CSE 312

Foundations of Computing II

Lecture 10: LOTUS, variance and independence among random variables.

slido.com/3680281

Agenda

- Recap
- LOTUS
- Variance
- Properties of Variance
- Independence of random variables

Review Random Variables

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

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

For a RV $X: \Omega \rightarrow \mathbb{R}$, 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 $X: \Omega \rightarrow \mathbb{R}$, 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

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

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)

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

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

Using LOE to compute complicated expectations

$E(X)$

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

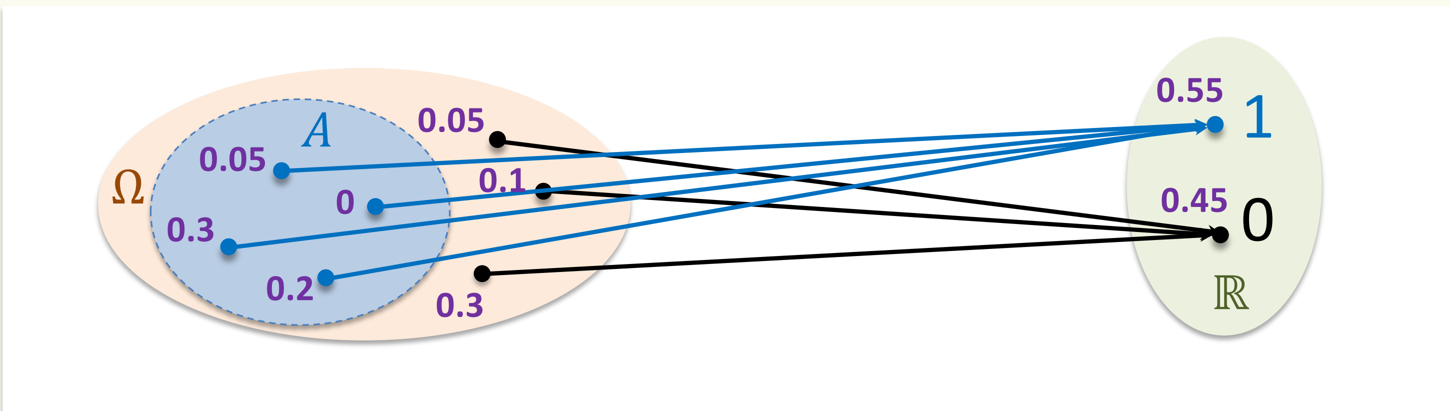
Often, X_i are **indicator** (0/1) random variables.

Indicator random variables – 0/1 valued

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

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

$$\boxed{\begin{aligned} P(X_A = 1) &= P(A) \\ P(X_A = 0) &= 1 - P(A) \end{aligned}}$$



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

Agenda

- Recap
- LOTUS
- Variance
- Properties of Variance
- Independence of random variables

Linearity of Expectation – Even stronger

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

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

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

$Y = g(X)$

$X = x$

$\downarrow$

$\forall x$

$g(X) \rightarrow g(x)$

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 $\mathbb{E}[g(X)]$?

$$E(X) = \sum_{\omega \in \Omega} X(\omega) P(\omega)$$

$$E(X) = \sum_{x \in \Omega_X} x\, P(X=x)$$

Expected Value of $g(X)$

$Y = g(X)$

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
- $\boxed{Y = X \bmod 4}$
- What is $\mathbb{E}[Y]$?

$\boxed{g(x) = x \bmod 4}$

$E(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6}$

$E(Y) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 0 \cdot \frac{1}{6} + 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6}$

$g(x)$

$\Pr(\omega)$	ω	X	Y
1/6	1	1	1
1/6	2	2	2
1/6	3	3	3
1/6	4	4	0
1/6	5	5	1
1/6	6	6	2

$P(Y=0) = \frac{1}{6}$

$1 = \frac{1}{3}$

$2 \quad \frac{1}{3}$

$3 \quad \frac{1}{6}$

Agenda

- Recap
- LOTUS
- Variance
- Properties of Variance
- Independence of random variables

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

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

$E(W_1) = 2 \cdot \frac{1}{3} + (-1)\frac{2}{3} = 0$

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

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

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

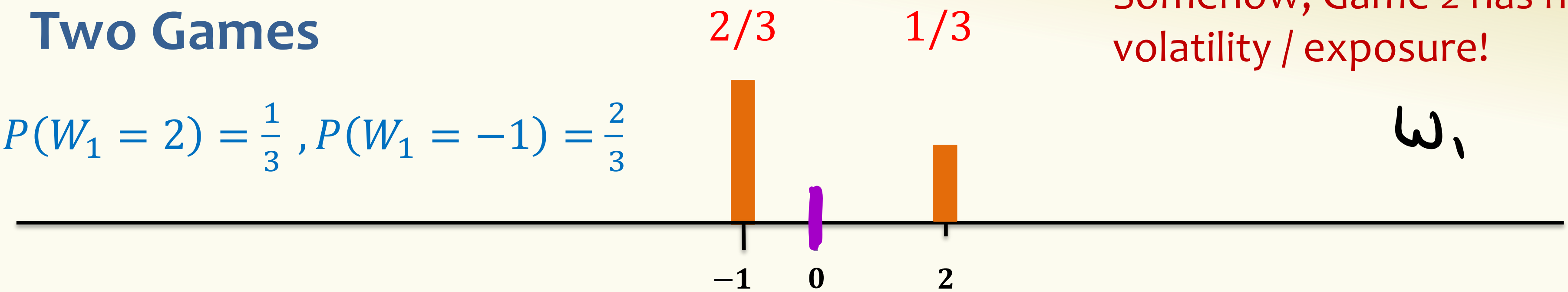
$$E(W_2) = 10 \cdot \frac{1}{3} + (-5)\frac{2}{3} = 0$$

Two Games

Somehow, Game 2 has higher volatility / exposure!

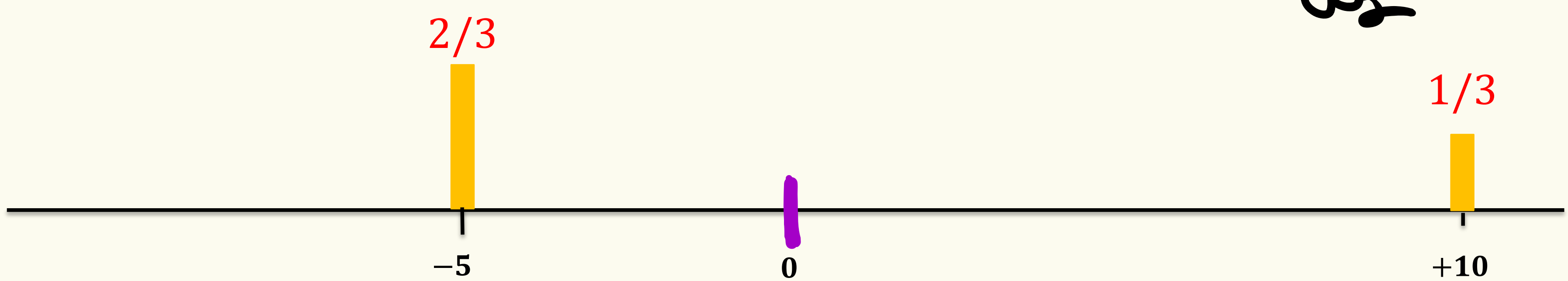
$P(W_1 = 2) = \frac{1}{3}, P(W_1 = -1) = \frac{2}{3}$

W_1



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

W_2



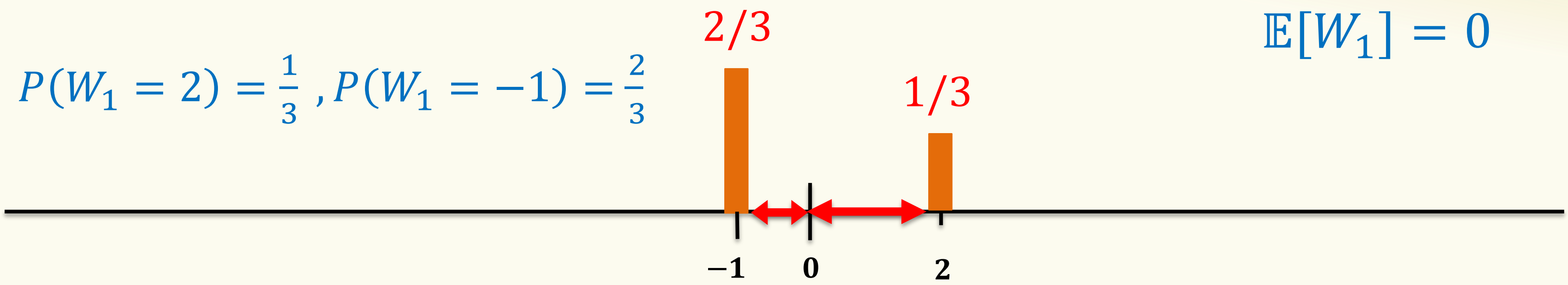
Same expectation, but clearly a very different distribution.

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?

$$E\left(W_1 - \mathbb{E}[W_1]\right) \underset{LOE}{=} E(W_1) - E(E(W_1))$$

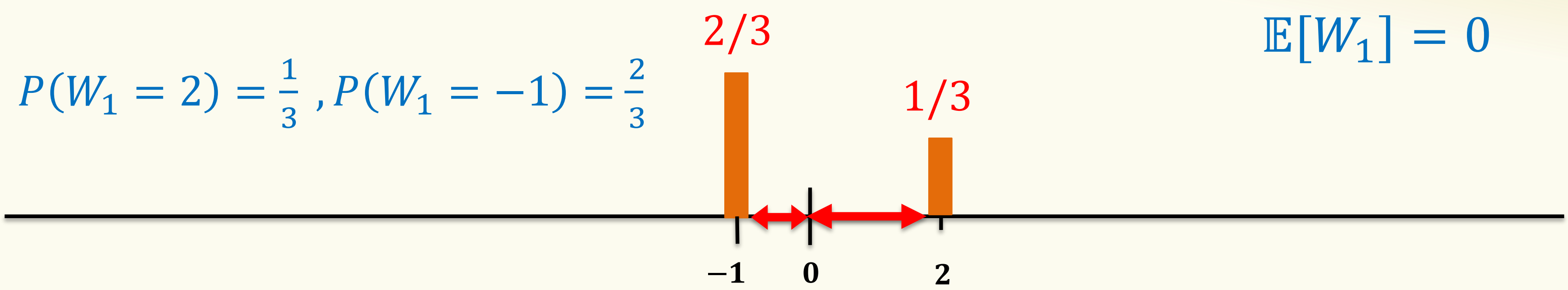
$$= E(W_1) - E(W_1) = 0$$

$$\cancel{E(|W_1 - E(W_1)|)}$$

Variance (Intuition, First Try)

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

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



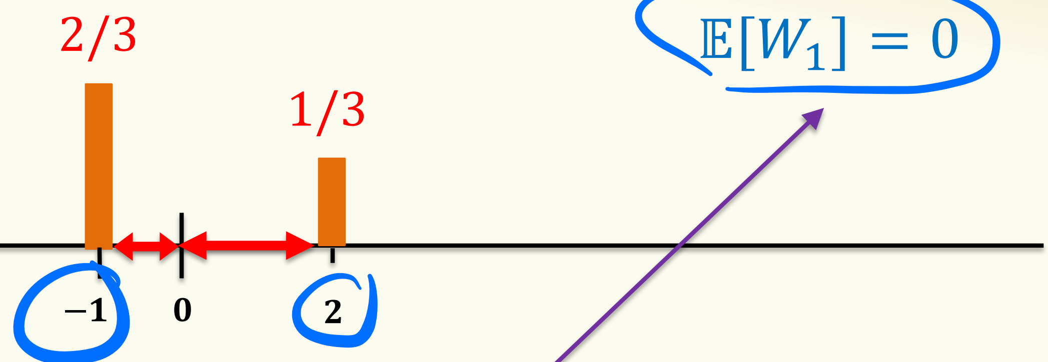
New quantity (random variable): How far from the expectation?

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

$$\begin{aligned}\mathbb{E}[W_1 - \mathbb{E}[W_1]] &= \mathbb{E}[W_1] - \mathbb{E}\big[\mathbb{E}[W_1]\big] \\ &= \mathbb{E}[W_1] - \mathbb{E}[W_1] \\ &= 0\end{aligned}$$

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?

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

$$= E(W_1^2)$$

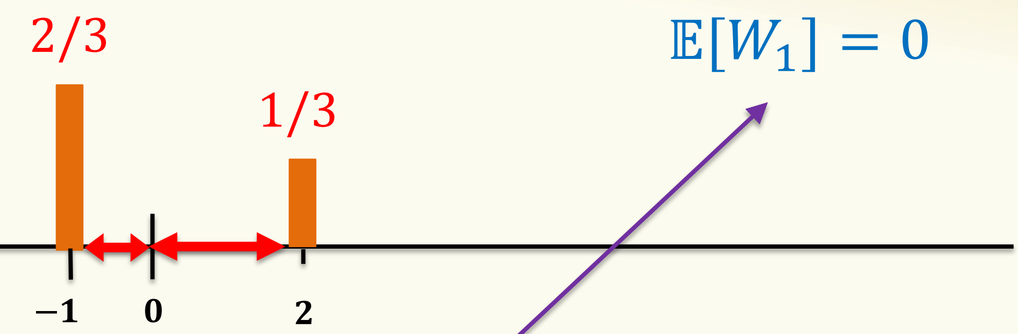
$$= 2^2 \cdot \frac{1}{3} + (-1)^2 \cdot \frac{2}{3} = 2$$

LOTUS:

$$E(g(X)) = \sum_{x \in \Omega_X} g(x) \Pr(X = x)$$

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?

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

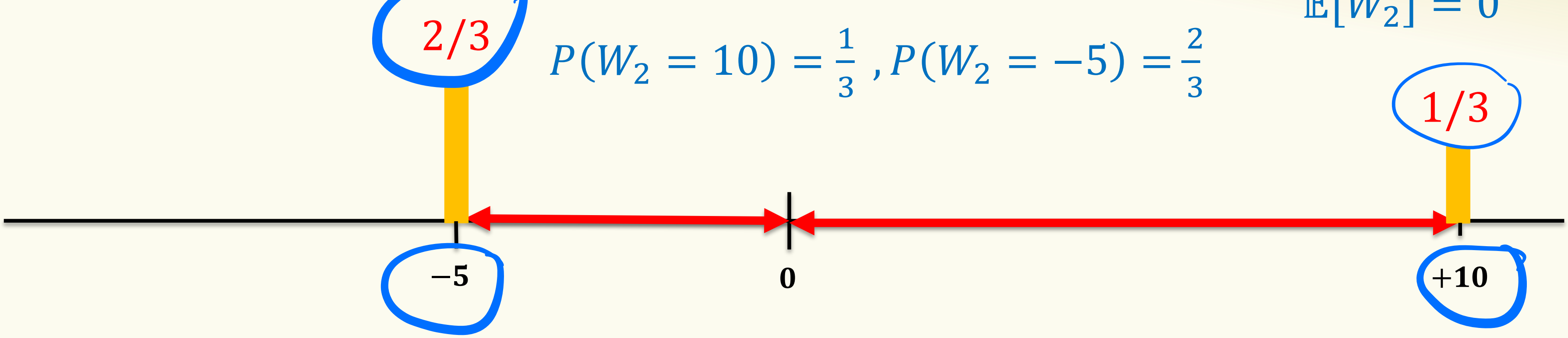
$$= \frac{2}{3} \cdot 1 + \frac{1}{3} \cdot 4$$

$$= 2$$

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?

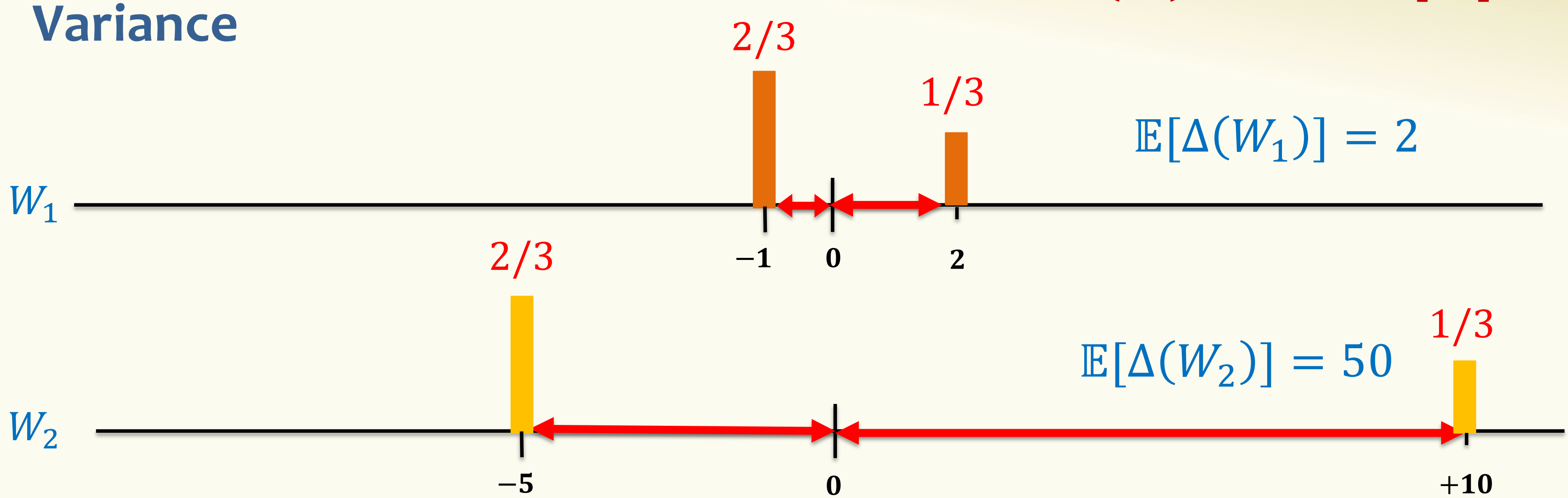
$$\mathbb{E}[(W_2 - \overset{0}{\cancel{\mathbb{E}[W_2]}})^2]$$

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

$$= 50$$

Variance

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



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

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

$g(x) = (x - \mathbb{E}(X))^2$

Variance

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

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

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

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

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

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

Variance – Example 1

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

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

X fair die

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

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

$$= \frac{1}{6}(1 - 3.5)^2 + \frac{1}{6}(2 - 3.5)^2 + \cdots + \frac{1}{6}(6 - 3.5)^2$$

Variance – Example 1

X fair die

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

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

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

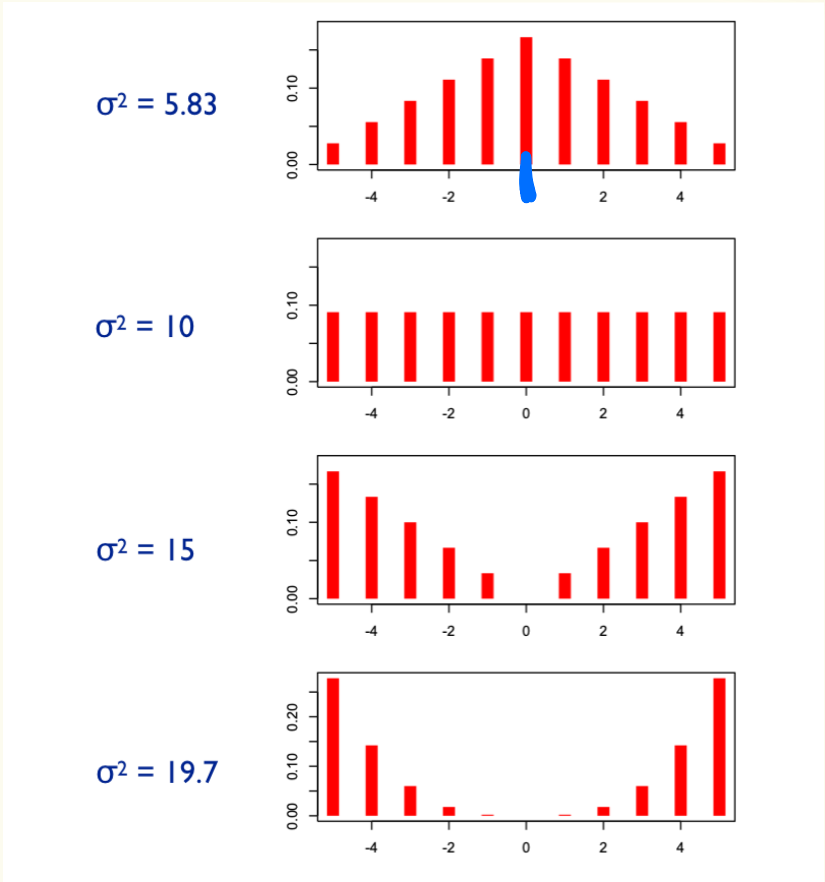
$$= \frac{2}{6}[2.5^2 + 1.5^2 + 0.5^2] = \frac{2}{6}\left[\frac{25}{4} + \frac{9}{4} + \frac{1}{4}\right] = \frac{35}{12} \approx 2.91677 \ldots$$

Variance in Pictures

E(X)=0 for all of these

Captures how much "spread' there is in a pmf

All pmfs have same expectation



Agenda

- Recap
- LOTUS
- Variance
- Properties of Variance
- Independence of random variables

Variance – Properties

$Var(Y)$ $\quad x+b \quad E(X)+b$

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

$X \quad Y = X + b$

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

$\boxed{Z = aX}$

$E(Z) = a E(X)$

$Z = aX$

X	prob	$Z=aX$	
x_1	p_1	ax_1	p_1
x_2	p_2	ax_2	
$\vdots$		$\vdots$	
x_k	p_k	ax_k	p_k

$$Var(Z) = \sum_{x \in \Omega_X} p_X(x) \left(ax - E(Z)\right)^2$$

$= \sum_x p_X(x) a^2 (x - E(X))^2 \quad a^2 E(X)$

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

Variance – Properties

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

$(a-b)^2 = a^2 - 2ab + b^2$

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

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

$$\underset{\text{LOE}}{=} E(X^2) + \underbrace{E(-2X E(X))}_{-2E(X)E(X)} + \underbrace{E(E(X)^2)}_{(E(X))^2}$$

$= E(X^2) - [E(X)]^2$

Variance

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

X fair die

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

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

$$E\left[\underline{(X - E(X))^2}\right]$$

Variance of Indicator Random Variables

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

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

$$X_A = \begin{cases} 1 & A \text{ happens} \\ 0 & \text{o.w.} \end{cases}$$

$$E(X_A^2) = 1^2 P(A) + 0^2(1 - P(A)) = p$$

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

Variance of Indicator Random Variables

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

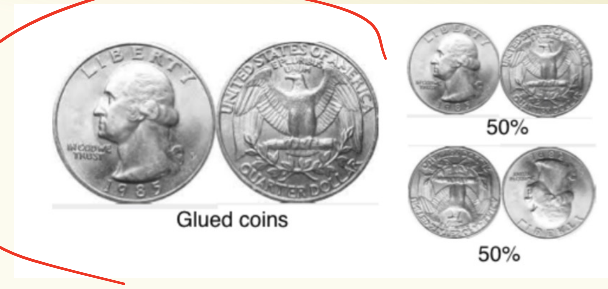
$$\mathbb{E}[X_A] = P(A) = p$$

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

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

$p = \frac{1}{2}$ $\quad$ $\mathrm{Var}(X_A) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$

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



Proof by counter-example:

Recall glued coins

- Let X_1 be a r.v. that indicates if the first coin comes up heads.
- Let X_2 be a r.v. that indicates if the second coin comes up heads.

HT $\frac{1}{2}$

TH $\frac{1}{2}$

first 2nd

$\mathrm{Var}(X_1) = \frac{1}{4}$

$\mathrm{Var}(X_2) = \frac{1}{4}$

$X_1 + X_2$ is a const w/ value 1

$\mathrm{Var}(X_1 + X_2) = 0 \neq \mathrm{Var}(X_1) + \mathrm{Var}(X_2)$

Brain Break

Agenda

- Recap
- LOTUS
- Variance
- Properties of Variance
- Independent Random Variables

Random Variables and Independence

Comma is shorthand for AND

Definition. Two random variables X, Y are **(mutually) independent** if for all x, y,

$$P(\underbrace{X = x}_{A}, \underbrace{Y = y}_{B}) = P(X = x) \cdot P(Y = y)$$

Intuition: Knowing X doesn't help you guess Y and vice versa

Definition. The random variables $X_1, \dots, X_n$ are **(mutually) independent** if for all $x_1, \dots, x_n$,

$$P(X_1 = x_1, \dots, X_n = x_n) = P(X_1 = x_1) \cdots P(X_n = x_n)$$

Note: No need to check for all subsets, but need to check <u>for all values</u>!

Example

Let X be the number of heads in n *independent* coin flips of the same coin with probability p of coming up heads.

Let $Y = X \bmod 2$ be the parity (even/odd) of X.

Are X and Y independent?

$P(X=3, Y=0) = 0$

even

$P(X=3) = \binom{n}{3} p^3 (1-p)^{n-3}$

$P(Y=0) = \frac{1}{2}$

Poll:
slido.com/3680281

A. Yes
B. No

Example

Make $2n$ *independent* coin flips of the same coin with probability p of coming up heads. .

Let X be the number of heads in the first n flips and Y be the number of heads in the last n flips.

Are X and Y independent?

Poll:
slido.com/3680281

A. Yes
B. No

Agenda

- Variance
- Properties of Variance
- Independent Random Variables
- Properties of Independent Random Variables

Important Facts about Independent Random Variables

Theorem. If X, Y independent, $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$

Theorem. If X, Y independent, $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$

Corollary. If $X_1, X_2, \ldots, X_n$ mutually independent,

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i}^{n} \mathrm{Var}(X_i)$$

(Not Covered) Proof of $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$

Theorem. If X, Y independent, $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$

Proof

Let $x_i, y_i, i = 1, 2, \ldots$ be the possible values of X, Y.

$$\mathbb{E}[X \cdot Y] = \sum_i \sum_j x_i \cdot y_j \cdot P(X = x_i \wedge Y = y_j)$$

independence

$$= \sum_i \sum_j x_i \cdot y_i \cdot P(X = x_i) \cdot P(Y = y_j)$$

$$= \sum_i x_i \cdot P(X = x_i) \cdot \left(\sum_j y_j \cdot P(Y = y_j) \right)$$

$$= \mathbb{E}[X] \cdot \mathbb{E}[Y]$$

Note: *NOT* true in general; see earlier example $\mathbb{E}[X^2] \neq \mathbb{E}[X]^2$

(Not Covered) Proof of $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$

Theorem. If X, Y independent, $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$

Proof

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

linearity

equal by independence

Example – Coin Tosses

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

- $X_i = \begin{cases} 1, & i^{\text{th}} \text{ outcome is heads} \\ 0, & i^{\text{th}} \text{ outcome is tails.} \end{cases}$
- $Z =$ number of heads

Fact. $Z = \sum_{i=1}^{n} X_i$

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

What is $\mathbb{E}[Z]$? What is $\text{Var}(Z)$?

$P(Z = k) = \binom{n}{k} p^k (1-p)^{n-k}$

Note: $X_1, \ldots, X_n$ are <u>mutually</u> independent! [Verify it formally!]

$$\text{Var}(Z) = \sum_{i=1}^{n} \text{Var}(X_i) = n \cdot p(1-p)$$

Note $\text{Var}(X_i) = p(1-p)$