CSE 312

Foundations of Computing II

Lecture 9: Linearity of expectation, LOTUS and variance

Agenda

- Recap
- Linearity of expectation
- LOTUS
- Variance

Review Random Variables

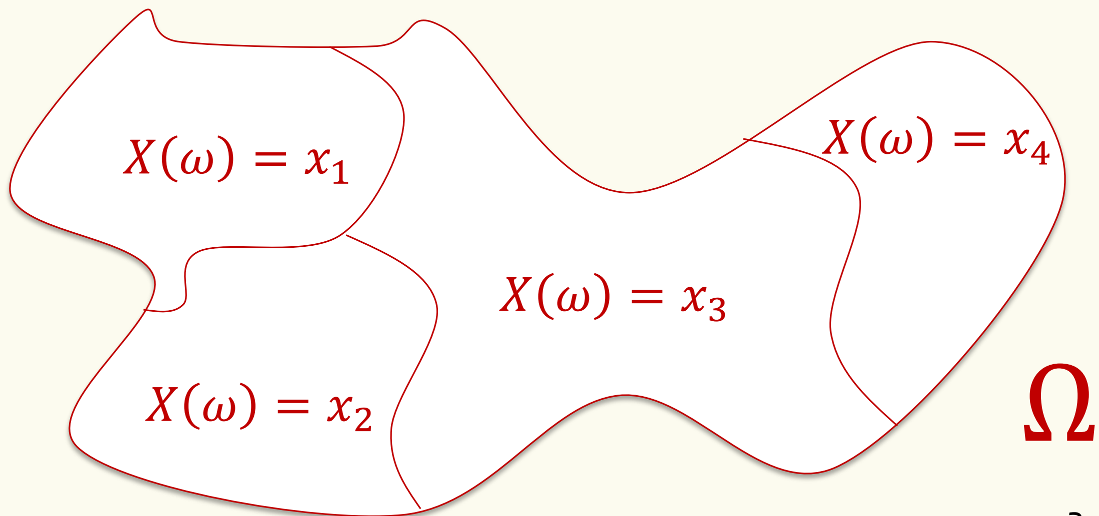
Definition. A **random variable (RV)** for a probability space (Ω, P) is a function $X: \Omega \to \mathbb{R}$.

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

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

Random variables **partition** the sample space.

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



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 $X = x$

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

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

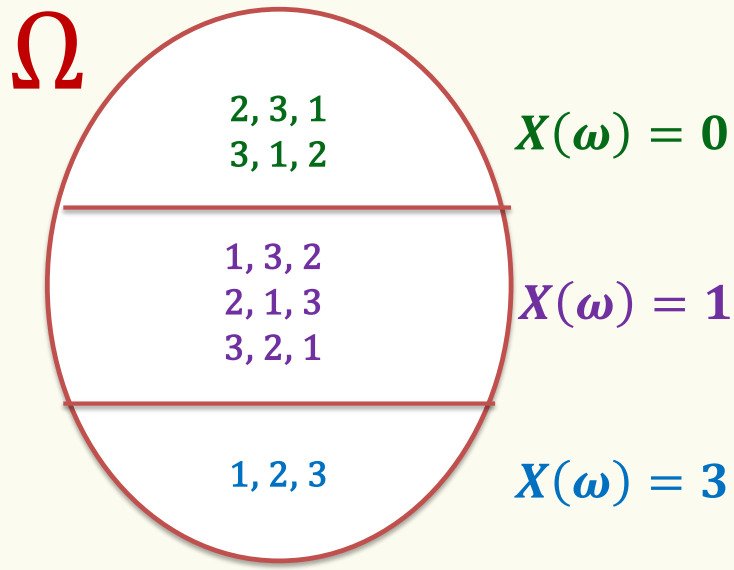
For a RV $X: \Omega \to \mathbb{R}$, the cumulative distribution function (cdf) of X specifies, for any real number x, the probability that $X \leq x$

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

Example: Returning Homeworks

- Class with 3 students, randomly hand back homeworks. All permutations equally likely.
- Let X be the number of students who get their own HW

$\Pr(\omega)$	ω	$X(\omega)$
$1/6$	$1, 2, 3$	3
$1/6$	$1, 3, 2$	1
$1/6$	$2, 1, 3$	1
$1/6$	$2, 3, 1$	0
$1/6$	$3, 1, 2$	0
$1/6$	$3, 2, 1$	1



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

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

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

Example: Returning Homeworks

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

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

- Class with 3 students, randomly hand back homeworks. All permutations equally likely.
- Let X be the number of students who get their own HW

$\Pr(\omega)$	ω	$X(\omega)$
$1/6$	$1, 2, 3$	3
$1/6$	$1, 3, 2$	1
$1/6$	$2, 1, 3$	1
$1/6$	$2, 3, 1$	0
$1/6$	$3, 1, 2$	0
$1/6$	$3, 2, 1$	1

$$\mathbb{E}[X] = 3 \cdot P(X = 3) + 1 \cdot P(X = 1) + 0 \cdot P(X = 0)$$

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

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

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)

Indicator random variable – 0/1 valued

- Class with 3 students, randomly hand back homeworks.
  All permutations equally likely.
- For any event, can define the indicator random variable for that event

$$X_1 = \begin{cases} 1 & \text{if person 1 gets their homework back} \\ 0 & \text{otherwise.} \end{cases}$$

$\Pr(\omega)$	ω	$X(\omega)$
$1/6$	$1, 2, 3$	3
$1/6$	$1, 3, 2$	1
$1/6$	$2, 1, 3$	1
$1/6$	$2, 3, 1$	0
$1/6$	$3, 1, 2$	0
$1/6$	$3, 2, 1$	1

$$P(X_1 = 1) =$$

$$P(X_1 = 0) =$$

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

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

Example – Coin Tosses – The brute force method

We flip n coins, each one heads with probability p,

Z is the number of heads, what is $\mathbb{E}[Z]$?

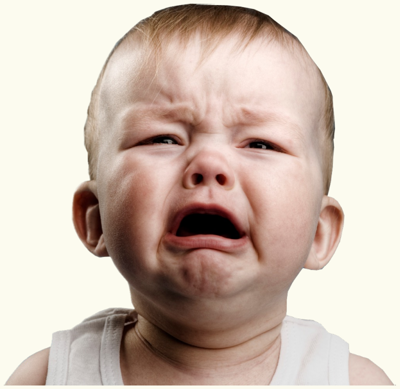
$$\mathbb{E}[Z] = \sum_{k=0}^{n} k \cdot P(Z=k) \quad = \sum_{k=0}^{n} k \cdot \binom{n}{k} p^k (1-p)^{n-k}$$

$$= \sum_{k=0}^{n} k \cdot \frac{n!}{k!\,(n-k)!} p^k (1-p)^{n-k} \quad = \sum_{k=1}^{n} \frac{n!}{(k-1)!\,(n-k)!} p^k (1-p)^{n-k}$$

$$= np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!\,(n-k)!} p^{k-1} (1-p)^{n-k}$$

$$= np \sum_{k=0}^{n-1} \frac{(n-1)!}{k!\,(n-1-k)!} p^k (1-p)^{(n-1)-k}$$

$$= np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{(n-1)-k} = np\big(p + (1-p)\big)^{n-1} = np \cdot 1 = np$$



Can we solve it more elegantly, please?

Example – Coin Tosses

We flip n coins, each toss independent, comes up heads with probability p

Z is the number of heads, what is $\mathbb{E}[Z]$?

$$X_i = \begin{cases} 1, & i^{\text{th}} \text{ coin flip is heads} \\ 0, & i^{\text{th}} \text{ coin flip is tails.} \end{cases}$$

Fact. $Z = X_1 + \cdots + X_n$

Outcomes	X_1	X_2	X_3	Z
TTT	0	0	0	0
TTH	0	0	1	1
THT	0	1	0	1
THH	0	1	1	2
HTT	1	0	0	1
HTH	**1**	**0**	**1**	**2**
HHT	1	1	0	2
HHH	1	1	1	3

Example – Coin Tosses

We flip n coins, each toss independent, comes up heads with probability p

Z is the number of heads, what is $\mathbb{E}[Z]$?

- $X_i = \begin{cases} 1, & i^{\text{th}} \text{ coin flip is heads} \\ 0, & i^{\text{th}} \text{ coin flip is tails.} \end{cases}$

Fact. $Z = X_1 + \cdots + X_n$

Linearity of Expectation:

$$\mathbb{E}[Z] = \mathbb{E}[X_1 + \cdots + X_n] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n] \quad = n \cdot p$$

$P(X_i = 1) = p$
$P(X_i = 0) = 1 - p$

$$\mathbb{E}[X_i] = p \cdot 1 + (1-p) \cdot 0 = p$$

Using LOE to compute complicated expectations

Often boils down to the following three steps:

- Decompose: Finding 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

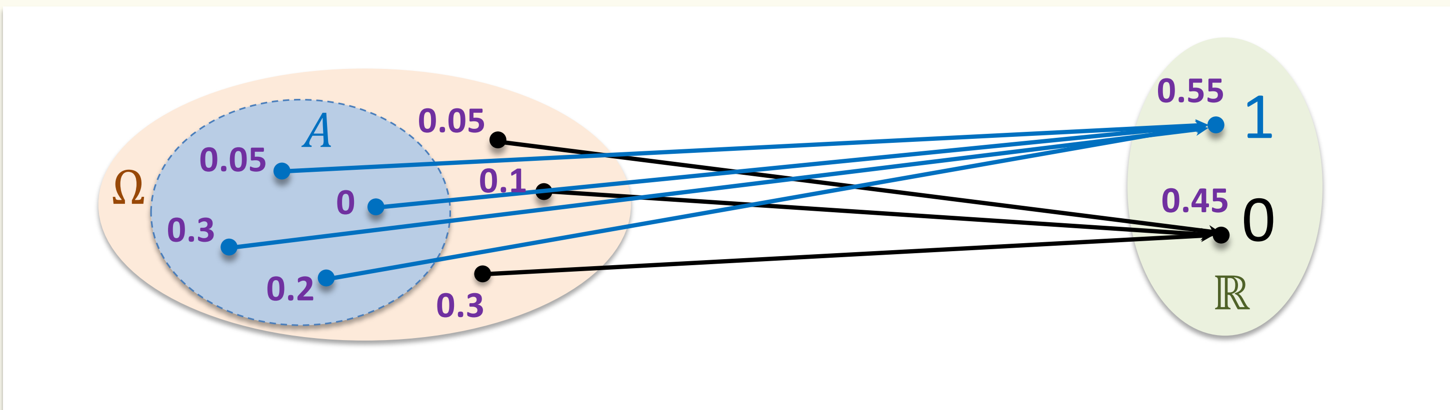
Often, X_i are **indicator** (0/1) random variables.

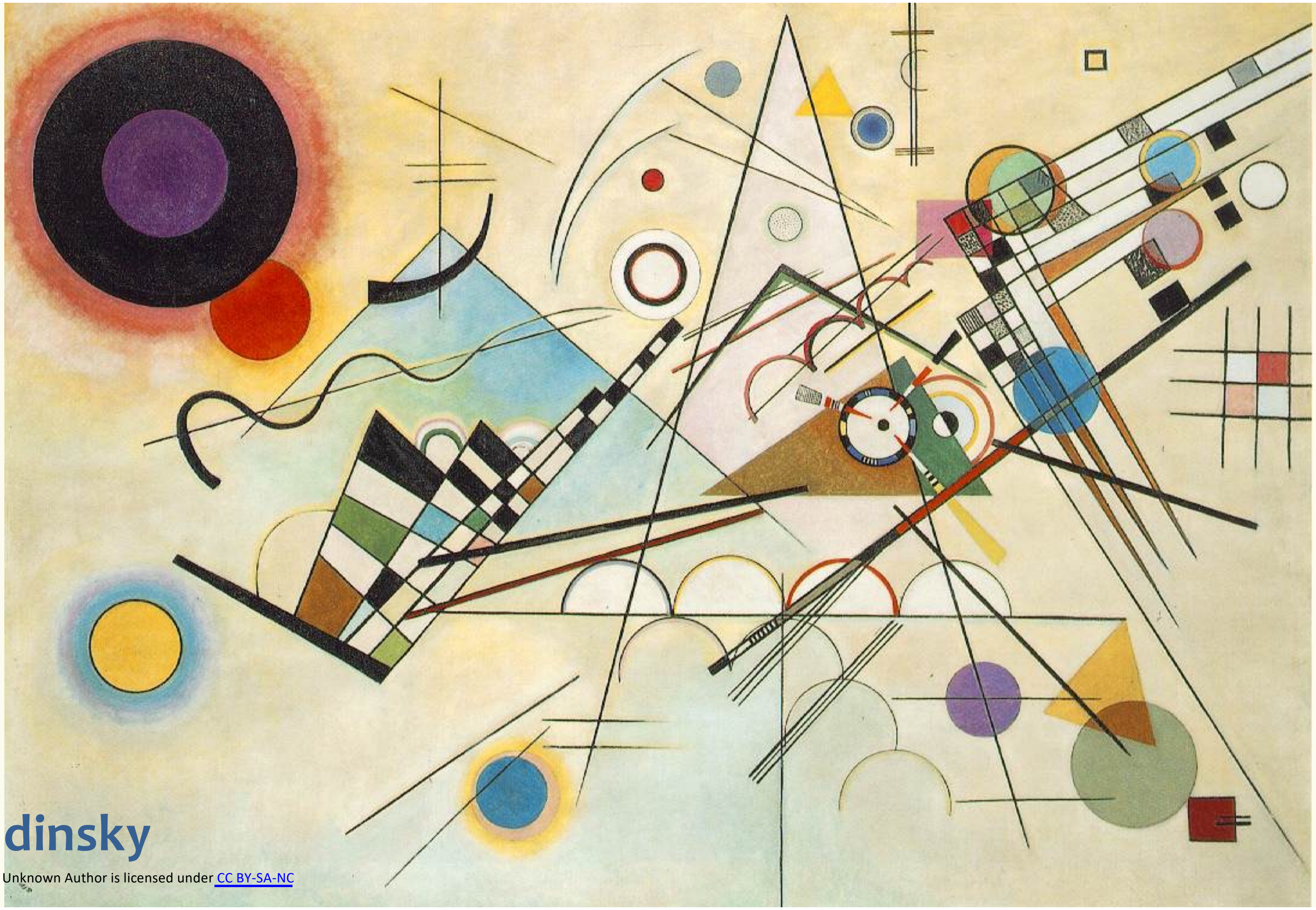
Indicator random variables – 0/1 valued

For any event A, can define the indicator random variable X_A for A

$$X_A = \begin{cases} 1 & \text{if event } A \text{ occurs} \\ 0 & \text{if event } A \text{ does not occur} \end{cases}$$

$$P(X_A = 1) = P(A)$$
$$P(X_A = 0) = 1 - P(A)$$





Kandinsky

Example: Returning Homeworks

- Class with n students, randomly hand back homeworks. All permutations equally likely.
- Let X be the number of students who get their own HW

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

$\Pr(\omega)$	ω	$X(\omega)$
$1/6$	$1, 2, 3$	3
$1/6$	$1, 3, 2$	1
$1/6$	$2, 1, 3$	1
$1/6$	$2, 3, 1$	0
$1/6$	$3, 1, 2$	0
$1/6$	$3, 2, 1$	1

Example: Returning Homeworks

- Class with n students, randomly hand back homeworks. All permutations equally likely.
- Let X be the number of students who get their own HW

What is $\mathbb{E}[X]$? Use linearity of expectation!

$\Pr(\omega)$	ω	$X(\omega)$
$1/6$	$1, 2, 3$	3
$1/6$	$1, 3, 2$	1
$1/6$	$2, 1, 3$	1
$1/6$	$2, 3, 1$	0
$1/6$	$3, 1, 2$	0
$1/6$	$3, 2, 1$	1

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!

Example: Returning Homeworks

- Class with n students, randomly hand back homeworks. All permutations equally likely.
- Let X be the number of students who get their own HW

What is $\mathbb{E}[X]$? Use linearity of expectation!

Decompose:

$\Pr(\omega)$	ω	$X(\omega)$
$1/6$	$1, 2, 3$	3
$1/6$	$1, 3, 2$	1
$1/6$	$2, 1, 3$	1
$1/6$	$2, 3, 1$	0
$1/6$	$3, 1, 2$	0
$1/6$	$3, 2, 1$	1

LOE:

Conquer:

Example: Returning Homeworks

- Class with n students, randomly hand back homeworks. All permutations equally likely.
- Let X be the number of students who get their own HW

What is $\mathbb{E}[X]$? Use linearity of expectation!

$\Pr(\omega)$	ω	$X(\omega)$
$1/6$	$1, 2, 3$	3
$1/6$	$1, 3, 2$	1
$1/6$	$2, 1, 3$	1
$1/6$	$2, 3, 1$	0
$1/6$	$3, 1, 2$	0
$1/6$	$3, 2, 1$	1

Decompose: What is X_i?

$X_i = 1$ iff i^{th} student gets own HW back; 0 o.w.

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

Conquer: $\mathbb{E}[X_i] = \frac{1}{n}$

Therefore, $\mathbb{E}[X] = n \cdot \frac{1}{n} = 1$

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

- In a class of m students, on average how many pairs of people have the same birthday (assuming 365 equally likely birthdays)?

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 $\frac{\binom{m}{2}}{365} = \frac{m(m-1)}{730}$ pairs

Agenda

- Recap
- Linearity of expectation
- LOTUS
- Variance

Linearity of Expectation – Even stronger

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

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

Very important: In general, we do not have $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$

Linearity is special!

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

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

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

How DO we compute $\mathbb{E}[g(X)]$?

Expected Value of $g(X)$

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

(nothing special going on in the discrete case)

Example: from concept check

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

- Toss a die; each side equally likely. X is the number showing
- $Y = X \bmod 4$
- What is $\mathbb{E}[Y]$?

$\mathbf{Pr}(\boldsymbol{\omega})$	$\boldsymbol{\omega}$	$\boldsymbol{X}$
$1/6$	1	1
$1/6$	2	2
$1/6$	3	3
$1/6$	4	4
$1/6$	5	5
$1/6$	6	6

Agenda

- Recap
- Linearity of expectation
- LOTUS
- Variance

Which game would you rather play?

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?

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.*

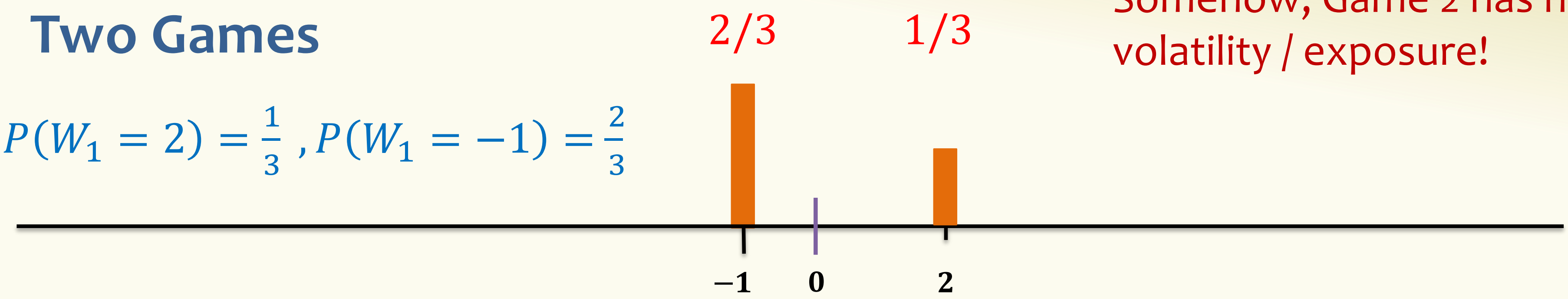
W_2 = payoff in a round of Game 2

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

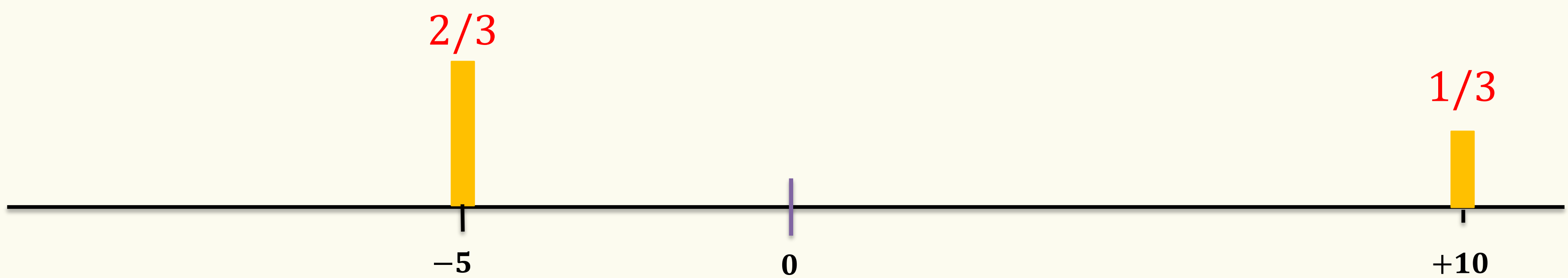
$$\mathbb{E}[W_2] = 0$$

Two Games

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



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



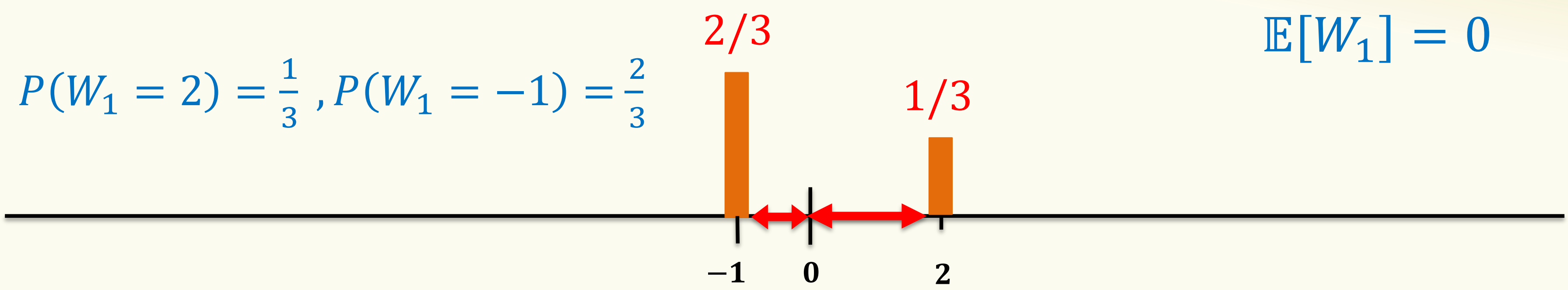
Same expectation, but clearly a very different distribution.

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

Variance (Intuition, First Try)

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

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



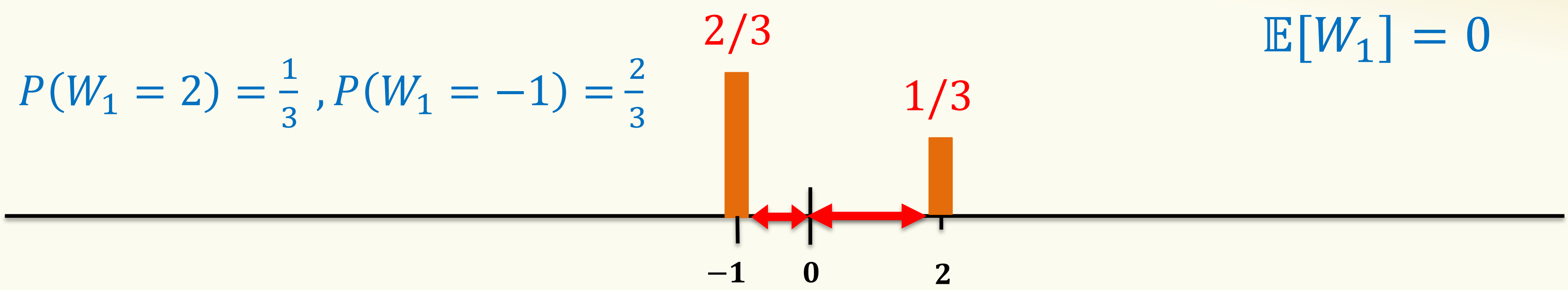
New quantity (random variable): How far from the expectation?

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

Variance (Intuition, First Try)

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

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



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

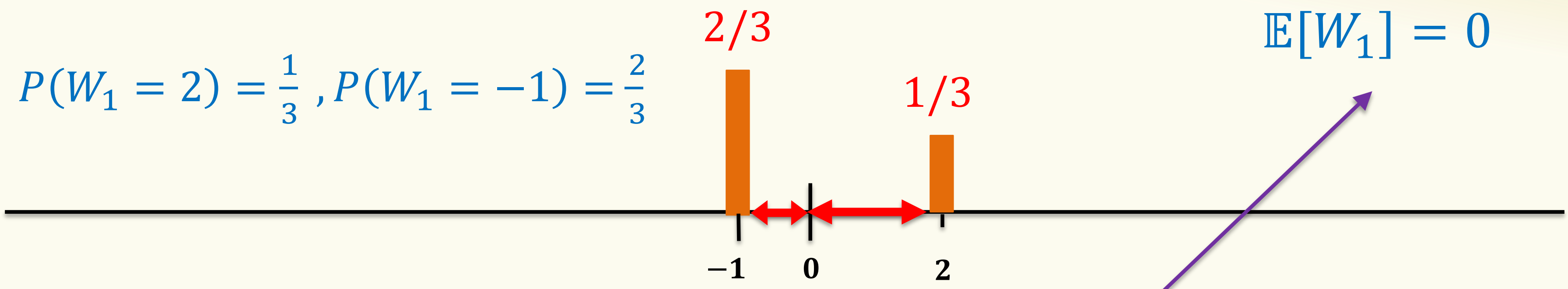
$W_1 - \mathbb{E}[W_1]$

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

Variance (Intuition, Better Try)

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

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



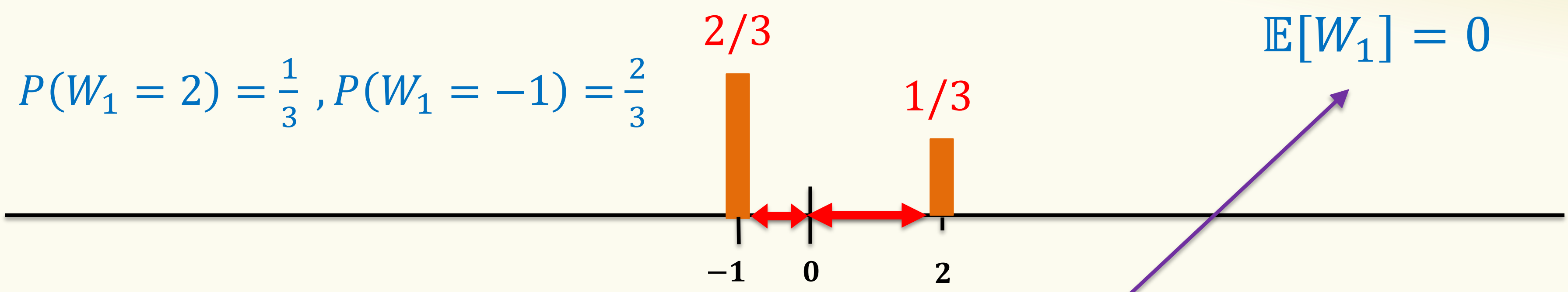
A better quantity (random variable): How far from the expectation?

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

Variance (Intuition, Better Try)

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

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



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

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

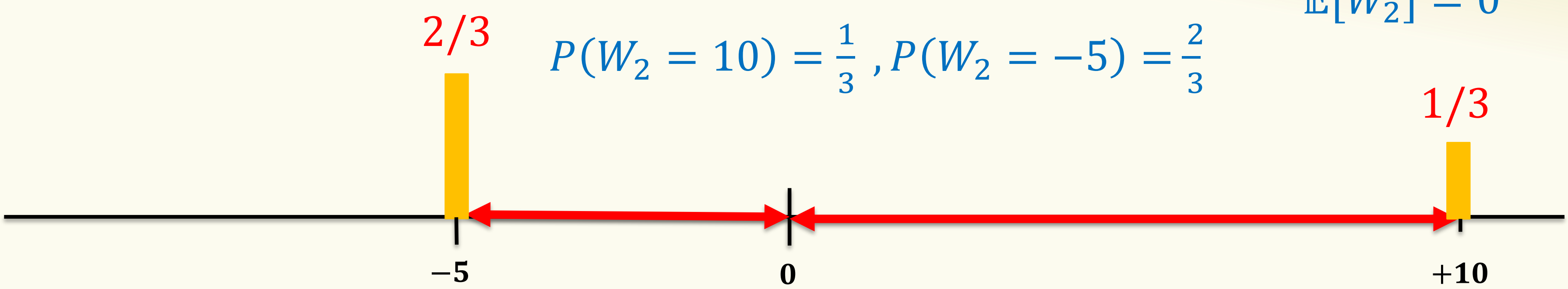
$$= \frac{2}{3} \cdot 1 + \frac{1}{3} \cdot 4$$

$$= 2$$

Variance (Intuition, Better Try)

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

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



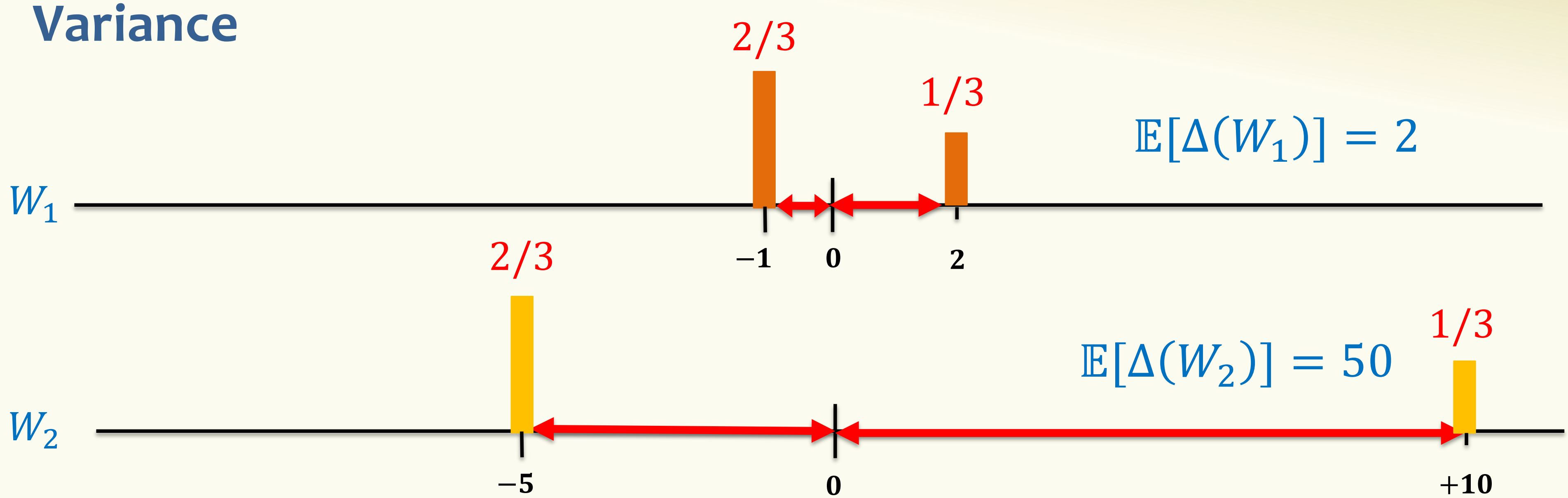
A better quantity (random variable): How far from the expectation?

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

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

$$= 50$$

Variance



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

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

Variance

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

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

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

X fair die

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

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

Variance – Example 1

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