CSE 312

Foundations of Computing II

Lecture 6: 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 ☺

Last Class:

- Conditional Probability $\quad \mathbb{P}(\mathcal{B}|\mathcal{A}) = \frac{\mathbb{P}(\mathcal{A} \cap \mathcal{B})}{\mathbb{P}(\mathcal{A})}$
- Bayes Theorem $\quad \mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}$
- Law of Total probability $\quad \mathbb{P}\,(F) = \sum_{i=1}^{n} \mathbb{P}(F|E_i)\mathbb{P}(E_i) \quad E_i$ partition Ω

Bayes Theorem with Law of Total Probability

Bayes Theorem with LTP: Let $E_1, E_2, \dots, E_n$ 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).

https://pollev.com/ annakarlin185

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 may yield 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).

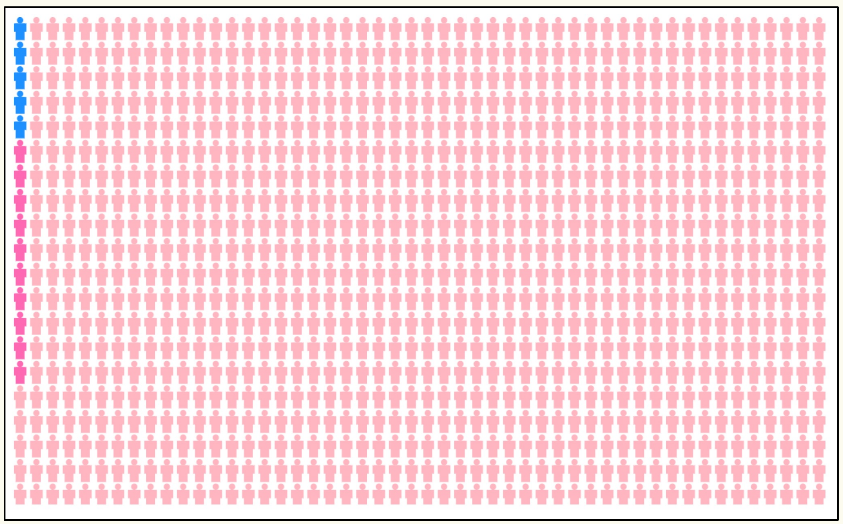
Example – Zika Testing

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% 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**

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

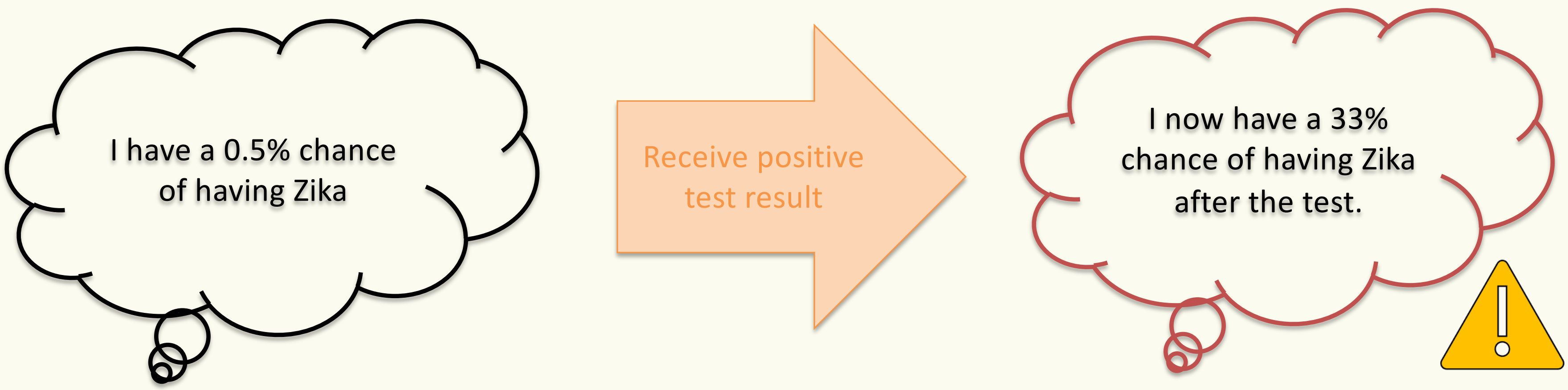
Demo

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 A follows the same properties as (unconditional) probability.

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

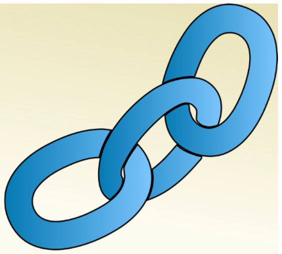
Formally. $(\Omega, \mathbb{P})$ is a probability space + $\mathbb{P}(\mathcal{A}) > 0$

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

Today:

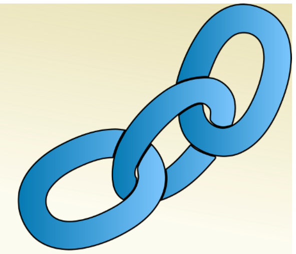
- Chain Rule
- Independence
- Sequential Process

Chain Rule



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

Chain Rule



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

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

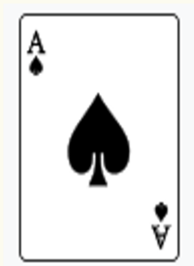
$$\mathbb{P}(\mathcal{A}_1 \cap \cdots \cap \mathcal{A}_n) = \mathbb{P}(\mathcal{A}_1) \cdot \mathbb{P}(\mathcal{A}_2|\mathcal{A}_1) \cdot \mathbb{P}(\mathcal{A}_3|\mathcal{A}_1 \cap \mathcal{A}_2)$$

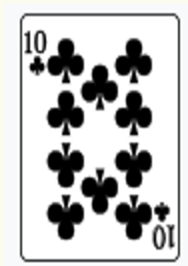
$$\cdots \mathbb{P}(\mathcal{A}_n|\mathcal{A}_1 \cap \mathcal{A}_2 \cap \cdots \cap \mathcal{A}_{n-1})$$

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 (

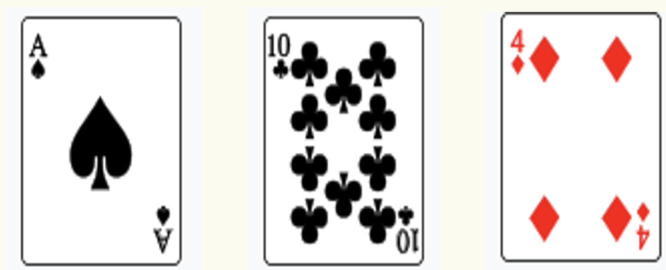


) = P(**A** ∩ **B** ∩ **C**)?

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

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

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

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

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?

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

Poll:

A. Yes, independent

B. No

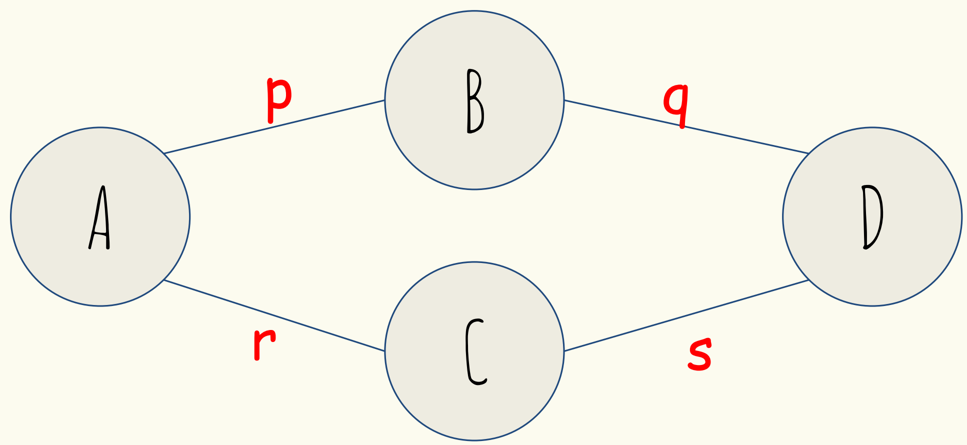
pollev/annakarlin185

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

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

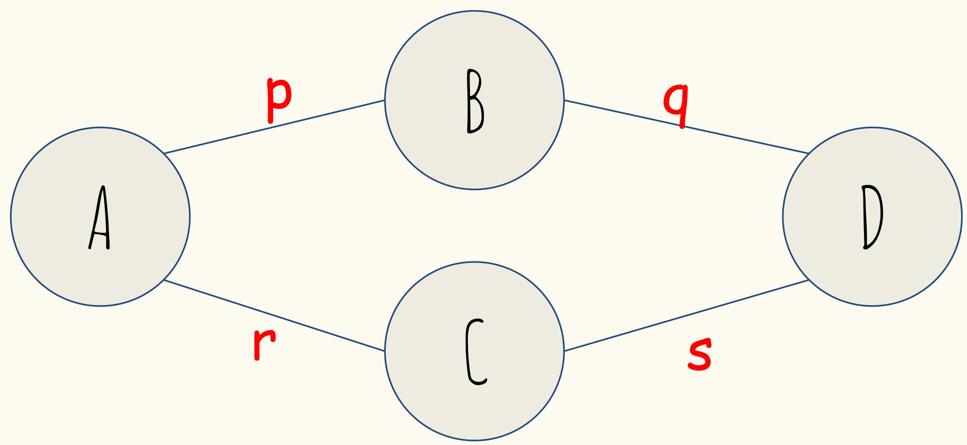
Each link works with the probability given, **independently.**
What's the probability A and D can communicate?

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

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

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

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



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

Example – Biased coin

We have a biased coin comes up Heads with probability 2/3; Each flip is independent of all other fips. 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.

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

https://pollev.com/ annakarlin185

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

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

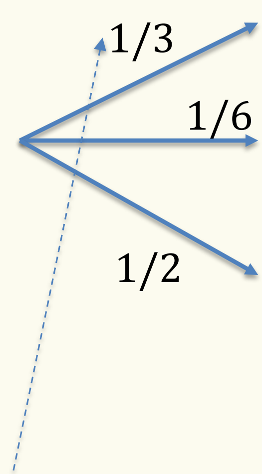
What is Pr(Alice wins on 1^{st} round) =
Pr(Alice wins on 2^{st} round) =
…
Pr(Alice wins on i^{th} round) = ?
Pr(Alice wins) = ?

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

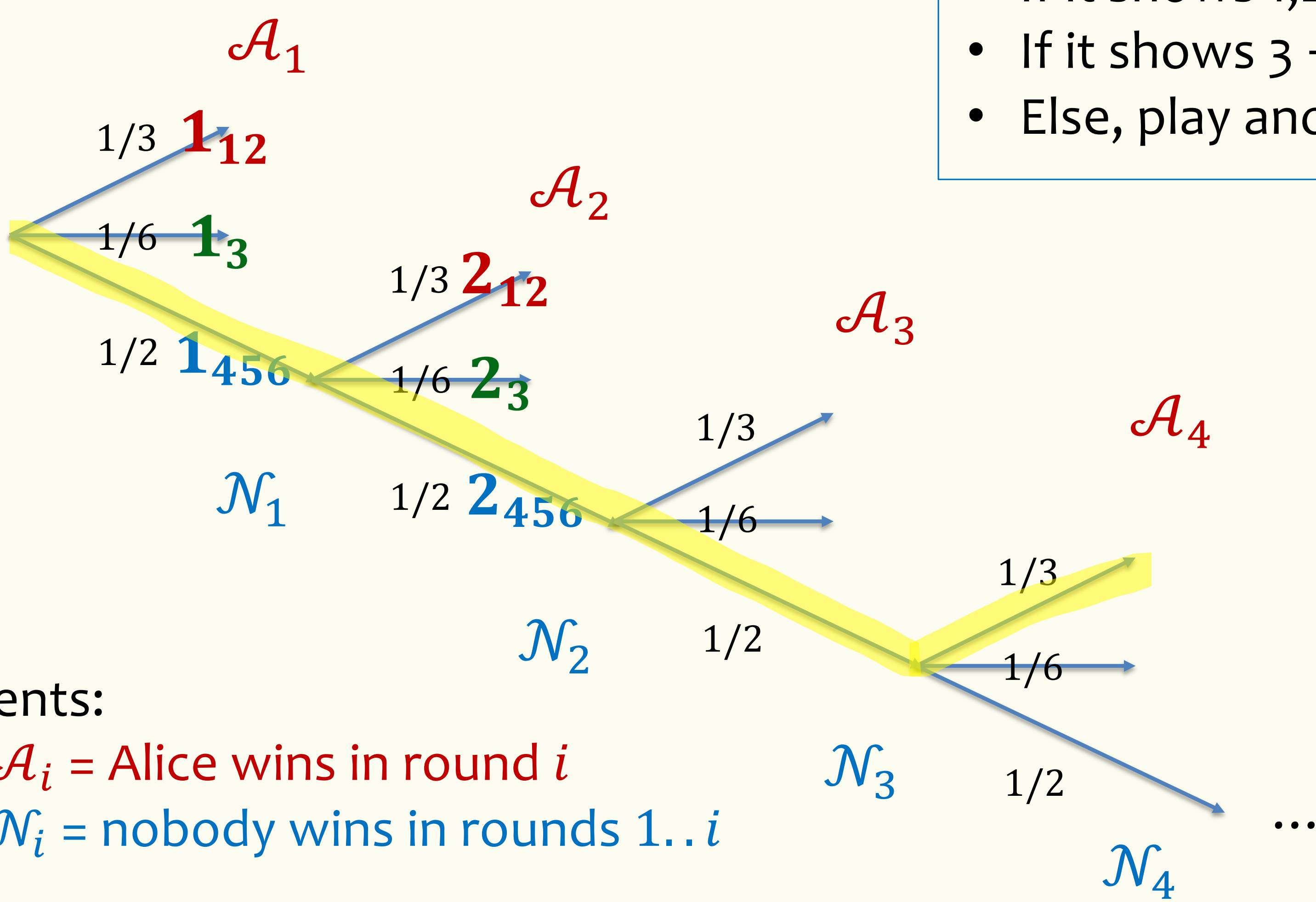
$\boldsymbol{i_{456}}$ = i th round toss 4 or 5 or 6

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:

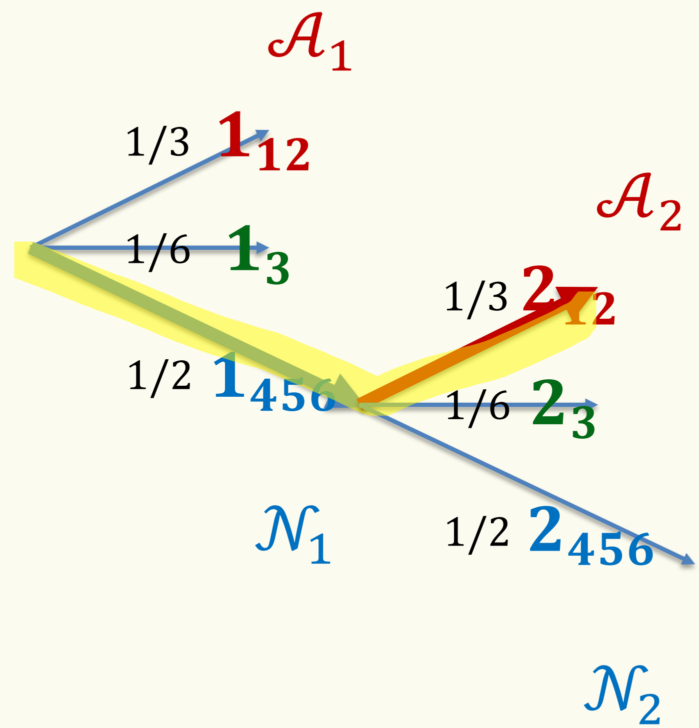
- $\mathcal{A}_i$ = Alice wins in round i
- $\mathcal{N}_i$ = nobody wins in rounds $1..i$

Sequential Process – Example

Events:

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

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

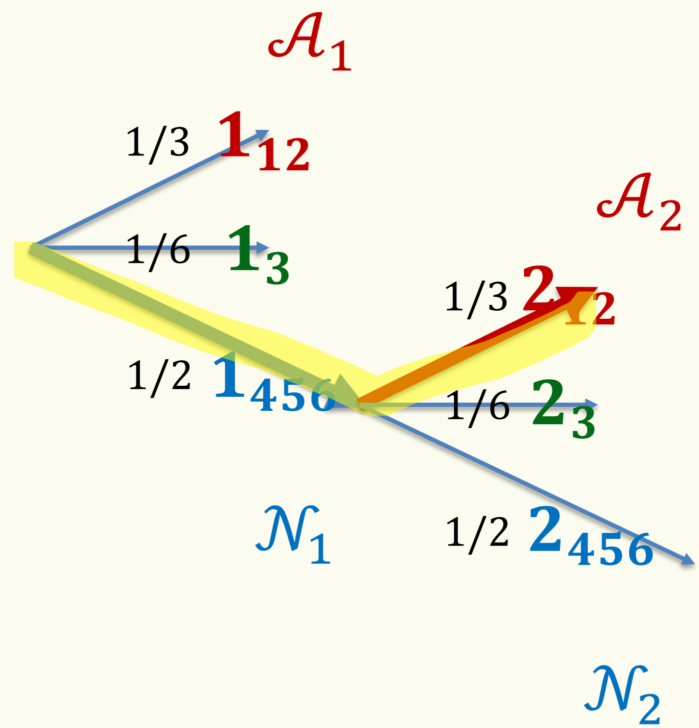


2nd roll indep of 1st roll

Sequential Process – Example

Events:

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



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

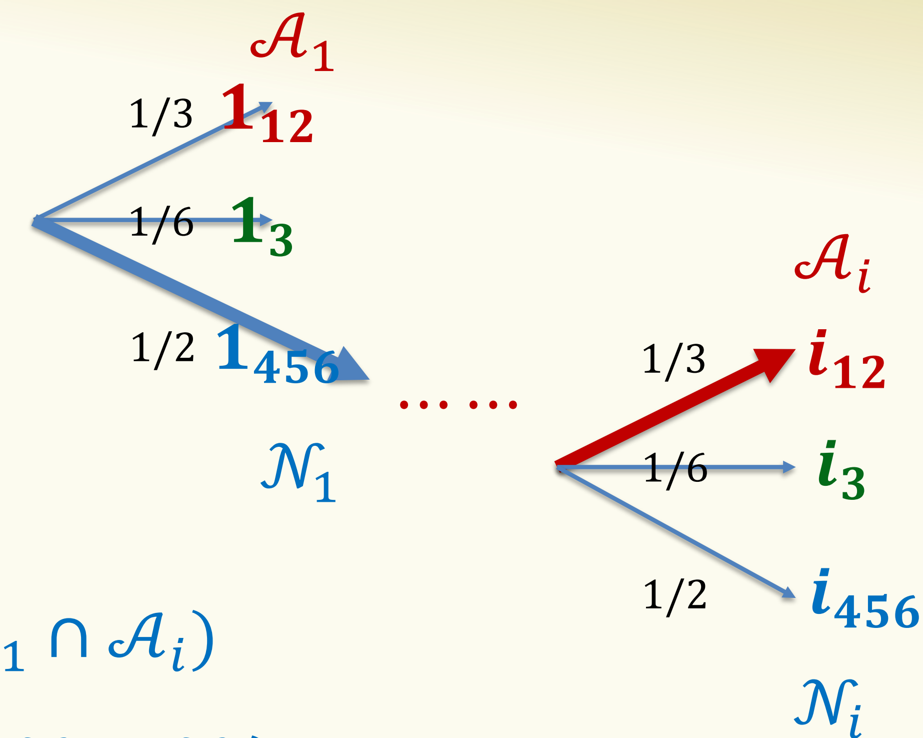
The event $\mathcal{A}_2$ implies $\mathcal{N}_1$, and this means that $\mathcal{A}_2 \cap \mathcal{N}_1 = \mathcal{A}_2$

2nd roll indep of 1st 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 $\quad \mathbb{P}(\mathcal{A}_i) = \left(\frac{1}{2}\right)^{i-1} \times \frac{1}{3}$

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?

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

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 $\sum_{i=0}^{\infty} x^i = \frac{1}{1-x}$.

Independence – Another Look

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

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

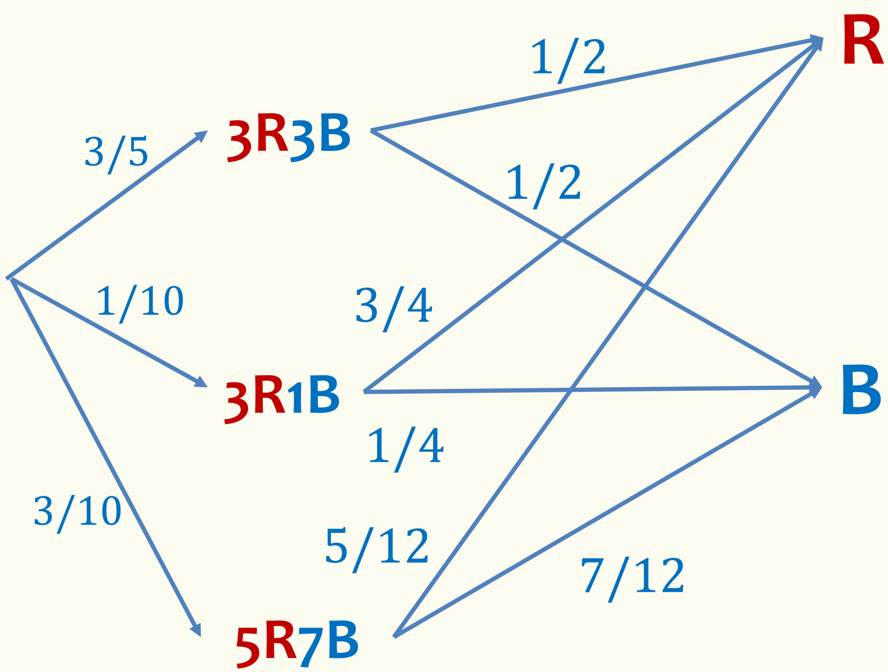
"Equivalently." $\mathbb{P}(\mathcal{A}|\mathcal{B}) = \mathbb{P}(\mathcal{A})$.

Events generated independently ➔ *their probabilities satisfy independence*

~~⇐~~ *Not necessarily*

This can be counterintuitive!

Sequential Process



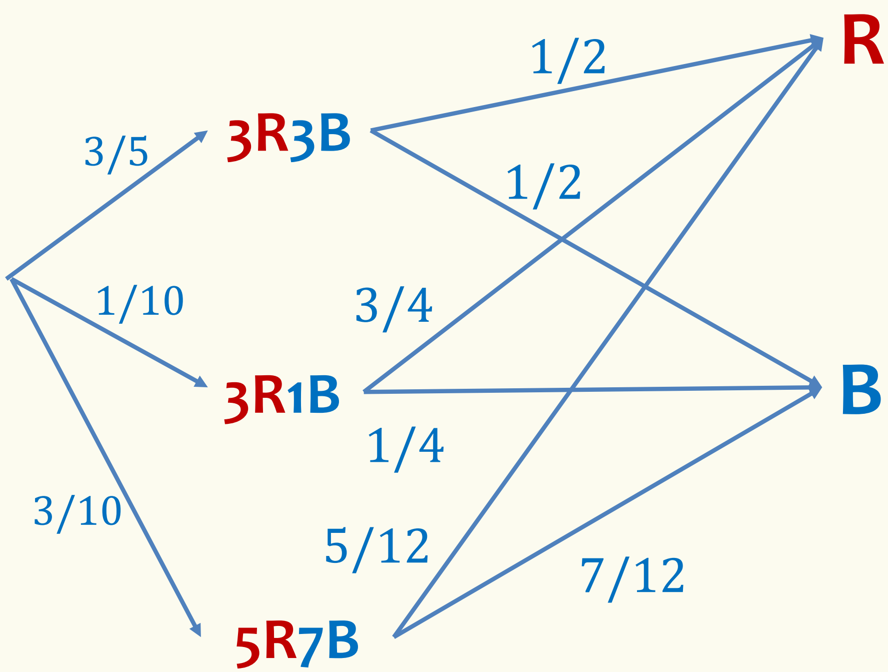
Setting: An urn contains:

- 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.

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

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

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

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

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

Equivalence:

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

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?

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

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

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

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