Lecture 6: Chain Rule and Independence

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

- Conditional Probability
- Bayes Theorem
- Law of Total probability

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

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

\[ P(F) = \sum_{i=1}^{n} P(F|E_i)P(E_i) \quad E_i \text{ partition } \Omega \]

\[ = \sum_{i=1}^{n} \Pr(E_i \cap F) \]
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 $F$ an 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 \( T \)).

\[
\begin{align*}
\Pr(Z) &= 0.005 \\
\Pr(T|Z) &= 0.98 \\
\Pr(T|\overline{Z}) &= 0.01
\end{align*}
\]

https://pollev.com/annakarlin185

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 $T$).

$$
Pr(Z|T) = \frac{Pr(T|Z)Pr(Z)}{Pr(T)} = \frac{0.98 \times 0.005}{0.01485}
$$

$$
Pr(T) = Pr(T|Z)Pr(Z) + Pr(T|Z^c)Pr(Z^c) = 0.98 \times 0.005 + 0.01 \times 0.995 = 0.01485
$$

$$
\approx 0.33
$$
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% 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
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: \( P(Z) \)
I have a 0.5% chance of having Zika

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")
- 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 $Z$)?

\[
Pr(\bar{T} | Z) = 1 - Pr(T | Z) = 0.02
\]
Conditional Probability Define a Probability Space

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

**Example.** $\Pr(B^c | A) = 1 - \Pr(B | A)$

$$
\Pr(B | A) + \Pr(B^c | A) = \frac{\Pr(B \cap A)}{\Pr(A)} + \frac{\Pr(B^c \cap A)}{\Pr(A)}
$$

$$
\frac{\Pr(B \cap A)}{\Pr(A)} + \frac{\Pr(B^c \cap A)}{\Pr(A)} = \frac{\Pr(A)}{\Pr(A)} = 1
$$
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 + $\mathbb{P}(\Omega) > 0$

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

• Chain Rule
• Independence
• Sequential Process
Chain Rule

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

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

\[ P(A_1 \cap A_2 \cap A_3) = P_r(A_3|A_1, A_2)P_r(A_1 \cap A_2) \]

\[ = P_r(A_3|A_1, A_2)P_r(A_2|A_1)P_r(A_1) \]
Chain Rule

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

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

**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 **in order**. (uniform probability space).

What is $P(\text{A} \cap \text{B} \cap \text{C}) = P(A \cap B \cap C)$?

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

\[
\Pr(\text{AS first}) \times \Pr(\text{10 Club and AS first}) \times \Pr(\text{4D | AS first and 10 Club})
\]

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

\[
= \frac{\text{# possible AS}}{\text{# total}} = \frac{51!}{52!}
\]
Chain Rule Example

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

What is \( P(\text{A} \cap \text{B} \cap \text{C}) = P(\text{A} \cap \text{B} \cap \text{C})? \)

\[
P(\text{A}) \cdot P(\text{B}|\text{A}) \cdot P(\text{C}|\text{A} \cap \text{B})
\]

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

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

**Definition.** Two events $A$ and $B$ are (statistically) **independent** if

$$P(A \cap B) = P(A) \cdot P(B).$$

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 = \{at most one T\} = \{HHH, HHT, HTH, THH\}
• B = \{at most 2 Heads\} = \{HHH\}

Independent?

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

\[ \mathbb{P}(A) = \frac{1}{2}, \quad \mathbb{P}(B) = \frac{3}{8} \]

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

\[ \times \]

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

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

\[
P(A \cap D) = ?
\]

\[
= P(A \cap B \cup A \cap C \cap D)
\]

\[
= P(A \cap B \cap D) + P(A \cap C \cap D) - P(A \cap B \cap D \cap C \cap D)
\]

\[
= p \cdot q \cdot r \cdot s + r \cdot s - (p \cdot q \cdot r \cdot s)
\]
Example – Network Communication

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

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

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

\[ P(AB \cap BD \cap AC \cap CD) = P(AB) \cdot P(BD) \cdot P(AC) \cdot P(CD) = pqrst \]
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.

\[
\begin{align*}
\mathbb{P}(HHH) &= \Pr(H) \Pr(H) \Pr(H) = \left(\frac{2}{3}\right)^3 \\
\mathbb{P}(TTT) &= \Pr(T) \Pr(T) \Pr(T) = \left(\frac{1}{3}\right)^3 \\
\mathbb{P}(HTT) &= \Pr(H) \Pr(T) \Pr(T) = \frac{2}{3} \left(\frac{1}{3}\right)^2 \\
\mathbb{P}(THT) &= \frac{2}{3} \left(\frac{1}{3}\right)^2
\end{align*}
\]
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.

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

\[ = P(\text{HHT, HTH, THT}) \]

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

https://pollev.com/ annakarlin185