Announcements

• Pset #1 is due on Friday
• Office hours, wahoo!
• First sections this week!
• Section assignments will be sent out today.
• Auditing? Cool, but just lectures and online resources.
• https://cs109psets.netlify.app/win22/lecture4/
Follow Along with Lecture

https://cs109psets.netlify.app/win22/lecture4/
Review
This class going forward

Last week
Equally likely events

Today and for most of this course
Not equally likely events
Recall: $S =$ all possible outcomes. $E =$ the event.

- Axiom 1: $0 \leq P(E) \leq 1$
- Axiom 2: $P(S) = 1$
- Identity 3: $P(E^c) = 1 - P(E)$
Review, Axioms of Probability

Recall: $S =$ all possible outcomes. $E =$ the event.

- Axiom 1: $0 \leq P(E) \leq 1$

- Axiom 2: $P(S) = 1$

- Axiom 3: If events $E$ and $F$ are mutually exclusive:

\[ P(E \cup F) = P(E) + P(F) \]
If events are mutually exclusive, probability of OR is simple:

\[ P(E \cup F) = P(E) + P(F) \]
If events are mutually exclusive, probability of OR is simple:

\[ P(E \cup F) = \frac{7}{50} + \frac{4}{5} = \frac{11}{50} \]
Review, Mutually Exclusive Events

\[ P(X_1 \cup X_2 \cup \cdots \cup X_n) = \sum_{i=1}^{n} P(X_i) \]

Wahoo! All my events are mutually exclusive
Review, Mutually Exclusive Events

If events are mutually exclusive, probability of OR is easy!
End Review
Learning Goal for Today: Conditional Probability

\[ P(E \text{ and } F) \]

Chain rule (Product rule)  Definition of conditional probability

\[ P(E | F) \]

Law of Total Probability  Bayes’ Theorem

\[ P(E) \quad P(F | E) \]
Conditional Probability
Roll two dice

Roll two 6-sided fair dice. What is $P(\text{sum} = 7)$?

$$S = \{(1,1) \ (1,2) \ (1,3) \ (1,4) \ (1,5) \ (1,6) \\
(2,1) \ (2,2) \ (2,3) \ (2,4) \ (2,5) \ (2,6) \\
(3,1) \ (3,2) \ (3,3) \ (3,4) \ (3,5) \ (3,6) \\
(4,1) \ (4,2) \ (4,3) \ (4,4) \ (4,5) \ (4,6) \\
(5,1) \ (5,2) \ (5,3) \ (5,4) \ (5,5) \ (5,6) \\
(6,1) \ (6,2) \ (6,3) \ (6,4) \ (6,5) \ (6,6) \}$$

$E = \text{In blue}$
Dice, our misunderstood friends

Roll two 6-sided dice, yielding values $D_1$ and $D_2$. You want them to sum to 4.

What is the best outcome for $P(D_1)$?

Your Choices:

- A. 1 and 3 tie for best
- B. 1, 2 and 3 tie for best
- C. 2 is the best
- D. Other/none/more than one
Dice, our misunderstood friends

Roll two 6-sided dice, yielding values $D_1$ and $D_2$.

Let $E$ be event: $D_1 + D_2 = 4$.

What is $P(E)$?

$|S| = 36$

$E = \{(1,3), (2,2), (3,1)\}$

$P(E) = \frac{3}{36} = \frac{1}{12}$

Let $F$ be event: $D_1 = 2$.

What is $P(E, \text{given } F \text{ already observed})$?

$S = \{(2,1), (2,2), (2,3), (2,4), (2,5), (2,6)\}$

$E = \{(2,2)\}$

$P(E) = \frac{1}{6}$
Conditional Probability

The conditional probability of \( E \) given \( F \) is the probability that \( E \) occurs given that \( F \) has already occurred. This is known as conditioning on \( F \).

Written as: \( P(E|F) \)

Means: “\( P(E, \text{given } F \text{ already observed}) \)”

Sample space \( \rightarrow \) all possible outcomes consistent with \( F \) (i.e. \( S \cap F \))

Event \( \rightarrow \) all outcomes in \( E \) consistent with \( F \) (i.e. \( E \cap F \))
Conditional Probability, visual intuition

The conditional probability of $E$ given $F$ is the probability that $E$ occurs given that $F$ has already occurred. This is known as conditioning on $F$.

\[ P(E) = \frac{8}{50} \approx 0.16 \]
\[ P(E|F) = \frac{3}{14} \approx 0.21 \]
Conditional Probability, equally likely outcomes

The conditional probability of $E$ given $F$ is the probability that $E$ occurs given that $F$ has already occurred. This is known as conditioning on F.

With equally likely outcomes:

$$\Pr(E|F) = \frac{\# \text{ of outcomes in } E \text{ consistent with } F}{\# \text{ of outcomes in } S \text{ consistent with } F} = \frac{|EF|}{|SF|} = \frac{|EF|}{|F|}$$

$P(E) = \frac{8}{50} \approx 0.16$

$P(E|F) = \frac{3}{14} \approx 0.21$
Conditional Probability, equally likely outcomes

The **conditional probability** of $E$ given $F$ is the probability that $E$ occurs given that $F$ has already occurred. This is known as conditioning on $F$.

With equally likely outcomes:

$$\Pr(E|F) = \frac{\text{# of outcomes in } E \text{ consistent with } F}{\text{# of outcomes in } S \text{ consistent with } F} = \frac{|EF|}{|SF|} = \frac{|EF|}{|F|}$$

- $P(E) = \frac{8}{50} \approx 0.16$
- $P(E|F) = \frac{3}{14} \approx 0.21$
Conditional probability in general

General definition of conditional probability:

\[ P(E | F) = \frac{P(EF)}{P(F)} \]

The Chain Rule (aka Product rule):

\[ P(EF) = P(F)P(E | F) \]

What if \( P(F) = 0 \)?

- \( P(E | F) \) undefined
- Congratulations! Observed impossible

These properties hold even when outcomes are not equally likely.
### Notation

<table>
<thead>
<tr>
<th>And</th>
<th>Or</th>
<th>Given</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(E \text{ and } F)$</td>
<td>$P(E \text{ or } F)$</td>
<td>$P(E</td>
</tr>
<tr>
<td>$P(E,F)$</td>
<td>$P(E \cup F)$</td>
<td>$P(E</td>
</tr>
<tr>
<td>$P(EF)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P(E \cap F)$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Probability of $E$ given $F$ and $G$
Baby Poop

\[ P(E|F) = \frac{P(EF)}{P(F)} \]

\[ P(EF) = P(F)P(E|F) \]

In the morning when she wakes up, a baby has a 50% chance of having pooped. The chance that a baby cries given that she has pooped is 50%. What is the probability that a baby has pooped, and cries.

https://cs109psets.netlify.app/win22/lecture4/poop
Generalized Chain Rule

\[ \Pr(E_1 \text{ and } E_2 \text{ and } E_3 \text{ and } \ldots E_n) \]

\[ = \Pr(E_1) \cdot \Pr(E_2|E_1) \cdot \Pr(E_3|E_1, E_2) \cdot \ldots \cdot \Pr(E_n|E_1, E_2 \ldots E_{n-1}) \]
NETFLIX

and Learn
What is the probability that a user will watch Life is Beautiful?

\[ P(E) \]

\[ S = \{ \text{Watch, Not Watch} \} \]

\[ E = \{ \text{Watch} \} \]

\[ P(E) = \frac{1}{2} ? \]
What is the probability that a user will watch Life is Beautiful?

\[ P(E) \]
What is the probability that a user will watch Life is Beautiful?

\[ P(E) \]

\[
P(E) = \lim_{n \to \infty} \frac{n(E)}{n} \approx \frac{\# \text{people who watched movie}}{\# \text{people on Netflix}}
\]

\[ P(E) = \frac{10,234,231}{50,923,123} = 0.20 \]
Let $E$ be the event that a user watches the given movie.

\[
P(E|F) = \frac{P(EF)}{P(F)}
\]

Definition of Cond. Probability
Netflix and Learn

Let $E$ = a user watches Life is Beautiful.
Let $F$ = a user watches CODA.

What is the probability that a user watches Life is Beautiful, given they watched Amelie?

$$P(E|F)$$

$$P(E|F) = \frac{P(EF)}{P(F)} = \frac{\text{# people who have watched both}}{\text{# people on Netflix}} \cdot \frac{\text{# people who have watched Amelie}}{\text{# people who have watched Amelie}}$$

$$= \frac{\text{# people who have watched both}}{\text{# people who have watched Amelie}}$$

$$\approx 0.42$$
Let $E$ be the event that a user watches the given movie. Let $F$ be the event that the same user watches CODA (2021).

\[
P(E) = 0.19 \quad P(E) = 0.32 \quad P(E) = 0.20 \quad P(E) = 0.09 \quad P(E) = 0.20
\]

\[
P(E|F) = 0.14 \quad P(E|F) = 0.35 \quad P(E|F) = 0.20 \quad P(E|F) = 0.72 \quad P(E|F) = 0.42
\]
Machine Learning is:

Probability + Data + Computers
Law of Total Probability
Baby Poop Redux

In the morning when she wakes up, a baby has a 50% chance of having pooped. The chance that a baby cries given that she has pooped is 50%. What is the probability that a baby has pooped, and cries.

Other interesting questions (coming soon):
Probability of crying

What information do you need?
Relationship Between Probabilities

\[ P(E \text{ and } F) \]

Chain rule (Product rule)

Definition of conditional probability

\[ P(E | F) \]

Law of Total Probability

\[ P(E) \]
Law of Total Probability

Say $E$ and $F$ are events in $S$

$E = EF \cup EF^c$

$P(E) = P(EF) + P(EF^c)$
Law of Total Probability

Thm Let $F$ be an event where $P(F) > 0$. For any event $E$,

$$P(E) = P(E|F)P(F) + P(E|F^C)P(F^C)$$

Proof

1. $E = (EF)$ or $(EF^C)$

2. $P(E) = P(EF) + P(EF^C)$ Since $F$ and $F^C$ are disjoint

3. $P(E) = P(E|F)P(F) + P(E|F^C)P(F^C)$ Probability of or for disjoint

Chain rule (product rule)
Law of Total Probability

Here $F$ is like a “background event”. You know your event conditioned on the background.
Law of Total Probability

Here $F$ is like a “background event”. You know your event conditioned on the background.
In the morning when she wakes up, a baby has a 50% chance of having pooped. The chance that a baby cries given that she has pooped is 50%. What is the probability that a baby has pooped, and cries.

Other interesting questions (coming soon):
Probability of crying (T)
What information do you need?
Probability of crying given no poop.

Recall that T is crying and E is poop

\[ P(T) = P(T|E)P(E) + P(T|E^C)P(E^C) \]
You have bacteria in your gut which is causing a disease. 10% have a mutation which makes them resistant to anti-biotics. You take half a course of anti-biotics...

Probability a bacteria survives given it has the mutation: 20%
Probability a bacteria survives given it doesn't have the mutation: 1%

What is the probability that a randomly chosen bacteria survives?

Let $E$ be the event that a bacterium survives. Let $M$ be the event that a bacteria has the mutation. By the Law of Total Probability (LOTP):

\[
Pr(E) = Pr(E \text{ and } M) + Pr(E \text{ and } M^C) \\
= Pr(E|M)Pr(M) + Pr(E|M^C)Pr(M^C) \\
= 0.20 \cdot 0.10 + 0.01 \cdot 0.90 \\
= 0.029
\]
Law of Total Probability

\[ P(E) = P(EF) + P(EF^C) \]
\[ = P(E|F)P(F) + P(E|F^C)P(F^C) \]
Law of Total Probability

**Thm**

For **mutually exclusive events** $B_1, B_2, \ldots, B_n$

s.t. $B_1 \cup B_2 \cup \cdots \cup B_n = S,$

\[
P(E) = \sum_i P(B_i \cap E)
\]

\[
= \sum_i P(E|B_i)P(B_i)
\]
Real question. What is the probability that a surviving bacteria has the mutation?

\[ \Pr(\text{Mutation} \mid \text{Survives}) \]

\[ \Pr(\text{M} \mid E) \]
Real Question: $\Pr(M \mid E)$?

You have bacteria in your gut which is causing a disease. 10% have a mutation which makes them resistant to anti-biotics. You take half a course of anti-biotics...

Probability a bacteria survives given it has the mutation: 20%
Probability a bacteria survives given it doesn't have the mutation: 1%
What is the probability that a randomly chosen bacteria survives?

Let $E$ be the event that our bacterium survives. Let $M$ be the event that a bacteria has the mutation. By the Law of Total Probability (LOTP):

$$\Pr(E) = \Pr(E \text{ and } M) + \Pr(E \text{ and } M^C)$$

LOTP

$$= \Pr(E \mid M)\Pr(M) + \Pr(E \mid M^C)\Pr(M^C)$$

Chain Rule

$$= 0.20 \cdot 0.10 + 0.01 \cdot 0.90$$

Substituting

$$= 0.029$$
Real Question: $\Pr(M \mid E)$?

You have bacteria in your gut which is causing a disease. 10% have a mutation which makes them resistant to anti-biotics. You take half a course of anti-biotics...

$\Pr(E \mid M) = 0.20$

$\Pr(E \mid M^C) = 0.01$

What is the probability that a randomly chosen bacteria survives?

Let $E$ be the event that our bacterium survives. Let $M$ be the event that a bacteria has the mutation. By the Law of Total Probability (LOTP):

$$\Pr(E) = \Pr(E \text{ and } M) + \Pr(E \text{ and } M^C)$$

$$= \Pr(E \mid M) \Pr(M) + \Pr(E \mid M^C) \Pr(M^C)$$

$$= 0.20 \cdot 0.10 + 0.01 \cdot 0.90$$

$$= 0.029$$
Relationship Between Probabilities

\[ P(E \text{ and } F) \]

Chain rule (Product rule) \[\uparrow\]
Definition of conditional probability \[\downarrow\]

\[ P(E|F) \]

Law of Total Probability

\[ P(E) \]
Relationship Between Probabilities

\[ P(E \text{ and } F) \]

Chain rule (Product rule) \[ \uparrow \quad \downarrow \]
Definition of conditional probability

\[ P(E|F) \]

Law of Total Probability

\[ P(E) \]

Bayes’ Theorem

\[ P(F|E) \]
Bayes’ Theorem
Thomas Bayes

Rev. Thomas Bayes (~1701-1761):
British mathematician and Presbyterian minister

He looked remarkably similar to Charlie Sheen
(but that’s not important right now)
I want to calculate
P(State of the world \( F \) | Observation \( E \))
It seems so tricky!…

The other way around is easy
P(Observation \( E \) | State of the world \( F \))
What options do I have, chief?
Thomas Bayes

Want $P(F | E)$. Know $P(E | F)$

\[
P(F|E) = \frac{P(EF)}{P(E)} \quad \text{Def. of Conditional Prob.}
\]

\[
= \frac{P(E|F)P(F)}{P(E)} \quad \text{Chain Rule}
\]

A little while later...
Bayes’ Theorem

For any events $E$ and $F$ where $P(E) > 0$ and $P(F) > 0$, we have

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

Proof

2 steps! See board

Expanded form:

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

Proof

1 more step! See board
Detecting spam email

We can easily calculate how many spam emails contain “Dear”:

\[ P(E | F) = P(\text{“Dear”} | \text{Spam email}) \]

But what is the probability that an email containing “Dear” is spam?

\[ P(F | E) = P(\text{Spam email} | \text{“Dear”}) \]
(silent drumroll)
Detecting spam email

- 60% of all email in 2016 is spam.
- 20% of spam has the word “Dear”
- 1% of non-spam (aka ham) has the word “Dear”

You get an email with the word “Dear” in it. What is the probability that the email is spam?

1. Define events & state goal
2. Identify known probabilities
3. Solve

Let: $E$: “Dear”, $F$: spam
Want: $P(\text{spam} | \text{“Dear”})$

$P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^c)P(F^c)}$ Bayes’ Theorem

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

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

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

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

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

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

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

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

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

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

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

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

$P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^c)P(F^c)}$
Bayes’ Theorem terminology

- 60% of all email in 2016 is spam. $P(F)$
- 20% of spam has the word “Dear” $P(E|F)$
- 1% of non-spam (aka ham) has the word “Dear” $P(E|F^C)$

You get an email with the word “Dear” in it. What is the probability that the email is spam?

Want: $P(F|E)$

\[
P(F|E) = \frac{P(E|F)P(F)}{P(E)}
\]

This is known as the posterior probability, which is the likelihood of the hypothesis given the evidence, multiplied by the prior probability of the hypothesis, divided by the normalization constant which is the probability of the evidence.
A test is 98% effective at detecting SARS
- However, test has a “false positive” rate of 1%
- 0.5% of US population has SARS
- Let $E =$ you test positive for SARS with this test
- Let $F =$ you actually have SARS
- What is $P(F \mid E)$?

Solution:

$$P(F \mid E) = \frac{P(E \mid F) \ P(F)}{P(E \mid F) \ P(F) + P(E \mid F^c) \ P(F^c)}$$

$$P(F \mid E) = \frac{(0.98)(0.005)}{(0.98)(0.005) + (0.01)(1 - 0.005)} \approx 0.330$$
Intuition Time
Bayes Theorem Intuition
Bayes Theorem Intuition

People with SARS

All People
Bayes Theorem Intuition

All People

People who test positive

All People
Bayes Theorem Intuition

People who test positive

People with SARS

All People
Bayes Theorem Intuition

Conditioning on a positive result changes the sample space to this:

- People who test positive and have SARS
- People who test positive

\[ \approx 0.330 \]
Bayes' Theorem Intuition

Conditioning on a positive result changes the sample space to this:

\[ P(F)P(E|F) + P(F^c)P(E|F^c) \]

\[ \approx 0.330 \]
Bayes’ Theorem Intuition

- People with positive test
- People with SARS
- All People
Bayes Theorem Intuition

Say we have 1000 people:

5 have SARS and test positive, 985 do not have SARS and test negative.
10 do not have SARS and test positive. \approx 0.333
Bayes Theorem Intuition

Conditioned on just those that test positive:

5 have SARS and test positive, 985 do not have SARS and test negative.
10 do not have SARS and test positive. \( \approx 0.333 \)

Notice that all the people with SARS are here, but the group is still mainly folks without SARS.
Why it is still good to get tested

<table>
<thead>
<tr>
<th></th>
<th>SARS +</th>
<th>SARS –</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test +</td>
<td>0.98 = P(E</td>
<td>F)</td>
</tr>
<tr>
<td>Test –</td>
<td>0.02 = P(E^c</td>
<td>F)</td>
</tr>
</tbody>
</table>

- Let $E^c$ = you test **negative** for SARS with this test
- Let $F$ = you actually have SARS
- What is $P(F | E^c)$?

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

$$P(F | E^c) = \frac{(0.02)(0.005)}{(0.02)(0.005) + (0.99)(1 - 0.005)} \approx 0.0001$$
Bayes' Theorem and Location

\[ P(L_1) \quad P(L_2) \]

\[ P(L_5) \]

Before Observation
Bayes' Theorem and Location

Know: $P(O|L_i)$

Before Observation

After Observation

$P(L_5)$
Bayes' Theorem and Location

Know: \( P(O|L_i) \)

\[
P(L_5|O) = \frac{P(O|L_5)P(L_5)}{P(O)}
\]
Bayes' Theorem and Location

Before Observation

After Observation

\[ P(L_5) \]

\[ P(L_5|O) = \frac{P(O|L_5)P(L_5)}{\sum_i P(O|L_i)P(L_i)} \]
Bayes' Theorem and Location

\[ P(L_5 | O) = \frac{P(O | L_5) P(L_5)}{\sum_i P(O | L_i) P(L_i)} \]
Bayes' Theorem and Location

\[ P(L_5 | O) = \frac{P(O | L_5) P(L_5)}{\sum_i P(O | L_i) P(L_i)} \]
Monty Hall Problem
Monty Hall Problem

and Wayne Brady
Behind one door is a prize (equally likely to be any door).

Behind the other two doors is nothing

1. We choose a door
2. Host opens 1 of other 2 doors, revealing nothing
3. We are given an option to change to the other door.

Should we switch?

Note: If we don’t switch, P(win) = 1/3 (random)

We are comparing P(win) and P(win|switch).
If we switch

Without loss of generality, say we pick A (out of Doors A,B,C).

\[\begin{align*}
\text{A = prize} & \quad \text{B = prize} & \quad \text{C = prize} \\
\text{• Host opens B or C} & \quad \text{• Host must open C} & \quad \text{• Host must open B} \\
\text{• We switch} & \quad \text{• We switch to B} & \quad \text{• We switch to C} \\
\text{• We always lose} & \quad \text{• We always win} & \quad \text{• We always win} \\
\text{P(win | A prize, picked A, switched) = 0} & \quad \text{P(win | B prize, picked A, switched) = 1} & \quad \text{P(win | C prize, picked A, switched) = 1}
\end{align*}\]

\[\text{P(win | picked A, switched) = } \frac{1}{3} \times 0 + \frac{1}{3} \times 1 + \frac{1}{3} \times 1 = \frac{2}{3}\]

You should switch.
Monty Hall, 1000 envelope version

Start with 1000 envelopes (of which 1 is the prize).

1. You choose 1 envelope.
   \[
   \frac{1}{1000} = P(\text{envelope is prize})
   \]
   \[
   \frac{999}{1000} = P(\text{other 999 envelopes have prize})
   \]

2. I open 998 of remaining 999 (showing they are empty).
   \[
   \frac{999}{1000} = P(998 \text{ empty envelopes had prize})
   \]
   \[
   + P(\text{last other envelope has prize})
   \]
   \[
   = P(\text{last other envelope has prize})
   \]

3. Should you switch?  
   No: \( P(\text{win without switching}) = \frac{1}{\text{original # envelopes}} \)
   Yes: \( P(\text{win with new knowledge}) = \frac{1}{\text{original # envelopes} - 1} \)
Marilyn Vos Savant

Ask Marilyn™

BY MARILYN VOS SAVANT