STA 130A: Mathematical Statistics: Brief Course

Lecture 2: Set Theory and Probabilistic Models

Dogyoon Song

Fall 2026, UC Davis

Announcements

Office hours:

- Instructor: Wed 2–3 PM (MSB 1143)
- TA: Thu 9–11 AM (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: Mon, Dec 7, 3:30–5:30 PM

* *No make-up exams; see the syllabus for exam and accommodation 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 7, 11:59 PM PT
  - Late submissions will not be accepted for any reason and will receive 0 points
- Submission instructions:
  - You may typeset your solutions in LaTeX (preferred), use a word processor, or handwrite and scan them; in all cases, make sure your work is legible and clearly organized
  - 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 130A," your name, and the last four digits of your student ID on the front page

* See the syllabus for more details about homework policy and instructions

Agenda

- Set theory
- Probabilistic models, probability laws, and basic properties
- Models vs. reality

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): $S = \{x_1, x_2, \ldots, x_n\}$, or $S = \{x_1, x_2, \ldots\}$

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

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

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

Sets may be finite, countably infinite, or uncountable

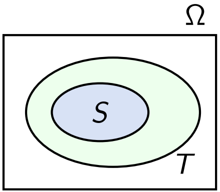
Set theory: Inclusion relations

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

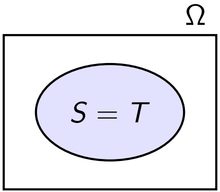
- We write $S \subseteq T$ (or $S \subset T$)
- Equivalently, T is a **superset** of S, denoted by $T \supseteq S$ (or $T \supset S$)

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

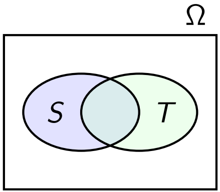
S is a **proper (strict)** subset of T, denoted by $S \subsetneq T$, if $S \subseteq T$ and $S \neq T$



$S \subsetneq T$



$S = T$



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

* *Remark:* The inclusion relation $\subseteq$ gives a partial order between sets

Set theory: Set operations

Often we can 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**

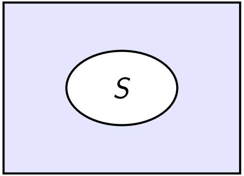
Two sets S and T are **disjoint** if $S \cap T = \emptyset$

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

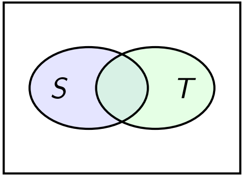
A collection of sets $\{S_1, S_2, \dots\}$ is a **partition** of S if

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

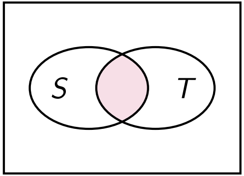
Illustration with Venn diagrams: Set operations



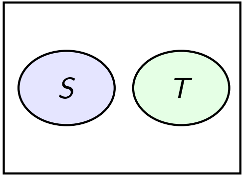
S^c



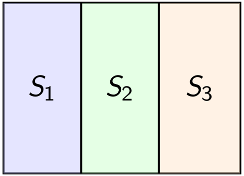
$S \cup T$



$S \cap T$



Disjoint



Partition of Ω

Set theory: The algebra of sets

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

- Commutativity:

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

- Associativity:

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

- Distributivity:

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

- ... and more:

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

Set theory: De Morgan's laws

The following properties are known as **De Morgan's laws**:

$$\left(\bigcup_n S_n\right)^c = \bigcap_n S_n^c, \qquad \left(\bigcap_n S_n\right)^c = \bigcup_n S_n^c$$

Venn diagrams help visualize these identities for two or three sets

However, formally establishing these laws requires a *rigorous mathematical proof*

- To show $(\bigcup_n S_n)^c \subseteq \bigcap_n S_n^c$: *"Want: every x in $(\bigcup_n S_n)^c$ also belongs to $\bigcap_n S_n^c$"*
  - Take $x \in (\bigcup_n S_n)^c$
  - Then $x \notin \bigcup_n S_n$, which implies that for every n, $x \notin S_n$
  - Thus, $x \in S_n^c$ for all n, and therefore, $x \in \bigcap_n S_n^c$
- To show $(\bigcup_n S_n)^c \supseteq \bigcap_n S_n^c$: proceed similarly (*Exercise!*)

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
  - Event[1]: 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

[1] For uncountable sample spaces, events belong to a suitable collection of subsets; such measure-theoretic details are beyond the scope of 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
  - The landing location or maximum height is unnecessary if we only care about heads or tails

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 $A_1, A_2, \ldots$ 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)
- $P(B \setminus A) = P(B) - P(A \cap B)$ $\qquad B \setminus A := B \cap A^c$
- ...

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

Constructing a probability law: Example 1

Example (Single coin toss)

Experiment: Tossing a coin once

Possible outcomes: Heads (H) and tails (T)

Sample space: $\Omega = \{H, T\}$

Events: $\{H, T\}, \{H\}, \{T\}, \emptyset$

If the coin is fair—meaning that we know or believe that heads and tails are "equally likely"—then we should assign equal probabilities to the two possible outcomes: $P(\{H\}) = P(\{T\})$

The additivity and the normalization axioms together imply that

$$P(\{H, T\}) = P(\{H\}) + P(\{T\}) = 1$$

Conclusion: The resulting probability law is given by

$$P\big(\{H, T\}\big) = 1, \qquad P\big(\{H\}\big) = \frac{1}{2}, \qquad P\big(\{T\}\big) = \frac{1}{2}, \qquad P\big(\emptyset\big) = 0$$

Constructing a probability law: Example 2

Example (Three coin tosses)

Experiment: Tossing a coin three times

Possible outcomes: length-3 sequences of H's and T's

Sample space: $\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$

Events: subsets of Ω ($2^8 = 256$ subsets)

We model the 8 sequences as equally likely (as for independent tosses of a fair coin), so each has probability $1/8$

Consider an event $A = \{\text{exactly 2 heads occur}\}$: the resulting probability law gives

$$\begin{aligned} P(A) &= P\big(\{HHT, HTH, THH\}\big) \\ &= P\big(\{HHT\}\big) + P\big(\{HTH\}\big) + P\big(\{THH\}\big) \\ &= \frac{1}{8} + \frac{1}{8} + \frac{1}{8} \\ &= \frac{3}{8} \end{aligned}$$

Discrete vs. continuous models

In a discrete model, the probabilities of singleton events determine the probability law:

$$P(A) = \sum_{a \in A} P(\{a\}).$$

In a continuous model, singleton probabilities typically do *not* determine the law

Example (A wheel of fortune)

Experiment: Spin a fair wheel and record a number in $[0, 1]$

Possible outcomes: the numbers in the interval $[0, 1]$

Sample space: $\Omega = [0, 1]$

Events: subsets of $[0, 1]$

Q: What is the probability $P(\{x\})$ for a single point $x \in [0, 1]$?

- Under the natural uniform model on $[0, 1]$, $P(\{x\}) = 0$ for every x
- Yet intervals can still have positive probability; $P([a, b]) = b - a$ for $0 \leq a \leq b \leq 1$

Models vs. reality

Using the framework of probability theory involves two distinct stages:

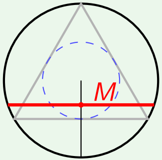
1. **Modeling:** Choose a sample space and a probability law to represent reality
   - The axioms ensure mathematical consistency; suitability depends on the situation
   - Different reasonable models may represent the same situation differently
   - Even a simplified "incorrect" model may be useful due to simplicity and tractability
2. **Analysis:** Once the model is fixed, compute probabilities and draw conclusions (*within the model*)
   - This stage is governed by the axioms of probability and ordinary logic
   - The main difficulty is often computation, not ambiguity

Some apparent paradoxes arise from an ambiguous or incomplete probabilistic model!

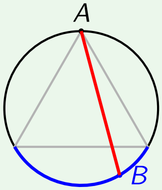
Example: Bertrand's paradox

Example (Bertrand's paradox)

Question: Given a circle and an inscribed equilateral triangle, what is the probability that the length of a randomly chosen chord of the circle is greater than the side of the triangle?



(a) Choose a random point uniformly along radius. A chord M is longer than the triangle side iff its midpoint lies within distance $R/2$ of the center. Hence $P = 1/2$.



(b) Fix one endpoint A, then choose the other endpoint uniformly on the circumference. The favorable arc has length $1/3$ of the circumference. Hence $P = 1/3$.

The phrase *"a randomly chosen chord"* does not specify a unique probabilistic model.

Takeaway: Different models of a "random chord" give different answers.
Probability theory itself is not inconsistent here.

Wrap-up

Set theory

- Set theory provides a formal language for probability theory; notation and basic operations
- Venn diagrams are useful for intuition, but mathematical proofs come from the definitions

Probabilistic models

- A probabilistic model consists of a sample space Ω and a probability law P
- Choose Ω to be exhaustive, mutually exclusive, and no finer than needed
- P must satisfy the three axioms: nonnegativity, countable additivity, and normalization
- In discrete models, singleton probabilities determine P; in continuous models, they may not

Models vs. reality

- Modeling requires judgment: the same real situation can admit different reasonable models
- Once the model is fixed, probability calculations are governed by the axioms and logic

Next lecture: Conditional probability, Bayes' rule, and an introduction to independence.

Suggested reading: [BT08, Ch. 1.1 & 1.2]

References

Dimitri Bertsekas and John N Tsitsiklis.
Introduction to probability, volume 1.
Athena Scientific, 2nd edition, 2008.