Stat 501

Conditional Probability and Bayes' Theorem

Dr. Maher Boudabra

August 31, 2025

Lecture Overview

Lecture 1: The Foundation of Conditional Probability

Lecture 1: Learning Objectives

By the end of this lecture, you will be able to:

- Understand the concept of a conditional probability.
- Motivate why conditional probability is necessary.
- State and apply the formal definition of conditional probability: $\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$.
- Distinguish between $\mathbb{P}(A \mid B)$ and $\mathbb{P}(B \mid A)$.
- Solve basic problems involving conditional probability.

A Motivating Example

Scenario: You roll a fair six-sided die. You are told the outcome is an **even number**. What is the probability that it is a 6?

A Motivating Example

Scenario: You roll a fair six-sided die. You are told the outcome is an **even number**. What is the probability that it is a 6?

- Original sample space: $\Omega = \{1, 2, 3, 4, 5, 6\}$

A Motivating Example

Scenario: You roll a fair six-sided die. You are told the outcome is an **even number**. What is the probability that it is a 6?

- Original sample space: $\Omega = \{1, 2, 3, 4, 5, 6\}$
- New information restricts the sample space: $B = \{\text{even}\} = \{2, 4, 6\}$

A Motivating Example

Scenario: You roll a fair six-sided die. You are told the outcome is an **even number**. What is the probability that it is a 6?

- Original sample space: $\Omega = \{1, 2, 3, 4, 5, 6\}$
- New information restricts the sample space: $B = \{\text{even}\} = \{2, 4, 6\}$
- The event of interest: $A = \{6\}$

A Motivating Example

Scenario: You roll a fair six-sided die. You are told the outcome is an **even number**. What is the probability that it is a 6?

- Original sample space: $\Omega = \{1, 2, 3, 4, 5, 6\}$
- New information restricts the sample space: $B = \{\text{even}\} = \{2, 4, 6\}$
- The event of interest: $A = \{6\}$
- Given we are in B, the chance of A is 1 out of 3.

A Motivating Example

Scenario: You roll a fair six-sided die. You are told the outcome is an **even number**. What is the probability that it is a 6?

- Original sample space: $\Omega = \{1, 2, 3, 4, 5, 6\}$
- New information restricts the sample space: $B = \{\text{even}\} = \{2, 4, 6\}$
- The event of interest: $A = \{6\}$
- Given we are in B, the chance of A is 1 out of 3.

We say: "The probability of a 6, *given* an even number, is $\frac{1}{3}$."

Definition of Conditional Probability

The conditional probability of event A given that event B has occurred (with $\mathbb{P}(B) > 0$) is:

Formula

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

- $\mathbb{P}(A \cap B)$ is the probability **both** A and B happen.
- $\mathbb{P}(B)$ is the probability of the **conditioning** event.
- This formula *rescales* the probability of the joint event $(A \cap B)$ to the new “world” where we know B is certain.

Back to the Die Example

Let's verify our intuitive answer with the formula.

$$
\begin{aligned}
A &= \{6\} \\
B &= \{2, 4, 6\} \\
A \cap B &= \{6\}
\end{aligned}
$$

Back to the Die Example

Let's verify our intuitive answer with the formula.

$$\begin{aligned} A &= \{6\} \\ B &= \{2, 4, 6\} \\ A \cap B &= \{6\} \end{aligned}$$

Applying the formula:

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(\{6\})}{\mathbb{P}(\{2, 4, 6\})} = \frac{1/6}{3/6} = \frac{1}{3}$$

Back to the Die Example

Let's verify our intuitive answer with the formula.

$$\begin{aligned} A &= \{6\} \\ B &= \{2, 4, 6\} \\ A \cap B &= \{6\} \end{aligned}$$

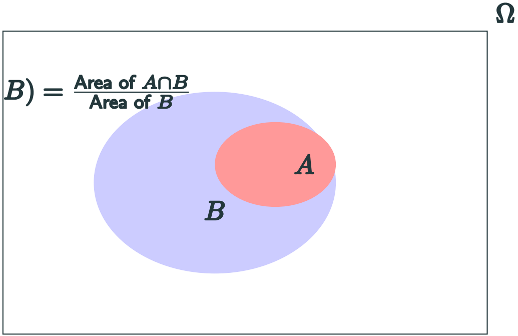
Applying the formula:

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(\{6\})}{\mathbb{P}(\{2, 4, 6\})} = \frac{1/6}{3/6} = \frac{1}{3}$$

It matches our intuition!

Visualizing Conditional Probability

$$\mathbb{P}(A \mid B) = \frac{\text{Area of } A \cap B}{\text{Area of } B}$$



Conditional probability focuses our attention on the part of event A that lies within the known event B.

Another Example: Cards

Scenario: Draw one card from a standard 52-card deck.

- A: The event that the card is a King. $\mathbb{P}(A) = 4/52 = 1/13$.
- B: The event that the card is a Face card (J, Q, K). $\mathbb{P}(B) = 12/52 = 3/13$.

Another Example: Cards

Scenario: Draw one card from a standard 52-card deck.

- A: The event that the card is a King. $\mathbb{P}(A) = 4/52 = 1/13$.
- B: The event that the card is a Face card (J, Q, K). $\mathbb{P}(B) = 12/52 = 3/13$.

Question: What is the probability it is a King, given that it is a Face card?

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(\text{King})}{\mathbb{P}(\text{Face card})} = \frac{4/52}{12/52} = \frac{4}{12} = \frac{1}{3}$$

Another Example: Cards

Scenario: Draw one card from a standard 52-card deck.

- A: The event that the card is a King. $\mathbb{P}(A) = 4/52 = 1/13$.
- B: The event that the card is a Face card (J, Q, K). $\mathbb{P}(B) = 12/52 = 3/13$.

Question: What is the probability it is a King, given that it is a Face card?

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(\text{King})}{\mathbb{P}(\text{Face card})} = \frac{4/52}{12/52} = \frac{4}{12} = \frac{1}{3}$$

This makes sense: Among the 12 face cards, 4 are Kings.

The Multiplication Rule

The definition of conditional probability can be rearranged to give a rule for finding the probability of the intersection of two events.

Multiplication Rule

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

Or equivalently:

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

This is an extremely useful tool for solving complex probability problems.

Example: Multiplication Rule

Scenario: An urn contains 5 blue and 3 red marbles. We draw two marbles **without replacement**.
Question: What is the probability that both marbles are blue?

Example: Multiplication Rule

Scenario: An urn contains 5 blue and 3 red marbles. We draw two marbles **without replacement**.
Question: What is the probability that both marbles are blue?

- Let A = first marble is blue. $\mathbb{P}(A) = 5/8$.
- Let B = second marble is blue. We want $\mathbb{P}(A \cap B)$.

Example: Multiplication Rule

Scenario: An urn contains 5 blue and 3 red marbles. We draw two marbles **without replacement**.
Question: What is the probability that both marbles are blue?

- Let $A =$ first marble is blue. $\mathbb{P}(A) = 5/8$.
- Let $B =$ second marble is blue. We want $\mathbb{P}(A \cap B)$.
- $\mathbb{P}(B \mid A) = ?$ After removing one blue marble, there are 4 blue and 3 red left. So $\mathbb{P}(B \mid A) = 4/7$.

Example: Multiplication Rule

Scenario: An urn contains 5 blue and 3 red marbles. We draw two marbles **without replacement**.
Question: What is the probability that both marbles are blue?

- Let A = first marble is blue. $\mathbb{P}(A) = 5/8$.
- Let B = second marble is blue. We want $\mathbb{P}(A \cap B)$.
- $\mathbb{P}(B \mid A) =$? After removing one blue marble, there are 4 blue and 3 red left. So $\mathbb{P}(B \mid A) = 4/7$.

Applying the multiplication rule:

$$\mathbb{P}(\text{Both Blue}) = \mathbb{P}(A \cap B) = \mathbb{P}(B \mid A) \cdot \mathbb{P}(A) = \frac{4}{7} \cdot \frac{5}{8} = \frac{20}{56} = \frac{5}{14}$$

A Common Pitfall: Order Matters!

Conditional probability is not symmetric.

$$\mathbb{P}(A \mid B) \neq \mathbb{P}(B \mid A)$$

A Common Pitfall: Order Matters!

Conditional probability is not symmetric.

$$\mathbb{P}(A \mid B) \neq \mathbb{P}(B \mid A)$$

Example: Medical Testing

- A: Person has a disease.
- B: Test for the disease is positive.

A Common Pitfall: Order Matters!

Conditional probability is not symmetric.

$$\mathbb{P}(A \mid B) \neq \mathbb{P}(B \mid A)$$

Example: Medical Testing

- A: Person has a disease.
- B: Test for the disease is positive.
- $\mathbb{P}(B \mid A)$: **Sensitivity** of the test (probability test is positive if you have the disease). This is usually high.

A Common Pitfall: Order Matters!

Conditional probability is not symmetric.

$$\mathbb{P}(A \mid B) \neq \mathbb{P}(B \mid A)$$

Example: Medical Testing

- A: Person has a disease.
- B: Test for the disease is positive.
- $\mathbb{P}(B \mid A)$: **Sensitivity** of the test (probability test is positive if you have the disease). This is usually high.
- $\mathbb{P}(A \mid B)$: What we *really* want to know (probability you have the disease if your test is positive). This can be surprisingly low if the disease is rare.

A Common Pitfall: Order Matters!

Conditional probability is not symmetric.

$$\mathbb{P}(A \mid B) \neq \mathbb{P}(B \mid A)$$

Example: Medical Testing

- A: Person has a disease.
- B: Test for the disease is positive.
- $\mathbb{P}(B \mid A)$: **Sensitivity** of the test (probability test is positive if you have the disease). This is usually high.
- $\mathbb{P}(A \mid B)$: What we *really* want to know (probability you have the disease if your test is positive). This can be surprisingly low if the disease is rare.

Confusing these two is a very common and serious error!

Independent vs. Dependent Events

Independent Events: The occurrence of B does *not* affect the probability of A.

$$\mathbb{P}(A \mid B) = \mathbb{P}(A)$$

Independent vs. Dependent Events

Independent Events: The occurrence of B does *not* affect the probability of A.

$$\mathbb{P}(A \mid B) = \mathbb{P}(A)$$

Dependent Events: The occurrence of B *does* affect the probability of A.

$$\mathbb{P}(A \mid B) \neq \mathbb{P}(A)$$

Independent vs. Dependent Events

Independent Events: The occurrence of B does *not* affect the probability of A.

$$\mathbb{P}(A \mid B) = \mathbb{P}(A)$$

Dependent Events: The occurrence of B *does* affect the probability of A.

$$\mathbb{P}(A \mid B) \neq \mathbb{P}(A)$$

Check: From the definition, if A and B are independent:

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A)\mathbb{P}(B)}{\mathbb{P}(B)} = \mathbb{P}(A)$$

Example: Independence

Scenario 1 (Dependent): Draw a card from a deck.

- A: The card is a king. ($\mathbb{P}(A) = 4/52$)
- B: The card is a heart. ($\mathbb{P}(B) = 13/52$)
- $A \cap B$: The card is the king of hearts. ($\mathbb{P}(A \cap B) = 1/52$)
- $\mathbb{P}(A \mid B) = \frac{1/52}{13/52} = 1/13$ $\neq \mathbb{P}(A) = 4/52$

Not independent. Knowing it's a heart changes the chance it's a king.

Example: Independence

Scenario 1 (Dependent): Draw a card from a deck.

- A: The card is a king. ($\mathbb{P}(A) = 4/52$)
- B: The card is a heart. ($\mathbb{P}(B) = 13/52$)
- $A \cap B$: The card is the king of hearts. ($\mathbb{P}(A \cap B) = 1/52$)
- $\mathbb{P}(A \mid B) = \frac{1/52}{13/52} = 1/13$ $\neq \mathbb{P}(A) = 4/52$

Not independent. Knowing it's a heart changes the chance it's a king. Scenario 2

(Independent): Roll a fair die twice.

- A: First roll is a 6. ($\mathbb{P}(A) = 1/6$)
- B: Second roll is a 6. ($\mathbb{P}(B) = 1/6$)
- $\mathbb{P}(A \mid B) = \mathbb{P}(A) = 1/6$. The second roll doesn't affect the first.

Conditional Probability with Tables

Scenario: Survey of 100 students on gender and preferred subject.

	Math	Art	Total
Male	20	25	45
Female	30	25	55
Total	50	50	100

Conditional Probability with Tables

Scenario: Survey of 100 students on gender and preferred subject.

	Math	Art	Total
Male	20	25	45
Female	30	25	55
Total	50	50	100

- **What is $\mathbb{P}(\textbf{Math})$?** $50/100 = 0.5$

Conditional Probability with Tables

Scenario: Survey of 100 students on gender and preferred subject.

	Math	Art	Total
Male	20	25	45
Female	30	25	55
Total	50	50	100

- **What is $\mathbb{P}$(Math)?** $50/100 = 0.5$
- **What is $\mathbb{P}$(Math | Male)?** Focus only on the Male row.

$$\frac{\mathbb{P}(\text{Math} \cap \text{Male})}{\mathbb{P}(\text{Male})} = \frac{20/100}{45/100} = \frac{20}{45} \approx 0.444$$

Conditional Probability with Tables

Scenario: Survey of 100 students on gender and preferred subject.

	Math	Art	Total
Male	20	25	45
Female	30	25	55
Total	50	50	100

- **What is $\mathbb{P}($Math$)$?** $50/100 = 0.5$
- **What is $\mathbb{P}($Math $\mid$ Male$)$?** Focus only on the Male row.

$$\frac{\mathbb{P}(\text{Math} \cap \text{Male})}{\mathbb{P}(\text{Male})} = \frac{20/100}{45/100} = \frac{20}{45} \approx 0.444$$

- **What is $\mathbb{P}($Female $\mid$ Art$)$?** Focus only on the Art column.

$$\frac{\mathbb{P}(\text{Female} \cap \text{Art})}{\mathbb{P}(\text{Art})} = \frac{25/100}{50/100} = \frac{25}{50} = 0.5$$

Key Takeaways: Lecture 1

- Conditional probability $\mathbb{P}(A \mid B)$ quantifies the probability of A **given the information** that B has occurred.
- It is defined as $\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$.
- This is fundamentally different from $\mathbb{P}(B \mid A)$.
- The Multiplication Rule: $\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B) \cdot \mathbb{P}(B)$.
- If $\mathbb{P}(A \mid B) = \mathbb{P}(A)$, then events A and B are **independent**.

Practice Problems: Lecture 1

1. From the student survey table, find $\mathbb{P}(\text{Art} \mid \text{Female})$.
2. You toss two fair coins. The first coin shows Heads. What is the probability that both are Heads?
3. A bag has 4 white and 2 black balls. Two draws are made without replacement. What is the probability the second ball is white, given the first ball was white?

Lecture 2: Bayes' Theorem and the Law of Total Probability

Lecture 2: Learning Objectives

By the end of this lecture, you will be able to:

- Understand and apply the **Law of Total Probability**.
- State and derive **Bayes' Theorem**.
- Use probability trees to visualize complex conditional scenarios.
- Apply Bayes' Theorem to simple problems to "invert" conditional probabilities.

The Need for a New Tool

Recall our medical testing example:

- We can often estimate $\mathbb{P}(\text{Test } + \mid \text{Disease})$ easily (test sensitivity).
- But a patient cares about $\mathbb{P}(\text{Disease} \mid \text{Test } +)$.

The Need for a New Tool

Recall our medical testing example:

- We can often estimate $\mathbb{P}(\text{Test} + \mid \text{Disease})$ easily (test sensitivity).
- But a patient cares about $\mathbb{P}(\text{Disease} \mid \text{Test} +)$.

How do we reverse the condition?
Bayes' Theorem provides the answer.

Prerequisite: Law of Total Probability

If events $B_1, B_2, \ldots, B_k$ form a **partition** of the sample space Ω (i.e., they are mutually exclusive and their union is Ω), then for any event A:

Law of Total Probability

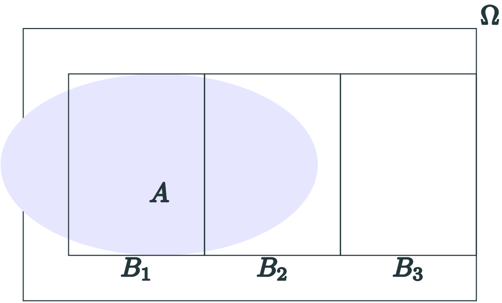
$$\mathbb{P}(A) = \mathbb{P}(A \mid B_1)\mathbb{P}(B_1) + \mathbb{P}(A \mid B_2)\mathbb{P}(B_2) + \cdots + \mathbb{P}(A \mid B_k)\mathbb{P}(B_k)$$

Prerequisite: Law of Total Probability

If events $B_1, B_2, \ldots, B_k$ form a **partition** of the sample space Ω (i.e., they are mutually exclusive and their union is Ω), then for any event A:

Law of Total Probability

$$\mathbb{P}(A) = \mathbb{P}(A \mid B_1)\mathbb{P}(B_1) + \mathbb{P}(A \mid B_2)\mathbb{P}(B_2) + \cdots + \mathbb{P}(A \mid B_k)\mathbb{P}(B_k)$$



$$\mathbb{P}(A) = \mathbb{P}(A \cap B_1) + \mathbb{P}(A \cap B_2) + \mathbb{P}(A \cap B_3) = \mathbb{P}(A|B_1)\mathbb{P}(B_1) + \mathbb{P}(A|B_2)\mathbb{P}(B_2) + \mathbb{P}(A|B_3)\mathbb{P}(B_3)$$

Example: Law of Total Probability

Scenario: Two factories supply light bulbs. Factory X supplies 60% of the bulbs, Factory Y supplies 40%. The defect rate is 2% for X and 4% for Y.
Question: What is the probability a randomly purchased bulb from this company is defective?

Example: Law of Total Probability

Scenario: Two factories supply light bulbs. Factory X supplies 60% of the bulbs, Factory Y supplies 40%. The defect rate is 2% for X and 4% for Y.
Question: What is the probability a randomly purchased bulb from this company is defective?

- Let $D =$ bulb is defective.
- $B_X =$ bulb from Factory X. $\mathbb{P}(B_X) = 0.6$, $\mathbb{P}(D \mid B_X) = 0.02$
- $B_Y =$ bulb from Factory Y. $\mathbb{P}(B_Y) = 0.4$, $\mathbb{P}(D \mid B_Y) = 0.04$

Example: Law of Total Probability

Scenario: Two factories supply light bulbs. Factory X supplies 60% of the bulbs, Factory Y supplies 40%. The defect rate is 2% for X and 4% for Y.
Question: What is the probability a randomly purchased bulb from this company is defective?

- Let D = bulb is defective.
- B_X = bulb from Factory X. $\mathbb{P}(B_X) = 0.6$, $\mathbb{P}(D \mid B_X) = 0.02$
- B_Y = bulb from Factory Y. $\mathbb{P}(B_Y) = 0.4$, $\mathbb{P}(D \mid B_Y) = 0.04$

Applying the Law of Total Probability:

$$\begin{aligned}\mathbb{P}(D) &= \mathbb{P}(D \mid B_X)\mathbb{P}(B_X) + \mathbb{P}(D \mid B_Y)\mathbb{P}(B_Y) \\ &= (0.02)(0.6) + (0.04)(0.4) \\ &= 0.012 + 0.016 = 0.028\end{aligned}$$

Deriving Bayes' Theorem

We start with the definition of conditional probability, twice:

1. $\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$
2. $\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$

Deriving Bayes' Theorem

We start with the definition of conditional probability, twice:

1. $\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$
2. $\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$

From (2), we can write: $\mathbb{P}(A \cap B) = \mathbb{P}(B \mid A) \cdot \mathbb{P}(A)$

Deriving Bayes' Theorem

We start with the definition of conditional probability, twice:

1. $\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$
2. $\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$

From (2), we can write: $\mathbb{P}(A \cap B) = \mathbb{P}(B \mid A) \cdot \mathbb{P}(A)$ Substitute this into (1):

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A) \cdot \mathbb{P}(A)}{\mathbb{P}(B)}$$

Deriving Bayes' Theorem

We start with the definition of conditional probability, twice:

1. $\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$
2. $\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$

From (2), we can write: $\mathbb{P}(A \cap B) = \mathbb{P}(B \mid A) \cdot \mathbb{P}(A)$ Substitute this into (1):

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A) \cdot \mathbb{P}(A)}{\mathbb{P}(B)}$$

This is the simplest form of **Bayes' Theorem**! It allows us to "invert" the conditional probability.

The Full Bayes' Theorem

Often, we use the Law of Total Probability to expand the denominator $\mathbb{P}(B)$.

Bayes' Theorem

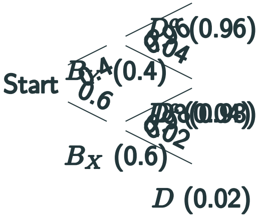
If $A_1, A_2, \ldots, A_k$ form a partition of the sample space, then for any event B:

$$\mathbb{P}(A_i \mid B) = \frac{\mathbb{P}(B \mid A_i) \cdot \mathbb{P}(A_i)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B \mid A_i) \cdot \mathbb{P}(A_i)}{\sum_{j=1}^{k} \mathbb{P}(B \mid A_j)\mathbb{P}(A_j)}$$

Numerator: Our prior belief about A_i updated by the likelihood of B given A_i.
Denominator: The total probability of observing B (a normalizing constant).

Visualizing with a Probability Tree

Bulb Factory Example: Let's find $\mathbb{P}$(Bulb is from X | it is Defective).



Follow the paths:

- $\mathbb{P}(D \cap B_X) = 0.6 \times 0.02 = 0.012$
- $\mathbb{P}(D \cap B_Y) = 0.4 \times 0.04 = 0.016$
- $\mathbb{P}(D) = 0.012 + 0.016 = 0.028$ (Law of Total Prob.)

Applying Bayes' Theorem to the Bulb Problem

We want $\mathbb{P}(B_X \mid D)$.

Applying Bayes' Theorem to the Bulb Problem

We want $\mathbb{P}(B_X \mid D)$.

$$\mathbb{P}(B_X \mid D) = \frac{\mathbb{P}(D \mid B_X) \cdot \mathbb{P}(B_X)}{\mathbb{P}(D)} = \frac{(0.02)(0.6)}{0.028} = \frac{0.012}{0.028} \approx 0.4286$$

Applying Bayes' Theorem to the Bulb Problem

We want $\mathbb{P}(B_X \mid D)$.

$$\mathbb{P}(B_X \mid D) = \frac{\mathbb{P}(D \mid B_X) \cdot \mathbb{P}(B_X)}{\mathbb{P}(D)} = \frac{(0.02)(0.6)}{0.028} = \frac{0.012}{0.028} \approx 0.4286$$

Interpretation: Even though Factory X has a lower defect rate, it supplies more bulbs.

Therefore, given a defective bulb, there's still a 43% chance it came from the *better* factory X. This often feels counterintuitive.

The Famous Medical Test Example

Scenario:

- Prevalence of a disease: $\mathbb{P}(D) = 0.01$ (1% of population has it).
- Test Sensitivity: $\mathbb{P}(T+ \mid D) = 0.99$ (99% accurate if you have it).
- Test Specificity: $\mathbb{P}(T- \mid D^c) = 0.95$ (95% accurate if you don't have it).

Question: What is $\mathbb{P}(D \mid T+)$? The probability you actually have the disease given a positive test result.

Solving the Medical Test Problem

Define the events:

- D: Has disease. $\mathbb{P}(D) = 0.01$
- D^c: Does not have disease. $\mathbb{P}(D^c) = 0.99$
- $T+$: Test positive.
- $T-$: Test negative.

We know:

- $\mathbb{P}(T+ \mid D) = 0.99$ (Sensitivity)
- $\mathbb{P}(T- \mid D^c) = 0.95$, so $\mathbb{P}(T+ \mid D^c) = 1 - 0.95 = 0.05$ (False Positive Rate)

Solving the Medical Test Problem

Define the events:

- D: Has disease.
- D^c: Does not have disease.
- $T+$: Test positive.
- $T-$: Test negative.

$\mathbb{P}(D) = 0.01$
$\mathbb{P}(D^c) = 0.99$

We know:

- $\mathbb{P}(T+ \mid D) = 0.99$ (Sensitivity)
- $\mathbb{P}(T- \mid D^c) = 0.95$, so $\mathbb{P}(T+ \mid D^c) = 1 - 0.95 = 0.05$ (False Positive Rate)

We want $\mathbb{P}(D \mid T+)$. Apply Bayes' Theorem:

$$
\begin{aligned}
\mathbb{P}(D \mid T+) &= \frac{\mathbb{P}(T+ \mid D) \cdot \mathbb{P}(D)}{\mathbb{P}(T+)} \\
&= \frac{\mathbb{P}(T+ \mid D) \cdot \mathbb{P}(D)}{\mathbb{P}(T+ \mid D)\mathbb{P}(D) + \mathbb{P}(T+ \mid D^c)\mathbb{P}(D^c)} \\
&= \frac{(0.99)(0.01)}{(0.99)(0.01) + (0.05)(0.99)} \\
&= \frac{0.0099}{0.0099 + 0.0495} = \frac{0.0099}{0.0594} \approx 0.1667
\end{aligned}
$$

Solving the Medical Test Problem

Define the events:

- D: Has disease.
- D^c: Does not have disease.
- $T+$: Test positive.
- $T-$: Test negative.

$\mathbb{P}(D) = 0.01$
$\mathbb{P}(D^c) = 0.99$

We know:

- $\mathbb{P}(T+ \mid D) = 0.99$ (Sensitivity)
- $\mathbb{P}(T- \mid D^c) = 0.95$, so $\mathbb{P}(T+ \mid D^c) = 1 - 0.95 = 0.05$ (False Positive Rate)

We want $\mathbb{P}(D \mid T+)$. Apply Bayes' Theorem:

$$\mathbb{P}(D \mid T+) = \frac{\mathbb{P}(T+ \mid D) \cdot \mathbb{P}(D)}{\mathbb{P}(T+)}$$

$$= \frac{(0.99)(0.01)}{(0.99)(0.01) + (0.05)(0.99)}$$

$$= \frac{0.0099}{0.0099 + 0.0495} = \frac{0.0099}{0.0594} \approx 0.1667$$

Solving the Medical Test Problem

Define the events:

- D: Has disease. $\mathbb{P}(D) = 0.01$
- D^c: Does not have disease. $\mathbb{P}(D^c) = 0.99$
- $T+$: Test positive.
- $T-$: Test negative.

We know:

- $\mathbb{P}(T+ \mid D) = 0.99$ (Sensitivity)
- $\mathbb{P}(T- \mid D^c) = 0.95$, so $\mathbb{P}(T+ \mid D^c) = 1 - 0.95 = 0.05$ (False Positive Rate)

We want $\mathbb{P}(D \mid T+)$. Apply Bayes' Theorem:

$$\mathbb{P}(D \mid T+) = \frac{\mathbb{P}(T+ \mid D) \cdot \mathbb{P}(D)}{\mathbb{P}(T+)}$$

$$= \frac{0.0099}{0.0099 + 0.0495} = \frac{0.0099}{0.0594} \approx 0.1667$$

Solving the Medical Test Problem

Define the events:

- D: Has disease.
- D^c: Does not have disease.
- $T+$: Test positive.
- $T-$: Test negative.

$\mathbb{P}(D) = 0.01$
$\mathbb{P}(D^c) = 0.99$

We know:

- $\mathbb{P}(T+ \mid D) = 0.99$ (Sensitivity)
- $\mathbb{P}(T- \mid D^c) = 0.95$, so $\mathbb{P}(T+ \mid D^c) = 1 - 0.95 = 0.05$ (False Positive Rate)

We want $\mathbb{P}(D \mid T+)$. Apply Bayes' Theorem:

$$\begin{aligned}
\mathbb{P}(D \mid T+) &= \frac{\mathbb{P}(T+ \mid D) \cdot \mathbb{P}(D)}{\mathbb{P}(T+)} \\
&= \frac{\mathbb{P}(T+ \mid D) \cdot \mathbb{P}(D)}{\mathbb{P}(T+ \mid D)\mathbb{P}(D) + \mathbb{P}(T+ \mid D^c)\mathbb{P}(D^c)} \\
&= \frac{(0.99)(0.01)}{(0.99)(0.01) + (0.05)(0.99)} \\
&= \frac{0.0099}{0.0099 + 0.0495} = \frac{0.0099}{0.0594} \approx 0.1667
\end{aligned}$$

Solving the Medical Test Problem

Define the events:

- D: Has disease.
- D^c: Does not have disease.
- $T+$: Test positive.
- $T-$: Test negative.

$\mathbb{P}(D) = 0.01$
$\mathbb{P}(D^c) = 0.99$

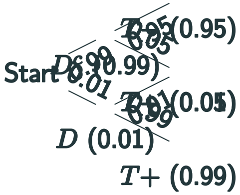
We know:

- $\mathbb{P}(T+ \mid D) = 0.99$ (Sensitivity)
- $\mathbb{P}(T- \mid D^c) = 0.95$, so $\mathbb{P}(T+ \mid D^c) = 1 - 0.95 = 0.05$ (False Positive Rate)

We want $\mathbb{P}(D \mid T+)$. Apply Bayes' Theorem:

$$\begin{aligned}
\mathbb{P}(D \mid T+) &= \frac{\mathbb{P}(T+ \mid D) \cdot \mathbb{P}(D)}{\mathbb{P}(T+)} \\
&= \frac{\mathbb{P}(T+ \mid D) \cdot \mathbb{P}(D)}{\mathbb{P}(T+ \mid D)\mathbb{P}(D) + \mathbb{P}(T+ \mid D^c)\mathbb{P}(D^c)} \\
&= \frac{(0.99)(0.01)}{(0.99)(0.01) + (0.05)(0.99)} \\
&= \frac{0.0099}{0.0099 + 0.0495} = \frac{0.0099}{0.0594} \approx 0.1667
\end{aligned}$$

Tree Diagram for Medical Test



- $\mathbb{P}(T + \cap D) = 0.01 \times 0.99 = 0.0099$
- $\mathbb{P}(T + \cap D^c) = 0.99 \times 0.05 = 0.0495$
- $\mathbb{P}(T+) = 0.0099 + 0.0495 = 0.0594$
- $\mathbb{P}(D \mid T+) = \frac{0.0099}{0.0594} \approx 0.167$

Interpreting the Result

$$\mathbb{P}(D \mid T+) \approx 16.7\%$$

- This seems shockingly low for a “99% accurate” test.
- The intuition comes from the **base rate** (prior probability) of the disease.
- The number of false positives (healthy people testing positive) is large because the disease is rare.
- The test is much better at **ruling in** the disease if the prevalence is higher.

This is the core lesson of Bayes’ Theorem: **Prior beliefs matter!**

Key Takeaways: Lecture 2

- The **Law of Total Probability** breaks down $\mathbb{P}(A)$ based on a partition of the sample space.
- **Bayes' Theorem** allows us to invert conditional probabilities:

$$\mathbb{P}(\text{Hypothesis} \mid \text{Evidence}) = \frac{\mathbb{P}(\text{Evidence} \mid \text{Hypothesis}) \cdot \mathbb{P}(\text{Hypothesis})}{\mathbb{P}(\text{Evidence})}$$

- **Probability trees** are an excellent tool for visualizing these problems.
- The **prior probability** $\mathbb{P}(\text{Hypothesis})$ has a huge impact on the result. Ignoring it leads to serious errors.

Practice Problems: Lecture 2

1. Using the medical test data, find $\mathbb{P}(D^c \mid T-)$ (the probability you are truly healthy given a negative test). This is called the **Negative Predictive Value**.
2. There are three warehouses. Warehouse 1 has 100 items, 10 defective. Warehouse 2 has 50 items, 5 defective. Warehouse 3 has 200 items, 10 defective. An item is chosen at random from the combined stock and is found to be defective. What is the probability it came from Warehouse 3?

Lecture 3: Advanced Applications and Interpretations

Lecture 3: Learning Objectives

By the end of this lecture, you will be able to:

- Apply Bayes' Theorem to problems with more than two partitions.
- Understand the Bayesian interpretation of probability as a **degree of belief**.
- Perform **Bayesian updating** with multiple pieces of evidence.
- Recognize applications of Bayes' Theorem in machine learning (Naive Bayes classifier), statistics, and everyday reasoning.

Recap: The Bayesian Framework

Bayes' Theorem provides a formal mechanism for updating beliefs in light of new evidence.

Terminology

- **Prior Probability** $\mathbb{P}(H)$: Initial degree of belief in hypothesis H *before* seeing evidence E.
- **Likelihood** $\mathbb{P}(E \mid H)$: How likely the evidence E is, assuming H is true.
- **Marginal Likelihood** $\mathbb{P}(E)$: The total probability of the evidence (across all hypotheses).
- **Posterior Probability** $\mathbb{P}(H \mid E)$: Revised degree of belief in H *after* seeing evidence E.

$$\underbrace{\mathbb{P}(H \mid E)}_{\text{Posterior}} = \frac{\overbrace{\mathbb{P}(E \mid H)}^{\text{Likelihood}} \cdot \overbrace{\mathbb{P}(H)}^{\text{Prior}}}{\underbrace{\mathbb{P}(E)}_{\text{Marginal Likelihood}}}$$

Example 1: The Three Prisoners Problem

Scenario: Three prisoners (A, B, C) are on death row. The governor randomly picks one to pardon. The warden knows who is pardoned but cannot say.

- Prisoner A asks the warden: "Which one of B or C will be executed?" (He knows at least one will be.)
- The warden says: "Prisoner B will be executed."

Question: What is the probability that Prisoner A is pardoned, *given* this new information?

Example 1: The Three Prisoners Problem

Scenario: Three prisoners (A, B, C) are on death row. The governor randomly picks one to pardon. The warden knows who is pardoned but cannot say.

- Prisoner A asks the warden: "Which one of B or C will be executed?" (He knows at least one will be.)
- The warden says: "Prisoner B will be executed."

Question: What is the probability that Prisoner A is pardoned, *given* this new information?
Intuition: It seems like it should be $1/2$ (between A and C). Let's use Bayes' Theorem.

Solving the Three Prisoners Problem

Let P_A, P_B, P_C be the events that A, B, or C is pardoned. $\mathbb{P}(P_A) = \mathbb{P}(P_B) = \mathbb{P}(P_C) = 1/3$.
Let W_B be the event that the warden says "B will be executed".
We need to model the warden's behavior. If A is pardoned, the warden can freely say B or C. Assume he picks randomly. If B is pardoned, he must say C is executed. If C is pardoned, he must say B is executed.

$$\mathbb{P}(W_B \mid P_A) = 1/2 \quad \text{(Random choice)}$$
$$\mathbb{P}(W_B \mid P_B) = 0 \quad \text{(Must say C)}$$
$$\mathbb{P}(W_B \mid P_C) = 1 \quad \text{(Must say B)}$$

Solving the Three Prisoners Problem

Let P_A, P_B, P_C be the events that A, B, or C is pardoned. $\mathbb{P}(P_A) = \mathbb{P}(P_B) = \mathbb{P}(P_C) = 1/3$.
Let W_B be the event that the warden says "B will be executed".
We need to model the warden's behavior. If A is pardoned, the warden can freely say B or C. Assume he picks randomly. If B is pardoned, he must say C is executed. If C is pardoned, he must say B is executed.

$$\mathbb{P}(W_B \mid P_A) = 1/2 \quad \text{(Random choice)}$$
$$\mathbb{P}(W_B \mid P_B) = 0 \quad \text{(Must say C)}$$
$$\mathbb{P}(W_B \mid P_C) = 1 \quad \text{(Must say B)}$$

We want $\mathbb{P}(P_A \mid W_B)$.

$$\begin{aligned}
\mathbb{P}(P_A \mid W_B) &= \frac{\mathbb{P}(W_B \mid P_A)\mathbb{P}(P_A)}{\mathbb{P}(W_B)} \\
&= \frac{\mathbb{P}(W_B \mid P_A)\mathbb{P}(P_A)}{\mathbb{P}(W_B|P_A)\mathbb{P}(P_A) + \mathbb{P}(W_B|P_B)\mathbb{P}(P_B) + \mathbb{P}(W_B|P_C)\mathbb{P}(P_C)} \\
&= \frac{(1/2)(1/3)}{(1/2)(1/3) + (0)(1/3) + (1)(1/3)} = \frac{1/6}{1/6 + 0 + 1/3} = \frac{1/6}{1/6 + 2/6} = \frac{1/6}{3/6} = \frac{1}{3}
\end{aligned}$$

Prisoner A's chance of being pardoned is still $1/3$! The probability for C is now $2/3$.

Solving the Three Prisoners Problem

Let P_A, P_B, P_C be the events that A, B, or C is pardoned. $\mathbb{P}(P_A) = \mathbb{P}(P_B) = \mathbb{P}(P_C) = 1/3$.
Let W_B be the event that the warden says "B will be executed".
We need to model the warden's behavior. If A is pardoned, the warden can freely say B or C. Assume he picks randomly. If B is pardoned, he must say C is executed. If C is pardoned, he must say B is executed.

$$\mathbb{P}(W_B \mid P_A) = 1/2 \quad \text{(Random choice)}$$
$$\mathbb{P}(W_B \mid P_B) = 0 \quad \text{(Must say C)}$$
$$\mathbb{P}(W_B \mid P_C) = 1 \quad \text{(Must say B)}$$

We want $\mathbb{P}(P_A \mid W_B)$.

$$\mathbb{P}(P_A \mid W_B) = \frac{\mathbb{P}(W_B \mid P_A)\mathbb{P}(P_A)}{\mathbb{P}(W_B)}$$

$$= \frac{(1/2)(1/3)}{(1/2)(1/3) + (0)(1/3) + (1)(1/3)} = \frac{1/6}{1/6 + 0 + 1/3} = \frac{1/6}{1/6 + 2/6} = \frac{1/6}{3/6} = \frac{1}{3}$$

Prisoner A's chance of being pardoned is still $1/3$! The probability for C is now $2/3$.

Solving the Three Prisoners Problem

Let P_A, P_B, P_C be the events that A, B, or C is pardoned. $\mathbb{P}(P_A) = \mathbb{P}(P_B) = \mathbb{P}(P_C) = 1/3$.
Let W_B be the event that the warden says "B will be executed".
We need to model the warden's behavior. If A is pardoned, the warden can freely say B or C. Assume he picks randomly. If B is pardoned, he must say C is executed. If C is pardoned, he must say B is executed.

$$\mathbb{P}(W_B \mid P_A) = 1/2 \quad \text{(Random choice)}$$
$$\mathbb{P}(W_B \mid P_B) = 0 \quad \text{(Must say C)}$$
$$\mathbb{P}(W_B \mid P_C) = 1 \quad \text{(Must say B)}$$

We want $\mathbb{P}(P_A \mid W_B)$.

$$\begin{aligned}
\mathbb{P}(P_A \mid W_B) &= \frac{\mathbb{P}(W_B \mid P_A)\mathbb{P}(P_A)}{\mathbb{P}(W_B)} \\
&= \frac{\mathbb{P}(W_B \mid P_A)\mathbb{P}(P_A)}{\mathbb{P}(W_B|P_A)\mathbb{P}(P_A) + \mathbb{P}(W_B|P_B)\mathbb{P}(P_B) + \mathbb{P}(W_B|P_C)\mathbb{P}(P_C)} \\
&= \frac{(1/2)(1/3)}{(1/2)(1/3) + (0)(1/3) + (1)(1/3)} = \frac{1/6}{1/6 + 0 + 1/3} = \frac{1/6}{1/6 + 2/6} = \frac{1/6}{3/6} = \frac{1}{3}
\end{aligned}$$

Solving the Three Prisoners Problem

Let P_A, P_B, P_C be the events that A, B, or C is pardoned. $\mathbb{P}(P_A) = \mathbb{P}(P_B) = \mathbb{P}(P_C) = 1/3$.
Let W_B be the event that the warden says "B will be executed".
We need to model the warden's behavior. If A is pardoned, the warden can freely say B or C. Assume he picks randomly. If B is pardoned, he must say C is executed. If C is pardoned, he must say B is executed.

$$\mathbb{P}(W_B \mid P_A) = 1/2 \quad \text{(Random choice)}$$
$$\mathbb{P}(W_B \mid P_B) = 0 \quad \text{(Must say C)}$$
$$\mathbb{P}(W_B \mid P_C) = 1 \quad \text{(Must say B)}$$

We want $\mathbb{P}(P_A \mid W_B)$.

$$\begin{aligned}\mathbb{P}(P_A \mid W_B) &= \frac{\mathbb{P}(W_B \mid P_A)\mathbb{P}(P_A)}{\mathbb{P}(W_B)} \\ &= \frac{\mathbb{P}(W_B \mid P_A)\mathbb{P}(P_A)}{\mathbb{P}(W_B|P_A)\mathbb{P}(P_A) + \mathbb{P}(W_B|P_B)\mathbb{P}(P_B) + \mathbb{P}(W_B|P_C)\mathbb{P}(P_C)} \\ &= \frac{(1/2)(1/3)}{(1/2)(1/3) + (0)(1/3) + (1)(1/3)} = \frac{1/6}{1/6 + 0 + 1/3} = \frac{1/6}{1/6 + 2/6} = \frac{1/6}{3/6} = \frac{1}{3}\end{aligned}$$

Prisoner A's chance of being pardoned is still $1/3$! The probability for C is now $2/3$.

Bayesian Updating with Multiple Evidence

What if we get more than one piece of evidence?
We can apply Bayes' Theorem iteratively.

Sequential Updating

1. Start with your **prior** belief: $\mathbb{P}(H)$
2. See evidence E_1. Update your belief to the **posterior**: $\mathbb{P}(H \mid E_1)$

Bayesian Updating with Multiple Evidence

What if we get more than one piece of evidence?
We can apply Bayes' Theorem iteratively.

Sequential Updating

1. Start with your **prior** belief: $\mathbb{P}(H)$
2. See evidence E_1. Update your belief to the **posterior**: $\mathbb{P}(H \mid E_1)$
3. This posterior becomes the **new prior** for the next piece of evidence.
4. See evidence E_2. Update again: $\mathbb{P}(H \mid E_1, E_2) \propto \mathbb{P}(E_2 \mid H, E_1) \cdot \mathbb{P}(H \mid E_1)$

Bayesian Updating with Multiple Evidence

What if we get more than one piece of evidence?
We can apply Bayes' Theorem iteratively.

Sequential Updating

1. Start with your **prior** belief: $\mathbb{P}(H)$
2. See evidence E_1. Update your belief to the **posterior**: $\mathbb{P}(H \mid E_1)$
3. This posterior becomes the **new prior** for the next piece of evidence.
4. See evidence E_2. Update again: $\mathbb{P}(H \mid E_1, E_2) \propto \mathbb{P}(E_2 \mid H, E_1) \cdot \mathbb{P}(H \mid E_1)$

This process is computationally efficient and mirrors how we learn in real life.

Example: Spam Filtering (Naive Bayes Concept)

Goal: Classify an email as Spam (S) or Not Spam (N).

Evidence: The presence of words like “free”, “money”, “viagra”.

Example: Spam Filtering (Naive Bayes Concept)

Goal: Classify an email as Spam (S) or Not Spam (N).

Evidence: The presence of words like "free", "money", "viagra".

- **Prior:** $\mathbb{P}(S)$ (base rate of spam, e.g., 0.8).

Example: Spam Filtering (Naive Bayes Concept)

Goal: Classify an email as Spam (S) or Not Spam (N).
Evidence: The presence of words like "free", "money", "viagra".

- **Prior:** $\mathbb{P}(S)$ (base rate of spam, e.g., 0.8).
- For a new email containing the word "free", we compute:

$$\mathbb{P}(S \mid \text{"free"}) = \frac{\mathbb{P}(\text{"free"} \mid S) \cdot \mathbb{P}(S)}{\mathbb{P}(\text{"free"})}$$

Example: Spam Filtering (Naive Bayes Concept)

Goal: Classify an email as Spam (S) or Not Spam (N).
Evidence: The presence of words like "free", "money", "viagra".

- **Prior:** $\mathbb{P}(S)$ (base rate of spam, e.g., 0.8).
- For a new email containing the word "free", we compute:

$$\mathbb{P}(S \mid \text{“free”}) = \frac{\mathbb{P}(\text{“free”} \mid S) \cdot \mathbb{P}(S)}{\mathbb{P}(\text{“free”})}$$

- The likelihood $\mathbb{P}(\text{“free”} \mid S)$ is estimated from training data (what fraction of spam emails contain "free"?).

Example: Spam Filtering (Naive Bayes Concept)

Goal: Classify an email as Spam (S) or Not Spam (N).
Evidence: The presence of words like "free", "money", "viagra".

- **Prior:** $\mathbb{P}(S)$ (base rate of spam, e.g., 0.8).
- For a new email containing the word "free", we compute:

$$\mathbb{P}(S \mid \text{"free"}) = \frac{\mathbb{P}(\text{"free"} \mid S) \cdot \mathbb{P}(S)}{\mathbb{P}(\text{"free"})}$$

- The likelihood $\mathbb{P}(\text{"free"} \mid S)$ is estimated from training data (what fraction of spam emails contain "free"?).
- Then the email contains "money". We update again, using $\mathbb{P}(S \mid \text{"free"})$ as our new prior:

$$\mathbb{P}(S \mid \text{"free"}, \text{"money"}) \propto \mathbb{P}(\text{"money"} \mid S, \text{"free"}) \cdot \mathbb{P}(S \mid \text{"free"})$$

Example: Spam Filtering (Naive Bayes Concept)

Goal: Classify an email as Spam (S) or Not Spam (N).
Evidence: The presence of words like "free", "money", "viagra".

- **Prior:** $\mathbb{P}(S)$ (base rate of spam, e.g., 0.8).
- For a new email containing the word "free", we compute:

$$\mathbb{P}(S \mid \text{“free”}) = \frac{\mathbb{P}(\text{“free”} \mid S) \cdot \mathbb{P}(S)}{\mathbb{P}(\text{“free”})}$$

- The likelihood $\mathbb{P}(\text{“free”} \mid S)$ is estimated from training data (what fraction of spam emails contain "free"?).
- Then the email contains "money". We update again, using $\mathbb{P}(S \mid \text{“free”})$ as our new prior:

$$\mathbb{P}(S \mid \text{“free”}, \text{“money”}) \propto \mathbb{P}(\text{“money”} \mid S, \text{“free”}) \cdot \mathbb{P}(S \mid \text{“free”})$$

- A "Naive" assumption is that words are independent given the class, simplifying the likelihood: $\mathbb{P}(\text{“money”} \mid S, \text{“free”}) \approx \mathbb{P}(\text{“money”} \mid S)$.

Bayesian vs. Frequentist Interpretation

- **Frequentist:** Probability is the long-run **frequency** of an event. (e.g., The probability of heads is 0.5 means if we flip the coin infinitely many times, 50% will be heads.)

Bayesian vs. Frequentist Interpretation

- **Frequentist:** Probability is the long-run **frequency** of an event. (e.g., The probability of heads is 0.5 means if we flip the coin infinitely many times, 50% will be heads.)
- **Bayesian:** Probability is a subjective **degree of belief**. (e.g., I am 50% confident the next flip will be heads. This belief can be updated with data.)

Bayesian vs. Frequentist Interpretation

- **Frequentist:** Probability is the long-run **frequency** of an event. (e.g., The probability of heads is 0.5 means if we flip the coin infinitely many times, 50% will be heads.)
- **Bayesian:** Probability is a subjective **degree of belief**. (e.g., I am 50% confident the next flip will be heads. This belief can be updated with data.)

Example: The probability it will rain tomorrow.

- A frequentist might have trouble with this (tomorrow only happens once).
- A Bayesian has no problem: they assign a prior belief based on weather models and historical data, and update it as new forecast models come in.

Application: Bayesian Inference in Statistics

Traditional (Frequentist)

- Population parameter (e.g., mean μ) is a fixed, unknown constant.
- Confidence intervals are interpreted as: "If I repeat this experiment many times, 95% of the calculated intervals will contain the true μ."

Application: Bayesian Inference in Statistics

Traditional (Frequentist)

- Population parameter (e.g., mean μ) is a fixed, unknown constant.
- Confidence intervals are interpreted as: "If I repeat this experiment many times, 95% of the calculated intervals will contain the true μ."

Bayesian

- Population parameter μ is itself a random variable with a **prior distribution** (e.g., μ could be anywhere between 0 and 100).
- We collect data and use Bayes' Theorem to find the **posterior distribution** of μ.
- We can then say: "Given the data, there is a 95% probability μ lies between X and Y." (This is called a **Credible Interval**).

Bayesian statistics is powerful but requires choosing a prior, which can be subjective.

Cognitive Biases and Bayes' Theorem

Humans are often bad at intuitive conditional reasoning. Bayes' Theorem helps explain why.

- **Base Rate Neglect:** Ignoring the prior probability $\mathbb{P}(H)$. (e.g., The medical test example feels surprising because we neglect how rare the disease is).

Cognitive Biases and Bayes' Theorem

Humans are often bad at intuitive conditional reasoning. Bayes' Theorem helps explain why.

- **Base Rate Neglect:** Ignoring the prior probability $\mathbb{P}(H)$. (e.g., The medical test example feels surprising because we neglect how rare the disease is).
- **Prosecutor's Fallacy:** Confusing $\mathbb{P}(\text{Evidence} \mid \text{Innocent})$ with $\mathbb{P}(\text{Innocent} \mid \text{Evidence})$. A prosecutor might argue: "The probability of this DNA match if the defendant were innocent is 1 in a million. Therefore, the probability he is innocent is 1 in a million."

Cognitive Biases and Bayes' Theorem

Humans are often bad at intuitive conditional reasoning. Bayes' Theorem helps explain why.

- **Base Rate Neglect:** Ignoring the prior probability $\mathbb{P}(H)$. (e.g., The medical test example feels surprising because we neglect how rare the disease is).
- **Prosecutor's Fallacy:** Confusing $\mathbb{P}(\text{Evidence} \mid \text{Innocent})$ with $\mathbb{P}(\text{Innocent} \mid \text{Evidence})$. A prosecutor might argue: "The probability of this DNA match if the defendant were innocent is 1 in a million. Therefore, the probability he is innocent is 1 in a million." This is a fatal error! It ignores the prior probability of guilt and the possibility of lab error.

Key Takeaways: Lecture 3

- Bayes' Theorem is a formal rule for **updating beliefs** with new evidence.
- The **prior** has a significant impact on the **posterior**.
- We can update beliefs **sequentially** with multiple pieces of evidence.
- It has wide applications: from spam filters and machine learning to medical diagnosis and judicial reasoning.
- It explains common **cognitive biases** like base rate neglect.
- It represents a different, **subjective interpretation** of probability compared to the frequentist view.

Final Thought

Bayes' Theorem is more than just a formula;
it is a framework for rational thinking
in an uncertain world.