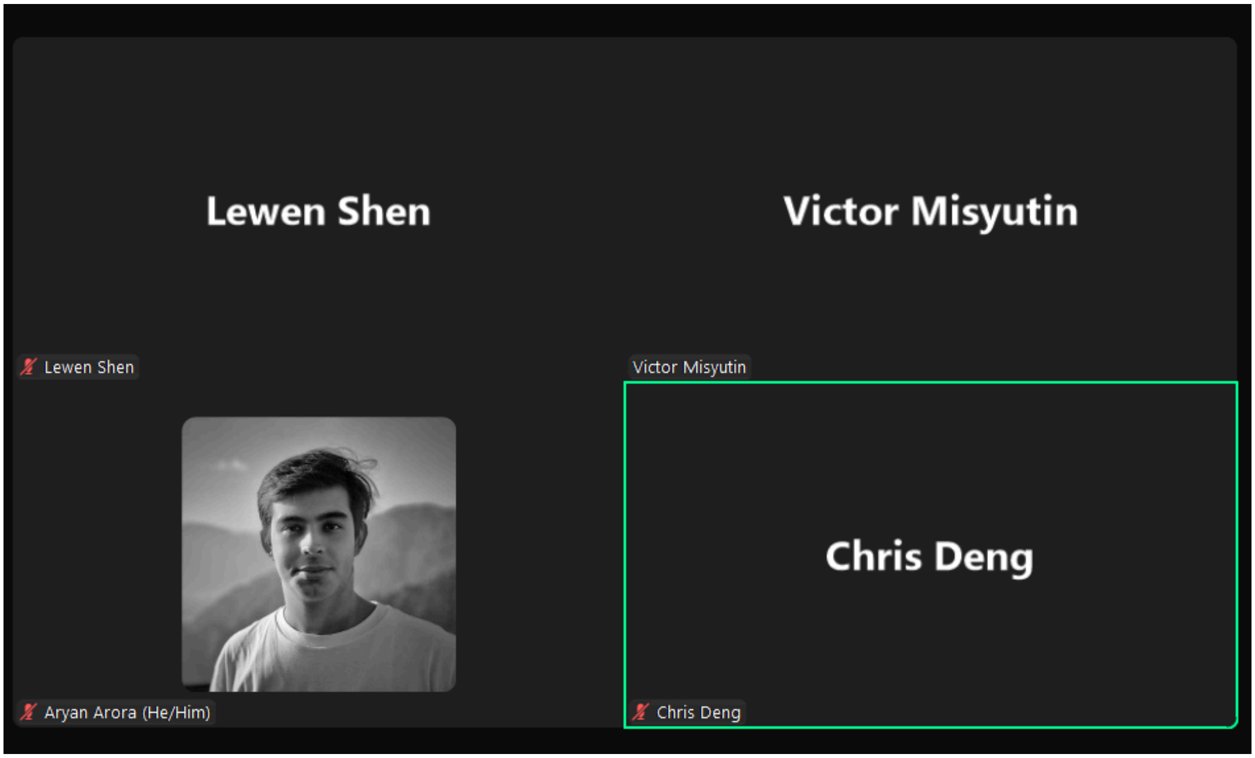
October 2nd



Questions to answer:

- Can you explain how WLLN can be derived?

The weak law of large numbers can be derived using random variables.

Let $\{X_1, X_2, ..., X_n\} = X$.

The Sample mean is

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

The variance of the sample mean is

$$Var(\{\overline{X}\}) = \frac{Var(\{X\})}{n}$$

Using Chebyshevs inequality we get:

$$P(\left|\overline{X}_n - mean(X)\right| \geq \epsilon) \leq \frac{Var(\{X\})}{n \times \epsilon^2}$$

As n goes to infinity $\frac{Var(\{X\})}{n \times \epsilon^2} = 0$

Therefore:

$$\lim_{n \to \infty} \left[P(\left|\overline{X}_n - mean(X)\right| \geq \epsilon)\right] = 0$$

- Can you describe what is IID samples? What is a sample mean?

IID samples are both independent and identically distributed. Meaning that knowing the outcome of one does not influence the outcome of another and each random variable has the same parameters.

- Can you describe the inequalities that we used to prove WLLN?

The inequality we used to prove the weak law of large numbers is the Chebyshev inequality. Which is:

$$P(|X - \mu| \geq \epsilon) \leq \frac{Var(X)}{\epsilon^2}$$

This inequality is used to describe the upper bound that a random variable deviates from its expected value.

- Can you explain why we say we can estimate probability using simulation because of WLLN? How is probability related to expected value?

We say this because the more times that you run a simulation, the closer you will get to the true probability of that event. For example, if you simulate a very large number of coin tosses, you will eventually find that the number of times it landed on heads and the number of times it landed on tails are equivalent. Which is the value of their true probability.

- Can you give an example of an indicator function?

Indicator functions are used to convert a large set of numbers or outcomes into a binary such as true and false or 0 and 1. They are often denoted using a piecewise function. Indicator functions are very common in many fields of mathematics, one example is electrical current. Let's say we have a circuit that requires 60 volts to be considered on. We can map turn this into an indicator function like so:

$$F(V_i) = \left\{ \begin{matrix} 1 & if\ V_i \geq 60 \\ 0 & if\ V_i \leq 60 \end{matrix} \right\}$$

Where 1 is the circuit being on and 0 is off.

Problems:

- We know that WLLN also applies when the random variable is continuous. Suppose we have a floor made of parallel strips of wood, each the same width, and we drop a needle onto the floor. We assume that the distance between the parallel strip is 2 and the needle's length 1. What is the probability that the needle will lie across a line between two strips? For this problem, we want to get the theoretical result and the simulating result of this problem.
    - For the theoretical result, let x be the distance from the center of the needle to the closest parallel line and let
    - θ
    - be the acute angle between the needle and one of the parallel lines. Try to use the integral of these two random variables to determine the probability that the needle will lie across a line between two strips.
    - For the simulating result, try to do the experiment by code and figure out what is the probability that the needle will lie across a line between two strips if you do the experiment 100 times in which you drop 100, 1000, 100000 needles. And compare the variances of results among the 100 experiments using different number of needles, and compare with the theoretical probability.

Let x be the distance from the center of the needle to the strip randomly, L be 1 and d be two. The prerequisite for crossing the line between two strips is: $P=2L/d \cdot \pi/1=1/\pi$, and this is the answer theoretically.

For a simulation,

```
def simulate_needles(num_needles):
    cross_count = 0
    for _ in range(num_needles):
        x = np.random.uniform(0, 1)
        theta = np.random.uniform(0, np.pi/2)
        if x <= (0.5 * np.sin(theta)):
            cross_count += 1
    return cross_count / num_needles
needle_counts = [100, 1000, 10000, 100000]
simulated_probs = [simulate_needles(n) for n in needle_counts]
theoretical_prob = 1 / np.pi
print("Theoretical probability: {:.4f}".format(theoretical_prob))
for n, p in zip(needle_counts, simulated_probs):
    print("Simulated probability for {} needles: {:.4f}".format(n, p))
```

```
Theoretical probability: 0.3183
Simulated probability for 100 needles: 0.2600
Simulated probability for 1000 needles: 0.3260
Simulated probability for 10000 needles: 0.3130
Simulated probability for 100000 needles: 0.3177
```

```
Needles: 100, Simulated Mean: 0.32390, Variance: 0.00239
Needles: 1000, Simulated Mean: 0.31986, Variance: 0.00017
Needles: 10000, Simulated Mean: 0.31799, Variance: 0.00002
Needles: 100000, Simulated Mean: 0.31817, Variance: 0.00000
```

This is another time of simulation covering the variance. Both of them are showing that, when the number of simulations increases, the mean is approaching the theoretical result more, which is $1/\pi$, and the variance is decreasing. When the simulation is large enough, 100000, the variance is almost 0.

Design: Design a problem that tests students' ability of using simulation to calculate probability.

Suppose you are playing a game where you roll two fair six-sided dice. Your goal is to calculate the probability that the sum of the numbers on the two dice is **7** or **11**. Since calculating exact probabilities might be challenging, you decide to use simulation to estimate the probability.

1. Simulate rolling two six-sided dice for 100, 1000, and 10000 times. Keep track of how many times the sum of the two dice is 7 or 11.
2. Use the results from your simulation to estimate the probability of rolling a sum of 7 or 11. Compare your simulated probability with the theoretical probability of the same event.
3. Now suppose you are rolling three six-sided dice instead of two. What is the probability that the sum of the numbers on all three dice is exactly 10? Go over the simulation process above and compare it with the theoretical probability.