CSE 312

Foundations of Computing II

Lecture 6: Bayesian Inference, Chain Rule, Independence

Review Conditional & Total Probabilities

- **Conditional Probability**

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

$P(A \cap B)$

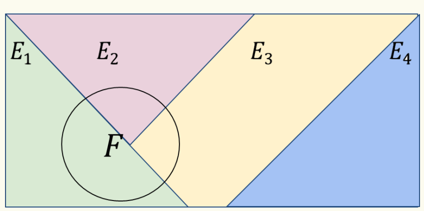
$= P(B \cap A)$

- **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**

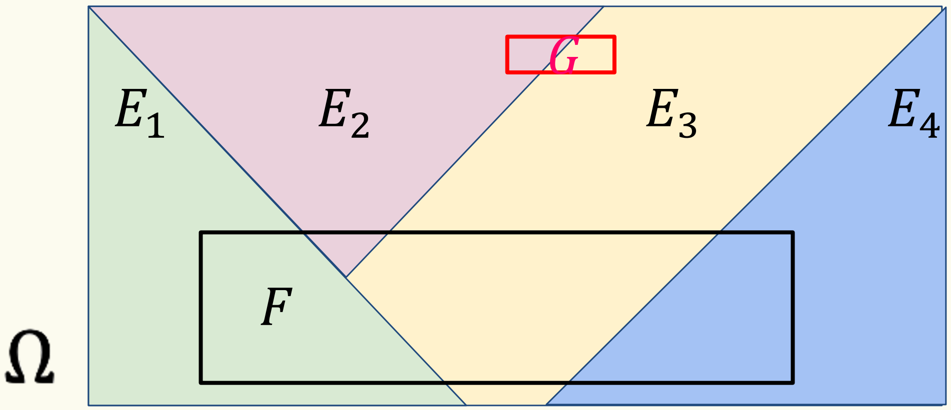
$E_1, \ldots, E_n$ partition Ω



Ω

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

Conditional Probability

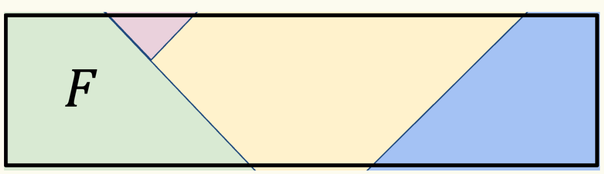


$P(X | B)$

$P(X | A)$

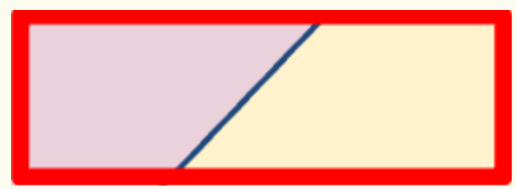
e.g. $P(E_3) = 0.3$

$P(E_4) = 0.25$



e.g. $P(E_3|F) = 0.4$

$P(E_4|F) = 0.3$



e.g. $P(E_3|G) = 0.5$

$P(E_3|G) = 0$

Conditional Probability Defines a Probability Space

Formally. (Ω, P) is a probability space and $P(\mathcal{A}) > 0$

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

The probability conditioned on $\mathcal{A}$ follows the same properties as (unconditional) probability.

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

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

Example – Zika Testing

Suppose we know the following Zika stats

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

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

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

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

$$= \frac{98}{100} \cdot \frac{5}{1000} + \frac{1}{100} \cdot \frac{995}{1000} = \frac{490}{100000} + \frac{995}{100000}$$

What is the probability that you do not have Zika (event Z^c)?

$P(Z^c) = 1 - P(Z) = 99.5\%$

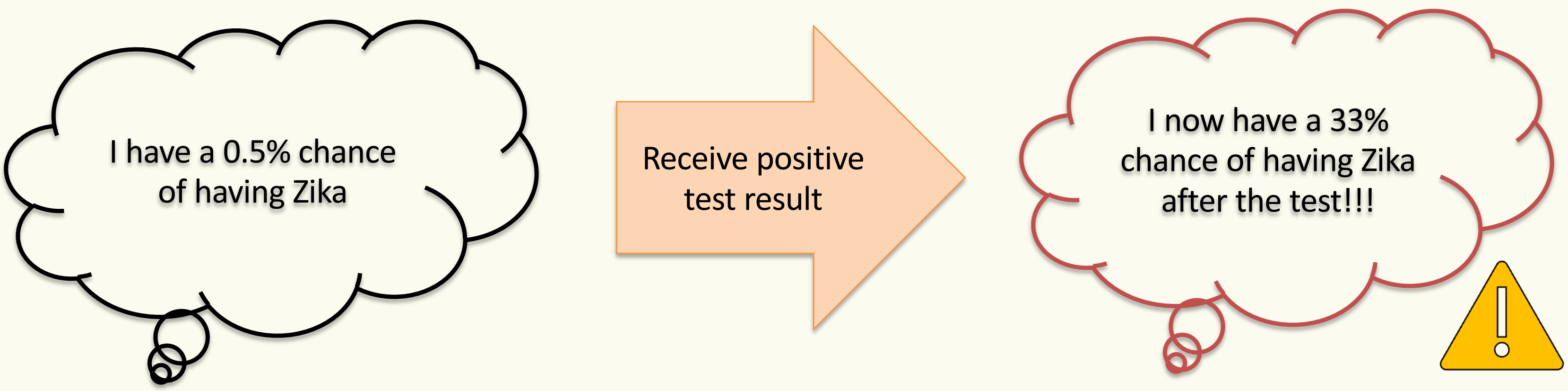
So, $P(Z|T) \approx 33\ \%$

Philosophy – Updating Beliefs

Your beliefs changed **drastically**

Z = you have Zika

T = you test positive for Zika



Prior: $P(Z)$

Posterior: $P(Z|T)$

What happened ?

The test seems excellent!

- Effectiveness $P(T|Z) = 0.98$

(False negative $P(T^c|Z) = 0.02$)

- False positive $P(T|Z^c) = 0.01$

But conditioned on positive test

Chance of actually having Zika is only $P(Z|T) = 0.33$

$$P(Z|T) = \frac{P(T|Z)P(Z)}{P(T)} = \frac{P(T \cap Z)}{P(T)}$$

Observe the ratio $\frac{P(Z|T)}{P(T|Z)} = \frac{P(Z)}{P(T)}$ can be very large in general

Example – Zika Testing [Demo]

$$P(Z|T) = \frac{P(T \cap Z)}{P(T)} = \frac{490}{490 + 995} \approx 0.33$$

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

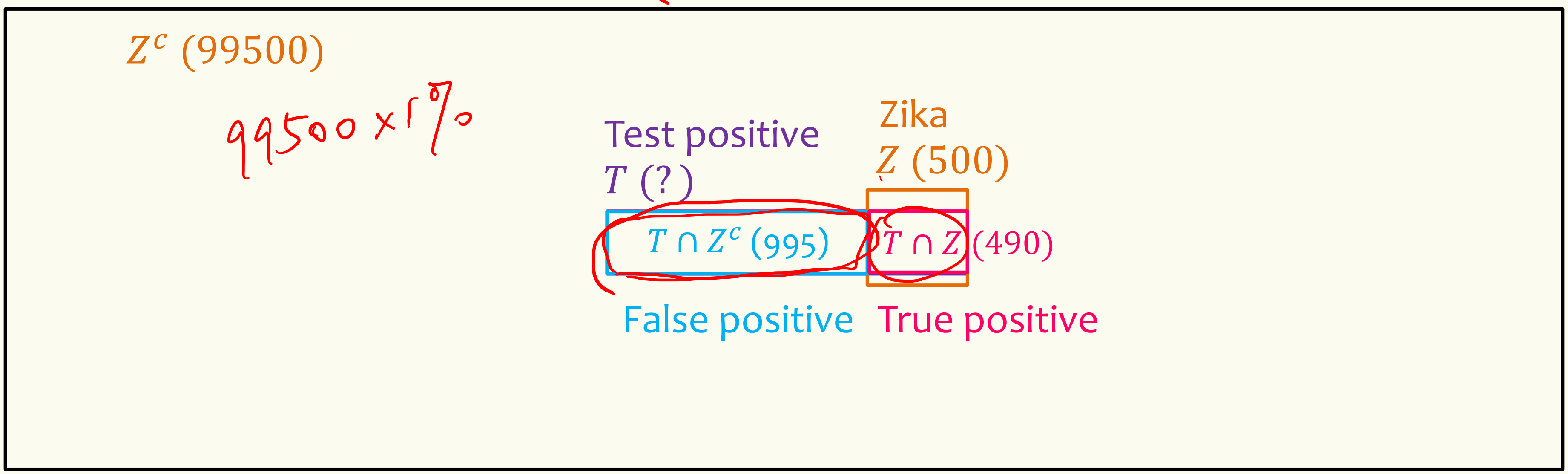
500 have Zika (0.5%)
99,500 do not

Suppose we had 100,000 people:

- **500 have Zika**
- 490 **have Zika** and **test positive**
- 995 **do not have Zika** and **test positive**

98% of those with Zika

1% of those without Zika



Ω=US population (100,000)

Take Home Exercise – Zika Testing

Suppose we know the following Zika stats

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

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

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

Take Home Exercise, Solution – Zika Testing

Suppose we know the following Zika stats

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

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

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

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

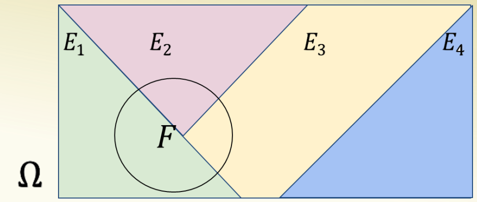
$$= \frac{2}{100} \cdot \frac{5}{1000} + \left(1 - \frac{1}{100}\right) \cdot \frac{995}{1000} = \frac{10}{100000} + \frac{98505}{100000}$$

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

$P(T^c|Z) = 1 - P(T|Z) = 2\%$

So, $P(Z|T^c) = \frac{10}{10+98505} \approx 0.01\ \%$

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,

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

We just used this implicity on the Zika test example with $E = Z$ and $F = T$

Simple Partition: In particu ... o probability, then

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

Our First Machine Learning Task: Spam Filtering

```
Subject: “FREE $$$ CLICK HERE”
```

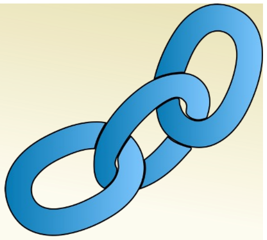
What is the probability this email is spam, given the subject contains “FREE”?
Some useful stats:

- 10% of ham (i.e., not spam) emails contain the word “FREE” in the subject.
- 70% of spam emails contain the word “FREE” in the subject.
- 80% of emails you receive are spam.

Agenda

- Bayes Theorem + Law of Total Probability
- Chain Rule
- Independence
- Infinite process and Von Neumann’s trick
- Conditional independence

Chain Rule



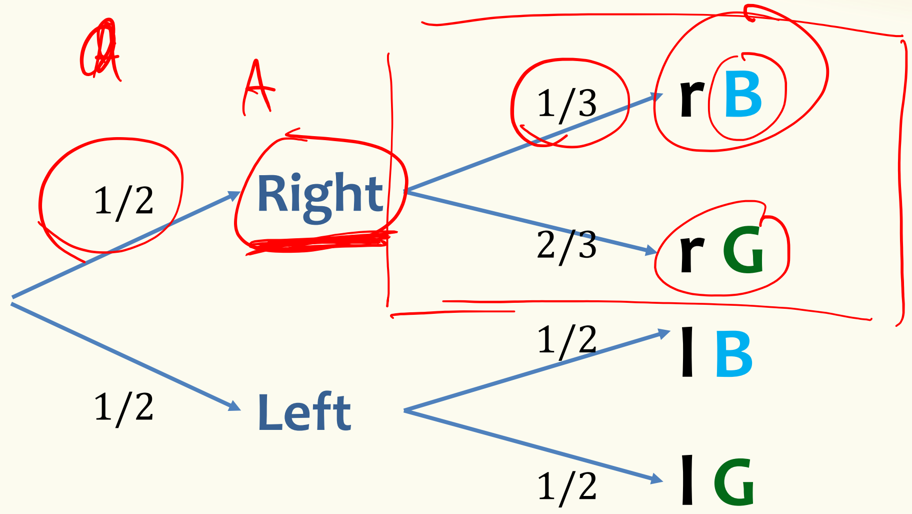
$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

$$\Rightarrow$$

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

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

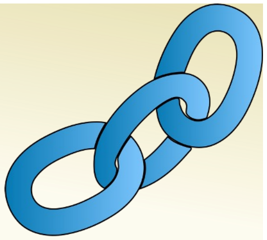
Recall from last time:



$$P(\mathbf{B}) = P(\mathbf{Left}) \times P(\mathbf{B}|\mathbf{Left}) + P(\mathbf{Right}) \times P(\mathbf{B}|\mathbf{Right})$$

What if we have more than two (e.g., n) steps?

Chain Rule



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

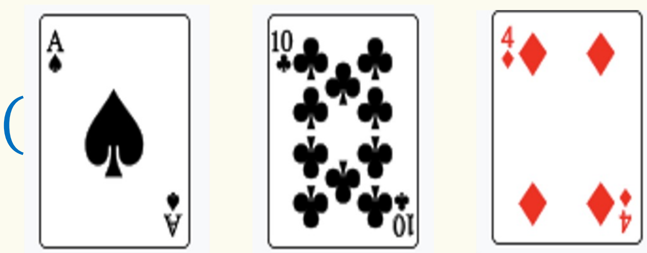
Theorem. (Chain Rule) For events $A_1, A_2, \ldots, A_n$,

$$P(A_1 \cap \cdots \cap A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2)$$
$$\cdots P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1})$$

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

Shuffle a standard 52-card deck and draw the top 3 cards.
(uniform probability space)

What is $P($  $) = P(A \cap B \cap C)$?

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

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

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

Agenda

- Bayes Theorem + Law of Total Probability
- Chain Rule
- Independence
- Infinite process and Von Neumann's trick
- Conditional independence

Independence

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

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

Equivalent formulations:

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

"The probability that B occurs after observing A" – Posterior
= "The probability that B occurs" – Prior

Independence - Example

Assume we toss two fair coins

"first coin is heads" $A = \{\mathrm{HH}, \mathrm{HT}\}$

"second coin is heads" $B = \{\mathrm{HH}, \mathrm{TH}\}$

$$P(A) = 2 \times \frac{1}{4} = \frac{1}{2}$$

$$P(B) = 2 \times \frac{1}{4} = \frac{1}{2}$$

$$P(A \cap B) = P(\{HH\}) = \frac{1}{4} = P(A) \cdot P(B)$$

Example – Independence

Toss a coin 3 times. Each of 8 outcomes equally likely.

- A = {at most one T} $= \{HHH, \underbrace{HHT, HTH, THH}_{A \cap B}\}$
- B = {at most 2 $H's$} $= \{HHH\}^c$

Independent?

$$\overset{\frac{3}{8}}{P(A \cap B)} \overset{?}{=} \overset{\frac{4}{8}}{P(A)} \cdot \overset{\frac{7}{8}}{P(B)}$$

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

Poll:

A. Yes, independent 10

B. No 60

pollev/rachel312

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

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

$n = 3$

$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2)$

$\vdots$

$A_2 \cap A_3$

$A_3 \cap A_1$

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

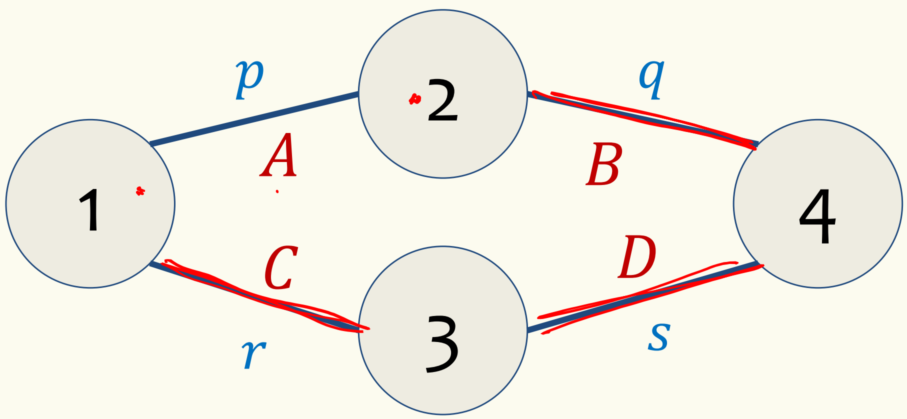
$= P(A_1) P(A_2) P(A_3)$.

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

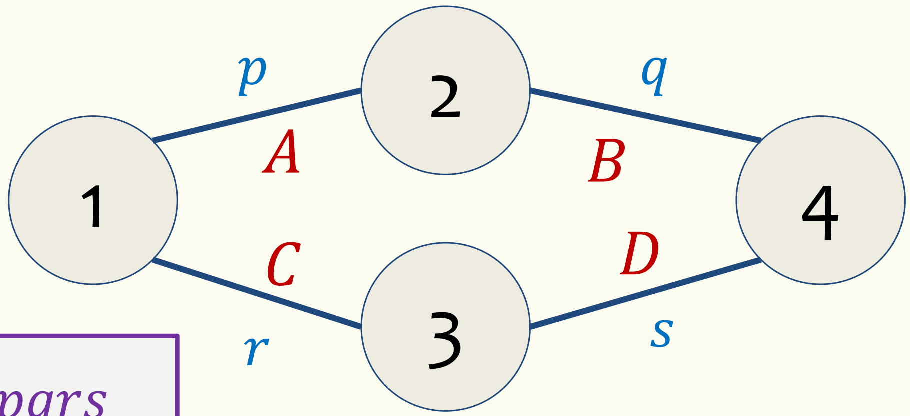
If each link works with the probability given, **independently**:
What's the probability that nodes 1 and 4 can communicate?

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

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

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

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



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

Independence 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.

Independence – Another Look

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

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

"Equivalently." If $P(B) \neq 0$, $P(A|B) = P(A)$.

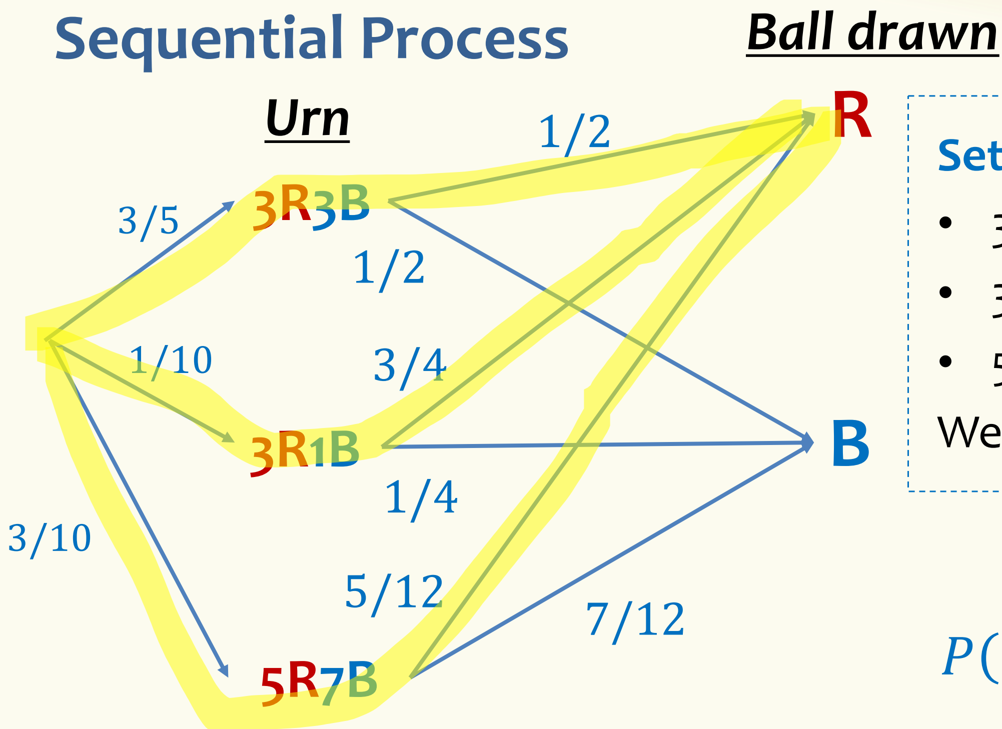
It is important to understand that independence is a property of probabilities of outcomes, not of the root cause generating these events.

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

$\not\Leftarrow$ *Not necessarily*

This can be counterintuitive!

Sequential Process



Are R and 3R3B independent?

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.

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

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

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

Agenda

- Bayes Theorem + Law of Total Probability
- Chain Rule
- Independence
- Infinite process and Von Neumann’s trick
- Conditional independence

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?

C_i = coin i was selected

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

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

Example – Throwing Dies

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 4^{st} round) ?

Often probability space $(\Omega, \mathbb{P})$ is given **implicitly** of the following form, using chain rule and/or independence

Experiment proceeds in n sequential steps, each step follows some local rules defined by conditional probability and independence.

– Allows for easy definition of experiments where $|\Omega| = \infty$ as in the previous game