0.1 Learning Objectives

By the end of this lecture, you will be able to:

- Apply properties of expectation: scaling, shift, additivity, LOTUS
- Recognize and work with Uniform, Bernoulli, and Geometric distributions
- Use indicator random variables to solve counting problems
- Apply linearity of expectation to simplify complex calculations

1. Properties of Expectation

1.1 Quick Recall

Recall: For a discrete random variable X with PMF $p_{X(x)}$:

$$E(X) = \sum_{x} x \cdot p_X(x)$$

1.1 Quick Recall

Recall: For a discrete random variable X with PMF $p_{X(x)}$:

$$E(X) = \sum_x x \cdot p_X(x)$$

Today: Properties that make computing expectations **much easier**.

1.1 Quick Recall

Recall: For a discrete random variable X with PMF $p_{X(x)}$:

$$E(X) = \sum_x x \cdot p_X(x)$$

Today: Properties that make computing expectations **much easier**.

- Constants

1.1 Quick Recall

Recall: For a discrete random variable X with PMF $p_{X(x)}$:

$$E(X) = \sum_x x \cdot p_X(x)$$

Today: Properties that make computing expectations **much easier**.

- Constants
- Scaling

1.1 Quick Recall

Recall: For a discrete random variable X with PMF $p_{X(x)}$:

$$E(X) = \sum_x x \cdot p_X(x)$$

Today: Properties that make computing expectations **much easier**.

- Constants
- Scaling
- Shift

1.1 Quick Recall

Recall: For a discrete random variable X with PMF $p_{X(x)}$:

$$E(X) = \sum_x x \cdot p_X(x)$$

Today: Properties that make computing expectations **much easier**.

- Constants
- Scaling
- Shift
- Additivity (Linearity)

1.1 Quick Recall

Recall: For a discrete random variable X with PMF $p_{X(x)}$:

$$E(X) = \sum_x x \cdot p_X(x)$$

Today: Properties that make computing expectations **much easier**.

- Constants
- Scaling
- Shift
- Additivity (Linearity)
- LOTUS

1.2 Property 1: Constants

Definition: Expectation of a Constant

If c is a constant, then $E(c) = c$

1.2 Property 1: Constants

Definition: Expectation of a Constant

If c is a constant, then $E(c) = c$

Proof: A constant c is a random variable that always equals c.

$$E(c) = c \cdot P(X = c) = c \cdot 1 = c$$

1.2 Property 1: Constants

Definition: Expectation of a Constant

If c is a constant, then $E(c) = c$

Proof: A constant c is a random variable that always equals c.

$$E(c) = c \cdot P(X = c) = c \cdot 1 = c$$

Note: This seems obvious, but it's useful in proofs!

1.3 Property 2: Scaling

Definition: Scaling Property

For any constant a: $E(aX) = aE(X)$

1.3 Property 2: Scaling

Definition: Scaling Property

For any constant a: $E(aX) = aE(X)$

Proof:

$$E(aX) = \sum_x ax \cdot p_{X(x)} = a \sum_x x \cdot p_{X(x)} = aE(X)$$

1.3 Property 2: Scaling

Definition: Scaling Property

For any constant a: $E(aX) = aE(X)$

Proof:

$$E(aX) = \sum_x ax \cdot p_{X(x)} = a \sum_x x \cdot p_{X(x)} = aE(X)$$

Example: Doubling a Die

If X is a fair die roll, $E(X) = \frac{7}{2}$.

If you double the result: $E(2X) = 2 \cdot \frac{7}{2} = 7$

1.4 Property 3: Shift

Definition: Shift Property

For any constant b: $E(X + b) = E(X) + b$

1.4 Property 3: Shift

Definition: Shift Property

For any constant b: $\quad E(X + b) = E(X) + b$

Proof:

$$E(X+b) = \sum_x (x+b) \cdot p_{X(x)} = \sum_x x \cdot p_{X(x)} + b \sum_x p_{X(x)} = E(X) + b$$

1.4 Property 3: Shift

Definition: Shift Property

For any constant b: $E(X+b) = E(X) + b$

Proof:

$$E(X+b) = \sum_x (x+b) \cdot p_{X(x)} = \sum_x x \cdot p_{X(x)} + b \sum_x p_{X(x)} = E(X) + b$$

Combining scaling and shift:

$$E(aX+b) = aE(X) + b$$

1.5 Property 4: Additivity

Definition: Additivity of Expectation

For any random variables X and Y:

$$E(X + Y) = E(X) + E(Y)$$

1.5 Property 4: Additivity

Definition: Additivity of Expectation

For any random variables X and Y:

$$E(X+Y) = E(X) + E(Y)$$

Note: This works even if X and Y are **dependent**!

1.5 Property 4: Additivity

Definition: Additivity of Expectation

For any random variables X and Y:

$$E(X + Y) = E(X) + E(Y)$$

Note: This works even if X and Y are **dependent**!

Combining with scaling gives us **linearity**:

$$E(aX + bY) = aE(X) + bE(Y)$$

2. Introducing Distributions

2.1 From Data to Histograms

Suppose we survey 50 students: “How many classes are you taking?”

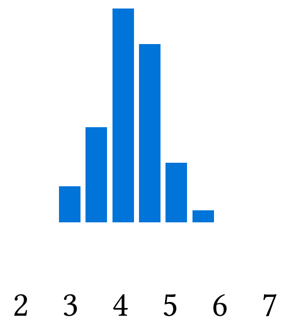
Raw data:

Student	Classes
1	4
2	5
3	4
4	3
5	4
⋮	⋮
50	5

Frequency table:

Classes	Count
2	3
3	8
4	18
5	15
6	5
7	1

Histogram:



2.1 From Data to Histograms

Suppose we survey 50 students: “How many classes are you taking?”

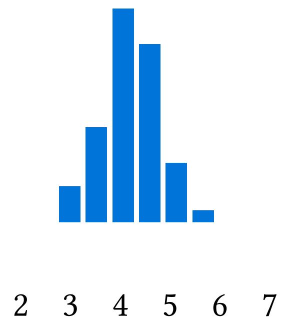
Raw data:

Student	Classes
1	4
2	5
3	4
4	3
5	4
⋮	⋮
50	5

Frequency table:

Classes	Count
2	3
3	8
4	18
5	15
6	5
7	1

Histogram:

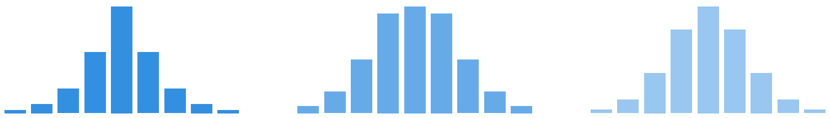


The histogram shows the **shape** of the data.

2.2 Patterns in Data: Family Resemblance

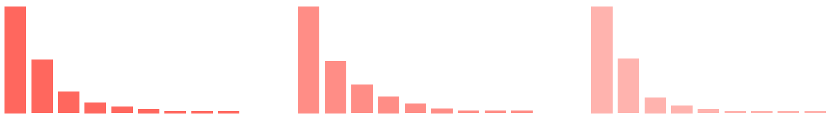
Different datasets, but similar shapes:

"Bell-shaped"



- Symmetric around center
- Most data near the middle
- Tails drop off smoothly

"Right-skewed"



- Peak on the left
- Long tail to the right
- Common: income, wait times

2.3 Why Name These Patterns?

2.3 Why Name These Patterns?

These "family resemblances" appear over and over in real data.

2.3 Why Name These Patterns?

These “family resemblances” appear over and over in real data.

Idea: Give them names and study their properties mathematically.

2.3 Why Name These Patterns?

These "family resemblances" appear over and over in real data.

Idea: Give them names and study their properties mathematically.

Definition: Probability Distribution

A mathematical description of a random variable's possible values and their probabilities.

2.3 Why Name These Patterns?

These "family resemblances" appear over and over in real data.

Idea: Give them names and study their properties mathematically.

Definition: Probability Distribution

A mathematical description of a random variable's possible values and their probabilities.

Instead of working with raw data, we can work with the **idealized pattern**.

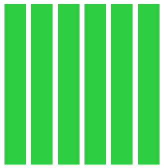
3. The Discrete Uniform Distribution

3.1 Equal Probability Outcomes

Definition: Discrete Uniform Distribution

$X \sim \text{Uniform}\{a, a+1, ..., b\}$ if X takes each value with equal probability:

$$P(X = k) = \frac{1}{b - a + 1}, \quad k = a, a+1, ..., b$$



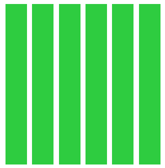
All bars equal height!

3.1 Equal Probability Outcomes

Definition: Discrete Uniform Distribution

$X \sim \text{Uniform}\{a, a+1, ..., b\}$ if X takes each value with equal probability:

$$P(X = k) = \frac{1}{b - a + 1}, \quad k = a, a+1, ..., b$$



All bars equal height!

Mean:

$$E(X) = \frac{a+b}{2}$$

(midpoint of the range)

Variance:

$$\text{Var}(X) = \frac{(b-a+1)^2 - 1}{12}$$

3.2 Example: Die Roll

A fair die is $X \sim \text{Uniform}\{1, 2, 3, 4, 5, 6\}$

3.2 Example: Die Roll

A fair die is $X \sim \text{Uniform}\{1, 2, 3, 4, 5, 6\}$

$$E(X) = \frac{1+6}{2} = \frac{7}{2} = 3.5$$

3.2 Example: Die Roll

A fair die is $X \sim \text{Uniform}\{1, 2, 3, 4, 5, 6\}$

$$E(X) = \frac{1+6}{2} = \frac{7}{2} = 3.5$$

$$\text{Var}(X) = \frac{6^2 - 1}{12} = \frac{35}{12} \approx 2.92$$

3.2 Example: Die Roll

A fair die is $X \sim \text{Uniform}\{1, 2, 3, 4, 5, 6\}$

$$E(X) = \frac{1+6}{2} = \frac{7}{2} = 3.5$$

$$\text{Var}(X) = \frac{6^2 - 1}{12} = \frac{35}{12} \approx 2.92$$

Note: We computed this variance earlier using the definition — the formula is much faster!

3.3 Parameters: Controlling the Specifics

$$P(X = k) = \frac{1}{b - a + 1}$$

3.3 Parameters: Controlling the Specifics

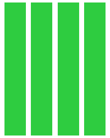
3.3 Parameters: Controlling the Specifics

$$P(X = k) = \frac{1}{b - a + 1}$$

The values a and b control the shape:

$\text{Uniform}\{1, 2, 3, 4\}$

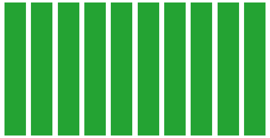
$a = 1$, $b = 4$



4 outcomes, each with prob $1/4$

$\text{Uniform}\{1, 2, \ldots, 10\}$

$a = 1$, $b = 10$



10 outcomes, each with prob $1/10$

3.3 Parameters: Controlling the Specifics

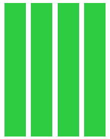
3.3 Parameters: Controlling the Specifics

$$P(X = k) = \frac{1}{b - a + 1}$$

The values a and b control the shape:

$\text{Uniform}\{1, 2, 3, 4\}$

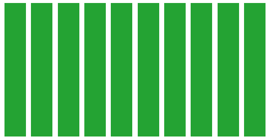
$a = 1$, $b = 4$



4 outcomes, each with prob $1/4$

$\text{Uniform}\{1, 2, \ldots, 10\}$

$a = 1$, $b = 10$



10 outcomes, each with prob $1/10$

3.3 Parameters: Controlling the Specifics

Definition: Parameters

Values that specify a particular distribution within a family.

For Uniform: a (min) and b (max) are the **parameters**.

4. LOTUS: Law of the Unconscious Statistician

4.1 The Problem

Suppose $X \sim \text{Uniform}\{1, ..., 6\}$ (a die roll) and we want $E(X^2)$.

4.1 The Problem

Suppose $X \sim \text{Uniform}\{1, ..., 6\}$ (a die roll) and we want $E(X^2)$.

The hard way: Find the PMF of $Y = X^2$, then compute $E(Y)$.

4.1 The Problem

Suppose $X \sim \text{Uniform}\{1, ..., 6\}$ (a die roll) and we want $E(X^2)$.

The hard way: Find the PMF of $Y = X^2$, then compute $E(Y)$.

X: 1, 2, 3, 4, 5, 6

Each with probability $\frac{1}{6}$

X^2: 1, 4, 9, 16, 25, 36

4.1 The Problem

Suppose $X \sim \text{Uniform}\{1, ..., 6\}$ (a die roll) and we want $E(X^2)$.

The hard way: Find the PMF of $Y = X^2$, then compute $E(Y)$.

X: 1, 2, 3, 4, 5, 6 Each with probability $\frac{1}{6}$

X^2: 1, 4, 9, 16, 25, 36

$$E(X^2) = \frac{1}{6}(1 + 4 + 9 + 16 + 25 + 36) = \frac{91}{6}$$

4.2 LOTUS: The Easy Way

Definition: LOTUS

Law of the Unconscious Statistician

For any function g and random variable X:

$$E(g(X)) = \sum_{x} g(x) \cdot p_X(x)$$

4.2 LOTUS: The Easy Way

Definition: LOTUS

Law of the Unconscious Statistician

For any function g and random variable X:

$$E(g(X)) = \sum_x g(x) \cdot p_X(x)$$

Key insight: Use the PMF of X, not the PMF of $g(X)$!

4.2 LOTUS: The Easy Way

Definition: LOTUS

Law of the Unconscious Statistician

For any function g and random variable X:

$$E(g(X)) = \sum_x g(x) \cdot p_X(x)$$

Key insight: Use the PMF of X, not the PMF of $g(X)$!

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

Same answer, but we didn't need to find the PMF of X^2

4.3 Why “Unconscious Statistician”?

4.3 Why "Unconscious Statistician"?

The formula looks like you're computing $E(g(X))$ by just plugging in $g(x)$…

4.3 Why "Unconscious Statistician"?

The formula looks like you're computing $E(g(X))$ by just plugging in $g(x)$...

...as if you forgot that $g(X)$ is a **different** random variable than X.

4.3 Why "Unconscious Statistician"?

The formula looks like you're computing $E(g(X))$ by just plugging in $g(x)$...

...as if you forgot that $g(X)$ is a **different** random variable than X.

But it actually works! The proof requires showing:

$$\sum_{x} g(x) p_{X(x)} = \sum_{y} y \cdot p_{g(X)}(y)$$

4.3 Why "Unconscious Statistician"?

The formula looks like you're computing $E(g(X))$ by just plugging in $g(x)$…

…as if you forgot that $g(X)$ is a **different** random variable than X.

But it actually works! The proof requires showing:

$$\sum_x g(x) p_{X(x)} = \sum_y y \cdot p_{g(X)}(y)$$

The "unconscious" approach gives the right answer.

4.4 LOTUS: Try It Yourself

Try it yourself

Talk to your neighbor and try to solve this problem.

Let $X \sim \text{Uniform}\{-2, -1, 0, 1, 2\}$.

Find $E(X^2)$.

4.4 LOTUS: Try It Yourself

Try it yourself

Talk to your neighbor and try to solve this problem.

Let $X \sim \text{Uniform}\{-2, -1, 0, 1, 2\}$.

Find $E(X^2)$.

$$E(X^2) = \frac{1}{5}((-2)^2 + (-1)^2 + 0^2 + 1^2 + 2^2)$$

4.4 LOTUS: Try It Yourself

Try it yourself

Talk to your neighbor and try to solve this problem.

Let $X \sim \text{Uniform}\{-2, -1, 0, 1, 2\}$.

Find $E(X^2)$.

$$E(X^2) = \frac{1}{5}((-2)^2 + (-1)^2 + 0^2 + 1^2 + 2^2)$$

$$= \frac{1}{5}(4 + 1 + 0 + 1 + 4) = \frac{10}{5} = 2$$

4.4 LOTUS: Try It Yourself

Try it yourself

Talk to your neighbor and try to solve this problem.

Let $X \sim \text{Uniform}\{-2, -1, 0, 1, 2\}$.

Find $E(X^2)$.

$$E(X^2) = \frac{1}{5}((-2)^2 + (-1)^2 + 0^2 + 1^2 + 2^2)$$

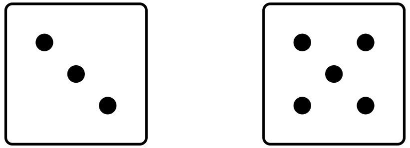
$$= \frac{1}{5}(4 + 1 + 0 + 1 + 4) = \frac{10}{5} = 2$$

Note: $E(X) = 0$ but $E(X^2) = 2$, so $E(X^2) \neq (E(X))^2$ in general!

5. Linearity in Action

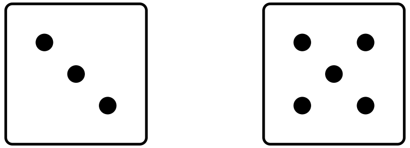
5.1 Motivating Problem

You roll two fair dice and add up the faces.



5.1 Motivating Problem

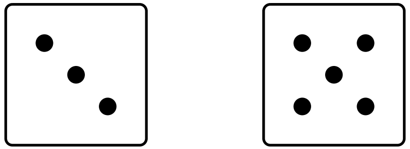
You roll two fair dice and add up the faces.



Question: What is the expected value of the sum?

5.1 Motivating Problem

You roll two fair dice and add up the faces.



Question: What is the expected value of the sum?

Think for 30 seconds: How would you approach this?

5.2 The Hard Way

We could enumerate all 36 outcomes...

5.2 The Hard Way

We could enumerate all 36 outcomes...

- $(1, 1) \to 2$
- $(1, 2) \to 3$
- $(1, 3) \to 4$
- $\vdots$
- $(6, 4) \to 10$
- $(6, 5) \to 11$
- $(6, 6) \to 12$

5.2 The Hard Way

We could enumerate all 36 outcomes…

- $(1, 1) \to 2$
- $(1, 2) \to 3$
- $(1, 3) \to 4$
- $\vdots$
- $(6, 4) \to 10$
- $(6, 5) \to 11$
- $(6, 6) \to 12$

Then compute: $E(X + Y) = \sum_{i=2}^{12} i \cdot P(X + Y = i)$

5.2 The Hard Way

We could enumerate all 36 outcomes…

- $(1, 1) \to 2$
- $(1, 2) \to 3$
- $(1, 3) \to 4$
- $\vdots$
- $(6, 4) \to 10$
- $(6, 5) \to 11$
- $(6, 6) \to 12$

Then compute: $E(X + Y) = \sum_{i=2}^{12} i \cdot P(X + Y = i)$

This is tedious! There must be a better way…

5.3 The Easy Way: Linearity

Let X = first die, Y = second die.

5.3 The Easy Way: Linearity

Let X = first die, Y = second die.

Step 1: Find $E(X)$ for a single die.

$$E(X) = \sum_{i=1}^{6} i \cdot \frac{1}{6} = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = \frac{21}{6} = \frac{7}{2}$$

5.3 The Easy Way: Linearity

Let X = first die, Y = second die.

Step 1: Find $E(X)$ for a single die.

$$E(X) = \sum_{i=1}^{6} i \cdot \frac{1}{6} = \frac{1}{6}(1+2+3+4+5+6) = \frac{21}{6} = \frac{7}{2}$$

Step 2: Apply linearity.

$$E(X+Y) = E(X) + E(Y) = \frac{7}{2} + \frac{7}{2} = 7$$

5.3 The Easy Way: Linearity

Let X = first die, Y = second die.

Step 1: Find $E(X)$ for a single die.

$$E(X) = \sum_{i=1}^{6} i \cdot \frac{1}{6} = \frac{1}{6}(1+2+3+4+5+6) = \frac{21}{6} = \frac{7}{2}$$

Step 2: Apply linearity.

$$E(X+Y) = E(X) + E(Y) = \frac{7}{2} + \frac{7}{2} = 7$$

The expected sum of two dice is **7**

6. Indicator Random Variables and the Bernoulli Distribution

6.1 A Powerful Trick

Definition: Indicator Random Variable

For an event A, the indicator $\mathbf{1}_A$ equals:

$$\mathbf{1}_A = \begin{cases} 1 \text{ if } \ A \ \text{ occurs} \\ 0 \text{ otherwise} \end{cases}$$

6.1 A Powerful Trick

Definition: Indicator Random Variable

For an event A, the indicator $\mathbf{1}_A$ equals:

$$\mathbf{1}_A = \begin{cases} 1 \text{ if } & A & \text{occurs} \\ 0 \text{ otherwise} \end{cases}$$

Key property: $E(\mathbf{1}_A) = P(A)$

6.1 A Powerful Trick

Definition: Indicator Random Variable

For an event A, the indicator $\mathbf{1}_A$ equals:

$$\mathbf{1}_A = \begin{cases} 1 \text{ if } \ A \ \text{ occurs} \\ 0 \text{ otherwise} \end{cases}$$

Key property: $E(\mathbf{1}_A) = P(A)$

Proof: $E(\mathbf{1}_A) = 1 \cdot P(A) + 0 \cdot P(A^c) = P(A)$

6.2 The Bernoulli Distribution

Indicators are a special case of a named distribution!

6.2 The Bernoulli Distribution

Indicators are a special case of a named distribution!

Definition: Bernoulli Distribution

$X \sim \text{Bernoulli}(p)$ if X takes value 1 with probability p and 0 with probability $1 - p$.

$$P(X = 1) = p, \quad P(X = 0) = 1 - p$$

6.2 The Bernoulli Distribution

Indicators are a special case of a named distribution!

Definition: Bernoulli Distribution

$X \sim \text{Bernoulli}(p)$ if X takes value 1 with probability p and 0 with probability $1 - p$.

$$P(X = 1) = p, \quad P(X = 0) = 1 - p$$

Mean: $E(X) = p$

Variance: $\text{Var}(X) = p(1 - p)$

Examples:

- Coin flip: heads or tails
- Email: spam or not spam
- Patient: has disease or not

6.3 Example: Counting Heads

You flip a fair coin 100 times. Let X = number of heads.

6.3 Example: Counting Heads

You flip a fair coin 100 times. Let X = number of heads.

Trick: Write X as a sum of indicators (Bernoulli RVs)!

$$X = \mathbf{1}_1 + \mathbf{1}_2 + \ldots + \mathbf{1}_{100}$$

where $\mathbf{1}_i \sim \text{Bernoulli}\left(\frac{1}{2}\right)$.

6.3 Example: Counting Heads

You flip a fair coin 100 times. Let X = number of heads.

Trick: Write X as a sum of indicators (Bernoulli RVs)!

$$X = \mathbf{1}_1 + \mathbf{1}_2 + \ldots + \mathbf{1}_{100}$$

where $\mathbf{1}_i \sim \text{Bernoulli}\left(\frac{1}{2}\right)$.

$$E(X) = E(\mathbf{1}_1) + E(\mathbf{1}_2) + \ldots + E(\mathbf{1}_{100})$$

6.3 Example: Counting Heads

You flip a fair coin 100 times. Let X = number of heads.

Trick: Write X as a sum of indicators (Bernoulli RVs)!

$$X = \mathbf{1}_1 + \mathbf{1}_2 + \ldots + \mathbf{1}_{100}$$

where $\mathbf{1}_i \sim \text{Bernoulli}\left(\frac{1}{2}\right)$.

$$E(X) = E(\mathbf{1}_1) + E(\mathbf{1}_2) + \ldots + E(\mathbf{1}_{100})$$

$$= P(\text{heads}) + P(\text{heads}) + \ldots = 100 \cdot \frac{1}{2} = 50$$

7. The Geometric Distribution

7.1 Waiting for Success

How many trials until the first success?

7.1 Waiting for Success

How many trials until the first success?

Example: Coin Flipping

Flip a coin until you get heads. How many flips on average?

7.1 Waiting for Success

How many trials until the first success?

Example: Coin Flipping

Flip a coin until you get heads. How many flips on average?

Example: Rolling a Six

Roll a die until you get a 6. How many rolls on average?

7.2 The Geometric Distribution

Definition: Geometric Distribution

$X \sim \text{Geometric}(p)$ counts the number of trials until the first success, where each trial succeeds with probability p.

$$P(X = k) = (1 - p)^{k-1} p, \quad k = 1, 2, 3, \ldots$$

7.2 The Geometric Distribution

Definition: Geometric Distribution

$X \sim \text{Geometric}(p)$ counts the number of trials until the first success, where each trial succeeds with probability p.

$$P(X = k) = (1 - p)^{k-1} p, \quad k = 1, 2, 3, \ldots$$

Interpretation: $(k - 1)$ failures, then 1 success.

7.3 Geometric: Mean and Variance

7.3 Geometric: Mean and Variance

$$E(X) = \frac{1}{p}$$

7.3 Geometric: Mean and Variance

$$E(X) = \frac{1}{p}$$

$$\mathrm{Var}(X) = \frac{1-p}{p^2}$$

7.3 Geometric: Mean and Variance

$$E(X) = \frac{1}{p}$$

$$\mathrm{Var}(X) = \frac{1-p}{p^2}$$

Coin flip until heads:

$p = \frac{1}{2}$, so $E(X) = 2$ flips

Die roll until 6:

$p = \frac{1}{6}$, so $E(X) = 6$ rolls

7.3 Geometric: Mean and Variance

$$E(X) = \frac{1}{p}$$

$$\mathrm{Var}(X) = \frac{1-p}{p^2}$$

Coin flip until heads:

$p = \frac{1}{2}$, so $E(X) = 2$ flips

Die roll until 6:

$p = \frac{1}{6}$, so $E(X) = 6$ rolls

Note: Intuition: If you succeed 1 in p of the time, you need about $\frac{1}{p}$ tries.

7.4 The Coupon Collector Problem

Adapted from Grimmett & Stirzaker

A cereal company puts one of n different toys in each box (equally likely).

Question: On average, how many boxes until you collect all n toys?

7.4 The Coupon Collector Problem

Adapted from Grimmett & Stirzaker

A cereal company puts one of n different toys in each box (equally likely).

Question: On average, how many boxes until you collect all n toys?

This seems hard... but linearity makes it easy!

7.5 Coupon Collector: Solution

Let T_j = boxes needed to get the jth **new** toy (after having $j - 1$).

7.5 Coupon Collector: Solution

Let T_j = boxes needed to get the jth **new** toy (after having $j-1$).

When you have $j-1$ toys, probability of new toy = $\frac{n-j+1}{n}$

7.5 Coupon Collector: Solution

Let T_j = boxes needed to get the jth **new** toy (after having $j - 1$).

When you have $j - 1$ toys, probability of new toy = $\frac{n-j+1}{n}$

So T_j is geometric with $p = \frac{n-j+1}{n}$, meaning $E(T_j) = \frac{n}{n-j+1}$

7.5 Coupon Collector: Solution

Let T_j = boxes needed to get the jth **new** toy (after having $j - 1$).

When you have $j - 1$ toys, probability of new toy = $\frac{n-j+1}{n}$

So T_j is geometric with $p = \frac{n-j+1}{n}$, meaning $E(T_j) = \frac{n}{n-j+1}$

Total boxes: $T = T_1 + T_2 + \ldots + T_n$

7.5 Coupon Collector: Solution

Let T_j = boxes needed to get the jth **new** toy (after having $j - 1$).

When you have $j - 1$ toys, probability of new toy = $\frac{n-j+1}{n}$

So T_j is geometric with $p = \frac{n-j+1}{n}$, meaning $E(T_j) = \frac{n}{n-j+1}$

Total boxes: $T = T_1 + T_2 + \ldots + T_n$

$$E(T) = \sum_{j=1}^{n} \frac{n}{n - j + 1} = n \sum_{k=1}^{n} \frac{1}{k} \approx n \ln n$$

7.5 Coupon Collector: Solution

Let T_j = boxes needed to get the jth **new** toy (after having $j-1$).

When you have $j-1$ toys, probability of new toy = $\frac{n-j+1}{n}$

So T_j is geometric with $p = \frac{n-j+1}{n}$, meaning $E(T_j) = \frac{n}{n-j+1}$

Total boxes: $T = T_1 + T_2 + \ldots + T_n$

$$E(T) = \sum_{j=1}^{n} \frac{n}{n-j+1} = n \sum_{k=1}^{n} \frac{1}{k} \approx n \ln n$$

$\boxed{\text{For } n = 50 \text{ toys: } E(T) \approx 225 \text{ boxes!}}$

7.6 Try It Yourself: Matching Problem

Try it yourself

Talk to your neighbor and try to solve this problem.

n people each put their hat in a box. Hats are randomly redistributed.

Let X = number of people who get their own hat back.

Find $E(X)$.

7.6 Try It Yourself: Matching Problem

> ***Try it yourself***
>
> Talk to your neighbor and try to solve this problem.

n people each put their hat in a box. Hats are randomly redistributed.

Let X = number of people who get their own hat back.

Find $E(X)$.

Hint: Let $\mathbf{1}_i = 1$ if person i gets their own hat.

7.6 Try It Yourself: Matching Problem

> ***Try it yourself***
>
> Talk to your neighbor and try to solve this problem.

n people each put their hat in a box. Hats are randomly redistributed.

Let X = number of people who get their own hat back.

Find $E(X)$.

Hint: Let $\mathbf{1}_i = 1$ if person i gets their own hat.

$$E(X) = \sum_{i=1}^{n} E(\mathbf{1}_i) = \sum_{i=1}^{n} P(\text{person } \ i \ \text{ gets own hat}) = n \cdot \frac{1}{n} = 1$$

7.6 Try It Yourself: Matching Problem

Try it yourself

Talk to your neighbor and try to solve this problem.

n people each put their hat in a box. Hats are randomly redistributed.

Let X = number of people who get their own hat back.

Find $E(X)$.

Hint: Let $\mathbf{1}_i = 1$ if person i gets their own hat.

$$E(X) = \sum_{i=1}^{n} E(\mathbf{1}_i) = \sum_{i=1}^{n} P(\text{person } \; i \; \text{ gets own hat}) = n \cdot \frac{1}{n} = 1$$

On average, exactly 1 person gets their own hat — regardless of n!

8. Variance Properties (Review)

8.1 Recall: Variance

Recall:

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

8.1 Recall: Variance

Recall:

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

Scaling:

$\mathrm{Var}(aX) = a^2\ \mathrm{Var}(X)$

$\sigma(aX) = |a|\ \sigma(X)$

Shift:

$\mathrm{Var}(X + b) = \mathrm{Var}(X)$

Adding a constant doesn't change spread!

8.1 Recall: Variance

Recall:

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

Scaling:

$\mathrm{Var}(aX) = a^2\ \mathrm{Var}(X)$

$\sigma(aX) = |a|\ \sigma(X)$

Shift:

$\mathrm{Var}(X + b) = \mathrm{Var}(X)$

Adding a constant doesn't change spread!

Combined: $\mathrm{Var}(aX + b) = a^2\ \mathrm{Var}(X)$

8.2 Why $\mathrm{Var}(X + b) = \mathrm{Var}(X)$?

8.2 Why Var(X + b) = Var(X)?

$$\mathrm{Var}(X + b) = E\big[(X + b - E(X + b))^2\big]$$

8.2 Why Var(X + b) = Var(X)?

$$\mathrm{Var}(X + b) = E\left[\left(X + b - E(X + b)\right)^2\right]$$

$$= E\left[\left(X + b - E(X) - b\right)^2\right]$$

8.2 Why Var(X + b) = Var(X)?

$$\mathrm{Var}(X+b) = E\big[(X+b-E(X+b))^2\big]$$

$$= E\big[(X+b-E(X)-b)^2\big]$$

$$= E\big[(X-E(X))^2\big] = \mathrm{Var}(X)$$

8.2 Why Var(X + b) = Var(X)?

$$\mathrm{Var}(X+b) = E\left[(X+b-E(X+b))^2\right]$$

$$= E\left[(X+b-E(X)-b)^2\right]$$

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

Intuition: Shifting all values by b moves the mean by b too.

The **distances** from the mean don't change!

9. Higher-Order Moments

9.1 Moments: A Generalization

Definition: nth Moment

The nth moment of X is $E(X^n)$

The nth **central** moment is $E[(X-\mu)^n]$ where $\mu = E(X)$

9.1 Moments: A Generalization

Definition: nth Moment

The nth moment of X is $E(X^n)$

The nth **central** moment is $E[(X-\mu)^n]$ where $\mu = E(X)$

- 1st moment: **Mean** μ
- 2nd central moment: **Variance** σ^2
- 3rd central moment: related to **Skewness**
- 4th central moment: related to **Kurtosis**

9.2 Skewness: Asymmetry

Definition: Skewness

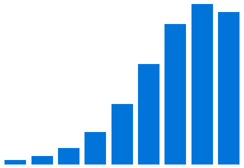
$$\text{Skew}(X) = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{E[(X-\mu)^3]}{\sigma^3}$$

9.2 Skewness: Asymmetry

Definition: Skewness

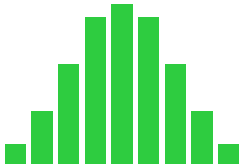
$$\text{Skew}(X) = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{E[(X-\mu)^3]}{\sigma^3}$$

Negative skew



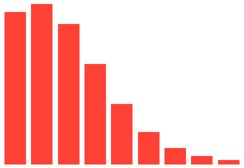
Left tail longer

Symmetric



Skew ≈ 0

Positive skew



Right tail longer

9.3 Kurtosis: Tail Heaviness

Definition: Kurtosis

$$\mathrm{Kurt}(X) = E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] = \frac{E[(X-\mu)^4]}{\sigma^4}$$

9.3 Kurtosis: Tail Heaviness

Definition: Kurtosis

$$\mathrm{Kurt}(X) = E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] = \frac{E[(X-\mu)^4]}{\sigma^4}$$

Normal distribution has kurtosis = 3.

Excess kurtosis = $\mathrm{Kurt}(X) - 3$

9.3 Kurtosis: Tail Heaviness

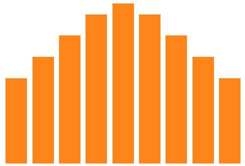
Definition: Kurtosis

$$\mathrm{Kurt}(X) = E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] = \frac{E[(X-\mu)^4]}{\sigma^4}$$

Normal distribution has kurtosis = 3.

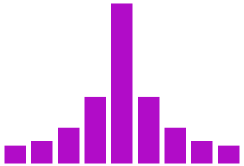
Excess kurtosis = $\mathrm{Kurt}(X) - 3$

Low kurtosis (platykurtic)



Thin tails, flat peak

High kurtosis (leptokurtic)



Heavy tails, sharp peak

9.4 Why Do Moments Matter?

9.4 Why Do Moments Matter?

1. **Descriptive statistics**: Summarize distribution shape beyond mean/variance

9.4 Why Do Moments Matter?

1. **Descriptive statistics**: Summarize distribution shape beyond mean/variance
2. **Risk analysis**: Skewness and kurtosis matter for financial returns

9.4 Why Do Moments Matter?

1. **Descriptive statistics**: Summarize distribution shape beyond mean/variance
2. **Risk analysis**: Skewness and kurtosis matter for financial returns
3. **Goodness of fit**: Compare empirical vs theoretical distributions

9.4 Why Do Moments Matter?

1. **Descriptive statistics**: Summarize distribution shape beyond mean/variance
2. **Risk analysis**: Skewness and kurtosis matter for financial returns
3. **Goodness of fit**: Compare empirical vs theoretical distributions
4. **Moment generating functions**: $M_{X(t)} = E(e^{tX})$ encodes all moments

9.4 Why Do Moments Matter?

1. **Descriptive statistics**: Summarize distribution shape beyond mean/variance
2. **Risk analysis**: Skewness and kurtosis matter for financial returns
3. **Goodness of fit**: Compare empirical vs theoretical distributions
4. **Moment generating functions**: $M_{X(t)} = E\left(e^{tX}\right)$ encodes all moments

Note: In practice, 3rd and 4th moments are most commonly used. Higher moments are rarely estimated reliably from data.

9.5 Recap

Today we covered:

- $E(aX + b) = aE(X) + b$; LOTUS: $E(g(X)) = \sum_x g(x)p(x)$
- Uniform$\{a, ..., b\}$: $E(X) = \frac{a+b}{2}$; die roll is Uniform$\{1, ..., 6\}$
- Bernoulli(p): $E(X) = p$; Geometric(p): $E(X) = 1/p$
- Indicators + linearity: powerful for counting problems
- Higher moments: skewness (asymmetry), kurtosis (tail weight)