Probability Theory

CS 70 Discussion 11B

Raymond Tsao

2025-04-11

Note: These slides are unofficial course materials. Please use the notes as the only single source of truth.

Expectation and Variance

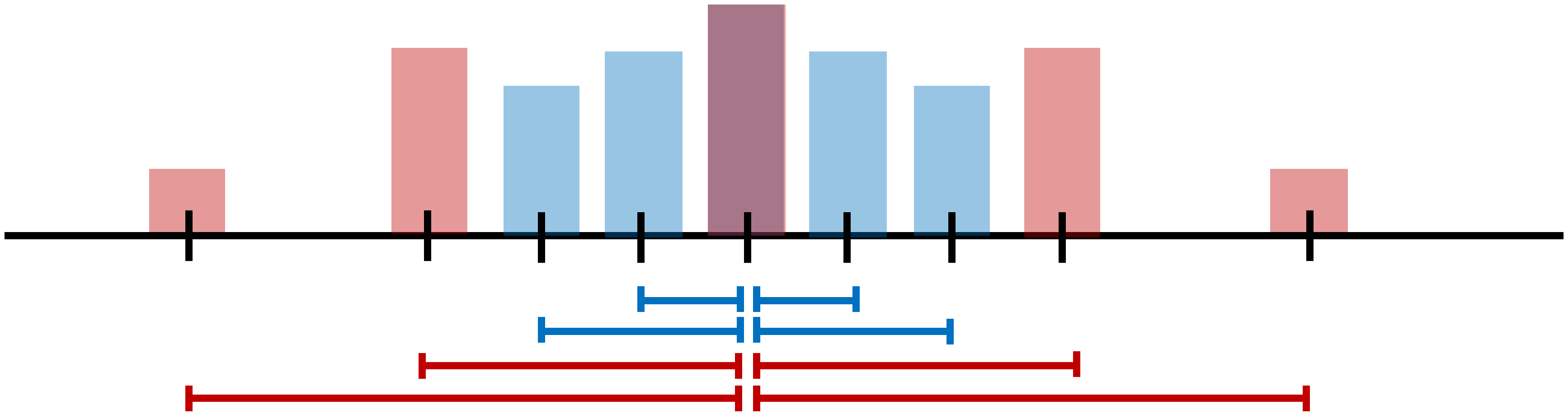
Expectation: Measures where the distribution is centered

$$\mathbb{E}[X] = \sum_x x\mathbb{P}[X = x]$$

Variance: Measures how “spread out” a distribution is

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

distance to the center of mass



Calculating Variance

Strategy 1: Follow the definition

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

Strategy 2: Alternative expression for variance (we usually use this)

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

Second moment

Aside: Expectation of functions of random variables

$$\mathbb{E}[g(X)] = \sum_x g(x)p(x)$$

For example:

$$\mathbb{E}[X^2] = \sum_x x^2 p(x)$$

values weighted by the probability

Calculating Variance

Strategy 3 (Powerful but a lot of calculations): Decompose X into sums of other R.V.

$$X = \sum_{i=1}^{n} X_i \qquad Var(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

??? We know how to calculate this

$$\mathbb{E}[X^2] = \mathbb{E}\left[\left(\sum_{i=1}^{n} X_i\right)^2\right] = \mathbb{E}\left[\sum_{i}^{n} X_i^2 + \sum_{i \neq j} X_i X_j\right]$$

$$= \sum_{i=1}^{n} \mathbb{E}[X_i^2] + \sum_{i \neq j} \mathbb{E}[X_i X_j]$$

Second moment of X_i Cross terms

Special case: if X_i are Bernoullis

$$\mathbb{E}[X_i^2] = \mathbb{E}[X_i]$$

$$\mathbb{E}[X_i X_j] = \mathbb{P}[X_i = 1, X_j = 1]$$

Problem 1: Pullout Balls

(a)

Way 1: Use the definition

$\mathbb{E}[X] = \frac{7}{2}$

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

$$= \left(1 - \frac{7}{2}\right)^2 \cdot \frac{1}{6} + \left(2 - \frac{7}{2}\right)^2 \cdot \frac{1}{6} + \cdots + \left(6 - \frac{7}{2}\right)^2 \cdot \frac{1}{6} = \frac{35}{12}$$

Way 2:

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

$$\mathbb{E}[X^2] = 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + \cdots 6^2 \cdot \frac{1}{6} = \frac{91}{6}$$

$$= \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$$

Problem 1: Pullout Balls

(b) Let X_i be the outcome of the ith roll, then

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

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

Only holds when independent!

Variance is not linear!

$Var(aX) = a^2 Var(X)$

$$= \frac{1}{n^2} Var\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i)$$

$$= \frac{1}{n^2} \cdot n \cdot \frac{35}{12} = \frac{35}{12n}$$

Problem 2: Elevator Variance

Step 1: Define appropriate sub-events

Observation: We can counting the number of occurrences out of n trials (floors)

Let X_i be a Bernoulli that detects whether the elevator stops at floor i or not.

$$X = \sum_{i=1}^{n} X_i$$

To compute the variance, we need the expectation and second moment:

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

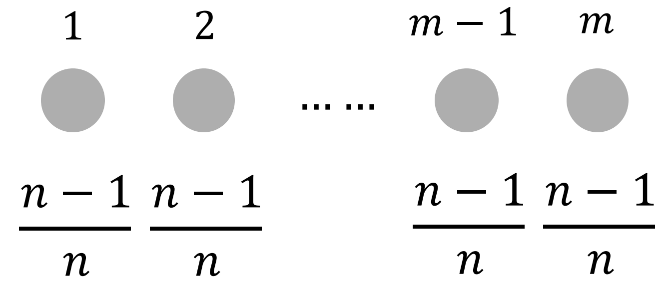
Problem 2: Elevator Variance

Step 2: Compute expectation

Probability that elevator does not stop at floor i

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} \mathbb{P}[X_i = 1] = n\boxed{\left(\frac{n-1}{n}\right)^m}$$

Step 3: Compute second moment



Cannot get off at floor i

$$\mathbb{E}[X^2] = \mathbb{E}\left[\left(\sum_{i=1}^{n} X_i\right)^2\right]$$

$$= \sum_{i=1}^{n} \mathbb{E}[X_i^2] + \sum_{i \neq j} \mathbb{E}[X_i X_j]$$

For Bernoulli, same as $\mathbb{E}[\mathrm{X_i}]$

Problem 2: Elevator Variance

$$\mathbb{E}[X_i X_j] = \mathbb{P}[X_i = 1, X_j = 1] = \left(\frac{n-2}{n}\right)^m$$

Probability that elevator does not stop at floor i and j

Plugging in:

$$\mathbb{E}[X^2] = \sum_{i=1}^{n} \mathbb{E}[X_i^2] + \sum_{i \neq j} \mathbb{E}[X_i X_j]$$

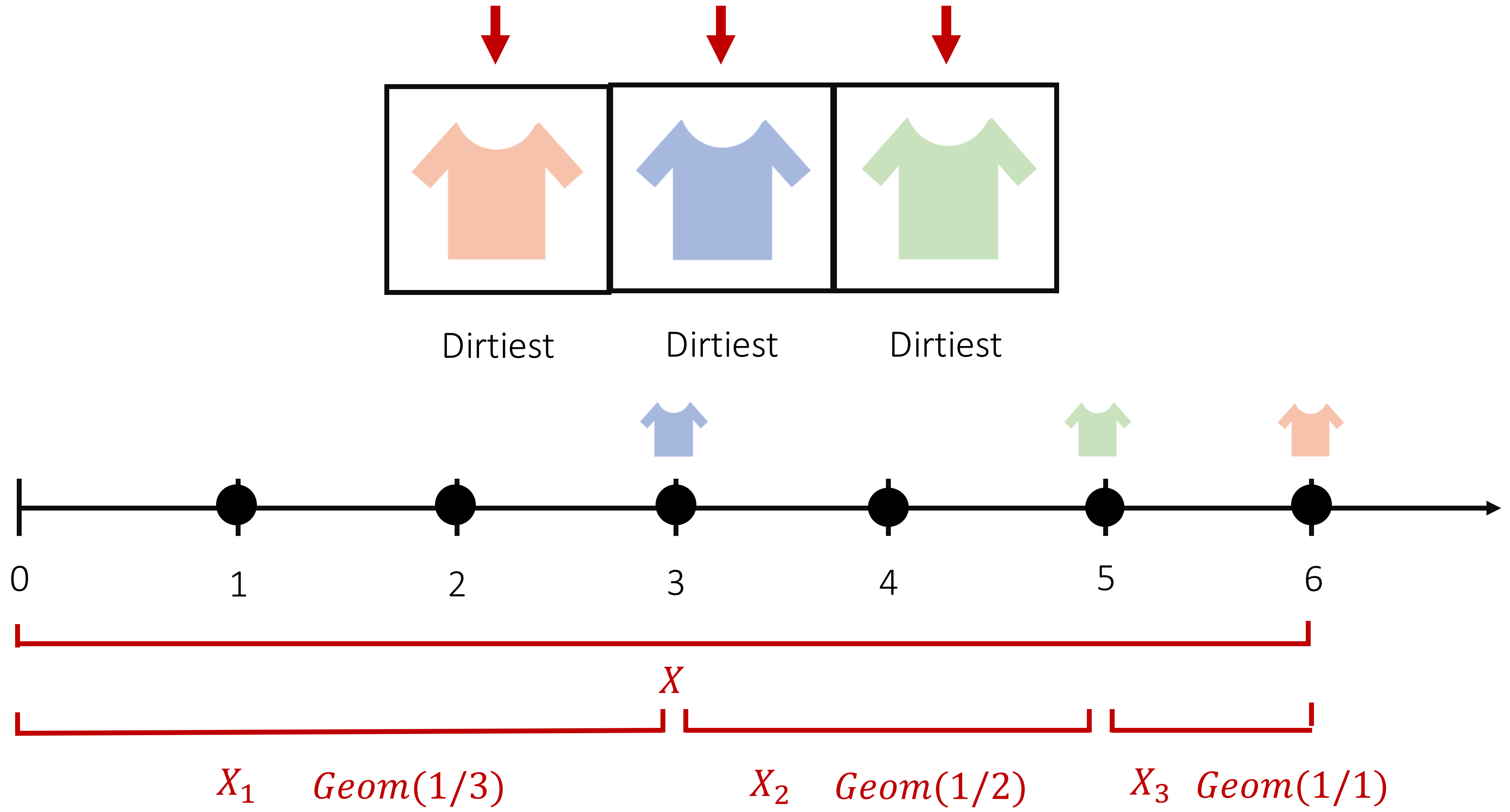
$$= n\left(\frac{n-1}{n}\right)^m + n(n-1)\left(\frac{n-2}{n}\right)^m$$

And finally

$$Var(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = n\left(\frac{n-1}{n}\right)^m + n(n-1)\left(\frac{n-2}{n}\right)^m - \left(n\left(\frac{n-1}{n}\right)^m\right)^2$$

Problem 3: Student Life

Let's try to understand the process with 3 shirts:



Problem 3: Student Life

Define X_i the same way, then $X_i \sim Geom(\frac{1}{n+i-1})$

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i]$$

$$= \sum_{i=1}^{n} (n+i-1) = \frac{n(n+1)}{2}$$

Since X_i are independent, the variance is easy…

If $X \sim Geom(p)$, then $\mathrm{Var}(X) = \frac{1-p}{p^2}$

$$Var(X) = Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) = \sum_{i=1}^{n} (n+i-1)^2 (1 - \frac{1}{n+i-1})$$

Only works with independent X_is