CSE 312

Foundations of Computing II

Lecture 6: Chain Rule and Independence

PAUL G. ALLEN SCHOOL
OF COMPUTER SCIENCE & ENGINEERING

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

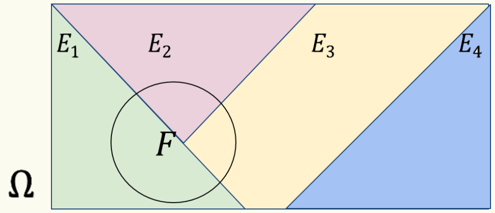
- **Conditional Probability**

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

- **Bayes Theorem**

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

- **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).

$P(T \cap Z) + P(T \cap \bar{Z})$

Example – Zika Testing

Z have Zika
T test positive

Suppose we know the following Zika stats

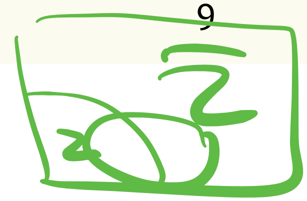
- A test is 98% effective at detecting Zika ("true positive") $P(T|Z) = 0.98$ $\frac{P(T \cap Z)}{P(Z)}$
- However, the test may yield a "false positive" 1% of the time $P(T|Z^c)=0.01$
- 0.5% of the US population has Zika. $P(Z)=0.005$

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

By Bayes Rule, $P(Z|T) = \dfrac{P(T|Z)P(Z)}{P(T)} = \dfrac{0.98 \cdot 0.005}{P(T)} \approx 0.33$

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

$0.98 \cdot 0.005 \quad 0.01 \cdot 0.995$

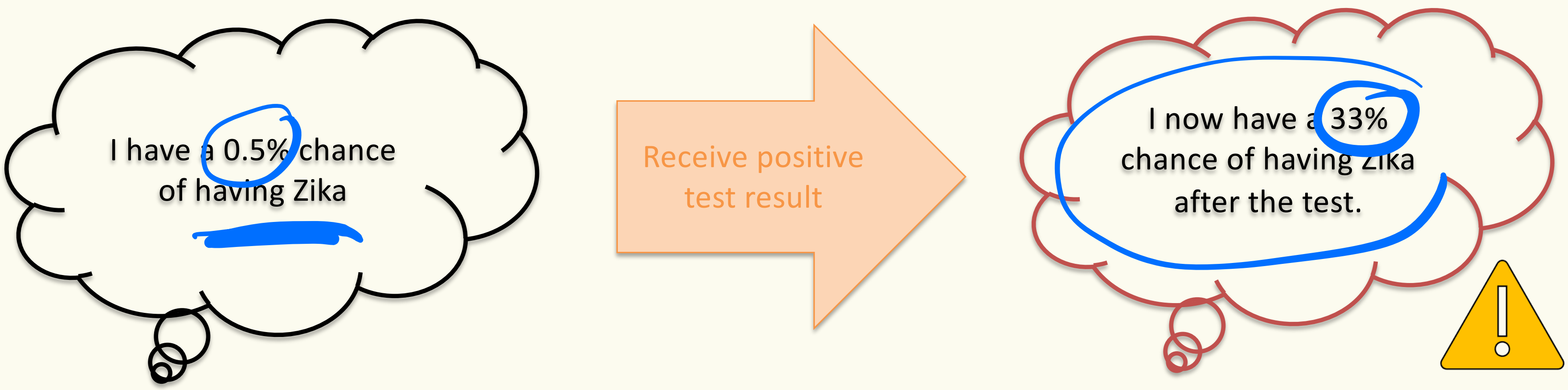


Philosophy – Updating Beliefs

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)

$P(T|Z) = 1$

Example – Zika Testing

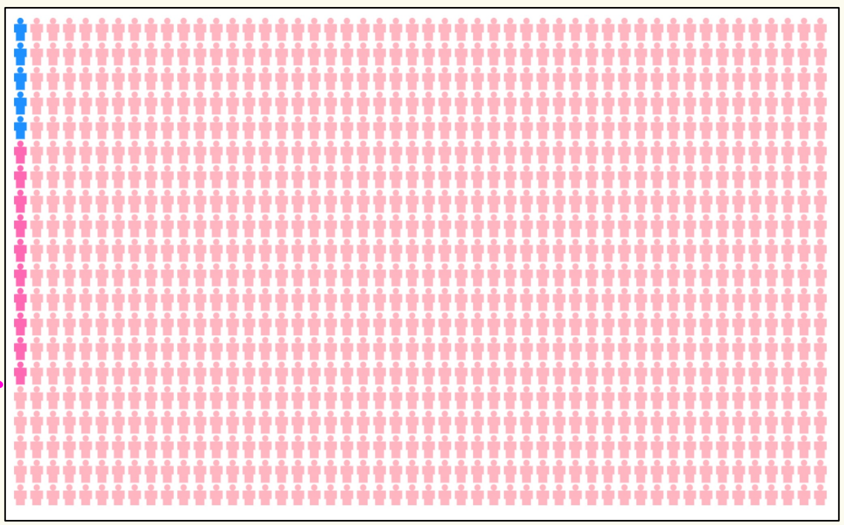
Have zika blue, don't pink

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

$P(T|Z) = 1$

$P(T|Z^c) = \frac{10}{995} \approx 0.01$

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



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

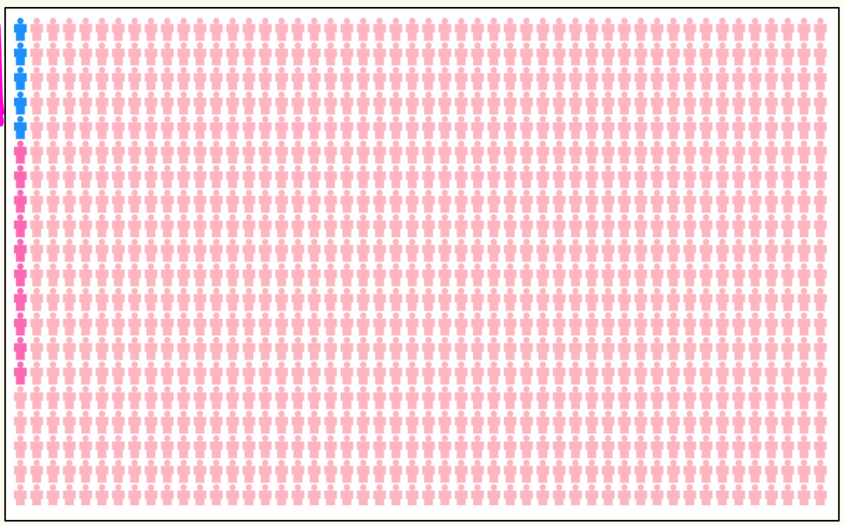
Have zika blue, don't pink

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.

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

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**

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

Example – Zika Testing

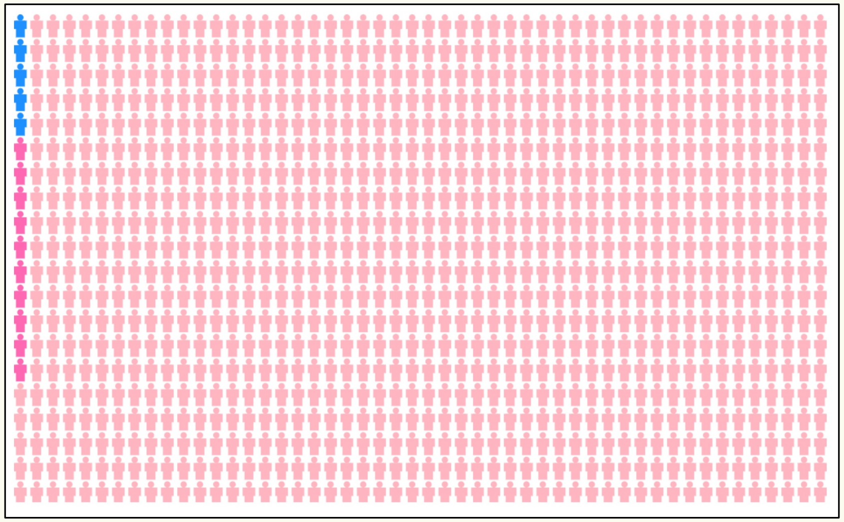
Have zika blue, don't pink

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") $\boxed{P(T|Z)} = 0.98$
- 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)?

$$P(\overline{T}|Z) = 1 - P(T|Z) = 0.02$$

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 + $\mathbb{P}(\mathcal{A}) > 0$

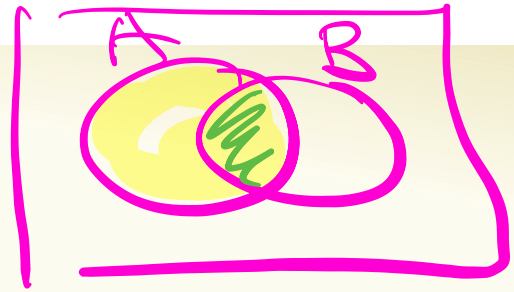
$\Rightarrow$ $(\mathcal{A}, \mathbb{P}(\cdot\,|\mathcal{A}))$ is a probability space

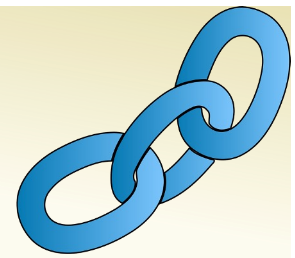
Axiom 1 (Non-negativity): $P(E) \geq 0$.
Axiom 2 (Normalization): $P(\Omega) = 1$
Axiom 3 (Countable Additivity): If E and F are mutually exclusive, then $P(E \cup F) = P(E) + P(F)$

Agenda

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

Chain Rule



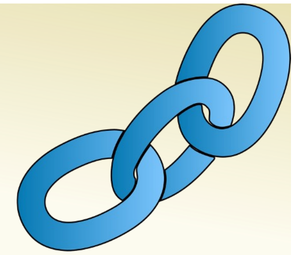


$$\boxed{\mathbb{P}(\mathcal{B}|\mathcal{A}) = \frac{\mathbb{P}(\mathcal{A}\cap\mathcal{B})}{\mathbb{P}(\mathcal{A})}} \Rightarrow \mathbb{P}(\mathcal{A})\mathbb{P}(\mathcal{B}|\mathcal{A}) = \mathbb{P}(\mathcal{A}\cap\mathcal{B})$$

$$P(A\cap B\cap C) = P(A)\underline{P(B|A)}\,\underline{P(C|A\cap B)}$$

$$= \cancel{P(A)}\,\frac{\cancel{P(A\cap B)}}{\cancel{P(A)}}\,\frac{P(A\cap B\cap C)}{\cancel{P(A\cap B)}}$$

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 $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$,

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

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

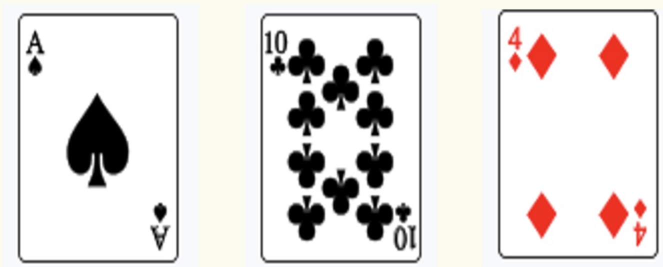
An easy way to remember: We have n tasks and we can do them sequentially, conditioning on the outcome of previous tasks

10c

$$P(\text{seq}) = \frac{1}{52!}$$

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**)?

1 2 3

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

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

$$= \frac{1}{52} \qquad = \frac{1}{51} \qquad = \frac{1}{50}$$

$$P(10 \text{ of clubs}) \neq P(10 \text{ club} \mid A♠ \text{ 1st}) = \frac{1}{51}$$

51!

$52!$

$$\frac{P(\text{perm } (A♠\ 10♣) \_\_)}{P(\text{first card } A♠)}$$

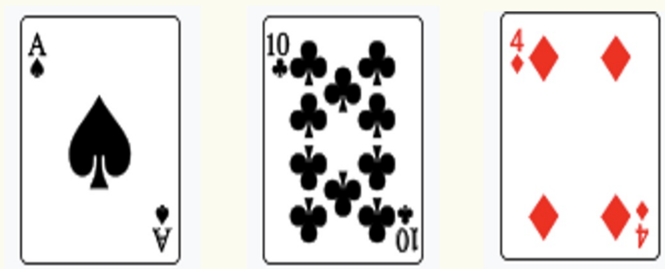
$\frac{50!}{52!}$

$\frac{1}{52}$

$= \frac{1}{51}$

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

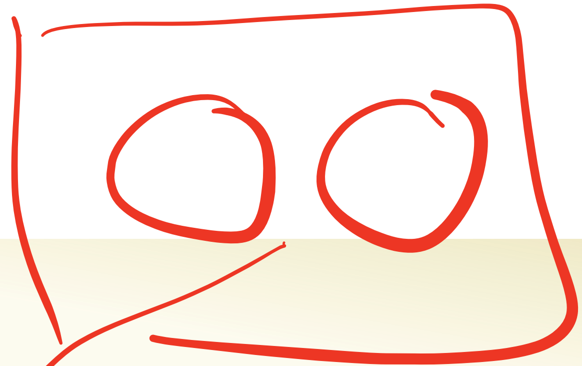
C: 4 of Diamonds Third

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

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

Agenda

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



Independence

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

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

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

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?

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

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

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

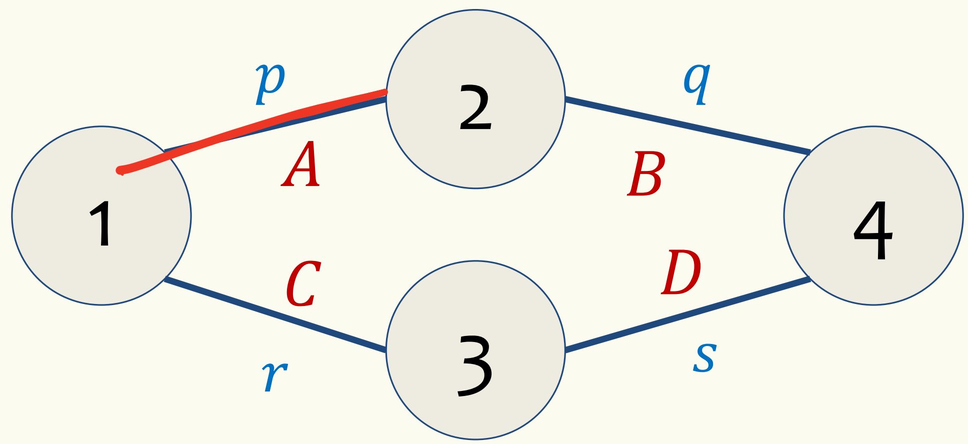
$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i).$$

Example – Network Communication

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

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

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



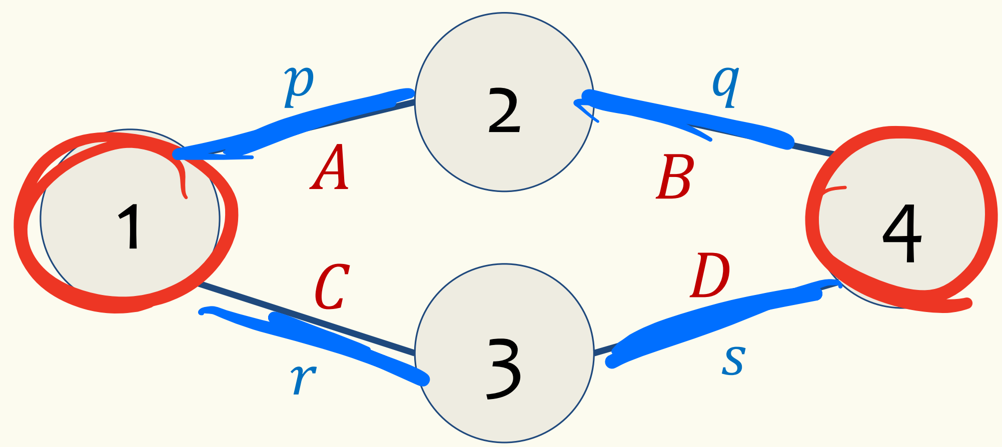
Example – Network Communication

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

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

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

What is P(1-4 connected)?



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

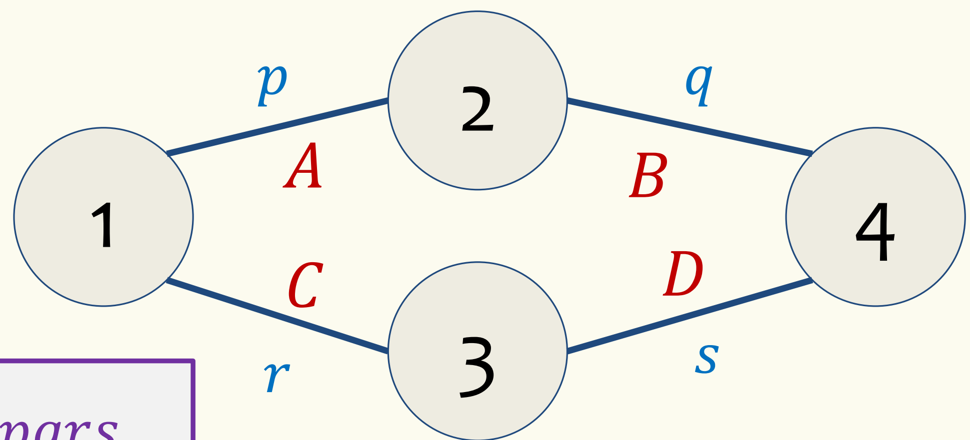
$$= \underbrace{P(A \cap B)}_{P(A)\cdot P(B) \;\; p\cdot q} + \underbrace{P(C \cap D)}_{P(C)\cdot P(D) \;\; r\cdot s} - \underbrace{P(A \cap B \cap C \cap D)}_{p\cdot q\cdot r\cdot s}$$

Example – Network Communication

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

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

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



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

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 $\Rightarrow$ *their probabilities satisfy independence*

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

$\cancel{\Leftarrow}$

This can be counterintuitive!

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.

$P(H_1)P(H_2)P(H_3)$

$\mathbb{P}(HHH) = \frac{2}{3} \cdot \frac{2}{3} \cdot \frac{2}{3} = \left(\frac{2}{3}\right)^3$

$\mathbb{P}(TTT) = \left(\frac{1}{3}\right)^3$

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

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.

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

$$P(\underline{HHT} \cup \underline{HTH} \cup THH)$$

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

$$P(k \text{ heads in } n \text{ tosses}) = \binom{n}{k} \left(\frac{2}{3}\right)^k \left(\frac{1}{3}\right)^{n-k}$$

slido.com/4171468

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

HHHT HTTT

$\boxed{HHHH\ TTTTT}$

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