Lab Assignment =7
Topic: Sampling
Sub topic: Parametric Estimation, CLT and WLL

Q1) Suppose that the annual incomes of a population of 10,000 people follow a normal distribution with a mean of $60,000 and a standard deviation of $10,000.

1. What is the probability that a randomly selected person from this population has an income between $50,000 and $70,000 per year?
2. What is the probability that the average income of a random sample of 100 people from this population is between $55,000 and $65,000 per year?
3. Create a histogram of the income distribution of the population, and add vertical lines to indicate the mean, standard deviation, and the income range in part 1.
4. Create a box plot of the income distribution of the population, and add a horizontal line to indicate the income range in part 1. [Read about box plots and how do you infer different parameters of the distribution from
5 .Create a histogram of the sampling distribution of the sample mean for sample sizes of 10, 50, and 100, and add vertical lines to indicate the mean and standard deviation of the sampling distribution. How do you infer your observation with regards to the Central Limit Theorem? [Use: plt.avxline to draw the lines]

Q2
Suppose we have a sequence of independent and identically distributed random variables X1, X2, X3, ..., Xn, where each Xi has a gamma distribution with shape parameter $\alpha=2$ and scale parameter $\beta=1/2$. We want to investigate the weak law of large numbers for this distribution using simulation.

a) Write a Python code to generate n independent and identically distributed samples from this distribution, and compute the sample mean Sn for each n.
b) Plot the sample means for various values of n from 10-10000.
c) Plot variance and standard error for the various values of n from 10-10000.

[Hint: use np.random.gamma to generate the samples/ or the parameters of the samples, Use subplots]

Q3)
You are given a dataset of exam scores from a population of 75 students. The mean score of the population is unknown, but you are interested in estimating it using a random sample of 5,10,15,25 scores from the population. Use pandas to open the ‘marks.csv’ dataset.

1. Write a Python function that takes in the dataset and calculates the point estimate of the population mean using the sample mean of 100 different samples of given samples sizes.

[Hint: Calculate the means of 100 samples of the given sizes and compute the overall mean of the 100 means]

2. Find the Population variance and standard deviation using the sample variance and standard deviation.
3. Plot sampling distribution of sample means for sample sizes(n=5,10,15, 25). Comment on what the sampling distribution looks like?

[incase of queries, Contact: Om Kshatriya- 8237788597]