Discrete Distributions

Bernoulli, Binomial, and Poisson

Scott Cunningham

Gov 2001 · Harvard University

Spring 2026

Where We Are

From abstract machinery to named distributions

Last week: The general framework

- Random variables, PMFs, PDFs, CDFs
- Expected value, variance, Jensen's inequality
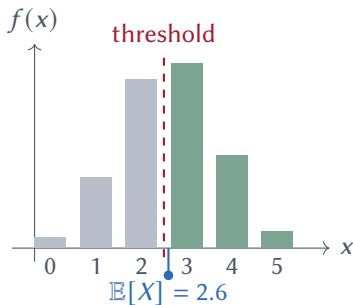- Independence and indicator variables

This week: Named distributions — the **vocabulary** of statistics

- **Today**: Discrete distributions (Bernoulli, Binomial, Poisson)
- **Wednesday**: Continuous distributions (Uniform, Normal, Exponential)

Reading: Aronow & Miller §1.2–1.3, Blackwell 2.2–2.3

A Puzzle from Thursday

How can you win 54% of the time but expect only 2.6 votes?

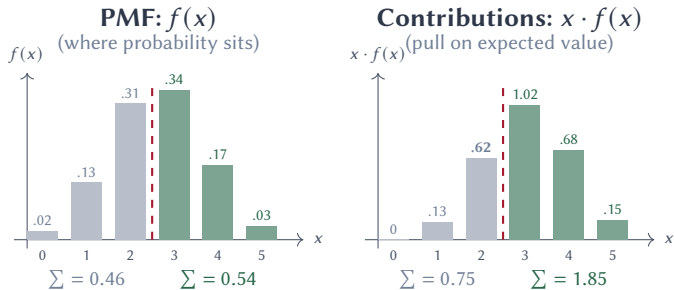**P(X ≥ 3) = 0.54**

Win more than half

E[X] = 2.60

Below threshold

How can both be true?

PMF vs. Contributions to E[X]

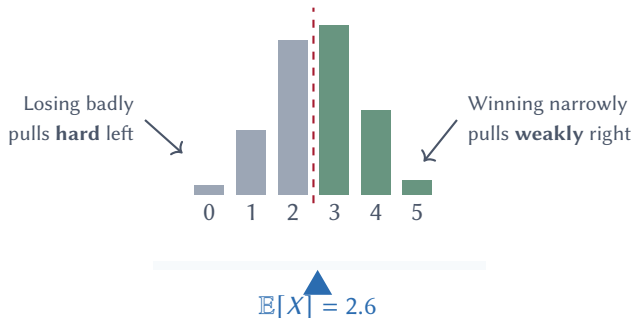The $x = 2$ bar looks small in the PMF, but contributes a lot to E[X]

PMF: $f(x)$
(where probability sits)

$f(x)$

.02   .13   .31   .34   .17   .03

0   1   2   3   4   5 x

$\sum = 0.46$ $\sum = 0.54$

Contributions: $x \cdot f(x)$
(pull on expected value)

$x \cdot f(x)$

0   .13   .62   1.02   .68   .15

0   1   2   3   4   5 x

$\sum = 0.75$ $\sum = 1.85$

Left: PMF shows probability mass. More mass on the right (54%), so the map is *usually* struck down.

Right: Contributions to $\mathbb{E}[X]$. The $x = 2$ bar is now substantial (0.62). That's a lot of "pull" just below threshold.

$\mathbb{E}[X] = 0.75 + 1.85 = 2.60$ (sum of all contributions)

Expected Value is a Center of Mass

E[X] is pulled by magnitude, not just frequency

Losing badly pulls **hard** left

Winning narrowly pulls **weakly** right

0 1 2 3 4 5

$\mathbb{E}[X] = 2.6$

Resolution: More than half the probability mass is above 3, but the *center of mass* is at 2.6. Losing badly (0, 1) pulls harder than winning narrowly (3). The asymmetry pulls $\mathbb{E}[X]$ below threshold.

The A-List and the Character Actors

All famous discrete distributions have non-negative support

All of today's distributions **count things**: successes, trials, events. Counts are non-negative, so support is $\{0, 1, 2, \ldots\}$ or subsets.

The A-List (Tom Hanks, Meryl Streep)

Distribution	Models
Bernoulli	Single yes/no
Binomial	Successes in n trials
Poisson	Rare event counts
Geometric	Trials to first success

You'll see these everywhere.

The B-List (Character actors)

Distribution	Models
Neg. Binomial	Overdispersed counts
Hypergeometric	Sample w/o replace

J.K. Simmons ("the insurance guy"), Margo Martindale ("that lady"). Solid. Not today.

Today: The A-list. For each: history → apps → properties

Why Named Distributions?

These aren't arbitrary — they model real phenomena

Three reasons to learn these:

1. **They appear everywhere**: Elections, counts, durations, waiting times
2. **They're tractable**: We can derive E[X], Var[X], and more
3. **They connect**: Understanding one helps you understand others

Today's cast:

- **Bernoulli** — yes/no, success/failure (the foundation)
- **Binomial** — counting successes in *n* trials
- **Poisson** — counting rare events
- **Geometric** — waiting for first success

The Big Picture

Why distributions matter for inference

What we're building toward:

1. We **describe populations** with distributions (this week)
2. We **estimate parameters** (p, λ, μ) from sample data (coming soon)
3. We **quantify uncertainty** about those estimates (requires understanding distributions)

Example: A poll samples 100 voters. 58 support candidate A.

- The *population* has some true support rate p (unknown)
- Our *estimate* is $\hat{p} = 0.58$ (from sample)
- How good is this estimate? *That depends on the Binomial distribution.*

Distributions aren't just mathematical objects — they're the foundation for learning from data.

The Bernoulli Distribution

The Simplest Random Variable

Jacob Bernoulli (1655–1705)

The mathematician who formalized probability

Jacob Bernoulli was a Swiss mathematician from a family that produced eight prominent mathematicians across three generations. His posthumous work *Ars Conjectandi* (1713) was foundational to probability theory.

Key contributions:

- Proved the first version of the **Law of Large Numbers**
- Showed that with enough trials, observed frequencies converge to true probabilities
- Established probability as a rigorous mathematical field

The Bernoulli distribution — a single binary trial — is named in his honor. Simple as it is, it's the building block for much of what follows.

Where You See Bernoulli in Social Science

Any yes/no outcome is a Bernoulli trial

Political science:

- Did this citizen vote? ($p \approx 0.6$ in US presidential elections)
- Did the incumbent win this district?
- Does this respondent approve of the president?

Economics:

- Did this worker find a job this month?
- Did this firm enter the market?

Sociology:

- Did this couple divorce within 5 years?
- Did this student graduate?

Whenever your outcome is binary (0 or 1), you're in Bernoulli territory.

The Bernoulli Distribution

The simplest random variable: just 0 or 1

The model: $X \sim \text{Bernoulli}(p)$ means:

$$X = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1 - p \end{cases}$$

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

Support: $\{0, 1\}$ — only two possible values

The Bernoulli is a single trial. What if we have many trials? That's where the Binomial comes in.

A Quick Definition: Support

What values can a random variable take?

Definition: The **support** of X is the set of values where the PMF (or PDF) is positive:

$$\text{Supp}[X] = \{x : f(x) > 0\}$$

Examples:

- Bernoulli(p): Support = $\{0, 1\}$
- Binomial(n, p): Support = $\{0, 1, 2, \ldots, n\}$
- Poisson(λ): Support = $\{0, 1, 2, \ldots\}$ (unbounded)

Why it matters: When computing expectations or conditioning on X, you sum/integrate over the support. Knowing what values are possible is fundamental.

For Bernoulli, $\mathbb{E}[X] = p$ **and** $\text{Var}[X] = p(1-p)$

For $X \sim \text{Bernoulli}(p)$:

Expected value (this is key — memorize it):
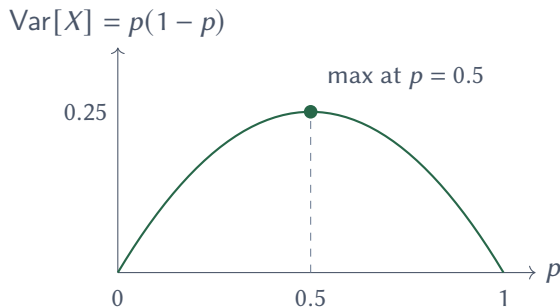
$$\mathbb{E}[X] = 0 \cdot (1-p) + 1 \cdot p = \boxed{p}$$

Variance:

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

Key insight: $\mathbb{E}[X] = \mathbb{P}(X = 1)$. For a 0/1 variable, the mean *is* the probability.
This is why we can estimate probabilities by computing sample means.

Bernoulli Variance is Maximized at $p = 0.5$

$\text{Var}[X] = p(1 - p)$

max at $p = 0.5$

0.25

0   0.5   1 p

When outcomes are most uncertain ($p = 0.5$), variance is highest.

Certainty ($p = 0$ or $p = 1$) means zero variance — no spread if there's only one possible outcome.

The Binomial Distribution

Counting Successes

Pascal, Fermat, and the Gamblers

Probability theory was invented to solve gambling problems

In 1654, a French gambler named **Chevalier de Méré** had a problem: how should you fairly divide the pot when a dice game is interrupted mid-play?

He asked **Blaise Pascal**, who wrote to **Pierre de Fermat**. Their letters that summer are often called *the birth of probability theory.*

The "problem of points": Each player has some probability of eventually winning the remaining rounds. Pascal and Fermat realized you need to count the *ways* each could win — the binomial coefficient:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Weirdly, gamblers drove these early innovations. And they still do: modern finance and speculation run on probability theory that traces back to these 1654 letters about dice.

The math has far exceeded gambling — but gambling remains where it started, and where a lot of the money still is.

Where You See Binomial in Social Science

Counting successes in a fixed number of trials

Political science:

- Out of 10 voters polled, how many support candidate A?
- Out of 50 precincts, how many have wait times over 1 hour?
- Out of 9 Supreme Court justices, how many vote to strike down?

Economics:

- Out of 100 loan applicants, how many default?
- Out of 20 firms, how many survive their first year?

Public health:

- Out of 500 vaccinated people, how many get infected?
- Out of 30 patients, how many respond to treatment?

Whenever you're counting successes in *n* independent trials, you're in Binomial territory.

From Bernoulli to Binomial

Sum of independent trials

Setup: Run n independent Bernoulli trials, each with success probability p.

Question: What's the distribution of the **total number of successes**?

If $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}(p)$, then:

$$Y = \sum_{i=1}^{n} X_i \sim \text{Binomial}(n, p)$$

You saw this logic last week with the redistricting court — same idea, but now p is the same for all trials.

The Binomial PMF Counts Ways to Achieve k Successes

Three ingredients: successes, failures, arrangements

For $Y \sim \text{Binomial}(n, p)$:

$$f(k) = \mathbb{P}(Y = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

for $k \in \{0, 1, 2, \ldots, n\}$

Decomposing the formula:

- p^k: probability that the k successes happen
- $(1 - p)^{n-k}$: probability that the $n - k$ failures happen
- $\binom{n}{k} = \frac{n!}{k!(n-k)!}$: number of ways to choose *which* k trials are successes

You saw this logic last week with the redistricting court — same idea, but now p is the same for all trials.

$\mathbb{E}[Y] = np$ Because It's a Sum of Bernoullis

Linearity of expectation does the heavy lifting

For $Y \sim \text{Binomial}(n, p)$:

Expected value: Since $Y = \sum_{i=1}^{n} X_i$ where $X_i \sim \text{Bernoulli}(p)$:

$$\mathbb{E}[Y] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \boxed{np}$$

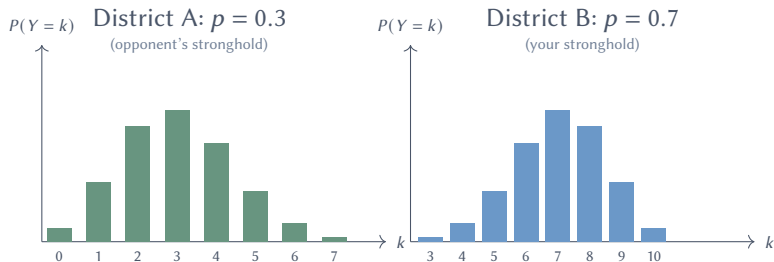**Variance**: Since the X_i are **independent**:

$$\text{Var}[Y] = \sum_{i=1}^{n} \text{Var}[X_i] = \boxed{np(1-p)}$$

Linearity of expectation **always** works. Additivity of variance requires **independence**.

How Likely Is Your Candidate's Support Level?

Visualizing the Binomial: same sample size, different support rates

Poll 10 voters in two different districts:

Same sample size $n = 10$, different p. The distribution shifts with the true support rate.

Key Property: Binomials Add

If they share the same p

Theorem: If $X \sim \text{Binomial}(n, p)$ and $Y \sim \text{Binomial}(m, p)$ are **independent**, then:

$$X + Y \sim \text{Binomial}(n + m, p)$$

Intuition: X is successes in n trials, Y is successes in m trials. Together, you have $n + m$ trials.

Example: Two precincts, same underlying turnout rate $p = 0.6$.

- Precinct A: 100 voters, $X \sim \text{Bin}(100, 0.6)$
- Precinct B: 150 voters, $Y \sim \text{Bin}(150, 0.6)$
- Combined: $X + Y \sim \text{Bin}(250, 0.6)$

This only works when p is the same across both. Different p's? No nice closed form.

Your Turn: Binomial Practice

Try these before we move on

A poll samples **200 voters**. Suppose the true approval rate is $p = 0.45$.

Let X = number who approve. Then $X \sim \text{Binomial}(200, 0.45)$.

Questions:

1. What is $\mathbb{E}[X]$?
2. What is $\text{Var}[X]$? What is the standard deviation?
3. What is $\mathbb{P}(X = 90)$? (Just set up the formula)

Answers:

1. $\mathbb{E}[X] = np = 200 \times 0.45 = \boxed{90}$
2. $\text{Var}[X] = np(1 - p) = 200 \times 0.45 \times 0.55 = 49.5$, so $\sigma = \sqrt{49.5} \approx \boxed{7.04}$
3. $\mathbb{P}(X = 90) = \binom{200}{90}(0.45)^{90}(0.55)^{110}$ (use software to compute)

The Poisson Distribution

Counting Rare Events

Siméon Denis Poisson (1781–1840)

From artillery to actuarial science

Poisson was a French mathematician who studied under Laplace and Lagrange. He introduced his distribution in 1837 in a work on criminal justice — studying the number of wrongful convictions in court cases.

But the distribution became famous through a morbid application: **Ladislaus Bortkiewicz** (1898) used it to model *deaths by horse kicks in the Prussian army*. He showed that these rare, random events followed the Poisson beautifully.

Why it works: When you have many opportunities for something rare to happen (many soldiers, many days), and each event is independent, the Poisson emerges naturally.

The "law of rare events" — Poisson appears whenever you're counting unlikely things in large populations.

Where You See Poisson in Social Science

Counting rare events in large populations or time windows

Political science:
- Supreme Court vacancies per presidential term (average ≈ 1.5)
- Coups in a region per decade
- Mass protests in a country per year
- Terrorist attacks per month

Economics:
- Number of bankruptcies in a sector per quarter
- Patent applications per firm per year

Criminology / Public health:
- Homicides per city per year
- Disease outbreaks per region per decade

Whenever you're counting events that happen rarely and independently, you're in Poisson territory.

The Poisson Distribution

Counting events that occur at a constant rate

The model: $X \sim \text{Poisson}(\lambda)$ counts events occurring at rate λ:

$$f(k) = \mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} \quad \text{for } k \in \{0, 1, 2, \ldots\}$$

Support: $\{0, 1, 2, 3, \ldots\}$ — can be any non-negative integer (no upper bound!)

The parameter λ is the *average rate* of events.

This formula tells us $\mathbb{P}(X = k)$ given λ. Later, we'll flip it: given observed counts, which λ makes the data most probable? That's **maximum likelihood** thinking.

Poisson Has $\mathbb{E}[X] = \mathrm{Var}[X] = \lambda$

For $X \sim \mathrm{Poisson}(\lambda)$:

Expected value: $\mathbb{E}[X] = \lambda$

Variance: $\mathrm{Var}[X] = \lambda$

The mean equals the variance. This is the defining characteristic.

Why this matters: If you see count data where variance $\approx$ mean, Poisson might be a good model. If variance $\gg$ mean ("overdispersion"), you need something else.

Poisson Emerges When *n* Is Large and *p* Is Small

The "law of rare events"

The Poisson approximation:
If *n* is large and *p* is small (with $np = \lambda$ moderate):

$$\text{Binomial}(n, p) \approx \text{Poisson}(\lambda = np)$$

Example: Rare disease in a large population

- $n = 1{,}000{,}000$ people
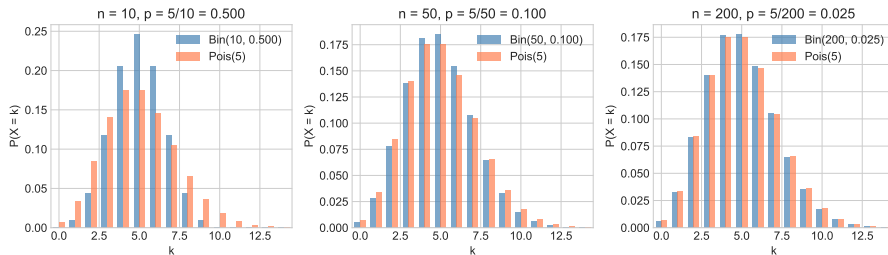- $p = 0.00001$ (1 in 100,000 chance)
- $\lambda = np = 10$

Number of cases $\approx$ Poisson(10), much easier to work with than Binomial(1M, 0.00001).

This is the "law of rare events" — Poisson emerges naturally when counting unlikely things in large populations.
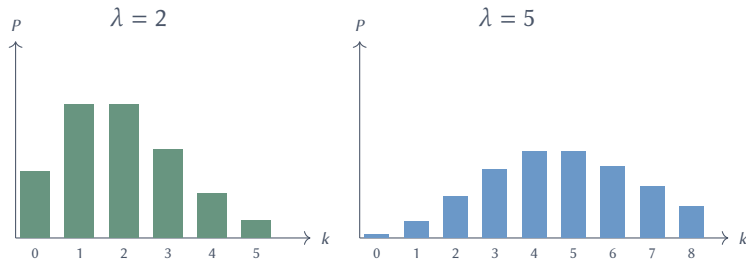
As *n* Grows and *p* Shrinks, Binomial → Poisson

Visualization: Kaixiao Liu

Binomial → Poisson: As $n \to \infty$, $p \to 0$, $np = \lambda$ fixed

Rule of thumb: Use Poisson approximation when $n \geq 20$ and $p \leq 0.05$.

Visualizing the Poisson

As λ increases, the distribution shifts right and becomes more symmetric.

For large λ, Poisson looks increasingly normal (another CLT application).

Key Property: Poissons Add

Theorem: If $X \sim \text{Poisson}(\lambda_1)$ and $Y \sim \text{Poisson}(\lambda_2)$ are **independent**, then:

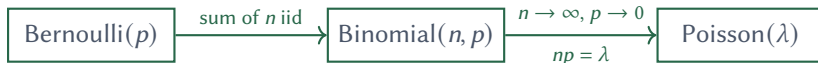$$X + Y \sim \text{Poisson}(\lambda_1 + \lambda_2)$$

Intuition: If events arrive at rate λ_1 from one source and λ_2 from another, the combined rate is $\lambda_1 + \lambda_2$.

Example: Counting legislative hearings

- House committees: $X \sim \text{Poisson}(12)$ per month
- Senate committees: $Y \sim \text{Poisson}(8)$ per month
- Total: $X + Y \sim \text{Poisson}(20)$

Unlike Binomial, Poisson addition works even when the rates differ.

How These Distributions Connect

```
┌──────────────┐  sum of n iid  ┌──────────────┐  n → ∞, p → 0  ┌──────────────┐
│ Bernoulli(p) │───────────────▶│ Binomial(n,p)│───────────────▶│ Poisson(λ)   │
└──────────────┘                └──────────────┘    np = λ       └──────────────┘
```

- **Bernoulli → Binomial**: Sum of independent 0/1 trials
- **Binomial → Poisson**: Many trials, small probability (law of rare events)

Wednesday: Continuous distributions (Uniform, Normal, Exponential) and the Poisson–Exponential connection.

The Geometric Distribution: Pascal Again

How long until the first success?

Blaise Pascal (1623–1662) — the same Pascal from the famous correspondence with Fermat.

The Problem of Points (1654):

Two players are in a game of chance. The game is interrupted before either wins. How should the stakes be divided fairly?

To solve this, Pascal needed to answer: **How many more trials until someone wins?**

This is the Geometric question: *How many independent trials until the first success?*

Once again, gambling drove mathematical innovation. The casinos of 17th-century France were inadvertent research labs.

Where You See Geometric in Social Science

Waiting for the first success

Political Science:

- **Electoral breakthroughs**: How many election cycles until a party wins a historically hostile state?
- **Negotiations**: Rounds of bargaining until first agreement
- **Legislative passage**: Congressional sessions until a bill passes
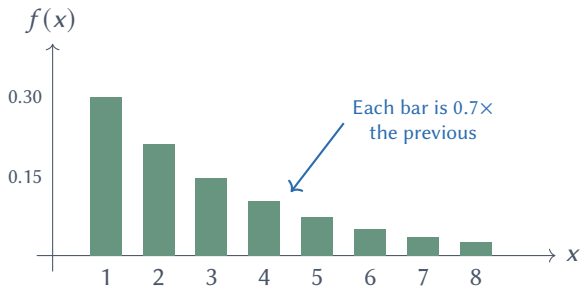- **Coalition formation**: Attempts until government forms

Other fields:

- Quality control: Items inspected until first defect
- Epidemiology: Contacts until disease transmission
- Job search: Applications until first offer

Anywhere you ask "how long until...?" in discrete time, Geometric is nearby.

The Geometric Distribution

Exponential decay in discrete time

$f(x)$

Each bar is $0.7\times$ the previous

PMF: $f(x) = p(1-p)^{x-1}$ for $x = 1, 2, 3, \ldots$

Support: $\{1, 2, 3, \ldots\}$ — you need at least one trial.

Interpretation: Probability of *first* success on trial x = (fail $x - 1$ times) $\times$ (succeed once).

Note: Some books use $\{0, 1, 2, \ldots\}$ (failures before first success). We count trial number.

Geometric Properties

Simple moments, remarkable memorylessness

Moments:

$$\mathbb{E}[X] = \frac{1}{p}$$

$$\text{Var}[X] = \frac{1-p}{p^2}$$

Intuition: If $p = 0.1$ (10% success rate), expect $\mathbb{E}[X] = 10$ trials on average.

The Memoryless Property:

$$\mathbb{P}(X > s + t \mid X > s) = \mathbb{P}(X > t)$$

"If you haven't succeeded after s trials, your expected remaining wait is *exactly the same* as when you started."

The **only** discrete memoryless distribution. Wednesday: its continuous cousin, the Exponential.

Your Turn: Geometric

Legislative persistence

A bill has a 20% chance of passing in any given Congress (independent across sessions).

Questions:

1. What's the expected number of Congresses until passage?
2. What's the probability it takes more than 5 Congresses?

Answers:

1. $\mathbb{E}[X] = 1/p = 1/0.2 = \mathbf{5}$ Congresses
2. $\mathbb{P}(X > 5) = (1 - p)^5 = 0.8^5 \approx \mathbf{0.328}$

About a third of bills like this would still be waiting after 5 Congresses (10 years).

Summary: Four Discrete Distributions

Distribution	Support	$\mathbb{E}[X]$	$\text{Var}[X]$	Use case
Bernoulli(p)	$\{0, 1\}$	p	$p(1-p)$	Binary outcomes
Binomial(n, p)	$\{0, \ldots, n\}$	np	$np(1-p)$	Count successes
Poisson(λ)	$\{0, 1, 2, \ldots\}$	λ	λ	Rare event counts
Geometric(p)	$\{1, 2, 3, \ldots\}$	$1/p$	$(1-p)/p^2$	Trials to 1st success

Sum properties:

- Binomial(n, p) + Binomial(m, p) = Binomial($n + m, p$)   (same p!)
- Poisson(λ_1) + Poisson(λ_2) = Poisson($\lambda_1 + \lambda_2$)
- Geometric + Geometric = Negative Binomial (a B-list distribution!)

Know these moments by heart. They will reappear constantly.

The C-List: Discrete but Negative

Less famous, but they exist

All the distributions we've covered count things, so they're non-negative. But there *are* discrete distributions that can take negative values — they model **differences** or **signed outcomes**, not counts.

Distribution	Support	Use case
Skellam	$\{\ldots, -2, -1, 0, 1, 2, \ldots\}$	Difference of two Poisson counts (e.g., goal differential in soccer)
Rademacher	$\{-1, +1\}$	Random sign assignment (e.g., randomization inference)

We won't cover these, but know they're out there. Discrete ≠ non-negative.

Looking Ahead

Wednesday: Continuous distributions

- Uniform — and the "universality of uniform"
- Normal — the star of the show
- Exponential — the continuous cousin of today's Geometric
- Both Geometric and Exponential are **memoryless** — we'll see why that matters

Reading:

- Aronow & Miller, §1.2 continued (pp. 25–50)
- Blackwell, Chapter 2.3