Linearity of Expectation, LOTUS and Variance

CSE 312 Spring 26

Lecture 10

Review Random Variables and Associated Events

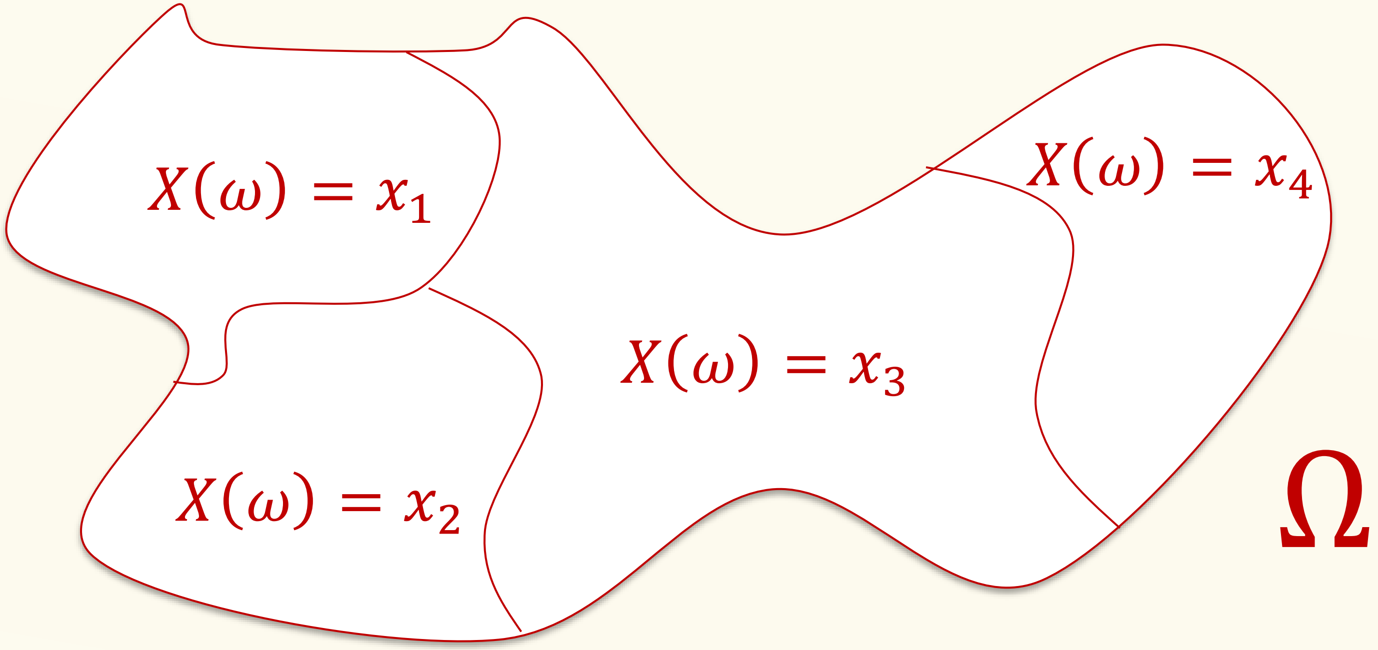
Definition. A **random variable (RV)** for a probability space (Ω, P) is a function $X: \Omega \to \mathbb{R}$.

The set of values that X can take on is its *range/support*: $\mathbf{\Omega}_X$

$\{X = x_i\} = \{\omega \in \Omega \mid X(\omega) = x_i\}$

Random variables **partition** the sample space.

$\Sigma_{x \in \mathbf{\Omega}_X} P(X = x) = 1$



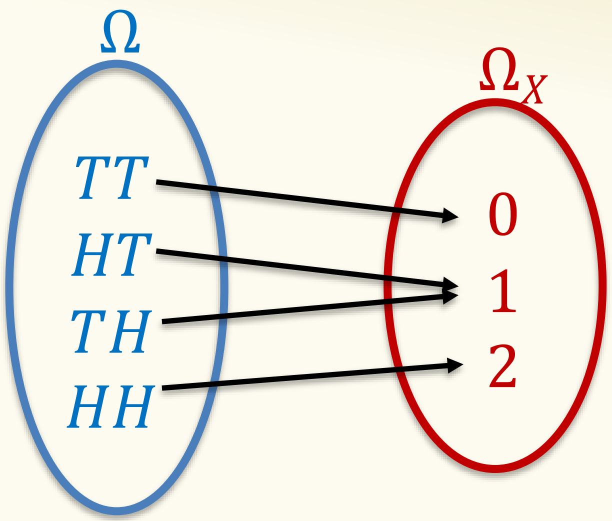
Example

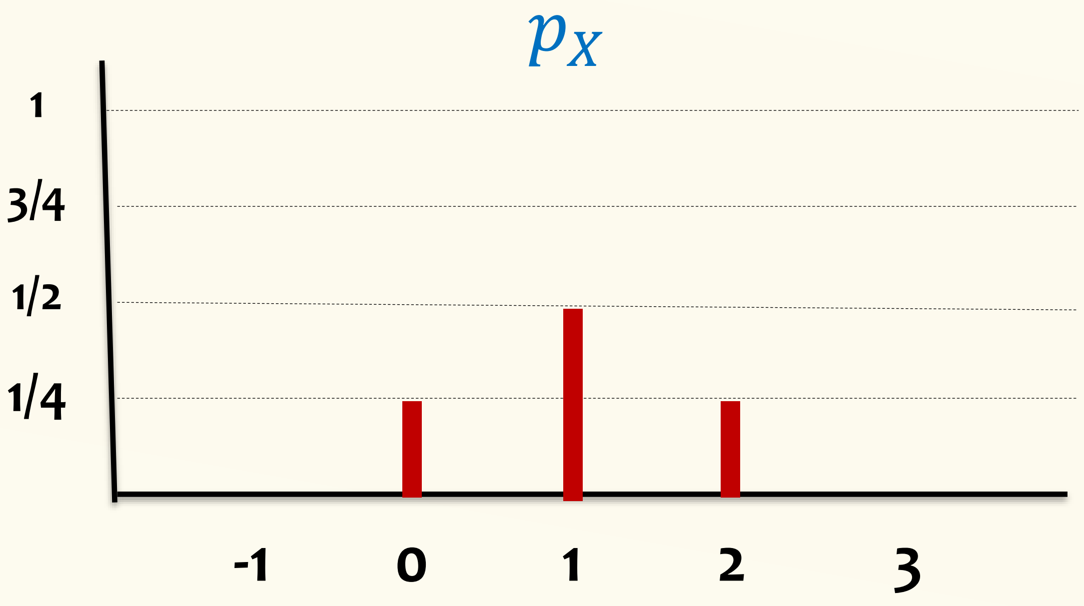
Two fair coin flips

$$\Omega = \{\mathrm{TT}, \mathrm{HT}, \mathrm{TH}, \mathrm{HH}\}$$

X = number of heads

$$\Omega_X = \{0,1,2\}$$





$$p_X(0) = \Pr(X = 0) = \Pr(\mathrm{TT})$$

$$p_X(1) = \Pr(X = 1) = \Pr(\{\mathrm{TH}, \mathrm{HT}\})$$

$$p_X(2) = \Pr(X = 2) = \Pr(\mathrm{HH})$$

Example – Two fair independent coin flips

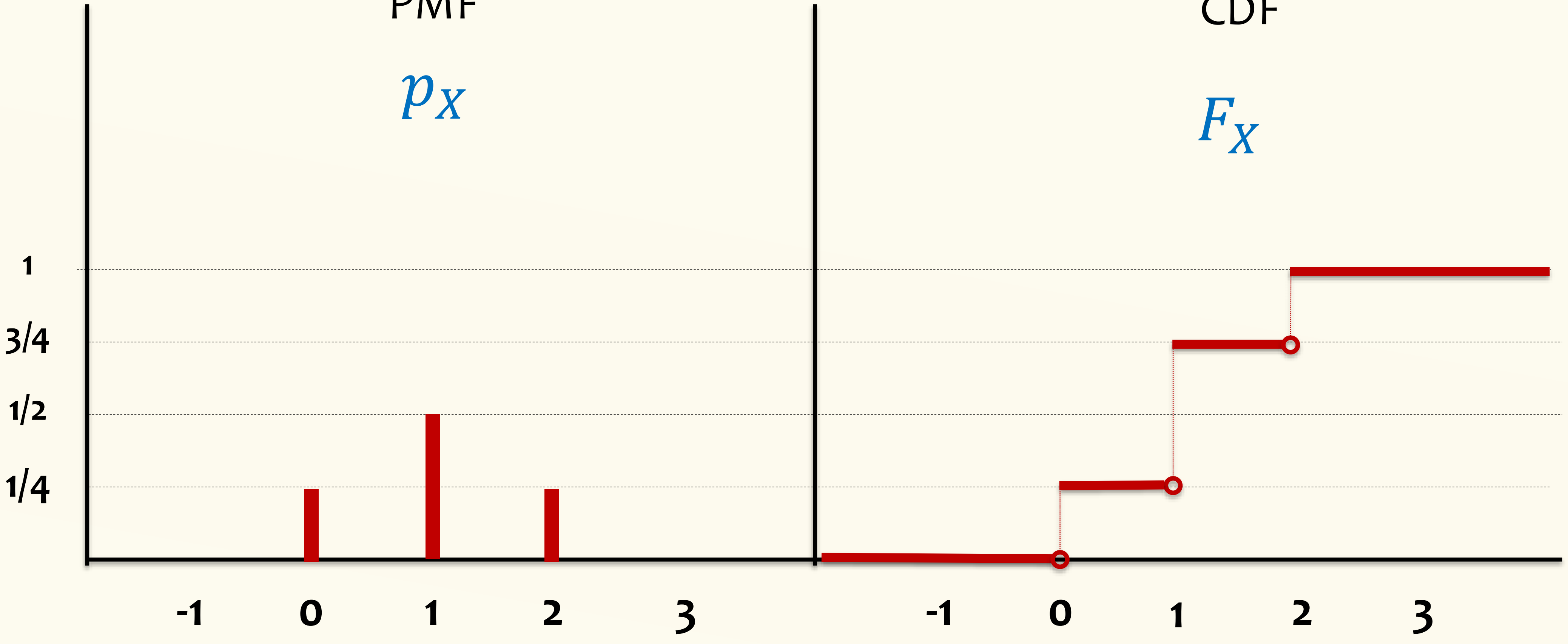
X = number of heads

Probability Mass Function
PMF

p_X

Cumulative Distribution Function
CDF

F_X



Review PMF and CDF

Definitions:

For a RV $X: \Omega \to \mathbb{R}$, the probability mass function (pmf) of X specifies, for any real number x, the probability that $X = x$

$$p_X(x) = P(X = x) = P(\{\omega \in \Omega \mid X(\omega) = x\})$$

$$\sum_{x \in \Omega_X} p_X(x) = 1$$

For a RV $X: \Omega \to \mathbb{R}$, the cumulative distribution function (cdf) of X specifies, for any real number x, the probability that $X \leq x$

$$F_X(x) = P(X \leq x)$$

Review Expected Value of a Random Variable

Definition. Given a discrete RV $X: \Omega \to \mathbb{R}$, the **expectation** or **expected value** or **mean** of X is

$$\mathbb{E}[X] = \sum_{x \in \Omega_X} x \cdot P(X = x) \qquad = \sum_{x \in \Omega_X} x \cdot p_X(x)$$

Intuition: "Weighted average" of the possible outcomes (weighted by probability)

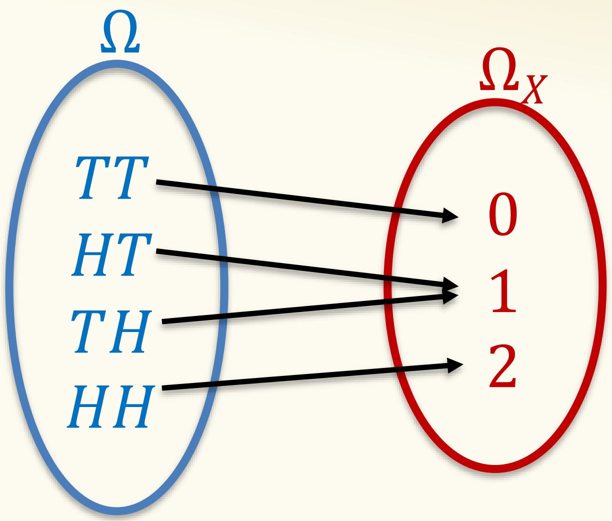
Example calculation

Two fair coin flips

$$\Omega = \{\mathrm{TT}, \mathrm{HT}, \mathrm{TH}, \mathrm{HH}\}$$

$$X = \text{number of heads}$$

$$\Omega_X = \{0,1,2\}$$

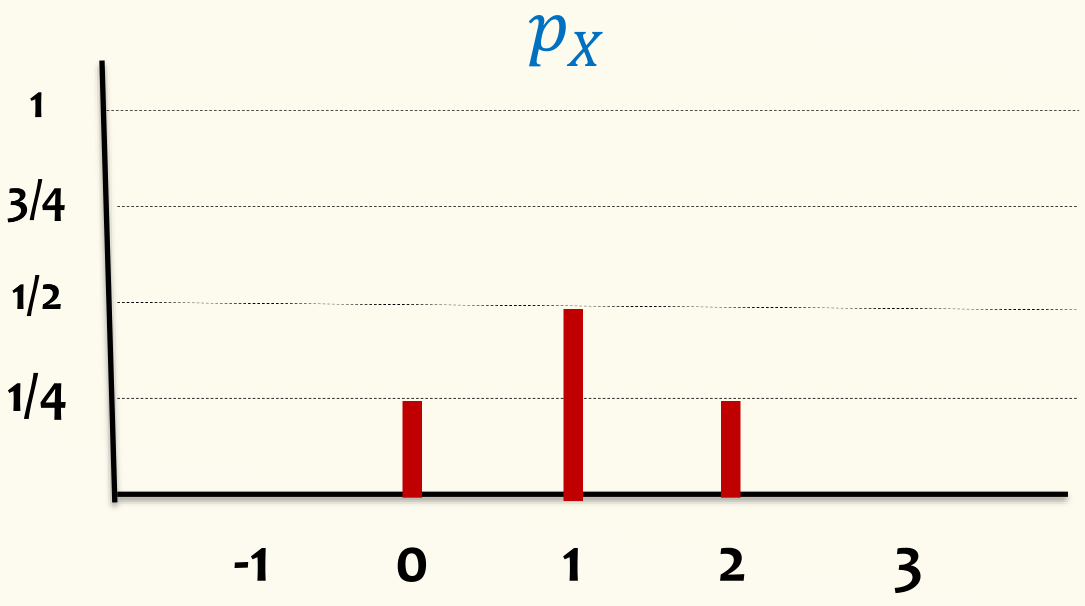


$$\mathbb{E}[X] = \sum_{x \in \Omega_X} x \cdot P(X = x)$$

$$p_X(0) = \Pr(X = 0) = \Pr(\mathrm{TT})$$

$$p_X(1) = \Pr(X = 1) = \Pr(\{\mathrm{TH}, \mathrm{HT}\})$$

$$p_X(2) = \Pr(X = 2) = \Pr(\mathrm{HH})$$



$$\mathbb{E}[X] = 0 \cdot p_X(0) + 1 \cdot p_X(1) + 2 \cdot p_X(2)$$

$$= 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 1$$

Expected Value of a Random Variable

Definition. Given a discrete RV $X: \Omega \to \mathbb{R}$, the **expectation** or **expected value** or **mean** of X is

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \cdot P(\omega)$$

or equivalently

$$\mathbb{E}[X] = \sum_{x \in \Omega_X} x \cdot P(X = x) \quad = \sum_{x \in \Omega_X} x \cdot p_X(x)$$

Intuition: "Weighted average" of the possible outcomes (weighted by probability)

Expectation

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \cdot P(\omega)$$

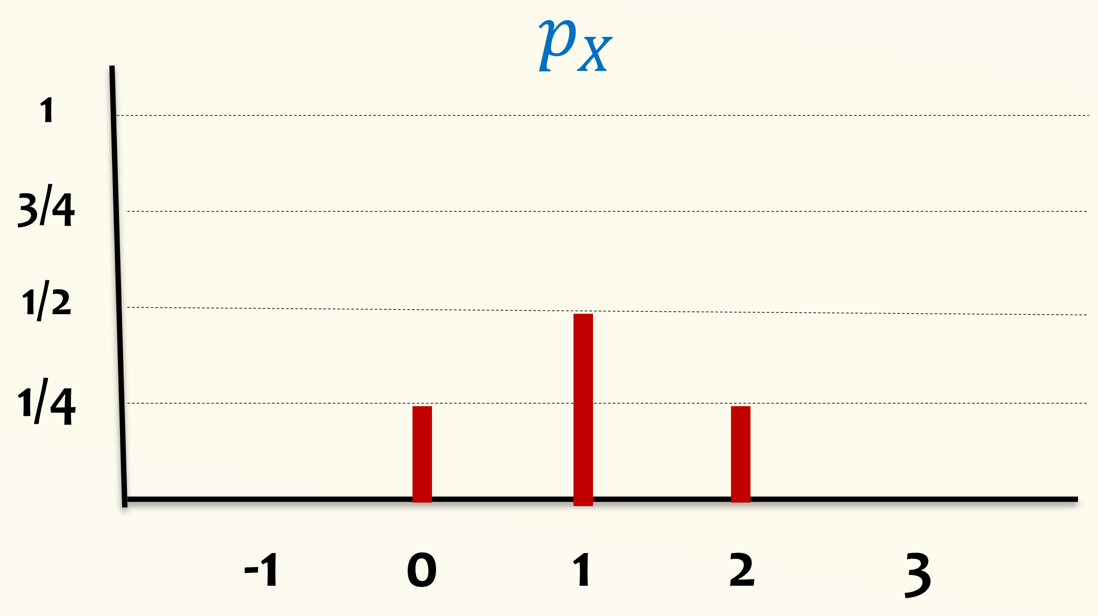
$$\mathbb{E}[X] = \sum_{x \in \Omega_X} x \cdot P(X = x)$$

Example. Two fair coin flips

$\Omega = \{\mathrm{TT}, \mathrm{HT}, \mathrm{TH}, \mathrm{HH}\}$

$X =$ number of heads

What is $\mathbb{E}[X]$?



$$\mathbb{E}[X] = X(TT)\,P(TT) + X(HT)\,P(HT) + X(TH)\,P(TH) + X(HH)\,P(HH)$$

$$= 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{4} + 1 \cdot \frac{1}{4} + 2 \cdot \frac{1}{4} = 1$$

$$\mathbb{E}[X] = 0 \cdot p_X(0) + 1 \cdot p_X(1) + 2 \cdot p_X(2)$$

$$= 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = \frac{1}{2} + \frac{1}{2} = 1$$

Recap Linearity of Expectation

Theorem. For any two random variables X and Y

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y].$$

Or, more generally: For any random variables $X_1, \dots, X_n$,

$$\mathbb{E}[X_1 + \cdots + X_n] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n].$$

Theorem. For any random variables X, and constants a and b

$$\mathbb{E}[aX + b] = a \cdot \mathbb{E}[X] + b.$$

Recap Linearity of Expectation (more)

Theorem. For any two random variables X and Y

$$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y].$$

Or, more generally: For any random variables $X_1, \dots, X_n$,

$$\mathbb{E}[X_1 + \cdots + X_n] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n].$$

Theorem. For any random variables X, and constants a and b

$$\mathbb{E}[aX + b] = a \cdot \mathbb{E}[X] + b.$$

Theorem. For any random variables $X_1, \dots, X_n$, and real numbers $a_1, \dots, a_n \in \mathbb{R}$,

$$\mathbb{E}[a_1X_1 + \cdots + a_nX_n] = a_1\mathbb{E}[X_1] + \cdots + a_n\mathbb{E}[X_n].$$

Agenda

- Recap
- Linearity of expectation
- LOTUS
- Variance

Pairs with the same birthday

- In a class of m students, on average how many pairs of people have the same birthday (assuming 365 equally likely birthdays)?
- Each person's birthday is equally likely to be any of the 365 possibilities and different people's bdays are independent.

Pairs with the same birthday - LOE

- In a class of m students, on average how many pairs of people have the same birthday (assuming 365 equally likely birthdays)?
- Each person's birthday is equally likely to be any of the 365 possibilities and different people's bdays are independent.

Decompose: Find the right way to decompose the random variable into sum of simple random variables

$$X = X_1 + \cdots + X_n$$

LOE: Apply linearity of expectation.

$$\mathbb{E}[X] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n].$$

Conquer: Compute the expectation of each X_i and sum!

Pairs with the same birthday – LOE (2)

- In a class of m students, let X be the number of pairs of people with the same birthday (assuming 365 equally likely birthdays)?
- What is $\mathbb{E}[X]$?

Decompose: Indicator events involve **pairs** of students (i, j) for $i \neq j$

$X_{ij} = 1$ iff students i and j have the same birthday

LOE: $\binom{m}{2}$ indicator variables X_{ij}

Conquer: $\mathbb{E}[X_{ij}] = \frac{1}{365}$ so total expectation is $\frac{\binom{m}{2}}{365}$ pairs

Agenda (2)

- Recap
- Linearity of expectation
- LOTUS
- Variance

Constructing new random variables from old ones.

1. $X \rightarrow Y = 2X - 1$
2. $X_1, X_2, \dots, X_n \rightarrow Z = X_1 + \cdots + X_n$
3. $X_1, X_2 \rightarrow W = 5X_1 - 3X_2 + 2$
4. $X \rightarrow U = X^3 \bmod 4$

How do we compute the expectation of the new r.v.s, given the expectations of the old r.v.s?

Example of constructing new r.v.

$\Pr(\omega)$	ω	X	$Y = X^2 \bmod 4$
1/4	ω_1	0	
1/6	ω_2	1	
1/4	ω_3	2	
1/3	ω_4	2	

$$\mathbb{E}[Y] = \sum_{y \in \Omega_Y} y \cdot P(Y = y)$$

$$\mathbb{E}[g(X)] = \sum_{x \in \Omega_X} g(x) \cdot P(X = x)$$

Expected Value of $g(X)$

Definition. Given a discrete RV $X: \Omega \to \mathbb{R}$, the **expectation** or **expected value** or **mean** of $Y = g(X)$ is

$$\mathbb{E}[g(X)] = \sum_{\omega \in \Omega} g\big(X(\omega)\big) \cdot P(\omega)$$

or equivalently

$$\mathbb{E}[g(X)] = \sum_{x \in \Omega_X} g(x) \cdot P(X = x) \quad = \sum_{x \in \Omega_X} g(x) \cdot p_X(x)$$

Also known as **LOTUS:** "Law of the unconscious statistician

Linearity is special!

$$\mathbb{E}[g(X)] = \sum_{x \in \Omega_X} g(x) \cdot P(X = x)$$

In general $\mathbb{E}[g(X)] \neq g\big(\mathbb{E}(X)\big)$

E.g., $X = \begin{cases} +1 \text{ with prob } 1/2 \\ -1 \text{ with prob } 1/2 \end{cases}$

$\mathbb{E}[X] =$

$\mathbb{E}[X^2] =$

$\mathbb{E}[X]^2 =$

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

Agenda (3)

- Recap
- Linearity of expectation
- LOTUS
- Variance

A game

$$\mathbb{E}[X] = \sum_{x \in \Omega_X} x \cdot P(X = x)$$

Game 1: *In every round, you win $2 with probability 1/3, lose $1 with probability 2/3.*

W_1 = payoff in a round of Game 1

$P(W_1 = 2) = \frac{1}{3}$, $P(W_1 = -1) = \frac{2}{3}$

Which game would you rather play?

$$\mathbb{E}[X] = \sum_{x \in \Omega_X} x \cdot P(X = x)$$

Game 1: *In every round, you win \$2 with probability 1/3, lose \$1 with probability 2/3.*

W_1 = payoff in a round of Game 1

$$P(W_1 = 2) = \frac{1}{3}, P(W_1 = -1) = \frac{2}{3}$$

$$\mathbb{E}[W_1] = 0$$

Game 2: *In every round, you win \$10 with probability 1/3, lose \$5 with probability 2/3.*

W_2 = payoff in a round of Game 2

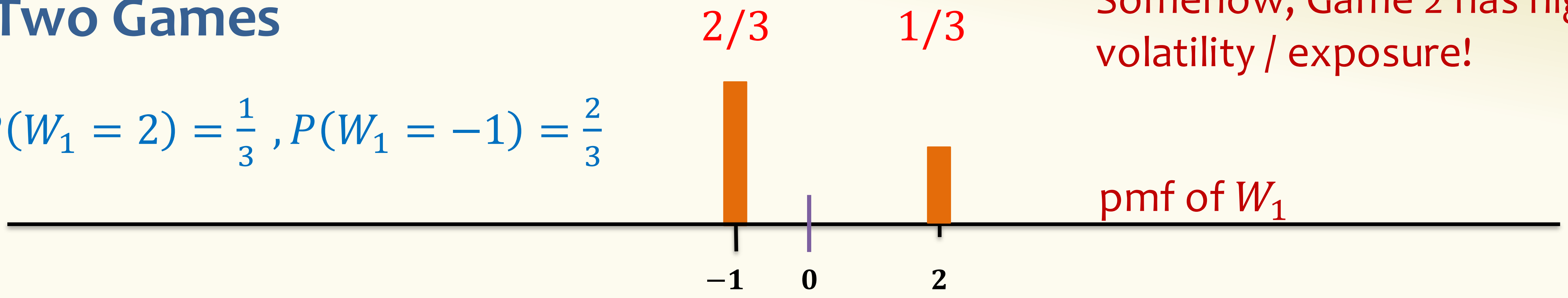
$$P(W_2 = 10) = \frac{1}{3}, P(W_2 = -5) = \frac{2}{3}$$

$$\mathbb{E}[W_2] = 0$$

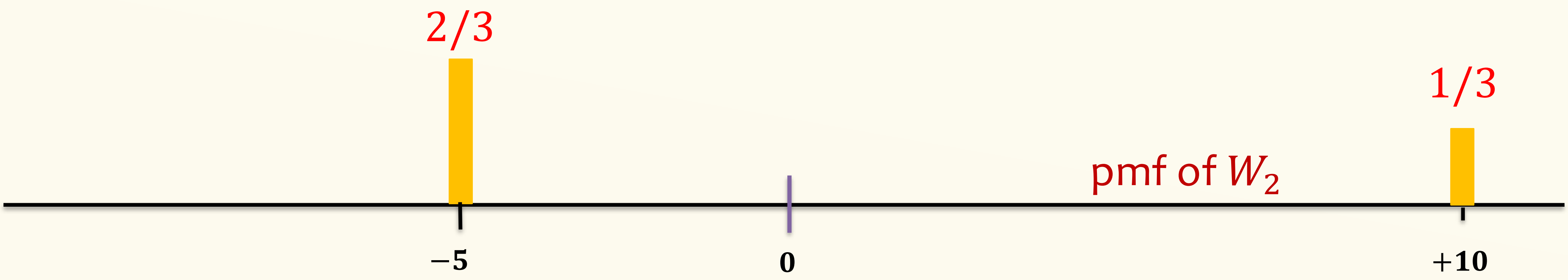
Two Games

Somehow, Game 2 has higher volatility / exposure!

$P(W_1 = 2) = \frac{1}{3}, P(W_1 = -1) = \frac{2}{3}$



$P(W_2 = 10) = \frac{1}{3}, P(W_2 = -5) = \frac{2}{3}$



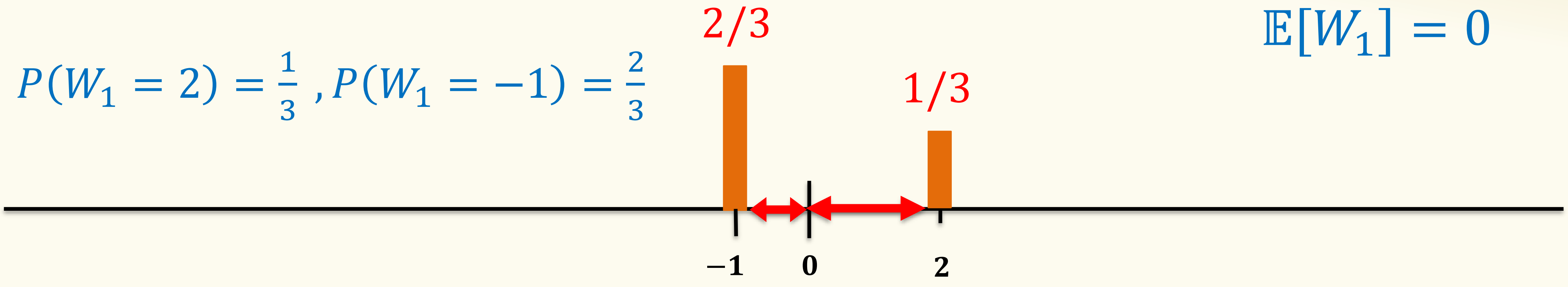
Same expectation, but clearly a very different distribution.

We want to capture the difference – New concept: **Variance**

Variance (Intuition, First Try)

$P(W_1 = 2) = \frac{1}{3}, P(W_1 = -1) = \frac{2}{3}$

$\mathbb{E}[W_1] = 0$

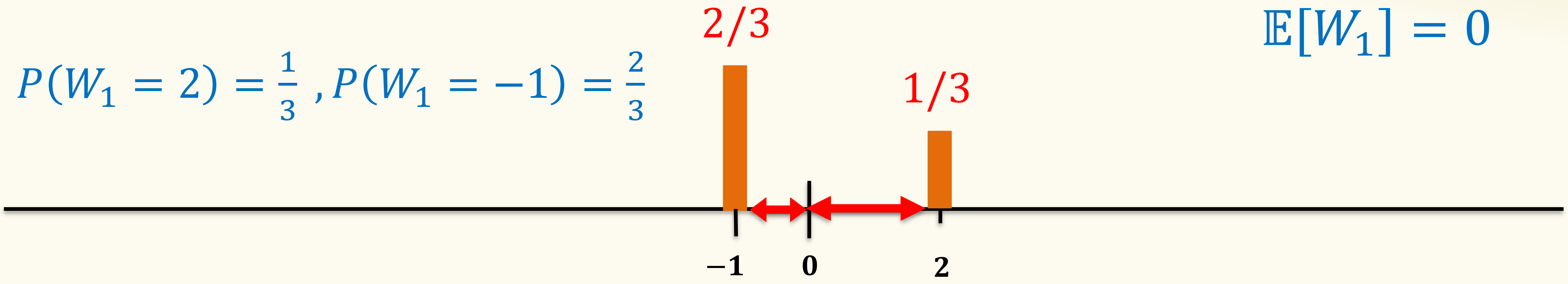


New quantity (random variable): How far from the expectation?

$W_1 - \mathbb{E}[W_1]$

Variance (Intuition, First Try with calculation)

$P(W_1 = 2) = \frac{1}{3}, P(W_1 = -1) = \frac{2}{3}$



$\mathbb{E}[W_1] = 0$

New quantity (random variable): How far from the expectation?

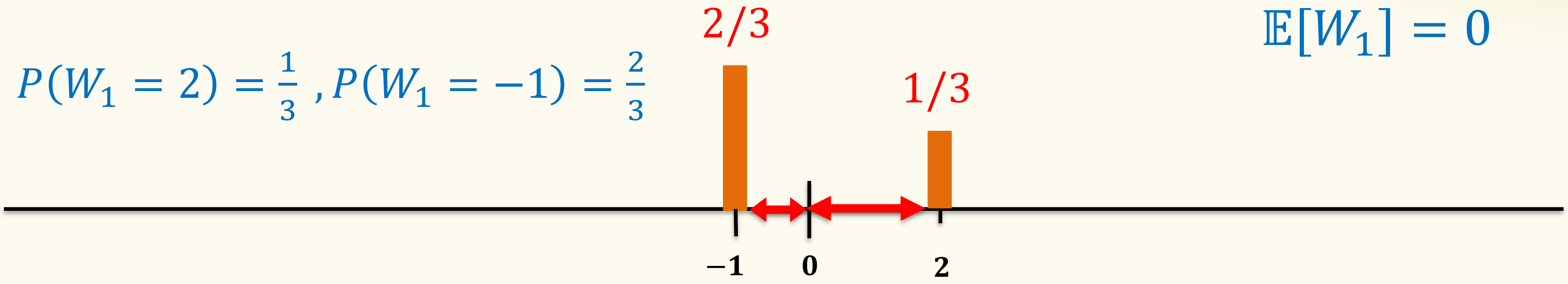
$W_1 - \mathbb{E}[W_1]$

$$\mathbb{E}[W_1 - \mathbb{E}[W_1]]$$
$$= \mathbb{E}[W_1] - \mathbb{E}[\mathbb{E}[W_1]]$$
$$= \mathbb{E}[W_1] - \mathbb{E}[W_1]$$
$$= 0$$

Variance (Intuition, Better Try)

$P(W_1 = 2) = \frac{1}{3}, P(W_1 = -1) = \frac{2}{3}$

$\mathbb{E}[W_1] = 0$



New quantity (random variable): How far from the expectation?

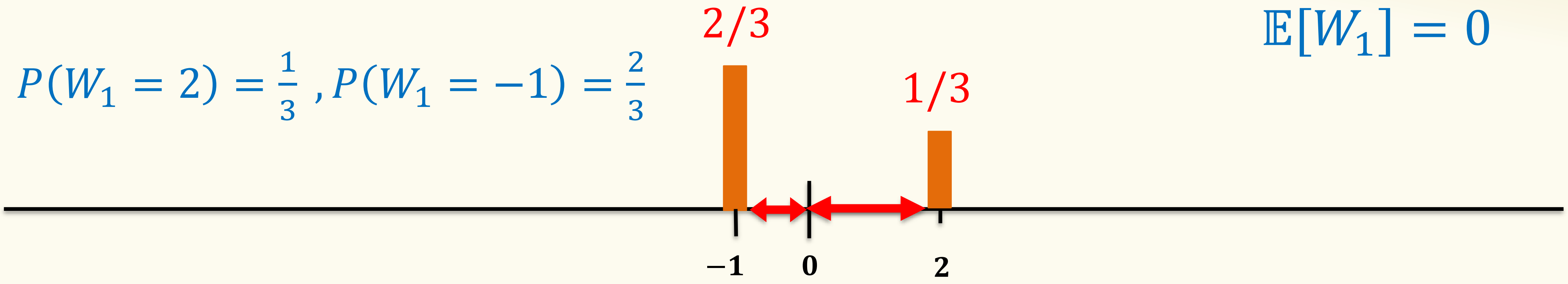
$W_1 - \mathbb{E}[W_1]$

$\mathbb{E}[(W_1 - \mathbb{E}[W_1])^2]$

$\mathbb{E}[g(X)] = \sum_{x \in \Omega_X} g(x) \cdot P(X = x)$

Variance (Intuition, Better Try with calculation)

$P(W_1 = 2) = \frac{1}{3}, P(W_1 = -1) = \frac{2}{3}$



$\mathbb{E}[W_1] = 0$

New quantity (random variable): How far from the expectation?

$W_1 - \mathbb{E}[W_1]$

$$\mathbb{E}[(W_1 - \mathbb{E}[W_1])^2]$$

$$= \frac{2}{3} \cdot 1 + \frac{1}{3} \cdot 4$$

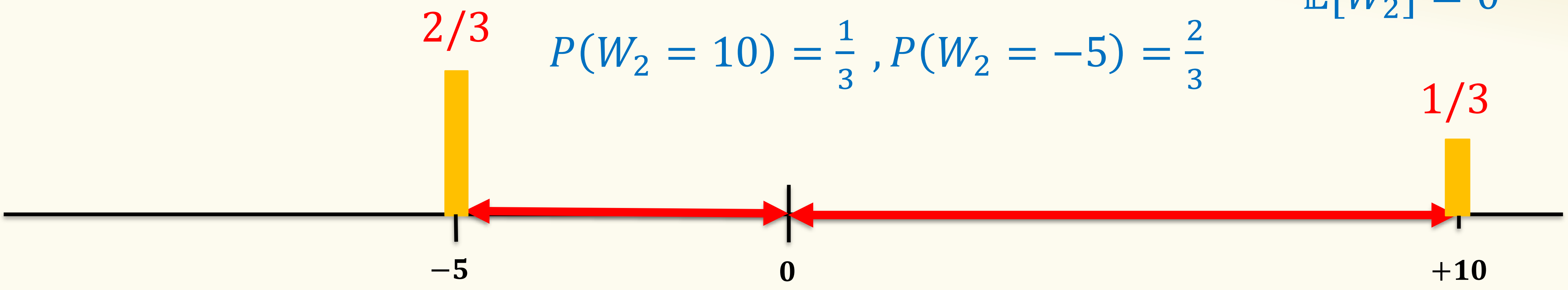
$$= 2$$

$\mathbb{E}[g(X)] = \sum_{x \in \Omega_X} g(x) \cdot P(X = x)$

Variance (Intuition, Better Try – W_2)

$\mathbb{E}[W_2] = 0$

$P(W_2 = 10) = \frac{1}{3}$, $P(W_2 = -5) = \frac{2}{3}$



A better quantity (random variable): How far from the expectation?

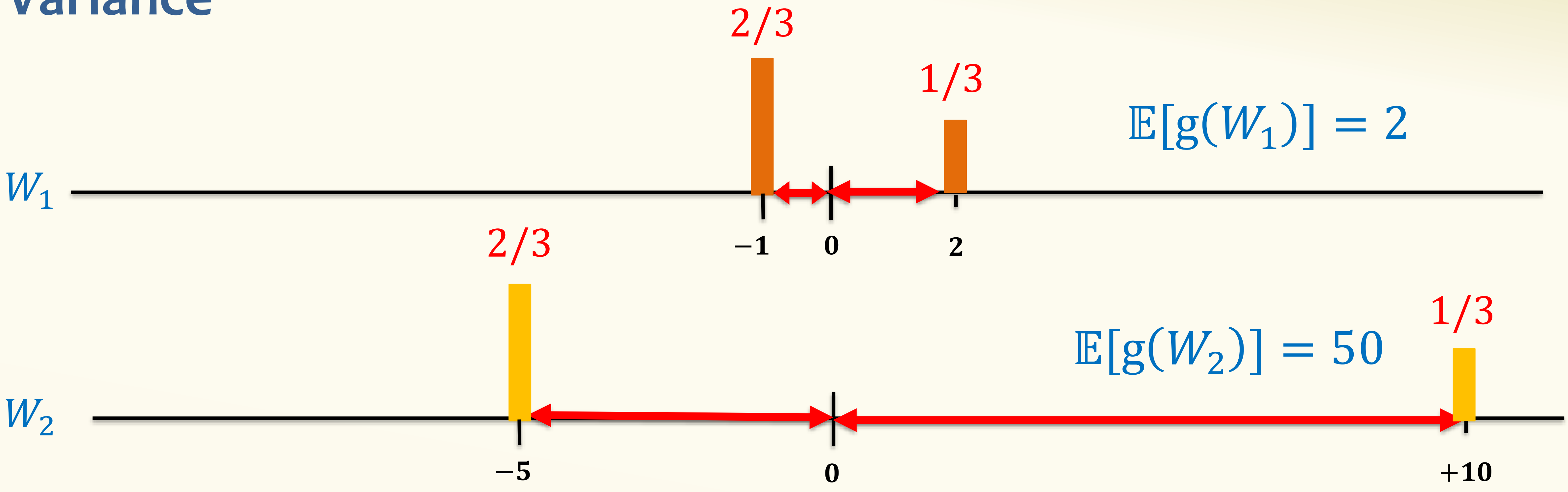
$$\mathbb{E}[(W_2 - \mathbb{E}[W_2])^2]$$

$$= \frac{2}{3} \cdot 25 + \frac{1}{3} \cdot 100$$

$$= 50$$

$$\mathbb{E}[g(X)] = \sum_{x \in \Omega_X} g(x) \cdot P(X = x)$$

Variance



We say that W_2 has **"higher variance"** than W_1.

$$g(W) = (W - \mathbb{E}[W])^2$$

Variance - summary

Definition. The **variance** of a (discrete) RV X is

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \sum_{x \in \Omega_X} p_X(x) \cdot (x - \mathbb{E}[X])^2$$

Recall $\mathbb{E}[X]$ is a **constant**, not a random variable itself.

Standard deviation: $\sigma(X) = \sqrt{\mathrm{Var}(X)}$

Intuition: Variance (or standard deviation) is a quantity that measures, in expectation, how "far" the random variable is from its expectation.

Variance – Example 1

$$\mathbb{E}[g(X)] = \sum_{x \in \Omega_X} g(x) \cdot P(X = x)$$

X fair die

- $P(X = 1) = \cdots = P(X = 6) = 1/6$
- $\mathbb{E}[X] = 3.5$

$$\mathrm{Var}(X) = \sum_{x \in \Omega \mathrm{X}} P(X = x) \cdot (x - \mathbb{E}[X])^2$$

Variance – Example 1 (with calculation)

X fair die

- $P(X = 1) = \cdots = P(X = 6) = 1/6$
- $\mathbb{E}[X] = 3.5$

$$\mathrm{Var(X)} = \sum_x P(X = x) \cdot (x - \mathbb{E}[X])^2$$

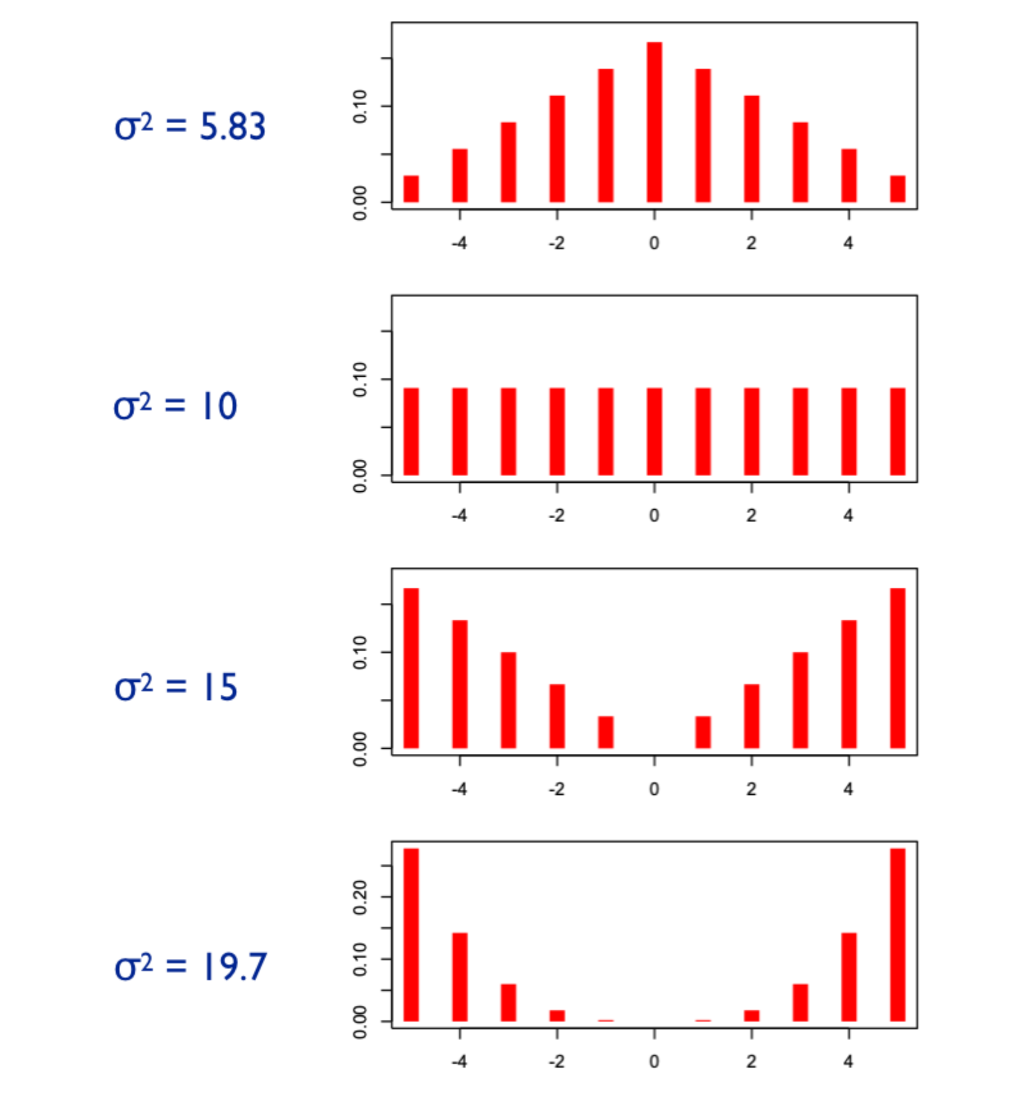
$$= \frac{1}{6}[(1 - 3.5)^2 + (2 - 3.5)^2 + (3 - 3.5)^2 + (4 - 3.5)^2 + (5 - 3.5)^2 + (6 - 3.5)^2]$$

$$= \frac{2}{6}[2.5^2 + 1.5^2 + 0.5^2] = \frac{2}{6}\left[\frac{25}{4} + \frac{9}{4} + \frac{1}{4}\right] = \frac{35}{12} \approx 2.91677 \ldots$$

Variance in Pictures

Captures how much "spread' there is in a pmf

All pmfs have same expectation = 0



Agenda (4)

- Variance
- Properties of Variance

Variance – Properties

$$\mathbb{E}[X] = \sum_{x \in \Omega_X} x \cdot P(X = x)$$

$$\mathbb{E}[g(X)] = \sum_{x \in \Omega_X} g(x) \cdot P(X = x)$$

Definition. The **variance** of a (discrete) RV X is

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \sum_x p_X(x) \cdot (x - \mathbb{E}[X])^2$$

Theorem. For any $a, b \in \mathbb{R}$, $\mathrm{Var}(a \cdot X + b) = a^2 \cdot \mathrm{Var}(X)$

Variance – Properties (2)

$$\mathbb{E}[X] = \sum_{x \in \Omega_X} x \cdot P(X = x)$$

$$\mathbb{E}[g(X)] = \sum_{x \in \Omega_X} g(x) \cdot P(X = x)$$

Definition. The **variance** of a (discrete) RV X is

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \sum_x p_X(x) \cdot (x - \mathbb{E}[X])^2$$

Theorem. For any $a, b \in \mathbb{R}$, $\mathrm{Var}(a \cdot X + b) = a^2 \cdot \mathrm{Var}(X)$

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

Variance (3)

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

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

Recall $\mathbb{E}[X]$ is a **constant**

$= \mathbb{E}[X^2 - 2\mathbb{E}[X] \cdot X + \mathbb{E}[X]^2]$

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

(linearity of expectation!)

$= \mathbb{E}[X^2] - \mathbb{E}[X]^2$

$\mathbb{E}[X^2]$ and $\mathbb{E}[X]^2$ are different !

Variance – Example 1 – calculated the other way

X fair die

- $\mathbb{P}(X = 1) = \cdots = \mathbb{P}(X = 6) = 1/6$
- $\mathbb{E}[X] = \frac{21}{6}$
- $\mathbb{E}[X^2] = \frac{91}{6}$

$$\mathrm{Var}(\mathrm{X}) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{91}{6} - \left(\frac{21}{6}\right)^2 = \frac{105}{36} \approx 2.91677$$

Variance of Indicator Random Variables

Suppose that X_A is an indicator RV for event A with $P(A) = p$ so

$$\mathbb{E}[X_A] = P(A) = p$$

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

Variance of Indicator Random Variables - calculation

Suppose that X_A is an indicator RV for event A with $P(A) = p$ so

$$\mathbb{E}[X_A] = P(A) = p$$

Since X_A only takes on values 0 and 1, we always have $X_A^2 = X_A$ so

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

In General, $\mathrm{Var}(X+Y) \neq \mathrm{Var}(X) + \mathrm{Var}(Y)$

$$\mathbb{E}[\boldsymbol{X}] = \sum_{x \in \Omega_X} \boldsymbol{x} \cdot \boldsymbol{P}(\boldsymbol{X} = \boldsymbol{x})$$

Let's see a counter-example:

- Let X be a r.v. with pmf $P(X=1) = P(X=-1) = 1/2$
  - What is $\mathbb{E}[X]$ and $\mathrm{Var}(X)$?
- Let $Y = -X$
  - What is $\mathbb{E}[Y]$ and $\mathrm{Var}(Y)$?

What is $\mathrm{Var}(X+Y)$?

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \sum_x p_X(x) \cdot (x - \mathbb{E}[X])^2$$