Expected Value

Gov 2001: Quantitative Social Science Methods I

Scott Cunningham

Harvard University

Spring 2026

Today's Reading

Required

- **Aronow & Miller**, §2.1: Expected value (pp. 45–66)
- **Blackwell**, Ch. 2.4: Summaries of distributions

Key concepts: definition of expected value, linearity, examples with named distributions.

The Problem of Summary

- A random variable X is characterized by its **distribution**
- But distributions can be complicated—many numbers, whole functions
- We often want to summarize: *What's a "typical" value?*

Examples in political science:

- What's the average vote share for incumbent candidates?
- What's a typical treatment effect in get-out-the-vote experiments?
- What proportion of the population supports this policy?

Enter Expected Value

The **expected value** $\mathbb{E}[X]$ provides a single-number summary.

Two interpretations (for now):

- The “center of mass” of the distribution
- The long-run average if we sample repeatedly

Third interpretation (we’ll prove this later):
The value that minimizes mean squared error as a prediction.

Definition: Discrete Case

Expected Value (Discrete)

If X is a discrete random variable with PMF $f(x) = \Pr(X = x)$:

$$\mathbb{E}[X] = \sum_x x \cdot f(x) = \sum_x x \cdot \Pr(X = x)$$

Intuition:

- Each value x is weighted by its probability $f(x)$
- Values that occur more often contribute more to the average
- This is a probability-weighted average

Also called **expectation** or **mean**; often written μ or μ_X.

Example: Fair Die

Setup: Roll a fair 6-sided die. Let X = number showing.

The PMF: $f(x) = \frac{1}{6}$ for $x \in \{1, 2, 3, 4, 5, 6\}$

Expected value:

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

Note: The expected value (3.5) is *not* a possible outcome!
This is fine—$\mathbb{E}[X]$ is a summary, not a prediction of a single roll.

Example: Bernoulli Random Variable

Setup: $X_i = 1$ if voter i supports the policy, $X_i = 0$ otherwise.
Let $X_i \sim \text{Bernoulli}(p)$ where p = true population proportion.

Expected value:

$$\begin{aligned} \mathbb{E}[X_i] &= 0 \cdot \Pr(X_i = 0) + 1 \cdot \Pr(X_i = 1) \\ &= 0 \cdot (1 - p) + 1 \cdot p = p \end{aligned}$$

Key Insight

$\mathbb{E}[X] = p = \Pr(X = 1)$
The expected value of a binary variable equals the probability it equals 1.

This is why we can estimate proportions with sample means. Surveys and experiments exploit this.

Definition: Continuous Case

Expected Value (Continuous)

If X is a continuous random variable with PDF $f(x)$:

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

Same intuition:

- Replace summation with integration
- Each value x is weighted by its density $f(x)$
- Regions with more probability mass contribute more

The integral is over the entire support where $f(x) > 0$.

Example: Uniform Distribution

Setup: $X \sim \text{Uniform}(a, b)$

PDF: $f(x) = \frac{1}{b-a}$ for $x \in [a, b]$

Expected value:

$$\begin{aligned}
\mathbb{E}[X] &= \int_a^b x \cdot \frac{1}{b-a}\, dx \\
&= \frac{1}{b-a} \cdot \frac{x^2}{2}\Big|_a^b \\
&= \frac{1}{b-a} \cdot \frac{b^2 - a^2}{2} = \frac{(b+a)(b-a)}{2(b-a)} = \frac{a+b}{2}
\end{aligned}$$

Key Result

For $X \sim \text{Uniform}(a, b)$: $\mathbb{E}[X] = \frac{a+b}{2}$ (the midpoint!)

Example: Normal Distribution

Setup: $X \sim N(\mu, \sigma^2)$

PDF: $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

Expected value: By symmetry of the normal distribution around μ:

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

Key Result

For $X \sim N(\mu, \sigma^2)$: $\mathbb{E}[X] = \mu$
The parameter μ *is* the expected value—that's why we call it the "mean parameter."

Expected Value of Functions of X

Law of the Unconscious Statistician (LOTUS)

For any function $g(X)$:

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

Why "unconscious"?

- You don't need to find the distribution of $g(X)$ first
- Just weight $g(x)$ by the original probabilities $f(x)$

This is extremely useful—we'll use it constantly.

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

Setup: Roll a fair die. What is $\mathbb{E}[X^2]$?

Using LOTUS with $g(x) = x^2$:

$$\mathbb{E}[X^2] = \sum_{x=1}^{6} x^2 \cdot \frac{1}{6}$$
$$= \frac{1}{6}(1 + 4 + 9 + 16 + 25 + 36)$$
$$= \frac{91}{6} \approx 15.17$$

Important observation:

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

In general: $\mathbb{E}[g(X)] \neq g(\mathbb{E}[X])$ (Jensen's inequality!)

The Linearity of Expectation

Linearity Property

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

$$\mathbb{E}[aX + bY + c] = a\mathbb{E}[X] + b\mathbb{E}[Y] + c$$

Key implications:

- $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ (always—even if X, Y dependent!)
- $\mathbb{E}[aX] = a\mathbb{E}[X]$ (constants factor out)
- $\mathbb{E}[c] = c$ (expected value of a constant is itself)

This is powerful! We don't need independence for additivity.

Linearity Example: Two Dice

Setup: Roll two fair dice. Let S = sum of the dice.

Let X_1 = first die, X_2 = second die, so $S = X_1 + X_2$.

Find $\mathbb{E}[S]$ directly?

- S can be $2, 3, \ldots, 12$
- Need to compute all 36 outcomes, find PMF, sum…tedious!

Using linearity:

$$\mathbb{E}[S] = \mathbb{E}[X_1 + X_2] = \mathbb{E}[X_1] + \mathbb{E}[X_2]$$
$$= 3.5 + 3.5 = 7$$

Done in one line. Linearity is your friend.

Linearity Example: Binomial Mean

Setup: $X \sim \text{Binomial}(n, p)$ = number of successes in n trials.

Key insight: Write $X = X_1 + X_2 + \cdots + X_n$ where each $X_i \sim \text{Bernoulli}(p)$.

Using linearity:

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}[X_1 + X_2 + \cdots + X_n] \\
&= \mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_n] \\
&= p + p + \cdots + p = np
\end{aligned}$$

Key Result

For $X \sim \text{Binomial}(n, p)$: $\mathbb{E}[X] = np$

Much simpler than computing $\sum_{x=0}^{n} x\binom{n}{x}p^x(1-p)^{n-x}$ directly!

$\mathbb{E}[X]$ as the Optimal Constant Predictor

Question: If you had to guess a single number c to predict X, what should you choose?

Criterion: Minimize expected squared error:

$$\min_c \mathbb{E}[(X-c)^2]$$

Solution: Take derivative, set to zero:

$$\frac{d}{dc}\mathbb{E}[(X-c)^2] = \mathbb{E}\left[\frac{d}{dc}(X-c)^2\right] = \mathbb{E}[-2(X-c)] = 0$$

$$-2\,\mathbb{E}[X] + 2c = 0$$

$$c^* = \mathbb{E}[X]$$

Key Result

The expected value $\mathbb{E}[X]$ is the **mean squared error minimizing** constant prediction of X.

Why MSE-Optimality Matters

What we just showed:
If you must predict X with a single constant, choose $\mathbb{E}[X]$.

The natural next question:
What if you can observe something *related* to X before predicting?

Preview: The answer is the **conditional expectation** $\mathbb{E}[X|Z]$.

We'll get there in Week 4. For now, master the unconditional case.

Expected Values: Quick Reference

Distribution	**Notation**	$\mathbb{E}[X]$
Bernoulli	$X \sim \text{Bernoulli}(p)$	p
Binomial	$X \sim \text{Binomial}(n, p)$	np
Poisson	$X \sim \text{Poisson}(\lambda)$	λ
Uniform	$X \sim \text{Uniform}(a, b)$	$\frac{a+b}{2}$
Normal	$X \sim N(\mu, \sigma^2)$	μ
Exponential	$X \sim \text{Exp}(\lambda)$	$\frac{1}{\lambda}$

These will come up repeatedly. Worth memorizing.

Key Takeaways

1. **Expected value** summarizes a distribution with one number
2. **Definition**: $\mathbb{E}[X] = \sum x \cdot f(x)$ or $\int x \cdot f(x)\, dx$
3. **LOTUS**: $\mathbb{E}[g(X)] = \sum g(x)f(x)$—don't need distribution of $g(X)$
4. **Linearity**: $\mathbb{E}[aX + bY + c] = a\mathbb{E}[X] + b\mathbb{E}[Y] + c$ (always!)
5. **Optimality**: $\mathbb{E}[X]$ minimizes $\mathbb{E}[(X - c)^2]$
6. **Warning**: $\mathbb{E}[g(X)] \neq g(\mathbb{E}[X])$ in general

Next time: Variance, covariance, and correlation.

For Wednesday

Topic: Variance, Covariance, Correlation

Expected value tells us the "center"—but how spread out is the distribution?

Reading:

- A&M §2.1 (rest of section) and §2.2.1–2.2.2
- Blackwell Ch. 2.4–2.5