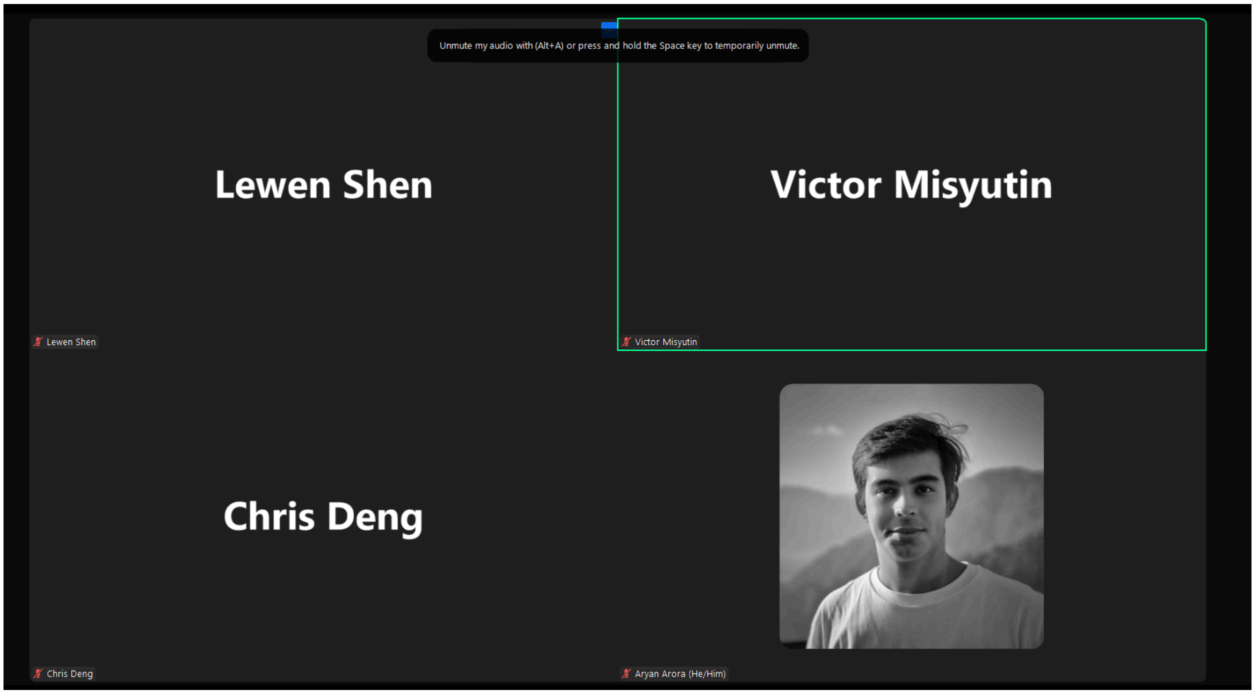
October 16th



Questions to answer:

- Can you explain why 2X is not the same as X+Y even if X and Y have the same distribution?

If X and Y are independent, then adding X and Y introduces additional variability, Var(2X)=4Var(X) which is different from Var(X)+Var(Y)+2Cov(X,Y).

- Can you describe what it means that two random variables have the same distribution?

This means that they have the same probability distribution, it can be PDF, CDF or CMF, and also the same mean and variance.

- Can you describe what is the Binomial distribution? Give an example of Binomial distri. in your description.

Binomial Distribution is a number of independent trials with either yes or no, and for each trial, the probability of yes and no does not change. For example, each time you flip a coin, it is either a head or not a head, and the previous time does not influence the result of later time.

- Can you describe what is the Geometric distribution? Give an example of Geometric distri. in your description.

Similar as the previous one, Geometric distribution is a number of discrete trials with either yes or no, and for each trial, the probability of yes and no does not change. Geometric distribution focuses on the first time to reach out yes. For example, you toss a die until the first time you get a six.

- Can you explain how different the Binomial distri. and Geometric distri. are even if they are both related to the Bernoulli trials?

The binomial distribution is concerned with the number of yes in a fixed number of independent Bernoulli trials, geometric distribution focuses on the first time of succession.

- Can you show with example how to calculate the probability of a normally distributed RV within certain interval [a,b] using the cdf table of standard normal? Do NOT use the same example in the video.

Suppose X is normally distributed with mean $\mu=50$ and standard deviation $\sigma=10$. We want to calculate the probability that X lies between 40 and 60, $P(40 \leq X \leq 60)$.

For $x = 40$, $z_1 = (40 - 50) / 10 = -1$

For $x = 60$, $z_2 = (60 - 50) / 10 = 1$

$P(z \leq -1) = 0.8413$

$P(z \leq 1) = 0.1587$

$P(40 \leq X \leq 60) = 0.8413 - 0.1587 = 0.6826$

- Can you describe what is the central limit theorem (CLT)?

  When you take a sufficiently large number of random samples from any population, the distribution of the sample means will approach a normal distribution, regardless of the original population's distribution.

- Can you explain how CLT is related to the normal distribution and sample mean?

  The Central Limit Theorem (CLT) states that, for any population distribution (whether it is normal, skewed, or uniform), the distribution of sample means will approximate a normal distribution as the sample size becomes sufficiently large. This means that regardless of the original data distribution, the mean of a large number of random samples will form a bell-shaped distribution. Moreover, the average of these sample means will be equal to the population mean, and the standard deviation of the sample means, called the standard error, decreases as the sample size increases.

Problems:

1. You can roll a 5-side dice up to 2 times. After the first roll, if you get a number, you can decide either to get dollars or to choose to continue rolling. But once you decide to continue, you forgot the number you just rolled. If you get to the second roll, you'll just get dollars if the second number is and the game stops.

Here are two examples: (i) If you get number 2 in the first roll and decide to stop, you will get 2 dollars. (ii) If you get number 2 in the first roll, decide to continue for the second roll, and get number 1 in the second roll, you will only get 1 dollar.

Then the questions are:

(1) What is your strategy for deciding whether to continue for the second roll?

(2) What is the game worth (expected profit if we follow the best strategy in (1))?

A. (1) Continue rolling if the first roll is 1, 2, or 3 (2) 3.6
B. (1) Continue rolling if the first roll is 4 (2) 4.75
C. (1) Continue rolling if the first roll is 1, 2, 3, or 4 (2) 4.4
D. (1) Continue rolling if the first roll is 1 or 2 (2) 4

The game expected profit for A – continue rolling if the first roll is 1, 2, 3 is

$P(4)*4+P(5)*5+P(1,2,3)*\sum_{i=1}^{5}\frac{1}{5}*i=\frac{1}{5}*4+\frac{1}{5}*5+\frac{3}{5}*3=3.6$

Similarly, the game expected profit for B – continue rolling if the first roll is 4 is

$\sum_{i=1,i\neq 4}^{5}P(i)*i+P(4)*\sum_{i=1}^{5}\frac{1}{5}*i=\frac{11}{5}+\frac{1}{5}*3=2.8$

The game expected profit for C – continue rolling if the first roll is 1, 2, 3, 4

$P(5)*5+P(1,2,3,4)*\sum_{i=1}^{5}\frac{1}{5}*i=1+\frac{4}{5}*3=3.4$

The game expected profit for D – continue rolling if the first roll is 1 or 2

$\sum_{i=3}^{5}\frac{1}{5}*i+P(1,2)*\sum_{i=1}^{5}\frac{1}{5}*i=\frac{12}{5}+\frac{2}{5}*3=3.6$

Based on calculating the expected profit, strategies A and D are the best ones.

2. Pat and Nat are dating, and all of their dates are scheduled to start at 9 p . m. Nat always arrives promptly at 9 p.m. Pat is highly disorganized and arrives at

a time that is uniformly distributed between 8 p.m. and 10 p.m. Let X be the time in hours between 8 p . m . and the time when Pat arrives. If Pat arrives before 9 p.m., their date will last exactly 3 hours. If Pat arrives after 9 p.m . , their date will last for a time that is uniformly distributed between 0 and 3 -X hours. The date starts at the time they meet. Nat gets irritated when Pat is late and will end the relationship after the second date on which Pat is late by more than 45 minutes. All dates are independent of any other dates.

(a) What is the expected number of hours Nat waits for Pat to arrive?

W = amount of time that nat waits for pat (in hours)

$$E[W] = \int_1^2 [(x - 1) \times \frac{1}{2} d]x = \frac{1}{2}\int_1^2 (x - 1) = \frac{1}{2}\left[\frac{(x-1)^2}{2}\right]_1^2$$

$$\frac{1}{2}\left[\frac{(x-1)^2}{2}\right]_1^2 = \frac{1}{2}\left[\frac{(2-1)^2}{2} - \frac{(1-1)^2}{2}\right] = \frac{1}{2}\left[\frac{1}{2}\right] = \frac{1}{4}$$

$$E[W] = \frac{1}{4}\ hours\ or\ 15\ minutes$$

(b) What is the expected duration of any particular date?

D = duration on any given date

$$E[D] = \int_0^1 \left[3 \times \frac{1}{2}\right] dx + \int_1^2 \left[\frac{3-(x-1)}{2} \times \frac{1}{2}\right] dx$$

$$E[D] = \frac{3}{2}\int_0^1 dx + \frac{1}{4}\int_1^2 (3 - x + 1)dx = \frac{3}{2} + \frac{1}{4}\int_1^2 (4x - 1)dx$$

$$= \frac{3}{2} + \frac{1}{4}\left[4x - \frac{x^2}{2}\right]_1^2 = \frac{3}{2} + \frac{1}{4}\left[(8 - 2) - (4 - \frac{1}{2})\right]$$

$$= \frac{3}{2} + \frac{1}{4}[2.5] = \frac{3}{2} + \frac{5}{8} = \frac{17}{8} = 2.125$$

The expected duration on any given day is 2.2125 hours or 127.5 minutes.

(c) What is the expected number of dates they will have before breaking up?

Since they would break up if Pat is more than 45 minutes late for 2 dates. We need to first calculate the probability he is more than 45 minutes late.

$$P(x > 1.75) = \frac{2-1.75}{2} = \frac{0.25}{2} = 0.125$$

Now we need to calculate the expected number of dates they go on.

$$E[number\ of\ dates] = \frac{2}{0.125} = 16$$

So, the expected number of dates they go on before they break up is 16.

Design: Design a problem that tests students' ability of calculating probability related to a normally distributed RV.

The heights of basketball players in a certain league are normally distributed with a mean of 78 inches (6 feet 6 inches) and a standard deviation of 3 inches.

1. What percentage of players in the league are shorter than 75 inches?
2. What is the probability that a randomly selected player is between 76 and 81 inches tall?
3. If the top 10% of the tallest players are considered for an elite training program, what is the minimum height required to qualify?