Class starts after this song

Dominic Fike, Weezer – Think Fast (2023)
requested by Nicole Errera (TA-of-CM8)

I'm a third-year majoring in CS with a concentration in AI/ML. I'm from Sacramento, CA and I love being outside.

Duke

UTA applications are open!

- Apply at https://cs.duke.edu/undergraduate/uta

- Apply by 4/22 (soft deadline)
- Can apply to multiple classes; can even work for multiple
- UTAs are hired to help people, not ace the material

CS230 Spring 2024
Module 08: Probability

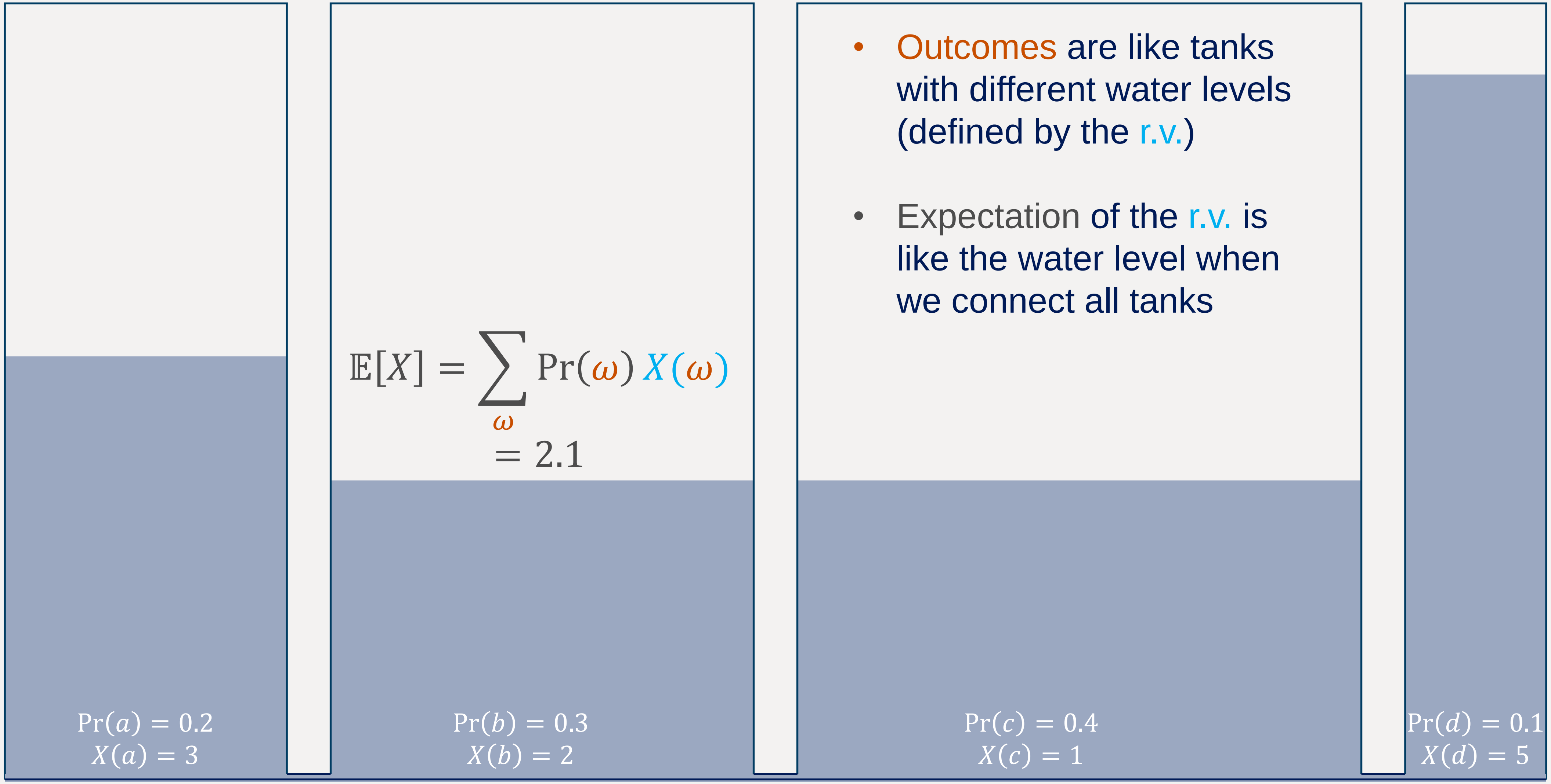
Duke

Expectation

- Nothing fancier than a weighted average
- May not “fall on” any existing outcome
  - There are 8 CMs in CS230
  - 3 of them have 3 class meetings; the other 5 have 2 class meetings
  - Weighted average $\frac{19}{8} = 2.375$ class meetings
  - No modules can have 2.375 class meetings and that is okay

- Outcomes are like tanks with different water levels (defined by the r.v.)
- Expectation of the r.v. is like the water level when we connect all tanks

$$\mathbb{E}[X] = \sum_{\omega} \Pr(\omega)\, X(\omega)$$
$$= 2.1$$



WOTO: expectation practice with familiar items

- Consider the Notable Women in Computing cards in CM7 recitation, where the card values are defined as $A = 1, J = 11, Q = 12, K = 13$ for letter cards (assume no jokers)
  - If we draw a card uniformly at random, what is the expected value of the card we draw?
  - If we draw a card such that somehow letter cards (A,J,Q,K) are twice likely to be drawn (as opposed to regular number cards), what is the expected value of the card we draw?

Duke

WOTO: expectation practice with familiar items

- If we draw a card uniformly at random, what is the expected value of the card we draw? 7
- If we draw a card such that somehow letter cards (A,J,Q,K) are twice likely to be drawn (as opposed to regular number cards), what is the expected value of the card we draw?
  - $\Pr(A) = \Pr(J) = \Pr(Q) = \Pr(K) = \frac{2}{17}, \Pr(2) = \cdots = \Pr(10) = \frac{1}{17}$
  - $(1 + 11 + 12 + 13) \times \frac{2}{17} + (2 + \cdots + 10) \times \frac{1}{17} = \frac{128}{17} \approx 7.53 > 7$

Duke

Peer discussion: Expectation

Recall Grace the grasshopper from the CM7 Gradescope assignment. (Try reading this sentence real quick.)

Anyway, as a reminder, at each step, here are the four options Grace can hop:

- +1 unit vertically and +3 units horizontally
- +1 unit vertically and -3 units horizontally
- -1 unit vertically and +3 units horizontally
- -1 unit vertically and -3 units horizontally

Now assume Grace takes each option with equal probability 1/4. **What is Grace's expected movement each step?**

Duke

Peer discussion: Expectation

- In expectation, Grace does not move each step
- In expectation, Grace moves for a distance of $\sqrt{10}$ each step
- Both are correct! The problem lies in we did not define the underlying random variable in an unambiguous way.
- Expectations are defined on random variables. No point in talking about expectations without a concrete random variable.

WOTO: linearity of expectation

- Recall from the recitation the Game of Coin: Alice and Bob are playing a game where they toss a fair coin until there is a sequence of three coin tosses resulting in HHH or THH.
  - Alice wins if the first HHH appears before the first THH.
  - Bob wins if the first THH appears before the first HHH.
- Question: in 100 tosses, how many times in expectation does HHH and THH happen?

WOTO: linearity of expectation

- Question: in 100 tosses, how many times in expectation does HHH and THH happen?
  - Define A_i to be the indicator variable that tosses $i, i+1, i+2$ are HHH
  - We are interested in $\mathbb{E}[\sum_{i=1}^{98} A_i]$
  - By linearity of expectation:

$$\mathbb{E}\left[\sum_{i=1}^{98} A_i\right] = \sum_{i=1}^{98} \mathbb{E}[A_i] = \sum_{i=1}^{98} \Pr(A_i = 1) = \sum_{i=1}^{98} \frac{1}{8} = \frac{98}{8}$$

Duke

WOTO: linearity of expectation

- Question: in 100 tosses, how many times in expectation does HHH and THH happen?
  - Define B_i to be the indicator variable that tosses $i, i+1, i+2$ are THH
  - We are interested in $\mathbb{E}[\sum_{i=1}^{98} B_i]$
  - By linearity of expectation:

$$\mathbb{E}\left[\sum_{i=1}^{98} B_i\right] = \sum_{i=1}^{98} \mathbb{E}[B_i] = \sum_{i=1}^{98} \Pr(B_i = 1) = \sum_{i=1}^{98} \frac{1}{8} = \frac{98}{8}$$

WOTO: linearity of expectation

- Question: in 100 tosses, how many times in expectation does HHH and THH happen?
  - They both happen $\frac{98}{8}$ times in expectation despite THH having an 87.5% chance to appear first!
  - Also note that the indicator variables are not independent.
    But linearity of expectation always holds regardless of independence.

- NOT "at the same scale of X"
- many good properties!
- better for calculation

- "at the same scale of X"
- not many good properties
- better for interpretation

Variance and Standard Deviation

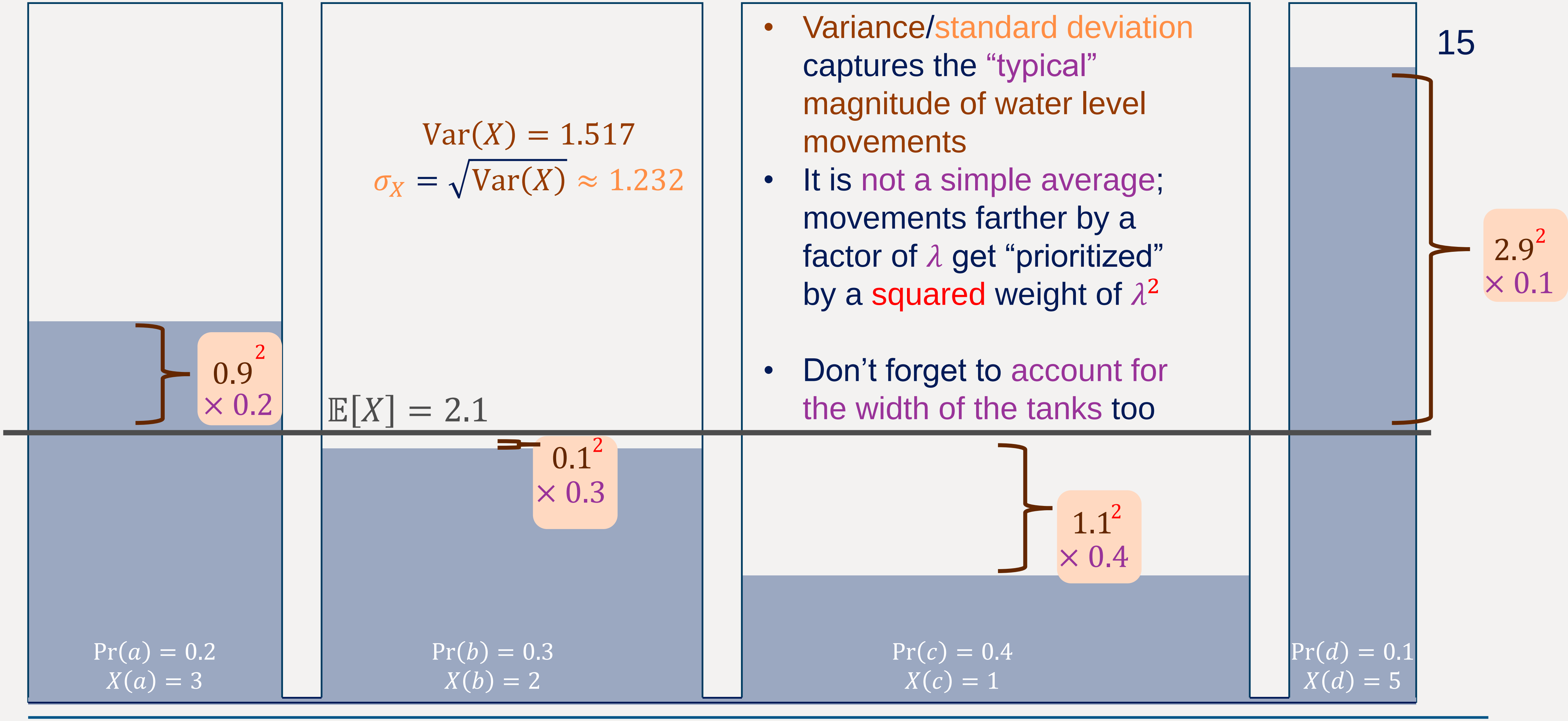
- $\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$
- $\text{Var}(X) = \mathbb{E}[(X - \mu)^2] = \mathbb{E}[X^2] - \mu^2$
- $\sigma_X = \text{std}(X) = \sqrt{\text{Var}(X)}$

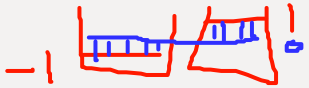
- Convention: $\mathbb{E}[X] = \mu$ (meaningful; this conveys μ is a constant and consists of no randomness)

Duke

- Variance/standard deviation captures the "typical" magnitude of water level movements
- It is not a simple average; movements farther by a factor of λ get "prioritized" by a squared weight of λ^2
- Don't forget to account for the width of the tanks too

$$\mathrm{Var}(X) = 1.517$$
$$\sigma_X = \sqrt{\mathrm{Var}(X)} \approx 1.232$$





WOTO: Variance and STD Examples

Random variable	X_1	X_2	X_3	$X_4 = X_2 + X_3$
Distribution	$\Pr[X_1 = 0] = 1$	$\Pr[X_2 = -1] = ½$ $\Pr[X_2 = +1] = ½$	$\Pr[X_3 = -10] = ½$ $\Pr[X_3 = +10] = ½$	$\Pr[X_4 = -11] = 1/4$ $\Pr[X_4 = -9] = 1/4$ $\Pr[X_4 = +9] = 1/4$ $\Pr[X_4 = +11] = 1/4$
μ_i	0	0	0	0
$\mathrm{Var}(X_i)$	0	1	100	101
σ_i	0	1	10	$\sqrt{101}$

NOT “at the same scale of X”

“at the same scale of X”

WOTO: Variance and STD Examples

Random variable	X_1	X_2	X_3	X_4
Distribution	$\Pr[X_1 = 0] = 1$	$\Pr[X_2 = -1] = ½$ $\Pr[X_2 = +1] = ½$	$\Pr[X_3 = -10] = ½$ $\Pr[X_3 = +10] = ½$	$\Pr[X_4 = -11] = 1/4$ $\Pr[X_4 = -9] = 1/4$ $\Pr[X_4 = +9] = 1/4$ $\Pr[X_4 = +11] = 1/4$
μ_i	0	0	0	0
$\mathrm{Var}(X_i)$	0	1	100	101
σ_i	0	1	10	$\sqrt{101}$

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

$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$

Warning: only if independent!

Duke

Expectation, Variance and STD Cheat Sheet

Random variable	Bernoulli(p)	Binomial(n, p)	Geometric(p)
μ	p	np	$1/p$
$\mathrm{Var}(X)$	$p(1-p)$	$np(1-p)$	$\frac{1-p}{p^2}$
σ	$\sqrt{p(1-p)}$	$\sqrt{np(1-p)}$	$\frac{\sqrt{(1-p)}}{p}$

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

$$\mathrm{Var}(X + \cdots + X) = \mathrm{Var}(X) + \cdots + \mathrm{Var}(X)$$

Warning: here we know they are independent!

- Really no need to memorize; you can always derive yourself
- Memorizing them does help; just like Java STL

Markov's Inequality/Bound

$$\Pr(X \geq c\mathbb{E}[X]) \leq \frac{1}{c}$$

$$\Pr(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

- Intuition:
  given the expectation (connected water level), each tank cannot be too high and too wide at the same time
- Otherwise water in that tank alone is too much
- Therefore, given the height, we have an upper bound for the width
- The equation holds only when all other tanks are empty: $\Pr(X < a) = 0$
- Warning: ***X* needs to be nonnegative** (otherwise the tank analogy does not work)

$c\mathbb{E}[X]$

$\leq \frac{1}{c}$

Chebyshev's Inequality/Bound

$$\Pr(|X - \mathbb{E}[X]| \geq c\sigma_x) \leq \frac{1}{c^2}$$

$$\Pr(|X - \mathbb{E}[X]| \geq \alpha) \leq \frac{\mathrm{Var}(X)}{\alpha^2}$$

$$\Pr((X - \mathbb{E}[X])^2 \geq \beta) \leq \frac{\mathrm{Var}(X)}{\beta}$$

- Intuition: given the variance (water level movements), a tank cannot be too high/low *(compared to the expectation)* and too wide at the same time
- Otherwise water in that tank alone **moves** too much
- Therefore, given the height *(compare to the expectation)*, we again have an upper bound for the width

- Many different variations – they all mean the same thing
- ***X* does not need to be nonnegative**

Proving Chebyshev's Inequality/Bound Using Markov's Inequality/Bound

$\Pr(|X - \mathbb{E}[X]| \geq c\sigma_x) \leq \frac{1}{c^2}$

$=$

$\Pr(|X - \mathbb{E}[X]| \geq \alpha) \leq \frac{\mathrm{Var}(X)}{\alpha^2}$

$=$

$\Pr((X - \mathbb{E}[X])^2 \geq \beta) \leq \frac{\mathrm{Var}(X)}{\beta}$ $=$ $\Pr(Y \geq \beta) \leq \frac{\mathbb{E}[Y]}{\beta}$

Chebyshev's Inequality/Bound — Markov's Inequality/Bound

- Intuition: let's now define the squared water level movement $(X - \mathbb{E}[X])^2$ as a new random variable Y
- What is $\mathbb{E}[Y]$?
- It's just $\mathrm{Var}(X)$
- Apply Markov's Inequality/Bound on Y

PI: Markov's and Chebyshev's Inequalities/Bounds

$$\Pr(X \geq c\mathbb{E}[X]) \leq \frac{1}{c}$$

$$=$$

$$\Pr(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

$$\Pr(|X - \mathbb{E}[X]| \geq c\sigma_x) \leq \frac{1}{c^2}$$

$$=$$

$$\Pr(|X - \mathbb{E}[X]| \geq \alpha) \leq \frac{\mathrm{Var}(X)}{\alpha^2}$$

$$=$$

$$\Pr((X - \mathbb{E}[X])^2 \geq \beta) \leq \frac{\mathrm{Var}(X)}{\beta}$$