Lecture 5: Conditional Probability Introduction
Belonging and CS Tas Research Study

Leah Perlmutter (she/her)
leahperl@uw.edu
tinyurl.com/belonging-study
Definition. A (discrete) probability space is a pair $(\Omega, \mathbb{P})$ where:

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

Some outcome must show up

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

Let $\Omega$ denote the sample space and $E, F \subseteq \Omega$ be events. Note this is more general to any probability space (not just uniform).

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

• Time Permitting:
  – Bayes Theorem
  – Law of Total Probability
  – Bayes Theorem + Law of Total Probability
  – More Examples
What’s the probability that someone likes ice cream \textit{given} they like donuts?
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

\[
P(A \cap B) = P(A|B)P(B)
\]

Example:

- $A =$ Ices cream
- $B =$ Donuts

\[
P(A \cap B) = P(B|A)P(A)P(A \cap B) = \frac{4}{20} \cdot \frac{20}{70} \cdot \frac{7}{70}
\]

\[
P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{4/70}{20/20} = \frac{7}{20}
\]
Reversing Conditional Probability

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

No! The following 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 $P(B|A)$?

B = red die is 1
A = sum is 4

| P(B) | P(B|A) |
|------|--------|
| 1/6  | 1/6    |
| 1/6  | 1/3    |
| 1/6  | 3/36   |
| 1/9  | 1/3    |

\[
P(B) = \frac{|B|}{|\Omega|} = \frac{6}{36} = \frac{1}{6}\]

\[
P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{\frac{|B \cap A|}{|\Omega|}}{\frac{|A|}{|\Omega|}} = \frac{\frac{1}{36}}{\frac{1}{3}} = \frac{1}{3}\]
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 $P(B|A)$?

$B = \text{red die is 1}$

$A = \text{sum is 4}$

|   | $P(B)$ | $P(B|A)$ |
|---|--------|----------|
| a) | 1/6    | 1/6      |
| b) | 1/6    | 1/3      |
| c) | 1/6    | 3/36     |
| d) | 1/9    | 1/3      |

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

$P(B|A) = \frac{1}{3}$
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 51st coin is “heads”?

\[ \mathcal{A} = \text{first 50 coins are “tails”} \]
\[ B = 51\text{st coin is ”heads”} \]

\[ \mathbb{P}(B | A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)} = \frac{|B \cap A|}{|A|} = \frac{1}{2} = \mathbb{P}(B) \]

\[ |\Omega| = 2^{51} \]
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 51st coin is “heads”?

\[ \mathcal{A} = \text{first 50 coins are “tails”} \]
\[ B = 51^{\text{st}} \text{ coin is ”heads”} \]

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

**Gambler’s fallacy** = Feels like it’s time for “heads”!?
Agenda

• Conditional Probability
• Time Permitting:
  – 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 $B$)
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

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.

\[
P(S|F) = \frac{P(F|S)P(S)}{P(F)} = \frac{0.7 \cdot 0.8}{P(F)}
\]

\[
P(F|S^c) = 0.1
\]

\[
P(F|S) = 0.7 \Rightarrow \frac{P(F)}{P(S)} = 0.8
\]
Agenda

• Conditional Probability
• Time Permitting:
  – 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 $\Omega$ if

(Exhaustive)

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

(Pairwise Mutually Exclusive)

$$\forall i \forall j \neq i E_i \cap E_j = \emptyset$$
Law of Total Probability (Idea)

If we know $E_1, E_2, ..., E_n$ partition $\Omega$, what can we say about $P(F)$?
Law of Total Probability (LTP)

Definition. If events $E_1, E_2, \ldots, E_n$ partition the sample space $\Omega$, 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 $P(R)$?
Sequential Process – Non-Uniform Case

- Right pocket: Two red, two green
- Right pocket: One red, two green
- Alice picks a random ball from a random pocket
Sequential Process – Non-Uniform Case

\[
P(R) = P(R \cap \text{Left}) + P(R \cap \text{Right}) \quad \text{(Law of total probability)}
\]

\[
= P(\text{Left}) \times P(R | \text{Left}) + P(\text{Right}) \times P(R | \text{Right})
\]

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

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

\[
1/3 = P(R | R) \quad \text{and} \quad 2/3 = P(G | R)
\]
Agenda

• Conditional Probability
• Time Permitting:
  – 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 $E_1, E_2, ..., E_n$ 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
• Time Permitting:
  – 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.
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 have Zika (event $Z$) if you test positive (event $T$).

A) Less than 0.25
B) Between 0.25 and 0.5
C) Between 0.5 and 0.75
D) Between 0.75 and 1
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 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”)
– However, the test may yield a “false positive” 1% of the time
– 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

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

Demo

Have zika blue, don’t pink
Philosophy – Updating Beliefs

While it’s not 98% that you have the disease, your beliefs changed **drastically**

\[ Z = \text{you have Zika} \]
\[ T = \text{you test positive for Zika} \]

Prior: \( P(Z) \)

I have a 0.5% chance of having Zika

Receive positive test result

I now have a 33% chance of having Zika after the test.

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 Define a Probability Space

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

Example. $\mathbb{P}(B^c | A) = 1 - \mathbb{P}(B | A)$
Conditional Probability Define a Probability Space

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

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

**Formally.** $(\Omega, \mathbb{P})$ is a probability space + $\mathbb{P}(A) > 0$

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