Probability Theory

CS 70 Discussion 11A

Raymond Tsao

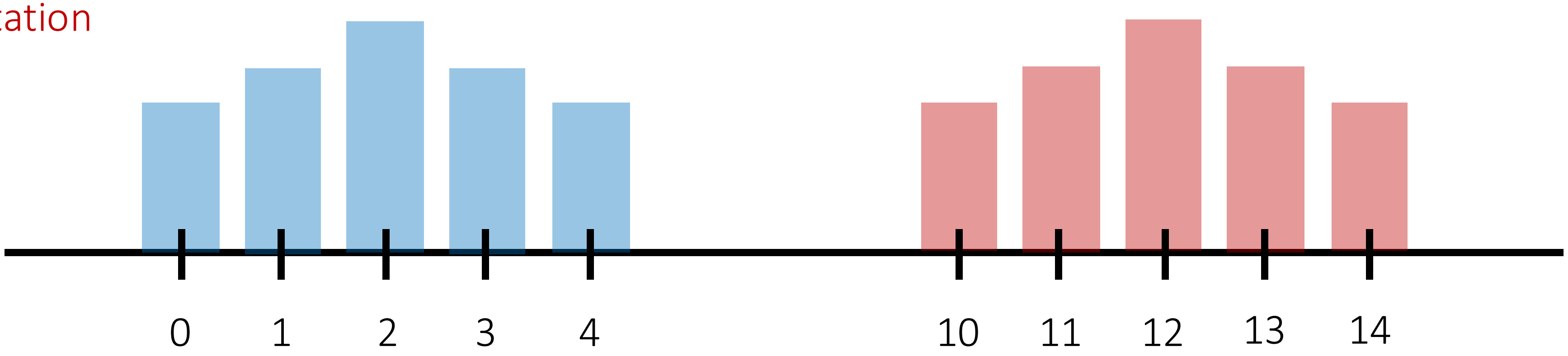
2025-04-09

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

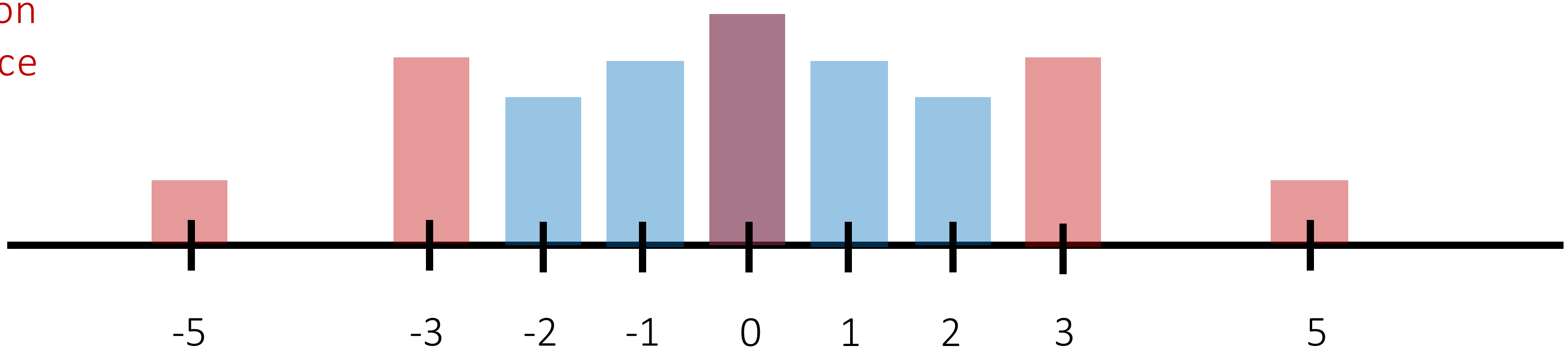
Expectation and Variance

Expectation and variance: two fundamental statistics of a distribution

Different expectation
Same variance



Same expectation
Different variance



Expectation: Three Strategies

Strategy 1: Follow the definition

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

Strategy 2: Tail sum formula

$$\mathbb{E}[X] = \sum_{t \geq 1} \mathbb{P}[X \geq t]$$

Aside: Tail sum formula

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

$$= 1 \cdot \mathbb{P}[X = 1] + 2 \cdot \mathbb{P}[X = 2] + 3 \cdot \mathbb{P}[X = 3] + \cdots$$

$$= 1 \cdot \mathbb{P}[X = 1] + 1 \cdot \mathbb{P}[X = 2] + 1 \cdot \mathbb{P}[X = 3] + \cdots \qquad \mathbb{P}[X \geq 1]$$

$$+ 1 \cdot \mathbb{P}[X = 2] + 1 \cdot \mathbb{P}[X = 3] + \cdots \qquad \mathbb{P}[X \geq 2]$$

$$+ 1 \cdot \mathbb{P}[X = 3] + \cdots \qquad \mathbb{P}[X \geq 3]$$

$$+ \cdots$$

$$= \sum_{t \geq 1} \mathbb{P}[X \geq t]$$

Expectation: Three Strategies

Strategy 1: Follow the definition

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

Strategy 2: Tail sum formula

$$\mathbb{E}[X] = \sum_{t\geq 1} \mathbb{P}[X \geq t]$$

Strategy 3 (Very powerful!): Linearity

Decompose X as sums of simpler random variables X_i

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_i X_i\right] = \sum_i \mathbb{E}[X_i]$$

Problem 1: Pullout Balls

(a) Let X be the number we picked up

- Sample space: $X = \{1, 2, 3, 4\}$
- Probability function: $\mathbb{P}[X = x] = 1/4$

$$\mathbb{E}[X] = 1 \cdot \frac{1}{4} + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{4} + 4 \cdot \frac{1}{4} = \frac{5}{2}$$

(b) Let Y be the product of the two numbers we picked up

- Sample space: $\mathrm{Y} = \{2, 3, 4, 6, 8, 12\}$
- Probability function: $\mathbb{P}[X = x] = 1/6$

$$\mathbb{E}[X] = 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} + 8 \cdot \frac{1}{6} + 12 \cdot \frac{1}{6} = \frac{35}{6}$$

Problem 2: Linearity

(a) Let X be the total number of tickets you receive

- Complicated probability function, so probably need to use linearity…

$$X = A_1 + A_2 + \cdots A_{10} + B_1 + B_2 + \cdots + B_{20}$$

1 if wins, 0 otherwise

To count the number of occurences in n trials:
Define a Bernoulli for the ith trial / position

$$\mathbb{E}[X] = \mathbb{E}[A_1] + \mathbb{E}[A_2] + \cdots \mathbb{E}[A_{10}] + \mathbb{E}[B_1] + \mathbb{E}[B_2] + \cdots + \mathbb{E}[B_{20}]$$

$$= 10 \cdot 3 \cdot \frac{1}{3} + 20 \cdot 4 \cdot \frac{1}{5} = 26$$

$$\mathbb{E}[A_i] = 1 \cdot \frac{1}{3} + 0 \cdot \frac{2}{3} = \frac{1}{3}$$

Problem 2: Linearity

(b) Let X be the number of times the sequence "book" appears

$$X = X_1 + X_2 + \cdots + X_{999997}$$

1 if there's an occurrence at position i, 0 otherwise

m	b	o	o	k	a	l	m	b	o	o	k
X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9			
0	1	0	0	0	0	0	0	1			

If $X_2 = 1$ then $X_3 \neq 1$! So these are not independent!

Problem 2: Linearity

Regardless of whether the X_i are independent or not, we can always apply linearity!

$$X = X_1 + X_2 + \cdots + X_{999997}$$

$$\mathbb{E}[X] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_{999997}]$$

$$= 999997 \cdot \frac{1}{26^4}$$

Extra: Expectation of Hypergeometric Distribution

Suppose n balls are sampled from an urn with N balls of which m are white.

Let X be number of white balls, what is the distribution of X?

$$\mathbb{P}[X = k] = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}$$

What is the expectation?

Assign each white ball a Bernoulli!

1 if the ith white ball is sampled

1   2   … m

X_1 X_2 X_m

Extra: Expectation of Negative Binomial Distribution

Suppose X is the number of trials needed to observe r success (with prob p).

$$\mathbb{P}[X = k] = \binom{k-1}{r-1} p^r (1-p)^{k-r}$$

$$X = X_1 + X_2 + X_3 + \cdots + X_r$$

Number of trials needed to observe 1st success