Thesis 2

Anna Lucia Lamacchia
Statistics

January 11, 2024

The Central Limit Theorem (CLT)

1 Meaning

The Central Limit Theorem (CLT) stands as a fundamental principle in statistical theory, offering profound insights into the behavior of sample means. At its core, the CLT asserts that, regardless of the original distribution of a set of random variables, the distribution of their sample means will tend to approach a normal (Gaussian) distribution as the sample size increases. This means that, under certain conditions, the average of a sufficiently large number of independent and identically distributed random variables, regardless of their underlying distribution, becomes increasingly well-described by a bell-shaped curve.
The beauty of the CLT lies in its universality, making it a powerful tool in statistical inference. It provides a mathematical explanation for the prevalence of normal distributions in various real-world scenarios. This theorem becomes particularly crucial in situations where the underlying distribution of the population is unknown or complex. The CLT allows researchers and practitioners to rely on the properties of the normal distribution when making statistical inferences, facilitating hypothesis testing, confidence intervals, and other analytical techniques.
Moreover, the CLT serves as a bridge between the microscale behavior of individual observations and the macroscopic properties of sample means, offering a statistical law of large numbers at the level of distributions. The implications of the CLT extend beyond the realm of academia, finding applications in fields such as quality control, finance, and public health. Its meaning lies not only in the convergence to normality but also in the assurance it provides

when dealing with diverse and unpredictable data, allowing for more robust and generalizable statistical analyses.

2 Proof

Central Limit Theorem. Let $X_1, X_2, \ldots, X_n$ be a sequence of independent and identically distributed random variables with mean μ and finite variance σ^2. Let $\bar{X}_n$ represent the sample mean of these random variables. Then, as n approaches infinity, the distribution of $\sqrt{n}(\bar{X}_n - \mu)$ converges in distribution to the standard normal distribution $\mathcal{N}(0, 1)$.

Proof.

2.1 Step 1: Moment Generating Function (MGF) and Characteristic Function (CF)

The characteristic function (CF) of X_i is given by:

$$\phi_{X_i}(t) = E[e^{itX_i}] = e^{i\mu t - \frac{1}{2}\sigma^2 t^2}$$

2.2 Step 2: Characteristic Function of Sample Mean

The CF of the sample mean $\bar{X}_n$ is given by:

$$\phi_{\bar{X}_n}(t) = \prod_{i=1}^{n} \phi_{X_i}\left(\frac{t}{n}\right)$$

Substituting the CF of X_i:

$$\phi_{\bar{X}_n}(t) = \left(e^{i\frac{\mu t}{n} - \frac{\sigma^2 t^2}{2n^2}}\right)^n$$

2.3 Step 3: Convergence of Characteristic Function

As n approaches infinity, the CF converges pointwise to the CF of the standard normal distribution:

$$\lim_{n\to\infty} \phi_{\bar{X}_n}(t) = \lim_{n\to\infty} \left(e^{i\frac{\mu t}{n} - \frac{\sigma^2 t^2}{2n^2}}\right)^n = e^{-\frac{1}{2}t^2}$$

2.4 Step 4: Continuous Mapping Theorem

Applying the Continuous Mapping Theorem, we conclude that as n approaches infinity, the distribution of $\sqrt{n}(\bar{X}_n - \mu)$ converges to the standard normal distribution. $\square$

3 Simulation in Python

This Python code simulates the Central Limit Theorem by generating random samples from a normal distribution and then plotting the distribution of sample means. As per the CLT, the resulting distribution of sample means converges to a normal distribution, even if the original population distribution is non-normal.

```python
import numpy as np
import matplotlib.pyplot as plt

# Parameters
population_mean = 50
population_std_dev = 10
sample_size = 30
num_simulations = 1000

# Simulate CLT
sample_means = []
for _ in range(num_simulations):
    # Generate random samples from a normal distribution
    samples = np.random.normal(population_mean, population_std_dev, sample_size)
    # Calculate the sample mean
    sample_mean = np.mean(samples)
    sample_means.append(sample_mean)

# Plot histogram of sample means
plt.hist(sample_means, bins=30, density=True, alpha=0.5, color='blue')

# Plot theoretical normal distribution
xmin, xmax = plt.xlim()
x = np.linspace(xmin, xmax, 100)
p = (1 / (population_std_dev * np.sqrt(2 * np.pi / sample_size))) * np.exp(-(x - population_mean)**2 / (2 * population_std_dev**2 / sample_size))
plt.plot(x, p, 'k', linewidth=2)

# Set plot labels and title
plt.xlabel('Sample Mean')
plt.ylabel('Probability Density')
plt.title('Central Limit Theorem Simulation')
plt.show()
```

Figure 1: Code

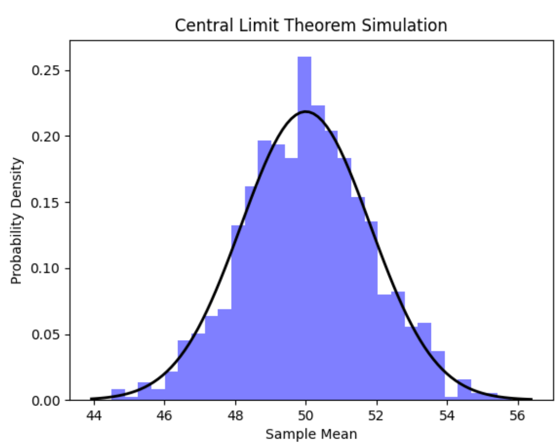

Figure 2: Chart