem sets.

Friday 10/8: Bayes Theorem with Law of Total Probability

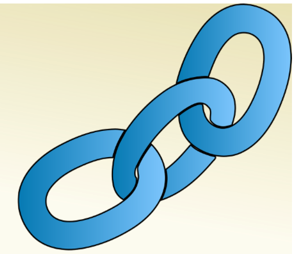
Bayes Theorem with LTP: Let $E_1, E_2, \ldots, E_n$ be a partition of the sample space, and F and event. Then,

$$P(E_1|F) = \frac{P(F|E_1)P(E_1)}{P(F)} = \frac{P(F|E_1)P(E_1)}{\sum_{i=1}^{n} P(F|E_i)P(E_i)}$$

Simple Partition: In particular, if E is an event with non-zero probability, then

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

Monday 10/10: Chain Rule



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

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

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

Monday: Independence

Definition. Two events $\mathcal{A}$ and $\mathcal{B}$ are (statistically) **independent** if

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

Alternatively,

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

Agenda

- A Sequential Process Defined Using Independence
- Independence As An Assumption
- Sometimes Independence Occurs for Nonobvious Reasons
- Conditional Independence
- Correlation vs Causation
- Information Cascades

Example – Throwing A Die Repeatedly

Alice and Bob are playing the following game.

A 6-sided die is thrown, and each time it's thrown, regardless of the history, it is equally likely to show any of the six numbers

If it shows 1, 2 → Alice wins.
If it shows 3 → Bob wins.
Otherwise, play another round

What is Pr(Alice wins on 1^{st} round) =

…

Pr(Alice wins on i^{th} round) = ?

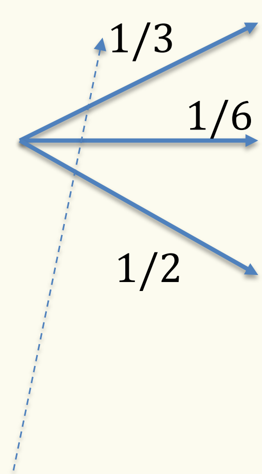
Pr(Alice wins) = ?

Sequential Process – defined in terms of independence

A 6-sided die is thrown, and each time it's thrown, regardless of the history, it is equally likely to show any of the six numbers

Local Rules: In each round

- If it shows 1,2 → Alice wins
- If it shows 3 → Bob wins
- Else, play another round



$\boldsymbol{i_{12}}$ = i th round toss 1 or 2

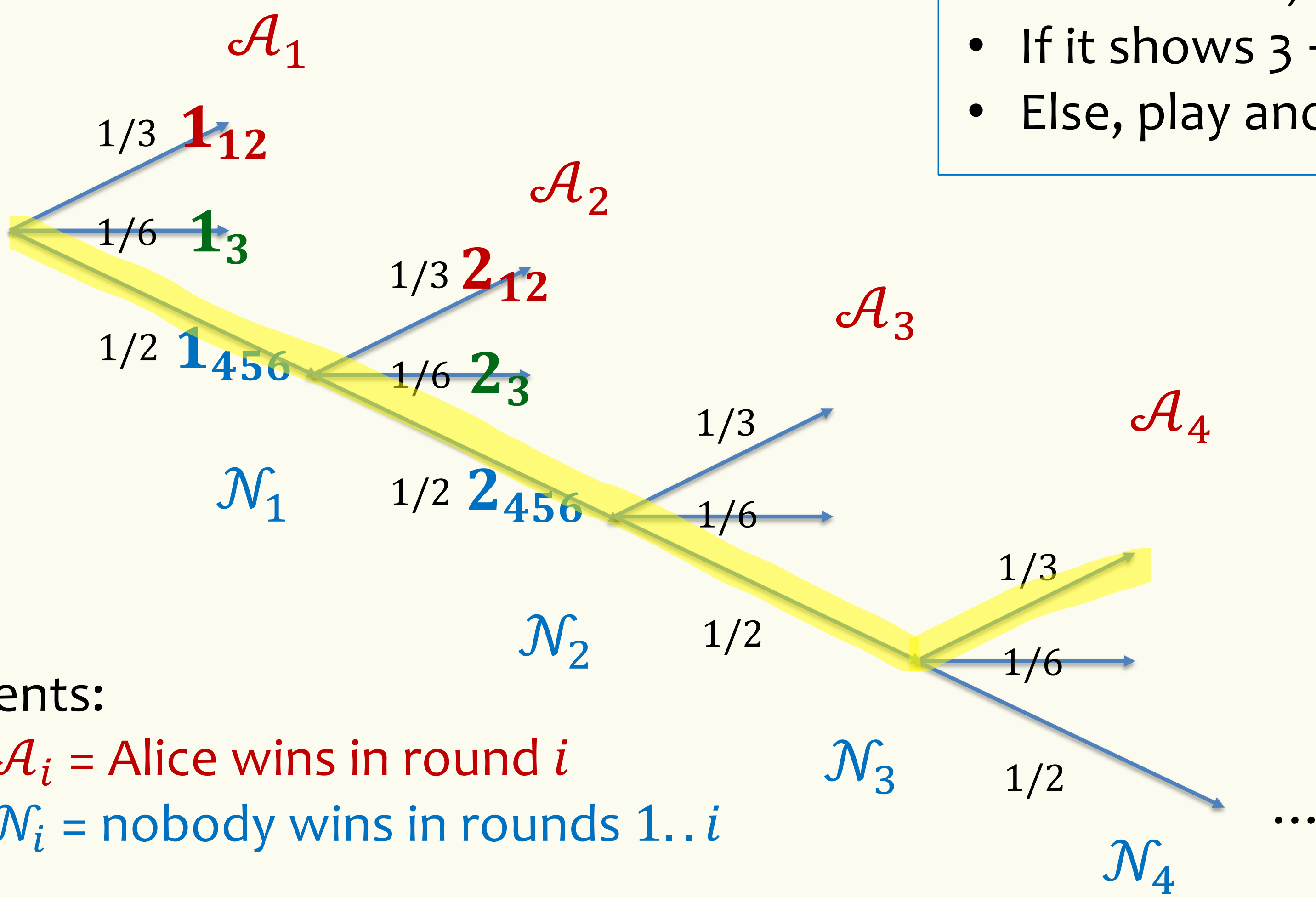
$\boldsymbol{i_{3}}$ = i th round toss 3

$\boldsymbol{i_{456}}$ = i th round toss 4 or 5 or 6

Pr (Alice wins on i -th round | nobody won in rounds $1..i$-1) = 1/3

Sequential Process – Example

Local Rules: In each round
- If it shows 1,2 → Alice wins
- If it shows 3 → Bob wins
- Else, play another round



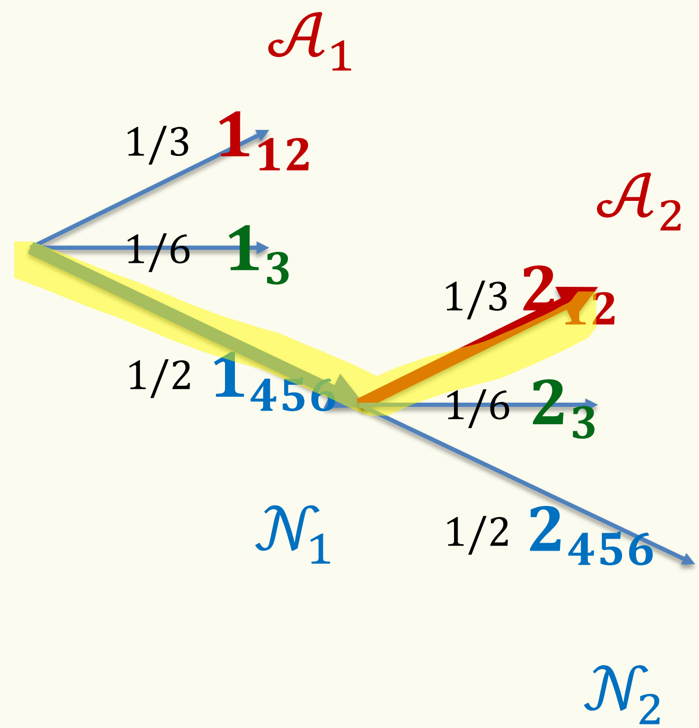
Events:
- $\mathcal{A}_i$ = Alice wins in round i
- $\mathcal{N}_i$ = nobody wins in rounds $1..i$

Sequential Process – Example

Events:

- $\mathcal{A}_i$ = Alice wins in round i
- $\mathcal{N}_i$ = nobody wins in rounds 1..i

$\mathbb{P}(\mathcal{A}_2) =$

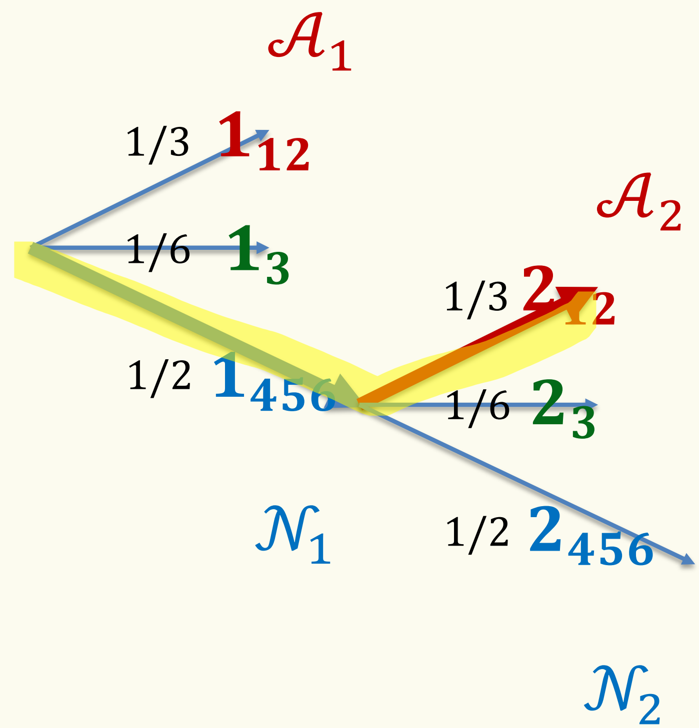


2nd roll indep of 1st roll

Sequential Process – Example

Events:

- $\mathcal{A}_i$ = Alice wins in round i
- $\mathcal{N}_i$ = nobody wins in rounds 1..i

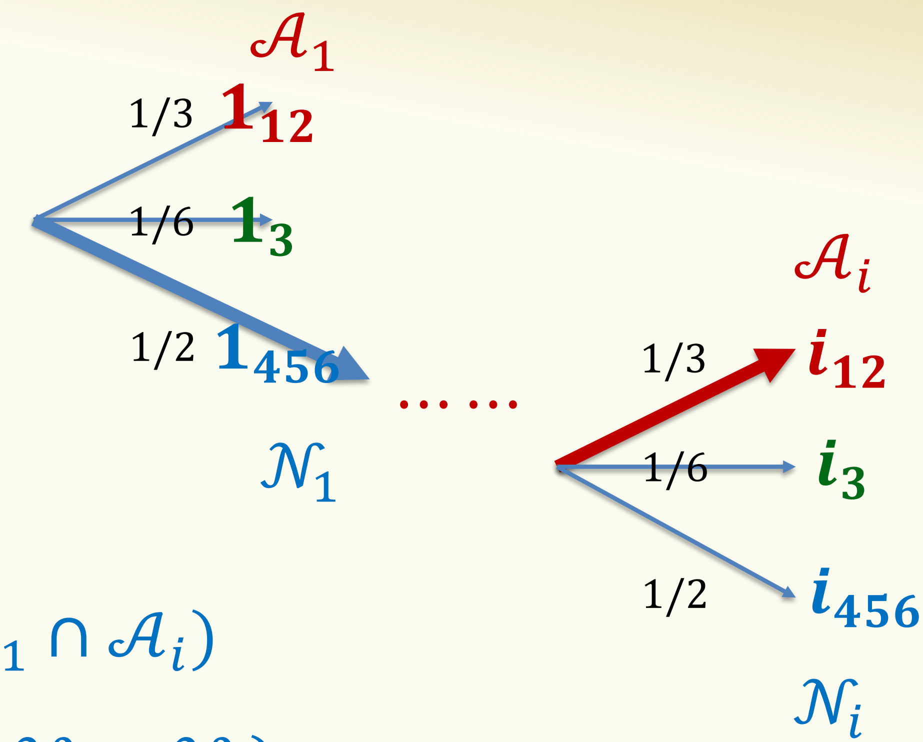


$$\mathbb{P}(\mathcal{A}_2) = \mathcal{P}(\mathcal{N}_1 \cap \mathcal{A}_2)$$
$$= \mathcal{P}(\mathcal{N}_1) \times \mathcal{P}(\mathcal{A}_2 | \mathcal{N}_1)$$
$$= \frac{1}{2} \times \frac{1}{3} = \frac{1}{6}$$

The event $\mathcal{A}_2$ implies $\mathcal{N}_1$, and this means that $\mathcal{A}_2 \cap \mathcal{N}_1 = \mathcal{A}_2$

2^{nd} roll indep of 1^{st} roll

Sequential Process – Example



Events:

- $\mathcal{A}_i$ = Alice wins in round i
- $\mathcal{N}_i$ = nobody wins in round i

$$\mathbb{P}(\mathcal{A}_i) = \mathcal{P}(\mathcal{N}_1 \cap \mathcal{N}_2 \cap \cdots \cap \mathcal{N}_{i-1} \cap \mathcal{A}_i)$$

$$= \mathcal{P}(\mathcal{N}_1) \times \mathcal{P}(\mathcal{N}_2 | \mathcal{N}_1) \times \mathcal{P}(\mathcal{N}_3 | \mathcal{N}_1 \cap \mathcal{N}_2)$$

$$\cdots \times \mathcal{P}(\mathcal{N}_{i-1} | \mathcal{N}_1 \cap \mathcal{N}_2 \cap \cdots \cap \mathcal{N}_{i-1}) \times \mathcal{P}(\mathcal{A}_i | \mathcal{N}_1 \cap \mathcal{N}_2 \cap \cdots \cap \mathcal{N}_{i-1})$$

$$= \left(\frac{1}{2}\right)^{i-1} \times \frac{1}{3}$$

Sequential Process -- Example

$\mathcal{A}_i$ = Alice wins in round i $\mathbb{P}(\mathcal{A}_i) = \left(\frac{1}{2}\right)^{i-1} \times \frac{1}{3}$

What is the probability that Alice wins?

Sequential Process -- Example

$\mathcal{A}_i$ = Alice wins in round i $\quad \mathbb{P}(\mathcal{A}_i) = \left(\frac{1}{2}\right)^{i-1} \times \frac{1}{3}$

What is the probability that Alice wins?

$$\mathbb{P}(\mathcal{A}_1 \cup \mathcal{A}_2 \cup \cdots) = \textstyle\sum_{i=1}^{\infty} \mathbb{P}(\mathcal{A}_i)$$

All $\mathcal{A}_i$'s are disjoint.

$$\sum_{i=1}^{\infty} \left(\frac{1}{2}\right)^{i-1} \times \frac{1}{3} = \frac{1}{3} \times 2 = \frac{2}{3}$$

Fact. If $|x| < 1$, then $\sum_{i=0}^{\infty} x^i = \frac{1}{1-x}$.

Agenda

- A Sequential Process Defined Using Independence
- Independence As An Assumption
- Sometimes Independence Occurs for Nonobvious Reasons
- Conditional Independence
- Correlation vs Causation
- Information Cascades

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 $\mathbb{P}(A) = 0.02$
  B : event that the backup doesn't open $\mathbb{P}(B) = 0.1$

- 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 $\mathbb{P}(A) = 0.02$

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

- 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

- A Sequential Process Defined Using Independence
- Independence As An Assumption
- Sometimes Independence Occurs for Nonobvious Reasons
- Conditional Independence
- Correlation vs Causation
- Information Cascades

Independence – Another Look

Definition. Two events $\mathcal{A}$ and $\mathcal{B}$ are (statistically) **independent** if

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

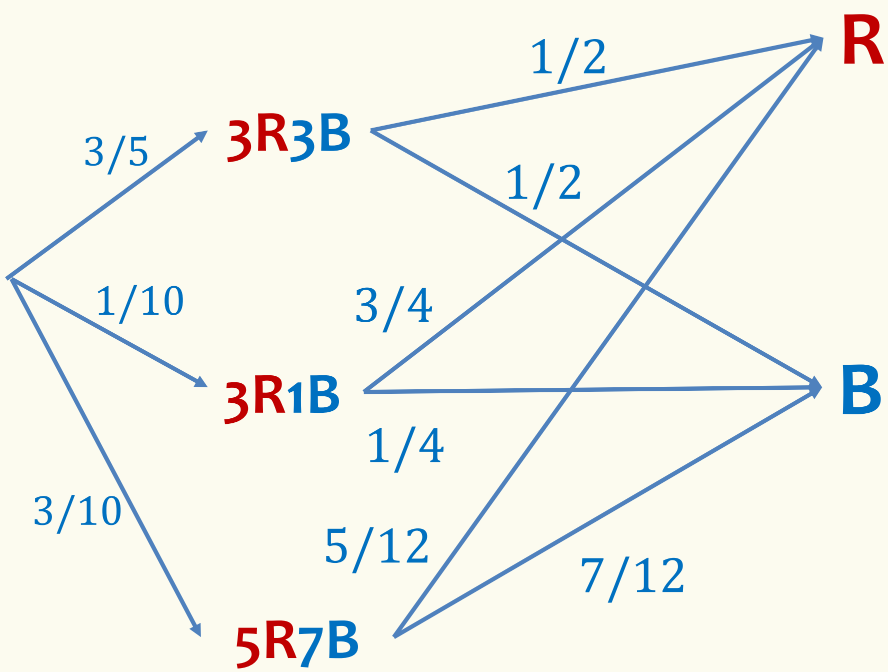
"Equivalently." $\mathbb{P}(\mathcal{A}|\mathcal{B}) = \mathbb{P}(\mathcal{A})$.

Events generated independently ➔ *their probabilities satisfy independence*

~~⇐~~ *Not necessarily*

This can be counterintuitive!

Sequential Process



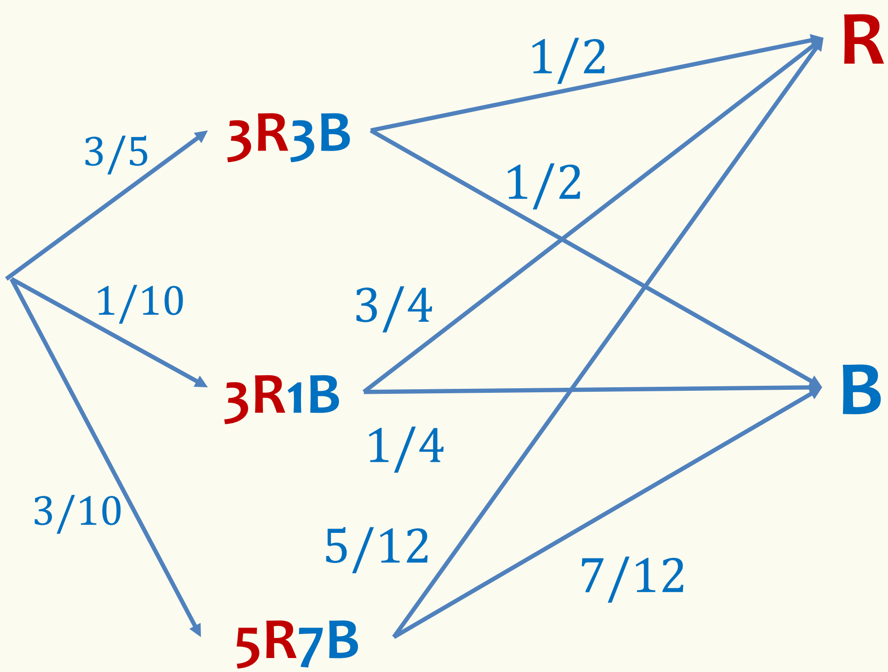
Setting: An urn contains:

- 3 **red** and 3 **blue** balls w/ probability 3/5
- 3 **red** and 1 **blue** balls w/ probability 1/10
- 5 **red** and 7 **blue** balls w/ probability 3/10

We draw a ball at random from the urn.

Are **R** and **3R3B** independent?

Sequential Process



Setting: An urn contains:

- 3 **red** and 3 **blue** balls w/ probability 3/5
- 3 **red** and 1 **blue** balls w/ probability 1/10
- 5 **red** and 7 **blue** balls w/ probability 3/10

We draw a ball at random from the urn.

$$\mathbb{P}(\mathbf{R}) = \frac{3}{5} \times \frac{1}{2} + \frac{1}{10} \times \frac{3}{4} + \frac{3}{10} \times \frac{5}{12} = \frac{1}{2}$$

$\mathbb{P}(\mathbf{3R3B}) \times \mathbb{P}(\mathbf{R} \mid \mathbf{3R3B})$

Are **R** and **3R3B** independent?

Independent! $\mathbb{P}(\mathbf{R}) = \mathbb{P}(\mathbf{R} \mid \mathbf{3R3B})$

Agenda

- A Sequential Process Defined Using Independence
- Independence As An Assumption
- Sometimes Independence Occurs for Nonobvious Reasons
- Conditional Independence
- Correlation vs Causation
- Information Cascades

Conditional Independence

Definition. Two events $\mathcal{A}$ and $\mathcal{B}$ are **independent** conditioned on C if $\mathbb{P}(C) \neq 0$ 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 $\mathbb{P}(C) \neq 0$ 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 $\mathbb{P}(\mathcal{A} | \mathcal{B} \cap C) = \mathbb{P}(\mathcal{A} \mid C)$

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?

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

- A Sequential Process Defined Using Independence
- Independence As An Assumption
- Sometimes Independence Occurs for Nonobvious Reasons
- Conditional Independence
- Correlation vs Causation
- Information Cascades

Correlation

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

- Fact: $\Pr(A|B) = 1.17 \Pr(A)$

- Conclusions?

The examples on the next few slides are from Berkeley CS 70.

Correlation

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

- Fact: $\Pr(A|B) = 1.17 \Pr(A)$

- Conclusions?
  - Smoking increases the probability of lung cancer by 17%.
  - Smoking causes lung cancer.

Correlation

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

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

- Let's take another look

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 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?

Proving Causality

- Very difficult!

You have to eliminate external causes of correlation and be able to test the cause/effect relationship (e.g. randomized clinical trials)

Some difficulties:

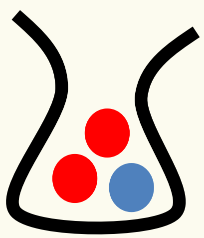
- A and B can be positively correlated because they have a common cause. E.g. being a rabbit.
- If B precedes A, then B is more likely to be the cause (e.g. smoking before lung cancer). However, they could have a common cause that induces both (e.g. studious, take CSE 312)

Agenda

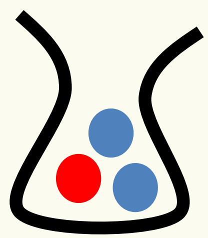
- A Sequential Process Defined Using Independence
- Independence As An Assumption
- Sometimes Independence Occurs for Nonobvious Reasons
- Conditional Independence
- Correlation vs Causation
- Information Cascades

Guessing game

Experiment: There are 3 balls in this urn. It is a ”Red urn” with probability ½ and a “Blue urn” with probability 1/2



Call this a
“Red urn”



Call this a
“Blue urn”

Guessing game

When I call your name:

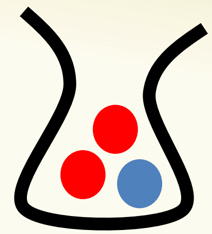
- You and only you will see a random ball.
- You must then guess if the urn is a Red urn or a Blue urn, and tell the class your guess.

You **do not** tell the class the color of the ball you pulled out

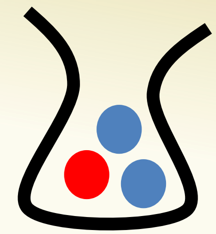
If you guess correctly you will earn one point.

What should the first student do?

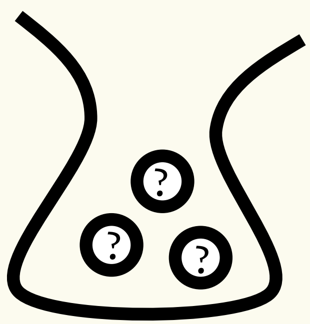
Prior



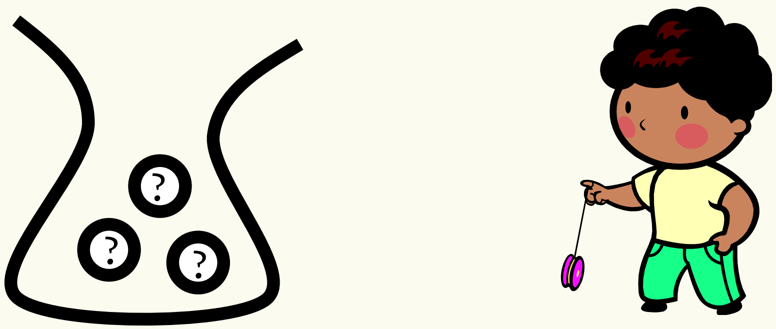
"Red urn"
Probability 1/2



"Blue urn"
Probability 1/2



What should the first student do?

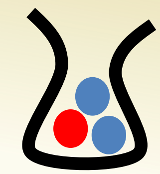


Guess that urn is the same color as the ball!

What should the second student do?

Prior

“Red urn”
Probability
1/2



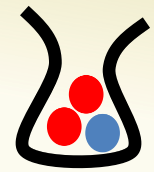
“Blue urn”
Probability
1/2

What should the second student do?

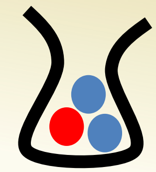
Guess that urn is the same color as the ball!

What should the 3rd student do?

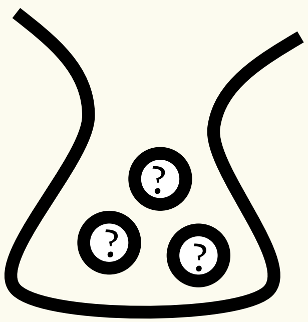
Prior



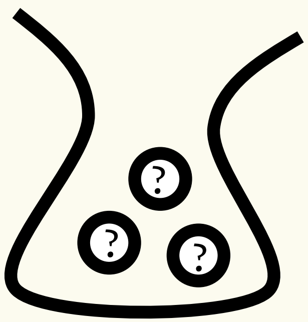
"Red urn"
Probability
1/2



"Blue urn"
Probability
1/2

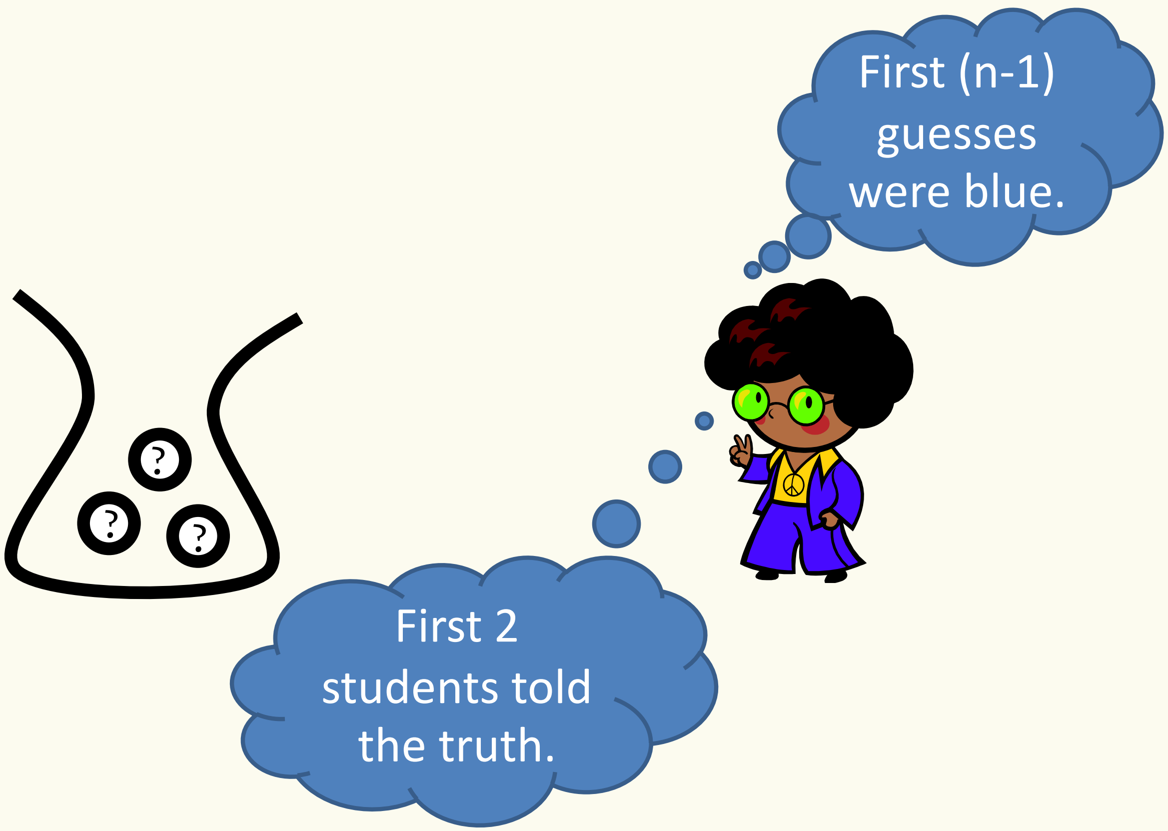


What should the 3rd student do?



Guess that urn is blue, ***no matter what she sees!!***

What should the n^{th} student do?



If the first two guesses are blue,

it is rational for ***everyone*** *to guess blue!*

Information cascades

- Staring up at the sky
- Choosing a restaurant in a strange city
- Self-reinforcing success of best-selling books.
- Voting for popular candidates

Information cascades

When

- People make decisions sequentially

- And observe actions of earlier people

Information Cascade: People abandon their own information in favor of inferences based on other's actions.

Information cascades

When

- People make decisions sequentially

- And observe actions of earlier people

This is rational!!!!!

Information Cascade: People abandon their own information in favor of inferences based on other's actions.

Observations about Information cascades

- Cascades can be wrong.

- Cascades can be based on very little information – if a cascade starts quickly in a large population, most of the private information that is collectively available is not being used.

- Cascades are fragile: Since they can be based on relatively little information, can be easy to stop if some people receive slightly superior information or people reveal more of their private information.

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 he selected or switching to the other unopened door

Should you switch or stay?