

**CSE 312**

# **Foundations of Computing II**

**Lecture 6: Chain Rule and Independence**




**[slido.com/4171468](https://slido.com/4171468)  
for polls and anonymous questions**

# Thank you for your feedback!!!



- Several people mentioned that I was going too fast.
  - *Slow me down! Ask questions!!! That's your job!!*
  - Watch Summer 2020 videos **before** class (at half speed)
  - Do the reading **before** class.
- If you want more practice
  - Do all the section problems!
  - Problems in all three readings.
  - MIT “Mathematics for Computer Science” 6.042J (sections on counting & probability)
  - Get the book “A First Course in Probability” by Sheldon Ross

## Agenda

- Recap 
- Chain Rule
- Independence
- Conditional independence
- Infinite process

## Review Conditional & Total Probabilities

- **Conditional Probability**

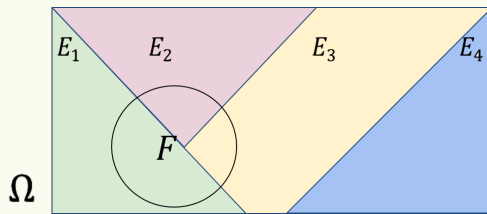
$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

- **Bayes Theorem**

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad \text{if } P(A) \neq 0, P(B) \neq 0$$

- **Law of Total Probability**

$E_1, \dots, E_n$  partition  $\Omega$



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

## 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  $F$  an 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

Zika fever

OVERVIEW SYMPTOMS SPECIALISTS

Fever  
Rash  
Joint pain  
Red eyes



Spread through mosquito bites *Source*

A disease caused by Zika virus that's spread through mosquito bites.

The image shows a woman with a red rash on her neck and shoulder. A circular inset shows a mosquito biting her arm. The text 'Spread through mosquito bites' and 'Source' is written below the inset.

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.

## Example – Zika Testing

$Z$  have Zika  
 $T$  test positive

Suppose we know the following Zika stats

- A test is 98% effective at detecting Zika (“true positive”)  $P(T|Z) = 0.98$
- However, the test may yield a “false positive” 1% of the time  $P(T|Z^c) = 0.01$
- 0.5% of the US population has Zika.  $P(Z) = 0.005$

What is the probability you have Zika (event  $Z$ ) given that you test positive (event  $T$ )?

By Bayes Rule,  $P(Z|T) = \frac{P(T|Z)P(Z)}{P(T)}$

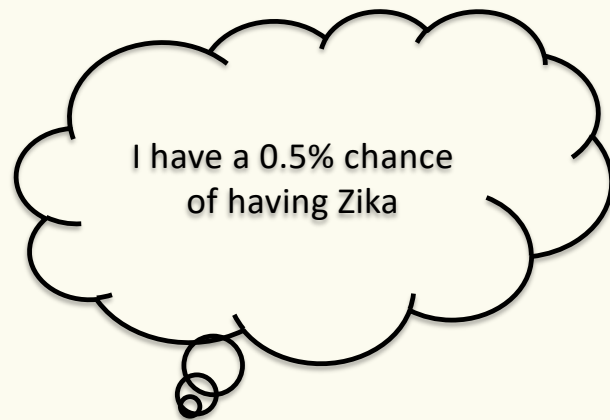
By the Law of Total Probability,  $P(T) = P(T|Z)P(Z) + P(T|Z^c)P(Z^c)$

## Philosophy – Updating Beliefs

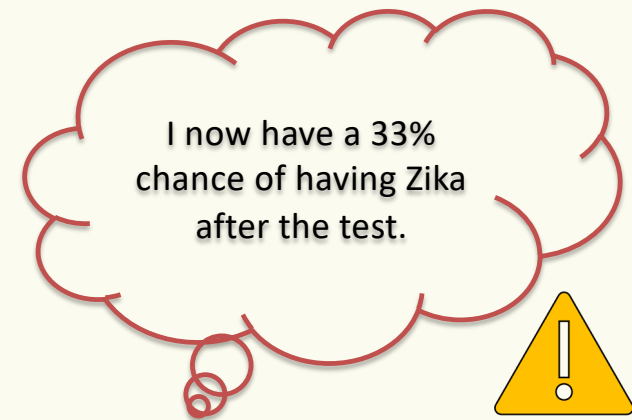
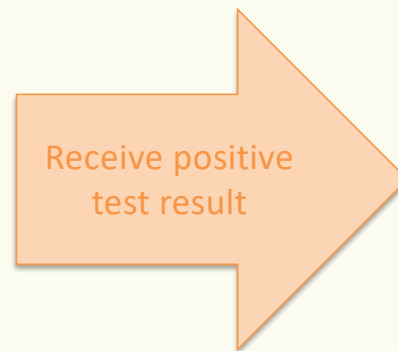
While it's not 98% that you have the disease, your beliefs changed **significantly**

Z = you have Zika

T = you test positive for Zika



**Prior:**  $P(Z)$



**Posterior:**  $P(Z|T)$



## Example – Zika Testing

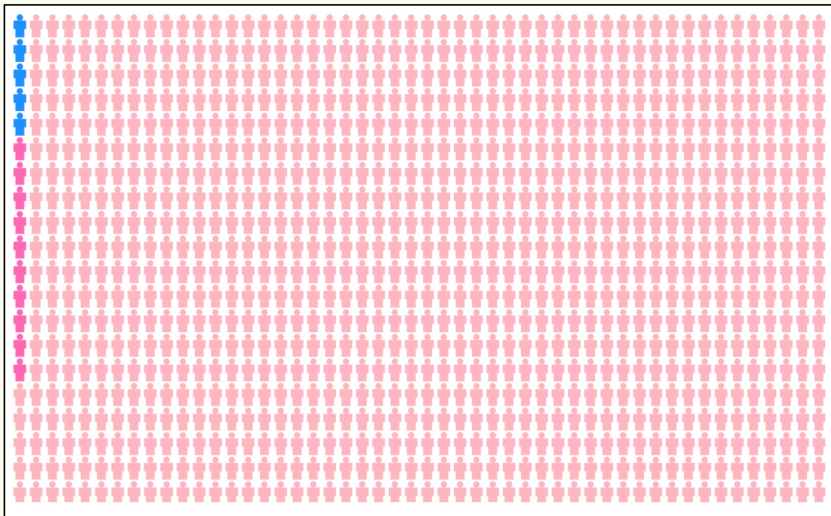
Have zika blue, don't pink

What is the probability you have Zika (event  $Z$ ) given that you test positive (event  $T$ ).

$$P(T|Z) =$$

$$P(T|Z^c) =$$

$$P(Z) =$$



Suppose we had 1000 people:

- 5 have Zika and all test positive
- 985 do not have Zika and test negative
- 10 do not have Zika and test positive

## Example – Zika Testing

Have zika blue, don't pink

Picture below gives us the following Zika stats

- A test is **100%** 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.

$$P(T|Z) = 5/5 = 1$$

$$P(T|Z^c) = 10/995$$

$$P(Z) = \frac{5}{1000} = 0.005$$

What is the probability you have Zika (event  $Z$ ) given that you test positive (event  $T$ )?



Suppose we had 1000 people:

- **5 have Zika and all test positive**
- **985 do not have Zika and test negative**
- **10 do not have Zika and test positive**

## Example – Zika Testing

Have zika blue, don't pink

Picture below gives us the following Zika stats

- A test is **100%** effective at detecting Zika (“true positive”).  $P(T|Z) = 5/5 = 1$
- However, the test may yield a “false positive” 1% of the time  $P(T|Z^c) = 10/995$
- 0.5% of the US population has Zika. 5% have it.  $P(Z) = \frac{995}{1000} = 0.005$

What is the probability you have Zika (event  $Z$ ) given that 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**

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

Suppose we know the following Zika stats

- A test is 98% effective at detecting Zika (“true positive”)  $P(T|Z) = 0.98$
- However, the test may yield a “false positive” 1% of the time  $P(T|Z^c) = 0.01$
- 0.5% of the US population has Zika.  $P(Z) = 0.005$

What is the probability you test negative (event  $T^c$ ) given you have Zika (event  $Z$ )?

## Conditional Probability Define a Probability Space

The probability conditioned on  $A$  satisfies the required axioms.


**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

**Axiom 1 (Non-negativity):**  $P(E) \geq 0$ .

**Axiom 2 (Normalization):**  $P(\Omega) = 1$

**Axiom 3 (Countable Additivity):** If  $E$  and  $F$  are mutually exclusive, then  $P(E \cup F) = P(E) + P(F)$

## Agenda

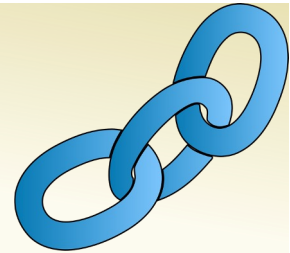
- Recap
- Chain Rule 
- Independence
- Conditional independence
- Infinite process

## Chain Rule

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

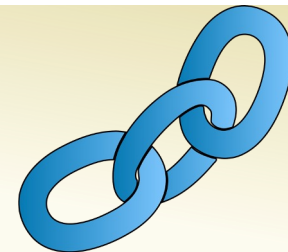


$$\mathbb{P}(\mathcal{A})\mathbb{P}(\mathcal{B}|\mathcal{A}) = \mathbb{P}(\mathcal{A} \cap \mathcal{B})$$





## Chain Rule



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

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

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

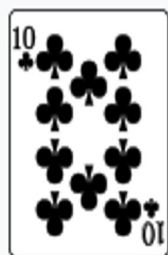
$$\dots \mathbb{P}(\mathcal{A}_n|\mathcal{A}_1 \cap \mathcal{A}_2 \cap \dots \cap \mathcal{A}_{n-1})$$

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  $P(\text{Ace of Spades First} \cap \text{10 of Clubs Second} \cap \text{4 of Diamonds Third}) = P(\mathbf{A} \cap \mathbf{B} \cap \mathbf{C})$ ?

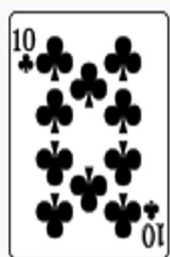


- A:** Ace of Spades First
- B:** 10 of Clubs Second
- C:** 4 of Diamonds Third

## Chain Rule Example

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

What is  $P(\text{Ace of Spades First} \cap \text{10 of Clubs Second} \cap \text{4 of Diamonds Third}) = P(\mathbf{A} \cap \mathbf{B} \cap \mathbf{C})$ ?



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

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

**A:** Ace of Spades First  
**B:** 10 of Clubs Second  
**C:** 4 of Diamonds Third

## Agenda

- Recap
- Chain Rule
- **Independence** ◀
- Conditional independence
- Infinite process

## 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  $\mathbb{P}(\mathcal{B}|\mathcal{A}) = \mathbb{P}(\mathcal{B})$
- 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.

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

Independent?

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

## Multiple Events – Mutual Independence

**Definition.** Events  $A_1, \dots, A_n$  are **mutually independent** if for every non-empty subset  $I \subseteq \{1, \dots, n\}$ , we have

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i).$$

## Example – Network Communication

Each link works with the probability given, **independently**

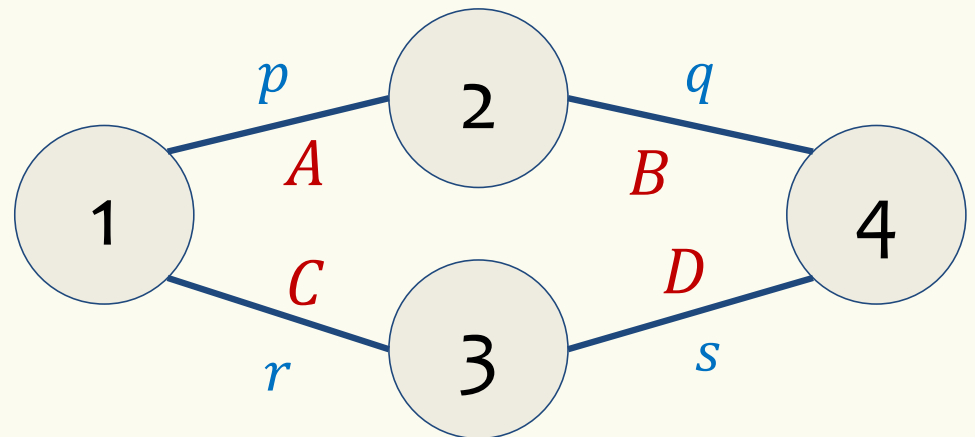
i.e., mutually independent events  $A, B, C, D$  with

$$P(A) = p$$

$$P(B) = q$$

$$P(C) = r$$

$$P(D) = s$$



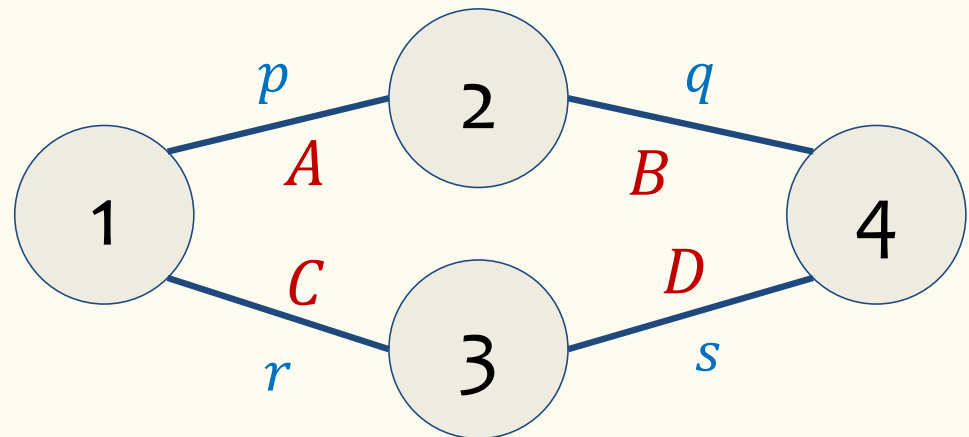


## Example – Network Communication

Each link works with the probability given, **independently**

i.e., mutually independent events  $A, B, C, D$

What is  $P(1-4 \text{ connected})$ ?



$$P(A) = p$$

$$P(B) = q$$

$$P(C) = r$$

$$P(D) = s$$

## Example – Network Communication

If each link works with the probability given, **independently**:  
What's the probability that nodes 1 and 4 can communicate?

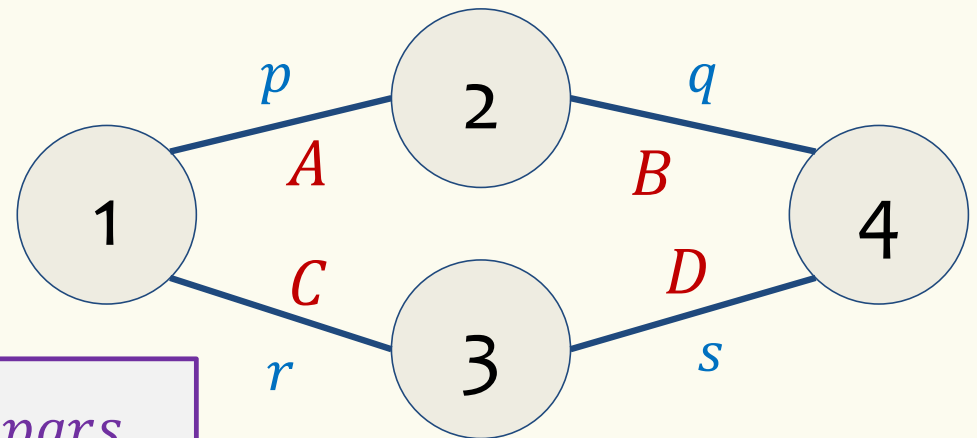
$$\begin{aligned}P(1-4 \text{ connected}) &= P((A \cap B) \cup (C \cap D)) \\ &= P(A \cap B) + P(C \cap D) - P(A \cap B \cap C \cap D)\end{aligned}$$

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

$$P(C \cap D) = P(C) \cdot P(D) = rs$$

$$P(A \cap B \cap C \cap D)$$

$$= P(A) \cdot P(B) \cdot P(C) \cdot P(D) = pqrs$$



$$P(1-4 \text{ connected}) = pq + rs - pqrs$$

## Independence – Another Look

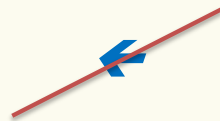
**Definition.** Two events  $A$  and  $B$  are (statistically) **independent** if

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

**“Equivalently.”**  $P(A|B) = P(A)$ .

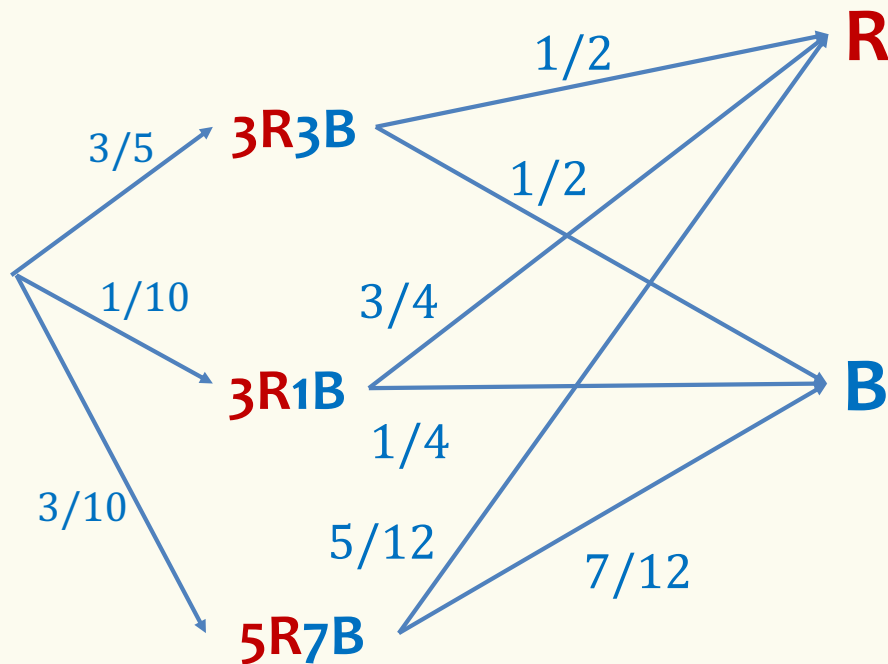
*Events generated independently  $\rightarrow$  their probabilities satisfy independence*

*But events can be independent without being generated by independent processes.*



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.

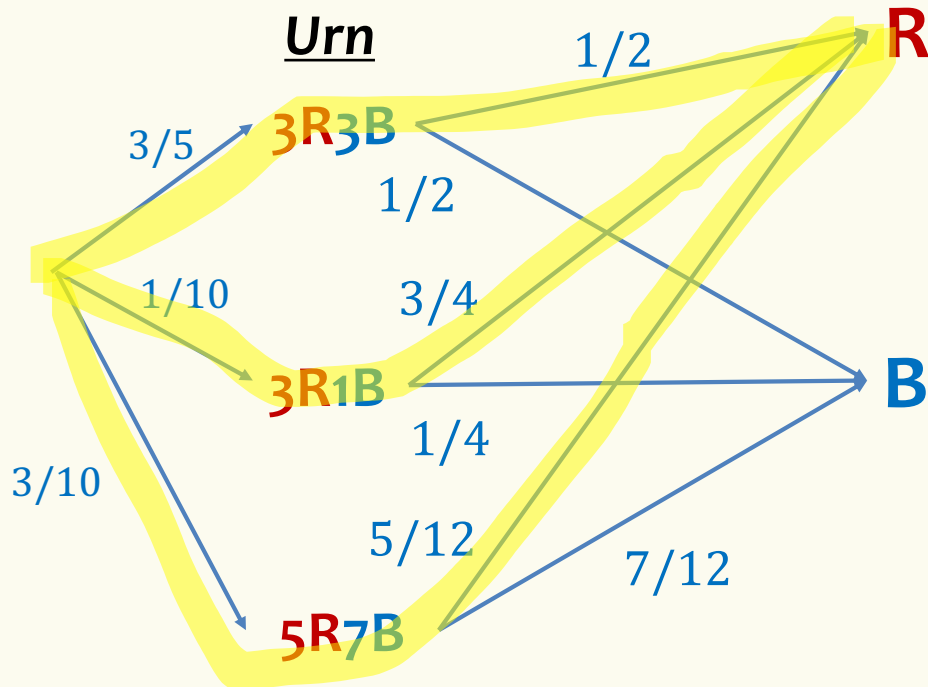
Are **R** and **3R3B** independent?

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

$$P(\mathbf{R}) =$$

## Sequential Process

## Ball drawn



**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  $3/10$

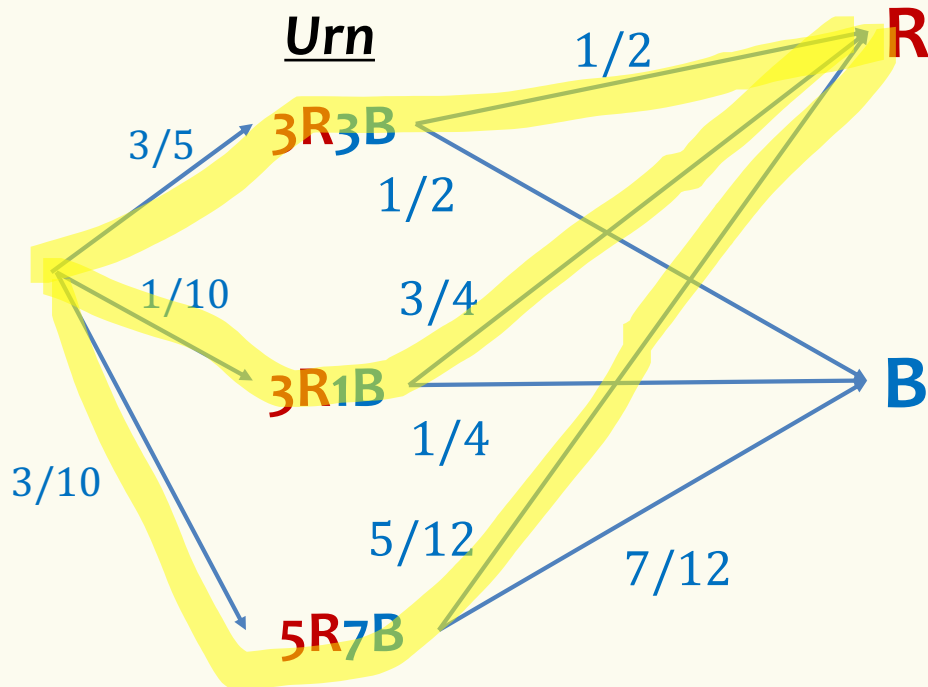
We draw a ball at random from the urn.

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

$$P(\mathbf{R} \mid \mathbf{3R3B}) = \frac{1}{2}$$

## Sequential Process

## Ball drawn



**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?

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

$$P(\mathbf{R} \mid \mathbf{3R3B}) = \frac{1}{2}$$

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



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



## 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.

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

[slido.com/4171468](https://slido.com/4171468)

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

## Independence as an assumption

- People often assume it **without justification**
- Example: A skydiver has two chutes
  - $A$ : event that the main chute doesn't open       $P(A) = 0.02$
  - $B$ : event that the back-up doesn't open       $P(B) = 0.1$
- What is the chance that at least one opens assuming independence?

## Independence as an assumption

- People often assume it **without justification**
- Both chutes could fail because of the same rare event e.g., freezing rain.

## Agenda

- Recap
- Chain Rule
- Independence
- **Conditional independence** ◀
- Infinite process

## Conditional Independence

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

- If  $P(A \cap C) \neq 0$ , equivalent to  $P(B|A \cap C) = P(B | C)$
- If  $P(B \cap C) \neq 0$ , equivalent to  $P(A|B \cap C) = P(A | C)$

**Contrast to Plain Independence.** Two events  $A$  and  $B$  are **independent** if

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

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

## Example – Throwing Dice

Suppose that Coin 1 has probability of heads 0.3  
and Coin 2 has probability of head 0.9.

We choose one of the coins randomly with equal probability and flip that coin 2 times independently. **What is the probability we get all heads?**

$$\begin{aligned} P(HH) &= P(HH | C_1) \cdot P(C_1) + P(HH | C_2) \cdot P(C_2) && \text{Law of Total Probability (LTP)} \\ &= P(H|C_1)^2 P(C_1) + P(H | C_2)^2 P(C_2) && \text{Conditional Independence} \\ &= 0.3^2 \cdot 0.5 + 0.9^2 \cdot 0.5 = 0.45 \end{aligned}$$

$C_i$  = coin  $i$  was selected

## Example – Throwing Dice

Suppose that Coin 1 has probability of heads 0.3  
and Coin 2 has probability of head 0.9.

We choose one coin randomly with equal probability and flip that coin 2 times independently. What is the probability we get all heads?

$$P(HH | C_1) = P(H|C_1)^2$$

$$P(HH) \neq P(H)^2$$

$$P(HH) = P(H|C_1)^2 P(C_1) + P(H | C_2)^2 P(C_2) = 0.3^2 \cdot 0.5 + 0.9^2 \cdot 0.5 = 0.45$$

$$P(H) = P(H|C_1) \cdot P(C_1) + P(H|C_2) \cdot P(C_2) \doteq 0.6 \qquad P(HH) = 0.36$$



## Agenda

- Recap
- Chain Rule
- Independence
- Conditional independence
- **Infinite process** ◀

Often probability space  $(\Omega, P)$  is given **implicitly** via sequential process

- *Experiment proceeds in  $n$  sequential steps, each step follows some **local rules** defined by the chain rule and independence*
- *Natural extension: Allows for easy definition of experiments where  $|\Omega| = \infty$*

## 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  $\Pr(\text{Alice wins on } 1^{\text{st}} \text{ round}) =$

$\Pr(\text{Alice wins on } 2^{\text{st}} \text{ round}) =$

...

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

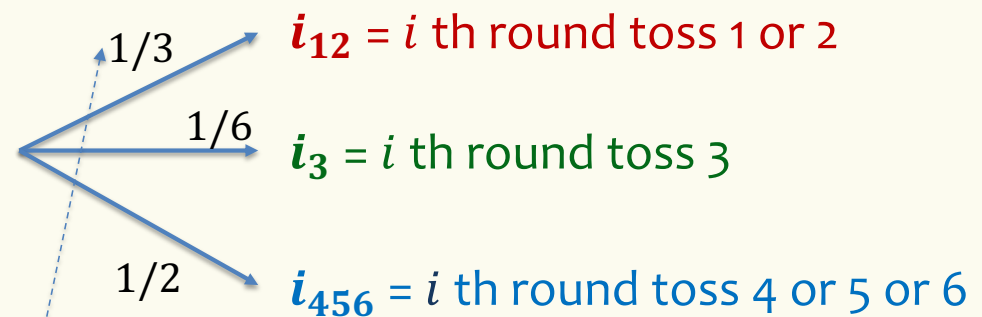
$\Pr(\text{Alice wins}) = ?$

## 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, toss a die

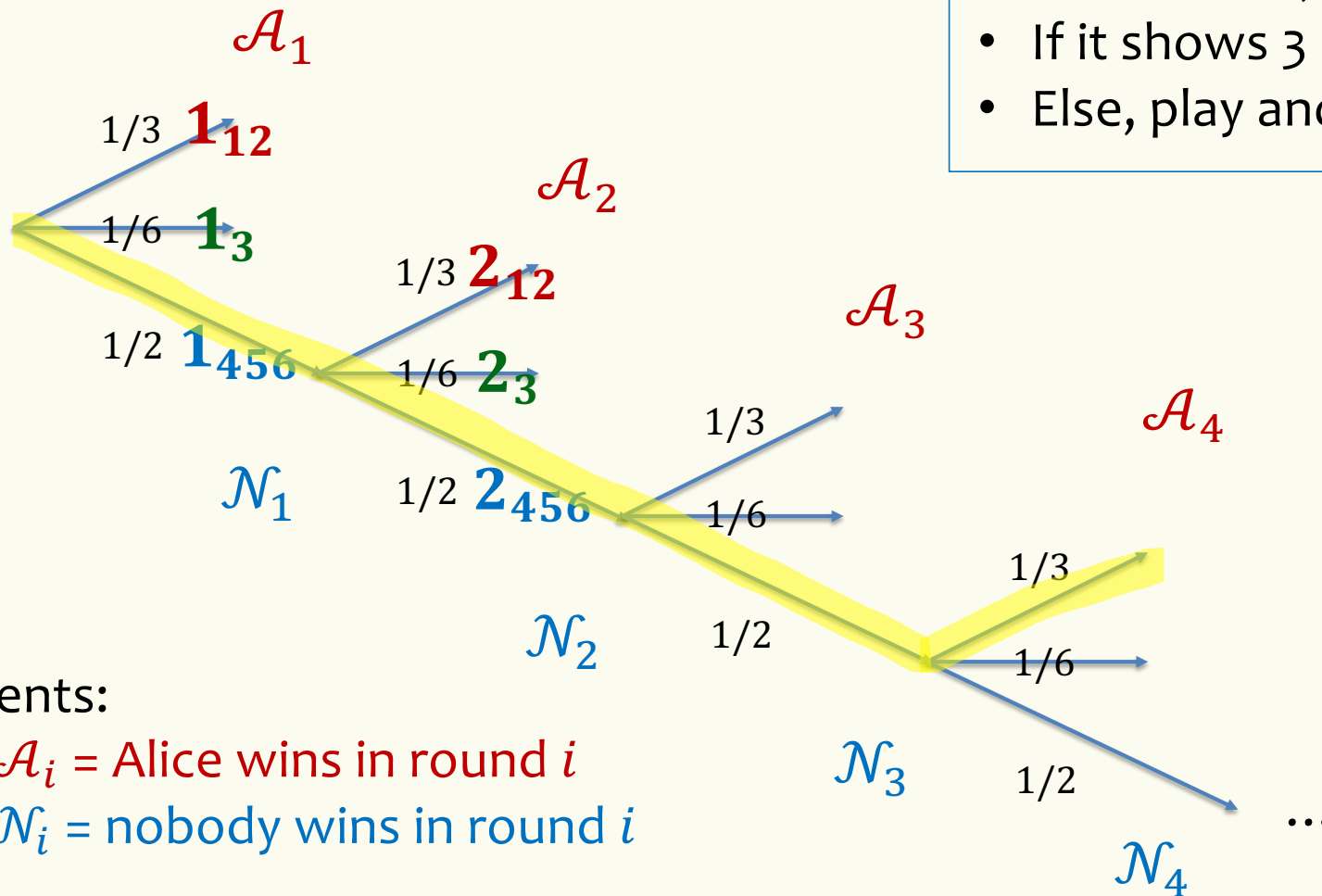
- If it shows 1,2 → **Alice wins**
- If it shows 3 → **Bob wins**
- Else, play another round



$\Pr(\text{Alice wins on } i\text{-th round} \mid \text{nobody won in rounds } 1..i-1) = 1/3$

## 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:

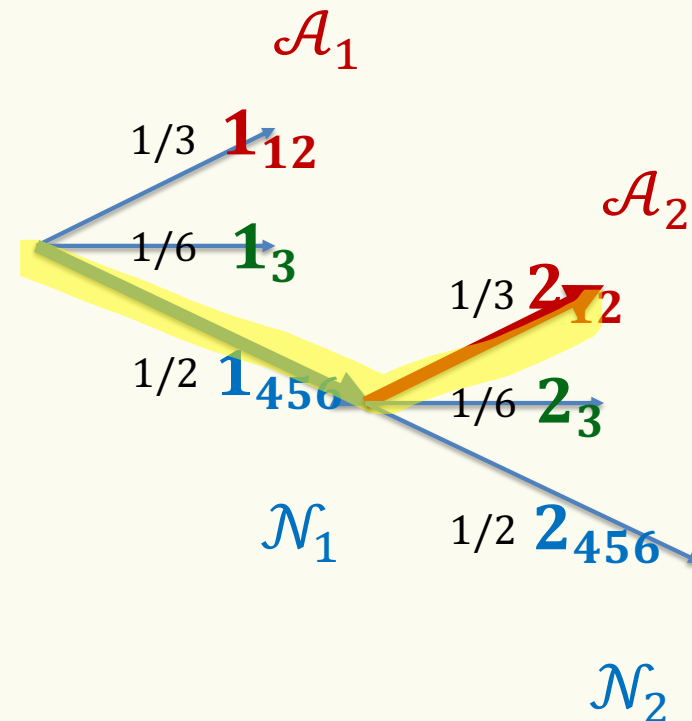
- $\mathcal{A}_i$  = Alice wins in round  $i$
- $\mathcal{N}_i$  = nobody wins in round  $i$

## Sequential Process – Example

Events:

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

$$\mathbb{P}(\mathcal{A}_2) = \mathbb{P}(\mathcal{N}_1 \cap \mathcal{A}_2)$$



2<sup>nd</sup> roll indep of 1<sup>st</sup> roll

## Sequential Process – Example

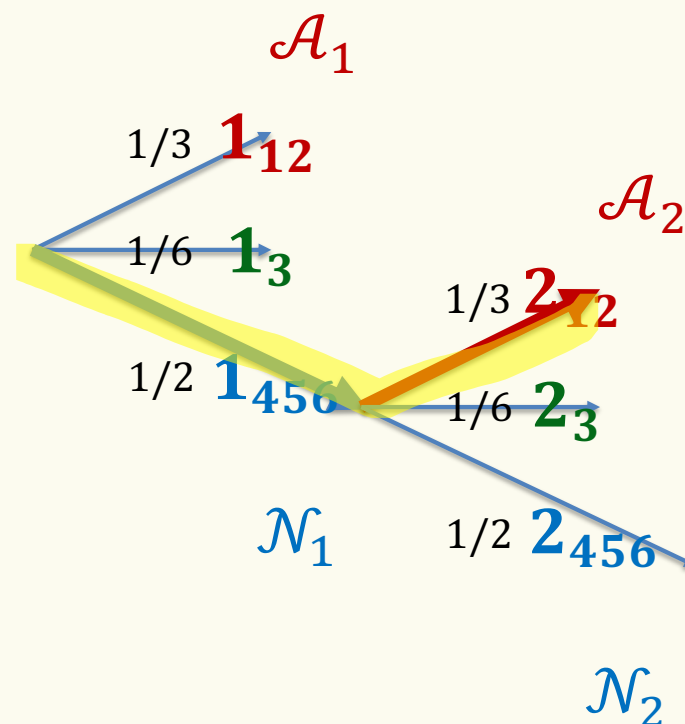
Events:

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

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

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

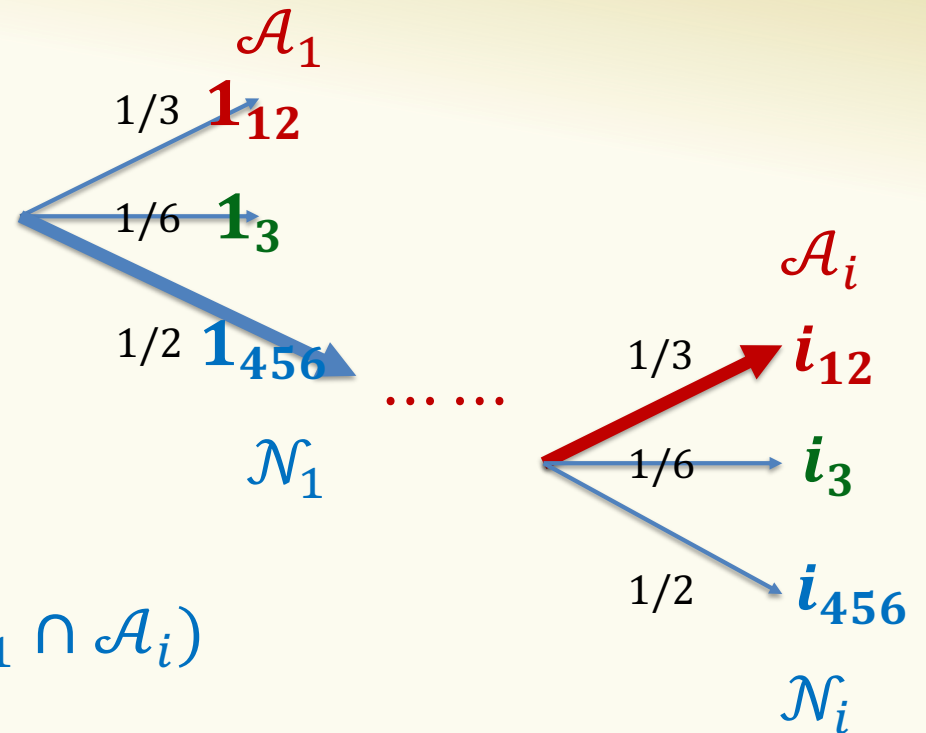
2<sup>nd</sup> roll indep of 1<sup>st</sup> roll



## Sequential Process – Example

Events:

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



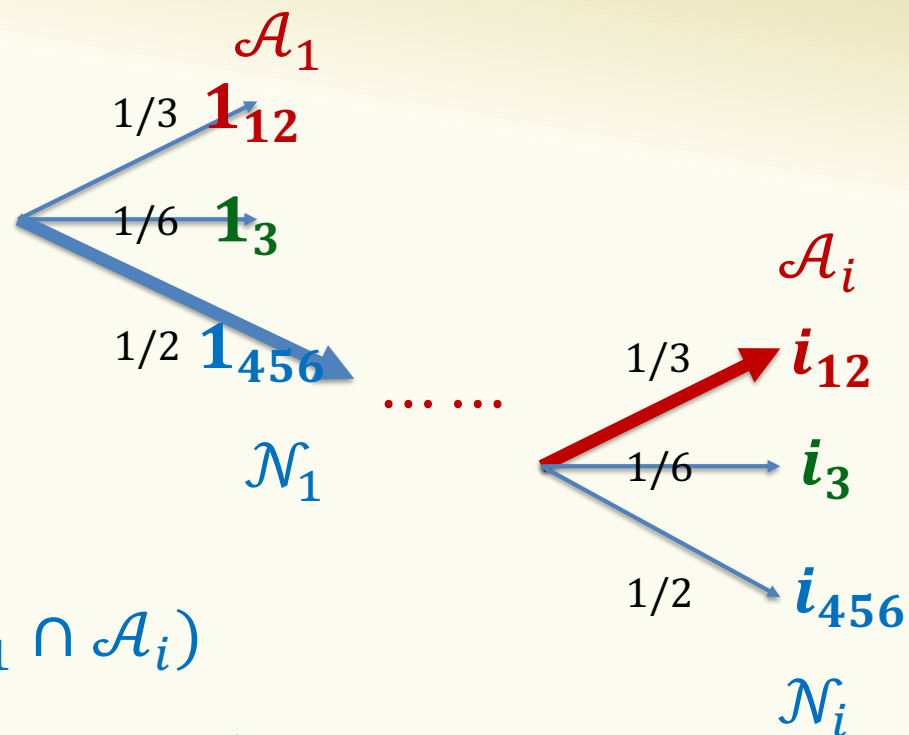
$$\mathbb{P}(\mathcal{A}_i) = \mathcal{P}(\mathcal{N}_1 \cap \mathcal{N}_2 \cap \dots \cap \mathcal{N}_{i-1} \cap \mathcal{A}_i)$$



## Sequential Process – Example

Events:

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



$$\begin{aligned}
 \mathbb{P}(\mathcal{A}_i) &= \mathcal{P}(\mathcal{N}_1 \cap \mathcal{N}_2 \cap \dots \cap \mathcal{N}_{i-1} \cap \mathcal{A}_i) \\
 &= \mathcal{P}(\mathcal{N}_1) \times \mathcal{P}(\mathcal{N}_2 | \mathcal{N}_1) \times \mathcal{P}(\mathcal{N}_3 | \mathcal{N}_1 \cap \mathcal{N}_2) \\
 &\quad \dots \times \mathcal{P}(\mathcal{N}_{i-1} | \mathcal{N}_1 \cap \mathcal{N}_2 \cap \dots \cap \mathcal{N}_{i-1}) \times \mathcal{P}(\mathcal{A}_i | \mathcal{N}_1 \cap \mathcal{N}_2 \cap \dots \cap \mathcal{N}_{i-1}) \\
 &= \left(\frac{1}{2}\right)^{i-1} \times \frac{1}{3}
 \end{aligned}$$

## Sequential Process -- Example

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

*What is the probability that Alice wins?*

## Sequential Process -- Example

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

*What is the probability that Alice wins?*

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

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

