CSE 312

Foundations of Computing II

Lecture 6: Bayesian Inference, Chain Rule, Independence

Review Conditional & Total Probabilities

- **Conditional Probability**

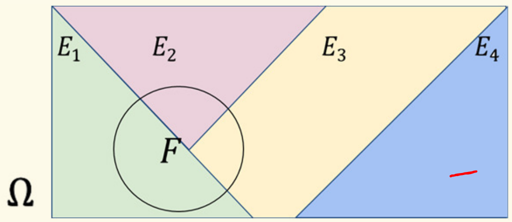
$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

- **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**

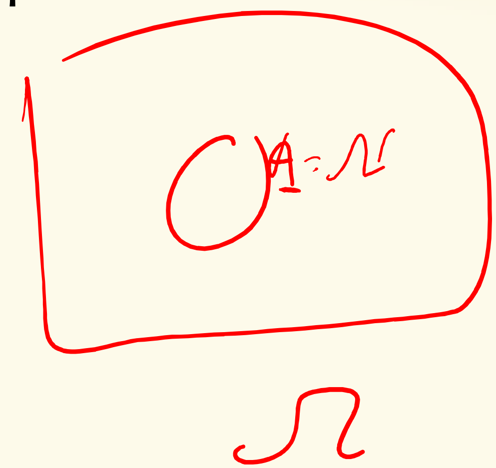
$E_1, \ldots, E_n$ partition Ω



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

Conditional Probability Defines a Probability Space

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



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

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

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

Agenda

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

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).?

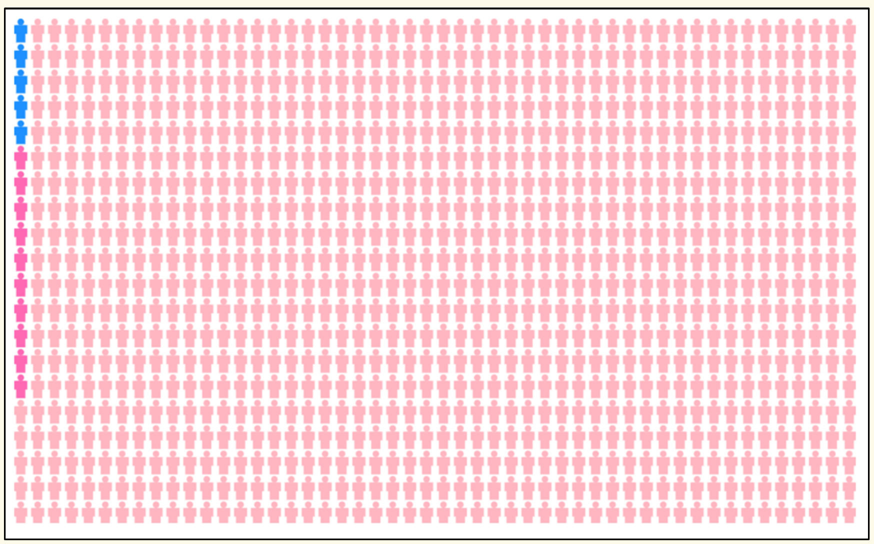
Example – Zika Testing

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. $P(Z)$

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

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



Suppose we had 100,000 people:

- 490 **have Zika** and **test positive**
- 10 **have Zika** and **test negative**
- 995 **do not have Zika** and **test positive**
- 98,505 **do not have Zika** and **test negative**

98% of those with Zika

2% of those with Zika

1% of those without Zika

$$P(Z|T) = \frac{\overset{500}{490}}{\underset{500}{490} + \underset{\approx 1000}{995}} \approx 0.33$$

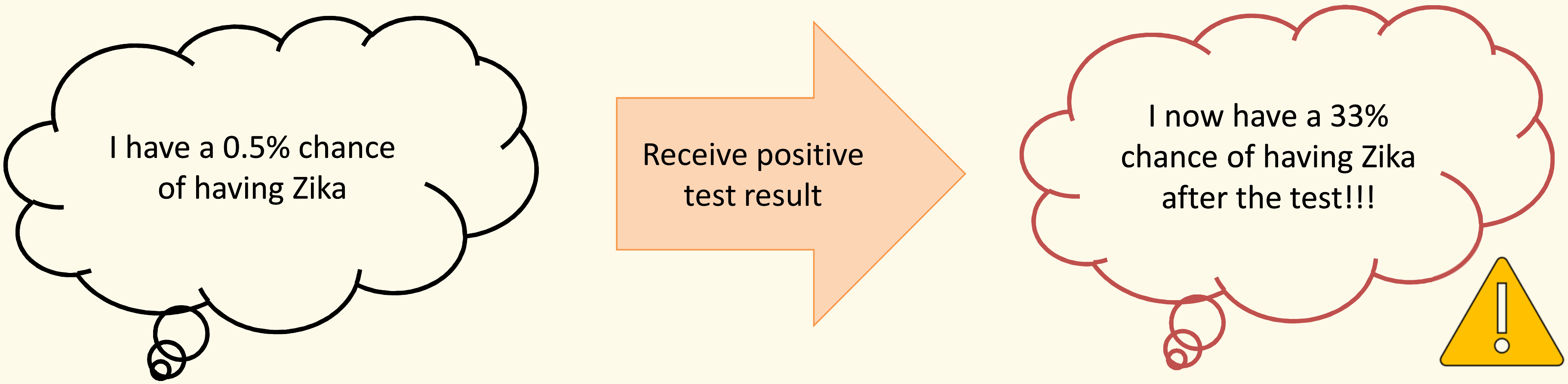
Demo

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") $P(T|Z)$
- 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 test negative (event T^c) if you have Zika (event Z)?

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

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 test negative (event T^c) if you have Zika (event Z)?

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

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

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

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

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

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

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

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

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

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 negative Zika test example with $E = Z$ and $F = T^c$

Simple Partition: In particul... 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"?

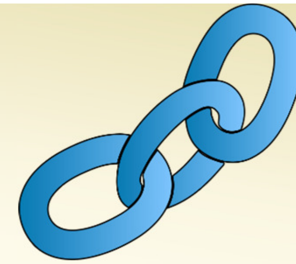
Some useful stats:

- 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

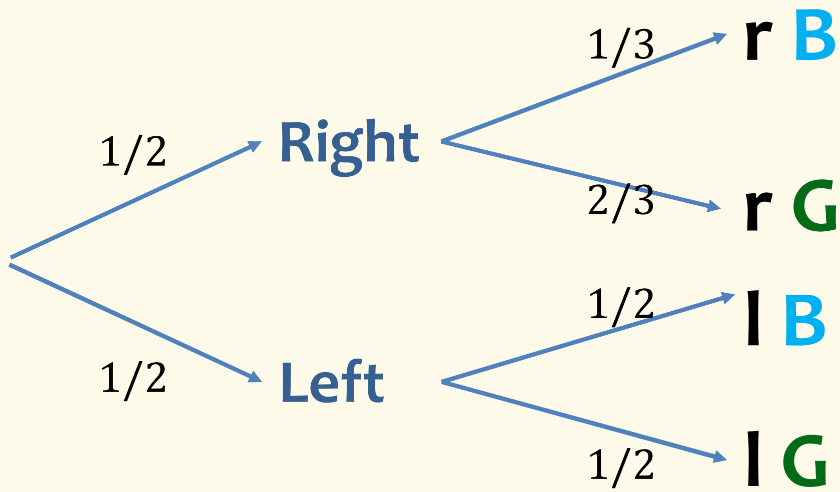
Chain Rule



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

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

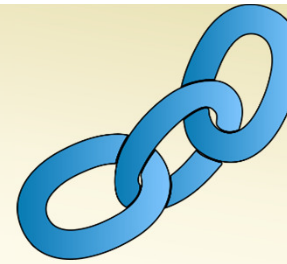
Recall from last time:



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

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 $A_1, A_2, \ldots, A_n$,

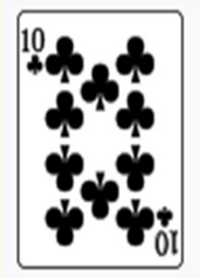
$$P(A_1 \cap \cdots \cap A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2)$$

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

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

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

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

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

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

Agenda

- 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

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

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

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} $= \{HHH, \underbrace{HHT, HTH, THH}\}$
- B = {at most 2 H's} $= \{\underline{HHH}\}^c$

Independent?

$$\underbrace{P(A \cap B)}_{= 3/8} \overset{?}{=} \underbrace{P(A)}_{\frac{1}{2}} \cdot \underbrace{P(B)}_{\frac{7}{8}}$$

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

Poll:

A. Yes, independent

B. No

pollev/paulbeame028

3.5 / 7 ✓

Multiple Events – Mutual Independence

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

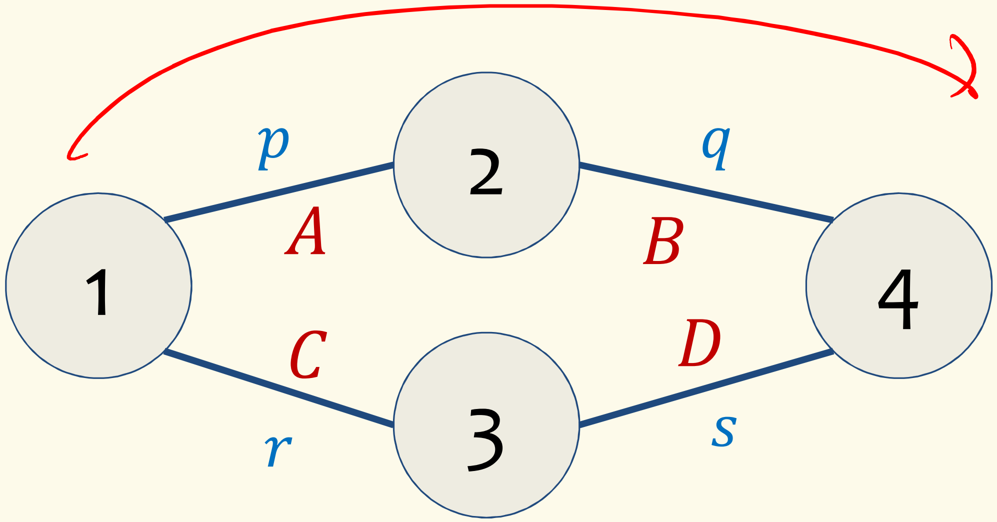
$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i).$$

Example – Network Communication

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

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

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

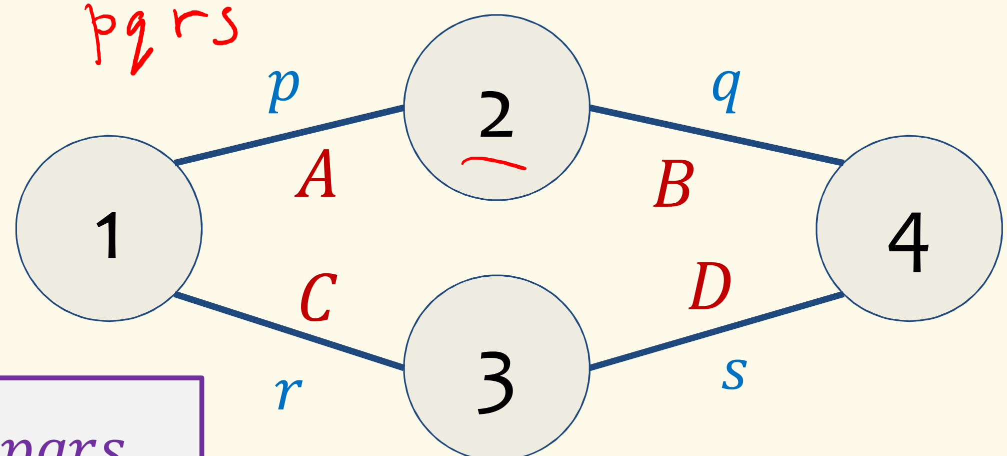


Example – Network Communication

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

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

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



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

Independence as an assumption

- People often assume it **without justification**

- Example: A skydiver has two chutes

  A: event that the main chute doesn't open $\quad P(A) = 0.02$

  B: event that the back-up doesn't open $\quad P(B) = 0.1$

- What is the chance that at least one opens assuming independence?

$1 - 0.02 \times 0.1 \qquad 1 - .002 = 99.8\%$

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

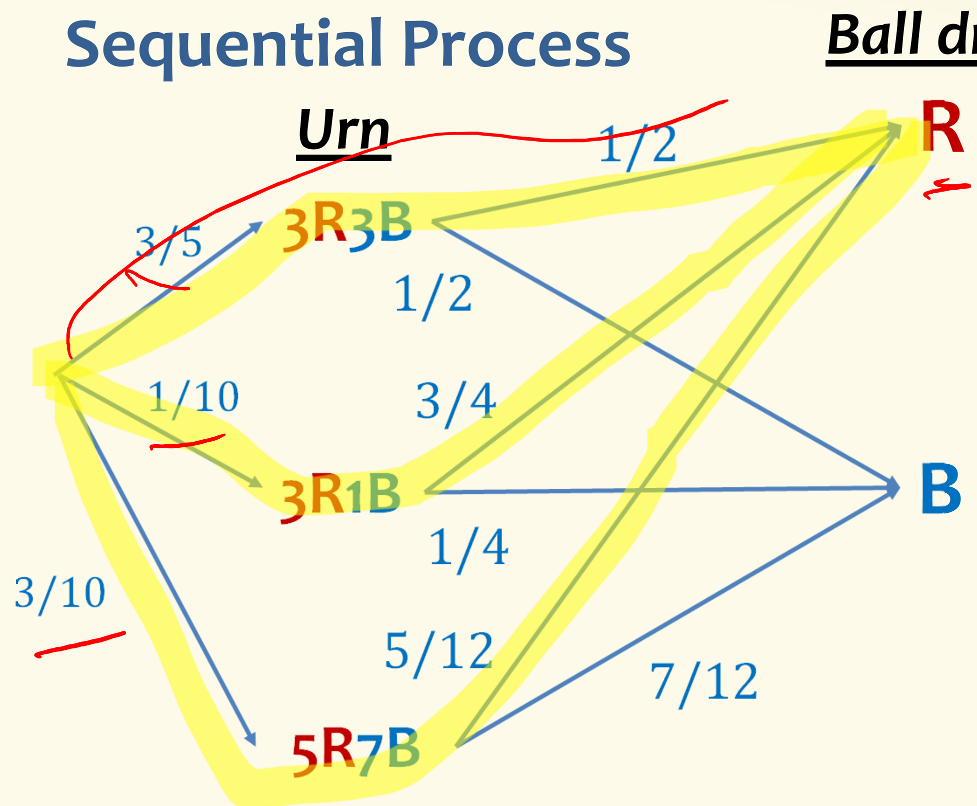
It is important to understand that independence is a property of probabilities of outcomes, not of the root cause generating these events.

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

$\not\Leftarrow$ *Not necessarily*

This can be counterintuitive!

Sequential Process



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?

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

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

$\boxed{\textbf{Independent! } P(\mathbf{R}) = P(\mathbf{R}\mid\mathbf{3R3B})}$ $= 1/2$

Agenda

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

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 $|\Omega| = \infty$

Fun: Von Neumann's Trick with a biased coin

- How to use a biased coin to get a fair coin flip:
  - Suppose that you have a biased coin:
    - $P(H) = p \qquad P(T) = 1 - p$

1. Flip coin twice: If you get HH or TT go to step 1
2. If you got $\underbrace{HT}_{p(1-p)}$ output H; if you got $\underbrace{TH}_{p(1-p)}$ output T.

Why is it fair? $P(H) = P(HT) = p(1-p) = (1-p)p = P(TH) = P(T)$

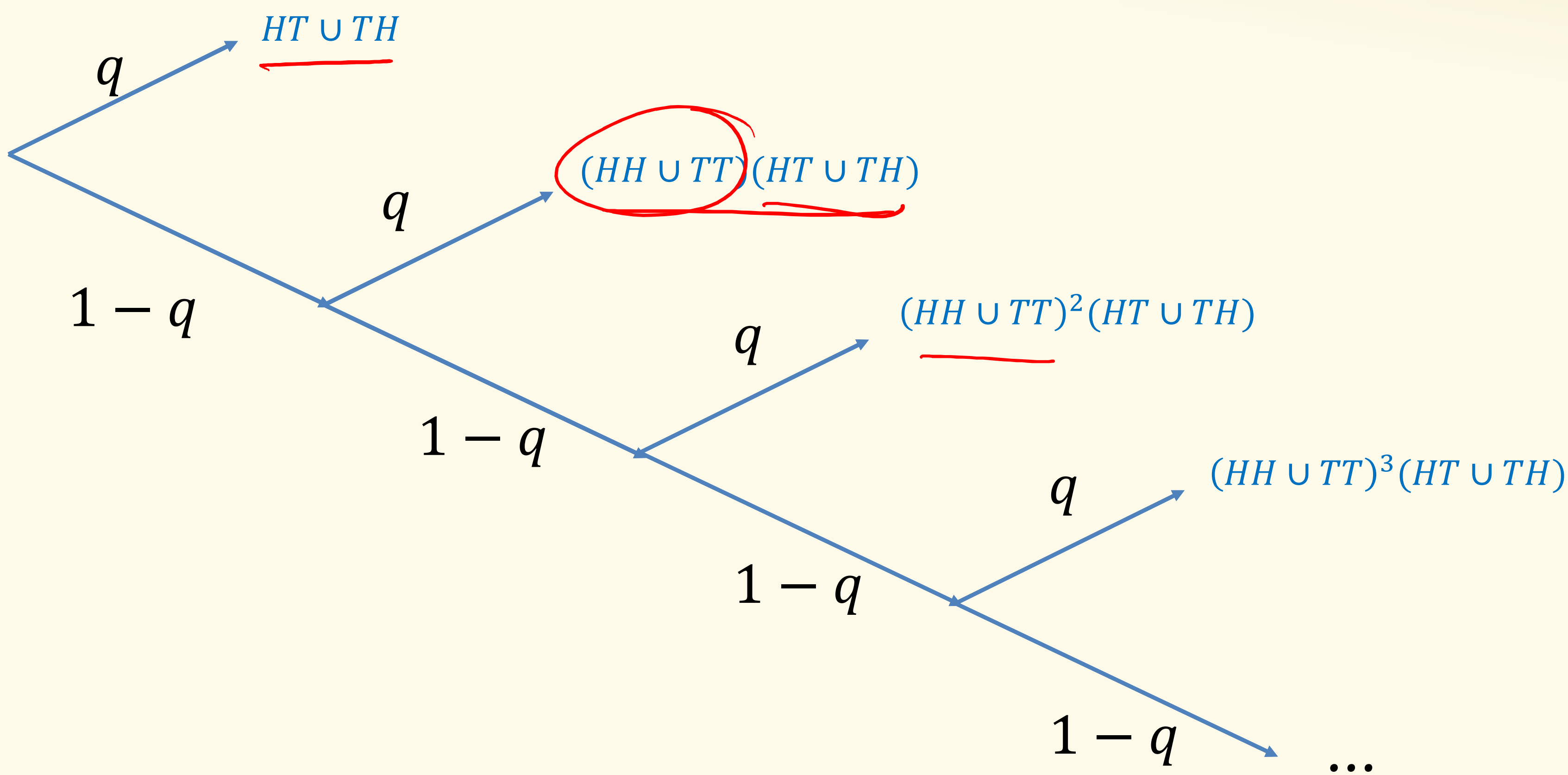
Drawback: You may never get to step 2.

The sample space for Von Neumann's trick

- For each round of Von Neumann's trick we flipped the biased coin twice.
  - If HT or TH appears, the experiment ends:
    - Total probability each round: $2p(1-p)$ call this q
  - If HH or TT appears, the experiment continues:
    - Total probability each round: $p^2 + (1-p)^2$ this is $1-q$

- Probability that flipping ends in round t is $(1-q)^{t-1} \cdot q$
  - Conditioned on ending in round t, $P(H) = P(T) = 1/2$

Sequential Process – Example



The sample space for Von Neumann's trick

More precisely, the sample space contains the successful outcomes:

$$\bigcup_{t=1}^{\infty}(HH \cup TT)^{t-1}(HT \cup TH)$$

which together have probability $\sum_{t=1}^{\infty}(1-q)^{t-1}q$ for $q = 2p(1-p)$

as well as all of the failing outcomes in $(HH \cup TT)^{\infty}$.

Observe that $q \neq 0$ iff $0 < p < 1$. We have two cases:

- If $q \neq 0$ then $\sum_{t=1}^{\infty}(1-q)^{t-1} = 1/q$ so successful outcomes account for total probability 1.

  $\frac{1}{1-r}$

- If $q = 0$ then either:
  - $p = 1$ and $(HH)^{\infty}$ has probability 1.
  - $p = 0$ and $(TT)^{\infty}$ has probability 1.