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.

Part I

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.

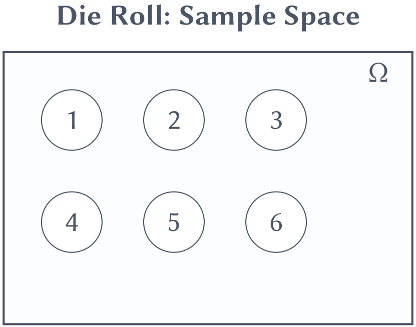
Examples:

- 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: $A \subseteq \Omega$.

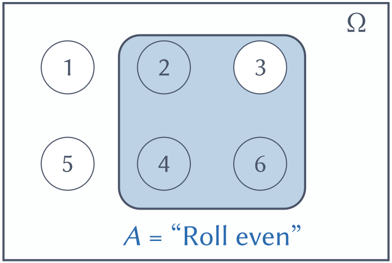
For a die roll ($\Omega = \{1, 2, 3, 4, 5, 6\}$):

- $A = \{6\}$: "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

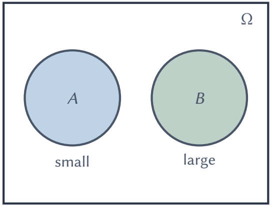
Example: Die roll, $A = \{2, 4, 6\}$ (even), $B = \{1, 2, 3\}$ (small)

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

Mutually Exclusive Events

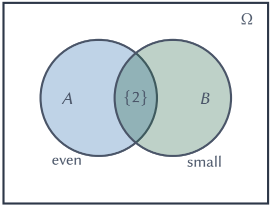
Two events are **mutually exclusive** (or *disjoint*) if they cannot both occur: $A \cap B = \emptyset$

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

Event Spaces

Which subsets can we assign probabilities to?

Election Night

A story about asking questions

You're a campaign manager. It's election night. The race is tight.

Three precincts remain uncounted: **North**, **South**, and **Central**.

Exactly one of them will push your candidate over the top—the *decisive* precinct.

You're asking questions:

- "Will North be the one that decides it?"
- "Will it be one of our strongholds?"

In probability language: $\Omega = \{\text{North}, \text{South}, \text{Central}\}$

Exactly one outcome will occur—whichever precinct decides the race.

Starting with the Bookends

The certain and the impossible

Before asking specific questions, we need two things:

The certain event $\Omega = \{N, S, C\}$:

"Some precinct will be decisive"—exactly one of them will decide the race. Probability = 1.

The impossible event $\emptyset$:

"No precinct decides it"—can't happen; one of them must. Probability = 0.

Our event space so far: $\mathcal{S} = \{\emptyset, \Omega\}$

This is the **minimal event space**. We can only say "some precinct decided it." We can't yet ask *which* one.

A Note on Notation

The empty set symbol

This symbol: $\emptyset$

It looks like the Greek letter theta (θ), but it's not.

It's the **empty set** symbol — a zero with a slash through it.

Pronunciation: Just say "the empty set."

The symbol comes from Scandinavian languages (Ø), introduced by French mathematicians in the 1930s.

What Is the Empty Set?

The impossible event

The empty set $\emptyset$ confuses people.

What does "no outcome" mean?

$\emptyset$ is **the event that contains no outcomes**.

Since *some* precinct must be decisive, $\emptyset$ cannot happen.

It's the **impossible event**. Probability = 0.

But where does it come from? Why must every event space include it?

Why Must $\emptyset$ Be in Every Event Space?

Complements again

Because $\emptyset$ is the complement of Ω:

- Ω = "something happens" (the certain event)
- Ω^c = "nothing happens" = $\emptyset$ (the impossible event)

If Ω is in our event space, then $\emptyset$ must be too.

$\emptyset$ and Ω are the bookends—every event space contains both.

Adding a Question

Building up the event space

Now you want to ask: “Was North the decisive precinct?”

That’s the event $\{\text{North}\}$—the set containing just North.

Can we just add $\{\text{North}\}$ to our event space?

Not by itself. If we can ask “Was it North?” we must also be able to ask “Was it *not* North?”

That’s the complement: $\{\text{North}\}^c = \{\text{South}, \text{Central}\}$

Adding a Question

We must add both

So we must add both the question and its opposite:

$\mathcal{S} = \{\emptyset, \{\text{North}\}, \{\text{South}, \text{Central}\}, \Omega\}$

Now we can ask:

- "Was North decisive?" ({North})
- "Was North *not* decisive?" ({South, Central})

How to read this: {South, Central} means "the decisive precinct was *either* South *or* Central." The set lists the possibilities.

Questions Come in Pairs

The complement rule

This is **Property 2** of event spaces: *closed under complements.*

If you can ask a question, you can ask its opposite.

Question	**Opposite**
Was North decisive? $\{\mathrm{N}\}$	Was North *not* decisive? $\{\mathrm{S}, \mathrm{C}\}$
Did some precinct decide it? Ω	Did no precinct decide it? $\emptyset$

Every event in $\mathcal{S}$ must have its complement in $\mathcal{S}$ too.

Combining Questions

Unions

What if you want to ask: "Was it North *or* South that decided it?"

That's a union: $\{\text{North}\} \cup \{\text{South}\} = \{\text{North}, \text{South}\}$

Property 3: If both $\{N\}$ and $\{S\}$ are in your event space, then $\{N, S\}$ must be too.

And if $\{N, S\}$ is in, its complement $\{C\}$ must be in (Property 2).

The event space grows:

Adding questions forces you to add their opposites and combinations.

The Power Set

The maximal event space

The **power set** 2^{Ω} is the collection of *all* subsets of Ω.

How many subsets? Each outcome is either *in* or *out* of a subset.

With n outcomes, that's $2 \times 2 \times \cdots \times 2 = 2^n$ possible subsets.

Our example: $\Omega = \{N, S, C\}$ has 3 outcomes, so $2^3 = 8$ subsets.

The power set is always a valid event space—it's "maximal" because it lets you ask *any* question.

The Power Set

All 8 subsets, in complement pairs

For $\Omega = \{N, S, C\}$, the power set contains:

Event	Complement
$\emptyset$	$\Omega = \{N, S, C\}$
$\{N\}$	$\{S, C\}$
$\{S\}$	$\{N, C\}$
$\{C\}$	$\{N, S\}$

That's 8 sets total (4 pairs).

Every element has a complement, and every possible question about which precinct was decisive can be asked.

Quick Notation Review

Translating the story to symbols

Now let's formalize what we just built:

Symbol	Meaning
$\Omega = \{N, S, C\}$	Sample space: all possible outcomes
$\{N\}$	A **singleton**: "North was decisive"
$\{N, S\}$	"North or South was decisive"
$\emptyset$	The empty set: the impossible event
$\mathcal{S}$	Event space: the collection of questions we can ask

Key point: $\mathcal{S}$ is a "set of sets." Each element of $\mathcal{S}$ is a subset of Ω.

Example: $\mathcal{S} = \{\emptyset, \{N\}, \{S, C\}, \Omega\}$ contains four sets (questions).

Why Do We Need Event Spaces?

A puzzle about infinite sample spaces

For finite sample spaces like $\Omega = \{1, 2, 3, 4, 5, 6\}$, we can assign probabilities to *every* subset (all $2^6 = 64$ of them).

But what about infinite sample spaces?

If $\Omega = [0, 1]$ (all real numbers from 0 to 1), there are *uncountably many* subsets. Some of these subsets are so pathological (i.e., so weird we can't assign probabilities to them) that we *cannot* assign probabilities to them consistently.

The solution: Don't try to assign probabilities to *every* subset. Instead, specify a collection of "well-behaved" subsets—an **event space**—and only assign probabilities to those.

For finite sample spaces, the power set (all subsets) always works. The machinery of event spaces becomes essential for continuous random variables.

What Is an Event Space?

The three properties

An **event space** (or σ-algebra) $\mathcal{S}$ on Ω is a *collection of subsets* of Ω. Think of $\mathcal{S}$ as a "set of sets"—each element of $\mathcal{S}$ is itself a subset of Ω. The notation $A \in \mathcal{S}$ means "A is one of the sets in $\mathcal{S}$."

A valid event space must satisfy three properties:

1. **Contains the sample space**: $\Omega \in \mathcal{S}$ (Ω is one of the sets in $\mathcal{S}$)
2. **Closed under complements**: If $A \in \mathcal{S}$, then $A^c \in \mathcal{S}$
3. **Closed under countable unions**: If $A_1, A_2, \ldots \in \mathcal{S}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{S}$

Key point: These properties ensure that when we combine events using "or", "and", or "not", we stay inside the event space.

Common Patterns

What to look for

The trivial event space $\mathcal{S} = \{\emptyset, \Omega\}$ *always* works.

Event spaces come in complement pairs:

If $A \in \mathcal{S}$, then A^c must also be in $\mathcal{S}$.

So if you see $\{A\}$ without $\{B, C, D\}$ (its complement)? Fails Property 2.

Unions can create new sets:

If $\{1\}$ and $\{2\}$ are both in $\mathcal{S}$, then $\{1, 2\}$ must be too.

So if you see $\{1\}$ and $\{2\}$ but not $\{1, 2\}$? Fails Property 3.

The power set 2^Ω (all subsets) is always valid—it's the "maximal" event space.

Consequences of the Three Properties

What follows automatically

From the three properties, we get additional facts *for free*:

The empty set is always in $\mathcal{S}$:

- By (1): $\Omega \in \mathcal{S}$
- By (2): $\Omega^c = \emptyset \in \mathcal{S}$ ✓

Closed under countable intersections (De Morgan's laws):

- If $A_1, A_2, \ldots \in \mathcal{S}$, then $\bigcap_{i=1}^{\infty} A_i \in \mathcal{S}$
- Why? $\bigcap A_i = \left(\bigcup A_i^c\right)^c$, and we can take complements and unions

Closed under set difference:

- If $A, B \in \mathcal{S}$, then $A \setminus B = A \cap B^c \in \mathcal{S}$

The three properties give us a "closed system" for doing set operations.

Singletons

Sets with exactly one element

A **singleton** is a set containing exactly one outcome.

Example: If $\Omega = \{A, B, C\}$, the singletons are $\{A\}$, $\{B\}$, and $\{C\}$.

Why do singletons matter?

- Singletons represent *elementary events*—specific outcomes
- If every singleton is in $\mathcal{S}$, we can ask "What's the probability of outcome A?"
- The power set (all 2^n subsets) always includes all singletons
- But smaller event spaces might not include all singletons!

Example: $\mathcal{S} = \{\emptyset, \Omega\}$ is a valid event space, but it contains *no* singletons. We can only ask about "something happens" or "nothing happens"—not about specific outcomes.

PS1 asks about singletons on infinite sample spaces. Stay tuned.

Checking Event Spaces: The Method

A systematic approach

Given: A sample space Ω and a collection of sets $\mathcal{S}$.

Question: Is $\mathcal{S}$ an event space?

The checklist:

1. **Property 1**: Is $\Omega \in \mathcal{S}$? (Also check: is $\emptyset \in \mathcal{S}$?)
2. **Property 2**: For *every* set $A \in \mathcal{S}$, compute A^c and check if $A^c \in \mathcal{S}$
3. **Property 3**: For *every pair* of sets in $\mathcal{S}$, compute their union and check if it's in $\mathcal{S}$

Key Skills for Checking Event Spaces

What you need to do be able to do

Computing complements:

If $\Omega = \{1, 2, 3\}$ and $A = \{1\}$, then $A^c = \{2, 3\}$

Computing unions:

$\{1\} \cup \{2, 3\} = \{1, 2, 3\} = \Omega$

Being systematic:

Check *every* element, not just one or two. If even one complement or union is missing, the collection fails to be an event space.

Let's work through some examples on the board...

Let's Practice

Board work

Examples on the board

We'll work through checking whether specific collections are event spaces:

1. A collection that *is* an event space (verify all three properties)
2. A collection that *fails* to be an event space (find where it breaks)
3. How to *construct* a new event space by adding sets carefully

These skills are directly relevant to Problem Set 1.

The Power Set

The "maximal" event space

The **power set** 2^Ω is the collection of *all* subsets of Ω.

Example: If $\Omega = \{1, 2, 3\}$, then 2^Ω has $2^3 = 8$ elements:

$$2^\Omega = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$$

The power set is always a valid event space:

- Property 1: $\Omega \in 2^\Omega$? ✓ (every set is a subset of itself)
- Property 2: Closed under complements? ✓ (A^c is always a subset of Ω)
- Property 3: Closed under unions? ✓ (union of subsets is a subset)

For finite Ω, we almost always use the power set. The distinction matters for infinite sample spaces.

Countable Infinite Sample Spaces

A puzzle to think about

Consider a **countably infinite** sample space: $\Omega = \{A_1, A_2, A_3, \ldots\}$

Question: Can we assign *equal* probability to every singleton $\{A_i\}$?

That is, can we have $\mathbb{P}(\{A_i\}) = p$ for the *same* value p for all i?

Hint: Think about what the axioms tell us:

- What does **normalization** say about $\mathbb{P}(\Omega)$?
- What does **countable additivity** say when events are mutually exclusive?
- What are the possible values p could take?

This is Problem Set 1, Question 2. Use the axioms to work through the answer carefully.

The axioms of probability have real consequences—they constrain what's possible.

Event Spaces: Summary

What you need to know for PS1

An event space $\mathcal{S}$ must satisfy three properties:

1. Contains Ω
2. Closed under complements
3. Closed under countable unions

To verify: Check each property explicitly—verify *every* complement and *every* union.

To construct: Start with $\{\emptyset, \Omega\}$, add sets in complement pairs.

The power set 2^{Ω} is always a valid event space.

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

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

   → 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:

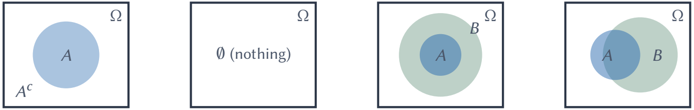
- **Complement rule**: $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
- **Impossible event**: $\mathbb{P}(\emptyset) = 0$
- **Monotonicity**: If $A \subseteq B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$
- **Subtraction rule**: $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$
- **Addition rule**: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

The addition rule corrects for double-counting the intersection.

Visualizing the Consequences

Quick reference

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

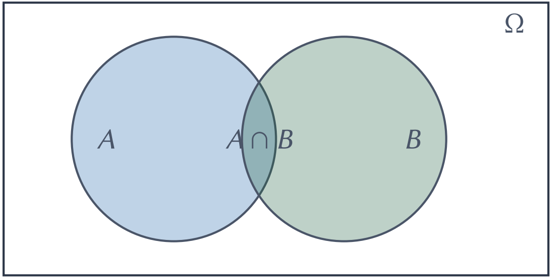


$\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ $\mathbb{P}(\emptyset) = 0$ $A \subseteq B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$ $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$

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.

Part II

Conditional Probability

Updating beliefs with new information

Conditional Probability

The key definition

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

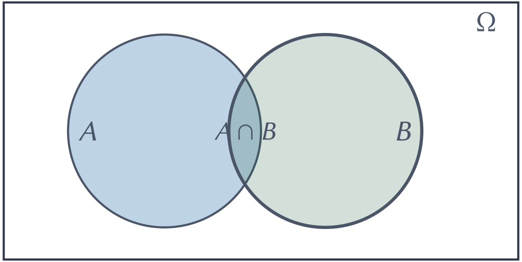
$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \qquad \text{provided } \mathbb{P}(B) > 0$$

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:

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

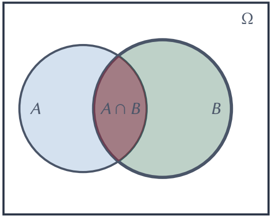
The chain rule (for three events):

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

Visualizing the Multiplicative Law

Two ways to compute $\mathbb{P}(A \cap B)$

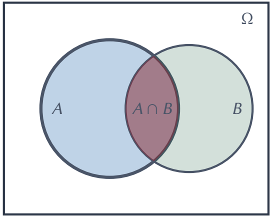
Start with B, then find A



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

"What fraction of B is also A?"

Start with A, then find B



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

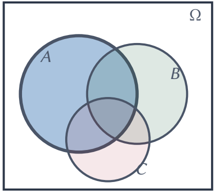
"What fraction of A is also B?"

Key insight: The intersection $A \cap B$ is the same region either way—we're just computing its probability through different "doors."

Visualizing the Chain Rule

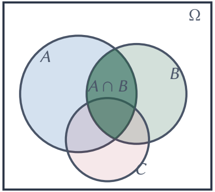
$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A) \cdot \mathbb{P}(B \mid A) \cdot \mathbb{P}(C \mid A \cap B)$

Step 1: Start with A



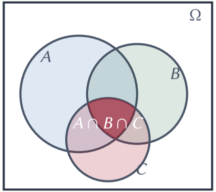
$\mathbb{P}(A)$

Step 2: B given A



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

Step 3: C given $A \cap B$



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

Intuition: Keep "zooming in." First restrict to A, then to $A \cap B$, then ask what fraction is also in C.

Each step narrows the universe; each conditional probability asks "what fraction of where we are is also in the next set?"

Part III

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:

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

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

The Reverend Thomas Bayes (1701–1761)

A thought experiment with billiard balls

Thomas Bayes was an English Presbyterian minister and amateur mathematician. His famous essay was published posthumously in 1763 by his friend Richard Price.

Bayes' thought experiment: Imagine a billiard table. A ball is rolled and comes to rest somewhere—you don't see where. Then more balls are rolled, and you're told whether each lands to the left or right of the first ball.

The question: Given this evidence, what can you infer about where the first ball is?

Key insight: Each new observation lets us *update* our beliefs about the unknown quantity. We start with uncertainty (a "prior"), observe evidence, and arrive at refined beliefs (a "posterior").

This is the logic of Bayesian inference: prior $\times$ likelihood $\longrightarrow$ posterior.

A Controversial Idea

The frequentist-Bayesian debate

Bayes' approach was **controversial** for over 200 years. Why?

Frequentist objection: Probability should describe long-run frequencies of *repeatable* events. "There's a 70% probability the ball is in the left half" seemed unscientific—the ball is either there or it isn't!

Bayesian response: Probability describes our *uncertainty*, not physical randomness. It's sensible to have beliefs about fixed but unknown quantities.

Historical irony: Statisticians dismissed Bayesian methods as "subjective" well into the 20th century. Yet Bayesian reasoning proved essential in:

- Breaking the Nazi Enigma code (1940s)
- Finding lost submarines and aircraft (1960s–today)
- Modern machine learning and AI

Today, most statisticians use both approaches as tools for different problems.

Bayes' Rule in History

From codebreaking to search and rescue

Alan Turing (1940s): Used sequential Bayesian updating to crack the Nazi Enigma code at Bletchley Park. Each piece of intercepted message updated beliefs about machine settings. He called it "banburismus" after the town of Banbury, where the paper strips were made.

Search Theory (1960s–today):

- **USS *Scorpion* (1968)**: Navy used Bayesian search to find the lost submarine
- **Air France 447 (2009)**: After two years of failed searches, Bayesian methods found the wreckage
- **Steve Fossett (2007)**: Updated probability maps based on search patterns
- **MH370 (2014)**: Bayesian analysis of satellite data guided the search

The logic: Start with prior beliefs about location. Each failed search in an area *lowers* the probability there and *raises* it elsewhere.

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."

A Hiring Problem

The firm's dilemma

You're a firm deciding whether to hire a job applicant.

The problem: Workers come in two types:

- **High type** — productive, will make your firm money
- **Low type** — unproductive, will cost you money

Unfortunately, you *can't directly observe* which type they are. But you *can* observe whether they have an MBA.

Your question: Given that this applicant has an MBA, what's the probability they're a high type?

This problem combines three Nobel Prizes: signaling (Spence), types (Harsanyi), and subgame perfect Bayesian Nash equilibrium (Selten). We'll just do the Bayes part.

Setting Up the Problem

Events and notation

Let's define two events:

- H = "Worker is high type" (productive)
- MBA = "Worker has an MBA"

The firm wants to know: $\mathbb{P}(H \mid MBA)$

"What's the probability this worker is productive, *given* that they have an MBA?"

The MBA is something economists call a **signal**—an observable action that might reveal hidden information about type.

What Can the Firm Observe?

The available data

The firm can observe:

- $\mathbb{P}(H)$ — the base rate of high types in the population
- $\mathbb{P}(MBA)$ — the fraction of applicants with MBAs
- $\mathbb{P}(MBA \mid H)$ — among high types, what fraction get MBAs?

But the firm wants $\mathbb{P}(H \mid MBA)$ — the probability of high type *given* they see the signal.

The problem: We know $\mathbb{P}(MBA \mid H)$, but we want $\mathbb{P}(H \mid MBA)$.

We need to "flip" the conditional. That's what Bayes' Rule does.

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

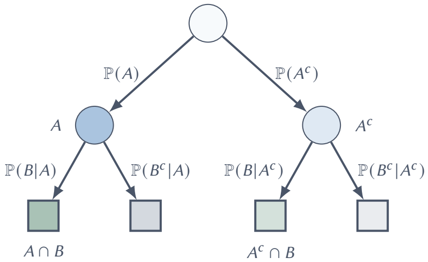
$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A) \cdot \mathbb{P}(A)}{\mathbb{P}(B)}$$

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

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

Visualizing Bayes' Rule

The tree diagram: forward vs. backward



Forward (what we often know):
Nature "picks" A or A^c first,
then B happens (or not).

Backward (what we want):
We *observe* B. Given that,
was it A or A^c?

Bayes' Rule: Of all the ways to reach B, what fraction came through A?

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A \cap B) + \mathbb{P}(A^c \cap B)} = \frac{\text{green path through } A}{\text{all green paths}}$$

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

Back to the Hiring Problem

Connecting terminology to the example

The firm asked: What is $\mathbb{P}(H \mid MBA)$?

In Bayes' terminology:

- $\mathbb{P}(H)$ is the **prior** — base rate of high types before seeing the signal
- $\mathbb{P}(MBA \mid H)$ is the **likelihood** — how likely high types are to get MBAs
- $\mathbb{P}(H \mid MBA)$ is the **posterior** — updated belief after seeing the signal

The likelihood $\mathbb{P}(MBA \mid H)$ captures the **signal strength**:

If high types are *much more likely* to get MBAs than low types, the signal is informative.

Back to the Hiring Problem

What data would the firm need?

Using Bayes' Rule:

$$\mathbb{P}(H \mid MBA) = \frac{\mathbb{P}(MBA \mid H) \cdot \mathbb{P}(H)}{\mathbb{P}(MBA)}$$

The firm would need:

- $\mathbb{P}(H)$ — base rate of high types in the applicant pool
- $\mathbb{P}(MBA)$ — fraction of applicants with MBAs
- $\mathbb{P}(MBA \mid H)$ — among high types, what fraction have MBAs?

With these three numbers, Bayes' Rule tells the firm how much to update.

Whether to actually *hire* is a separate decision—but now they know the probability.

Two Applications

Bayes' Rule in action

Now let's see Bayes' Rule applied to two famous problems:

1. **The Monty Hall Problem** — a game show that fooled 1,000 PhDs
2. **Medical Testing** — when a positive test doesn't mean what you think

Both involve "flipping" conditional probabilities.

“Ask Marilyn” and the Monty Hall Firestorm

September 9, 1990

Marilyn vos Savant—listed in the *Guinness Book of World Records* for highest recorded IQ—wrote a column in *Parade* magazine. A reader sent this puzzle:

> *“Suppose you’re on a game show, and you’re given the choice of three doors. Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what’s behind the doors, opens another door, say No. 3, which has a goat. He then says, ‘Do you want to pick door No. 2?’ Is it to your advantage to switch?”*

Marilyn’s answer: **Yes, you should switch.** Switching gives you a $\frac{2}{3}$ chance of winning.

What happened next would reveal something uncomfortable about how experts respond to being corrected—especially by a woman.

The Backlash

When mathematicians get it wrong

Marilyn received approximately **10,000 letters**, nearly 1,000 from PhDs. Many were hostile:

> *"You blew it, and you blew it big! Since you seem to have difficulty grasping the basic principle at work here, I'll explain..." —PhD, Georgetown University*
>
> *"You are utterly incorrect... How many irate mathematicians are needed to get you to change your mind?" —PhD, U.S. Army Research Institute*
>
> *"You made a mistake, but look at the positive side. If all those PhDs were wrong, the country would be in very serious trouble." —PhD, Univ. of Florida*

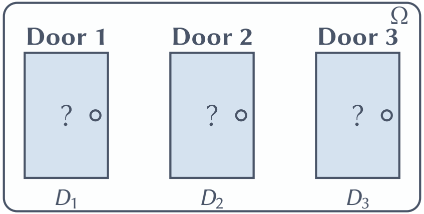
The twist: Marilyn was right. The PhDs were wrong.

Let's work through why, using Bayes' Rule.

The Monty Hall Problem

Step 1: The prior

Three doors. Behind one is \$1 million; behind the other two are goats.



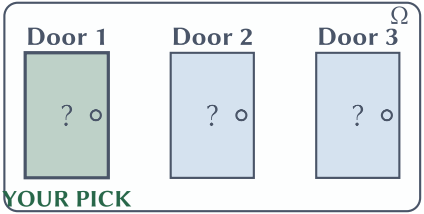
The events D_1, D_2, D_3 ("money behind door i") are **mutually exclusive** and **exhaustive**. With no information, each is equally likely:

$$\mathbb{P}(D_1) = \mathbb{P}(D_2) = \mathbb{P}(D_3) = \frac{1}{3} \qquad \text{(the } \textbf{prior}\text{)}$$

The Monty Hall Problem

Step 2: You choose

You pick Door 1. Since the doors are equally likely, it doesn't matter which you choose.



At this point, what's the probability you picked the winning door?

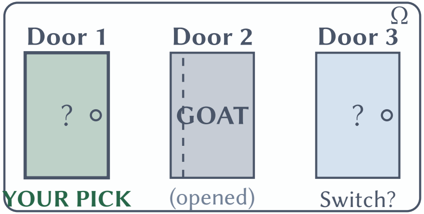
$$\mathbb{P}(D_1) = \frac{1}{3} \qquad \mathbb{P}(D_2 \cup D_3) = \frac{2}{3}$$

You have a 1/3 chance of being right. The money is *more likely* behind one of the other two doors.

The Monty Hall Problem

Step 3: Monty opens a door

Monty Hall, the host, *knows* where the money is. He opens Door 2: a goat.



Monty asks: *"Would you like to switch to Door 3?"*

The question: Has the probability changed? Is $\mathbb{P}(D_3 \mid O) = \frac{1}{3}$ still, or has Monty's action given us new information?

The Monty Hall Problem

The key insight

Key: Monty's choice is *not random*—it's constrained by his knowledge.

- Monty *knows* where the money is
- Monty will *never* open the door with the money
- Monty will *never* open the door you picked

This means his action **reveals information**. We need to update our beliefs.

Let O = "Monty opened door 2." We want to compute:

$$\mathbb{P}(D_3 \mid O) = ?$$

Is this still $\frac{1}{3}$, or has it changed? Let's use Bayes' Rule to find out.

The Monty Hall Problem

Setting up Bayes' Rule

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

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).

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:

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

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

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 was right. The angry PhDs were wrong. Bayes' Rule settles it.

Bayes and Strategic Behavior

When actions reveal information

In Monty Hall, the host's action (opening door 2) **reveals information** because it's constrained by what he knows. This is the foundation of **signaling theory**:

- **Michael Spence (1973)**: Education as a signal of ability. If degrees are *easier* for high-ability workers, employers can use education to update beliefs. (Nobel Prize, 2001)
- **Amotz Zahavi (1975)**: The peacock's tail is *costly*—only fit males can afford it. Costly signals are credible.
- **Diego Gambetta (2009)**: Criminals use tattoos and rituals as costly signals of commitment. (*Codes of the Underworld*)

In game theory, we solve for **Perfect Bayesian Equilibrium**: strategies are optimal given beliefs, and beliefs are updated via Bayes' Rule. David Lewis discovered signaling in his 1969 dissertation—before Spence.

Bayes' Rule in Action

Medical testing example (we'll work through this together)

Classic application: A medical test for a rare disease.

- Tests have **sensitivity** (true positive rate) and **specificity** (true negative rate)
- But what we really want is: given a positive test, what's $\mathbb{P}(\text{disease} \mid \text{positive})$?
- This requires *flipping* the conditional—exactly what Bayes' Rule does

Key ingredients:

- The **prior**: How common is the disease? (base rate)
- The **likelihoods**: How good is the test at detecting disease/non-disease?

Let's work through a specific example on the board...

Base rates matter. When people ignore priors, this is called the "base rate fallacy."

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?