Section 2: Probability

STA 35C – Statistical Data Science III

Instructor: Akira Horiguchi

Fall Quarter 2025 (Sep 24 – Dec 12)
MWF, 12:10 PM – 1:00 PM, Olson 158
University of California, Davis

Section 1: Overview

Based on Chapter 1 of textbook: `https://www.probabilitycourse.com/`

- Section 1.1: Introduction

Section 1: Basics in probability theory

Section 1.1: Introduction

Probability is a way to quantify randomness and/or uncertainty.

- e.g., coin flips, dice rolls, stocks, weather.
- Rules of probability should be intuitive and self-consistent.
- Self-consistent: the rules shouldn't lead to contradictions.
- Thus these rules must be constructed in a certain way.
- Suppose we want to assign a probability to each event in a set of possible events.
- We would like, at the very least:
  1. each probability to be a value between 0 and 1 (inclusive)
  2. the probability assigned to the full set of events to be 1
     - close to 1 $\Rightarrow$ very likely that A occurs.
  3. the probability assigned to the empty set to be 0
     - close to 0 $\Rightarrow$ very unlikely that A occurs.
- We need more restrictions to ensure self-consistency.

The following definition will lead to intuitive and self-consistent rules of probability.

- We assign a *probability* measure $P(A)$ to an event A.

Definition 1: Probabilty measure $P(\cdot)$

For a nonempty sample space Ω, the set function $P\colon \Omega \to [0, 1]$ is a *probability measure*, if

- $P(\Omega) = 1$,
- for any pairwise disjoints events $A_1, A_2, A_3, \dots \subset \Omega$ (i.e. $A_i \cap A_j = \varnothing$ for all i, j with $i \neq j$), holds:

$$P\left(A_1 \cup A_2 \cup A_3 \cup \cdots\right) = P(A_1) + P(A_2) + P(A_3) + \cdots \tag{1}$$

This definition fulfills the three desirable properties:

- $P(\Omega) = 1$: the probability of the biggest possible set is equal to 1.
- Property (1) – called the *countable additivity* property – allows us to add probabilities of disjoint sets.

Finding probabilities

Given a random experiment with a sample space Ω, how do we find the probability of an event of interest? Use:

- the specific information that we have about the random experiment.
- the probability rules induced by Definition 1.

Finding probabilities: example

Example: Roll a fair four-sided die. What is the probability of $E = \{1, 3\}$?

- Information about experiment (fair die): $P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\})$.
- Probability rules:

$$\begin{aligned} 1 &= P(S) \\ &= P\left(\{1\} \cup \{2\} \cup \{3\} \cup \{4\}\right) \\ &= P\left(\{1\}\right) + P\left(\{2\}\right) + P\left(\{3\}\right) + P\left(\{4\}\right) \\ &= 4P\left(\{1\}\right). \end{aligned}$$

Thus $P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = \frac{1}{4}$. Finally,

$$P(E) = P(\{1, 3\}) = P\left(\{1\}\right) + P\left(\{3\}\right) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

Finding probabilities: notation

Annoying to write e.g., $P(\{2\})$

- Simplify to $P(2)$
- But always keep in mind that P is a function on sets, not on individual outcomes.

Finding probabilities: more tools

Definition 1 implies the following additional properties:

Properties of $P(\cdot)$

Given a sample space Ω and arbitrary events $A, B \subset \Omega$, Definition 1 implies

1. $P(\varnothing) = 0$
2. $P(A^c) = 1 - P(A)$
3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
4. $P(B \setminus A) = P(B) - P(A \cap B)$
5. $P(A) \leq P(B)$ if $A \subset B$.

(Pictures for intuition; for formal proofs, see “Example 1.10” in §1.3.3 of textbook)

Finding probabilities: example

Suppose we have the following information:

1. There is a 60 percent chance that it will rain today.
2. There is a 50 percent chance that it will rain tomorrow.
3. There is a 30 percent chance that it does not rain either day.

Find the following probabilities:

a. The probability that it will rain today or tomorrow.
b. The probability that it will rain today and tomorrow.
c. The probability that it will rain today but not tomorrow.
d. The probability that it either will rain today or tomorrow, but not both.

Probability models: discrete vs continuous

Distinguish between two different types of sample spaces: *discrete* and *continuous*.

- Will discuss in more detail in Section 3 of the course.
- Discrete: can compute the probability of an event by adding all outcomes in the event.
- Continuous: need to use integration instead of summation.

Probability models: discrete

If a sample space Ω is a countable set, this refers to a *discrete* probability model.

- Can list all elements: $\Omega = \{s_1, s_2, s_3, \cdots\}$.
- For an event $A \subset \Omega$, by countable additivity (1) we can write

$$P(A) = P\left(\bigcup_{s \in A} \{s\}\right) = \sum_{s \in A} P(s)$$

  Thus, to find probability of an event, just need to sum the probability of individual elements in that event.

Probability models: discrete (example)

Consider a gambling game: win $k-2$ dollars with probability $\frac{1}{2^k}$ for any $k \in \mathbb{N}$.

- What is the probability of winning at least \$1 and less than \$4?
- What is the probability of winning more than \$1?

Probability models: discrete (equally likely outcomes)

Important special case: finite sample space Ω where each outcome is equally likely.

- Thus for any outcome $s \in \Omega$, we must have

$$P(s) = \frac{1}{|\Omega|}.$$

- In such a case, for any event A, we can write

$$P(A) = \sum_{s \in A} P(s) = \sum_{s \in A} \frac{1}{|\Omega|} = \frac{|A|}{|\Omega|}.$$

Consider a sample space that is an *uncountable* set.

- E.g., a 50-minute exam (so $\Omega = [0, 50]$), and let T_{Ant} be the time it takes Ant to finish the exam.
- What is the probability of $T_{Ant} \in [40, 45)$?