Central Limit Theorem

Submitted to Dr. Anil Kumar Sao
Assignment for DSL201
Prepared by Amay Dixit - 12340220

INTRODUCTION

The Central Limit Theorem (CLT) is a fundamental concept in statistics that states that the distribution of sample means approaches a normal distribution as the sample size increases, regardless of the original distribution's shape. This report demonstrates the CLT using three different types of distributions: Uniform, Normal, and Exponential. The sample means are calculated for various sample sizes (N = 2, 10, 30, 100) to illustrate the effect of increasing sample size on the shape of the distribution of means.

METHODOLOGY

1. Distribution Generation

Three types of distributions were generated with a sample size of 10,000:

- **Uniform Distribution**: Values were uniformly distributed between 0 and 100.
- **Normal Distribution**: Values were normally distributed with a mean of 50 and a standard deviation of 15.
- **Exponential Distribution**: Values followed an exponential distribution with a scale parameter of 50.

2. Sampling and Mean Calculation

For each distribution, the following steps were performed:

- **Sampling**: Random samples of size N (where N = 2, 10, 30, 100) were drawn 1,000 times.
- **Mean Calculation**: The mean of each sample was calculated.

The above steps were repeated for each distribution type and each sample size to generate a distribution of sample means.

3. Plotting

The distribution of sample means for each distribution type and sample size was plotted using histograms with KDE (Kernel Density Estimation) overlays. The x-axis represents the sample means, and the y-axis represents the frequency of occurrence.

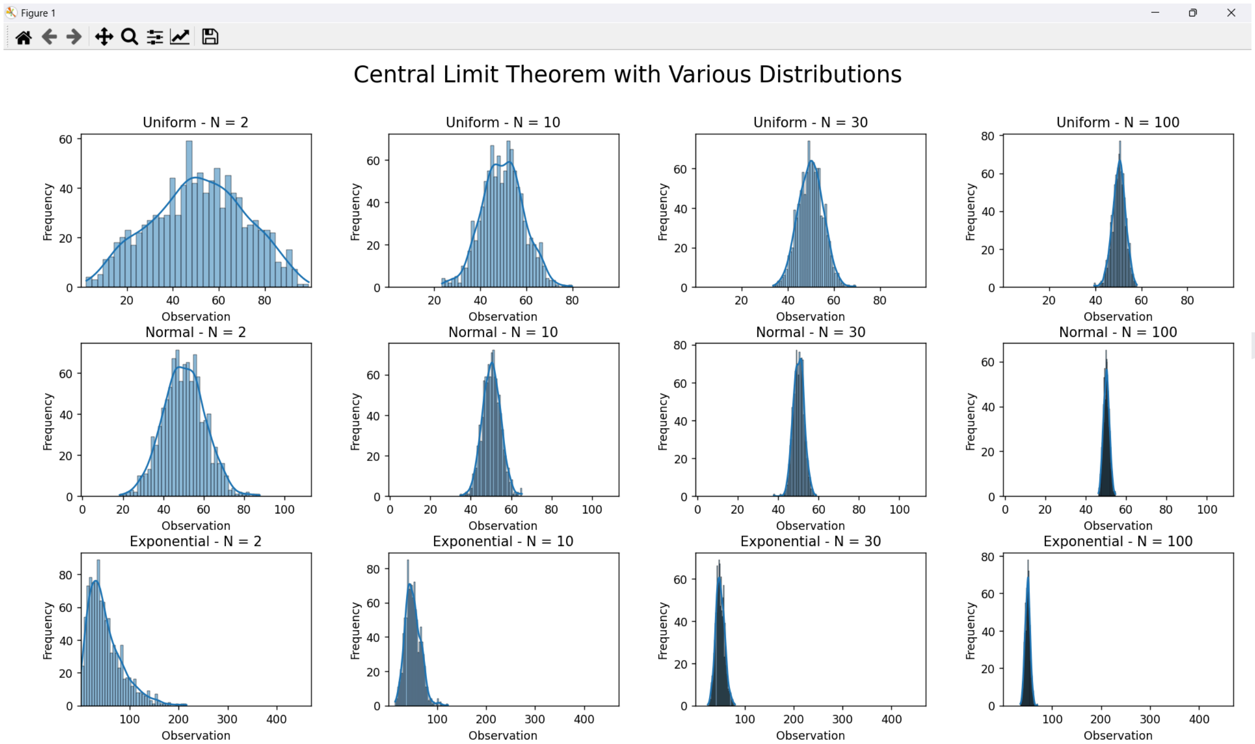
SOURCE CODE

```
import random

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns


class CentralLimitTheorem:

    def __init__(self, distribution):

        self.distribution = distribution

        self.dist_min = min(distribution)

        self.dist_max = max(distribution)


    def _sample(self, N):

        sample = random.choices(self.distribution, k=N)

        return np.mean(sample)


    def run_sample(self, N, num_samples=1000):

        return [self._sample(N) for _ in range(num_samples)]
```

```python
def generate_distribution(distribution_type, size):

    if distribution_type == 'uniform':

        return list(np.random.uniform(0, 100, size))

    elif distribution_type == 'normal':

        return list(np.random.normal(50, 15, size))

    elif distribution_type == 'exponential':

        return list(np.random.exponential(50, size))




def plot_distribution(distribution, title=None, bin_min=None, bin_max=None,
num_bins=None, ax=None):

    sns.histplot(distribution, bins=num_bins, kde=True, ax=ax)



    if title:

        ax.set_title(title)

    ax.set_xlim(bin_min, bin_max)

    ax.set_xlabel("Observation")

    ax.set_ylabel("Frequency")



def main():

    fig, axes = plt.subplots(3, 4, figsize=(10, 10))

    fig.suptitle("Central Limit Theorem with Various Distributions", fontsize=20)



    distribution_types = ['uniform', 'normal', 'exponential']
```

```
    n_vals = [2, 10, 30, 100]

    for i, dist_type in enumerate(distribution_types):
        sample_distribution = generate_distribution(dist_type, 10000)
        clt = CentralLimitTheorem(sample_distribution)

        for j, N in enumerate(n_vals):
            means = clt.run_sample(N=N)
            plot_distribution(means, f"{dist_type.capitalize()} - N = {N}",
                              bin_min=min(sample_distribution),
bin_max=max(sample_distribution),
                              num_bins=40, ax=axes[i, j])

    plt.tight_layout(rect=[0, 0, 1, 0.96])
    plt.show()

if __name__ == "__main__":
    main()
```

OBSERVATIONS



RESULTS

The results clearly demonstrate the Central Limit Theorem. As the sample size increases, the distribution of sample means approaches a normal distribution, regardless of the original distribution type:

- **Uniform Distribution**: Even with N = 2, the distribution of means begins to take a symmetric shape, and as N increases, it becomes more bell-shaped.
- **Normal Distribution**: The distribution of means quickly becomes more concentrated around the population mean as N increases, maintaining a normal shape.
- **Exponential Distribution**: Initially, the distribution of means is skewed; however, as N increases, it becomes more symmetric and bell-shaped.

These results confirm the power of the Central Limit Theorem in explaining the behavior of sample means.

CONCLUSION

The Central Limit Theorem is a robust concept that holds true for different types of distributions. As sample size increases, the distribution of sample means converges to a normal distribution, highlighting the significance of the CLT in statistical analysis.