Random Variables

From Outcomes to Numbers

Scott Cunningham

Harvard University
Department of Government

February 2, 2026

Where We Are

Week 1: Probability → **Week 2: Random Variables** → Week 3: Expectations

Last week: Sample spaces, events, probability axioms, conditional probability, Bayes' rule

This week: Random variables — the bridge from events to **numbers**

Why? Mathematics works with numbers. To use calculus, linear algebra, and optimization, we need to translate qualitative outcomes into quantitative values.

What Is a Random Variable?

A function that assigns numbers to outcomes

The Pollster's Problem

A story to set the stage

It's October 2024. You're a pollster trying to predict the presidential election.

You call a randomly selected voter. *Before* they answer, several things are uncertain:

- Will they answer the phone? (Yes/No)
- If they answer, who do they support? (Harris/Trump/Other/Undecided)
- How certain are they? (Very/Somewhat/Not at all)
- Will they actually vote? (Definitely/Probably/Probably not/Definitely not)

These are all **qualitative** outcomes.

But you need to produce a **number**: "Harris 48%, Trump 47%"

From Outcomes to Numbers

The translation problem

To do statistics, we need to convert qualitative outcomes to numbers.

Outcome	Possible Encoding
Supports Harris	$X = 1$
Supports Trump	$X = 0$
Definitely will vote	$Y = 1$
Probably will vote	$Y = 0.75$
Probably won't vote	$Y = 0.25$
Definitely won't vote	$Y = 0$

The function that does this translation is called a **random variable**.

Definition: Random Variable

Random Variable

A **random variable** is a function $X : \Omega \to \mathbb{R}$ that assigns a real number to each outcome in the sample space.

Key insight: A random variable is *neither random nor a variable.* It's a **function**.

- The *input* is an outcome $\omega \in \Omega$ (which outcome occurs)
- The *output* is a number $X(\omega) \in \mathbb{R}$ (the numerical value)

The "randomness" comes from not knowing which ω will occur.

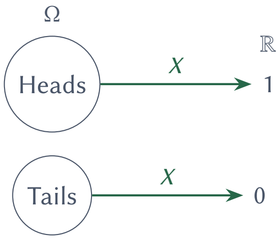
Convention: Capital letters (X, Y, Z) for random variables, lowercase (x, y, z) for specific values.

Example: Coin Flip

The simplest random variable

Sample space: $\Omega = \{\text{Heads}, \text{Tails}\}$

Random variable: $X(\omega) = \begin{cases} 1 & \text{if } \omega = \text{Heads} \\ 0 & \text{if } \omega = \text{Tails} \end{cases}$



Before the flip: X is uncertain (could be 0 or 1). After: X takes a specific value.

Example: Die Roll

A slightly richer case

Sample space: $\Omega = \{1, 2, 3, 4, 5, 6\}$ (outcomes are already numbers!)

Natural random variable: $X(\omega) = \omega$ (the identity function)

But we could define *other* random variables on the same sample space:

ω	1	2	3	4	5	6
$X(\omega)$ = face value	1	2	3	4	5	6
$Y(\omega)$ = is it even?	0	1	0	1	0	1
$Z(\omega)$ = is it ≥ 5?	0	0	0	0	1	1

Key point: Many different random variables can be defined on the same sample space. Each asks a different question about the outcome.

Functions of Random Variables

Building new random variables from old ones

If X is a random variable and $g : \mathbb{R} \to \mathbb{R}$ is any function, then $g(X)$ is also a random variable.

Example: Roll a die, let X = face value.

- X^2 is a random variable (squared face value)
- $2X + 3$ is a random variable (linear transformation)
- $1_{X>3}$ is a random variable (indicator: 1 if $X > 3$, else 0)

How it works: $(g \circ X)(\omega) = g(X(\omega))$

The composition of two functions is still a function.

Part II

Discrete Random Variables

When you can list all possible values

Discrete Random Variables

Definition and examples

Discrete Random Variable

A random variable X is **discrete** if its range (the set of values it can take) is **countable** — either finite or countably infinite.

Examples of discrete random variables:

- Coin flip: $X \in \{0, 1\}$ (finite)
- Die roll: $X \in \{1, 2, 3, 4, 5, 6\}$ (finite)
- Number of voters contacted before finding a Harris supporter (countably infinite)
- Number of children in a household (finite, but the bound is large)

Key feature: You can *list* all possible values, even if the list is infinitely long.

Three Questions You'll Want to Answer

Before the definitions: what are we actually trying to calculate?

When working with random variables, you'll constantly ask three types of questions:

1. **Exact value**: "What's $P(X = x)$?" — probability of a specific value
2. **Cumulative**: "What's $P(X \leq x)$?" — probability up to a threshold
3. **Density**: "How concentrated is probability near x?" (i.e., probability per unit)

But here's the catch: **not every question works for every type of random variable.**

	Exact value $P(X = x)$	**Cumulative** $P(X \leq x)$	**Density**
Discrete	✓ Yes	✓ Yes	× No
Continuous	× Always 0	✓ Yes	✓ Yes

Three Tools to Answer Them

PMF, CDF, and PDF

Each question has a corresponding tool:

Question	Tool	Works for
$P(X = x)$	**PMF** (probability mass function)	Discrete only
$P(X \leq x)$	**CDF** (cumulative distribution function)	Both
Density at x	**PDF** (probability density function)	Continuous only

Examples:

- Die roll: "$P(\text{roll} = 4)$?" $\rightarrow$ Use PMF
- Temperature: "$P(\text{temp} \leq 70)$?" $\rightarrow$ Use CDF
- Temperature: "How dense is probability near 72°?" $\rightarrow$ Use PDF

Keep these questions in mind. Every definition we introduce is a tool to answer them.

Probability Mass Function (PMF)

How probability is distributed across values

For a discrete random variable, we can ask: "What's the probability that X equals exactly x?"

Probability Mass Function

The **PMF** of a discrete random variable X is the function $f(x) = P(X = x)$ that gives the probability that X takes the value x.

The PMF tells you where the "probability mass" (i.e., the chunk of probability assigned to each point) is located — hence the name.

Properties of any PMF

What makes a valid PMF?

Properties of any PMF:

1. $f(x) \geq 0$ for all x (probabilities are non-negative)
2. $\sum_x f(x) = 1$ (probabilities sum to 1)

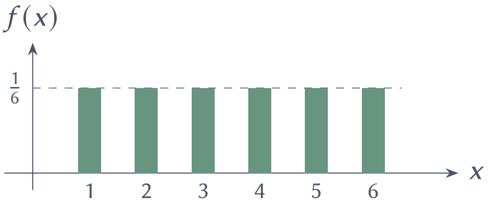
Any function satisfying these two properties is a valid PMF.

These mirror the axioms of probability from last week — just applied to the values of a random variable.

Example: Fair Die

PMF of a uniform discrete distribution

Let X = outcome of rolling a fair six-sided die. **PMF**: $f(x) = \frac{1}{6}$ for $x \in \{1, 2, 3, 4, 5, 6\}$



Check: $f(1) + f(2) + \cdots + f(6) = 6 \times \frac{1}{6} = 1$ ✓

Why must this equal 1? **Law of total probability** from Week 1: $\sum_x P(X = x) = P(\Omega) = 1$.
The PMF values partition the sample space, so they must sum to 1. Always.

The Bernoulli Distribution

The building block of binary outcomes

The simplest non-trivial random variable: two outcomes, one parameter.

Bernoulli Distribution

X has a **Bernoulli distribution** with parameter $p \in [0, 1]$, written $X \sim \text{Bernoulli}(p)$, if:

$$f(x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \end{cases}$$

Compact notation: $f(x) = p^x(1-p)^{1-x}$ for $x \in \{0, 1\}$

Interpretation: $X = 1$ is "success" (probability p); $X = 0$ is "failure" (probability $1 - p$)

Examples: coin flip ($p = 0.5$), voter supports Harris ($p = 0.48$), patient survives surgery.

A Key Insight About Bernoulli

The mean equals the probability

For a Bernoulli random variable, let's compute the expected value:

$$\begin{aligned}
\mathbb{E}[X] &= \sum_x x \cdot P(X = x) \\
&= 0 \cdot P(X = 0) + 1 \cdot P(X = 1) \\
&= 0 \cdot (1 - p) + 1 \cdot p \\
&= p
\end{aligned}$$

The insight: When X is a 0/1 indicator, $\mathbb{E}[X] = P(X = 1)$.

This connects expectations (a summary of distributions) to probabilities (what we learned last week).

Part III

Continuous Random Variables

When values form a continuum

The Problem with Continuous Random Variables

Why PMFs don't work

Consider: X = height of a randomly selected adult in inches.

X can take *any* value in some interval, say $[48, 84]$. There are **uncountably many** possible values.

Problem: What is $P(X = 67.3842719...)$?

The Problem with Continuous Random Variables

The answer

Problem: What is $P(X = 67.3842719...)$?

- There are infinitely many possible heights
- If each had positive probability, they'd sum to > 1
- So each specific value must have probability **zero**

Consequence: For continuous random variables, $P(X = x) = 0$ for every x.

This seems paradoxical: "The probability of any specific outcome is zero, yet some outcome must occur."

Solution: Don't ask about specific values. Ask about *intervals*.

Cumulative Distribution Function (CDF)

The universal description of random variables

Instead of asking "What's $P(X = x)$?" ask "What's $P(X \leq x)$?"

Cumulative Distribution Function

The **CDF** of a random variable X is the function $F(x) = P(X \leq x)$

The CDF works for ALL random variables — discrete, continuous, or mixed.

Properties of any CDF

What makes a valid CDF?

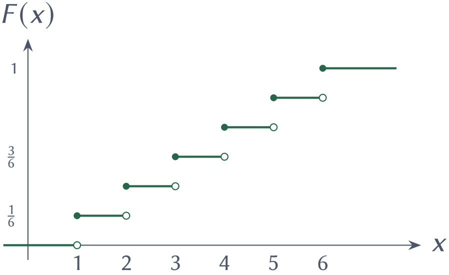
Properties of any CDF:

1. F is **non-decreasing**: if $a < b$, then $F(a) \leq F(b)$
2. $\lim_{x \to -\infty} F(x) = 0; \quad \lim_{x \to \infty} F(x) = 1$
3. $P(X > x) = 1 - F(x); \quad P(a < X \leq b) = F(b) - F(a)$

Property 3 is especially useful: we can compute probabilities for any interval using just the CDF.

Example: CDF of a Fair Die

A discrete CDF is a step function



Reading the CDF: $P(X \leq 3) = F(3) = \frac{1}{2}$ $P(X > 4) = 1 - F(4) = \frac{1}{3}$
$P(2 < X \leq 5) = F(5) - F(2) = \frac{1}{2}$

Continuous Random Variables

Smooth CDFs

Continuous Random Variable

A random variable X is **continuous** if its CDF $F(x)$ can be written as

$$F(x) = \int_{-\infty}^{x} f(u)\, du$$

for some non-negative function f.

The function f is called the probability density function (PDF).

Key Features of Continuous Random Variables

What makes them different from discrete

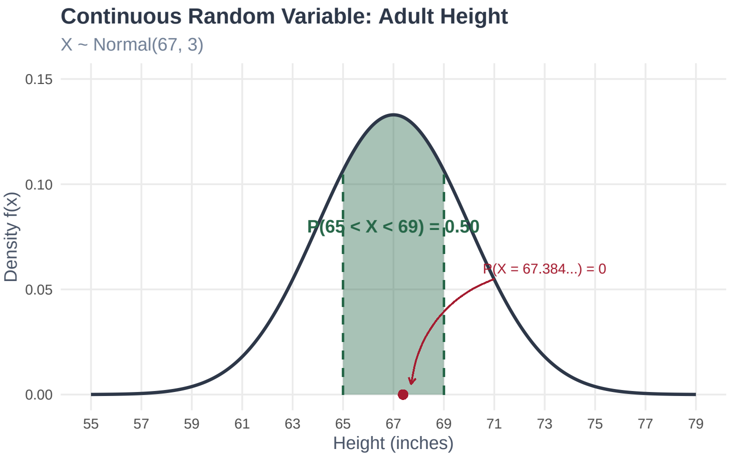
Key features:

- The CDF is smooth (no jumps)
- $P(X = x) = 0$ for any specific value x
- $P(a \leq X \leq b) = P(a < X < b)$ (strict vs. weak inequality doesn't matter)

That last point is important: since $P(X = a) = 0$ and $P(X = b) = 0$, including or excluding the endpoints doesn't change the probability.

Visualizing Continuous Random Variables

Height example: $X \sim \text{Normal}(67, 3)$



`pnorm(69, 67, 3) - pnorm(65, 67, 3)` $= 0.495$ (shaded area)

Probability Density Function (PDF)

The continuous analog of the PMF

Probability Density Function

For a continuous random variable with CDF F, the **PDF** is $f(x) = \frac{dF(x)}{dx}$ when this derivative exists.

Properties of any PDF: (1) $f(x) \geq 0$ for all x (2) $\int_{-\infty}^{\infty} f(x)\, dx = 1$

Critical warning: $f(x)$ is NOT a probability!

- $f(x)$ can be greater than 1 — it's a *density* (i.e., probability per unit of x), not a probability
- Probabilities come from *integrating*: $P(a \leq X \leq b) = \int_a^b f(x)\, dx$

Thinking About Density

A physical analogy

Physical density: You have exactly 1 kg of butter. How is it spread?

- Spread over 1 meter of bread → density = 1 kg/m   (thin layer)
- Spread over 0.5 meters of bread → density = 2 kg/m   (thick layer)
- Same total butter, different concentration

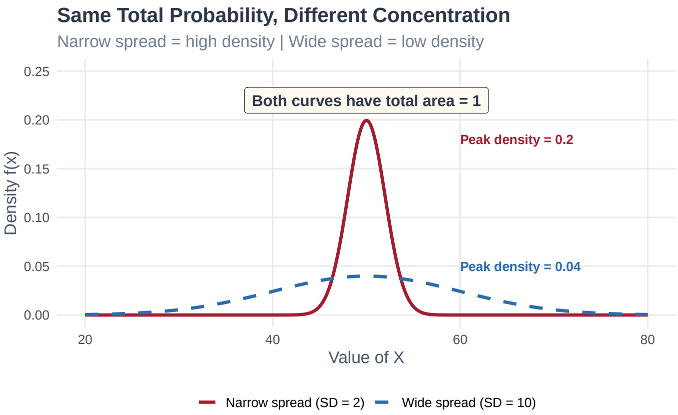
Probability density: You have exactly **1 unit of probability**. How is it spread?

- Spread over a wide range → low density (flat curve)
- Concentrated in a narrow range → high density (tall, narrow curve)
- **Total area under the curve always equals 1**

Density can exceed 1 (like 2 kg/m of butter) because it's "amount per unit," not the total amount.

Seeing the Tradeoff

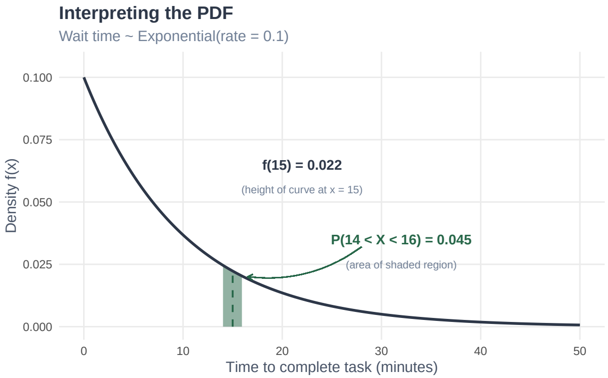
Same total probability, different concentration



The narrow curve is 5× **taller** because probability is packed into a region 5× **smaller**. Both curves contain exactly 1 unit of probability — they just distribute it differently.

Interpreting the PDF

What does the height of the curve mean?



Plain English: The PDF tells you how “crowded” probability is near each value.

- **Height** $f(x)$ = probability per unit (not a probability itself)
- **Area** under curve = actual probability

A Quick Note on Terminology

"Area under the curve" means different things in different contexts

When we say "area under the curve" here, we mean area under the **PDF**.

That area equals probability. That's what we're talking about.

However: You will hear (or may have already heard) people in machine learning talk about "area under the curve" or "AUC" when discussing classification models.

That's something completely different — it refers to the **ROC curve**, which measures predictive accuracy. We're not doing classification here, so that's not what we mean.

Just flagging this so you don't get confused when the same phrase shows up in different contexts.

Support of a Random Variable

Where can X actually land?

Support

The **support** of a random variable X is the set of values where the PMF or PDF is positive: $\text{Supp}[X] = \{x \in \mathbb{R} : f(x) > 0\}$

Interpretation: The support tells you which values X can actually take.

In plain English: "What values are even possible?"

Support: Die Roll Example

What's possible vs. what's impossible

For a fair six-sided die, the support is $\{1, 2, 3, 4, 5, 6\}$.

- $X = 1$ **is in the support** — this value is possible
- $X = 7$ **is not in the support** — a six-sided die can't land on 7
- $X = 3.5$ **is not in the support** — you can't roll "three and a half"

For a die, this isn't terribly interesting — the support is obvious from the context.

But the concept becomes important when we introduce famous distributions:

- Uniform$(0, 1)$: support is $[0, 1]$ — can't be negative or exceed 1
- Exponential: support is $[0, \infty)$ — can't be negative
- Normal: support is $(-\infty, \infty)$ — any real number is possible

The Uniform Distribution

The simplest continuous distribution

Uniform Distribution

X has a **Uniform distribution** on $[a, b]$, written $X \sim \text{Uniform}(a, b)$, if:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

Support: $[a, b]$ — the random variable can only land between a and b. Outside this interval, the PDF $f(x) = 0$.

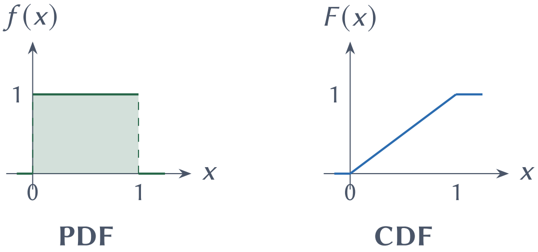
The standard uniform: $X \sim \text{Uniform}(0, 1)$

- PDF: $f(x) = 1$ for $x \in [0, 1]$, and 0 otherwise
- **Intuition**: Every value in $[0, 1]$ is equally likely

"Equally likely" means equal density, not equal probability (always zero for continuous RVs).

Visualizing the Uniform(0,1)

PDF and CDF



Probability as area:

$P(0.2 \leq X \leq 0.5) = \int_{0.2}^{0.5} 1\, dx = 0.5 - 0.2 = 0.3$

Or using the CDF: $F(0.5) - F(0.2) = 0.5 - 0.2 = 0.3$ ✓

Expected Value

The center of mass of a distribution

Summarizing a Distribution

Why we need summary statistics

A PMF or PDF gives us complete information about a random variable.

But sometimes we want a **single number** that summarizes the distribution:

- "What's the average income in this population?"
- "What's the typical approval rating?"
- "What should I predict for the next observation?"

The most common summary: the **expected value** (or *mean*).

Intuition: If you could repeat the random process infinitely many times and average the results, what would you get?

Texas Hold'em

Sequential decision-making under uncertainty

The setup: 52-card deck. Each player gets 2 private "hole cards."

The betting rounds:

1. **Pre-flop**: You see only your 2 hole cards. 50 cards unknown. Bet.
2. **Flop**: 3 community cards revealed (anyone can use). 47 unknown. Bet.
3. **Turn**: 4th community card. 46 unknown. Bet.
4. **River**: 5th and final community card. 45 unknown. Bet.

The goal: Make the best 5-card hand from your 7 available cards.

No-limit: You can bet any amount up to "all-in" (everything you have).

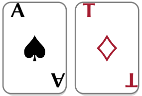
The skill? Maximize **expected value** of each decision. You'll lose individual hands, but win in the long run if you consistently take +EV actions.

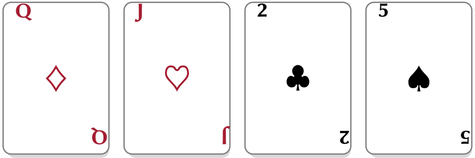
Scenario 1: The Straight Draw

Single card probability (PMF)

You're playing no-limit Texas Hold'em. You hold A♠T:

The Board (through the turn)

Your hand: 



You have A-T-J-Q. One card to come. A **King** gives you the **nut straight**.

Opponent goes **all-in** for $200 into a $100 pot.

Scenario 1: Computing P(King)

What are our outs?

Cards that help: Any King (4 in the deck)

Cards we've seen: 2 hole + 4 board = 6 cards

Unknown cards: 52 − 6 = 46

$$P(\text{King on river}) = \frac{4}{46} \approx 8.7\%$$

This is a **PMF calculation**: $P(X = \text{King})$ where X is the river card.

Scenario 1: Expected Value

Should we call \$200?

If we win: We get the \$100 pot + opponent's \$200 = \$300 profit

If we lose: We lose our \$200 call

Outcome	Probability	Profit
Win (hit King)	0.087	+\$300
Lose (miss)	0.913	−\$200

$$\mathbb{E}[X] = (0.087)(300) + (0.913)(-200) = 26.10 - 182.60 = -\$156.50$$

FOLD. This is a losing call.

More poker examples covering either/or and joint probability are in the appendix.

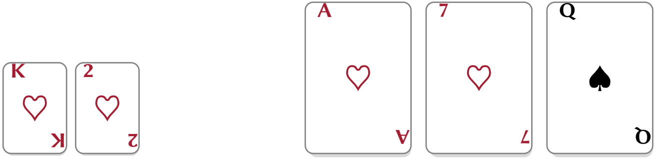
Scenario 2: The Flush Draw

Either/Or probability (two chances)

New hand. You hold K♡2♡:

The Flop

Your hand:



You have **four hearts**. One more gives you the **nut flush** (A-K high).

Opponent goes **all-in** for $100 into a $100 pot. You get to see *two* more cards.

Scenario 2: Two Cards to Come

P(heart on turn OR river)

You need a heart on the turn **or** the river (or both).

Hearts remaining: $13 - 4 = 9$ outs **Unknown cards**: $52 - 5 = 47$

Easier: Compute P(miss both), then subtract from 1.

$$P(\text{miss turn}) = \frac{38}{47}$$

$$P(\text{miss river} \mid \text{miss turn}) = \frac{37}{46} \quad (onenon - heartrevealed)$$

$$P(\text{miss both}) = \frac{38}{47} \times \frac{37}{46} \approx 65.0\%$$

$P(\text{heart on turn OR river}) = 1 - 0.65 = 35.0\%$

Scenario 2: Expected Value

Should we call \$100?

If we win: \$100 pot + \$100 bet = \$200 profit

If we lose: We lose our \$100 call

Outcome	Probability	Profit
Win (make flush)	0.35	+\$200
Lose (miss)	0.65	−\$100

$$\mathbb{E}[X] = (0.35)(200) + (0.65)(-100) = 70 - 65 = +\$5$$

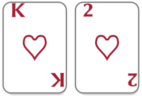
CALL. This is a winning play (in expectation).

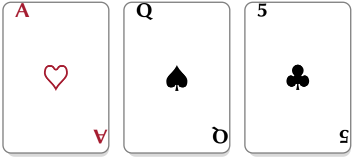
Scenario 3: Runner-Runner

Joint probability (both/and)

New hand. You hold K♡2♡ again, but the flop is different:

The Flop

Your hand: 



Now you only have **three hearts**. You need hearts on **both** the turn **and** river.

Opponent goes **all-in** for $200 into a $100 pot.

Scenario 3: P(heart on turn AND river)

The multiplication rule

This is **joint probability**: we need both events to occur.

Hearts remaining: $13 - 3 = 10$ **Unknown cards**: $52 - 5 = 47$

$$P(\text{heart on turn}) = \frac{10}{47}$$

$$P(\text{heart on river} \mid \text{heart on turn}) = \frac{9}{46} \quad \textbf{(one heart used up!)}$$

$$P(\text{both}) = \frac{10}{47} \times \frac{9}{46} = \frac{90}{2162} \approx 4.2\%$$

Notice: after using a heart on the turn, only **9 hearts remain** in 46 cards.

Scenario 3: Expected Value

Should we call \$200?

If we win: \$100 pot + \$200 bet = \$300 profit

If we lose: We lose our \$200 call

Outcome	**Probability**	**Profit**
Win (runner-runner flush)	0.042	+\$300
Lose (miss)	0.958	−\$200

$$\mathbb{E}[X] = (0.042)(300) + (0.958)(-200) = 12.60 - 191.60 = -\$179$$

FOLD. Runner-runner is almost never worth chasing.

The Puzzle

When *would* runner-runner be worth it?

We showed that calling \$200 to win \$300 with 4.2% odds is $-$EV.

Question: How big would the pot need to be to justify calling?

Set $\mathbb{E}[X] = 0$ and solve for the pot:

$$0 = P(\text{win}) \times \text{Pot} - P(\text{lose}) \times \text{Call}$$

$$0 = (0.042)(\text{Pot}) - (0.958)(200)$$

$$\text{Pot} = \frac{(0.958)(200)}{0.042} = \$4,562$$

You'd need a pot of **\$4,562** to justify a \$200 call — a ratio of **23:1**.

This is why "chasing" low-probability draws loses money. The math doesn't lie.

Flipping the Script

Using EV offensively

Now suppose *you* have the strong hand and suspect your opponent is on a draw.

The insight: If they need 23:1 odds to call profitably, but you only offer them 3:1...

Bet Sizing as a Weapon

By betting **large**, you force your opponent to either:

- **Fold** — you win the pot immediately
- **Call incorrectly** — you profit from their −EV decision

Either way, you win. The math works in both directions.

This is why professional poker isn't just about having good cards — it's about understanding how **expected value** shapes every decision, for you *and* your opponent.

Three Types of Probability Questions

What we just learned

Scenario	Probability Type	Formula
Straight (one card)	PMF	$P(X = x)$
Flush (turn or river)	Union (either/or)	$1 - P(\text{miss both})$
Runner-runner	Joint (both/and)	$P(A) \times P(B \mid A)$

Each maps to a **decision**:

- 8.7% with bad odds $\Rightarrow$ **Fold**
- 35% with good odds $\Rightarrow$ **Call**
- 4.2% with any reasonable odds $\Rightarrow$ **Fold**

Expected value turns probability into action.

From Poker to Theory

What we just computed

What did we actually calculate?

$$\mathbb{E}[X] = \sum_x x \cdot P(X = x) = (\text{profit if win}) \cdot P(\text{win}) + (\text{loss if lose}) \cdot P(\text{lose})$$

This is exactly the **expected value** formula:

- List each possible value of X (profit outcomes)
- Multiply each value by its probability
- Sum them up

The poker examples are random variables with two outcomes: $X : \Omega \to \mathbb{R}$, where $\Omega = \{\text{win}, \text{lose}\}$

Let's now write down the formal definition...

Definition: Expected Value

Expected Value

The **expected value** (or **mean**) of a random variable X is:

$$\mathbb{E}[X] = \begin{cases} \sum_x x \cdot f(x) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} x \cdot f(x)\, dx & \text{if } X \text{ is continuous} \end{cases}$$

Interpretation: The expected value is a **weighted average** of possible values, where the weights are the probabilities (or densities).

Notation: $\mathbb{E}[X]$, μ, μ_X all denote the same thing.

The expected value need not be a possible value of X. The expected value of a fair die roll is 3.5, which you can never actually roll.

Example: Fair Die

Computing the expected value

Let X = outcome of a fair die roll. PMF: $f(x) = \frac{1}{6}$ for $x \in \{1, 2, 3, 4, 5, 6\}$.

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{x=1}^{6} x \cdot f(x) \\
&= 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} \\
&= \frac{1+2+3+4+5+6}{6} = \frac{21}{6} = 3.5
\end{aligned}$$

Interpretation: If you rolled a die many times and averaged the outcomes, you'd get approximately 3.5.

Example: Bernoulli

The mean equals the probability

Let $X \sim \text{Bernoulli}(p)$. PMF: $f(1) = p, f(0) = 1 - p$.

$$\mathbb{E}[X] = \sum_{x \in \{0,1\}} x \cdot f(x) = 0 \cdot (1 - p) + 1 \cdot p = p$$

This confirms our earlier insight: For a 0/1 indicator variable, $\mathbb{E}[X] = P(X = 1)$

This is why sample proportions (averages of 0/1 indicators) estimate population proportions.

The Power of Indicator Variables

Turning counting problems into expectations

Since $\mathbb{E}[D] = P(D = 1)$ for indicators, we can compute probabilities via expectations.

Example: Expected number of heads in n coin flips (each with $P(\text{heads}) = p$)?

Let $D_i = \begin{cases} 1 & \text{if flip } i \text{ is heads} \\ 0 & \text{otherwise} \end{cases}$ and $X = \sum_{i=1}^{n} D_i$ = total heads

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{n} D_i\right] = \sum_{i=1}^{n} \mathbb{E}[D_i] = \sum_{i=1}^{n} p = np$$

Key insight: $\mathbb{E}[\sum D_i] = \sum \mathbb{E}[D_i]$ — the expectation of a sum is the sum of expectations.

This property (**linearity of expectation**) lets us skip finding the distribution of X.

Example: Uniform(0,1)

Expected value of a continuous random variable

Let $X \sim \text{Uniform}(0, 1)$. PDF: $f(x) = 1$ for $x \in [0, 1]$.

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot f(x)\, dx = \int_0^1 x \cdot 1\, dx$$

$$= \left[\frac{x^2}{2}\right]_0^1 = \frac{1}{2} - 0 = \frac{1}{2}$$

Intuition: The integral $\int x \cdot f(x)\, dx$ is the continuous version of $\sum x \cdot p(x)$ — outcomes weighted by how likely they are. Here, the distribution is symmetric around $\frac{1}{2}$, so the mean lands at the center.

More generally, for $X \sim \text{Uniform}(a, b)$: $\mathbb{E}[X] = \frac{a+b}{2}$

Law of the Unconscious Statistician (LOTUS)

Expected value of a function of X

What if we want $\mathbb{E}[g(X)]$ for some function g?

LOTUS

For any function g:

$$\mathbb{E}[g(X)] = \begin{cases} \sum_x g(x) \cdot f(x) & \text{if } X \text{ is discrete} \\ \int g(x) \cdot f(x)\, dx & \text{if } X \text{ is continuous} \end{cases}$$

Key insight: You don't need to find the PMF/PDF of $g(X)$ first. Just weight $g(x)$ by the original probability of x.

Example: For a fair die, $\mathbb{E}[X^2] = \frac{1}{6}(1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2) = \frac{91}{6} \approx 15.17$

Why "Unconscious"?

The strange name has a story

That formula is so intuitive that statisticians use it *without thinking* — often treating it as the *definition* of expected value rather than a theorem requiring proof.

Sheldon Ross coined the name in his textbook *A First Course in Probability* (1988) to gently mock this habit. The joke: statisticians apply it "unconsciously."

The controversy: Casella & Berger pushed back, arguing that calling it "unconscious" trivializes a result that actually requires careful measure-theoretic justification.

Ross later removed the name from subsequent editions — but it stuck anyway.

For us: the formula is indeed intuitive. But knowing it's a *theorem* (not a definition) will matter when we hit more advanced topics.

Properties of Expected Value

Linearity is the key

Properties of Expectation

For any random variables X and Y, and constants a, b, c:

1. $\mathbb{E}[c] = c$ (expectation of a constant is itself)
2. $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$ (linearity for one RV)
3. $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ (always true, even if X and Y are dependent!)

Property 3 is remarkable: **linearity of expectation does not require independence**.

Example: Expected sum of two dice = $\mathbb{E}[X_1] + \mathbb{E}[X_2] = 3.5 + 3.5 = 7$

This works even though the dice might be weighted together in some complicated way.

Part V

Variance

How spread out is the distribution?

Warning: Nonlinear Functions

A common mistake

Linearity does NOT extend to nonlinear functions:

$$\mathbb{E}[g(X)] \neq g(\mathbb{E}[X]) \text{ in general}$$

Example: Fair die roll.

- $\mathbb{E}[X] = 3.5$
- $(\mathbb{E}[X])^2 = 3.5^2 = 12.25$
- But $\mathbb{E}[X^2] = \frac{91}{6} \approx 15.17$

So $\mathbb{E}[X^2] \neq (\mathbb{E}[X])^2$

This inequality will be crucial when we define variance.

Jensen's Inequality

The direction of the inequality

We know $\mathbb{E}[g(X)] \neq g(\mathbb{E}[X])$ in general. But can we say which is bigger?

Jensen's Inequality

If g is **convex** (curves upward, like x^2): $\mathbb{E}[g(X)] \geq g(\mathbb{E}[X])$

If g is **concave** (curves downward, like $\log x$): $\mathbb{E}[g(X)] \leq g(\mathbb{E}[X])$

Example: Since $g(x) = x^2$ is convex, Jensen tells us $\mathbb{E}[X^2] \geq (\mathbb{E}[X])^2$

This means $\mathbb{E}[X^2] - (\mathbb{E}[X])^2 \geq 0$ — which is why **variance is always non-negative**!

Application: Risk aversion. If utility u is concave (diminishing marginal utility):

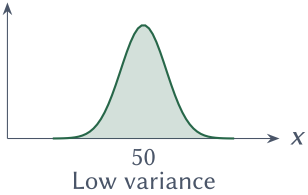
$$\mathbb{E}[u(X)] \leq u(\mathbb{E}[X])$$

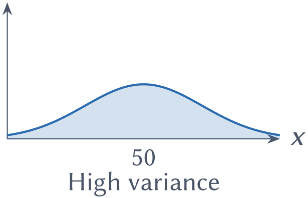
You prefer a sure \$50 over a 50/50 gamble between \$0 and \$100.

Spread Matters

Why the mean isn't enough

Consider two random variables, both with $\mathbb{E}[X] = 50$:





Both have the same mean, but they behave very differently.

Variance measures how spread out the distribution is around its mean.

Definition: Variance

Variance

The **variance** of a random variable X is:

$$\text{Var}[X] = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$$

Interpretation: The average squared deviation from the mean.

Why squared?

- Deviations above and below the mean would cancel out
- Squaring makes all deviations positive
- Larger deviations get penalized more heavily

Notation: $\text{Var}[X]$, σ^2, σ_X^2, $V[X]$ all mean the same thing.

Computational Formula for Variance

The version you'll actually use

Alternative Formula

$$\text{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

Proof:

$$\begin{aligned}
\text{Var}[X] &= \mathbb{E}[(X - \mathbb{E}[X])^2] \\
&= \mathbb{E}[X^2 - 2X\,\mathbb{E}[X] + (\mathbb{E}[X])^2] \\
&= \mathbb{E}[X^2] - 2\,\mathbb{E}[X]\,\mathbb{E}[X] + (\mathbb{E}[X])^2 \\
&= \mathbb{E}[X^2] - (\mathbb{E}[X])^2
\end{aligned}$$

Mnemonic: "The variance is the mean of the square minus the square of the mean."

Why Variance Is Always Non-Negative

Jensen's inequality in action

From the computational formula:

$$\text{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

Question: Why can't this ever be negative?

Answer: Jensen's inequality! Since $g(x) = x^2$ is **convex**:

$$\mathbb{E}[X^2] \geq (\mathbb{E}[X])^2$$

Therefore:

$$\text{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \geq 0$$

Variance equals zero only when X is constant (no randomness). Otherwise, $\mathbb{E}[X^2] > (\mathbb{E}[X])^2$ strictly.

Seeing Jensen's Inequality

A concrete example with three observations

Suppose X takes values 2, 5, 8 with equal probability $\frac{1}{3}$ each.

x	x^2	$(x - \bar{x})^2$
2	4	$(2-5)^2 = 9$
5	25	$(5-5)^2 = 0$
8	64	$(8-5)^2 = 9$
Mean: $\mathbb{E}[X] = 5$	$\mathbb{E}[X^2] = 31$	$\text{Var}[X] = 6$

Jensen's inequality in action:

$$\mathbb{E}[X^2] = 31 \quad > \quad (\mathbb{E}[X])^2 = 5^2 = 25$$

Variance: $\text{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = 31 - 25 = 6 \geq 0$ ✓

Example: Variance of Bernoulli

Let $X \sim \text{Bernoulli}(p)$.

We know: $\mathbb{E}[X] = p$

Compute $\mathbb{E}[X^2]$: Since $X \in \{0, 1\}$, we have $X^2 = X$, so $\mathbb{E}[X^2] = \mathbb{E}[X] = p$

Therefore:

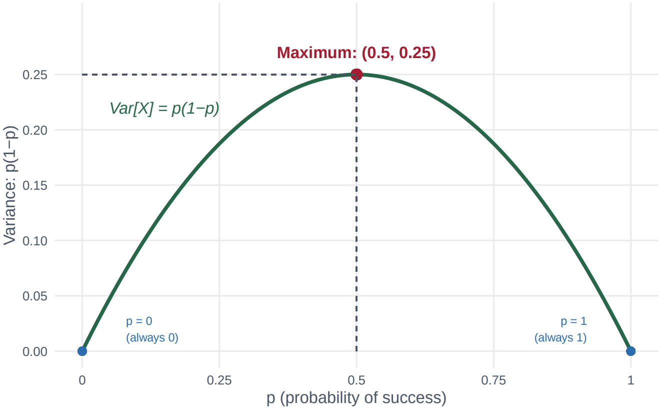
$$\text{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = p - p^2 = p(1 - p)$$

Notice:

- Maximum variance at $p = 0.5$: $\text{Var}[X] = 0.25$
- Variance is 0 when $p = 0$ or $p = 1$ (no uncertainty)

Visualizing Bernoulli Variance

Maximum uncertainty at $p = 0.5$



The parabola $p(1-p)$ is maximized at $p = 0.5$ — when the outcome is most uncertain.

Example: Variance of Uniform(0,1)

Let $X \sim \text{Uniform}(0, 1)$.

We know: $\mathbb{E}[X] = \frac{1}{2}$

Compute $\mathbb{E}[X^2]$:

$$\mathbb{E}[X^2] = \int_0^1 x^2 \cdot 1\, dx = \left[\frac{x^3}{3}\right]_0^1 = \frac{1}{3}$$

Therefore:

$$\text{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$$

For $X \sim \text{Uniform}(a, b)$: $\text{Var}[X] = \frac{(b-a)^2}{12}$

Properties of Variance

Not as nice as expectation!

Properties of Variance

For any random variable X and constants a, b:

1. $\text{Var}[X] \geq 0$ (variance is non-negative)
2. $\text{Var}[c] = 0$ (constants have no variability)
3. $\text{Var}[X + c] = \text{Var}[X]$ (shifting doesn't change spread)
4. $\text{Var}[aX] = a^2 \text{Var}[X]$ (note the square!)

Combining: $\text{Var}[aX + b] = a^2 \text{Var}[X]$

Warning: $\text{Var}[X + Y] \neq \text{Var}[X] + \text{Var}[Y]$ in general!

Variances only add when X and Y are independent. We'll cover this with joint distributions.

Standard Deviation

Back to the original units

Variance has a problem: the units are squared. If X is in dollars, $\text{Var}[X]$ is in dollars^2. Not interpretable!

Standard Deviation

The **standard deviation** of X is: $\sigma[X] = \sqrt{\text{Var}[X]}$

Interpretation: Average distance from the mean (roughly).

Examples:

- Bernoulli(p): $\sigma = \sqrt{p(1-p)}$. At $p = 0.5$: $\sigma = 0.5$
- Uniform$(0, 1)$: $\sigma = \sqrt{1/12} \approx 0.289$

General Moments

Beyond mean and variance

Mean and variance are special cases of a general concept:

Moments

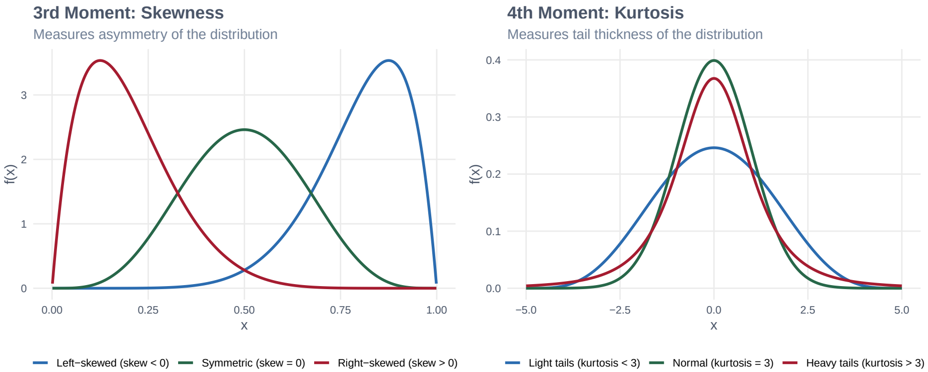
The kth **moment** of X: $\mathbb{E}[X^k]$ The kth **central moment**: $\mathbb{E}[(X-\mu)^k]$

k	**Central Moment**	**Measures**
1	$\mathbb{E}[X-\mu] = 0$	(always zero)
2	$\mathbb{E}[(X-\mu)^2] = \text{Var}[X]$	Spread
3	$\mathbb{E}[(X-\mu)^3]$ (normalized: skewness)	Asymmetry
4	$\mathbb{E}[(X-\mu)^4]$ (normalized: kurtosis)	Tail thickness

For now, focus on moments 1 and 2. Skewness and kurtosis appear in finance and distributional diagnostics.

Seeing Skewness and Kurtosis

What the 3rd and 4th moments look like



Left: Skewness measures which direction the tail stretches. Right: Kurtosis measures how heavy the tails are.

Part VI

Independence (Preview)

When knowing X tells you nothing about Y

Independence of Random Variables

A preview before joint distributions

Independence (CDF Definition)

Two random variables X and Y are **independent**, written $X \perp\!\!\!\perp Y$, if:

$$P(X \leq x, Y \leq y) = P(X \leq x) \cdot P(Y \leq y)$$

for all x and y.

Interpretation: The events $\{X \leq x\}$ and $\{Y \leq y\}$ are independent events (in the sense from last week) for every choice of x and y.

Intuition: Knowing the value of X gives you no information about Y.

We'll develop this more fully when we cover joint distributions. For now, this lets us state one key fact...

Key Fact: Variance of a Sum

Independence makes variances add

Variance of Independent Sum

If $X \perp\!\!\!\perp Y$, then:

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$$

This does NOT hold if X and Y are dependent.

Example: Roll two independent fair dice.

- $\mathbb{E}[X_1 + X_2] = \mathbb{E}[X_1] + \mathbb{E}[X_2] = 3.5 + 3.5 = 7$ (always works)
- $\text{Var}[X_1 + X_2] = \text{Var}[X_1] + \text{Var}[X_2] = \frac{35}{12} + \frac{35}{12} = \frac{35}{6}$ (requires independence)

When we introduce covariance, we'll see the general formula:
$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2\,\text{Cov}[X, Y]$

Summary

What we covered today

Random Variables: Functions $X : \Omega \to \mathbb{R}$. Discrete (countable) vs. Continuous (uncountable).

Distributions:

- PMF: $f(x) = P(X = x)$ for discrete RVs
- CDF: $F(x) = P(X \leq x)$ for any RV
- PDF: $f(x) = F'(x)$ for continuous RVs

Summary Statistics:

- Expected value: $\mathbb{E}[X]$ = weighted average of values
- Variance: $\text{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$ = spread around mean
- Standard deviation: $\sigma = \sqrt{\text{Var}[X]}$

Key Facts: Linearity of expectation (always). Variance adds only under independence.

Coming Up

Next lecture: More on expectations

- Famous distributions (Binomial, Poisson, Geometric, Exponential)
- The Normal distribution
- Why these distributions appear everywhere

Reading: Aronow & Miller, Chapter 1 (sections 1.2)

Problem Set 1: Due Tuesday — covers probability and random variables