

**CSE 312**

# **Foundations of Computing II**

**Lecture 10: Variance and Independence of RVs**

**Anonymous questions: [www.slido.com/3296240](http://www.slido.com/3296240)**

# Agenda

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

## Review Expected Value of a Random Variable

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

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

or equivalently

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

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

## Linearity of Expectation – Even stronger

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

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

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

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

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

(nothing special going on in the discrete case)

## Example: Returning Homeworks

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

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

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

$\Pr(\omega)$	$\omega$	$X(\omega)$
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

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

# Agenda

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

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

## 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

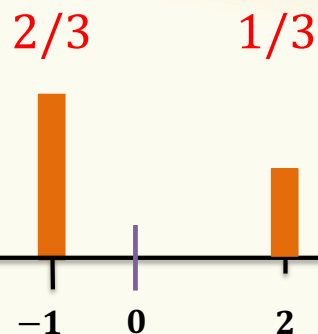
$$P(W_2 = 10) = \frac{1}{3}, P(W_2 = -5) = \frac{2}{3}$$

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

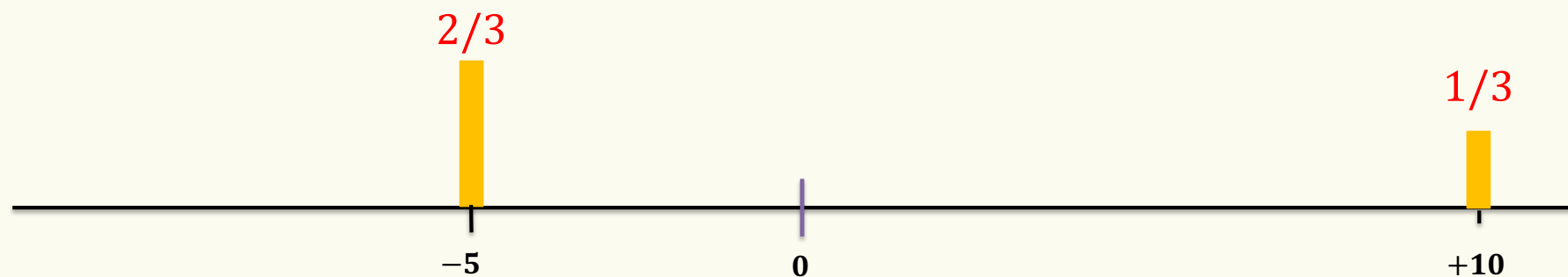
## Two Games

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

Somehow, Game 2 has higher volatility / exposure!



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



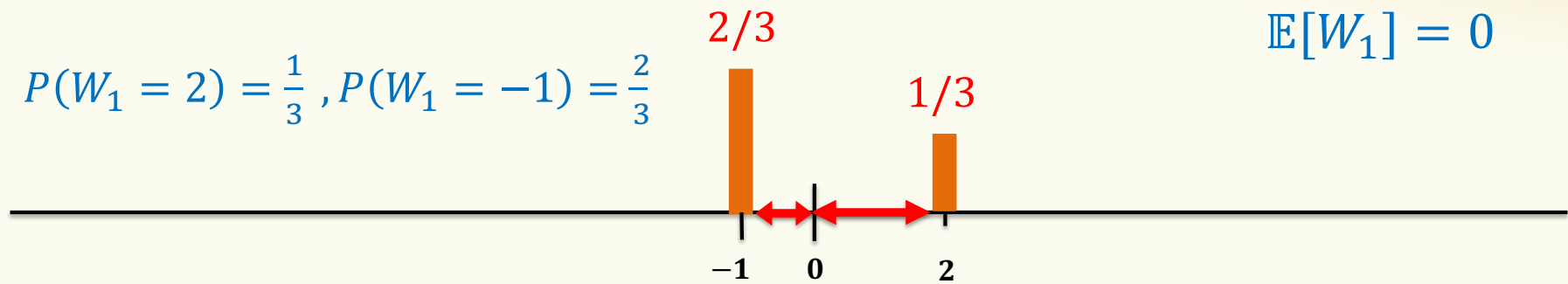
Same expectation, but clearly a very different distribution.

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

## Variance (Intuition, First Try)

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

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



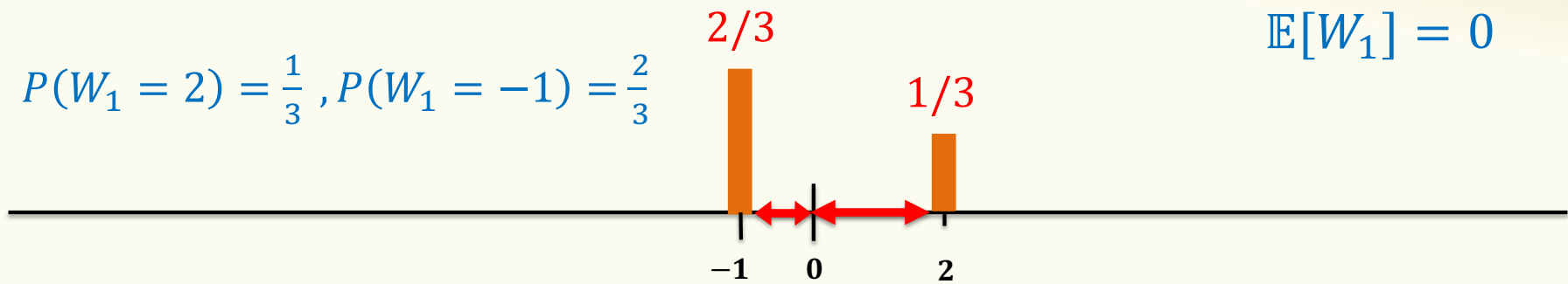
New quantity (random variable): How far from the expectation?

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

## 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?

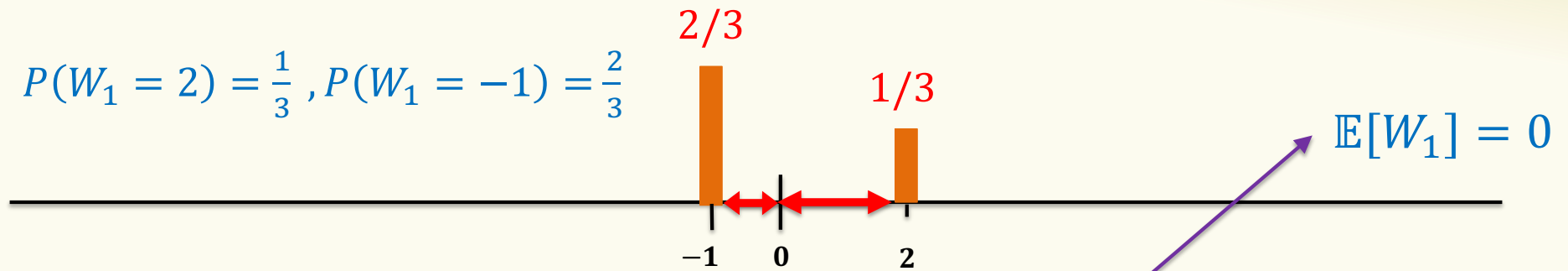
$$W_1 - \mathbb{E}[W_1]$$

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

## Variance (Intuition, Better Try)

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

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

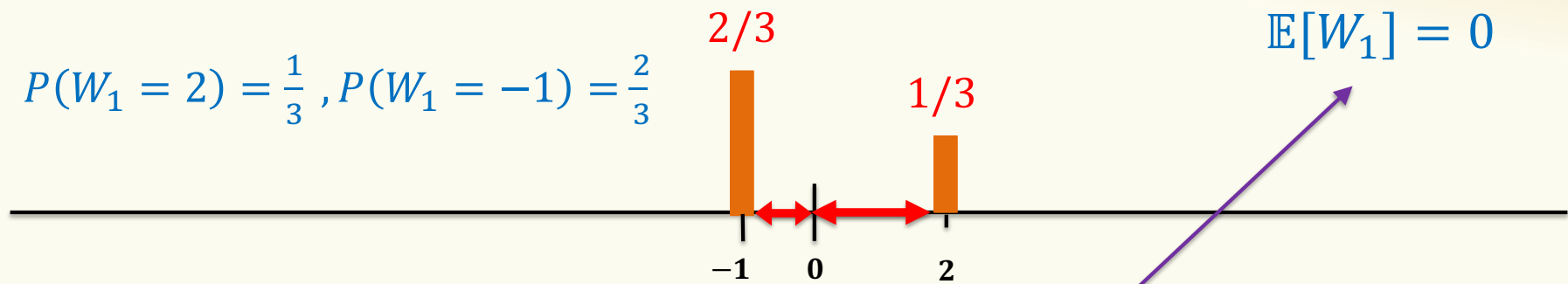


A better quantity (random variable): How far from the expectation?

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

## Variance (Intuition, Better Try)

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



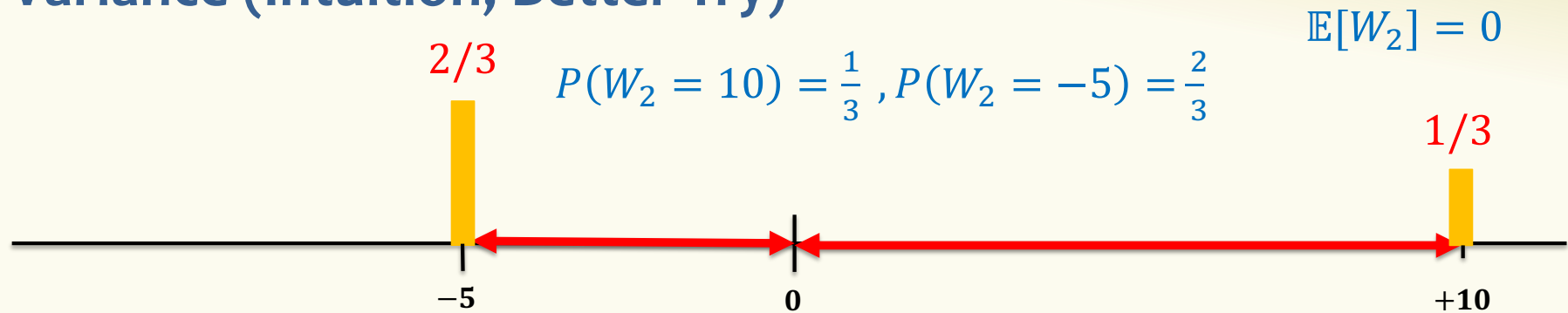
A better quantity (random variable): How far from the expectation?

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

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

$$= 2$$

## Variance (Intuition, Better Try)

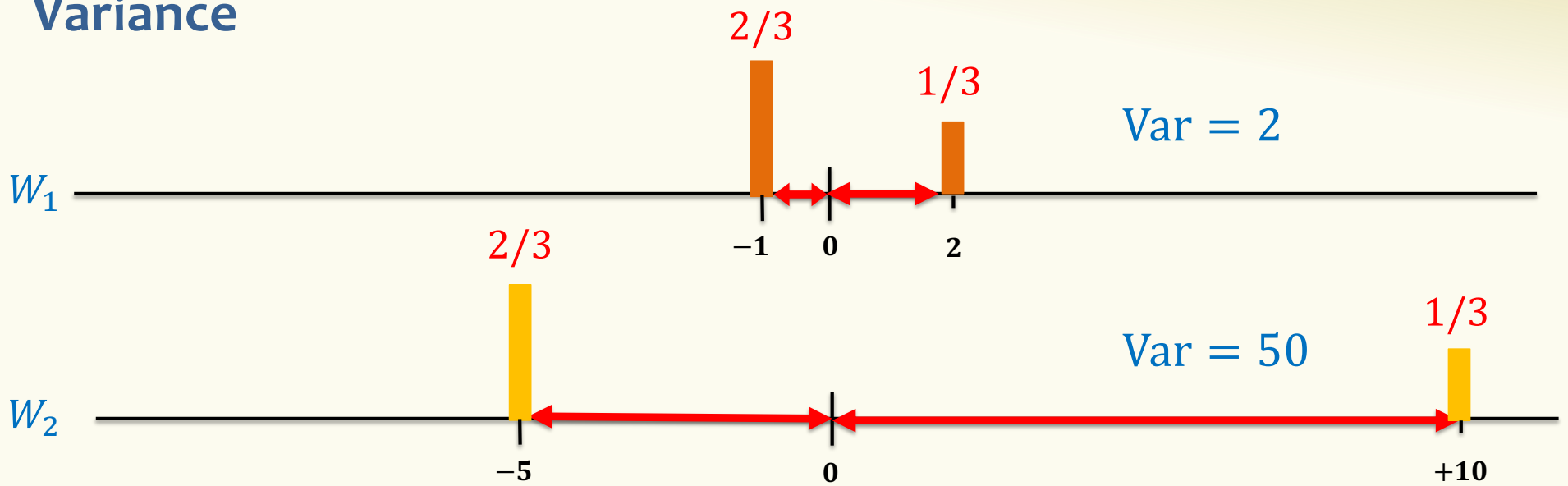


A better quantity (random variable): How far from the expectation?

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



## Variance



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

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

## 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

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

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

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

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

## Variance – Example 1

$X$  fair die

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

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

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

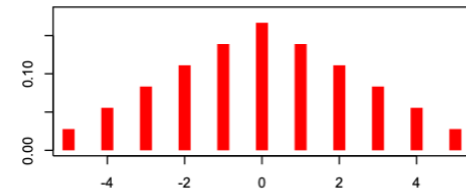
$$= \frac{2}{6} [2.5^2 + 1.5^2 + 0.5^2] = \frac{2}{6} \left[ \frac{25}{4} + \frac{9}{4} + \frac{1}{4} \right] = \frac{35}{12} \approx 2.91677 \dots$$

## Variance in Pictures

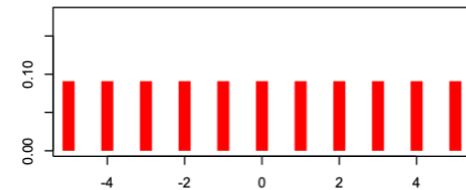
Captures how much  
“spread” there is in a pmf

All pmfs have same  
expectation

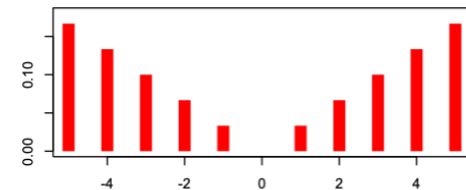
$$\sigma^2 = 5.83$$



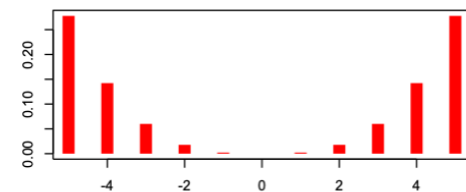
$$\sigma^2 = 10$$



$$\sigma^2 = 15$$



$$\sigma^2 = 19.7$$



# Agenda

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

## Variance

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

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

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

## 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

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

- LOTUS
- 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

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

## (Not Covered) 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  $p$

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

$$P(Z = k) =$$

## 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


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

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

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! [Verify it formally!]


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

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



## Questions

The **variance** of a (discrete) RV  $X$  is

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

- Can the variance of a random variable be negative?
- Is  $\text{Var}(X + 5) = \text{Var}(X) + 5$ ?
- Is it true that if  $\text{Var}(X) = 0$ , then  $X$  is a constant?
- What is the relationship between  $\mathbb{E}(X^2)$  and  $[\mathbb{E}(X)]^2$  ?

## Independence of random variables

- Suppose  $X$  and  $Y$  are independent indicator random variables taking the value 1 with probability  $\frac{1}{2}$ , and let  $Z = XY$ 
  - Are  $X$  and  $Z$  independent?
  - Are  $Y$  and  $Z$  independent?
- Is it true that if  $X$  and  $Y$  are independent, and  $Y$  and  $Z$  are independent, then  $X$  and  $Z$  are independent?

## Agenda

- LOTUS
- Variance
- Properties of Variance
- Independent Random Variables
- Properties of Independent Random Variables
- An Application: Bloom Filters! ◀

## Basic Problem

**Problem:** Store a subset  $S$  of a large set  $U$ .

**Example.**  $U$  = set of 128 bit strings  
 $S$  = subset of strings of interest

$$|U| \approx 2^{128}$$

$$|S| \approx 1000$$

### Two goals:

1. **Very fast** (ideally constant time) answers to queries “Is  $x \in S$ ?” for any  $x \in U$ .
2. **Minimal storage** requirements.

## Naïve Solution I – Constant Time

**Idea:** Represent  $S$  as an array  $A$  with  $2^{128}$  entries.

$$A[x] = \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{if } x \notin S \end{cases}$$

$S = \{0, 2, \dots, K\}$

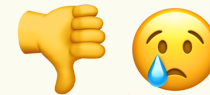


0	1	2	...	K	...		
1	0	1	0	1	...	0	0

**Membership test:** To check  $x \in S$  just check whether  $A[x] = 1$ .

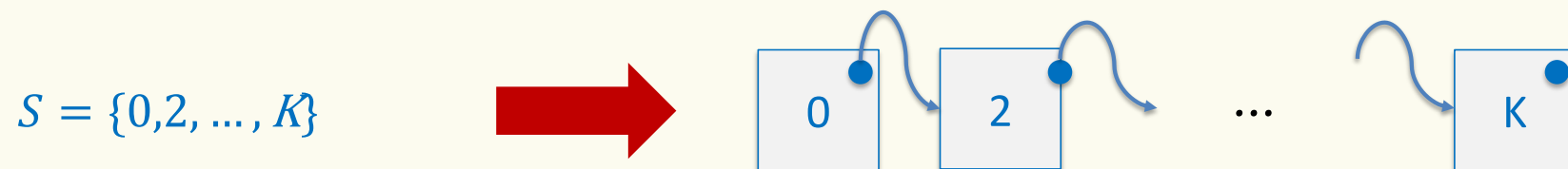
→ **constant time!** 👍 😊

**Storage:** Require storing  $2^{128}$  bits, even for small  $S$ .



## Naïve Solution II – Small Storage

**Idea:** Represent  $S$  as a list with  $|S|$  entries.



**Storage:** Grows with  $|S|$  only 👍 😊

**Membership test:** Check  $x \in S$  requires time linear in  $|S|$

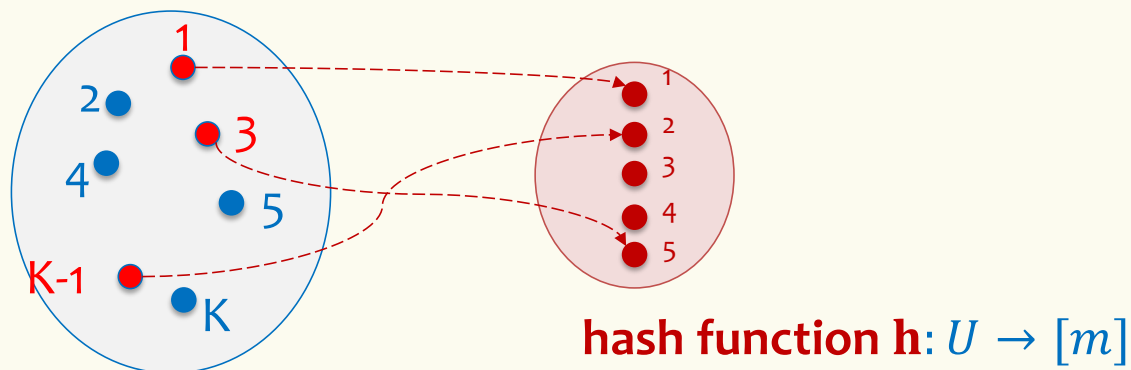
(Can be made logarithmic by using a tree) 👎 😓

## Hash Table

**Idea:** Map elements in  $S$  into an array  $A$  of size  $m$  using a hash function  $h$

**Membership test:** To check  $x \in S$  just check whether  $A[h(x)] = x$

**Storage:**  $m$  elements (size of array)

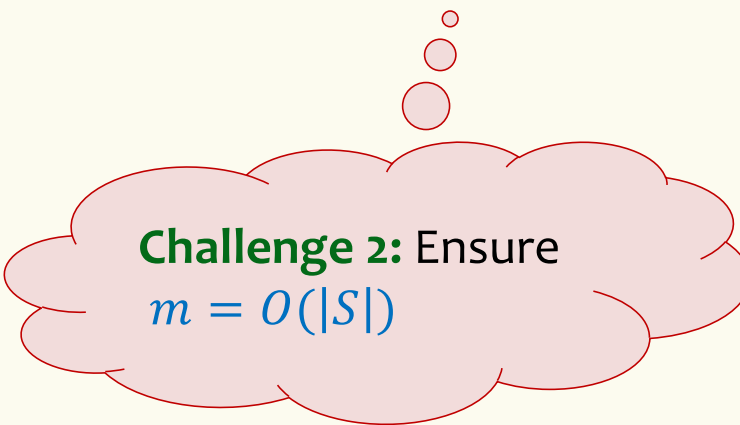


## Hash Table

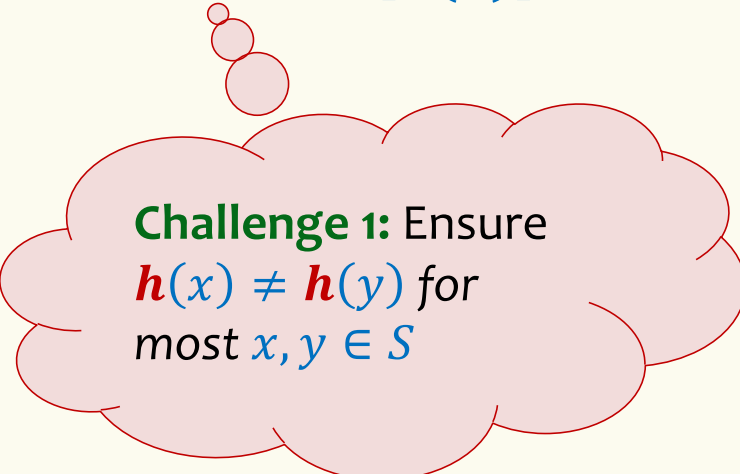
**Idea:** Map elements in  $S$  into an array  $A$  of size  $m$  using a hash function  $h$

**Membership test:** To check  $x \in S$  just check whether  $A[h(x)] = x$

**Storage:**  $m$  elements (size of array)



**Challenge 2:** Ensure  
 $m = O(|S|)$



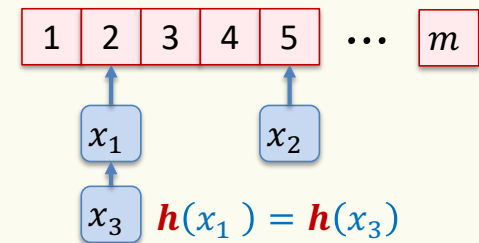
**Challenge 1:** Ensure  
 $h(x) \neq h(y)$  for  
most  $x, y \in S$



## Hashing: collisions

**Collisions** occur when  $h(x) = h(y)$  for some distinct  $x, y \in S$ , i.e., two elements of set map to the same location

- Common solution: chaining – at each location (bucket) in the table, keep linked list of all elements that hash there.



## Good hash functions to keep collisions low

- The hash function  $h$  is good iff it
  - distributes elements uniformly across the  $m$  array locations so that
  - pairs of elements are mapped independently

“Universal Hash Functions” – see CSE 332

## Hashing: summary

### Hash Tables

- They store the data itself
- With a good hash function, the data is well distributed in the table and lookup times are small.
- However, they need at least as much space as all the data being stored, i.e.,  $m = \Omega(|S|)$

In some cases,  $|S|$  is huge, or not known a-priori ...

Can we do better!?

## Next time: Bloom Filters

- Probabilistic data structure.
- Close cousins of hash tables.
  - But: Ridiculously space efficient
- Occasional errors, specifically false positives.





