Chapter 1F: Axiomatic Probability and the Addition Rule

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:

- state the axioms of probability;
- understand probability as a mathematical function on events;
- derive basic consequences of the axioms;
- apply the addition rule;
- use complements and inclusion–exclusion;
- identify mutually exclusive events;
- understand why axiomatic probability generalizes naive probability.

Why Axiomatize Probability (or anything)?

Limitations of Naive Probability

Previously, we defined (naive) probability by

$$\mathbb{P}(A) = \frac{|A|}{|\mathcal{S}|}.$$

This works well when:

- the sample space is finite;
- all outcomes are equally likely.

But many important situations do not satisfy these assumptions.

Examples:

- infinitely many outcomes;
- continuous measurements;
- unequal likelihoods;
- etc.

The Axiomatic Point of View

Instead of defining probability by counting, we define probability abstractly using properties that probabilities should satisfy.

Idea

A probability measure assigns a number to each event:

$$A \longmapsto \mathbb{P}(A),$$

subject to certain axioms (i.e., principles, rules) .

The naive definition will become a special case of the general theory.

The Axioms of Probability

Probability Space

A probability model consists of:

1. $\mathcal{S}$: sample space (set of all possible outcomes in an experiment);
2. $\mathcal{F}$: a collection of subsets of $\mathcal{S}$ (aka "events") ;
3. $\mathbb{P}(\cdot)$: a probability function $\mathbb{P} : \mathcal{F} \to [0, 1]$.

So for any event $A \subseteq \mathcal{S}$ (or $A \in \mathcal{F}$), $\mathbb{P}(A)$ is the probability that the function $\mathbb{P}(\cdot)$ assigns to A.
These 3 elements together form a Probability Model:

$$(\mathcal{S}, \mathcal{F}, \mathbb{P})$$

Note that the probability function assigns numbers to sets (events), not to numbers (this might be new for us). So we will see things like $\mathbb{P}([1, 2])$, $\mathbb{P}((-\infty, 4])$, or $\mathbb{P}(\{2, 4, 6\})$

Axiom 1: Nonnegativity

Axiom 1

For every event $A \subseteq \mathcal{S}$,

$$\boxed{\mathbb{P}(A) \geq 0.}$$

Interpretations:

- probabilities cannot be negative;
- impossible events have probability 0;
- larger probabilities correspond to more likely events.

Axiom 2: Normalization

Axiom 2

The probability that "something happens", is 1:

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

In other words, $\mathbb{P}(\cdot)$ assigns the number 1 to the event $\mathcal{S} \in \mathcal{F}$.

Axiom 3: Additivity

Axiom 3

If

$$A_1, A_2, A_3, \ldots$$

are pairwise disjoint events, then

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

Pairwise disjoint means:

$$A_i \cap A_j = \emptyset \qquad \text{whenever } i \neq j.$$

Finite Additivity

Axiom 3 immediately implies the finite version:
If

$$A_1, \ldots, A_n$$

are disjoint, then

$$\mathbb{P}(A_1 \cup \cdots \cup A_n) = \mathbb{P}(A_1) + \cdots + \mathbb{P}(A_n).$$

In particular, if

$$A \cap B = \emptyset,$$

then

$$\boxed{\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B).}$$

Basic Consequences of the Axioms

Probability of the Empty Set

Proposition

$$\mathbb{P}(\emptyset) = 0.$$

Reason:

$$\mathcal{S} = \mathcal{S} \cup \emptyset,$$

and

$$\mathcal{S} \cap \emptyset = \emptyset.$$

Using additivity,

$$\mathbb{P}(\mathcal{S}) = \mathbb{P}(\mathcal{S}) + \mathbb{P}(\emptyset).$$

Since $\mathbb{P}(\mathcal{S}) = 1$, we obtain $\mathbb{P}(\emptyset) = 0$.

Proposition (Complement Rule)

For every event A,

$$\boxed{\mathbb{P}(A^c) = 1 - \mathbb{P}(A).}$$

Reason:

$$A \cup A^c = \mathcal{S},$$

and

$$A \cap A^c = \emptyset.$$

Thus,

$$1 = \mathbb{P}(\mathcal{S}) = \mathbb{P}(A) + \mathbb{P}(A^c).$$

Probabilities Are Between 0 and 1

Proposition

For every event A,
$$0 \leq \mathbb{P}(A) \leq 1.$$

Why?

Axiom 1 gives:
$$\mathbb{P}(A) \geq 0.$$

Also,
$$\mathbb{P}(A^c) \geq 0.$$

Using
$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A),$$

we obtain
$$1 - \mathbb{P}(A) \geq 0$$

Monotonicity

Proposition (Monotonicity)

For any event $A, B \subseteq \mathcal{S}$ if $A \subseteq B$, then
$$\boxed{\mathbb{P}(A) \leq \mathbb{P}(B).}$$

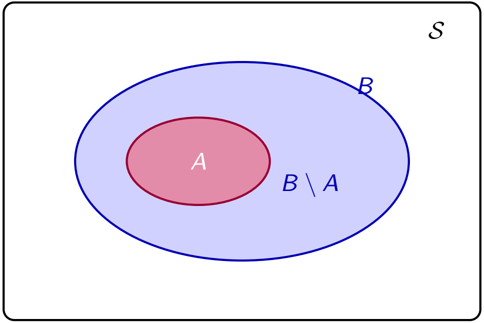
Proof.

If $A \subseteq B$, then $B = A \cup (B \setminus A)$, where the two events ($A$ and $B - A$) are disjoint (why?). Thus,

$$\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A),$$

since $\mathbb{P}(B - A) \geq 0$ according to Axiom 1. $\square$

Monotonicity: Picture



$$B = A \cup (B \setminus A)$$

$$A \cap (B \setminus A) = \emptyset$$

Since B contains all of A plus the extra part $B \setminus A$,

$$\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A).$$

The Addition Rule

Why Finite Additivity Is Not Enough

Suppose A and B are not disjoint.
Then:

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

Why?

Because outcomes in

$$A \cap B$$

are counted twice.

The Addition Rule

Proposition: Addition Rule

For any events A and B,

$$\boxed{\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).}$$

Proof.

Partition $A \cup B$ into three disjoint subsets and use a similar trick that was appkied to prove Monotonicity previously. $\square$

Why the Addition Rule Works

When adding

$$\mathbb{P}(A) + \mathbb{P}(B),$$

outcomes in

$$A \cap B$$

are counted twice.

Therefore we subtract one copy:

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

Mutually Exclusive Events

Definition

Events A and B are called **mutually exclusive** (or disjoint) if

$$A \cap B = \emptyset.$$

In that case,

$$\mathbb{P}(A \cap B) = 0,$$

so the addition rule becomes:

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B).$$

Example: Die Roll

Roll one fair die.
Let

$$A = \{\text{roll is even}\} = \{2, 4, 6\},$$
$$B = \{\text{roll is odd}\} = \{1, 3, 5\}.$$

Then

$$A \cap B = \emptyset.$$

Thus

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B).$$

Indeed,

$$\frac{1}{2} + \frac{1}{2} = 1.$$

Inclusion–Exclusion

Three-Set Inclusion–Exclusion

For three events:

$$\mathbb{P}(A \cup B \cup C)$$

we first add:

$$\mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C).$$

Then subtract pairwise intersections:

$$-\mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C).$$

But now the triple intersection has been subtracted too many times, so we add it back.

Three-Set Inclusion–Exclusion

$$\boxed{\begin{aligned}\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C)\\ - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C)\\ + \mathbb{P}(A \cap B \cap C).\end{aligned}}$$

Partitions

Partitions of the Sample Space

Definition

Events

$$A_1, \ldots, A_n$$

form a partition of $\mathcal{S}$ if:

1. $$A_i \cap A_j = \emptyset \qquad (i \neq j),$$
2. $$A_1 \cup \cdots \cup A_n = \mathcal{S}.$$

A partition divides the sample space into non-overlapping pieces.

Probability of a Partition

If

$$A_1, \ldots, A_n$$

form a partition of $\mathcal{S}$, then

$$\boxed{\mathbb{P}(A_1) + \cdots + \mathbb{P}(A_n) = 1.}$$

Reason:
The events are disjoint and their union is $\mathcal{S}$.

Naive Probability as a Special Case

Suppose:

- $\mathcal{S}$ is finite;
- all outcomes are equally likely.

Let $\mathcal{S} = \{s_1, s_2, \ldots, s_n\}$
Assign equal probability weights to each outcome:

$$\mathbb{P}(s_1) = \mathbb{P}(s_2) = \cdots = \mathbb{P}(s_n) = \frac{1}{n}.$$

Then for an event A, we have

$$\mathbb{P}(A) = \mathbb{P}(\cup_{s_j \in A} s_j) = \sum_{i \in A} \frac{1}{n} = |A| \times \frac{1}{n} = \frac{|A|}{|\mathcal{S}|}.$$

Thus the naive definition satisfies the probability axioms.

Mini-Exercises

1. If
$$\mathbb{P}(A) = 0.3,$$
what is
$$\mathbb{P}(A^c)?$$
2. If
$$\mathbb{P}(A) = 0.6, \qquad \mathbb{P}(B) = 0.5, \qquad \mathbb{P}(A \cap B) = 0.1,$$
compute
$$\mathbb{P}(A \cup B).$$
3. Can two disjoint events both have probability 1?

Answers

$$0.7, \qquad 1.0.$$

Summary

- Probability is defined axiomatically.
- The axioms are:

$$\mathbb{P}(A) \geq 0, \qquad \mathbb{P}(\mathcal{S}) = 1,$$

  and countable additivity.
- Important consequences include:

$$\mathbb{P}(\emptyset) = 0, \qquad \mathbb{P}(A^c) = 1 - \mathbb{P}(A), \qquad 0 \leq \mathbb{P}(A) \leq 1.$$

- The addition rule is:

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

- Naive probability is a special case of axiomatic probability.

Next lecture

Conditional probability and independence.