

STA 35C: Statistical Data Science III

Lecture 2: Probability Review

Dogyoon Song

Fall 2026, UC Davis

Announcements

Office hours:

- Instructor: Wed 1–2 PM (MSB 1143)
- TA: Thu 12–1 PM (MSB 1117)
- * Please also use Piazza for course questions

Important dates:

- Midterm 1: Mon, Oct 19 (in class)
- Midterm 2: Mon, Nov 9 (in class)
- Final exam: Wed, Dec 9, 8:00–10:00 AM
- * *No make-up exams; see the syllabus for missed-exam policies*

SDC accommodations: If you need accommodations, please submit requests through the [Student Disability Center \(SDC\)](#) as early as possible

* See the [course webpage](#) and [syllabus](#) for more details and additional information

Announcements: Homework 1

Homework 1 is now posted

- Due: Thu, October 1, 11:59 PM PT
 - Late submissions will not be accepted for any reason and will receive 0 points
- Submission instructions:
 - Upload a **single PDF file** to Canvas (*Assignments* → *Homework 1*)
 - Name the file using the prefix of your UC Davis email ID and homework number (e.g., dgsong_hw1.pdf)
 - Please make sure to include “STA 35C,” your name, and the last four digits of your student ID on the front page
 - For coding problems, prepare your solutions in R Markdown; for non-coding problems, you may typeset your solutions in $\text{\LaTeX}$ (preferred), use a word processor, or handwrite and scan them; in all cases, make sure your work is legible and clearly organized

* See the [syllabus](#) for more details about homework policy and requirements

Agenda¹

- Probability basics
 - Set theory
 - Probabilistic models
- Conditional probability
- Independence

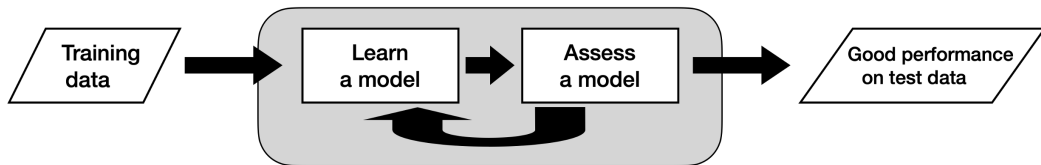
¹Most of today's topics were covered in [STA 35A](#); see Lectures 13–15

Motivation: Why probability?

Recall the goals of statistical learning:

- Predict Y from X (learn a function $f : \mathcal{X} \rightarrow \mathcal{Y}$)
- Identify patterns in data X

Standard workflow:



Key challenge:

- We aim for good predictions or insights on **new, unseen data**
- How can we reason about performance on future data, not just on the training data?

Solution: Probabilistic tools and viewpoints offer a formal way to handle uncertainty

Formalizing probability: Sample space and events

Ingredients to formalize probability:

- **Outcome:** a possible result of an experiment or trial
- **Sample space:** the set of all possible outcomes, often denoted by Ω
 - e.g., $\{H, T\}$, $\{1, 2, 3, 4, 5, 6\}$
- **Event:** a subset of Ω
 - e.g., for $\Omega = \{H, T\}$: $\emptyset$, $\{H\}$, $\{T\}$, $\{H, T\}$
 - e.g., for $\Omega = \{1, 2, 3, 4, 5, 6\}$: $\{6\}$, $\{1, 2\}$, $\{2, 4, 6\}$, ...
- **Probability²:** a map P that assigns each event A a number $P(A) \in [0, 1]$
 - We will shortly state the axioms that such assignments must satisfy

Probability theory makes extensive use of set notation and set operations!

²A formal, mathematically rigorous definition of probability measure is beyond the scope of STA 35C

Set theory: Notation and terminology

A **set** is a collection of distinct objects, called the **elements** of the set

- $x \in S$: x is an element of S
- $x \notin S$: x is not an element of S

The **empty set** is a set having no elements, denoted by $\emptyset$

Sets can be specified in various ways

- Roster notation (enumeration): $S = \{x_1, x_2, \dots, x_n\}$, or $S = \{x_1, x_2, \dots\}$

e.g., $\{H, T\}$, $\{1, 2, 3, 4, 5, 6\}$, $\{2, 4, 6, 8, \dots\}$

- Set-builder notation (logical formula): $S = \{x \mid x \text{ satisfies } Q\}$

e.g., $\{2k \mid k \text{ is a positive integer}\}$, $\{x \in \mathbb{R} \mid 0 \leq x \leq 1\}$

Sets may be finite, countably infinite, or uncountable

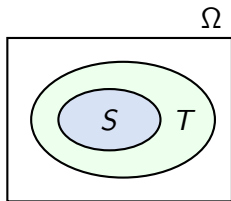
Set theory: Inclusion relations

S is a **subset** of T if every element of S is an element of T , i.e., $x \in S \implies x \in T$

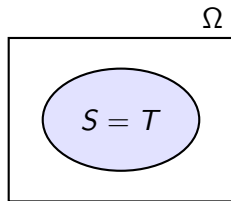
- We write $S \subseteq T$ (or $S \subset T$)
- Equivalently, T is a **superset** of S , denoted by $T \supseteq S$ (or $T \supset S$)

The two sets are **equal**, written $S = T$, if $S \subseteq T$ and $T \subseteq S$

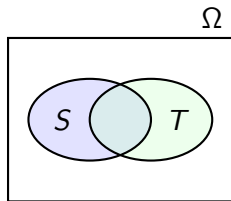
S is a **proper (strict) subset** of T , denoted by $S \subsetneq T$, if $S \subseteq T$ and $S \neq T$



$S \subsetneq T$



$S = T$



Neither $S \subseteq T$ nor $T \subseteq S$

Set theory: Set operations

Often, we fix the **universal set (universe)** Ω , which contains all objects of interest

- $S^c = \{x \in \Omega \mid x \notin S\}$ is the **complement** of S with respect to Ω

Given two sets S and T ,

- $S \cup T = \{x \mid x \in S \text{ or } x \in T\}$ is their **union**
- $S \cap T = \{x \mid x \in S \text{ and } x \in T\}$ is their **intersection**

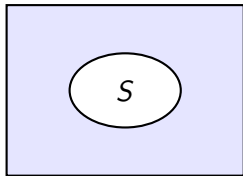
Two sets S and T are **disjoint** if $S \cap T = \emptyset$

- A collection of sets is **pairwise disjoint** if no two distinct sets have a common element

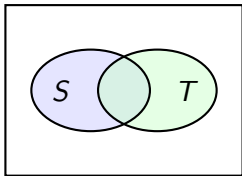
A collection of nonempty sets $\{S_1, S_2, \dots\}$ is a **partition** of S if

- $\bigcup_n S_n = S$ and
- the sets $S_1, S_2, \dots$ are pairwise disjoint

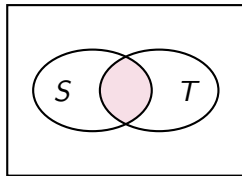
Illustration with Venn diagrams: Set operations



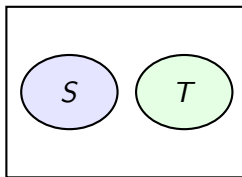
S^c



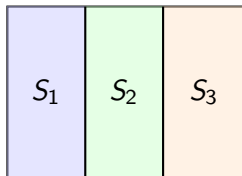
$S \cup T$



$S \cap T$



Disjoint



Partition of Ω

Set theory: The algebra of sets

Set operations have several properties, which are direct consequences of the definitions

- Commutativity:

$$S \cup T = T \cup S, \quad S \cap T = T \cap S$$

- Associativity:

$$S \cup (T \cup U) = (S \cup T) \cup U, \quad S \cap (T \cap U) = (S \cap T) \cap U$$

- Distributivity:

$$S \cap (T \cup U) = (S \cap T) \cup (S \cap U), \quad S \cup (T \cap U) = (S \cup T) \cap (S \cup U)$$

- ... and more:

$$(S^c)^c = S, \quad S \cap S^c = \emptyset, \quad S \cup S^c = \Omega, \quad S \cup \Omega = \Omega, \quad S \cap \Omega = S$$

Elements of a probabilistic model

A **probabilistic model** is a mathematical description of an uncertain situation

- **Experiment**: an underlying process producing exactly one out of several possible outcomes

A probabilistic model consists of a sample space Ω and a probability law P

- **Sample space** Ω : the set of all possible outcomes
 - Event³: a subset $A \subseteq \Omega$; it occurs when the realized outcome ω belongs to A
- **Probability law** P : a map that assigns a number $P(A)$ encoding our knowledge or belief about the collective likelihood of the event A , satisfying *certain axioms*

For a legitimate and useful probabilistic model, Ω and P must satisfy some requirements

³For uncountable sample spaces, we may need to restrict which subsets are events; such measure-theoretic details are beyond this course.

Choosing an appropriate sample space

When choosing a sample space Ω , keep three principles in mind:

- **Collectively exhaustive:** Ω should include every outcome the experiment can produce
 - Choosing $\Omega = \{1, 2, 3, 4, 5\}$ for a die roll will miss the outcome 6
- **Mutually exclusive:** the outcomes in Ω should be distinguishable from one another
 - The sample space for a die roll cannot have both "1" and "1 or 4" simultaneously as possible outcomes
- **No unnecessary detail:** make Ω fine enough to answer the questions of interest, but no finer
 - If we only care about heads or tails, recording the landing location or maximum height is unnecessary

Probability laws

Once the sample space Ω has been chosen, a probability law assigns to each event A a number $P(A)$, called the **probability** of A , subject to three axioms:

1. **Nonnegativity:** $P(A) \geq 0$ for every event A
2. **(Countable) Additivity:** if A and B are disjoint events, then

$$P(A \cup B) = P(A) + P(B)$$

More generally, if $A_1, A_2, \dots$ are pairwise disjoint, then

$$P\left(\bigcup_n A_n\right) = \sum_n P(A_n)$$

3. **Normalization:** $P(\Omega) = 1$

Some properties of probability laws

There are many natural properties of a probability law, derivable from the axioms

- $P(A^c) = 1 - P(A)$
- $P(\emptyset) = 0$
- If $A \subseteq B$, then $P(A) \leq P(B)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $P(A \cup B) \leq P(A) + P(B)$ (union bound)
- $P(B \setminus A) = P(B) - P(A \cap B)$ $B \setminus A := B \cap A^c$
- ...

It is often useful to visualize and verify these using a Venn diagram
(*You may also want to prove them!*)

A quick exercise: A die roll example

Setup:

- $\Omega = \{1, 2, 3, 4, 5, 6\}$ and $P(\{1\}) = P(\{2\}) = \dots = P(\{6\}) = 1/6$
- $A = \{2, 3, 5\}$ (prime faces)
- $B = \{2, 4, 6\}$ (even faces)

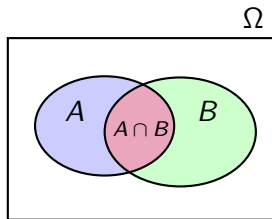
Questions:

- Draw a Venn diagram to visualize Ω , A and B
- Identify $A \cup B$, $A \cap B$ and $A \cap B^c$ on the Venn diagram
- Compute $P(A \cup B)$, $P(A \cap B)$, and $P(A \cap B^c)$

Conditional probability: Definition

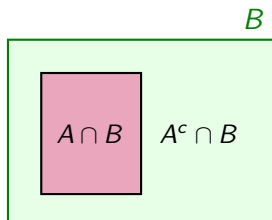
For an event⁴ B with $P(B) > 0$, the **conditional probability** of A given B is

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



Ordinary view inside Ω

$\Rightarrow$
condition on B



Restrict attention to outcomes inside B

Example: Compare $P(A)$ vs $P(A | B)$ in the die roll example on the previous slide

- $P(A) = 3/6 = 1/2$, whereas $P(A | B) = 1/3$, since within $B = \{2, 4, 6\}$ only 2 is prime

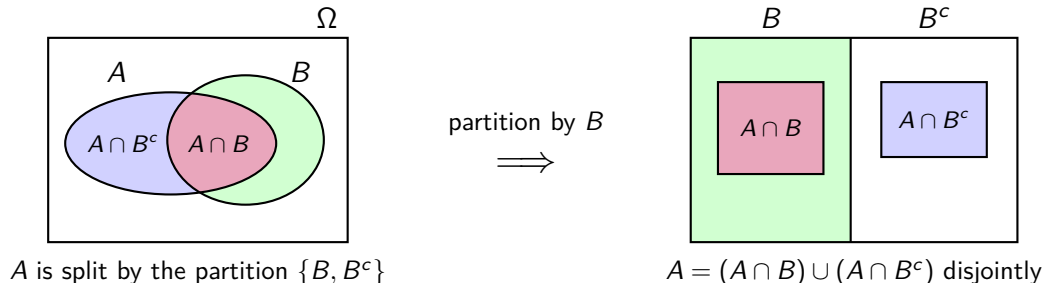
⁴In this course, we only condition on events with positive probability

Conditional probability: Key identities

Two key identities for conditional probability:

- **Multiplication rule:** $P(A \cap B) = P(B)P(A | B)$
- **Law of total probability:** If $0 < P(B) < 1$, then

$$\begin{aligned}P(A) &= P(A \cap B) + P(A \cap B^c) \\ &= P(A | B)P(B) + P(A | B^c)P(B^c)\end{aligned}$$



Independence

Events A and B are **independent** if $P(A \cap B) = P(A)P(B)$

- Recall that $P(A \cap B) = P(B)P(A | B)$
- When $P(B) > 0$, independence is equivalent to $P(A | B) = P(A)$
 - Knowing that B occurred does not change the probability that A occurs
- Similarly, when $P(A) > 0$, independence is equivalent to $P(B | A) = P(B)$

Example: A coin toss and a die roll, with all 12 outcome pairs equally likely

- Knowing the coin shows heads does not change the probability of an even die roll

Counter-example: Seeing someone wearing sunglasses and a rainy day

- If we see someone wearing sunglasses, it is less likely to be a rainy day

Wrap-up

- Events are sets, so set theory provides the basic language of probability
- A probabilistic model consists of a sample space Ω and a probability law P such that
 - $P(A) \geq 0$ for every event A
 - $P(\bigcup_n A_n) = \sum_n P(A_n)$ for pairwise disjoint $A_1, A_2, \dots$
 - $P(\Omega) = 1$
- Conditional probability $P(A \mid B)$ means we restrict attention to outcomes inside B
 - Independence: $P(A \cap B) = P(A)P(B)$