CSE 312

Foundations of Computing II

Lecture 5: Conditional Probability and Bayes Theorem

PAUL G. ALLEN SCHOOL
OF COMPUTER SCIENCE & ENGINEERING

Anna R. Karlin

Slide Credit: Based on Stefano Tessaro's slides for 312 19au
incorporating ideas from Alex Tsun, Rachel Lin, Hunter Schafer & myself ☺

$(\Omega, P(\cdot))$

Review Probability

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

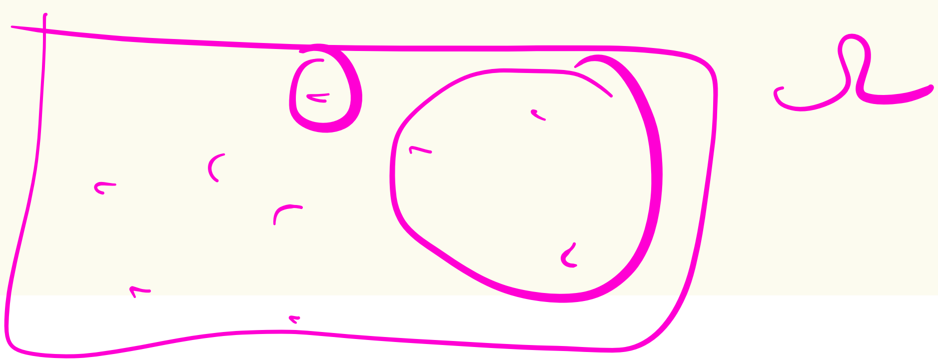
Examples:

- 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 : $E = \{2, 4, 6\}$



$P(\{\omega\}) = P(\omega)$

Probability space

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

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

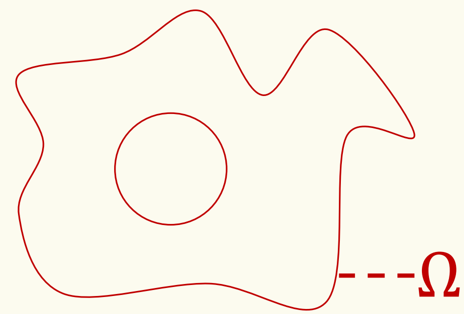
- Ω is a set called the **sample space.**
- $\mathbb{P}$ is the **probability measure,**

  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.

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

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, 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): $P(E \cup F) = P(E) + P(F) - P(E \cap F)$

Agenda

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

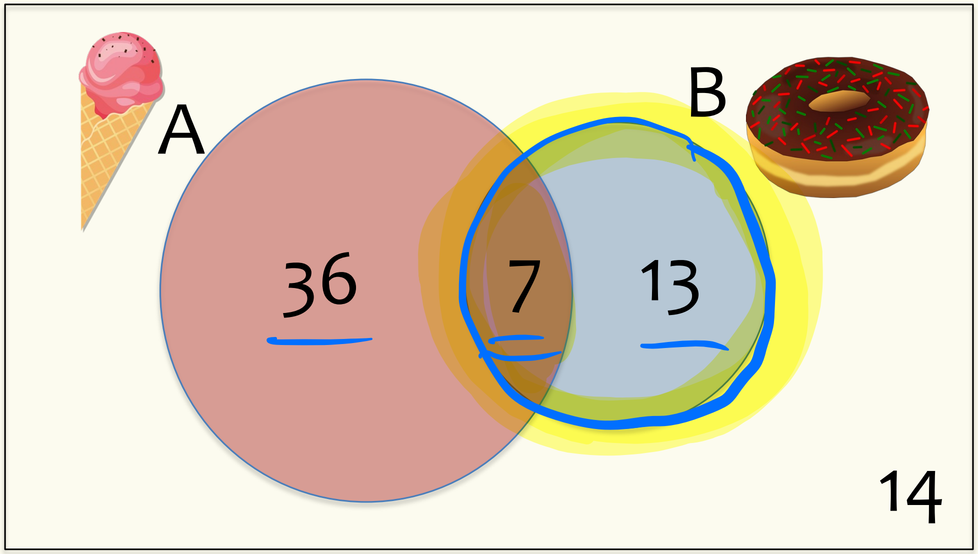
$Pr(E) = \frac{|E|}{|\Omega|}$

$36 + 7 + 13 + 14 = 70$

Conditional Probability (Idea)

A likes ice cream

B likes donuts



$|\Omega| = 70$

$Pr(w) = \frac{1}{70}$

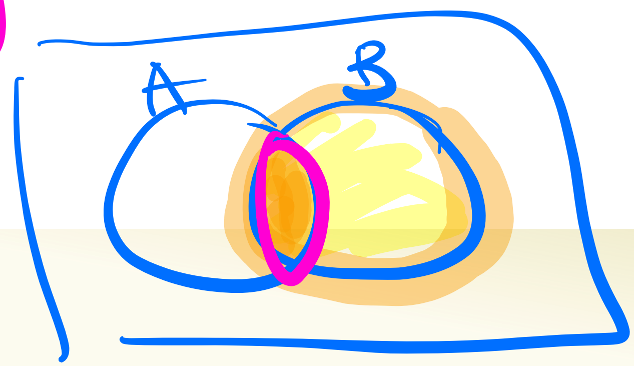
$Pr(B) = \frac{13+7}{70}$

What's the probability that someone likes ice cream **given** they like donuts?

$P(A|B) = \frac{7}{20}$

given
conditioned on

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



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

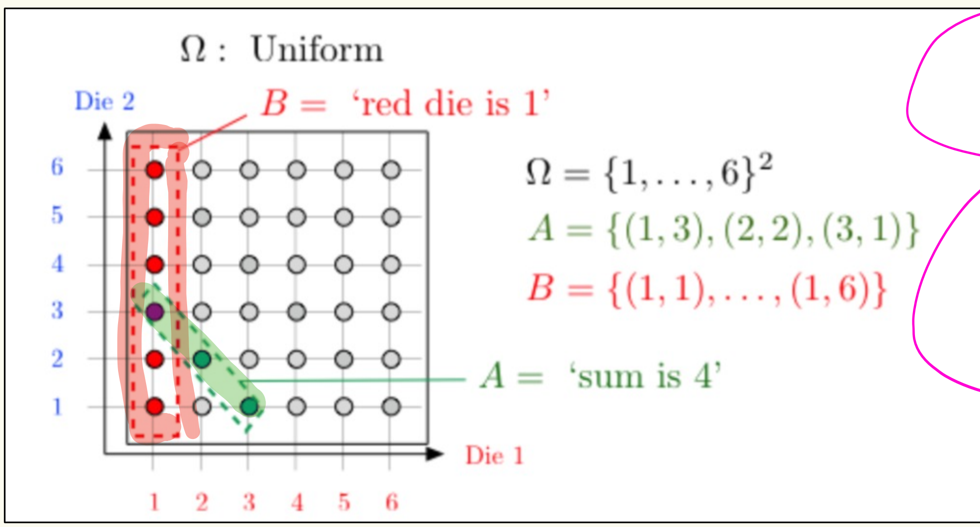
No! The following analogy is purely for intuition and makes no sense in terms of probability

- 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 $P(B|A)$?



$|\Omega| = 36$

A sum is 4

B red is 1

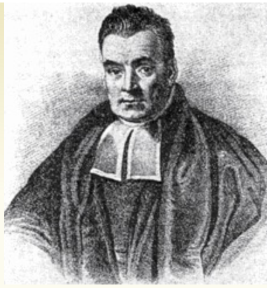
$P(B) = \frac{6}{36} = \frac{1}{6}$

$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{1/36}{3/36} = \frac{1}{3} = \frac{|B \cap A|}{|A|}$

Agenda

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

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)

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

$\cdot / P(B)$ $\qquad\qquad\qquad\qquad \cdot P(A)$

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

$$P(A|B)P(B) = P(B|A)P(A) \quad / \cdot \frac{1}{P(B)}$$

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

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

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

Our First Machine Learning Task: Spam Filtering

S: email is spam
$\bar{S}$: email is not spam (ham)

Subject: "FREE $$$ CLICK HERE"

F: subject contains "FREE"

What is the probability this email is spam, given the subject contains "FREE"?

Some useful stats:

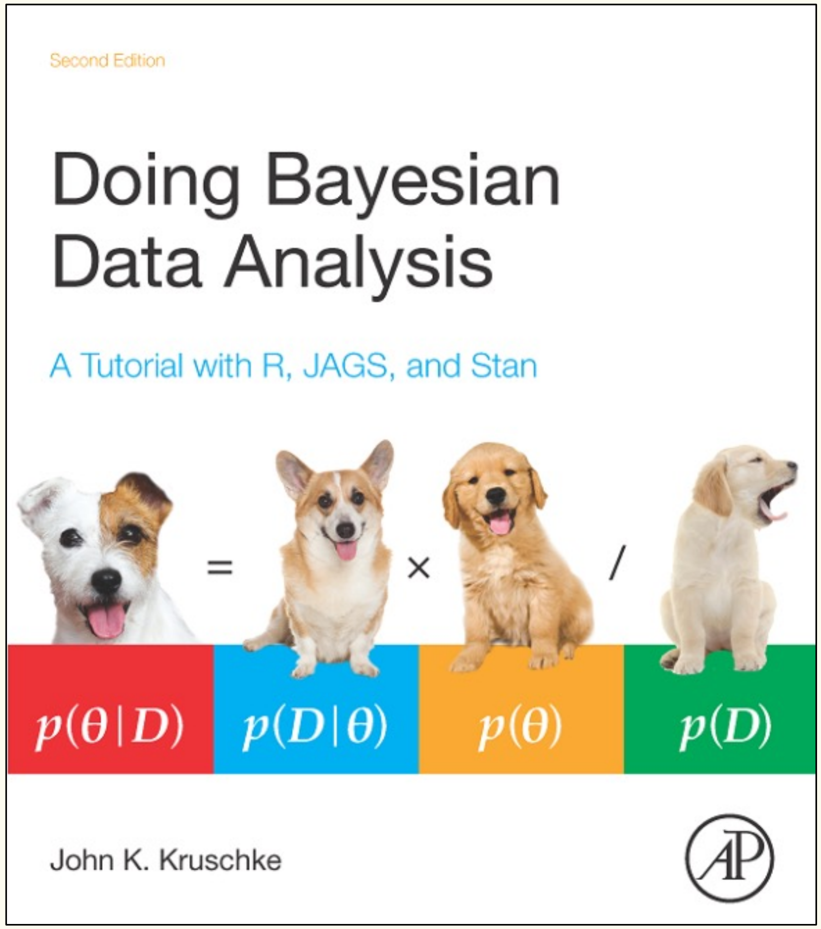
- 10% of ham (i.e., not spam) emails contain the word "FREE" in the subject. $\Pr(F \mid \bar{S}) = 0.1$
- 70% of spam emails contain the word "FREE" in the subject. $\Pr(F \mid S) = 0.7$
- 80% of emails you receive are spam.

$P(S) = 0.8$

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

$$\Pr(S \mid F) \underset{\text{Bayes}}{=} \frac{P(F \mid S)\,P(S)}{P(F)} = \frac{0.7 \cdot 0.8}{??}$$

Brain Break


Second Edition
Doing Bayesian
Data Analysis
A Tutorial with R, JAGS, and Stan
=
×
/
$p(\theta|D)$
$p(D|\theta)$
$p(\theta)$
$p(D)$
John K. Kruschke

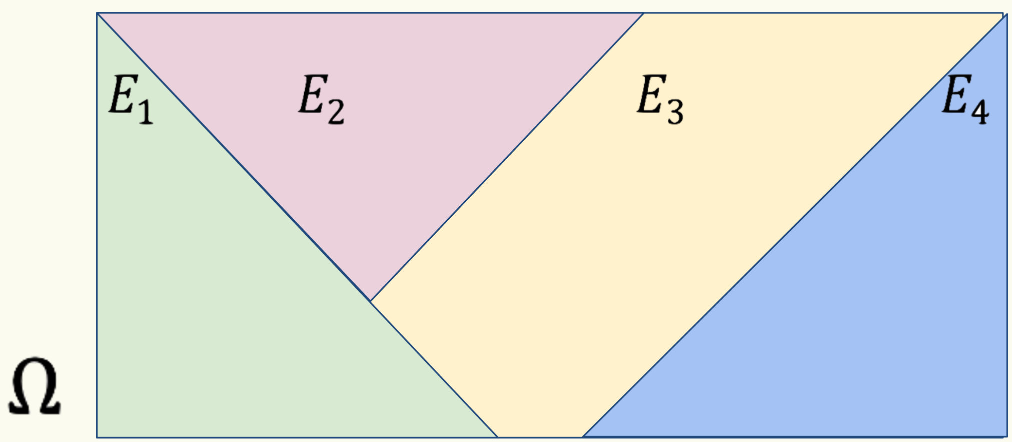
Agenda

- 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

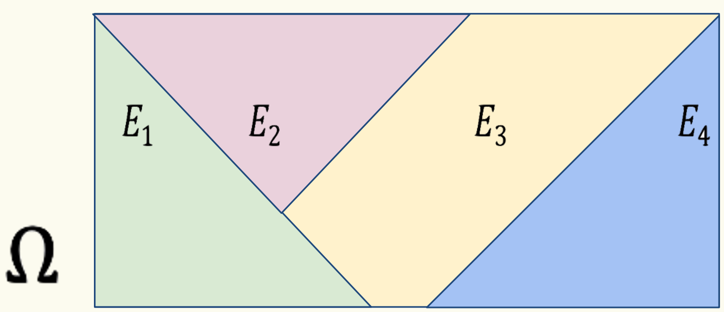
Definition. Non-empty events $E_1, E_2, \ldots, E_n$ **partition** the sample space Ω if

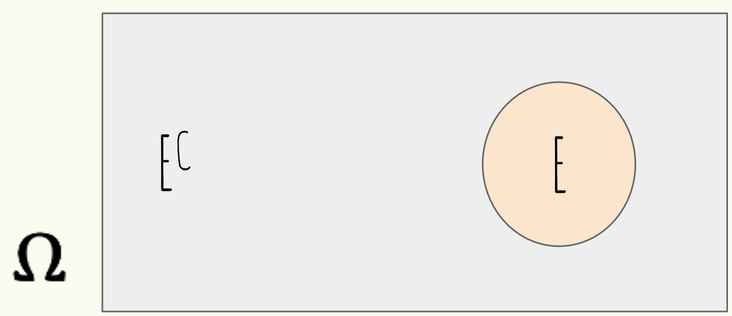
(Exhaustive)

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

(Pairwise Mutually Exclusive)

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



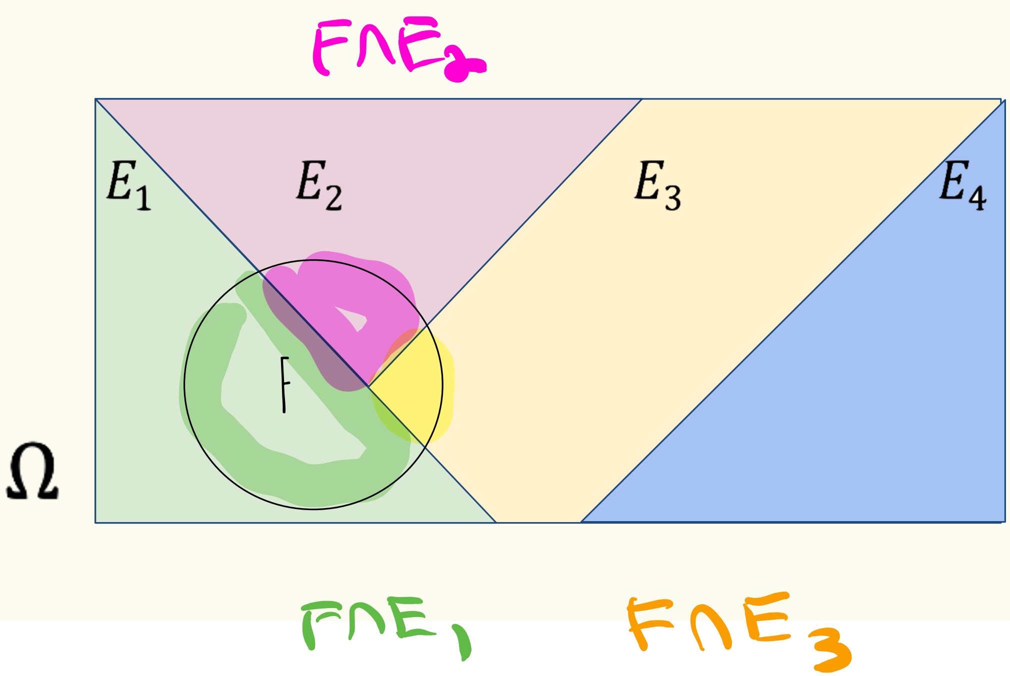


$$E^c = \overline{E}$$

Law of Total Probability (Idea)

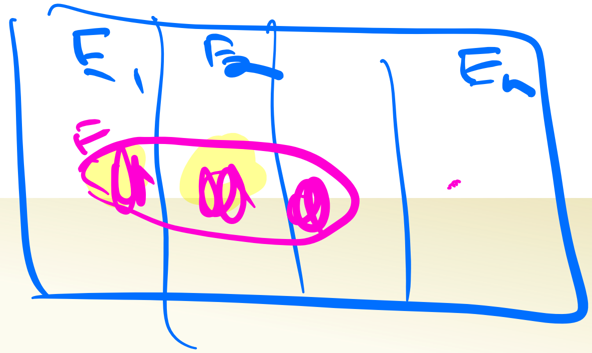
If we know $E_1, E_2, \dots, E_n$ partition Ω, what can we say about $P(F)$

$F = (F \cap E_1) \cup (F \cap E_2) \cup (F \cap E_3)$



$P(F) =$

$P(F \cap E_1) + P(F \cap E_2) + P(F \cap E_3)$



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

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

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

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

Another Contrived Example

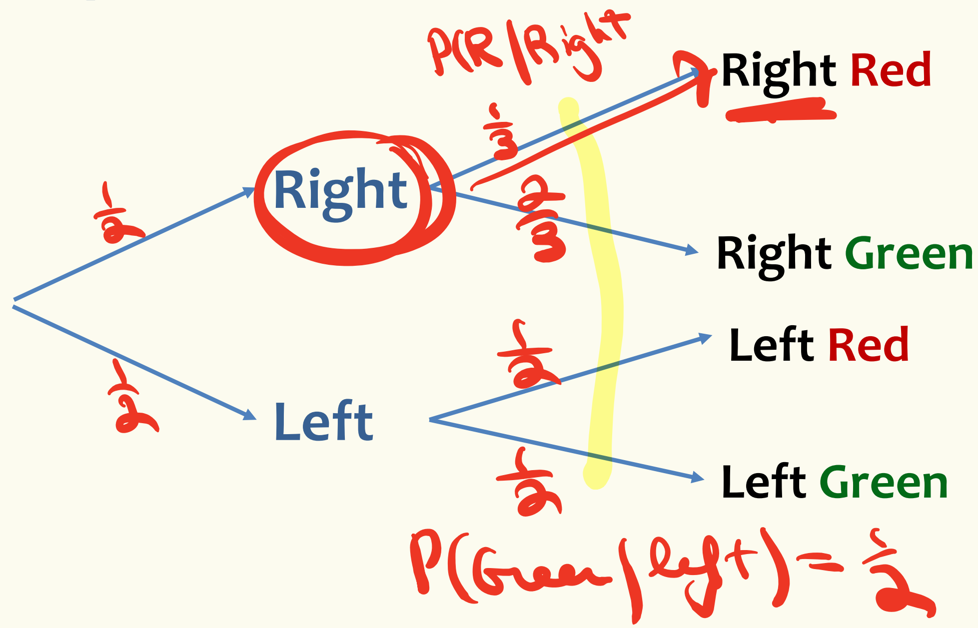
Alice has two pockets:

- **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.]

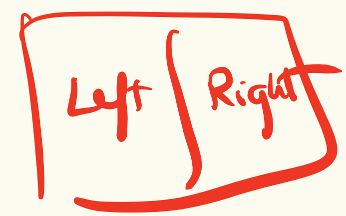
What is $\mathbb{P}(\mathbf{R})$?

Sequential Process – Non-Uniform Case



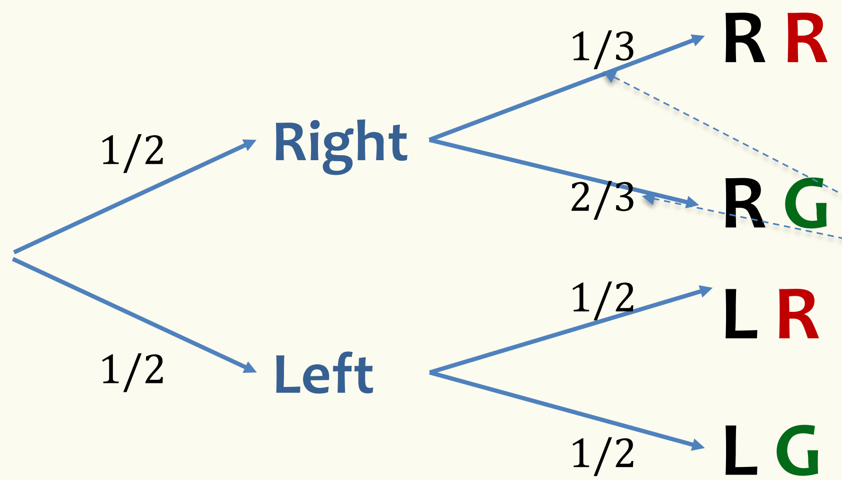
- **Left pocket:** Two red, two green
- **Right pocket:** One red, two green.

- Alice picks a random ball from a random pocket



$$\Pr(red) = P(red \mid Right)\,P(Right) + P(red \mid Left)\,P(Left)$$
$$= \frac{1}{3} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2}$$

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

S: email is spam

$\bar{S}$: email is not spam (ham)

Subject: "FREE $$$ CLICK HERE"

F: subject contains "FREE"

What is the probability this email is spam, given the subject contains "FREE"?

Some useful stats:

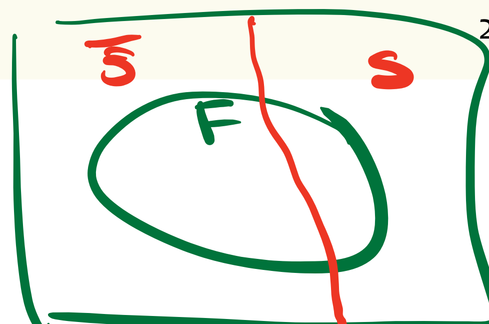
- 10% of ham (i.e., not spam) emails contain the word "FREE" in the subject. $\Pr(F \mid \bar{S}) = 0.1$
- 70% of spam emails contain the word "FREE" in the subject. $\Pr(F \mid S) = 0.7$
- 80% of emails you receive are spam. $P(S) = 0.8$

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

$$\Pr(S \mid F) \underset{\text{Bayes}}{=} \frac{P(F \mid S)\,P(S)}{P(F)} = \frac{0.7 \cdot 0.8}{??}$$

Use Bayes to compute $P(F)$

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



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 any 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 : have Zika
T : test positive.

Suppose we know the following Zika stats

- A test is 98% 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.

$P(T \mid Z) = 0.98$

$P(T \mid \bar{Z}) = 0.01$

$P(Z) = 0.005$

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

$$P(Z \mid T) = \frac{P(T \mid Z)\,P(Z)}{P(T)} = \frac{0.98 \cdot 0.005}{P(T)}$$

LTP: $P(T) = P(T \mid Z)\,P(Z) + P(T \mid \bar{Z})\,P(\bar{Z})$

$= 0.98 \cdot 0.005 + 0.01 \cdot 0.995$

$P(Z \mid T) \approx 0.33$

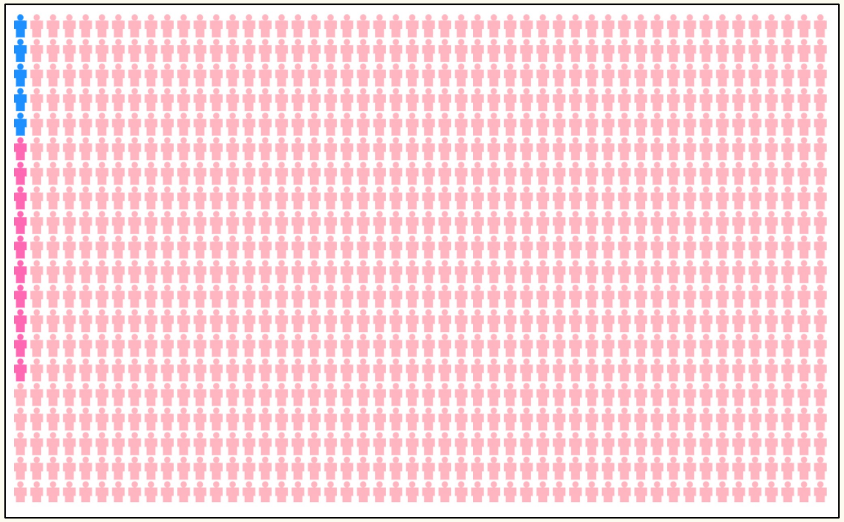
Example – Zika Testing

Have zika blue, don't pink

Suppose we know the following Zika stats

- A test is 98% effective at detecting Zika ("true positive") 100%
- However, the test may yield a "false positive" 1% of the time 10/995 = approximately 1%
- 0.5% of the US population has Zika. 5% have it.

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



Suppose we had 1000 people:

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

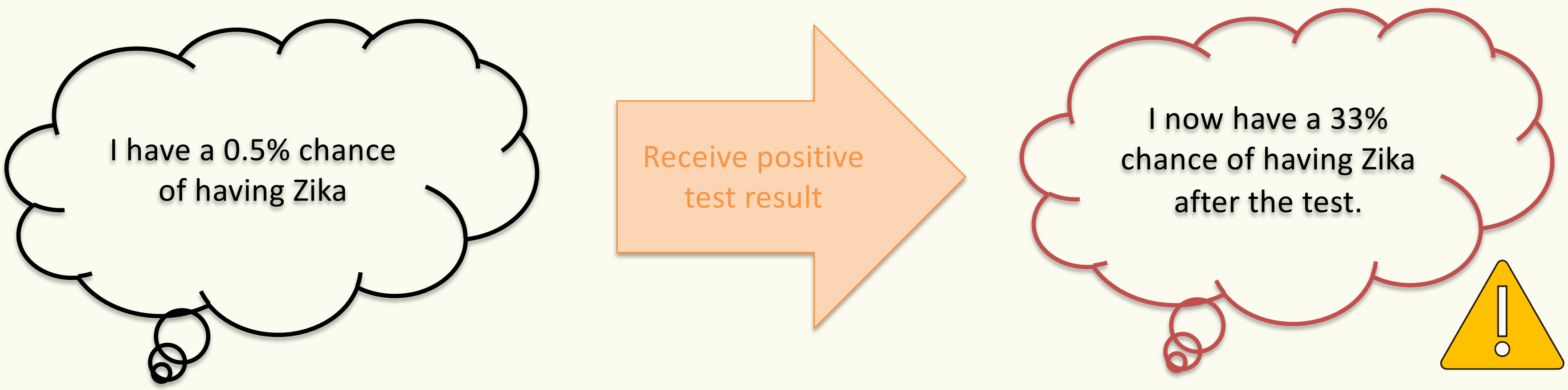
$$\frac{5}{5+10} = \frac{1}{3} \approx 0.33$$

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")
- However, the test may yield a "false positive" 1% of the time
- 0.5% of the US population has Zika.

What is the probability you test negative (event $\bar{T}$) if you have Zika (event Z)?

Conditional Probability Defines a Probability Space

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

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

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

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

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

Gambler's fallacy

Assume we toss **51** fair coins. Each outcome equally likely.
Assume we have seen **50** coins, and they are all "tails".
What are the odds the **51**$^{\text{st}}$ coin is "heads"?

$\mathcal{A} =$ first 50 coins are "tails"

$B =$ first 50 coins are "tails", 51$^{\text{st}}$ coin is "heads"

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

Gambler's fallacy

Assume we toss **51** fair coins.
Assume we have seen **50** coins, and they are all "tails".
What are the odds the **51^st** coin is "heads"?

$\mathcal{A} =$ first 50 coins are "tails"

$B =$ first 50 coins are "tails", 51^st coin is "heads"

$$\mathbb{P}(\mathcal{B}|\mathcal{A}) = \frac{\mathbb{P}(\mathcal{A} \cap \mathcal{B})}{\mathbb{P}(\mathcal{A})} = \frac{1/2^{51}}{2/2^{51}} = \frac{1}{2}$$

51^st coin is independent of outcomes of first 50 tosses!

Gambler's fallacy = Feels like it's time for " heads"!?