Lecture 6: More Conditional Probability
Agenda

- Review: Conditional Probability, Bayes
- Law of Total Probability (w/ Bayes)
- Chain Rule
- Independence
- Conditional Independence
- Assumptions and Correlation
- Bonus: Monty Hall Problem
Last Class:

• Conditional Probability
• Bayes Theorem

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

\[ P(A|B) = \frac{P(B|A)P(A)}{P(B)} \]

\[ P(A|B) \neq P(B|A) \]
Agenda

• Review: Conditional Probability, Bayes
• **Law of Total Probability (w/ Bayes)**
• Chain Rule
• Independence
• Conditional Independence
• Assumptions and Correlation
• Bonus: Monty Hall Problem
Law of Total Probability (Idea)

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

**Definition.** If events $E_1, E_2, ..., E_n$ 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 shows

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

\[ P(R) = \frac{1}{2} \]

What is \( P(R) \)?
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

\[
P(R) = \frac{1}{2} \cdot \frac{1}{2} + \frac{2}{3} \cdot \frac{1}{2} = \frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{2}
\]
Sequential Process – Non-Uniform Case

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

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

(Law of total probability)

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

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

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

$P(Z|T) =$ ?

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”) \( P(T|Z) = 0.98 \)
- However, the test yields a “false positive” 1% of the time \( P(T|\neg Z) = 0.01 \)
- 0.5% of the US population has Zika. \( P(Z) = 0.005 \)

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

\[
P(Z|T) = \frac{P(T|Z)P(Z)}{P(T)} = \frac{P(T|Z)P(Z)}{P(T|Z)P(Z) + P(T|\neg Z)P(\neg Z)}
\]

\[
= \frac{0.98 \cdot 0.005}{0.98 \cdot 0.005 + 0.01 \cdot 0.995} = 0.33
\]
Example – Zika Testing

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 people 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

\[
\frac{P(Z|\neg T)}{P(Z|Z, T)}
\]

\[
P(Z|T)
\]

\[
\frac{P(Z|T, \neg T)}{P(Z|Z, T)}
\]

I have a 0.5% chance of having Zika

T = you test positive for Zika

Prior: \( P(Z) \)

Receive positive test result

Posterior: \( P(Z|T) \)

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

Suppose we know the following Zika stats

– A test is 98% effective at detecting Zika (“true positive”)
  \[ P(T|Z) = 0.98 \]
– 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 \( \overline{T} \)) if you have Zika (event \( Z \))? 

\[ P(T^c | Z) = 0.02 \]

\[ P(T^c | Z) = 1 - P(T | Z) 
\[ 1 - 0.98 \]
Conditional Probability Define a Probability Space

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

\[ P(B^c | A) = 1 - P(B | A) \]

Example. $P(B^c | A) = 1 - P(B | A)$
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 + $\mathbb{P}(A) > 0$

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

• Review: Conditional Probability, Bayes
• Law of Total Probability (w/ Bayes)
• **Chain Rule**
• Independence
• Conditional Independence
• Assumptions and Correlation
• Bonus: Monty Hall Problem
Chain Rule

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

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

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

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

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

**Theorem. (Chain Rule)** For events \( A_1, A_2, \ldots, A_n \),

\[ P(A_1 \cap \cdots \cap A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \]

\[ \cdots \cdot P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1}) \]

An easy way to remember: We have \( n \) tasks and we can do them **sequentially**, conditioning on the outcome of previous tasks.
Chain Rule Example

Have a Standard 52-Card Deck. Shuffle It, and draw the top 3 cards \textbf{in order}. (uniform probability space).

What is $P(A \cap B \cap C)$?

- $A$: Ace of Spades First
- $B$: 10 of Clubs Second
- $C$: 4 of Diamonds Third

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

\[
= \frac{1}{52} \cdot \frac{1}{51} \cdot \frac{1}{50}
\]

\[
P(52, 3)
\]
Chain Rule Example

Have a Standard 52-Card Deck. Shuffle It, and draw the top 3 cards in order. (uniform probability space).

What is \( P(A \cap B \cap C) \)?

- **A**: Ace of Spades First
- **B**: 10 of Clubs Second
- **C**: 4 of Diamonds Third

\[
P(A) \cdot P(B|A) \cdot P(C|A \cap B) = \frac{1}{52} \cdot \frac{1}{51} \cdot \frac{1}{50}
\]
Agenda

• Review: Conditional Probability, Bayes
• Law of Total Probability (w/ Bayes)
• Chain Rule
• **Independence**
• Conditional Independence
• Assumptions and Correlation
• Bonus: Monty Hall Problem
**Independence**

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

\[ \frac{P(B)}{P(A)} = P(B) \]

**Definition.** If two events $A$ and $B$ are independent then

\[ P(A \cap B) = P(A) \cdot P(B). \]

Alternatively,

- If $P(A) \neq 0$, equivalent to $P(B|A) = P(B)$
- If $P(B) \neq 0$, equivalent to $P(A|B) = P(A)$

“The probability that $B$ occurs after observing $A$” -- Posterior

= “The probability that $B$ occurs” -- Prior
Example -- Independence

Toss a coin 3 times. Each of 8 outcomes equally likely.

• \( A = \{ \text{at most one T} \} = \{\text{HHH}, \text{HHT}, \text{HTH}, \text{THH}\} \)

• \( B = \{ \text{at most 2 Heads}\} = \{\text{HHH}\}^c \)

Independent?

\[
P(A \cap B) = P(A) \cdot P(B) \quad ?
\]

\[
P(A) = \frac{1}{2}
\]

\[
P(B) = \frac{2}{3}
\]

\[
P(A \cap B) = \frac{3}{8}
\]

Poll:
A. Yes, independent
B. No
Often probability space \((\Omega, \mathbb{P})\) is **defined** using independence

*Events generated independently ➔ their probabilities satisfy independence*

Not necessarily

This can be counterintuitive!
Example – Network Communication

Each link works with the probability given, independently. What’s the probability A and D can communicate?

$$P(AD) = ? P((AB \lor BD) \lor (AC \lor CD))$$

$$= P(AB \lor BD) + P(AC \lor CD) - P(AB \lor BD \land AC \lor CD)$$

$$= P(AB)P(BD) + P(AC)P(CD) - P(AB \land BD \land AC \land CD)$$

$$\frac{pq + rs - pqs}{pq + rs - pqs}$$
Example – Network Communication

Each link works with the probability given, **independently**. What’s the probability A and D can communicate?

\[
P(AD) = P(AB \cap BD \text{ or } AC \cap CD)
\]

\[
= P(AB \cap BD) + P(AC \cap CD) - P(AB \cap BD \cap AC \cap CD)
\]

\[
P(AB \cap BD) = P(AB) \cdot P(BD) = pq
\]

\[
P(AC \cap CD) = P(AC) \cdot P(CD) = rs
\]

\[
P(AB \cap BD \cap AC \cap CD) = P(AB) \cdot P(BD) \cdot P(AC) \cdot P(CD) = pqrs
\]
Example – Biased coin

We have a biased coin comes up Heads with probability $2/3$; Each flip is independent of all other flips. Suppose it is tossed 3 times.

\[
\mathbb{P}(HHH) = \\
\mathbb{P}(TTT) = \\
\mathbb{P}(HTT) = \\
\text{will go over next lecture}
\]
Example – Biased coin

We have a biased coin comes up Heads with probability 2/3, independently of other flips. Suppose it is tossed 3 times.

\[ P(2 \text{ heads in 3 tosses}) = \]

A) \((2/3)^2 \cdot 1/3\)
B) \(2/3\)
C) \(3 \cdot (2/3)^2 \cdot 1/3\)
D) \((1/3)^2\)
Agenda

- Review: Conditional Probability, Bayes
- Law of Total Probability (w/ Bayes)
- Chain Rule
- Independence
- Conditional Independence
- Assumptions and Correlation
- Bonus: Monty Hall Problem

ended here for today
Conditional Independence

**Definition.** Two events $\mathcal{A}$ and $\mathcal{B}$ are **independent** conditioned on $\mathcal{C}$ if $\mathbb{P}(\mathcal{C}) \neq 0$ and $\mathbb{P}(\mathcal{A} \cap \mathcal{B} \mid \mathcal{C}) = \mathbb{P}(\mathcal{A} \mid \mathcal{C}) \cdot \mathbb{P}(\mathcal{B} \mid \mathcal{C})$.

**Plain Independence.** Two events $\mathcal{A}$ and $\mathcal{B}$ are **independent** if

$$\mathbb{P}(\mathcal{A} \cap \mathcal{B}) = \mathbb{P}(\mathcal{A}) \cdot \mathbb{P}(\mathcal{B}).$$

**Equivalence:**
- If $\mathbb{P}(\mathcal{A}) \neq 0$, equivalent to $\mathbb{P}(\mathcal{B} \mid \mathcal{A}) = \mathbb{P}(\mathcal{B})$
- If $\mathbb{P}(\mathcal{B}) \neq 0$, equivalent to $\mathbb{P}(\mathcal{A} \mid \mathcal{B}) = \mathbb{P}(\mathcal{A})$
**Conditional Independence**

**Definition.** Two events $\mathcal{A}$ and $\mathcal{B}$ are independent conditioned on $\mathcal{C}$ if
\[
P(\mathcal{C}) \neq 0 \text{ and } P(\mathcal{A} \cap \mathcal{B} | \mathcal{C}) = P(\mathcal{A} | \mathcal{C}) \cdot P(\mathcal{B} | \mathcal{C}).
\]

**Equivalence:**
- If $P(\mathcal{A} \cap \mathcal{C}) \neq 0$, equivalent to $P(\mathcal{B} | \mathcal{A} \cap \mathcal{C}) = P(\mathcal{B} | \mathcal{C})$.
- If $P(\mathcal{B} \cap \mathcal{C}) \neq 0$, equivalent to $P(\mathcal{A} | \mathcal{B} \cap \mathcal{C}) = P(\mathcal{A} | \mathcal{C})$.

**Plain Independence.** Two events $\mathcal{A}$ and $\mathcal{B}$ are independent if
\[
P(\mathcal{A} \cap \mathcal{B}) = P(\mathcal{A}) \cdot P(\mathcal{B}).
\]

**Equivalence:**
- If $P(\mathcal{A}) \neq 0$, equivalent to $P(\mathcal{B} | \mathcal{A}) = P(\mathcal{B})$.
- If $P(\mathcal{B}) \neq 0$, equivalent to $P(\mathcal{A} | \mathcal{B}) = P(\mathcal{A})$. 

34
Example – More coin tossing

Suppose there is a coin C1 with Pr(Head) = 0.3 and a coin C2 with Pr(Head) = 0.9. We pick one randomly with equal probability and flip that coin twice independently. What is the probability we get all heads?

\[
\Pr(HH) = \Pr(HH \mid C1) \Pr(C1) + \Pr(HH \mid C2) \Pr(C2)
\]
Example – More coin tossing

Suppose there is a coin $C_1$ with $\Pr(\text{Head}) = 0.3$ and a coin $C_2$ with $\Pr(\text{Head}) = 0.9$. We pick one randomly with equal probability and flip that coin 2 times independently. What is the probability we get all heads?

\[
\Pr(\text{HH}) = \Pr(\text{HH} \mid C_1) \Pr(C_1) + \Pr(\text{HH} \mid C_2) \Pr(C_2)
\]

\[
= \Pr(H \mid C_2)^2 \Pr(C_1) + \Pr(H \mid C_2)^2 \Pr(C_2)
\]

\[
= 0.3^2 \cdot 0.5 + 0.9^2 \cdot 0.5 = 0.45
\]

\[
\Pr(H) = \Pr(H \mid C_1) \Pr(C_1) + \Pr(H \mid C_2) \Pr(C_2) = 0.6
\]
Agenda

• Review: Conditional Probability, Bayes
• Law of Total Probability (w/ Bayes)
• Chain Rule
• Independence
• Conditional Independence

• Assumptions and Correlation

• Bonus: Monty Hall Problem
Correlation

- Pick a person at random
  - $A$: event that the person has lung cancer
  - $B$: event that the person is a heavy smoker

- Fact: $\mathbb{P}(A|B) = 1.17 \cdot \mathbb{P}(A)$

- Conclusions?
Correlation

- Pick a person at random
- $A$ : event that the person has lung cancer
- $B$ : event that the person is a heavy smoker

- Fact: $\mathbb{P}(A|B) = 1.17 \cdot \mathbb{P}(A)$

- Conclusions?
  - Lung cancer increases the probability of smoking by 17%.
  - Lung cancer causes smoking.
Causality vs. Correlation

- Events $A$ and $B$ are **positively correlated** if

$$\Pr(A \cap B) > \Pr(A) \cdot \Pr(B)$$

- E.g. smoking and lung cancer.

- But $A$ and $B$ being positively correlated does not mean that $A$ causes $B$ or $B$ causes $A$. 
Causality vs. Correlation

- Events $A$ and $B$ are **positively correlated** if

\[ \mathbb{P}(A \cap B) > \mathbb{P}(A) \cdot \mathbb{P}(B) \]

- But $A$ and $B$ being positively correlated does not mean that $A$ causes $B$ or $B$ causes $A$.

Other examples:
- Tesla owners are more likely to be rich. That does not mean poor people should buy a Tesla to get rich.
- People who go to the opera are more likely to have a good career. That does not mean that going to the opera will improve your career.
- Rabbits eat more carrots and do not wear glasses. Are carrots good for eyesight?
Independence as an assumption

- People often assume it **without justification**.
- Example: A sky diver has two chutes
  
  \[
  A : \text{event that the main chute doesn’t open} \quad \mathbb{P}(A) = 0.02
  
  B : \text{event that the backup doesn’t open} \quad \mathbb{P}(B) = 0.1
  \]

- What is the chance that at least one opens assuming independence?
Independence as an assumption

- People often assume it **without justification**.
- Example: A sky diver has two chutes

  \[ A : \text{event that the main chute doesn’t open} \quad \mathbb{P}(A) = 0.02 \]
  \[ B : \text{event that the backup doesn’t open} \quad \mathbb{P}(B) = 0.1 \]

- What is the chance that at least one opens assuming independence?

- Assuming independence doesn’t justify the assumption! Both chutes could fail because of the same rare event e.g., freezing rain.
• Review: Conditional Probability, Bayes
• Law of Total Probability (w/ Bayes)
• Chain Rule
• Independence
• Conditional Independence
• Assumptions and Correlation
• **Bonus: Monty Hall Problem**
Monty Hall Problem

Suppose you’re on a game show, and you’re given the choice of three doors. Behind one of the doors is a car, behind the other, goats. You pick a door, say number 1, and the host, who knows what’s behind the doors, opens another door, say number 3, which has a goat. He says to you, “Do you want to switch to door number 2?” Is it to your advantage to switch your choice of doors?

Assumptions

- The player is equally likely to pick each of the three doors.
- After the player picks a door, the host must open a different door with a goat behind it and offer the player the choice of staying with the original door or switching.
- If the host has a choice of which door to open, then he is equally likely to select each of them.
Should you switch or stay?