

CSE 312

Foundations of Computing II

Lecture 9: Variance and Independence of RVs

Recap Linearity of Expectation

Theorem. For **any** two random variables X and Y (X, Y do not need to be independent)

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

Theorem. For any random variables X_1, \dots, X_n ,

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

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

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

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

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)?
- Call this number 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} $X = \sum_{i \neq j} X_{ij}$

Conquer: $\mathbb{E}[X_{ij}] = \frac{1}{365}$ so total expectation is $\frac{\binom{m}{2}}{365} = \frac{m(m-1)}{730}$ pairs

Linearity of Expectation – Even stronger

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

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

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

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 $X: \Omega \rightarrow \mathbb{R}$, the **expectation** or **expected value** or **mean** of $g(X)$ is

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

or equivalently

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

Example: Expectation of $g(X)$

Suppose we rolled a fair, 6-sided die in a game.

You will win the cube of the number rolled in dollars, times 10.

Let X be the result of the dice roll.

What is your expected winnings?

$$\mathbb{E}[10X^3] = 10 \sum_{k=1}^6 k^3 \cdot \frac{1}{6}$$

Agenda

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

Two Games

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

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

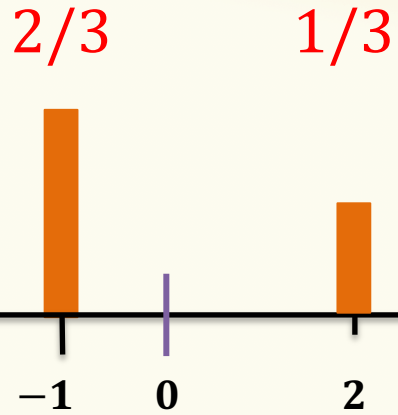
$$\mathbb{E}[W_2] = 0$$

Which game would you rather play?

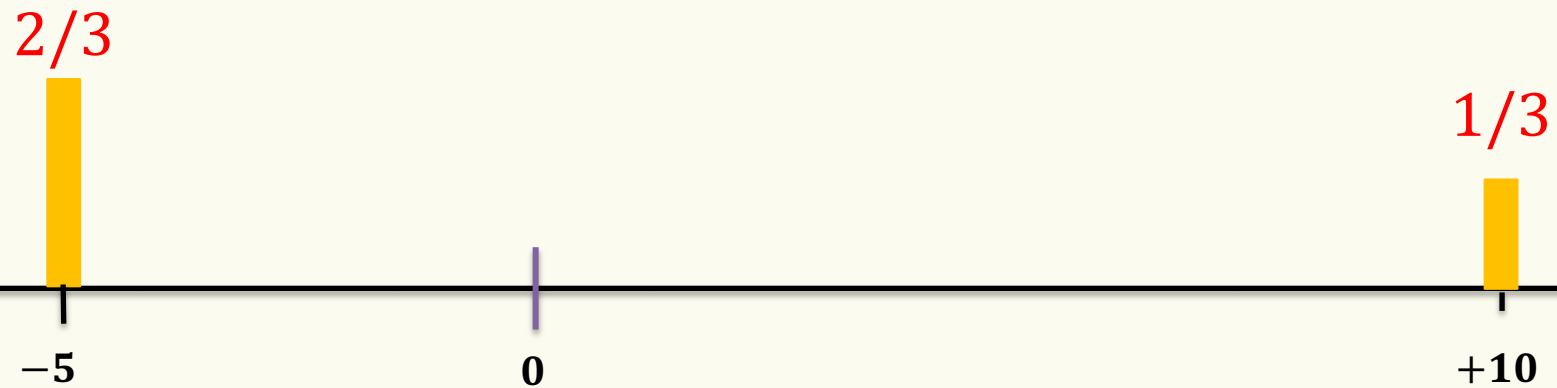
Somehow, Game 2 has higher volatility / exposure!

Two Games

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



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



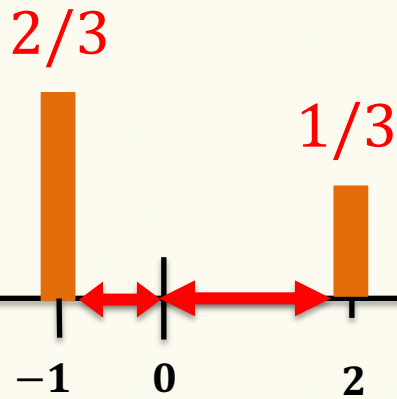
Same expectation, but clearly a very different distribution.

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?

$$\Delta(W_1) = W_1 - \mathbb{E}[W_1]$$

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

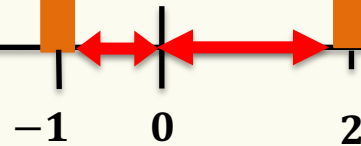
Variance (Intuition, Better Try)

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

$\frac{2}{3}$

$\frac{1}{3}$

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



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

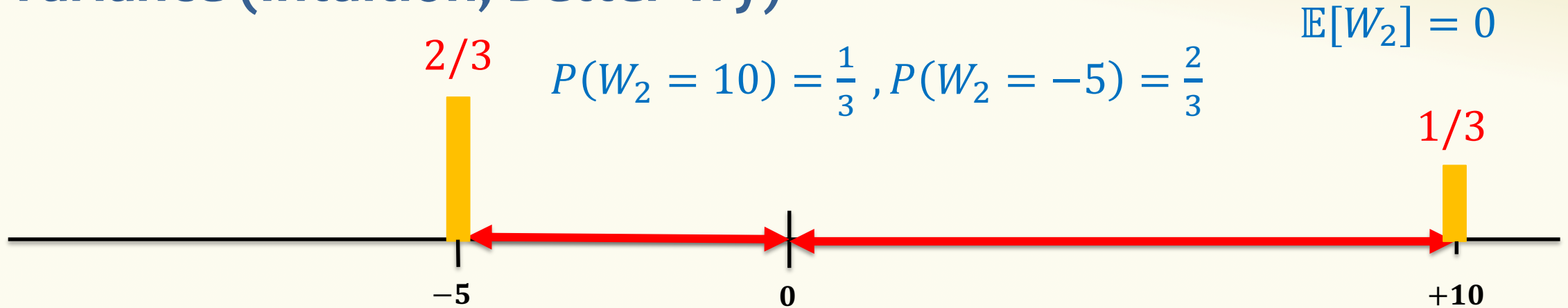
$$\Delta(W_1) = (W_1 - \mathbb{E}[W_1])^2$$

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

$$P(\Delta(W_1) = 4) = \frac{1}{3}$$

$$\begin{aligned}\mathbb{E}[\Delta(W_1)] &= \mathbb{E}[(W_1 - \mathbb{E}[W_1])^2] \\ &= \frac{2}{3} \cdot 1 + \frac{1}{3} \cdot 4 \\ &= 2\end{aligned}$$

Variance (Intuition, Better Try)



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

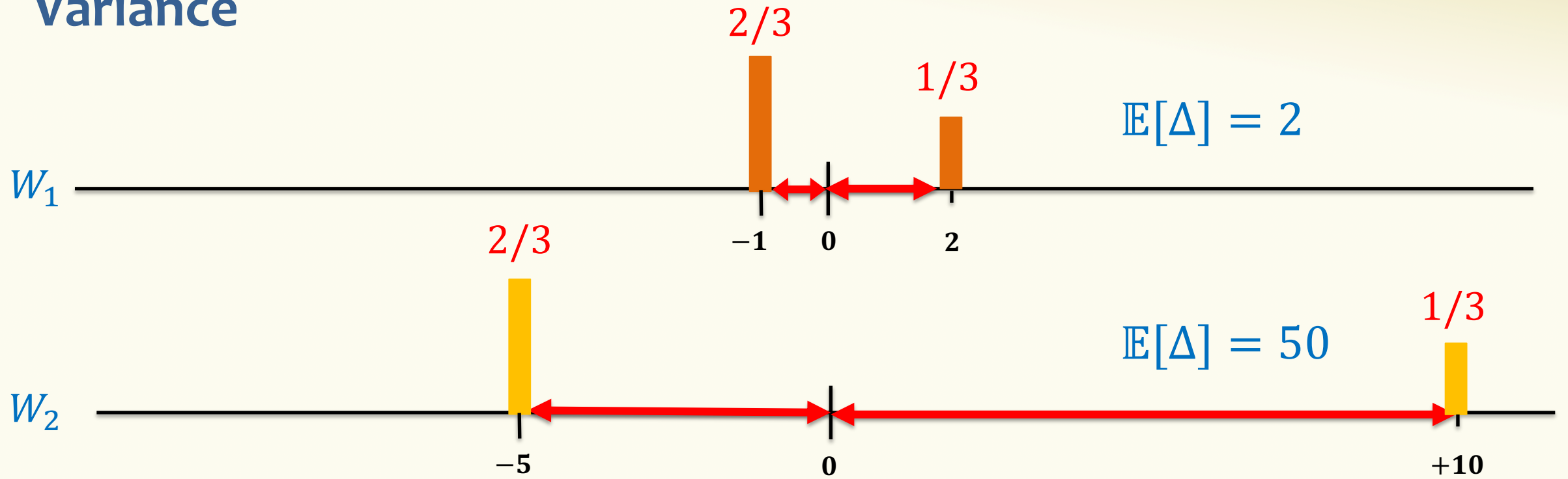
$$\Delta(W_2) = (W_2 - \mathbb{E}[W_2])^2$$

$$\mathbb{P}(\Delta(W_2) = 25) = \frac{2}{3}$$

$$\mathbb{P}(\Delta(W_2) = 100) = \frac{1}{3}$$

$$\begin{aligned}\mathbb{E}[\Delta(W_2)] &= \mathbb{E}[(W_2 - \mathbb{E}[W_2])^2] \\ &= \frac{2}{3} \cdot 25 + \frac{1}{3} \cdot 100 \\ &= 50\end{aligned}$$

Variance



We say that W_2 has “**higher variance**” than W_1 .

Variance

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

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

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

X fair die

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

$$\text{Var}(X) = ? \left(x - \mathbb{E}(X) \right)^2$$

Variance – Example 1

X fair die

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

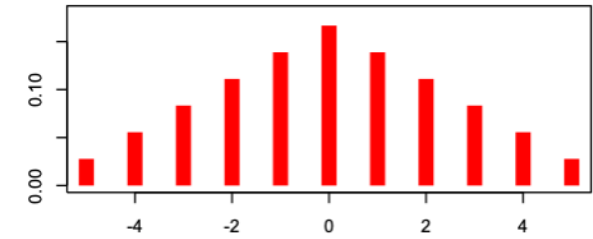
$$\begin{aligned}\text{Var}(X) &= \sum_{\mathbf{x}} P(X = \mathbf{x}) \cdot (\mathbf{x} - \mathbb{E}[X])^2 \\ &= \frac{1}{6} [(1 - 3.5)^2 + (2 - 3.5)^2 + (3 - 3.5)^2 + (4 - 3.5)^2 + (5 - 3.5)^2 + (6 - 3.5)^2] \\ &= \frac{2}{6} [2.5^2 + 1.5^2 + 0.5^2] = \frac{2}{6} \left[\frac{25}{4} + \frac{9}{4} + \frac{1}{4} \right] = \frac{35}{12} \approx 2.91677 \dots\end{aligned}$$

Variance in Pictures

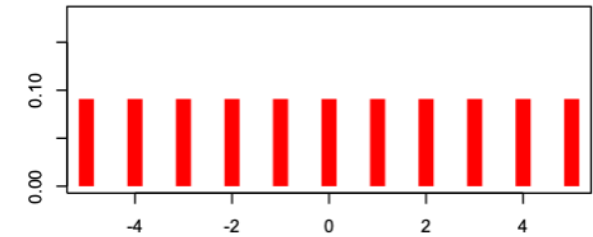
Captures how much
“spread” there is in a pmf

All pmfs have same
expectation

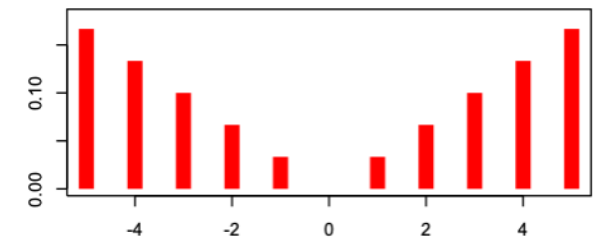
$$\sigma^2 = 5.83$$



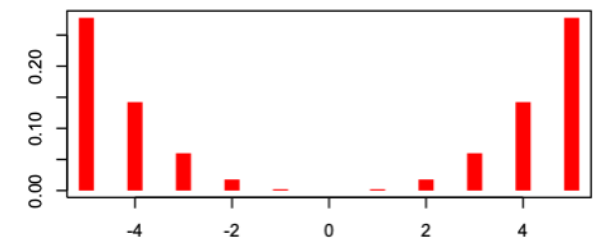
$$\sigma^2 = 10$$



$$\sigma^2 = 15$$



$$\sigma^2 = 19.7$$



Agenda

- Variance
- **Properties of Variance** ◀
- Independent Random Variables
- Properties of Independent Random Variables

Variance – Properties

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

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

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

(Proof: Exercise!)

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

Variance

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

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

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

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

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

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

(linearity of expectation!)

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

Variance – Example 1

X fair die

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

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

Variance of Indicator Random Variables

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

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

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

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

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

Proof by counter-example:

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

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

Brain Break



Agenda

- Variance
- Properties of Variance
- Independent Random Variables ◀
- Properties of 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 ,

$$P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$$

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 for all outcomes!

Example

Let X be the number of heads in n independent coin flips of the same coin. Let $Y = X \bmod 2$ be the parity (even/odd) of X .

Are X and Y independent?

Example

Make $2n$ independent coin flips of the same coin.

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?

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, $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$

Corollary. If X_1, X_2, \dots, X_n mutually independent,

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

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

$$\begin{aligned}\mathbb{E}[X \cdot Y] &= \sum_i \sum_j x_i \cdot y_j \cdot P(X = x_i \wedge Y = y_j) \\ &= \sum_i \sum_j x_i \cdot y_j \cdot P(X = x_i) \cdot P(Y = y_j) \quad \text{independence} \\ &= \sum_i x_i \cdot P(X = x_i) \cdot \left(\sum_j y_j \cdot P(Y = y_j) \right) \\ &= \mathbb{E}[X] \cdot \mathbb{E}[Y]\end{aligned}$$

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

Proof of $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$

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

Proof

$$\begin{aligned} & \text{Var}(X + Y) \\ &= \mathbb{E}[(X + Y)^2] - (\mathbb{E}[X + Y])^2 \\ &= \mathbb{E}[X^2 + 2XY + Y^2] - (\mathbb{E}[X] + \mathbb{E}[Y])^2 \\ &= \mathbb{E}[X^2] + 2 \mathbb{E}[XY] + \mathbb{E}[Y^2] - (\mathbb{E}[X]^2 + 2 \mathbb{E}[X] \mathbb{E}[Y] + \mathbb{E}[Y]^2) \\ &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 + \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 + 2 \mathbb{E}[XY] - 2 \mathbb{E}[X] \mathbb{E}[Y] \\ &= \text{Var}(X) + \text{Var}(Y) + 2 \mathbb{E}[XY] - 2 \mathbb{E}[X] \mathbb{E}[Y] \\ &= \text{Var}(X) + \text{Var}(Y) \end{aligned}$$

linearity

equal by independence

Example – Coin Tosses

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

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

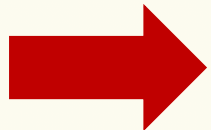
$$\text{Fact. } Z = \sum_{i=1}^n X_i$$

$$\begin{aligned} P(X_i = 1) &= p \\ P(X_i = 0) &= 1 - p \end{aligned}$$

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, \dots, X_n are mutually independent!


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

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