# Lecture 4 (June 29)

Wednesday, June 24, 2026 12:34 PM

chapter2_conditional_...

## Chapter 2: Conditional Probability

MATH/STAT 394: Probability I


Arman Jahangiri

University of Washington Department of Mathematics


Summer 2026



## Learning Goals

By the end of this lecture, students should be able to:

- interpret conditional probability intuitively;
- apply Bayes' rule and the Law of Total Probability;
- distinguish between $P(A|B)$ and $P(B|A)$;
- understand independence and conditional independence;
- use conditioning as a strategy in solving in probability problems.



## Conditional Probability



## Why Do We Need Conditional Probability?

In many situations, probabilities change once additional information becomes available. $\mathbb{P}(A)$

**Before observing new information**

Let

$$R = \{\text{it rains today}\}.$$

Without any additional information, we assign the probability

prior "belief"

$$\mathbb{P}(R).$$

This represents our initial uncertainty about rain.

**After observing new information**

Suppose we now observe dark clouds:

$$C = \{\text{dark clouds are present}\}.$$

The probability of rain may now change to

posterior "belief" $\mathbb{P}(R \mid C),$ given

read as "probability of $R$ given $C$.".

An important observation here is that we do not necessarily have $\mathbb{P}(R \mid C) = \mathbb{P}(R)$, because the information that $C$ occurred changes what outcomes now appear more plausible.

**Main Idea**

Conditional probability quantifies how probabilities are updated after new information or evidence is observed.



## Definition of Conditional Probability

**Definition**

If

$$\mathbb{P}(B) > 0,$$

then the conditional probability of $A$ given $B$ is

$$\boxed{\mathbb{P}(A|B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}.} \leq 1$$

$A \cap B \subseteq B$

$P(A\cap B) \leq P(B)$

$\frac{P(A\cap B)}{P(B)} \leq 1$

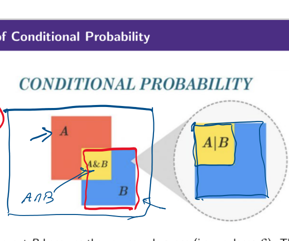
Interpretation:

- $B$ is the observed evidence;
- we restrict attention to outcomes in $B$;
- probabilities are renormalized within $B$.



## Visual Interpretation of Conditional Probability

$B = \{1,3,5\}$, $A = \{1\}$, dice toss: $P(A) = 1/6$, $P(A\cap B) = 1/6$, $P(B) = 3/6 = 1/2$




yellow / blue $= P(A|B)$

When conditioning on $B$, the event $B$ becomes the new sample space (i.e., replaces $\mathcal{S}$). The conditional probability $\mathbb{P}(A \mid B)$ measures the proportion of $B$ that also belongs to $A$.

$P(A|B) = \frac{P(A\cap B)}{P(B)} = \frac{1/6}{1/2} = \frac{1}{3}$

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} \quad \text{provided } \mathbb{P}(B) > 0.$$

$P(A) = \frac{|A|}{|S|} = \frac{|A\cap S|}{|S|}$

$P(A|B) = \frac{|A\cap B|}{|B|}$ ← new sample space

Image adapted from: ExcelR — Conditional Probability.



## Useful Equations from Conditional Probability

$P(A|B) = \frac{P(A\cap B)}{P(B)} \Rightarrow \boxed{P(A\cap B) = P(B) \times P(A|B)}$ ①

$P(B|A) = \frac{P(B\cap A)}{P(A)} \Rightarrow \boxed{P(A\cap B) = P(A) \times P(B|A)}$

$$\mathbb{P}(A) = \mathbb{P}(A \mid B)\mathbb{P}(B) + \mathbb{P}(A \mid B^c)\mathbb{P}(B^c).$$

$P(A) = P(A\cap S) = P(A\cap [B\cup B^c]) = P(A\cap B \cup A\cap B^c) = P(A\cap B) + P(A\cap B^c) = P(A|B)P(B) + P(A|B^c)P(B^c)$

$$\mathbb{P}(B) = \mathbb{P}(B \mid A)\mathbb{P}(A) + \mathbb{P}(B \mid A^c)\mathbb{P}(A^c).$$

$A_1, \dots, A_n$ is a partition if 1) $\bigcup_{i=1}^n A_i = S$ 2) $A_i \cap A_j = \emptyset$ for all $i \neq j$

**Law of Total Probability**

Each equation partitions the sample space into two disjoint cases and adds the corresponding conditional probabilities.

$A_1, \dots, A_n$: $\sum_{i=1}^n P(A_i) = 1$

example ②: event $A$: $A, A^c$ ⇒ $A \cup A^c = S$ ⇒ $A \cap A^c = \emptyset$

example ①: $\emptyset, S$: $\emptyset \cup S = S$, $\emptyset \cap S = \emptyset$

examples of partitions



## Example: Two Cards

Two cards are drawn without replacement.

Define

$$A = \{\text{first card is a heart}\}, \qquad B = \{\text{second card is red}\}.$$

Using the multiplication rule,

$$\mathbb{P}(A\cap B) = \mathbb{P}(B \mid A)\mathbb{P}(A) = \frac{25}{51}\cdot\frac{13}{52} = \frac{25}{204}.$$

Also,

$$\mathbb{P}(A) = \frac{1}{4},$$

and

$$\mathbb{P}(B) = \mathbb{P}(B \mid A)\mathbb{P}(A) + \mathbb{P}(B \mid A^c)\mathbb{P}(A^c) = \frac{25}{51}\cdot\frac{1}{4} + \frac{17}{34}\cdot\frac{3}{4} = \frac{1}{2}, \; = 0.5$$

$52 - 13 = 39$

Therefore,

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(A)} = \frac{25/204}{1/4} = \frac{25}{51} \approx 0.490, \; < P(B)$$

and

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} = \frac{25/204}{1/2} = \frac{25}{102} \approx 0.245. \; < P(A)$$



## Important Warning

**Conditional probabilities are directional**

In general,

$$\boxed{\mathbb{P}(A|B) \neq \mathbb{P}(B|A).}$$

$\frac{P(A\cap B)}{P(B)}$, $\frac{P(A\cap B)}{P(A)}$

The conditioning event matters.

$$\mathbb{P}(A|B)$$

means:

probability of $A$ after learning $B$.



## Bayes' Rule and LOTP



## Intersection Formula

From the definition:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}.$$ Def

Multiplying both sides by $\mathbb{P}(B)$ gives:

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

Similarly,

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

The last two equalities together imply an important equality in Probability, known as the Bayes' rule, given in the next slide.



## Bayes' Rule: Updating Beliefs with Evidence

**Bayes' Rule**

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

$P(A|B) = \frac{P(A\cap B)}{P(B)} = \frac{P(B|A)\cdot P(A)}{P(B)}$

$P(B|A)P(A) + P(B|A^c)P(A^c)$

Bayes' rule describes how probabilities "evolve" when new information is observed.

$$\underbrace{\mathbb{P}(A)}_{\text{before observing } B} \Longrightarrow \underbrace{\mathbb{P}(A \mid B)}_{\text{after observing } B} = \mathbb{P}(A) \times \underbrace{\frac{\mathbb{P}(B \mid A)}{\mathbb{P}(B)}}_{\text{updating weight}}$$

$\overset{?}{=} 1$ ($P(B|A) = P(B)$)

The observation of $B$ changes our belief about whether $A$ occurred.

**Interpretation**

Bayes' rule updates a prior belief $\mathbb{P}(A)$ into a posterior belief $\mathbb{P}(A \mid B)$ by multiplying by a factor that measures how strongly the evidence $B$ supports $A$.

important



## Law of Total Probability

$(n=2)$ $\{A \,\&\, A^c\}$ is a partition for $S$ (reason) $\Rightarrow P(B) = P(B|A)P(A) + P(B|A^c)P(A^c)$

Suppose

$$A_1, \dots, A_n$$

form a partition of $\mathcal{S}$.

This is a specific example of LOTP (when $n=2$ & $A_1 = A$, $A_2 = A^c$)

Then:

$$\boxed{\mathbb{P}(B) = \sum_{i=1}^n \mathbb{P}(B|A_i)\mathbb{P}(A_i).}$$ LOTP

**Interpretation**

Break the sample space into cases, compute probabilities within each case, then combine them.



## Rare Disease Example

Suppose we are analyzing the disease rate in a certain area.

Let

$$D = \{\text{person has the disease}\}, \qquad T = \{\text{test result is positive}\},$$

and assume

$$\mathbb{P}(D) = 0.01, \quad \mathbb{P}(T \mid D) = 0.95, \quad \mathbb{P}(T^c \mid D^c) = 0.95.$$

$\to P(T|D^c) = 1 - P(T^c|D^c)$

Interpretation:

- The disease prevalence is 1%;
- The test correctly detects the disease 95% of the time;
- The test correctly returns a negative result 95% of the time for healthy individuals.

Question: A patient tests positive. What is $\mathbb{P}(D \mid T)$?

Using Bayes' Rule and LOTP:

Using Bayes' Rule and the LOTP,

$$\mathbb{P}(D \mid T) = \frac{\mathbb{P}(T \mid D)\mathbb{P}(D)}{\mathbb{P}(T)} = \frac{\mathbb{P}(T \mid D)\mathbb{P}(D)}{\mathbb{P}(T \mid D)\mathbb{P}(D) + \mathbb{P}(T \mid D^c)\mathbb{P}(D^c)} = \frac{0.95(0.01)}{0.95(0.01) + 0.05(0.99)} \approx 0.16.$$



## Why This Feels Surprising

Most people expect the answer to be close to 95%.

But the disease is very rare.

Among many healthy people:

$$5\%$$

still test positive.

Those false positives greatly outnumber the true positives.

**Lesson**

Posterior probabilities depend both on:

- the evidence,
- and the prior probability.



## Conditional Probability as Probability



## Conditional Probability is a Probability Measure

Fix an event $E$ with

$$\mathbb{P}(E) > 0.$$

Then the conditional probability function

$$\mathbb{P}(\cdot \mid E)$$

satisfies the three probability axioms.

1) $P(A|B) \geq 0$ 2) $= \frac{P(A\cap B) \geq 0}{P(B) \geq 0}$ 3) additivity

**Axiom 1 (Nonnegativity)**

$$\mathbb{P}(A \mid E) \geq 0 \quad \text{for every event } A.$$

**Axiom 2 (Normalization)**

$$\mathbb{P}(\mathcal{S} \mid E) = 1.$$

$\to$ 2) $P(S|E) = \frac{P(S\cap E)}{P(E)} = \frac{P(E)}{P(E)} = 1$ (sample space)

**Axiom 3 (Countable Additivity)** If $A_1, A_2, \dots$ are pairwise disjoint events, then

$$\mathbb{P}\left(\bigcup_{i=1}^\infty A_i \,\middle|\, E\right) = \sum_{i=1}^\infty \mathbb{P}(A_i \mid E).$$

$\frac{P([\bigcup A_i] \cap E)}{P(E)} = \frac{P(\bigcup (A_i \cap E))}{P(E)}$



## Independence



## Definition of Independence

**Definition (Independence)**

Events $A$ and $B$ are independent if

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

provided $\mathbb{P}(B) > 0$.

lhs $\frac{P(A\cap B)}{P(B)} = \frac{P(A)}{1}$ rhs

**Interpretation:**

Learning that $B$ occurred gives no information about $A$ (and vice versa*).

(*) We will later show that

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

Thus independence is symmetric:

$$A \text{ is independent of } B \iff B \text{ is independent of } A.$$

Think of it this way: $\mathbb{P}(A$ with new (additional) information$) = \mathbb{P}(A$ with no additional information$)$

**Lemma (Equivalent characterization of independence)**

Events $A$ and $B$ are independent if and only if

$$\boxed{\mathbb{P}(A\cap B) = \mathbb{P}(A) \times \mathbb{P}(B).}$$



## Conditional Independence

(skip)

**Conditional Independence**

Events $A$ and $B$ are conditionally independent given $E$ if

$$\boxed{\mathbb{P}(A\cap B|E) = \mathbb{P}(A|E)\mathbb{P}(B|E).}$$

Conditional independence is different from ordinary independence.

Two events may:

- be independent but not conditionally independent;
- be conditionally independent but not independent.



## Example: A Hidden Variable Creates Dependence

please (at home) example of independence

A coin is selected at random:

- Fair coin with probability 1/2;
- Biased coin ($P(H) = 0.9$) with probability 1/2.

**If we know which coin was selected:**

The tosses are independent.

**If we do not know which coin was selected:**

Observing early tosses changes our belief about which coin we have.

Therefore, earlier tosses affect our prediction of later tosses, so the tosses are *not* independent.

$$P(H_2 \mid H_1) \neq P(H_2).$$



## Conditioning as a Strategy



## Monty Hall Problem: Setup

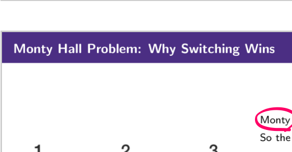


Image source: Brilliant: Monty Hall Problem.

There are three doors:

- one door hides a car;
- two doors hide goats.

You choose one door, but it is not opened yet.

Assume, without loss of generality, that you initially choose door 1.

$$C_i = \{\text{car is behind door } i\}.$$

Before Monty opens a door,

$$\mathbb{P}(C_1) = \mathbb{P}(C_2) = \mathbb{P}(C_3) = \frac{1}{3}.$$



## Monty Hall Problem: Why Switching Wins

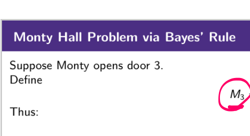



Image source: Brilliant: Monty Hall Problem.

Monty knows where the car is and always opens a door with a goat.

So the important question is:

Should you switch?

| Car location | First choice | Switching |
|---|---|---|
| $C_1$ | correct | loses |
| $C_2$ | wrong | wins |
| $C_3$ | wrong | wins |

Therefore,

$$\mathbb{P}(\text{win by switching}) = \mathbb{P}(C_2 \cup C_3) = \mathbb{P}(C_2) + \mathbb{P}(C_3) = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}.$$



## Monty Hall Problem via Bayes' Rule

Suppose Monty opens door 3.

Define

$$M_3 = \{\text{Monty opens door 3}\}.$$

Thus:

$$\underbrace{\mathbb{P}(\text{initial choice correct})}_{\text{door 1}} = \frac{1}{3}, \quad \text{and} \quad \underbrace{\mathbb{P}(\text{initial choice wrong})}_{\text{door 2 or 3}} = \frac{2}{3}.$$

After Monty reveals a goat behind door 3, the entire remaining probability mass concentrates on door 2.

Using Bayes' Rule,

$$\mathbb{P}(\text{win by switching}|M_3) = \mathbb{P}(C_2 \mid M_3) = \frac{\mathbb{P}(M_3 \mid C_2)\mathbb{P}(C_2)}{\mathbb{P}(M_3)}.$$ Bayes rule

Now, $\mathbb{P}(C_2) = \frac{1}{3}$, $\mathbb{P}(M_3 \mid C_2) = 1$, and

$$\mathbb{P}(M_3) = \frac{1}{2}\cdot\frac{1}{3} + 1\cdot\frac{1}{3} + 0\cdot\frac{1}{3} = \frac{1}{2}.$$

$P(M_3|C_1)P(C_1) + P(M_3|C_2)P(C_2) + P(M_3|C_3)P(C_3)$

Therefore,

$$\mathbb{P}(C_2 \mid M_3) = \frac{1\cdot\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}, \text{ and thus , and thus } \mathbb{P}(\text{win without switching}|M_3) = \mathbb{P}(C_1|M_3) = \frac{1}{3}$$



## Summary

- Conditional probability updates probabilities using observed evidence.
- $$\mathbb{P}(A|B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}.$$
- Bayes' Rule relates: $$\mathbb{P}(A|B) \quad \text{and} \quad \mathbb{P}(B|A).$$
- LOTP breaks problems into cases.
- Conditional probabilities themselves satisfy the probability axioms.
- Independence means: $$\mathbb{P}(A|B) = \mathbb{P}(A).$$
- Conditioning is a powerful problem-solving strategy.

**Next lecture**

Random variables and distributions.