Simulation of

CLT under Sampling Bias

Last time, we simulated CLT on a uniform distribution and showed how sample means tend to normality as sample size increases.

But…

What happens when sampling bias is introduced?

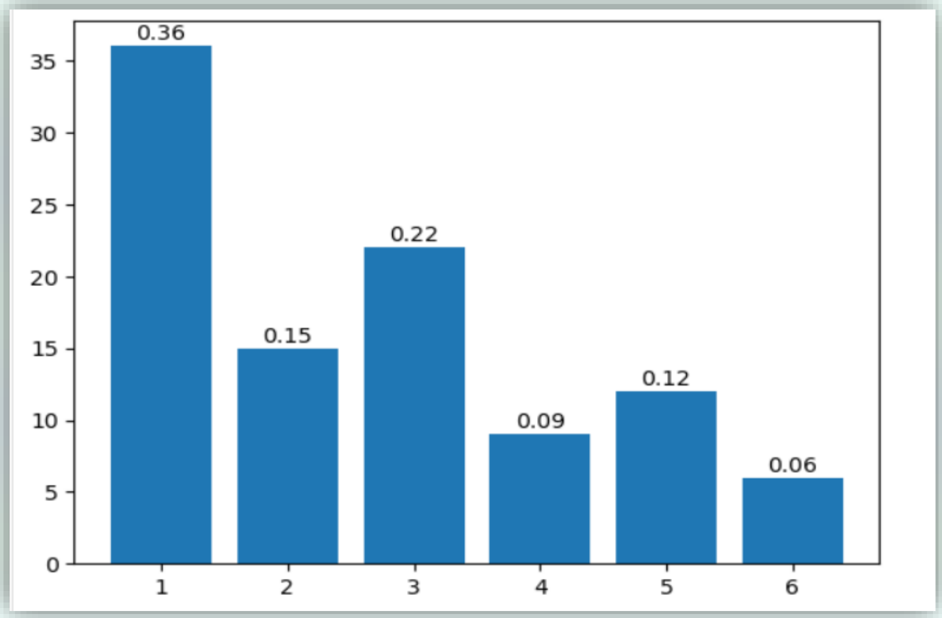
CLT states that regardless of the parent distribution, the distribution of sample means will be approximately normal for large samples.

But what if the samples themselves are biased? Will the CLT still hold?

- by Ruturaj T. Saravane

Experiment Setup

Rolling an **unfair** six-sided die (1-6) multiple times. The probability function for this biased die is given as,



Here, probability of getting 1 on die is higher than others and following the same logic for the remaining outcomes.

For better understanding, we can say that if we roll this biased die 100 times, there is a higher chance of getting "1" as the outcome approximately 36 times, and so on…

Weighted finite population -

For this experiment I created finite empirical population based on the given PMF from which we are drawing random samples.

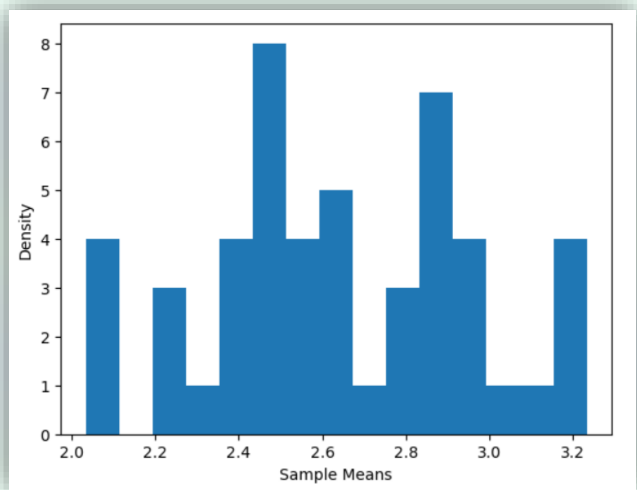
Specifically, I created a dataset with:

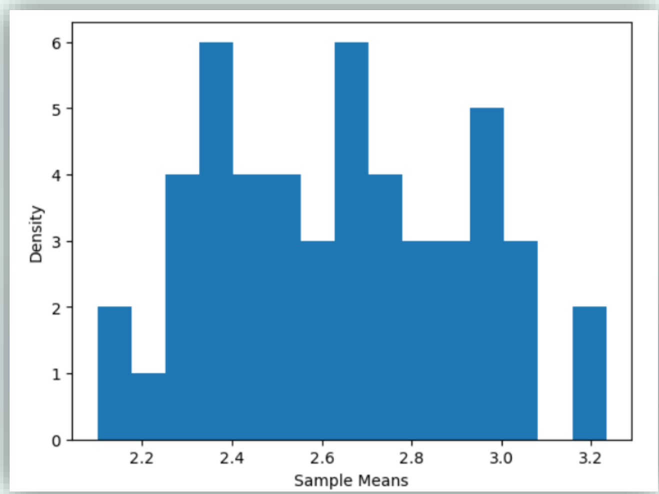
- 36 occurrences of "1" (since P(1) = 0.36)
- 15 occurrences of "2" (P(2) = 0.15)
- 22 occurrences of "3" (P(3) = 0.22)
- 9 occurrences of "4" (P(4) = 0.09)
- 12 occurrences of "5" (P(5) = 0.12)
- 6 occurrences of "6" (P(6) = 0.06)

Sampling Strategy:

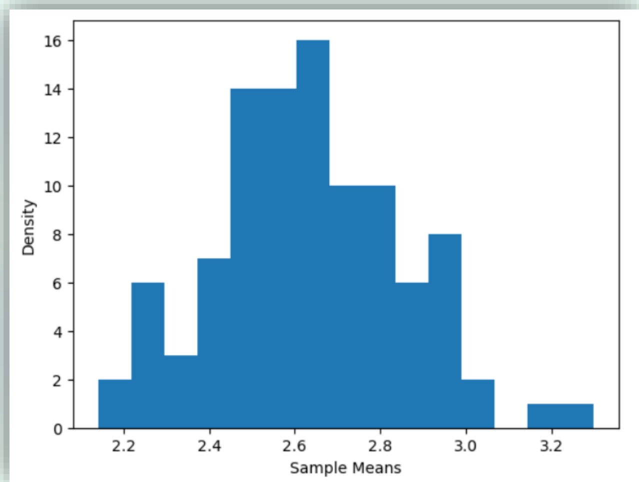
- The finite population serves as my sampling frame, from which I randomly draw multiple samples.
- Starting with 50 samples of 30 rolls each, then progressively increase both the number of samples and sample size.
- Compute the sample means and analyse their distribution to observe the effects of the Central Limit Theorem.

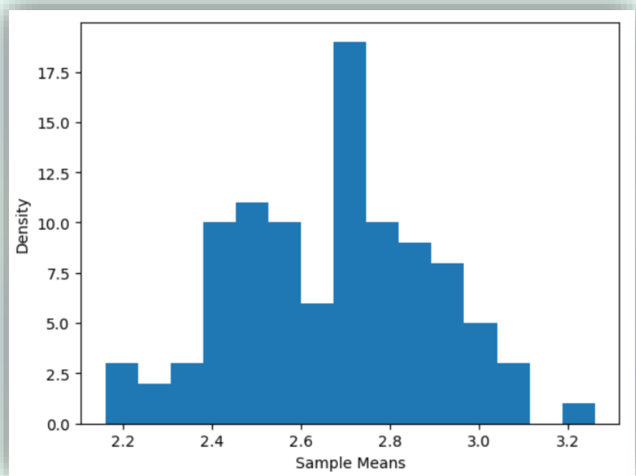
Sample size = 30, Number of samples = 50





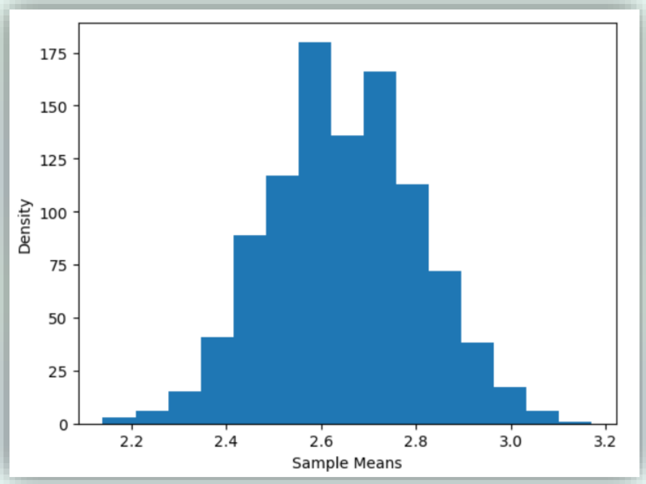
Sample size = 50, Number of samples = 100

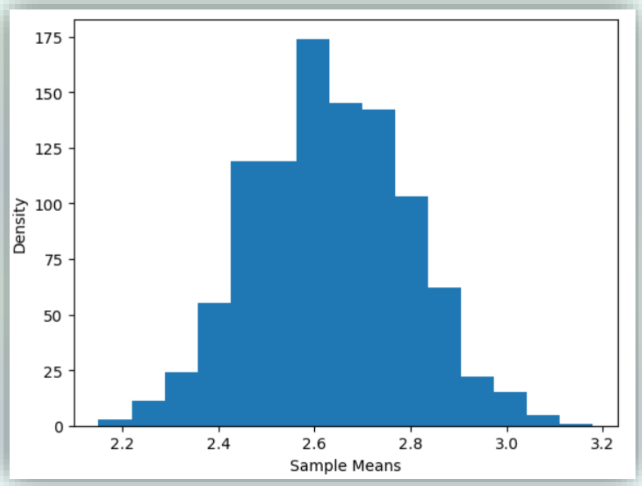




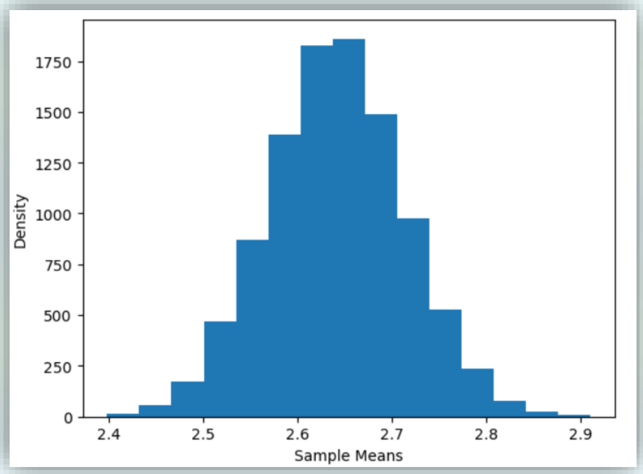
- by Ruturaj T. Saravane

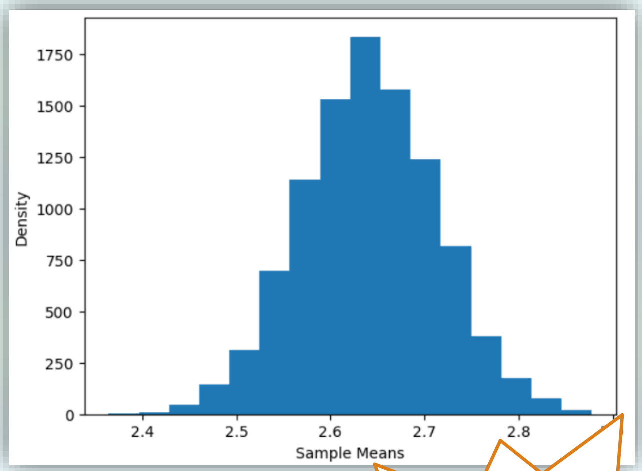
Sample size = 100, Number of samples = 1000

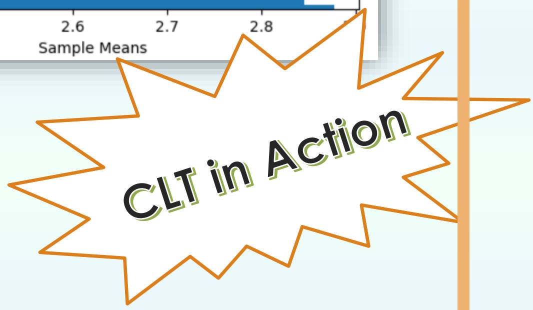




Sample size = 500, Number of samples = 10000







- by Ruturaj T. Saravane

Conclusion

- ✓ Despite the strong bias in individual die rolls, our experiment confirms that as sample size or number of samples increases, the distribution of sample means becomes approximately normal, demonstrating the Central Limit Theorem in action

- ✓ Regardless of the population distribution, as long as we take sufficiently large and independent samples, the sampling distribution of the mean will approximate normality.

But wait—how does this experiment differ from our previous one? 🤔

Next, we'll explore **how bias affects the expected value –**

- Does it cause a shift in sample means from the true population mean?
- If so, by how much?

stay connected…

- by Ruturaj T. Saravane