Conditional Probability and Bayes
Announcements

• Pset #1 is due on Friday

• Section assignments will be sent out today. Can’t make your time or need a swap? See the ed post!

• PSet Party Wed at 9pm

• https://cs109psets.netlify.app/win22/lecture4/
Peer Teaching by Ali and Juliette
Review
Combinatorial

How many unique shuffles of a card deck are there?

52!

8065817517094387857166063685
6403766975289505440883277824
000000000000
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!
\[ P(E^c) = 1 - P(E) \]

\[ P(E \cup E^c) = P(E) + P(E^c) \]

Axiom 3. Since \( E \) and \( E^c \) are mutually exclusive

\[ P(S) = P(E) + P(E^c) \]

Since everything must either be in \( E \) or \( E^c \)

\[ 1 = P(E) + P(E^c) \]

Axiom 2

\[ P(E^c) = 1 - P(E) \]

Rearrange
Let it find you.

SERENDIPITY

the effect by which one accidentally stumbles upon something truly wonderful, especially while looking for something entirely unrelated.
WHEN YOU MEET YOUR BEST FRIEND

Somewhere you didn't expect to.
Serendipity

- Say the population of Stanford is 17,000 people
  - You are friends with?
  - Walk into a room, see 268 random people.
  - What is the probability that you see someone you know?
  - Assume you are equally likely to see each person at Stanford
Many times it is easier to calculate $P(E^C)$.
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 $P(\text{sum} = 7)$?

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

$$E = \text{In blue}$$
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
Sum of Two Die = 4?

Roll two 6-sided dice. What is probability the sum = 4?
Let E be the event that the sum is 4

\[
S = \{ [1,1], [1,2], [1,3], [1,4], [1,5], [1,6], [2,1], [2,2], [2,3], [2,4], [2,5], [2,6], [3,1], [3,2], [3,3], [3,4], [3,5], [3,6], [4,1], [4,2], [4,3], [4,4], [4,5], [4,6], [5,1], [5,2], [5,3], [5,4], [5,5], [5,6], [6,1], [6,2], [6,3], [6,4], [6,5], [6,6] \}
\]

\[
E = \text{Each outcome}
\]
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 $P(E)$?

$|S| = 36$

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

$P(E) = 3/36 = 1/12$

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

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

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

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

$P(E) = 1/6$
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 $F$.

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

$$P(E|F) = \frac{3}{14} \approx 0.21$$
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*
NETFLIX and Learn
What is the probability that a user will watch Life is Beautiful?

\[ P(E) \]

S = \{Watch, Not Watch\}

E = \{Watch\}

\[ P(E) = \frac{1}{2} \]
Netflix and Learn

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.

$P(E) = 0.19$  
$P(E) = 0.32$  
$P(E) = 0.20$  
$P(E) = 0.09$  
$P(E) = 0.20$
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?

$$P(E|F)$$

$$P(E|F) = \frac{P(EF)}{P(F)} = \frac{\# \text{ people who have watched both}}{\# \text{ people who have watched Amelie}} \approx 0.42$$
Let $E$ be the event that a user watches the given movie. Let $F$ be the event that the same user watches CODA (2021).

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

\[
P(E|F) = 0.14 \quad P(E|F) = 0.35 \quad P(E|F) = 0.20 \quad P(E|F) = 0.72 \quad P(E|F) = 0.42
\]
Machine Learning is:
Probability + Data + Computers
### 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)$</td>
<td></td>
</tr>
<tr>
<td>$P(EF)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P(E \cap F)$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Given

- Probability of E given F and G

- $P(E|F, G)$
In the morning when she wakes up, a baby has a 50% chance of having pooped. The chance that a baby cries \textbf{given} that she has pooped is 50%. What is the probability that a baby \textbf{has pooped, and cries}.  

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

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

\[ = \Pr(E_1) \cdot \Pr(E_2|E_1) \cdot \Pr(E_3|E_1, E_2) \cdots \Pr(E_n|E_1, E_2 \ldots E_{n-1}) \]
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 of crying, unconditioned?

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 $S$

$$P(E) = P(EF) + P(EF^C)$$
$$= P(E|F)P(F) + P(E|F^C)P(F^C)$$
Law of Total Probability

Say E and F are events in S

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

Say E and F are events in S

\[
P(E) = P(EF) + P(EF^C)
\]

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

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

\[ P(E) = P(EF) + P(EF^C) \]
\[ = P(E|F)P(F) + P(E|F^C)P(F^C) \]
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 $(EF^C)$  
   Since $F$ and $F^C$ are disjoint

2. $P(E) = P(EF) + P(EF^C)$  
   Probability of or for disjoint

3. $P(E) = P(E|F)P(F) + P(E|F^C)P(F^C)$  
   Chain rule (product rule)
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%.

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):

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

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):

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

LOTP

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

Chain Rule

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

Substituting

$$= 0.029$$
Law of Total Probability

\[ P(E) = P(E|F)P(F) + P(E|F^C)P(F^C) \]
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?

\[
\Pr \left( \text{Mutation} \mid \text{Survives} \right) \quad \Pr \left( M \mid E \right)
\]
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):

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

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

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

$$
= 0.029
$$

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...

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?
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)

Definition of
conditional probability

\[ P(E | F) \]

Law of Total
Probability

\[ P(E) \]
Relationship Between Probabilities

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

Chain rule
(Product rule)

Definition of conditional probability

\[ P(E|F) \]

Law of Total Probability

Bayes’ Theorem

\[ P(E) \]

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

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

He looked remarkably similar to Sean Astin
(but that’s not important right now)
I want to calculate $P(\text{State of the world } F \mid \text{Observation } E)$

It seems so tricky!…

The other way around is easy
$P(\text{Observation } E \mid \text{State of the world } F)$

What options do I have, chief?

$P( F \mid E )$

$P( E \mid F )$
Thomas Bayes

Want \( P(F \mid E) \). Know \( P(E \mid F) \)

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

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

\[
= \frac{P(E \mid F)P(F)P(F)}{P(E\mid F)P(F)P(F) + P(E \mid F)P(F^C)P(F^C)} \quad \text{LOTP}
\]
Bayes’ Theorem

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

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

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

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

Bayes’ Theorem
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.

Want: $P(F|E)$

What is the probability that the email is spam?

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

Solution:

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

\[
P(F | E) = \frac{(0.98)(0.005)}{(0.98)(0.005) + (0.01)(1 - 0.005)}
\]

\[
\approx 0.330
\]
Intuition Time
Bayes Theorem Intuition

All People
Bayes Thorem Intuition

People with SARS

All People
Bayes Theorem Intuition

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
- People who test positive and have SARS

\[ \approx 0.330 \]
Bayes Thorem Intuition

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

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

\[
\approx 0.330
\]
Bayes Theorem Intuition

People with positive test

People with SARS

All People
Bayes Thorem 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:

Notice that all the people with SARS are here, but the group is still mainly folks without SARS

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
Why it is still good to get tested

Let $E^c = \text{you test negative for SARS with this test}$
Let $F = \text{you actually have SARS}$
What is $P(F \mid E^c)$?

<table>
<thead>
<tr>
<th></th>
<th>SARS +</th>
<th>SARS –</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test +</td>
<td>$0.98 = P(E \mid F)$</td>
<td>$0.01 = P(E \mid F^c)$</td>
</tr>
<tr>
<td>Test –</td>
<td>$0.02 = P(E^c \mid F)$</td>
<td>$0.99 = P(E^c \mid F^c)$</td>
</tr>
</tbody>
</table>

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

\[
P(F \mid E^c) = \frac{(0.02)(0.005)}{(0.02)(0.005) + (0.99)(1 - 0.005)} \approx 0.0001
\]
Multiple Choice Theory

Let's consider the relationship between knowing the concepts used in a multiple choice midterm question, and getting the question correct, taking into account guessing and making silly mistakes.

Let 3/4 be the probability that a learner knows the concepts to a midterm question.

Let 1/4 be the probability that a learner gets the answer correct if they don't know the concepts.

Let 1/10 be the probability that a learner gets the question incorrect given they do know the concepts.

What is the probability they know the concept, given they answered correct?
Advanced
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
Bayes' Theorem and Location

Know: $P(O|L_i)$

Before Observation: $P(L_5)$

After Observation: $P(L_5|O)$

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

$$P(L_5) \rightarrow P(L_5|O)$$

Before Observation

After Observation

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

Before Observation

\[ P(L_5) \]

After Observation

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

\[
P(L_5 | O) = \frac{P(O | L_5)P(L_5)}{\sum_i P(O | L_i)P(L_i)}
\]
Come on Wednesday and we will gamble!