

Chapter 2: Conditional Probability

MATH/STAT 394: Probability I

Arman Jahangiri

University of Washington
Department of Mathematics

Summer 2026

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.

Before observing new information

Let

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

Without any additional information, we assign the probability

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

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

$$\mathbb{P}(R \mid C),$$

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

If

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

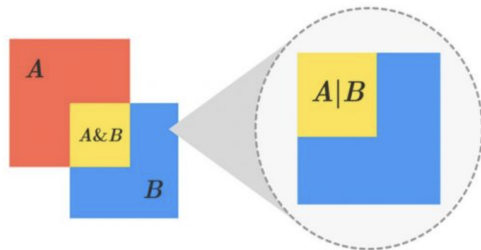
then the conditional probability of A given B is

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

Interpretation:

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

CONDITIONAL PROBABILITY



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

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

Image adapted from: ExcelR — Conditional Probability.

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

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

Law of Total Probability

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

Example: Two Cards

Two cards are drawn without replacement.

Define

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

Using the multiplication rule,

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

Also,

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

and

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

Therefore,

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

and

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

Conditional probabilities are directional

In general,

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

The conditioning event matters.

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

means:

probability of A after learning B .

Bayes' Rule and LOTP

From the definition:

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

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

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

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

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

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 .

Law of Total Probability

Suppose

$$A_1, \dots, A_n$$

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

Then:

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

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

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

and assume

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

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 $\mathbb{P}(D | T)$?

Using Bayes' Rule and the LOTP,

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

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.

Axiom 1 (Nonnegativity)

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

Axiom 2 (Normalization)

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

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

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

Independence

Definition of Independence

Definition (Independence)

Events A and B are independent if

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

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

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: $\mathbb{P}(A \text{ with new (additional) information}) = \mathbb{P}(A \text{ with no additional information})$

Lemma (Equivalent characterization of independence)

Events A and B are independent if and only if

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

Conditional Independence

Events A and B are conditionally independent given E if

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

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

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 (Example)

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

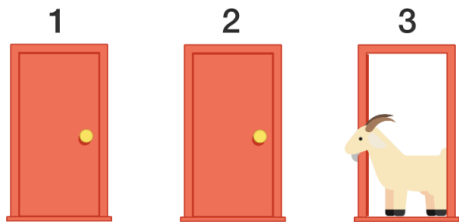


Image source: Brilliant: Monty Hall Problem.

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,

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

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

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

Therefore,

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

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.

Temporary page!

$\text{\LaTeX}$ was unable to guess the total number of pages correctly. As there was some unprocessed data have been added to the final page this extra page has been added to receive it.
If you rerun the document (without altering it) this surplus page will go away, because $\text{\LaTeX}$ now knows the number of pages to expect for this document.