CSE 312
Foundations of Computing II

Lecture 5: Conditional Probability and Bayes Theorem

PAUL G. ALLEN SCHOOL
OF COMPUTER SCIENCE & ENGINEERING

slido.com/2402743
for polls and anonymous questions

Review Probability

Definition. A **sample space** Ω is the set of all possible outcomes of an experiment.

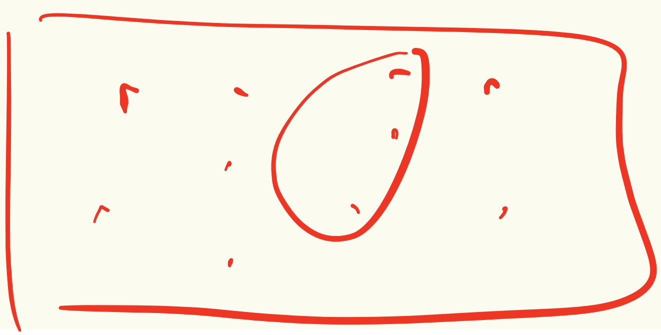
Examples:

- Single coin flip: $\Omega = \{H, T\}$
- Two coin flips: $\Omega = \{HH, HT, TH, TT\}$
- Roll of a die: $\Omega = \{1, 2, 3, 4, 5, 6\}$

Definition. An **event** $E \subseteq \Omega$ is a subset of possible outcomes.

Examples:

- Getting at least one head in two coin flips:
  $E = \{HH, HT, TH\}$
- Rolling an even number on a die :
  $E = \{2, 4, 6\}$



Ω

$P(E) = \frac{3}{4}$

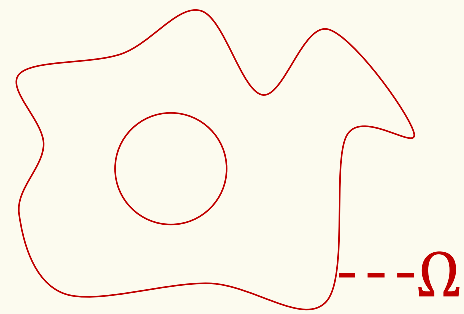
Probability space

Either finite or infinite countable (e.g., integers)

Definition. A (discrete) **probability space** is a pair $(\Omega, \mathbb{P})$ where:

- Ω is a set called the **sample space**.
- $\mathbb{P}$ is the **probability measure,** a function $\mathbb{P}: \Omega \to [0,1]$ such that:
  - $\mathbb{P}(E) \geq 0$ for all $E \subseteq \Omega$
  - $\sum_{\omega \in \Omega} \mathbb{P}(\omega) = 1$

Set of possible **elementary outcomes**



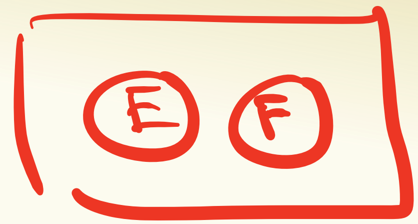
Specifies Likelihood (or probability) of each **event in the sample space**

Some outcome must show up

The likelihood (or probability) of each outcome is non-negative.

$$P(E) = \sum_{w \in E} P(w)$$

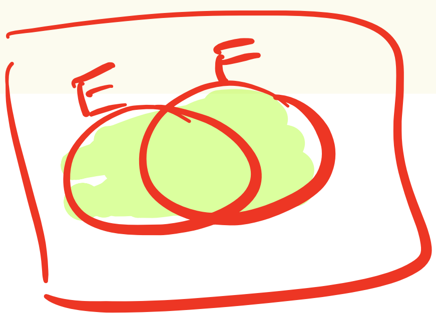
Review Axioms of Probability



Let Ω denote the sample space and $E, F \subseteq \Omega$ be events.

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 events,
then $P(E \cup F) = P(E) + P(F)$

Corollary 1 (Complementation): $P(E^c) = 1 - P(E)$
Corollary 2 (Monotonicity): If $E \subseteq F$, $P(E) \leq P(F)$
Corollary 3 (Inclusion-Exclusion): $P(E \cup F) = P(E) + P(F) - P(E \cap F)$



Review Equally Likely Outcomes

$P(\omega) = \frac{1}{|\Omega|}$

If (Ω, P) is a **uniform** probability space, then for any event $E \subseteq \Omega$, then

$$P(E) = \frac{|E|}{|\Omega|}$$

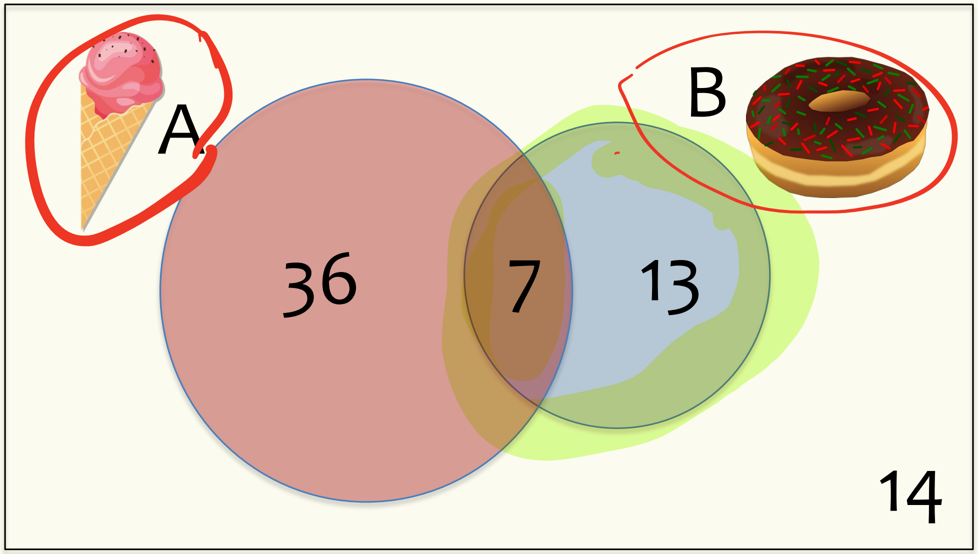
$$\sum_{\omega \in E} P(\omega) = \sum_{\omega \in E} \frac{1}{|\Omega|} = \frac{|E|}{|\Omega|}$$

Agenda

- Conditional Probability
- Bayes Theorem
- Law of Total Probability
- Bayes Theorem + Law of Total Probability
- More Examples

$36 + 7 + 13 + 14 = \underline{70}$

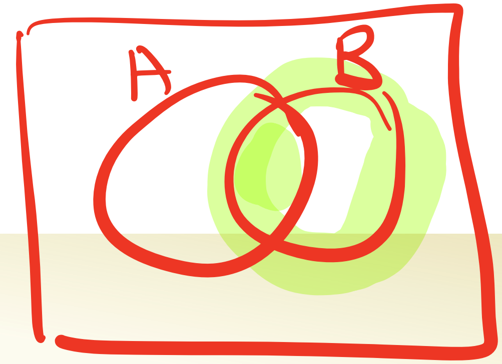
Conditional Probability (Idea)



What’s the probability that a uniformly random person likes ice cream **given** they like donuts?

$P(\text{person selected likes donuts}) = \frac{20}{70}$

$Pr(A \mid B) = \frac{7}{20}$



Conditional Probability

Definition. The **conditional probability** of event A **given** an event B happened (assuming $P(B) \neq 0$) is

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

An equivalent and useful formula is

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

Reversing Conditional Probability

Question: Does $P(A|B) = P(B|A)$?

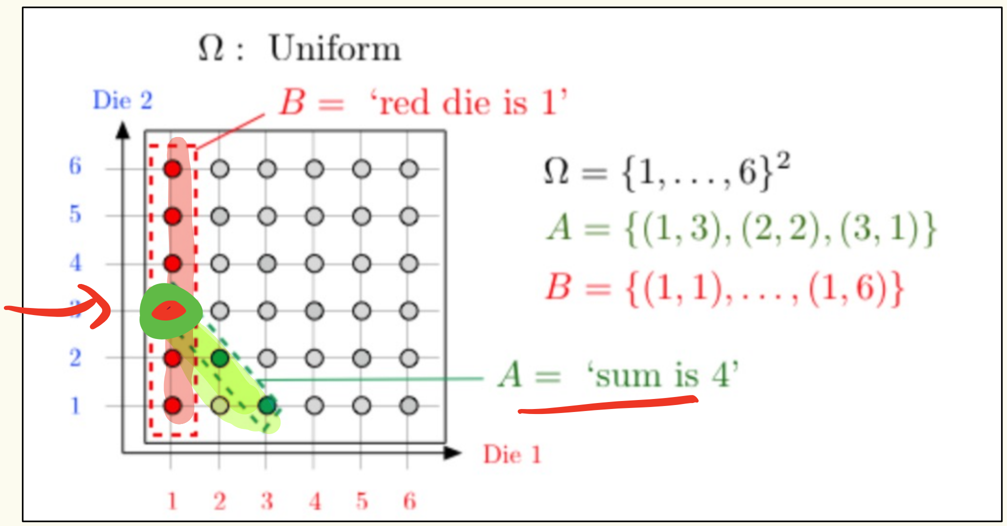
No! The following analogy is purely for intuition and makes no sense in terms of probability

- Let A be the event you are wet
- Let B be the event you are swimming

$$P(A|B) = 1$$
$$P(B|A) \neq 1$$

Example with Conditional Probability

Toss a red die and a blue die (both 6 sided and all outcomes equally likely). What is $P(B)$? What is $P(B|A)$?



$P(\omega) = \frac{1}{36}$

A: sum is 4

B: red die is 1

$Pr(B) = \frac{6}{36} = \frac{1}{6}$.

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

$= \frac{1/36}{3/36} = \frac{1}{3}$

$\frac{|A \cap B|}{|A|}$

Conditional Probability in a uniform probability space

Definition. The **conditional probability** of event A **given** an event B happened (assuming $P(B) \neq 0$) is

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

If the probability space is uniform, then $P(A \mid B) = \frac{|A \cap B|}{|B|}$

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

Agenda

- Conditional Probability
- Bayes Theorem
- Law of Total Probability
- Bayes Theorem + Law of Total Probability
- More Examples

Our First Machine Learning Task: Spam Filtering

`Subject: “FREE $$$ CLICK HERE”`

Suppose you know that 80% of emails you receive are spam.

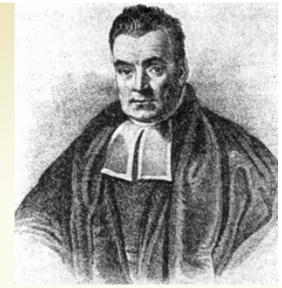
$P(\text{email is spam}) = 0.8$

So a priori, our belief is that any email has an 80% chance of being spam.

How do you update that belief when you see that the subject line contains the phrase “FREE $$$”?

$P(\text{email is spam} \mid \text{contains FREE \$})$

Bayes Theorem



A formula to let us "reverse" the conditional.

Theorem. (Bayes Rule) For events A and B, where $P(A), P(B) > 0$,

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

$P(A)$ is called the **prior** (our belief without knowing anything)

$P(A|B)$ is called the **posterior** (our belief after learning B)

Bayes Theorem follows from the definition of conditional probability

Definition. The **conditional probability** of event A **given** an event B happened (assuming $P(B) \neq 0$) is

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

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

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

Bayes Theorem Proof

By definition of conditional probability

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

Swapping A, B gives

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

But $P(A \cap B) = P(B \cap A)$, so

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

Dividing both sides by $P(B)$ gives

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

S email is spam
F subject line contains

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. $\Rightarrow P(F \mid \bar{S}) = 0.1$
- 70% of spam emails contain the word "FREE" in the subject. $\Rightarrow P(F \mid S) = 0.7$
- 80% of emails you receive are spam. $P(S) = 0.8$

$P(\bar{S}) = 0.2$

$$P(S \mid F) = \frac{P(F \mid S)P(S)}{P(F)}$$

$$= \frac{0.7 \cdot 0.8}{P(F)}$$

$$= \frac{0.7 \cdot 0.8}{0.7 \cdot 0.8 + 0.1 \cdot 0.2} =$$

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

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

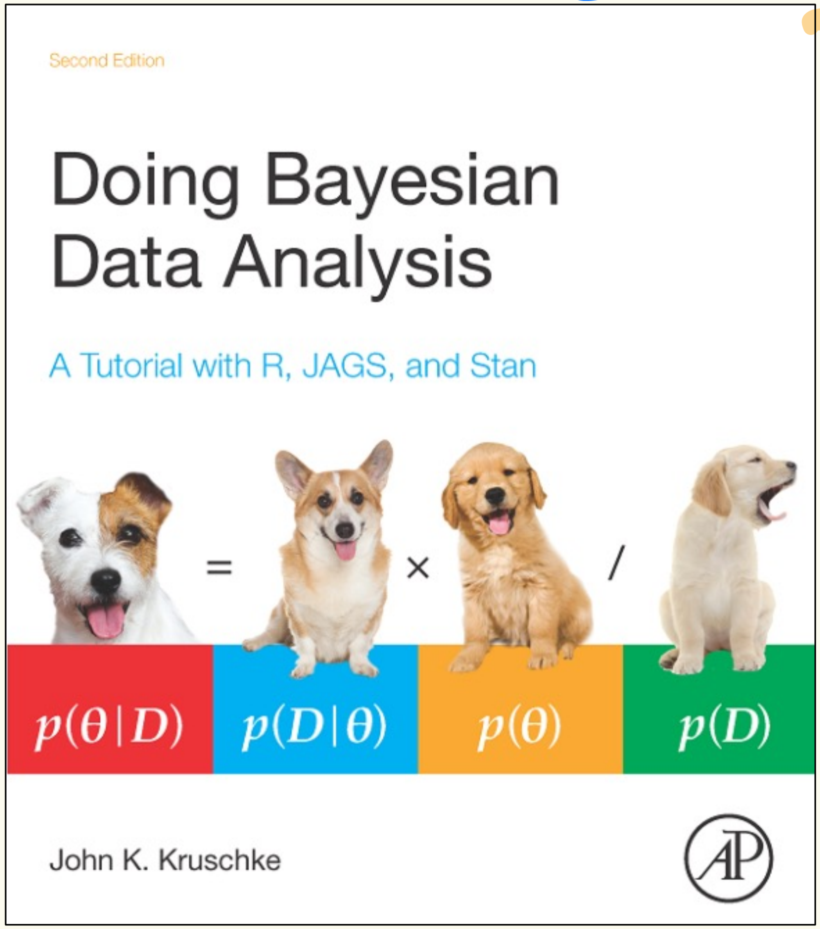
Brain Break

Law of total probability

$$\boxed{P(F) = P(F\cap S) + P(F\cap \bar{S})}$$

green $\Downarrow$ orange $\Downarrow$

$$= P(F|S)P(S) + P(F|\bar{S})P(\bar{S})$$



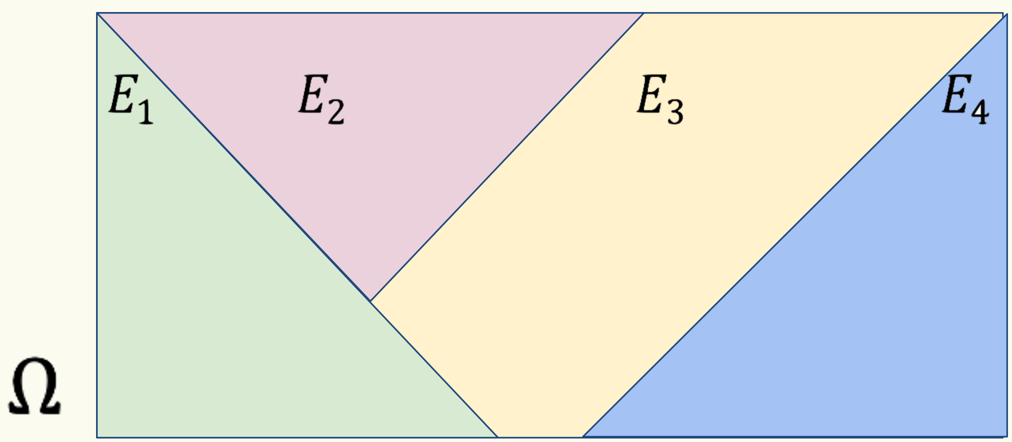
Agenda

- Conditional Probability
- Bayes Theorem
- Law of Total Probability
- Bayes Theorem + Law of Total Probability
- More Examples

Partitions (Idea)

These events **partition** the sample space

1. They "cover" the whole space
2. They don't overlap



Partition

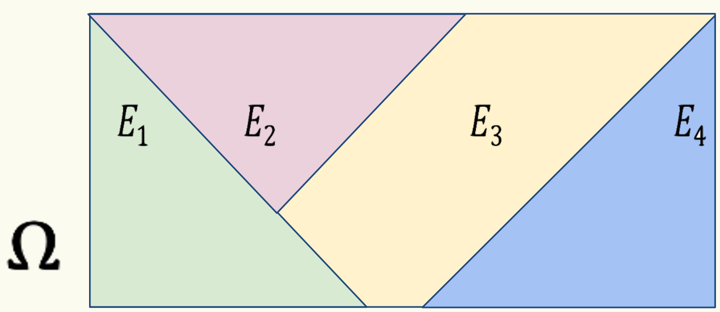
Definition. Non-empty events $E_1, E_2, \dots, E_n$ **partition** the sample space Ω if

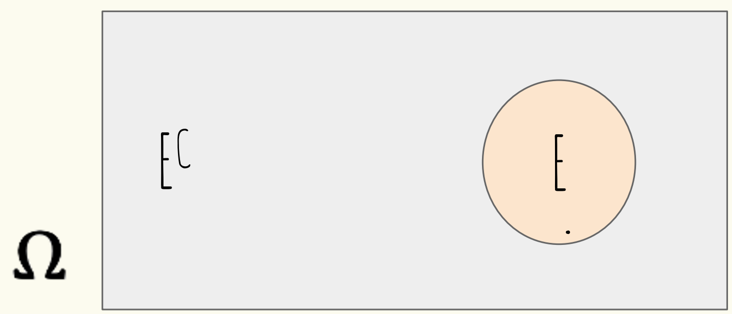
(Exhaustive)

$$E_1 \cup E_2 \cup \cdots \cup E_n = \bigcup_{i=1}^{n} E_i = \Omega$$

(Pairwise Mutually Exclusive)

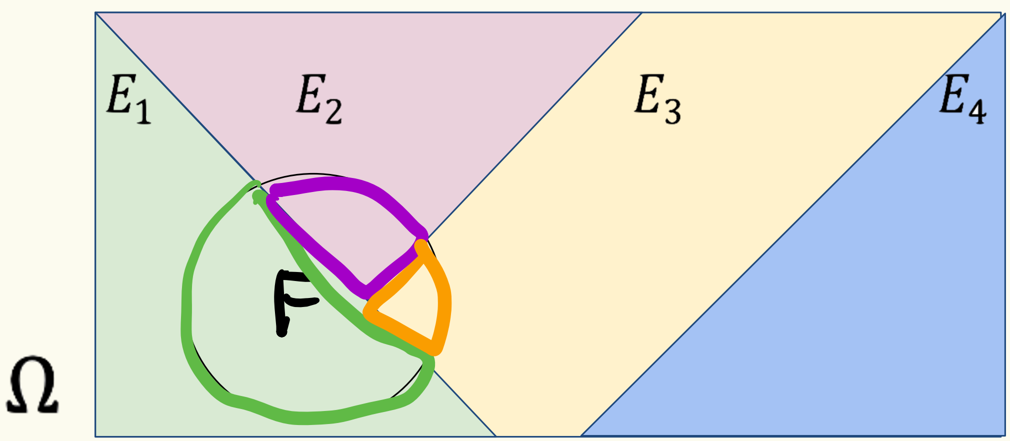
$$\forall_i \forall_{i \neq j} E_i \cap E_j = \emptyset$$





Law of Total Probability (Idea)

If we know $E_1, E_2, \dots, E_n$ partition Ω, how does that help us compute $P(F)$?



Law of Total Probability (LTP)

Definition. If events $E_1, E_2, \ldots, E_n$ partition the sample space Ω, then for any event F

$$P(F) = P(F \cap E_1) + \ldots + P(F \cap E_n) = \sum_{i=1}^{n} \underbrace{P(F \cap E_i)}_{P(F|E_i)P(E_i)}$$

Using the definition of conditional probability $P(F \cap E) = P(F|E)P(E)$
We can get the alternate form of this that show

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

An Example

Alice has two pockets:

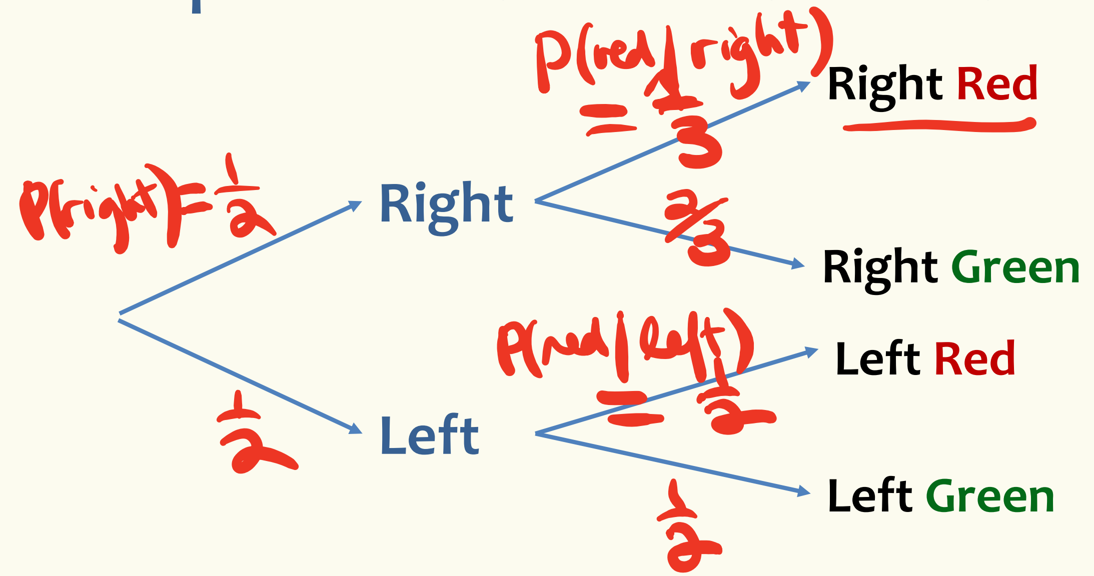
- **Left pocket:** Two red balls, two green balls
- **Right pocket:** One red ball, two green balls.

Alice picks a random ball from a random pocket.
[Both pockets equally likely, each ball equally likely.]

What is $\mathbb{P}(\mathbf{R})$?

red.

Sequential Process – Non-Uniform Case

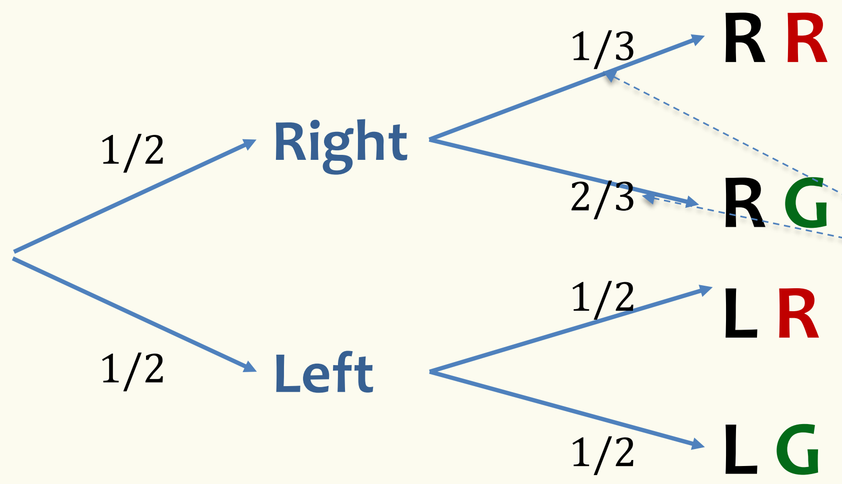


- **Left pocket:** Two red, two green
- **Right pocket:** One red, two green.

- Alice picks a random ball from a random pocket

$$P(\text{red}) = P(\text{right})P(\text{red} \mid \text{right}) + P(\text{left})P(\text{red} \mid \text{left})$$
$$= \frac{1}{2} \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{2}$$

Sequential Process – Non-Uniform Case



- **Left pocket:** Two red, two green
- **Right pocket:** One red, two green.

1/3 = $\mathcal{P}(R \mid \mathrm{R})$ and 2/3 = $\mathcal{P}(G \mid \mathrm{R})$

$$\mathbb{P}(\mathbf{R}) = \mathbb{P}(\mathbf{R} \cap \mathbf{Left}) + \mathbb{P}(\mathbf{R} \cap \mathbf{Right})$$

(Law of total probability)

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

$$= \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{3} = \frac{1}{4} + \frac{1}{6} = \frac{5}{12}$$

Agenda

- Conditional Probability
- Bayes Theorem
- Law of Total Probability
- Bayes Theorem + Law of Total Probability
- More Examples

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.

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,

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

Simple Partition: In particular, if E is an event with non-zero probability, then

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

Agenda

- Conditional Probability
- Bayes Theorem
- Law of Total Probability
- Bayes Theorem + Law of Total Probability
- More Examples

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

Z: has Zika
T: tests positive.

Suppose we know the following Zika stats

- A test is 98% effective at detecting Zika ("true positive")
- However, the test yields a "false positive" 1% of the time
- 0.5% of the US population has Zika.

$P(T|Z) = 0.98$

$P(T|\overline{Z}) = 0.01$

$P(Z) = 0.005$

What is the probability a random person has Zika (event *Z*) given that they test positive (event *T*).

$P(Z|T)$

$P(T|Z) + P(\overline{T}|Z)$

Summary

- Conditional Probability
- Bayes Theorem
- Law of Total probability

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

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

$$\mathbb{P}\,(F) = \sum_{i=1}^{n} \mathbb{P}(F|E_i)\mathbb{P}(E_i)$$ E_i partition Ω