What Is Probability?

Probability is just a way to measure **how likely something is to happen**. It's like saying:

- "There's a 50% chance of rain" → Not certain, but possible.
- "I always get heads when I flip this coin" → That would be surprising, since most coins are fair.

We use probability to make smart guesses in games, weather, science, and even in daily life.

Counting: The First Step in Probability

Before we talk about chances, we need to **count possibilities**.

Example: Ice Cream Choices

Imagine you're at an ice cream shop:

- 3 flavors: chocolate, vanilla, strawberry
- 2 toppings: sprinkles, nuts

How many different combos can you make?

Multiply: 3 flavors × 2 toppings = **6 total combinations**.

This is called the **Multiplication Rule**: if you make several choices one after another, multiply the number of options at each step.

When All Outcomes Are Equally Likely (Naive Probability)

Sometimes, every outcome has the **same chance** — like flipping a fair coin or rolling a fair die.

Probability = (Number of ways you want) / (Total number of possible outcomes)

Example: Rolling a Die

What's the chance of rolling an even number?

- Even numbers: 2, 4, 6 → 3 outcomes
- Total outcomes: 1, 2, 3, 4, 5, 6 → 6 outcomes

Probability = 3 ÷ 6 = 1/2 or 50%

Thinking Conditionally: "Given That..."

Sometimes, new information changes the odds.

> "What's the chance it rains **given that** the sky is cloudy?"

This is **conditional probability**.

Example:

- What's the chance you pass the test **if** you studied?
- What's the chance you're late **if** there's traffic?

New info → new probability.

Independent vs. Dependent Events

- **Independent:** One event doesn't affect the other.
  > Flipping a coin twice: getting heads on the first flip doesn't change the second.
- **Dependent:** One event affects the next.
  > Drawing cards from a deck without putting them back: once you draw the Ace of Spades, you can't draw it again.

Combining Events: AND, OR, NOT

Let's say:
- A = it rains
- B = you bring an umbrella

- **A and B happen** → both rain and umbrella
- **A or B happens** → either rain, or umbrella, or both
- Not **A** → it does **not** rain

There's a trick (called **De Morgan's Law**) that helps with "not":
> Not (A or B) = Not A **and** Not B
> Not (A and B) = Not A **or** Not B

Like saying: "I didn't eat cake or ice cream" means "I didn't eat cake **and** I didn't eat ice cream."

Bayes' Rule: Updating Beliefs with New Info

Imagine you think a coin is fair. But then you flip it 10 times and get heads every time.

Should you still think it's fair? Probably not!

Bayes' Rule helps you **update your beliefs** when you see new evidence.

> Start with a guess (prior) → see data → update your guess (posterior)

Used in medical tests, spam filters, and detective work!

Random Variables: Turning Outcomes into Numbers

A **random variable** is just a number that depends on chance.

Examples:
- Number of heads in 3 coin flips → could be 0, 1, 2, or 3
- Time until your next text message → could be 1 minute, 10 minutes, etc.

We give these variables names like **X** or **Y**, and study:
- What values they can take
- How likely each value is

Two Types of Random Variables

1. **Discrete** – counts things (whole numbers)
   - Example: Number of goals in a soccer game
2. **Continuous** – measures things (any number, including decimals)
   - Example: Height of a person, time to finish a race

The Two Main Tools: PMF and CDF

These are like instruction manuals for random variables.

> **PMF (Probability Mass Function) – for discrete variables**
> Tells you the chance of each specific outcome.
> Like: "Chance of 2 heads = 37.5%"

> **CDF (Cumulative Distribution Function) – for any variable**
> Tells you the chance the result is **less than or equal to** a number.
> Like: "Chance of scoring 5 or fewer points = 60%"

Expected Value: The Long-Term Average

Think of this as the **average result** if you repeated something many times.

Example:
- You play a game where you win $10 half the time and $0 the other half.
- Expected value = (½ × $10) + (½ × $0) = **$5 per game**

You won't win $5 every time, but over 100 games, you'd expect to win about $500.

> *It's not a guarantee — it's a long-term average.* <

Indicator Variables: Yes-or-No Questions

These are super useful tools: a variable that is:
- **1** if something happens
- **0** if it doesn't

Example:
- I = 1 if it rains today, 0 if it doesn't
- Then the **average of I** is just the **probability it rains**

This is called the **Fundamental Bridge**:
> **Average of indicator = Probability of the event**

Variance: How Spread Out the Results Are

Expected value tells you the center. **Variance** tells you how much things vary.

- Low variance → results are close to average
- High variance → results jump around

Example:
- Two jobs pay $50/day on average.
  - Job A: always pays $50 → low variance
  - Job B: pays $0 or $100 randomly → high variance

You make the same on average, but Job B is riskier.

<u>**Continuous Random Variables (Like Time or Weight)**</u>
These can take **any value** in a range — not just whole numbers.

Examples:
- Time until your phone battery dies
- Amount of sugar in a soda

Instead of listing probabilities for each value (impossible!), we use:
- **PDF (Probability Density Function)** – shows how likely different values are
- **Area under the curve = probability**

You don't ask, "What's P(time = 2.000000)?"
You ask, "What's P(time between 2 and 3)?" → that's an area.

<u>**The Universality of the Uniform (UoU)**</u>
This is a cool trick:
- If you take **any continuous random process** (like height, weight, time), plug it into its own CDF, you get a **Uniform(0,1)** number.

And vice versa:
- If you take a random number between 0 and 1 and plug it into the inverse CDF, you get a value from your original distribution.

This is how computers **simulate random data** from any distribution.

<u>**Moment Generating Functions (MGFs): The "DNA" of a Distribution**</u>
Think of this as a **fingerprint** for a probability distribution.

- It can tell you the mean, variance, and more.
- If two distributions have the same MGF, they are the **same type**.
- It also makes it easier to find the distribution of sums of random variables.

$ Don't worry about the math — just know it's a powerful tool experts use.

<u>**Joint, Marginal, and Conditional Distributions**</u>
When you have **two things happening at once**, like:
- Height and weight
- Rain and traffic

You can ask:
- **Joint:** What's the chance both happen?
- **Marginal:** What's the chance of just one, ignoring the other?
- **Conditional:** What's the chance of one **given** the other?

Covariance and Correlation: Do Two Things Move Together?

> **Covariance** = how two variables change together
  - *Positive:* both go up or down together (e.g., height and weight)
  - *Negative:* one goes up, the other down (e.g., exercise and body fat)
  - *Zero:* no pattern

> **Correlation** = a "standardized" version of covariance
  - Always between -1 and 1
  - 1 = perfect positive relationship
  - 0 = no linear relationship
  - -1 = perfect negative relationship

Independent → correlation is 0, but **zero correlation doesn't mean independent**!

Transformations: Changing Variables

Sometimes you want to know the distribution of a **function** of a random variable.

Example:
- X = time until a phone dies (Exponential)
- Y = 2X = twice that time → still Exponential, but scaled

There's a method (using the **Jacobian**) to find the new distribution when you change variables — useful in advanced stats.

Convolutions: Adding Random Variables

If you want to know the distribution of **X + Y**, like:
- Total time of two phone batteries
- Combined score of two quiz parts

You use a **convolution** — a special kind of integral or sum.

"***For independent normal variables, the sum is also normal!***"

Poisson Process: Counting Rare Events Over Time

Used for things that happen randomly over time:
- Emails arriving
- Customers entering a store
- Earthquakes

Key ideas:
- Events happen at a steady average rate (e.g., 2 emails per hour).
- Time between events is **Exponential**.
- Number of events in a time period is **Poisson**.

And the process has **no memory**: just because you waited a long time doesn't mean the next event is "due."

<u>**Order Statistics: Ranking Random Values**</u>

If you take 5 random people and line them up by height:
- The shortest is the **1st order statistic**
- The tallest is the **5th**

Used in competitions, quality control, and data analysis.

Fun fact: If you take random numbers between 0 and 1 and pick the 3rd smallest, it follows a **Beta distribution**.

<u>**Conditional Expectation: Average Given Some Info**</u>

Instead of asking, "What's the average test score?"
You ask, "What's the average score **given** the student studied?"

This is **conditional expectation**.

And there's a powerful rule:
> **Average of (Average given X) = Overall Average**

Called **Adam's Law** (Law of Total Expectation).

There's also **Eve's Law** (Law of Total Variance):
> **Total variability = average of within-group variability + variability between group averages**

Super useful in data science and machine learning.

<u>**Law of Large Numbers (LLN)**</u>

If you flip a fair coin many times:
- The **proportion of heads** will get closer and closer to 50%

The more trials, the closer the average gets to the true expected value.

It's why casinos always win in the long run.

<u>**Central Limit Theorem (CLT): Why Averages Are Normal**</u>

Even if individual things are **not** normally distributed, the **average** of many of them **is** approximately normal.

Example:
- One person's income might be skewed
- But the average income of 100 people is roughly normal

This is why the **bell curve** shows up everywhere — in test scores, heights, errors, etc.

CLT is one of the most important ideas in statistics.

Markov Chains: Random Walks with Memory Rules

Imagine a robot moving between rooms. At each step, it picks the next room based only on **where it is now**, not where it's been.

That's a **Markov chain**.

Used in:
- Google's PageRank algorithm
- Predicting weather
- Game AI

Key ideas:
- **Transition matrix:** chance of going from one state to another
- **Stationary distribution:** long-term probability of being in each state
- **Reversible:** going forward or backward looks the same

Common Probability Distributions (The Usual Suspects)

DISTRIBUTION	WHEN TO USE	EXAMPLE
Bernoulli	One yes/no trial	Coin flip
Binomial	Number of successes in fixed trials	10 coin flips, how many heads?
Geometric	Number of failures before first success	How many lottery tickets until you win?
Poisson	Rare events over time/space	Emails per hour
Normal	Many natural phenomena	Heights, test scores
Exponential	Time between events	Time until next bus
Uniform	All outcomes equally likely	Random number between 1 and 10
Gamma	Total time for several events	Time to see 5 shooting stars
Beta	Probabilities themselves (Bayesian)	Estimating fairness of a coin

---

Common Mistakes (Biohazards!)

1. Assuming everything is equally likely – Not all combinations are equally probable!
2. Confusing conditional and joint probability – "Given that" is not the same as "and"
3. Assuming independence – Just because two things seem unrelated doesn't mean they are
4. Pulling functions out of averages – $E(X^2) \neq (E(X))^2$
5. Forgetting to check if answers make sense – Probabilities can't be over 1 or negative!

Problem-Solving Tips
1. Define your variables and events clearly
2. Draw/Plot a picture/graph or make up numbers to understand the setup
3. **Try small cases** (e.g., 2 people instead of 100)
4. **Use symmetry** – sometimes all outcomes are equally likely
5. **Break big problems into smaller ones** (use indicator variables!)
6. **Use known distributions** – does this story match Binomial? Poisson?
7. **Check your answer** – does it make sense? What if n=1? n=∞?

Take Aways
- **Start simple** – use examples and stories
- **Think in terms of real-world situations**
- **Practice with small numbers**
- Trust your intuition, but verify with logic
- **Revisit the cheat-sheet for reference**, and textbook for exam preparation
---
Summary: The Big Ideas
1. Probability = long-term frequency of random events
2. Counting is the foundation
3. Expectation = average outcome
4. Variance = spread
5. Independence matters
6. Bayes' Rule updates beliefs
7. LLN and CLT explain why averages behave nicely
8. Distributions model real-world randomness
9. Conditional thinking is powerful
10. Always check your work!