CSE 312

Foundations of Computing II

Lecture 6: Chain Rule and Independence

Announcements

- Section tomorrow is important with new content that you will need on pset 3. Bring your laptops.
- **Anonymous questions: www.slido.com/1891306**

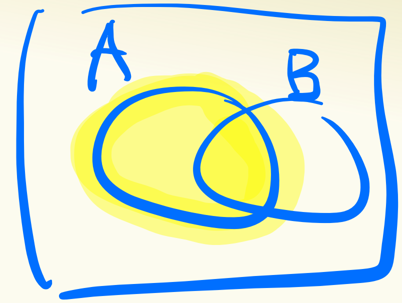
Agenda

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

Review Conditional & Total Probabilities

- **Conditional Probability**

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



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

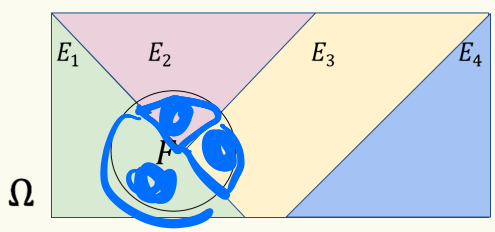
- **Bayes Theorem**

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

if $P(A) \neq 0, P(B) \neq 0$

- **Law of Total Probability**

$E_1, \ldots, E_n$ partition Ω



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

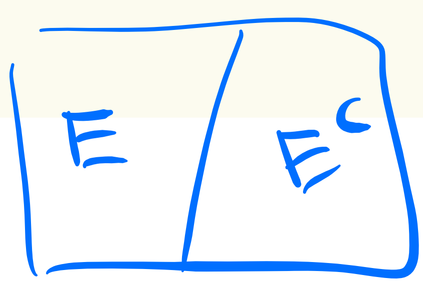
Bayes Theorem with Law of Total Probability

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

T test positive
Z have Zika.

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. $P(Z)=0.005$

$P(Z^c) = 1 - P(Z)$

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

Last time:

$$P(Z|T) = 0.33$$

How?

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

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

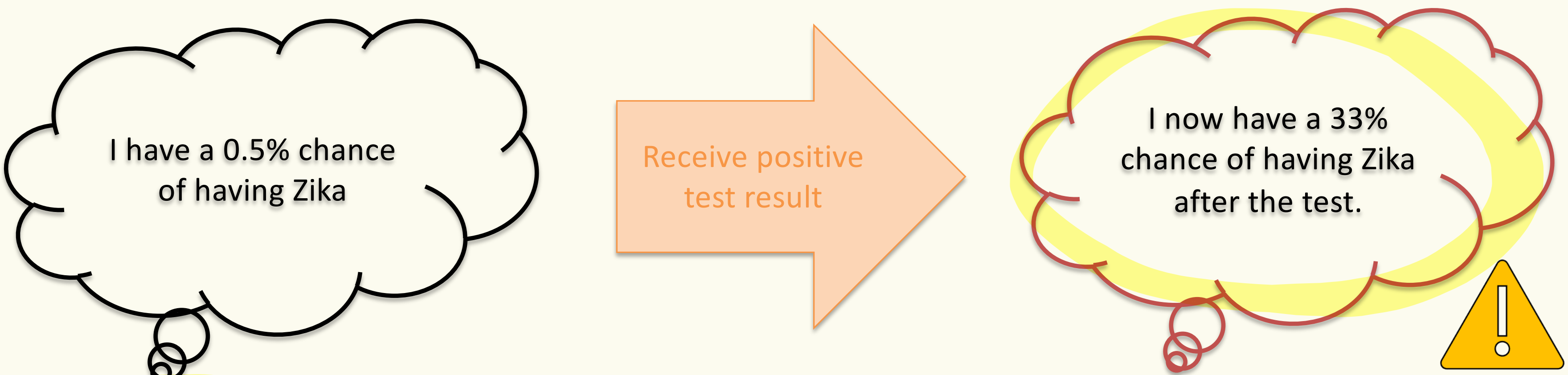
$P(T|Z) + P(T^c|Z) = 1$

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)

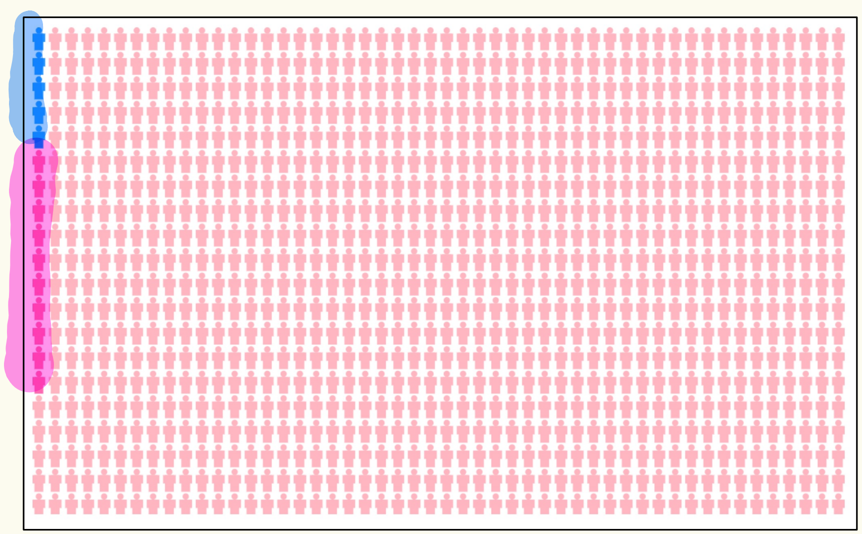
Posterior: P(Z|T)

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*).

$P(T|Z) = 1$



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**

$\Pr(T|Z^c) = \frac{10}{985+10}$

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. 5% have it.

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

$P(T|Z^c) = 10/995$ ≈ 0.01

$P(Z) = \frac{995}{1000} = 0.005$

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

$P(Z|T) = \frac{5}{5+10} = \frac{1}{3}$

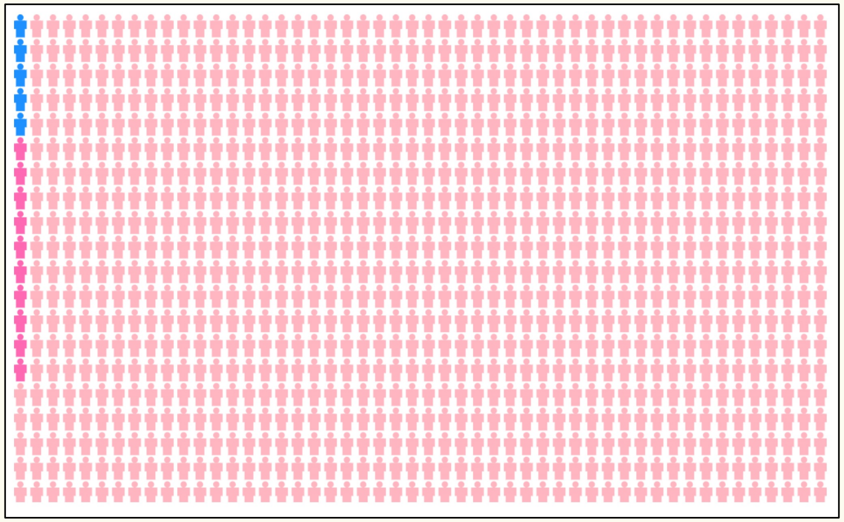
Example – Zika Testing

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 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) give you have Zika (event Z)?

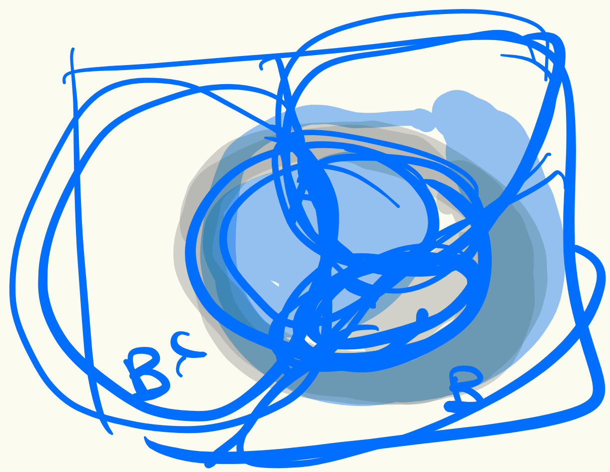
P(T|Z)

$$P(T^c|Z) = 1 - P(T|Z) = 0.02$$

Conditional Probability Define a Probability Space

The probability conditioned on A follows the same properties as (unconditional) probability.

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



$$P(B|A) + P(B^c|A)$$

$$= \frac{P(B\cap A)}{P(A)} + \frac{P(B^c \cap A)}{P(A)}$$

$$= \frac{P(B\cap A) + P(B^c\cap A)}{P(A)} = \frac{P(A)}{P(A)} = 1$$

Conditional Probability Define a Probability Space

The probability conditioned on A follows the same properties as (unconditional) probability.

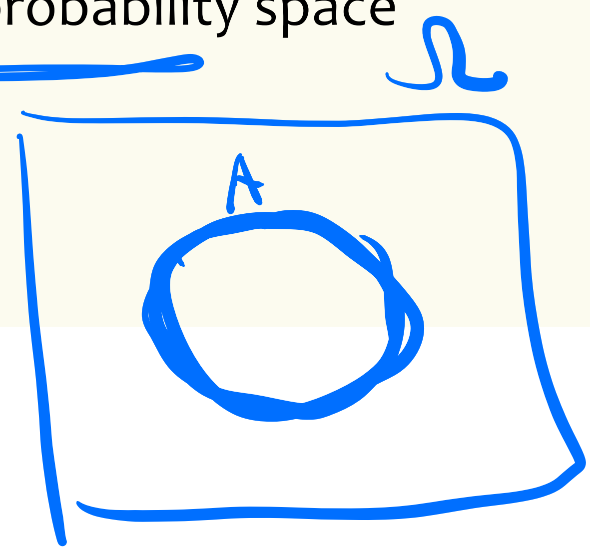
Example. $\mathbb{P}(\mathcal{B}^c|\mathcal{A}) = 1 - \mathbb{P}(\mathcal{B}|\mathcal{A})$

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

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

E

P(E|A)

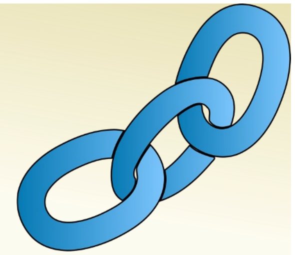


Agenda

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

NOTE See Lecture 6 Megathread
I fixed what's written on this slide after lecture

Chain Rule



Fixed what's written

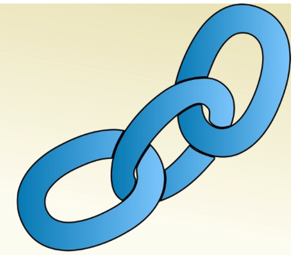
$$\mathbb{P}(\mathcal{B}|\mathcal{A}) = \frac{\mathbb{P}(\mathcal{A} \cap \mathcal{B})}{\mathbb{P}(\mathcal{A})} \Rightarrow \mathbb{P}(\mathcal{A})\mathbb{P}(\mathcal{B}|\mathcal{A}) = \mathbb{P}(\mathcal{A} \cap \mathcal{B})$$

$$P(\underbrace{A_1 \cap A_2}_{A} \cap \underbrace{A_3}_{B}) = \underbrace{P(A_1 \cap A_2)}_{P(A_1)P(A_2|A_1)} P(A_3 | A_1 \cap A_2)$$

$$= P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)$$

But order doesn't matter

Chain Rule



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

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

$$\mathbb{P}(\mathcal{A}_1 \cap \cdots \cap \mathcal{A}_n) = \mathbb{P}(\mathcal{A}_1) \cdot \mathbb{P}(\mathcal{A}_2|\mathcal{A}_1) \cdot \mathbb{P}(\mathcal{A}_3|\mathcal{A}_1 \cap \mathcal{A}_2) \cdots \mathbb{P}(\mathcal{A}_n|\mathcal{A}_1 \cap \mathcal{A}_2 \cap \cdots \cap \mathcal{A}_{n-1})$$

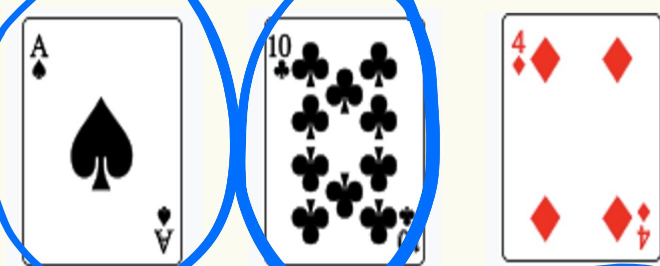
An easy way to remember: We have n tasks and we can do them sequentially, conditioning on the outcome of previous tasks

Chain Rule Example

$\Omega = \{$all possibilities for first 3 cards in order$\}$

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

$|\Omega| = 52 \cdot 51 \cdot 50$

What is P() = P(**A** ∩ **B** ∩ **C**)? $= \frac{1}{52 \cdot 51 \cdot 50}$

A: Ace of Spades First

B: 10 of Clubs Second

C: 4 of Diamonds Third

$P(A)\,P(B|A)\,P(C|A \cap B)$

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

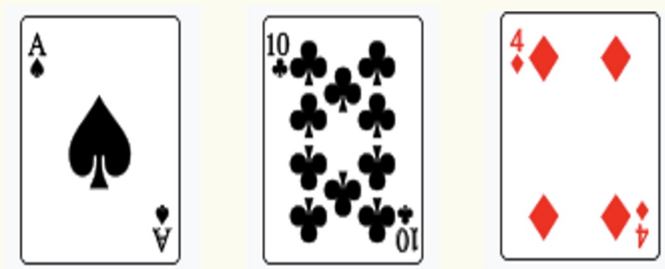
$P(B) = \frac{1}{52} = \frac{|B|}{52 \cdot 51 \cdot 50} = \frac{51 \cdot 50}{52 \cdot 51 \cdot 50}$

10C

$= \frac{1}{52}$

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

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

$|\Omega| = 8$

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}

$Pr(A) = \frac{4}{8} = \frac{1}{2}$

$Pr(B) = \frac{7}{8}$

Independent?

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

$$\frac{3}{8} \neq \frac{1}{2} \cdot \frac{7}{8}$$

$\frac{1}{2}$ A first toss H

$\frac{1}{2}$ B 2nd toss H

{HHH, HHT, HTH, HTT}

$A \cap B = \{HHT, HHH\}$

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

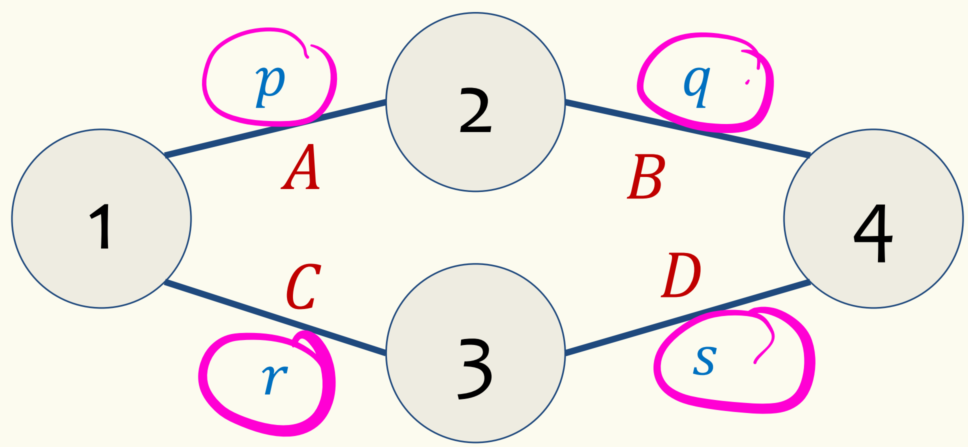
$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i).$$

Example – Network Communication

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

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

$$P(A) = p$$
$$P(B) = q$$
$$P(C) = r$$
$$P(D) = s$$



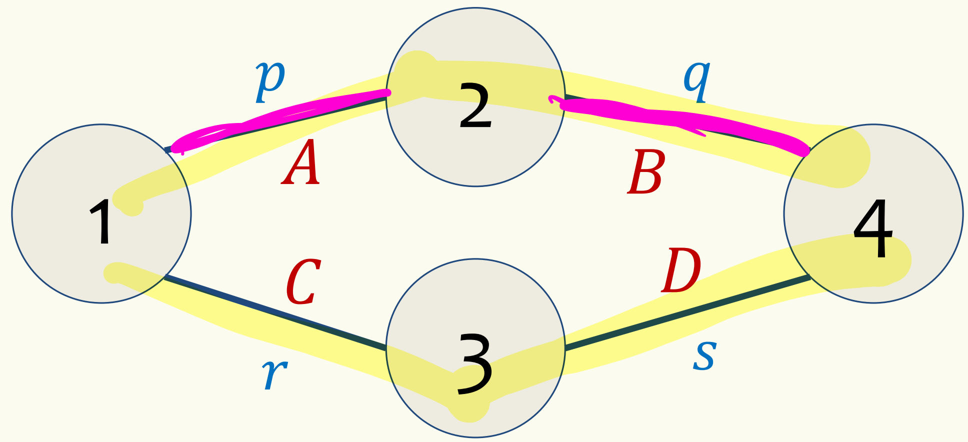
Example – Network Communication

$P(A) = p$
$P(B) = q$
$P(C) = r$
$P(D) = s$

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

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

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



$P(A \cap B \text{ or } C \cap D)$

$= P(A \cap B) + P(C \cap D) - P(A \cap B \cap C \cap D)$

$P(A)P(B) = p \cdot q$, $\quad r \cdot s$, $\quad P(A)P(B)P(C)P(D) = p \cdot q \cdot r \cdot s$

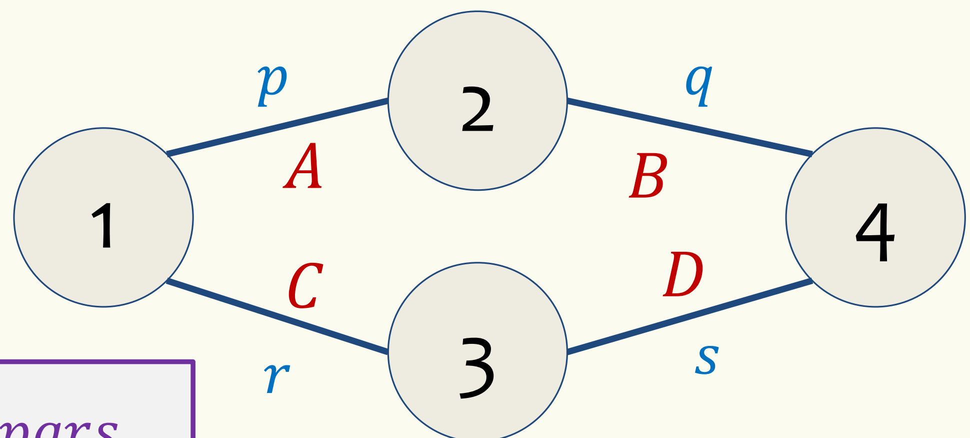
$P(A) = 0.3$, $P(B) = 0.4$ $\neq$ $P(A \cap B) = 0$

Example – Network Communication

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

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

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



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

Independence – Another Look

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

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

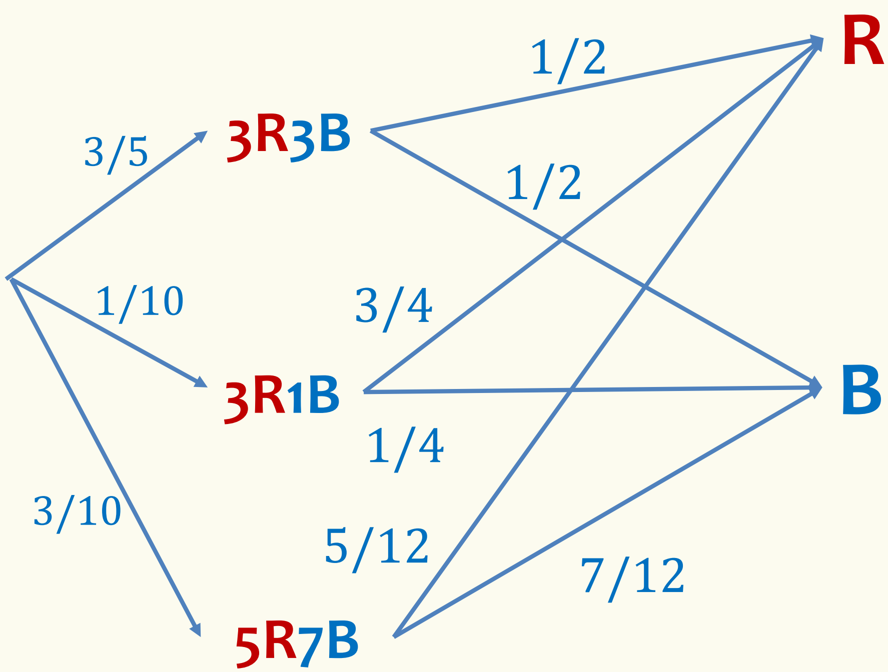
"Equivalently." $P(A|B) = P(A)$.

Events generated independently ➔ *their probabilities satisfy independence*

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

This can be counterintuitive!

Sequential Process



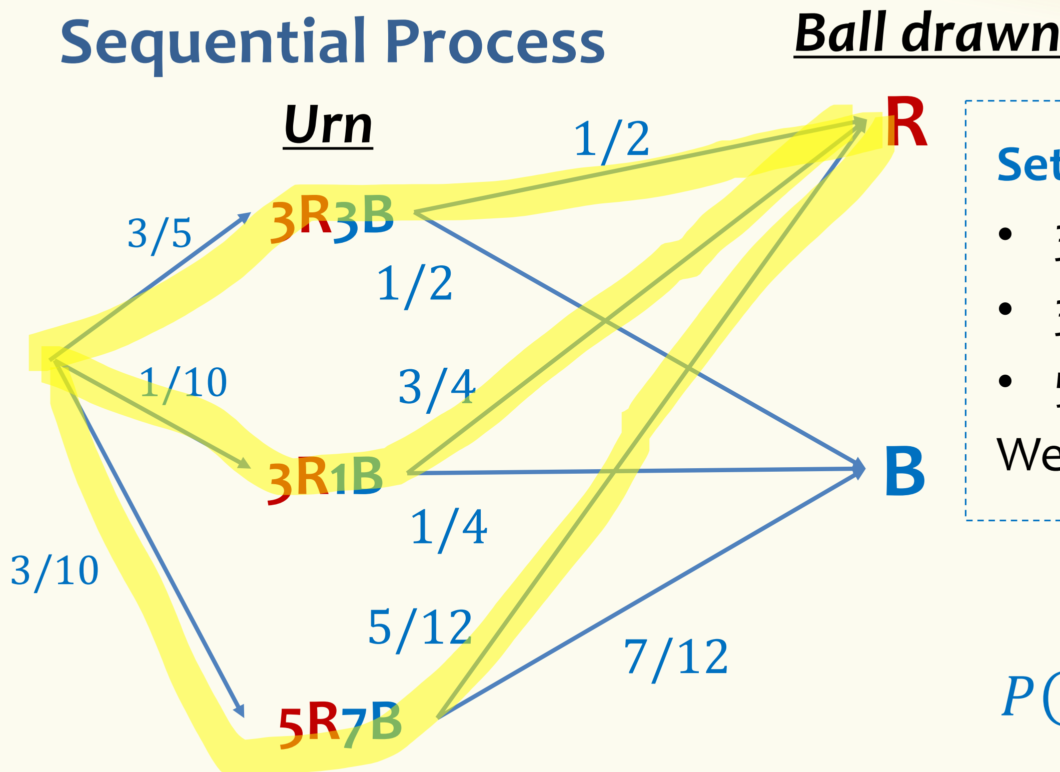
Setting: An urn contains:

- 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?

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.

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

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

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

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 fips. Suppose it is tossed 3 times.

$\mathbb{P}(HHH) =$ $\frac{2}{3} \cdot \frac{2}{3} \cdot \frac{2}{3}$ using

$\mathbb{P}(TTT) =$

$\mathbb{P}(HTT) =$ $\frac{2}{3} \cdot \frac{1}{3} \cdot \frac{1}{3}$

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.

$\mathbb{P}(2\ heads\ in\ 3\ tosses) =$ $Pr(\{HHT, HTH, THH\})$

$\left(\frac{2}{3}\right)^2 \frac{1}{3}$

$=$

$= 3 \cdot \left(\frac{2}{3}\right)^2 \frac{1}{3}$

HTH

TTT

TTH

THH

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

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

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 coin randomly with equal probability and flip that coin 3 times independently. What is the probability we get all heads?

$$P(HHH) = P(HHH \mid C_1) \cdot P(C_1) + P(HHH \mid C_2) \cdot P(C_2)$$

Law of Total Probability (LTP)

$$= P(H \mid C_1)^3 \, P(C_1) + P(H \mid C_2)^3 \, P(C_2)$$

Conditional Independence

$$= 0.3^3 \cdot 0.5 + 0.9^3 \cdot 0.5 = 0.378$$

C_i = coin i was selected

Agenda

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

Often probability space (Ω, 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

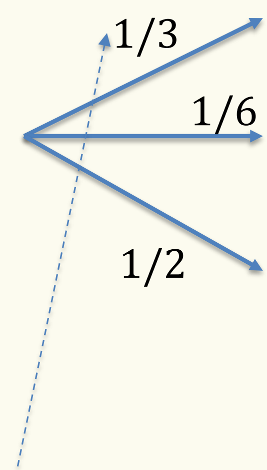
What is Pr(Alice wins on 1^{st} round) =
Pr(Alice wins on 2^{st} round) =
…
Pr(Alice wins on i^{th} round) = ?
Pr(Alice wins) = ?

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, toss a die

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



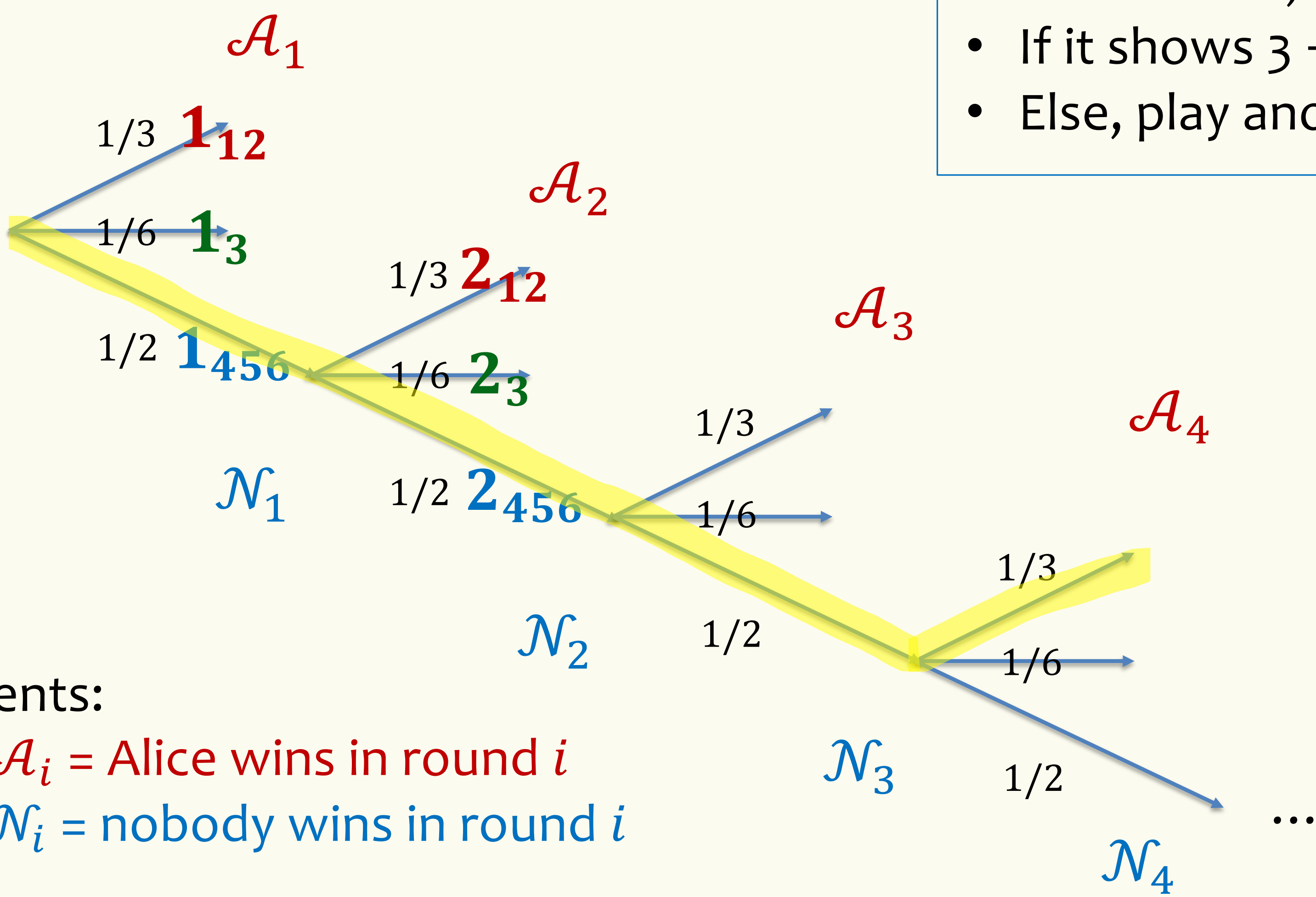
$\boldsymbol{i_{12}}$ = i th round toss 1 or 2

$\boldsymbol{i_{3}}$ = 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

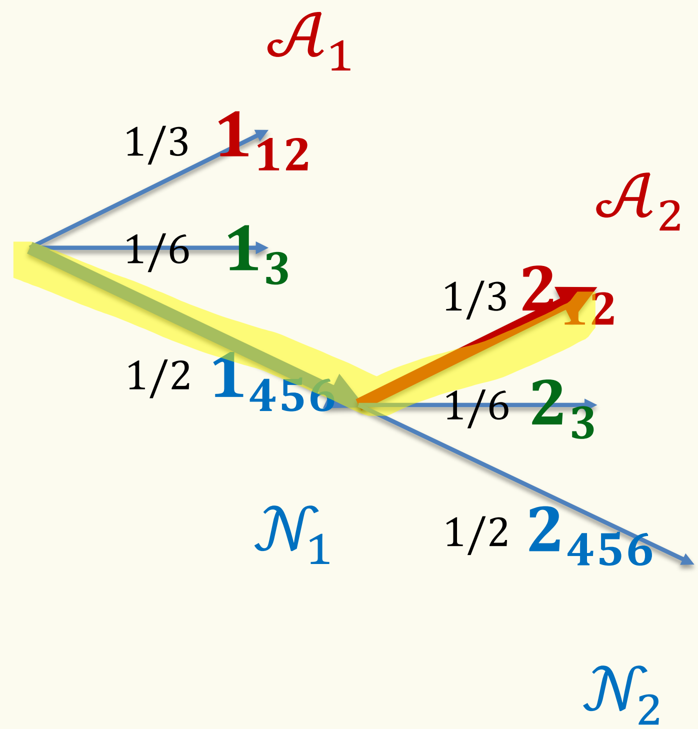
Events:
- $\mathcal{A}_i$ = Alice wins in round i
- $\mathcal{N}_i$ = nobody wins in round i

Sequential Process – Example

Events:

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

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

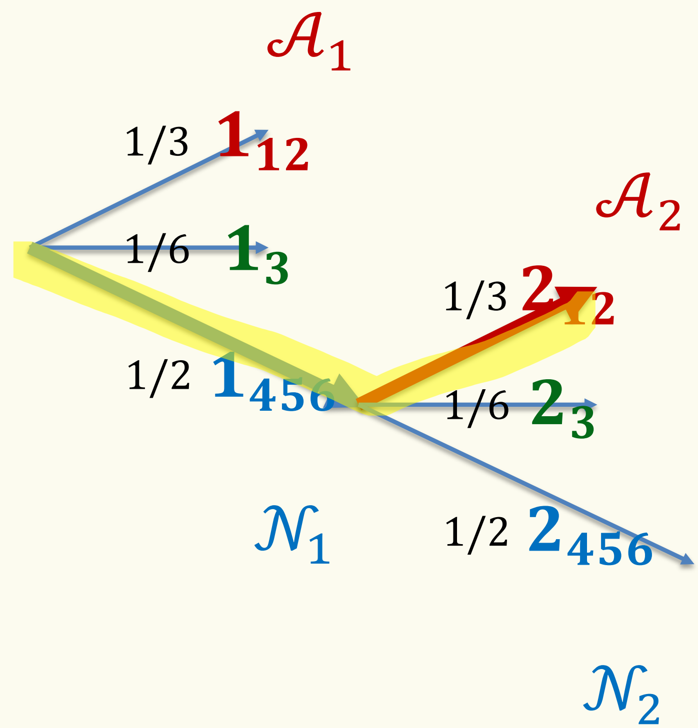


2[nd] roll indep of 1[st] roll

Sequential Process – Example

Events:

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



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

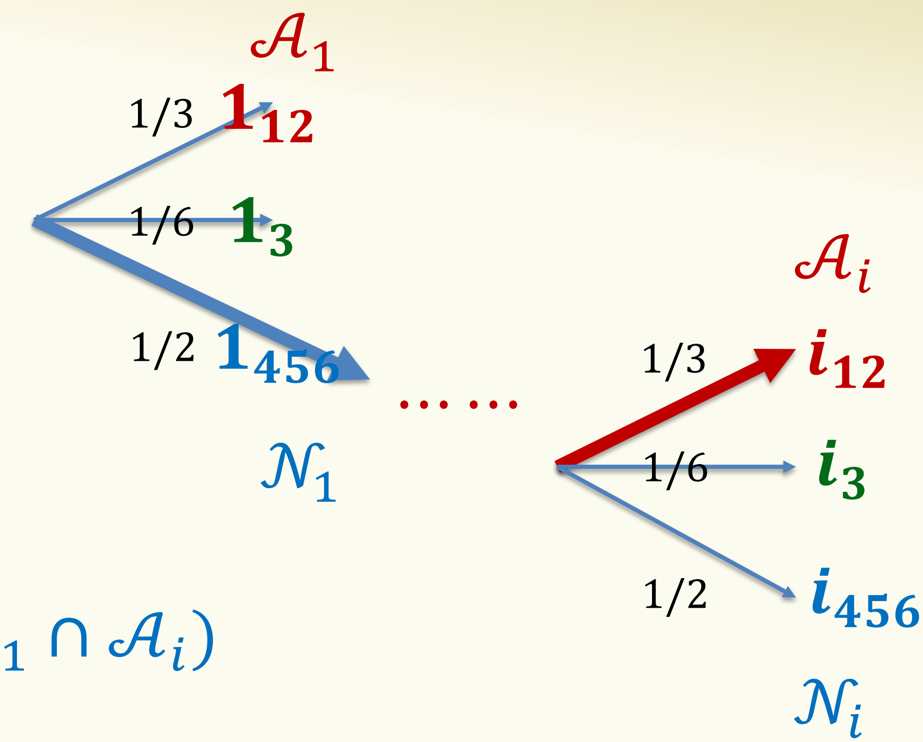
The event $\mathcal{A}_2$ implies $\mathcal{N}_1$, and this means that $\mathcal{A}_2 \cap \mathcal{N}_1 = \mathcal{A}_2$

2nd roll indep of 1st roll

Sequential Process – Example

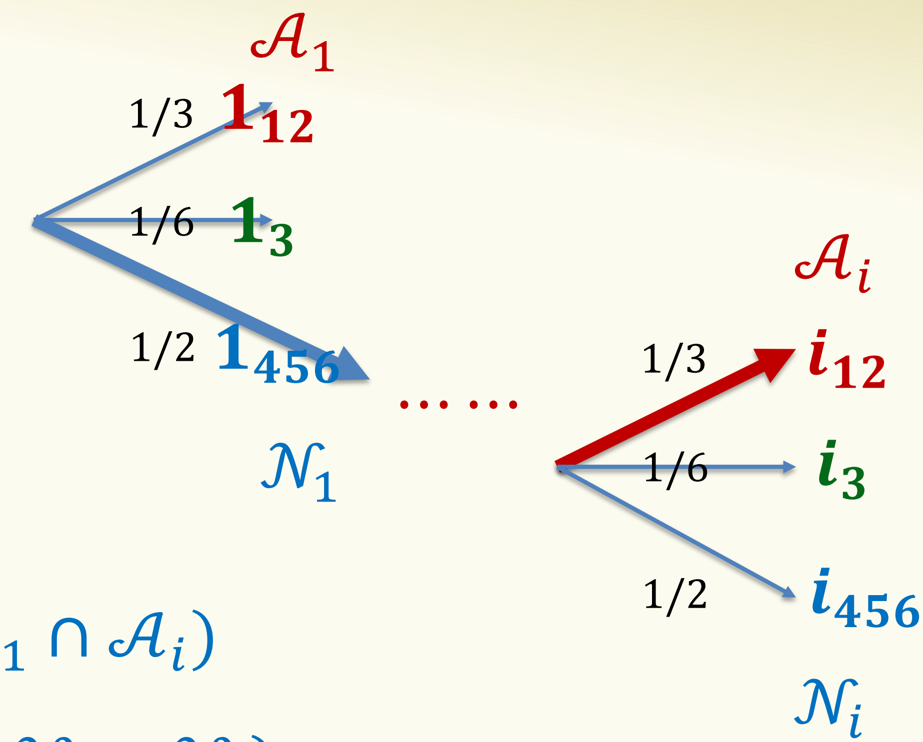
Events:

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



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

Sequential Process – Example



Events:

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

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

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

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

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

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 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?

Assuming independence doesn't justify the assumption!

Both chutes could fail because of the same rare event e.g., freezing rain.