Thesis 1

Anna Lucia Lamacchia
Statistics

January 11, 2024

The Law of Large Numbers (LLN)

1 Meaning

The Law of Large Numbers (LLN) fundamentally encapsulates the statistical concept that as the size of a sample increases, the sample average converges towards the expected or true population mean. In simpler terms, LLN highlights the remarkable tendency for randomness to smooth out when dealing with larger sets of observations. This principle assures us that, given a sufficiently large sample size, the average value derived from that sample becomes a highly reliable approximation of the overall population average. In essence, LLN provides a mathematical grounding for our intuition that more data tends to yield more accurate and representative insights into the underlying characteristics of a population or phenomenon, forming a bedrock principle in statistical theory and practical applications. There are two main versions of the LLN:

1. **Weak Law of Large Numbers (WLLN)**: This version asserts that the sample mean converges in probability to the population mean. In other words, as the sample size grows larger, the probability that the sample mean deviates significantly from the population mean approaches zero.

2. **Strong Law of Large Numbers (SLLN)**: This stronger version states that with probability one, the sample mean converges almost surely to the population mean. It implies that, almost surely, the sample mean will approach the population mean as the sample size increases.

2 Proof of WLLN

Theorem Weak Law of Large Numbers. For a sequence of independent and identically distributed random variables $X_1, X_2, \ldots, X_n$ with expected value μ and finite variance σ^2, the sample mean $\bar{X}_n$ converges in probability to μ.

Proof.

2.1 Step 1: Definition and Assumptions

Let $S_n = X_1 + X_2 + \ldots + X_n$ denote the sum of the first n random variables. The sample mean is defined as $\bar{X}_n = \frac{S_n}{n}$.
Assumptions:

1. $E[X_i] = \mu$ for all i (identically distributed).
2. $Var[X_i] = \sigma^2$ for all i (finite variance).
3. X_i are independent.

2.2 Step 2: Expectation and Variance Calculation

By the linearity of expectations:

$$E[S_n] = n\mu$$

Therefore, the expectation of the sample mean is:

$$E[\bar{X}_n] = \frac{1}{n}E[S_n] = \mu$$

The variance of S_n is $Var[S_n] = n\sigma^2$, and the variance of the sample mean is:

$$Var[\bar{X}_n] = \frac{1}{n^2}Var[S_n] = \frac{\sigma^2}{n}$$

2.3 Step 3: Application of Chebyshev's Inequality

For any $\epsilon > 0$, by Chebyshev's Inequality:

$$P\left(|\bar{X}_n - \mu| \geq \epsilon\right) \leq \frac{Var[\bar{X}_n]}{\epsilon^2}$$

Substituting the variance and simplifying:

$$P\left(|\bar{X}_n - \mu| \geq \epsilon\right) \leq \frac{\sigma^2}{n\epsilon^2}$$

2.4 Step 4: Convergence in Probability

The result above implies that as n approaches infinity, the probability of the sample mean deviating from the population mean by more than ϵ approaches zero:

$$\lim_{n \to \infty} P\left(|\bar{X}_n - \mu| \geq \epsilon\right) = 0$$

Therefore, the Weak Law of Large Numbers holds, demonstrating that the sample mean converges in probability to the population mean as the sample size grows indefinitely. $\square$

3 Simulations in Python

3.1 First simulation

This script defines a function *lln_simulation* to generate random samples and calculate the running sample mean. The *plot_lln_simulation* function is used to display the results with a line chart, comparing the running sample mean with the true mean.

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

def lln_simulation(sample_size):
    # Generate random samples from a standard normal distribution
    samples = np.random.randn(sample_size)

    # Calculate the running sample mean
    running_mean = np.cumsum(samples) / np.arange(1, sample_size + 1)

    return running_mean

def plot_lln_simulation(sample_sizes, running_means):
    plt.figure(figsize=(10, 6))
    plt.plot(sample_sizes, running_means, label='Running Sample Mean', color='blue')
    plt.axhline(0, color='red', linestyle='--', label='True Mean (0)')
    plt.title('Law of Large Numbers (LLN) Simulation')
    plt.xlabel('Sample Size')
    plt.ylabel('Sample Mean')
    plt.legend()
    plt.show()

# Set the maximum sample size
max_sample_size = 1000

# Generate sample sizes from 1 to max_sample_size
sample_sizes = np.arange(1, max_sample_size + 1)

# Run the LLN simulation
running_means = lln_simulation(max_sample_size)

# Plot the results
plot_lln_simulation(sample_sizes, running_means)
```

Figure 1: Code

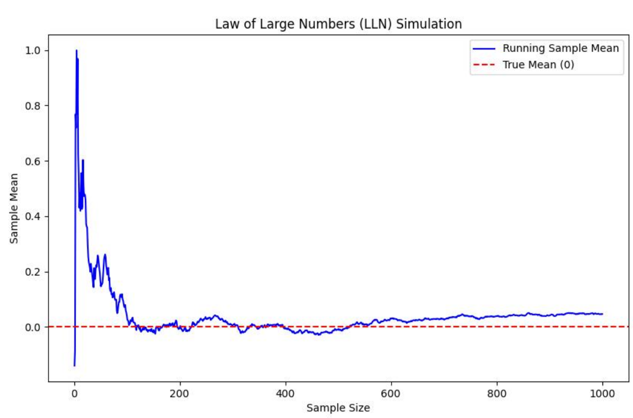

Figure 2: Chart

3.2 Second simulation

This Python code simulates the Law of Large Numbers by generating random samples from a normal distribution with a known population mean. As the sample size increases, the histogram of sample means should converge toward the population mean. The red dashed line represents the true population mean.

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

# Parameters
population_mean = 5
sample_sizes = [10, 50, 100, 500, 1000]
num_simulations = 1000

# Simulate LLN
for sample_size in sample_sizes:
    sample_means = []
    for _ in range(num_simulations):
        # Generate random samples from a normal distribution with mean population_mean
        samples = np.random.normal(population_mean, 1, 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, label=f'Sample Size = {sample_size}', alpha=0.5)

# Plot population mean as a vertical line
plt.axvline(population_mean, color='red', linestyle='dashed', linewidth=2, label='Population Mean')

# Set plot labels and title
plt.xlabel('Sample Mean')
plt.ylabel('Frequency')
plt.title('Law of Large Numbers Simulation')
plt.legend()
plt.show()
```

Figure 3: Code

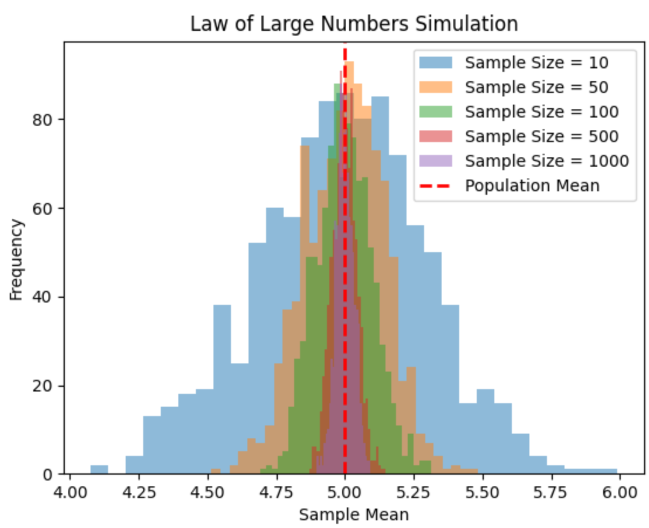

Figure 4: Chart