Central Limit Theorem

The central limit theorem tells us that the sample mean will be approximately normally distributed for large sample sizes, regardless of the distribution from which we are sampling.

The central limit theorem states (given that sample size >= 30):-

- The sampling distribution of the “sample mean” has an approximately normal distribution.
- The mean of the sampling distribution is equal to the population mean.
- The standard deviation of the sampling distribution equals the standard deviation in the population divided by the square root of the sample size (i.e. standard error).

Let’s illustrate this in Python.

Python implementation of the Central Limit Theorem

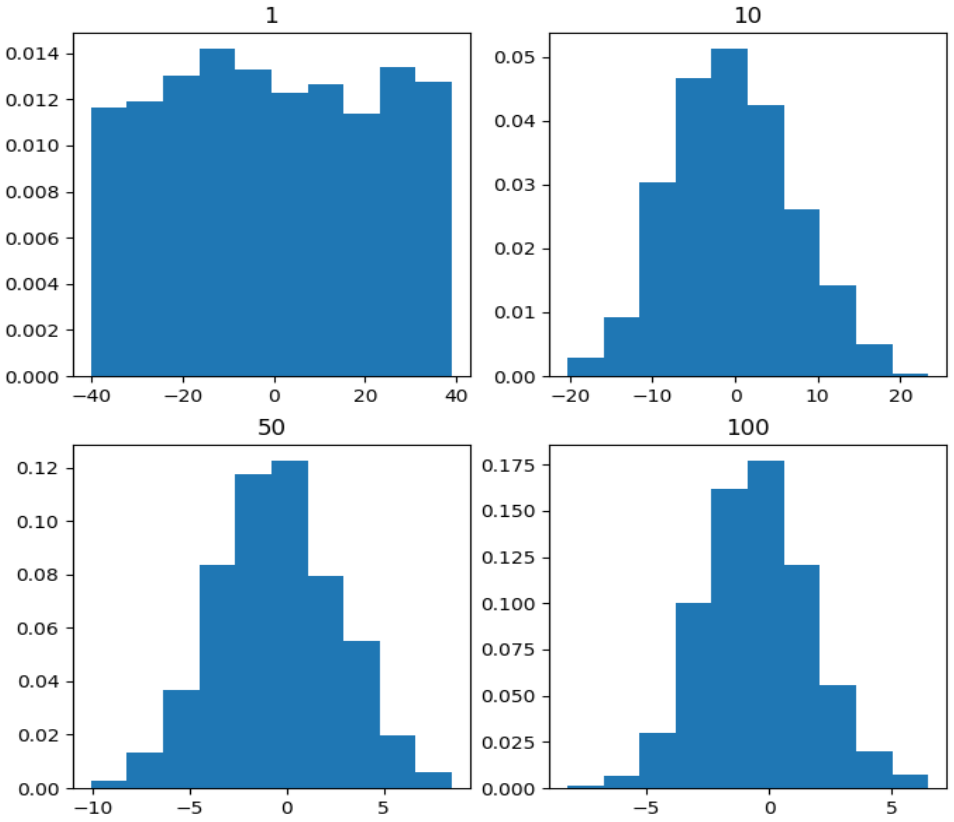
```
import numpy
import matplotlib.pyplot as plt

# number of sample
num = [1, 10, 50, 100]
# list of sample means
means = []

# Generating 1, 10, 30, 100 random numbers from -40 to 40
# taking their mean and appending it to list means.
for j in num:
    # Generating seed so that we can get same result
    # every time the loop is run...
    numpy.random.seed(1)
    x = [numpy.mean(
        numpy.random.randint(
            -40, 40, j)) for _i in range(1000)]
    means.append(x)
k = 0

# plotting all the means in one figure
fig, ax = plt.subplots(2, 2, figsize=(8, 8))
for i in range(0, 2):
    for j in range(0, 2):
        # Histogram for each x stored in means
        ax[i, j].hist(means[k], 10, density=True)
```

```
        ax[i, j].set_title(label=num[k])
        k = k + 1
plt.show()
```



Points to note -

- Central Limit Theorem holds irrespective of the type of distribution of the population.
- Now, we have a way to estimate the population mean by just making repeated observations of samples of a fixed size.
- Greater the sample size, lower the standard error, and greater accuracy in determining the population mean from the sample mean.

Significance of Central Limit Theorem:-

- Analyzing data involves statistical methods like hypothesis testing and constructing confidence intervals. These methods assume that the population is normally distributed. In the case of unknown or non-normal distributions, we treat the sampling distribution as normal according to the central limit theorem.

- If we increase the samples drawn from the population, the standard deviation of sample means will decrease. This helps us estimate the population mean much more accurately.