The law of large number

Pasciullo Marco

December 2023

1 INTRODUCTION

In probability theory, the law of large numbers (LLN) is a theorem that describes the result of performing the same experiment a large number of times. According to the law, the average of the results obtained from a large number of independent identical trials should be close to the expected value and tends to become closer to the expected value as more trials are performed. The LLN only applies to the average of the results obtained from repeated trials and claims that this average converges to the expected value; it does not claim that the sum of n results gets close to the expected value times n as n increases. There are two main versions of the law of large numbers. They are called the **weak** (WLLN) and **strong** (SLLN) laws of the large numbers. The difference between them is mostly theoretical. Before discussing the WLLN, let us define the sample mean

Definition. For i.i.d. random variables $X_1, X_2, ..., X_n$, the sample mean, denoted by X, is defined as

$$\overline{X} = \frac{X_1 + X_2 + ... + X_n}{n}$$

Another common notation for the sample mean is M_n. If the X_i's have CDF FX(x), we might show the sample mean by $M_n(X)$ to indicate the distribution of the Xi's.

Note that since the X_i's are random variables, the sample mean, $\overline{X} = M_n(X)$, is also a random variable. In particular, we have:

$$E[\overline{X}] = \frac{EX_1 + EX_2 + ... + EX_n}{n} = \frac{nEX}{n} = EX$$

Also, the variance of $\overline{X}$ is given by

$$Var(\overline{X}) = \frac{Var(X_1 + X_2 + ... + X_n)}{n^2}$$

$$= \frac{Var(X_1) + Var(X_2) + ... + Var(X_n)}{n^2}$$

Using the fact that the variance of the sum of independent variables X_i is the sum of their variances:

$$= \frac{n * Var(X)}{n^2}$$
$$= \frac{Var(X)}{n}$$

This simplification is based on the fact that the variance of a constant multiplied by a random variable is the square of the constant multiplied by the variance of the random variable.

now before going deeper into the two different branches of LLN (WLLN and SLLN), let's define the historical developments of the LLN. [1].

[1] Wikipedia LLN

2 HISTORICAL PROGESSION OF LAW LARGE NUMBER

- **Bernoulli and the Stochastic Limit (1713):** Swiss mathematician Jakob Bernoulli was a pioneer in formulating the law of large numbers in the 18th century. In his treatise Ars Conjectandi in 1713, Bernoulli introduced the concept of the stochastic limit, stating that as the number of trials in a random experiment increases, the relative frequency of events converges to the expected value.

- **Laplace and the Central Limit Theorem (1810):** Pierre-Simon Laplace further contributed to the development of the law of large numbers with his work on the central limit theorem. Laplace demonstrated that, regardless of the initial distribution, the sum of a large number of independently and identically distributed random variables converges to a normal distribution.

- **Chebyshev and the Inequality (1867):** Russian mathematician Pafnuty Chebyshev contributed to the law of large numbers with the formulation of Chebyshev's inequality in 1867. This inequality connects the standard deviation and the probability of significant deviations in a distribution.

- **Borel and the Concept of Convergence (1909):** Emile Borel introduced the concept of almost sure convergence in the law of large numbers in 1909. He demonstrated that the relative frequency of an event, in a large number of independent trials, almost surely converges to the expected value.

- **Kolmogorov and Mathematical Formalization (1933):** Andrey Kolmogorov provided a rigorous mathematical formalization of the law of large numbers in his foundational works on probability in 1933. Kolmogorov defined the concepts of convergence in probability and almost sure convergence, establishing a solid foundation for modern probability theory.

These contributions have shaped the theory of large numbers over centuries, making it an essential foundation of statistics and probability. Its evolution is characterized by a series of mathematical developments that have rendered the law of large numbers a fundamental tool for the analysis and interpretation of experimental data. Now we let's see in detail WLLN and SLLN.

3 WEAK LAW OF THE LARGE NUMBER

The Weak Law of Large Numbers, also known as Bernoulli's theorem, states that if you have a sample of independent and identically distributed random variables, as the sample size grows larger, the sample mean will tend toward the population mean.

To put this in formal mathematical notation, it looks like this:

$$\lim_{n \to \propto} P(|\overline{X}n - \mu| \geq \epsilon) = 0$$

As the sample size n grows to infinity, the probability that the sample mean x-bar differs from the population mean μ by some small amount epsilon is equal to 0.

We can prove this using Chebyshev's inequality, which says the probability that a random variable **X** differs from its mean by some small constant k is less than or equal to the variance of X divided by the the square of the constant k.

$$P(|X - \mu| \geq k) \leq \frac{var(X)}{k^2}$$

Since the random variable X and the constant k can be anything, we can replace X with the sample mean and replace k with epsilon.

$$P(|\overline{X}_n - \mu| \geq \epsilon) \leq \frac{var(\overline{X}_n)}{\epsilon^2}$$

Because we assumed that our sample contains independent and identically distributed random variables, we can simplify the right side of the equation.

$$\frac{var(\overline{X}_n)}{\epsilon^2} = \frac{\frac{1}{n^2}\sum_{i=1}^{n} var(x_i)}{\epsilon^2} = \frac{\frac{n\sigma^2}{n^2}}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$$

Replacing the right side of Chebyshev's inequality, we have the following.

$$P(|\overline{X}n - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}$$

As n tends to infinity, follows that the right side of the inequality equals 0. [2]. [3].

[2] Statelec.com
[3] medium.com

4 STRONG LAW LARGE NUMBER

The strong law of large numbers (SLLN) states that if $X_1,\ X_2,\ .\ .\ .\ ,\ X_n$ are i.i.d. random variables with a finite expected value $\mathrm{E}(X_i) = \mu < \infty$, then almost surely.

$$\lim_{n\to\infty} \frac{X_1, X_2, ..., X_n}{n} = \mu$$

We give a relatively simple proof of the strong law under the assumptions that the random variables X_i are i.i.d., $\mathrm{E}[X_i] =: \mu < \infty$, $Var(X_i) = \sigma^2 < \infty$, and $E[X_i^4] =: \tau < \infty$. Let us first note that without loss of generality we can assume that $\mu = 0$ by centering. In this case, the strong law says that

$$Pr\left(\lim_{n\to\infty} \overline{X}_n = 0\right) = 1$$

or

$$Pr\left(\omega : \lim_{n\to\infty} \frac{s_n(\omega)}{n} = 0\right) = 1$$

It is equivalent to show that

$$Pr\left(\omega : \lim_{n\to\infty} \frac{s_n(\omega)}{n} \neq 0\right) = 0$$

note that

$$\lim_{n\to\infty} \frac{s_n(\omega)}{n} \neq 0 \iff \exists \epsilon > 0, |\frac{s_n(\omega)}{n}| \geq \epsilon \quad inifinitly\ \ often,$$

and thus to prove the strong law we need to show that for every $\epsilon > 0$, we have

$$Pr\left(\omega : |S_n(\omega)| \geq n\epsilon \quad inifinitly\ \ often,\right) = 0$$

Define the events $A_n = \{\omega : |S_n| \geq n\epsilon\}$, and if we can show that

$$\sum_{n=1}^{\infty} Pr(A_n) < \infty,$$

then the Borel-Cantelli Lemma implies the result. So let us estimate $\Pr(A_n)$. We compute:

$$\mathrm{E}[S_n^4] = \mathrm{E}\left[\left(\sum_{i=i}^{n} X_i\right)^4\right] = \mathrm{E}\left[\sum_{1\leq i,j,k,l\leq n} X_i X_j X_k X_l\right]$$

We first claim that every term of the form

$$X_i^3 X_j, \;\; X_i^2 X_j X_k, \;\; X_i X_j X_k X_l$$

where all subscripts are distinct must have zero expectation. This is because

$$\mathrm{E}[X_i^3 X_j] = \mathrm{E}[X_i^3] E[X_j] \;\; by \;\; indipendence,$$

and the last term is zero — and similarly for the other terms. Therefore, the only terms in the sum with nonzero expectation are

$$\mathrm{E}[X_i^4] \;\; and \;\; \mathrm{E}[X_i^2 X_j^2]$$

.

Since the X_i are identically distributed, all of these are the same, and moreover

$$\mathrm{E}[X_i^2 X_j^2] = (\mathrm{E}[X_i^2])^2$$

.

There are n terms of the form

$$\mathrm{E}[X_i^4] \;\; and \;\; 3n(n-1) \;\; terms \;\; of \;\; the \;\; form \;\; (\mathrm{E}[X_i^2])^2,$$

and so

$$\mathrm{E}[S_n^4] = n\tau + 3n(n-1)\sigma^4$$

Note that the right-hand side is a quadratic polynomial in n, and as such, there exists a $C > 0$ such that

$$\mathrm{E}[S_n^4] \leq Cn^2$$

for n sufficiently large. By Markov's inequality,

$$Pr\left(|S_n| \geq n\epsilon\right) \leq \frac{1}{(n\epsilon)^4}\mathrm{E}[S_n^4] \leq \frac{C}{\epsilon^4 n^2}$$

for n sufficiently large, and therefore, this series is summable. Since this holds for any $\epsilon > 0$, we have established the Strong LLN.

4.1 Difference Between WLLN and SLLN

The difference between weak and strong laws of large numbers is very subtle and theoretical. The Weak law of large numbers suggests that it is a probability that the sample average will converge towards the expected value whereas Strong law of large numbers indicates almost sure convergence. Weak law has a probability near to 1 whereas Strong law has a probability equal to 1. As per Weak law, for large values of n, the average is most likely near is likely near μ. Thus there is a possibility that $(- \mu)> \varepsilon$ happens a large number of times albeit at infrequent intervals. With Strong Law, it is almost certain that $(- \mu)> \varepsilon$ will not occur i.e the probability is 1.

5 APPLICATION OF LAW LARGE NUMBER

now let's see various application of law large number

5.1 Casino's Profit

A Casino may lose money for small number of trials but its earning will move towards the predictable percentage as number of trials increases, so over a longer period of time, the odds are always in favor of the house, irrespective of the Gambler's luck over a short period of time as the law of large numbers apply only when number of observations is large.

5.2 Monte Carlo Problem

Monte Carlo Problems is based on the law of large numbers and it is a type of computational problem algorithm that relies on random sampling to get a numerical result. The main concept of Monte Carlo Problem is to use randomness to solve a problem that appears deterministic in nature. They are often used in computational problems which are otherwise difficult to solve using other techniques. Monte Carlo methods are mainly used in three categories of problem namely: **Optimization problem**, **Integration of numerals** and **draws generation from a probability distribution**.

5.3 Limitations of Law of Large Numbers

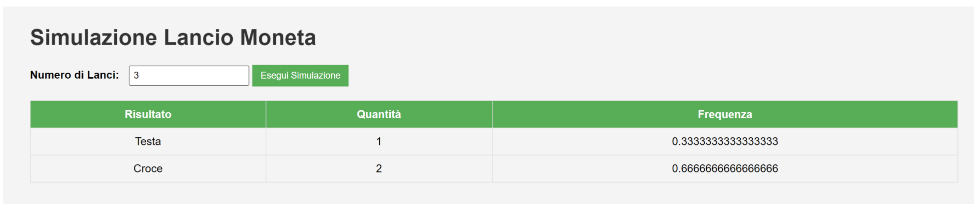
In some cases, the average of a large number of trials may not converge towards the expected value. This happens especially in the case of Cauchy Distribution or Pareto Distribution ($\alpha < 1$) as they have long tails. Cauchy Distribution doesn't have expectation value while as for Cauchy Distribution the expectation value is infinite for $\alpha < 1$. These distributions don't converge towards the expected value as n approaches infinity.

6 SIMULATIONS

Now, we'll take a look at various simulations to gain a simple understanding of the applications of the law of large numbers. This fundamental principle in probability and statistics states that, as the number of observations in a random experiment increases, the average of the results gradually approaches the theoretical expected value.

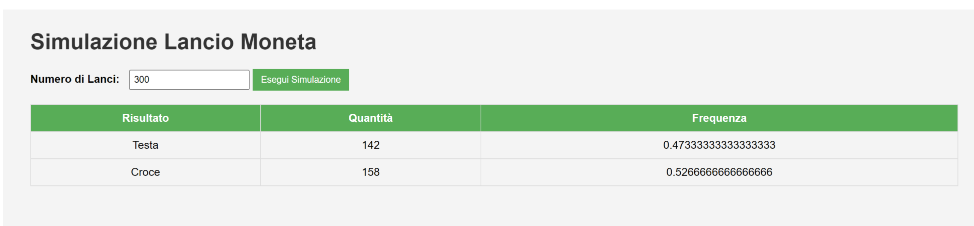
6.1 Coin Flip

the simulation represents the tossing of a fair coin a specified number of times (input provided by the user). The simulation's outcome includes the number of heads, the number of tails, the relative frequency of heads, and the relative frequency of tails. The connection to the law of large numbers is highlighted by the tendency of the relative frequencies of heads and tails to stabilize around the expected theoretical values (0.5 for heads and 0.5 for tails) as the number of tosses increases. In other words, the law of large numbers suggests that with a sufficiently large number of tosses of a fair coin, the relative frequency of heads will increasingly approach 0.5, and the relative frequency of tails will also approach 0.5. [4].



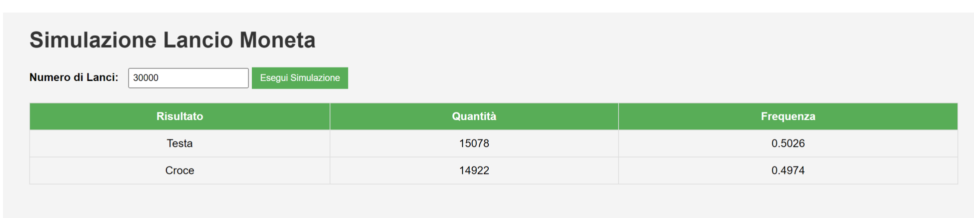
Risultato	Quantità	Frequenza
Testa	1	0.3333333333333333
Croce	2	0.6666666666666666

Figure 1: 5 shot



Risultato	Quantità	Frequenza
Testa	142	0.47333333333333333
Croce	158	0.5266666666666666

Figure 2: 200 shot



Risultato	Quantità	Frequenza
Testa	15078	0.5026
Croce	14922	0.4974

Figure 3: 30000 shot

[4] https://www.educba.com/weak-law-of-large-numbers/

6.2 Simulation Code

```
function lancioMoneta(numeroLanci) {
    let testa = 0;
    let croce = 0;

    for (let i = 0; i < numeroLanci; i++) {
        const risultato = Math.random();

        if (risultato < 0.5) {
            testa++;
        } else {
            croce++;
        }
    }

    const frequenzaTesta = testa / numeroLanci;
    const frequenzaCroce = croce / numeroLanci;

    return {
        testa,
        croce,
        frequenzaTesta,
        frequenzaCroce
    };
}

function eseguiSimulazione() {
    const numeroLanci = parseInt(document.getElementById('numeroLanci').value);
    const risultati = lancioMoneta(numeroLanci);

    document.getElementById('testa').innerText = risultati.testa;
    document.getElementById('croce').innerText = risultati.croce;
    document.getElementById('frequenzaTesta').innerText = risultati.frequenzaTesta;
    document.getElementById('frequenzaCroce').innerText = risultati.frequenzaCroce;
}


eseguiSimulazione();
```

Figure 4: JavaScript Code

7 APPLICATION OF LLN IN CYBERSECURITY

In the context of cybersecurity, the application of the law of large numbers can be fundamental to understanding and mitigating digital threats. Here are some areas where statistical principles can be successfully employed:

- **Future Threat Prediction:** - Through statistical analysis of historical data, predictive models can be developed to identify trends and predict potential future threats. This proactive approach allows organizations to better prepare and adopt preventive measures.

- **User Activity Monitoring:** - The law of large numbers can be employed to monitor user activity and identify significant changes in their digital behaviors. This may contribute to preventing insider threats or unauthorized access by legitimate users.

- **Vulnerability Analysis and Patch Management:** - The law of large numbers can be involved in assessing security vulnerabilities. Analyzing large datasets related to known vulnerabilities and applied patches can help identify trends and patterns, improving patch management strategies.

- **Estimating Threat Impact:** - Using statistical models, it is possible to estimate the magnitude of cybersecurity threats and assess their potential impact. For instance, the law of large numbers can be used to determine the probability of a successful attack, thereby aiding decisions on countermeasures and risk mitigation.

- **Log and Security Data Analysis:** - Examining extensive sets of log data can be challenging, but the law of large numbers can simplify analysis by identifying significant trends. This can help analysts quickly pinpoint relevant security events amid a large amount of information.

- **Intrusion Detection:** - Using statistical techniques, intrusion detection algorithms can be developed based on the evaluation of user activity patterns. For example, the law of large numbers can be employed to establish typical user behaviors and identify statistically significant deviations that might indicate an intrusion attempt.

- **Network Traffic Pattern Analysis:** - The law of large numbers can be applied to analyze network traffic and identify anomalous behavioral patterns. By collecting a vast amount of data on traffic, it is possible to detect significant deviations from the norm that may indicate a cyberattack or unauthorized behavior.

8 CONCLUSION

The law of large numbers is among the most important theorem in statistics. The law of large numbers not only helps us find the expectation of the unknown distribution from a sequence but also helps us in proving the fundamental laws of probability. There are two main versions of the law of large numbers Weak Law and Strong Law, with both being very similar to each other varying only on its relative strength.