02-680 Module 12

Essentials of Mathematics and Statistics

Aidan Jan

November 6, 2025

Introduction to Probability

Construction of a Probability Space

Objective:

- Define a mathematical framework to describe **random outcomes**
- Enable automated reasoning **under uncertainty**

Example: Coin Toss

- A single toss $\rightarrow$unpredictable outcome
- Many tosses $\rightarrow$observable pattern in average results

Use mathematics to describe and analyze patterns in random events.

Frequentist vs. Bayesian Interpretations of Probability

Frequentist Interpretation

- Probability is the **long-run frequency** of an event occurring over repeated trials.
- Defined in the limit as the number of trials goes to infinity.
- Example: "The probability of heads is 0.5" means that in a large number of coin tosses, about half will be heads.

Bayesian Interpretation

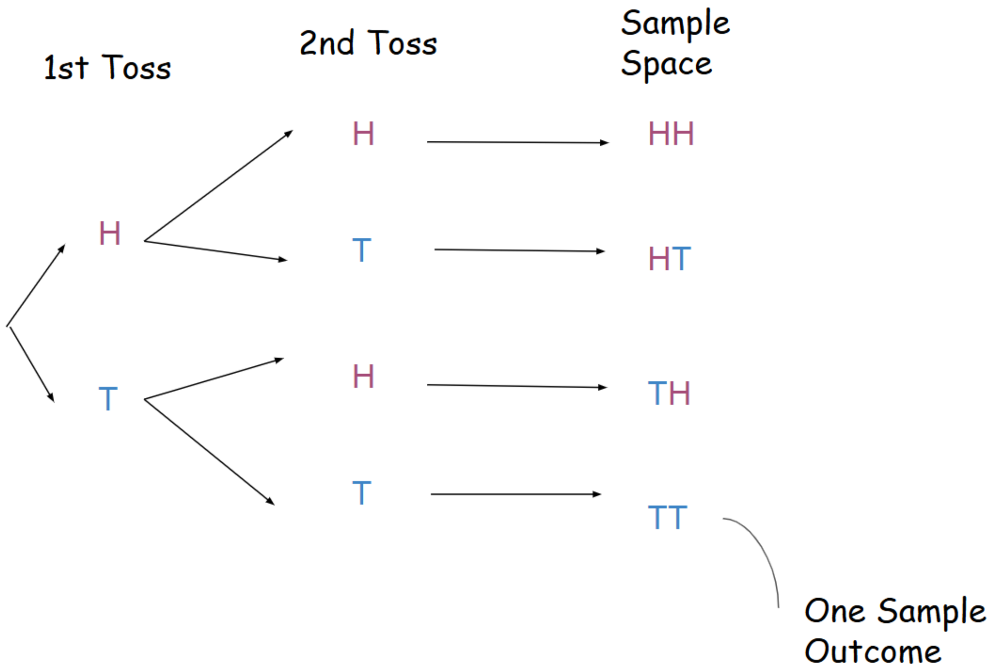
- Probability reflects a **degree of belief** or **uncertainty** about an event.
- Based on **subjective knowledge** it can be **updated** as new information becomes available.
- Example: "I am 80% confident it will rain tomorrow" (based on current data + personal judgment)
- Also called **subjective probability**

Sample Space

The sample space (Ω), sometimes called the universe, is the set of all possible mutually exclusive outcomes of an event.

$$\Omega_{coin} = \{\text{Heads, Tails}\}$$

$$\Omega_{twocoin} = \Omega_{coin} \times \Omega_{coin} = \{\langle\text{Heads, Heads}\rangle, \langle\text{Heads, Tails}\rangle, \langle\text{Tails, Heads}\rangle, \langle\text{Tails, Tails}\rangle\}$$



Event

A event is a subset of the sample space with some condition. For instance if the event is that the first of two coin tosses is heads.

$$A_{firstheads} = \{\langle \text{Heads, Heads} \rangle, \langle \text{Heads, Tails} \rangle\} \subset \Omega_{twocoin}$$

We say the probability $p(A)$ of an event A is the fraction of all outcomes that A covers. Therefore, in the example above, $p(A_{firstheads}) = \frac{1}{2}$.

Event Space

An **event space**, A, is the set of all possible events. For discrete events (such as coin flips) this can be thought of typically as the power set of Ω.

In continuous spaces it is typically thought of as the Borel field of Ω (details of this are beyond the scope of the class for now).

Another way of saying this: consider non-empty Ω and A:

1. $A \in \mathcal{A} \rightarrow \bar{A} \in \mathcal{A}$
2. $A_1, A_2, \ldots, \in \mathcal{A} \rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$

Example: Coin Toss

- Sample space (Ω): {Heads, Tails}
- Event space ($\mathcal{A}$): All subsets of Ω,
  - such as: $\mathcal{A} = \{\emptyset, \{\text{Heads}\}, \{\text{Tails}\}, \{\text{Heads, Tails}\}\}$

0.0.1 Example: Rolling a Die

- Sample space (Ω): $\{1, 2, 3, 4, 5, 6\}$
- Event space ($\mathcal{A}$): All subsets of Ω,
  - such as: $\mathcal{A} = \{\emptyset, \{1\}, \{2, 4\}, \{1, 2, 3\}, \{1, 3, 5\}, \Omega, \dots\}$

The Three Axioms

Axiom 1: Positive Probability

For any event A, $p(A) \geq 0$.

That is, we cannot have a negative probability. In the previous example, if $A_{threetails}$ is the event that you get 3 tails when a coin is tossed twice, $p(A_{threetails}) = 0$, but this can't be negative.

Axiom 2: Total Probability

For any sample space, Ω, $p(\Omega) = 1$. That is, the probability of something in the sample space happening is 1. So for the example above, the probability of a single coin flip being either Heads or Tails is 1, (as defined), and there are no other possible outcomes.

$$p(\Omega) = p(\text{Head}) + p(\text{Tail}) = 1$$

Axiom 3: Disjoint Event Space

For disjoint event spaces $A_1, A_2, \dots, A_n$,

$$p(A_1 \cup A_2 \cup \cdots \cup A_n) = p(A_1) + p(A_2) + \cdots + p(A_n)$$

So in a single coin flip if A_{heads} and A_{tails} are the events of a single coin flip being heads and tails respectively, we can see the event spaces are disjoint (they don't share any outcomes).

So,

$$p(A_{\text{heads}} \cup A_{\text{tails}}) = p(A_{\text{heads}}) + p(A_{\text{tails}}) = \frac{1}{2} + \frac{1}{2} = 1$$

As a counter example, let's define

$$A_{lastheads} = \{\langle \text{Heads, Heads} \rangle, \langle \text{Tails, Heads} \rangle\} \subset \Omega_{twocoin}$$

We can see that

$$p(A_{firstheads} \cup A_{lastheads}) \neq p(A_{firstheads}) + p(A_{lastheads})$$

because $A_{firstheads} \cap A_{lastheads} = \{\langle \text{Heads, Heads} \rangle\} \neq \emptyset$, thus they are not disjoint.

Axioms Summary

- Axiom 1: $P(A) \geq 0 \forall A$
- Axiom 2: $P(\Omega) = 1$
- Axiom 3: $P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + \cdots + P(A_n)$ for disjoint sets.

Perspectives

For a coin toss effect, there are two methods for explaining P(Heads) = 1/2.

- Frequentists: You flip a coin 100 times, you get Heads approximately 50 of those tosses. Frequentists are very objective in the explanation of things.
- Bayesian: You believe you will get tails 50% of the times you flip the coin. Bayesian logic is subject to its interpretation.

Continuous Random Variables

A continuous random variable can take any value in an interval of real numbers. The sample space Ω is uncountably infinite.

- Discrete: $\Omega = \{1, 2, 3, 4, 5, 6, \dots\} = \mathbb{N}$. $\mathcal{A} = \mathcal{P}(\Omega)$
- Continuous: $\Omega = [0, 1] = \mathbb{R}$. $\mathcal{A} = \mathcal{B}(\Omega)$
- in this case, $\mathcal{P}(\Omega)$ is a power set, and $\mathcal{B}(\Omega)$ is a Borel field.

For **discrete** sample spaces, every subset is measurable. The event space is the power set of Ω.

- For uncountable sets (like [0, 1], the power set includes non-measurable sets.)
- We use the **Borel field** for **continuous** random variables.

Continuous Sample Space

Suppose $\Omega = [0, 1]$. This means that

- The random outcome could be any **real number** between 0 and 1.
- There are infinitely many possible values - uncountably many!
- We cannot assign positive probability to a single value.
  - $P(X = 0.5) = 0$
  - $P(0.3 \le X \le 0.4) > 0$