Probability Foundations

The Language of Uncertainty

Scott Cunningham

Harvard University
Department of Government

January 26, 2026

Why Start Here?

Today: The language of probability

Probability is the vocabulary for describing **populations** and **uncertainty**.

Before we can estimate anything, we need language to describe *what we're trying to learn.*

This course takes a "population-first" approach: define what you want to know about the population before worrying about estimation.

Sample Spaces and Events

The building blocks

What Is Probability?

A **model** for describing uncertainty about outcomes.

Three ingredients:

1. A **sample space** Ω: all possible outcomes
2. An **event space** $\mathcal{S}$: subsets of outcomes we care about
3. A **probability measure** $\mathbb{P}$: assigns numbers to events

Together, $(\Omega, \mathcal{S}, \mathbb{P})$ is a **probability space**.

Probability Is a Model

Not a property of the world

Consider flipping a coin. If you knew *everything*—the exact force applied, the coin's initial orientation, air resistance, the surface it lands on—you could predict exactly whether it lands heads or tails. There's nothing inherently "random" about a coin flip.

So what is probability?

It's a **model of our uncertainty**, not a feature of physical reality. We use probability because we *don't* know everything—it describes what we believe given our ignorance.

This is a key feature of the **agnostic approach** we take in this course (following Aronow & Miller). It's worth noting upfront.

"All models are wrong, but some are useful." — George Box

Sample Space

All possible outcomes

The **sample space** Ω is the set of all possible outcomes of a random process.

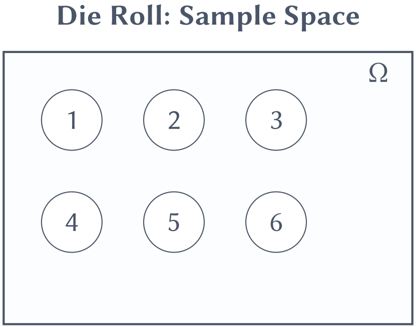
Examples:

- Coin flip: $\Omega = \{\text{Heads}, \text{Tails}\}$
- Die roll: $\Omega = \{1, 2, 3, 4, 5, 6\}$
- Two coin flips: $\Omega = \{HH, HT, TH, TT\}$
- Temperature tomorrow: $\Omega = \mathbb{R}$ (or some interval)

The sample space can be finite, countably infinite, or uncountable.

Visualizing the Sample Space

The universe of possibilities



The **sample space** $\Omega = \{1, 2, 3, 4, 5, 6\}$ contains *every* possible outcome.

Think of Ω as the "universe" — nothing can happen outside of it.

Events

Questions we can ask

An **event** is a subset of the sample space: $A \subseteq \Omega$.

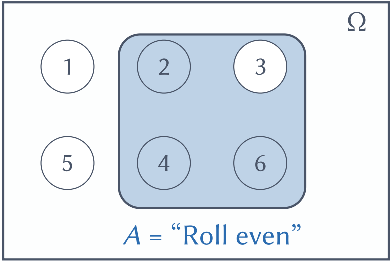
For a die roll ($\Omega = \{1, 2, 3, 4, 5, 6\}$):

- $A = \{6\}$: "Roll a six"
- $B = \{2, 4, 6\}$: "Roll an even number"
- $C = \{1, 2\}$: "Roll less than three"
- Ω: "Something happens" (the *certain* event)
- $\emptyset$: "Nothing happens" (the *impossible* event)

Events are the things we assign probabilities to.

Visualizing Events

Subsets of the sample space



The event $A = \{2, 4, 6\}$ is a **subset** of Ω: we write $A \subseteq \Omega$.

We assign probabilities to events: $\mathbb{P}(A) = \mathbb{P}(\text{"Roll even"}) = 3/6 = 1/2$

Sample Space vs. Events

$A \subseteq \Omega$

Sample Space Ω	**Event A**
All possible outcomes	*Some* possible outcomes
The "universe"	A subset of the universe
Fixed for a given experiment	Many different events possible
$\mathbb{P}(\Omega) = 1$ always	$0 \leq \mathbb{P}(A) \leq 1$

Die roll example:

- Sample space: $\Omega = \{1, 2, 3, 4, 5, 6\}$ — all six faces
- Event "roll even": $A = \{2, 4, 6\}$ — three of the six faces
- Event "roll a six": $B = \{6\}$ — just one face

Operations on Events

Events are sets, so we can combine them:

Operation	Notation	Meaning
Union	$A \cup B$	A or B (or both)
Intersection	$A \cap B$	A and B
Complement	A^c	not A
Difference	$A \setminus B$	A but not B

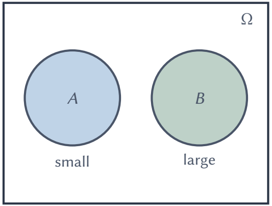
Example: Die roll, $A = \{2, 4, 6\}$ (even), $B = \{1, 2, 3\}$ (small)

- $A \cap B = \{2\}$ (even AND small)
- $A \cup B = \{1, 2, 3, 4, 6\}$ (even OR small)
- $A^c = \{1, 3, 5\}$ (odd)

Mutually Exclusive Events

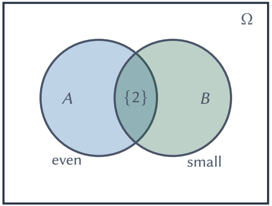
Two events are **mutually exclusive** (or *disjoint*) if they cannot both occur: $A \cap B = \emptyset$

Mutually Exclusive



$A \cap B = \emptyset$

NOT Mutually Exclusive



$A \cap B = \{2\} \neq \emptyset$

Example: Die roll — $A = \{1, 2, 3\}$ (small), $B = \{4, 5, 6\}$ (large) are mutually exclusive.

Why does this matter? It simplifies probability calculations: $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$

Kolmogorov Axioms

The rules probability must follow

A **probability measure** $\mathbb{P} : \mathcal{S} \to [0, 1]$ satisfies three axioms:

1. **Non-negativity**: $\mathbb{P}(A) \geq 0$ for all events A
   - $\rightarrow$ Probabilities can't be negative
2. **Normalization**: $\mathbb{P}(\Omega) = 1$
   - $\rightarrow$ Something must happen; probabilities sum to 1
3. **Countable additivity**: For mutually exclusive events $A_1, A_2, \ldots$:

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

   - $\rightarrow$ If events can't overlap, just add their probabilities

Everything else we'll derive follows from these three axioms.

Consequences of the Axioms

From the three axioms, we can prove:

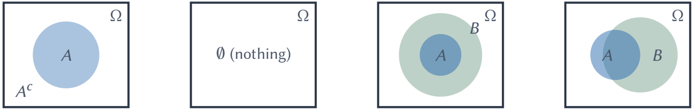
- **Complement rule**: $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
- **Impossible event**: $\mathbb{P}(\emptyset) = 0$
- **Monotonicity**: If $A \subseteq B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$
- **Subtraction rule**: $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$
- **Addition rule**: $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

The addition rule corrects for double-counting the intersection.

Visualizing the Consequences

Quick reference

Complement | **Impossible** | **Monotonicity** | **Subtraction**



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

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

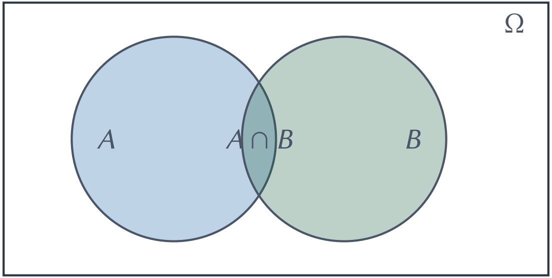
$A \subseteq B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$

$\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$

Each follows from the three axioms. Proofs are in the readings.

The Addition Rule

Visualizing inclusion-exclusion



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

If we add $\mathbb{P}(A)$ and $\mathbb{P}(B)$, we count the intersection twice.

Conditional Probability

Updating beliefs with new information

Conditional Probability

The key definition

The **conditional probability** of A given B is:

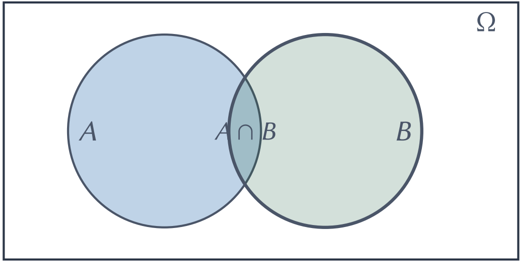
$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \qquad \text{provided } \mathbb{P}(B) > 0$$

Interpretation: The probability of A, *given that we know B occurred.*

We "zoom in" on the world where B happened and ask: how much of that world is A?

Conditional Probability

Visual intuition



$$\mathbb{P}(A \mid B) = \frac{\text{Probability of being in both } A \text{ and } B}{\text{Probability of being in } B}$$

Given that we're in B, what fraction is also in A?

Example: Two Dice

Roll two fair dice. What is $\mathbb{P}(\text{sum} = 8 \mid \text{first die} = 3)$?

Solution:

- Let $A = \{\text{sum} = 8\}$ and $B = \{\text{first die} = 3\}$
- $\mathbb{P}(B) = 6/36 = 1/6$ (six outcomes where first die is 3)
- $A \cap B = \{(3, 5)\}$ (only way to get sum 8 with first die 3)
- $\mathbb{P}(A \cap B) = 1/36$

$$\mathbb{P}(A \mid B) = \frac{1/36}{1/6} = \frac{1}{6}$$

Compare to $\mathbb{P}(\text{sum} = 8) = 5/36 \approx 0.14$. Knowing the first die changes things!

The Multiplicative Law

Rearranging the definition of conditional probability:

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

Or equivalently:

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

The chain rule (for three events):

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A) \cdot \mathbb{P}(B \mid A) \cdot \mathbb{P}(C \mid A \cap B)$$

Independence

When knowing one thing tells you nothing about another

Independence of Events

Definition

Events A and B are **independent** if:

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

Equivalent statement (when $\mathbb{P}(B) > 0$):

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

Knowing B occurred doesn't change the probability of A.

Independence means information is irrelevant. Learning B happened gives you no information about whether A happened.

Notation: $A \perp\!\!\!\perp B$ means "A is independent of B"

Independence vs. Mutual Exclusivity

These are NOT the same thing!

Mutually exclusive: $A \cap B = \emptyset$ (can't both happen)

Independent: $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ (knowing one doesn't affect the other)

In fact, they're almost opposites!

If A and B are mutually exclusive with $\mathbb{P}(A) > 0$ and $\mathbb{P}(B) > 0$:

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{0}{\mathbb{P}(B)} = 0 \neq \mathbb{P}(A)$$

So mutually exclusive events are **dependent** (strongly so!).

If I know B happened, I know A didn't happen.

Example: Coin Flips

Flip a fair coin twice. Let $A = \{\text{first flip is Heads}\}$ and $B = \{\text{second flip is Heads}\}$.

Are A and B independent?

Check:

- $\mathbb{P}(A) = 1/2, \quad \mathbb{P}(B) = 1/2$
- $\mathbb{P}(A \cap B) = \mathbb{P}(\{HH\}) = 1/4$
- $\mathbb{P}(A) \cdot \mathbb{P}(B) = (1/2)(1/2) = 1/4$ ✓

Yes, they are independent. The outcome of one flip doesn't affect the other.

Example: Drawing Cards

Draw two cards from a deck **without replacement**. Let:

- $A = \{\text{first card is an Ace}\}, \quad B = \{\text{second card is an Ace}\}$

Are A and B independent?

Check:

- $\mathbb{P}(A) = 4/52$
- $\mathbb{P}(B \mid A) = 3/51$ (if first was Ace, only 3 Aces left in 51 cards)
- $\mathbb{P}(B \mid A^c) = 4/51$ (if first wasn't Ace, still 4 Aces in 51 cards)

Since $\mathbb{P}(B \mid A) \neq \mathbb{P}(B \mid A^c)$, knowing A changes $\mathbb{P}(B)$.

No, they are **not** independent.

Bayes' Rule

Reversing conditional probabilities

Motivation: Strategic Thinking Under Uncertainty

Example: You're playing poker, and the person in front of you raises.

What's your best response?

- It depends on what you *learned* from that raise
- And what cards you're holding

This requires us to **update our beliefs** based on new information.

We need to calculate conditional probabilities—but often we know them "backwards."

Rev. Thomas Bayes (1701–1761), English statistician and Presbyterian minister.

Deriving Bayes' Rule

Step by step from definitions

Let A and B be two events. We want $\mathbb{P}(A \mid B)$.

Start with the definition of conditional probability:

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \quad \Longrightarrow \quad \mathbb{P}(A \cap B) = \mathbb{P}(A \mid B) \cdot \mathbb{P}(B)$$

Similarly:

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)} \quad \Longrightarrow \quad \mathbb{P}(B \cap A) = \mathbb{P}(B \mid A) \cdot \mathbb{P}(A)$$

Deriving Bayes' Rule

The key insight

Since $\mathbb{P}(A \cap B) = \mathbb{P}(B \cap A)$:

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

Solve for $\mathbb{P}(A \mid B)$:

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

This is **Bayes' Rule** (naive form).

It lets us "flip" conditional probabilities: from $\mathbb{P}(B \mid A)$ to $\mathbb{P}(A \mid B)$.

The Law of Total Probability

A consequence of the additivity axiom

Observation: We can decompose B using a **partition** of Ω: $B = (B \cap A) \cup (B \cap A^c)$

A partition is a collection of mutually exclusive, exhaustive "bins." Here $\{A, A^c\}$ partitions Ω.

These pieces are mutually exclusive, so by the **additivity axiom**:

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

Apply the multiplicative law:

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

The unconditional probability is a weighted average of conditional probabilities.

This gives us what we need for Bayes' denominator.

Bayes' Rule: Full Form

Substituting the Law of Total Probability into Bayes' Rule:

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

Terminology:

- $\mathbb{P}(A)$: **Prior** — belief before seeing B
- $\mathbb{P}(B \mid A)$: **Likelihood** — how likely is B if A is true?
- $\mathbb{P}(A \mid B)$: **Posterior** — updated belief after seeing B

Example: The Monty Hall Problem

Setup

The scenario:

There are three doors labeled 1, 2, and 3.

Behind one door is a million dollars; behind the other two are goats.

1. You select door 1
2. The host, Monty Hall, opens door 2 and shows you a goat
3. Monty asks: "Would you like to switch from door 1 to door 3?"

Question: What's the probability that the money is behind door 3?

Should you switch?

Example: The *Monty* Hall Problem

Setting up Bayes' Rule

Define events:

- D_1 = "money is behind door 1"
- D_2 = "money is behind door 2"
- D_3 = "money is behind door 3"
- O = "Monty opened door 2"

We want $\mathbb{P}(D_3 \mid O)$ using Bayes' Rule:

$$\mathbb{P}(D_3 \mid O) = \frac{\mathbb{P}(O \mid D_3) \cdot \mathbb{P}(D_3)}{\mathbb{P}(O \mid D_1)\mathbb{P}(D_1) + \mathbb{P}(O \mid D_2)\mathbb{P}(D_2) + \mathbb{P}(O \mid D_3)\mathbb{P}(D_3)}$$

Priors: $\mathbb{P}(D_1) = \mathbb{P}(D_2) = \mathbb{P}(D_3) = \frac{1}{3}$

Example: The Monty Hall Problem

The likelihoods

Key insight: Monty *knows* where the money is and will *never* open a door with money.

What is $\mathbb{P}(O \mid D_i)$? (Given the money is behind door i, what's the probability Monty opens door 2?)

1. $\mathbb{P}(O \mid D_1) = 0.5$
   Money behind door 1. Monty can choose door 2 or 3 randomly.
2. $\mathbb{P}(O \mid D_2) = 0$
   Money behind door 2. Monty would never open door 2!
3. $\mathbb{P}(O \mid D_3) = 1$
   Money behind door 3. Monty must open door 2 (can't open door 1 or 3).

Example: The Monty Hall Problem

The calculation

$$\mathbb{P}(D_3 \mid O) = \frac{\mathbb{P}(O \mid D_3) \cdot \mathbb{P}(D_3)}{\mathbb{P}(O \mid D_1)\mathbb{P}(D_1) + \mathbb{P}(O \mid D_2)\mathbb{P}(D_2) + \mathbb{P}(O \mid D_3)\mathbb{P}(D_3)}$$

Substituting:

$$\mathbb{P}(D_3 \mid O) = \frac{1 \cdot \frac{1}{3}}{\frac{1}{2} \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}}$$

$$= \frac{\frac{1}{3}}{\frac{1}{6} + 0 + \frac{1}{3}} = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$$

Example: The Monty Hall Problem

The answer

$$\mathbb{P}(D_3 \mid O) = \frac{2}{3} \qquad \mathbb{P}(D_1 \mid O) = \frac{1}{3}$$

Definitely switch to door 3!

Intuition:

- When you picked door 1, you had a $\frac{1}{3}$ chance of being right
- The other two doors collectively had $\frac{2}{3}$ probability
- Monty's action *concentrates* that $\frac{2}{3}$ onto door 3

Marilyn vos Savant published this solution in Parade magazine. Thousands of readers—including mathematicians—wrote to say she was wrong. She was right.

Example: Medical Testing

Setup

A disease affects 1% of the population. A test has:

- 95% sensitivity: $\mathbb{P}(\text{positive} \mid \text{disease}) = 0.95$
- 90% specificity: $\mathbb{P}(\text{negative} \mid \text{no disease}) = 0.90$

Question: If you test positive, what's the probability you have the disease?

Given:

- $\mathbb{P}(D) = 0.01$, so $\mathbb{P}(D^c) = 0.99$
- $\mathbb{P}(+ \mid D) = 0.95$
- $\mathbb{P}(+ \mid D^c) = 0.10$ (false positive rate = 1 − 0.90)

Example: Medical Testing

Applying Bayes' Rule

$$\mathbb{P}(D \mid +) = \frac{\mathbb{P}(+ \mid D) \cdot \mathbb{P}(D)}{\mathbb{P}(+ \mid D)\mathbb{P}(D) + \mathbb{P}(+ \mid D^c)\mathbb{P}(D^c)}$$

First, find $\mathbb{P}(+)$ using the Law of Total Probability:

$$\mathbb{P}(+) = (0.95)(0.01) + (0.10)(0.99) = 0.0095 + 0.099 = 0.1085$$

Then:

$$\mathbb{P}(D \mid +) = \frac{(0.95)(0.01)}{0.1085} = \frac{0.0095}{0.1085} \approx 0.088$$

Example: Medical Testing

The surprising result

Even with a positive test, there's only an **8.8%** chance you have the disease!

Why so low?

- The disease is rare (1% prevalence)
- Most positive tests are false positives from the 99% without disease
- The 10% false positive rate applied to 99% $\gg$ the 95% true positive rate applied to 1%

Base rates matter. This is called the "base rate fallacy" when people ignore priors.

Why Independence Matters

Independence dramatically simplifies calculations:

- $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$ (no need to find conditional)
- For n independent events: $\mathbb{P}(A_1 \cap \cdots \cap A_n) = \prod_{i=1}^{n} \mathbb{P}(A_i)$

In this course:

- The **i.i.d. assumption** (coming in a few weeks) assumes observations are independent
- Many of our results depend on independence
- When independence fails, we need different tools (clustering, time series)

Today's Key Ideas

1. **Sample spaces and events**: The vocabulary for describing outcomes
2. **Kolmogorov axioms**: Non-negativity, normalization, additivity
3. **Conditional probability**: $\mathbb{P}(A \mid B) = \mathbb{P}(A \cap B)/\mathbb{P}(B)$
4. **Independence**: $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$
5. **Law of Total Probability**: Follows from additivity axiom
6. **Bayes' Rule**: Derived from conditional probability; flips conditionals

This is the language. Next: the objects we'll actually work with.

Looking Ahead

Wednesday: Conditional probability and Bayes' Rule (continued)

- More examples of Bayes' Rule
- Law of Total Probability applications

Next week: Random variables, expectation, and variance

Week 3: Famous distributions (two full lectures)

We're building the vocabulary to describe populations precisely.

For Wednesday

Reading:

- Aronow & Miller, §1.1: Review today's material
- Blackwell, Chapter 2.1: Probability foundations

Think about:

- In the medical testing example, what would happen if prevalence were 10% instead of 1%?
- Can you think of real-world examples where base rate neglect causes problems?

Questions?