The Law of Large Numbers

Sarth Chavan

1 Introduction to Law of Large Numbers

The Law of Large Numbers is an important concept in statistics that illustrates the result when the same experiment is performed in a large number of times. As per the theorem, the average of the results obtained from conducting experiments a large number of times should be near to the Expected value (Population Mean) and will converge more towards the expected value as the number of trials increases. There are two different versions of the Law of Large numbers which are Strong Law of Large Numbers and Weak Law of Large Numbers, both have very minute differences among them. Weak Law of Large Number also termed as "Khinchin's Law" states that for a sample of an identically distributed random variable, with an increase in sample size, the sample means converge towards the population mean. For sufficiently large sample size, there is a very high probability that the average of sample observation will be close to that of the population mean (Within the Margin) so the difference between the two will tend towards zero or probability of getting a positive number ε when we subtract sample mean from the population mean is almost zero when the size of the observation is large.
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. Over the years, many more mathematicians contributed to the evolution of the Law of Large Numbers, refining it to make it what it is today. These mathematicians include: Andrey Markov, Pafnuty Chebyshev who proved a more general case of the Law of Large Numbers for averages, and Khinchin who was the first to provide a complete proof for the case of arbitrary random variables. Additionally, several mathematicians created their own variations of the theorem and that's what really makes this topic and paper so much interesting.

2 Weak Law of Large Numbers.

Before we begin our discussion of the Weak Law of Large Numbers, we first introduce suitable notation and define important terms needed for the discussion. In this chapter some elementary definitions with corresponding examples will also be provided. Additionally, this section will outline the notation that will be seen throughout the remainder of the paper. In order to best understand the content of the paper, one should have an appropriate grasp on important concepts such as expected value or mean, variance, random variables, some standard probability distributions. So this paper assumes a basic knowledge of all those basic concepts in the probability Theory.

2.1 Chebychev's Inequality (In Probability Theory).

Chebychev's Inequality . Let X be a discrete random variable with expected value $\mathbb{E}(X)$ and let $\varepsilon > 0$ be any positive real number. Then we have

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \varepsilon) \leq \frac{Var(X)}{\varepsilon^2}.$$

Proof. Let $m(x)$ denote the distribution function of X. Then the probability that X differs from $\mathbb{E}(X)$ by at least ε is given by

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \varepsilon) = \sum_{|x - \mathbb{E}(X)| \geq \varepsilon} m(x).$$

We already know that

$$Var(X) = \sum_x (x - \mathbb{E}(X))^2 m(x).$$

and this is clearly at least as large as

$$\sum_{|x - \mathbb{E}(X)| \geq \varepsilon} (x - \mathbb{E}(X))^2 m(x)$$

since all the summands are positive and we have restricted the range of summation in the second sum. But this last sum is at least

$$\sum_{|x - \mathbb{E}(X)| \geq \varepsilon} \varepsilon^2 m(x) = \varepsilon^2 \sum_{|x - \mathbb{E}(X)| \geq \varepsilon} m(x) = \varepsilon^2 \mathbb{P}(|x - \mathbb{E}(X)| \geq \varepsilon)$$

$$\sum_{|x-\mathbb{E}(X)|\geq\varepsilon} \varepsilon^2 m(x) = \varepsilon^2 \mathbb{P}(|x-\mathbb{E}(X)| \geq \varepsilon) \implies \mathbb{P}(|X-\mathbb{E}(X)| \geq \varepsilon) \leq \frac{Var(X)}{\varepsilon^2}.$$

Hence we have proved the chebychev inequality. ■

We are now ready to begin our main discussion of the Law of Large Numbers. In this chapter we will state the Weak and Strong Law of Large Numbers as well as Kolmogorov's Law of Large Numbers and prove the Weak Law using two different methods. The first proof uses Chebyshev's inequality, and the second uses what is known as characteristic functions. We will first state the Weak Law and prove it using Chebyshev's Inequality.

2.2 Proof of Weak Law of Large Numbers using the Chebychev's Inequality assuming the Finite Variance Case.

Weak Law of Large numbers . For any $\varepsilon > 0$ we have

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{X_1 + X_2 + X_3 + \ldots + X_n}{n} - \mathbb{E}(X_i)\right| > \varepsilon\right) = 0.$$

The proof of weak law of large numbers directly follows from the chebychev inequality. Chebychev inequality states that

$$\mathbb{P}(|X-\mathbb{E}(X)| \geq \varepsilon) \leq \frac{Var(X)}{\varepsilon^2}$$

Note that from now on we willl denote

$$\frac{1}{n}\sum_{i=1}^{n} X_i = Z_n, \text{ and } Var(Y_i) = \sigma^2.$$

Thus we have:

$$\mathbb{P}(|X-\mathbb{E}(X)| \geq \varepsilon) \leq \frac{Var(X)}{\varepsilon^2} \implies \mathbb{P}(|Z_n-\mathbb{E}(Z_n)| \geq \varepsilon) \leq \frac{Var(Z_n)}{\varepsilon^2}.$$

Now we will compute $\mathbb{E}(Z_n)$ and $Var(Z_n)$.

$$\mathbb{E}(Z_n) = \mathbb{E}\left(\frac{X_1 + X_2 + X_3 + \ldots + X_n}{n}\right) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\left(\sum_{i=1}^{n}\mathbb{E}(X_i)\right)$$

$$\frac{1}{n}\left(\sum_{i=1}^{n}\mathbb{E}(X_i)\right) = \frac{n\mathbb{E}(X_i)}{n} = \mathbb{E}(X_i) \implies \mathbb{E}(Z_n) = \mathbb{E}(X_i).$$

Now let's compute $Var(Z_n)$.

$$Var(Z_n) = Var\left(\frac{X_1 + X_2 + X_3 + \ldots + X_n}{n}\right) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}Var\left(\sum_{i=1}^{n} X_i\right)$$

$$\frac{1}{n^2}Var\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\left(\sum_{i=1}^{n} Var(X_i)\right) = \frac{1}{n^2}\left(\sum_{i=1}^{n} \sigma^2\right) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

$$\implies Var(Z_n) = \frac{\sigma^2}{n}.$$

Therefore we have

$$\mathbb{P}(|Z_n - \mathbb{E}(X_i)| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \ \ \forall\, \varepsilon > 0.$$

Now take the limit as $n \to \infty$, we have

$$\lim_{n\to\infty} \left[\mathbb{P}(|Z_n - \mathbb{E}(X_i)| \geq \varepsilon)\right] \leq \lim_{n\to\infty} \left(\frac{\sigma^2}{n\varepsilon^2}\right) = 0 \implies \lim_{n\to\infty} \left[\mathbb{P}(|Z_n - \mathbb{E}(X_i)| \geq \varepsilon)\right] = 0.$$

$$\lim_{n\to\infty} \left[\mathbb{P}(|Z_n - \mathbb{E}(X_i)| \geq \varepsilon)\right] = 0$$

$$\implies \lim_{n\to\infty} \mathbb{P}\left(\left|\frac{X_1 + X_2 + X_3 + \ldots + X_n}{n} - \mathbb{E}(X_1)\right| > \varepsilon\right) = 0.$$

Thus we have proved the weak law of large numbers. Next we will prove the Weak Law of Large Numbers using characteristic functions.

2.3 Proof of Weak Law of Large Numbers using Characteristic function with Infinite variance.

The proof for the finite variance case is pretty simple and is more widely known. However, since finite variance is not a necessary condition for the Weak Law of Large Numbers (WLLN), there's utility in knowing the proof for the infinite variance case in the interest of completeness. First, let's define the Characteristic function of an arbitrary random variable, and provide some properties for independent and identically distributed random variables that we might find helpful:

Definition 1.3.1 (The Characteristic Function of Random Variables)**.** For a Random Variable α, the characteristic function $\varphi_\alpha(t)$ is defined as

$$\varphi_\alpha(t) = \mathbb{E}[e^{itA}]$$

Also Note that for two independent random variables α and β we have

$$\varphi_{\alpha+\beta}(t) = \mathbb{E}\left[e^{it(A+B)}\right] = \mathbb{E}\left[e^{(itA+itB)}\right] = \mathbb{E}\left[e^{itA}e^{itB}\right] = \mathbb{E}\left[e^{itA}\right]\mathbb{E}\left[e^{itB}\right] = \varphi_\alpha(t)\varphi_\beta(t).$$

Also Note that for a random variable α we have

$$\varphi_{\alpha/n}(t) = \mathbb{E}\left[e^{(itA)/n}\right] = \varphi_{\alpha}\left(\frac{t}{n}\right).$$

Now let's Evaluate $e^{it\alpha}$ using the Taylors Theorem for Complex Functions **Corollary 1.3.2** (Corollary of Taylors Theorem for Complex Functions)**.**

$$e^{itA} = 1 + itA + \frac{(itA)^2}{2!} + \frac{(itA)^3}{3!} + \frac{(itA)^4}{4!} + \ldots + \frac{(itA)^n}{n!} + \ldots = 1 + itA + o(t).$$

Now taking Expected value on both the sides we have

$$\mathbb{E}\left(e^{itA}\right) = \mathbb{E}(1 + itA + o(t)) = 1 + \mathbb{E}(itA) + o(t).$$

We're now ready for the proof. Let's begin with the Characteristic function of our sample average of the n independent and identically distributed X random variables.

$$\varphi_{Z_n}(t) = \varphi_{\frac{1}{n}\sum_{i=1}^{n} X_i}(t) = \varphi_{\sum_{i=1}^{n}\frac{X_i}{n}}(t)$$

$$= \left(\varphi_{X/n}(t)\right)^n = \left(\varphi_X\left(\frac{t}{n}\right)\right)^n \impliedby \varphi_{\alpha/n}(t) = \varphi_{\alpha}\left(\frac{t}{n}\right)$$

$$\left(\varphi_X\left(\frac{t}{n}\right)\right)^n = \left(\mathbb{E}\left(e^{(itX)/n}\right)\right)^n \impliedby \varphi_{\alpha/n}(t) = \mathbb{E}\left[e^{(itA)/n}\right] = \varphi_{\alpha}\left(\frac{t}{n}\right).$$

$$\left(\mathbb{E}\left(e^{(itX)/n}\right)\right)^n = \left(1 + \frac{it\mathbb{E}(X)}{n} + o\left(\frac{t}{n}\right)\right)^n \impliedby \mathbb{E}\left(e^{(it\alpha)/n}\right) = 1 + \frac{it\mathbb{E}(\alpha)}{n} + o\left(\frac{t}{n}\right).$$

$$\left(1 + \frac{it\mathbb{E}(X)}{n} + o\left(\frac{t}{n}\right)\right)^n = \left(1 + \frac{it\mathbb{E}(X_i)}{n} + o\left(\frac{t}{n}\right)\right)^n \impliedby \mathbb{E}(X) = \mathbb{E}(X_i)$$

$$\implies \varphi_{Z_n}(t) = \left(1 + \frac{it\mathbb{E}(X_i)}{n} + o\left(\frac{t}{n}\right)\right)^n$$

$$\implies \lim_{n\to\infty}\left(\varphi_{Z_n}(t)\right) = \lim_{n\to\infty}\left(\left(1 + \frac{it\mathbb{E}(X_i)}{n} + o\left(\frac{t}{n}\right)\right)^n\right) = e^{itZ_n} = \mathbb{E}(e^{itZ_n}) = \varphi_{\mathbb{E}(X_i)}(t).$$

Thus we have shown that $\varphi_{Z_n}(t)$ has pointwise convergence to $\varphi_{\mathbb{E}(X_i)}(t)$.
With the help of Levy's continuity theorem we know that Z_n has a convergence in the distribution to $\mathbb{E}(X_i)$. Because $\mathbb{E}(X_i)$ is fixed we get that

$$Z_n \to \mathbb{E}(X_i) \implies \lim_{n\to\infty} \mathbb{P}\left(\left|\frac{X_1 + X_2 + X_3 + \ldots + X_n}{n} - \mathbb{E}(X_i)\right| > \varepsilon\right) = 0.$$

Thus we have proved The Weak Law of Large Numbers using Characteristic function with Infinite variance.

Interpretation: As per Weak Law of large numbers for any value of non-zero margins, when the sample size is sufficiently large, there is a very high chance that the average of observation will be nearly equal to the expected value within the margins. The weak law in addition to independent and identically distributed random variables also applies to other cases. For example, if the variance is different for each random variable but the expected value remains constant then also the rule applies. If the variance is bounded then also the rule applies as proved by Chebyshev in 1867. Chebyshev's proof works as long as the variance of the first n average value converges to zero as n move towards infinity.

3 Strong Law of Large Numbers.

Now we will state the more powerful variation of our theorem; The Strong Law of Large Numbers.

Strong Law of Large numbers . Let $X_1, X_2, X_3, X_4, \ldots$ be a sequence of independent and idenically distributed random variables, then we have

$$\mathbb{P}\left(\lim_{n\to\infty}\left(\frac{X_1 + X_2 + X_3 + \ldots + X_n}{n}\right) = \mathbb{E}(X_1)\right) = 1$$

Before we begin our discussion of the Strong Law of Large Numbers, we first introduce suitable notation and define some very important terms and concepts needed for the discussion and proof of the Strong Law of Large numbers. In this section some definitions will be provided. Additionally, this section will outline the notation that will be seen throughout the remainder of the paper. In order to best understand the content of the paper, one should have an appropriate grasp on important concepts such as Markov's Inequality, The First Borel-Cantelli Lemma, The Second Borel-Cantelli Lemma, so we would be seeing all those prelimnaries in this section so that it would be very easy to understand the latter part (that is the proof of the Strong Law of Large Numbers) of the paper. So without making any further delay, let's see the Markov's Inequality.

3.1 Statement and Proof of The Markov Inequality.

Markov's Inequality . Let X be a non-negative random variable, then we have

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

Proof. Let us define a random variable Y such as

$$Y = \begin{cases} 1 & \iff X \geq a \\ 0 & \text{otherwise} \end{cases}$$

Now if $X < a$, $Y = 0$. Otherwise, $X \geq a$, in which case $Y = 1$. In both the cases we get that $Y \leq X/a$. Note that we use the fact that X is a non-negative random variable in the first case.Therefore

$$\implies \mathbb{E}[Y] \leq \frac{\mathbb{E}[X]}{a}.$$

However since Y is an indicator random variable, we have

$$\mathbb{E}[Y] = \mathbb{P}(Y = 1) = \mathbb{P}(X \geq a) \implies \mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

■

Thus we have proved the Markov's Inequality. Now since were done with the markov Inequality, let's move towards a very crucial part in the Probability theory that is The First Borel-Cantelli Lemma and The Second Borel-Cantelli Lemma. Both will play a very crucial role in the proof of Strong Law of Large Numbers.

3.2 Statement and Proof of The First Borel-Cantelli Lemma.

The First Borel-Cantelli Lemma . Let $A_1, A_2, A_3, \ldots$ be a sequence of events in some probability space. The Borel–Cantelli lemma states that if:

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty \implies \mathbb{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m\right) = 0$$

Proof. We should note that

$$\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m \subset \bigcup_{m=n}^{\infty} A_m \quad \forall\ n \geq 1.$$

$$\implies \mathbb{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m\right) \leq \mathbb{P}\left(\bigcup_{m=n}^{\infty} A_m\right) \leq \sum_{m=n}^{\infty} \mathbb{P}(A_m) \to 0 \impliedby \lim_{n\to\infty} \sum_{m=n}^{\infty} \mathbb{P}(A_m) = 0.$$

$$\lim_{n\to\infty} \sum_{m=n}^{\infty} \mathbb{P}(A_m) = 0 \impliedby \text{ whenever } \sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty \implies \mathbb{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m\right) = 0.$$

■

3.3 Statement and Proof of The Second Borel-Cantelli Lemma.

The Second Borel-Cantelli Lemma . Let $E_1, E_2, E_3, \dots$ be a sequence of independant events in some probability space. The Second Borel–Cantelli lemma states

$$\sum_{n=1}^{\infty} \mathbb{P}(E_n) = \infty \implies \mathbb{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} E_k\right) = 1$$

Proof. We need to show that the E_n's did not occur infinitely many value of n has probability 0. Thus we need to prove that

$$1 - \mathbb{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} E_k\right) = 0 \text{ Note that } 1 - \mathbb{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} E_k\right) = \mathbb{P}\left(\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} E_k\right)^c\right).$$

$$\mathbb{P}\left(\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} E_k\right)^c\right) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} E_k^c\right) = \mathbb{P}\left(\liminf_{k\to\infty} E_k^c\right) = \lim_{k\to\infty} \mathbb{P}\left(\bigcap_{k=n}^{\infty} E_k^c\right)$$

It suffices to show that $\mathbb{P}\left(\bigcap_{k=n}^{\infty} E_k^c\right) = 0$, Since $E_1, E_2, E_3, E_4, \dots$ are independent.

$$\mathbb{P}\left(\bigcap_{k=n}^{\infty} E_k^c\right) = \prod_{k=n}^{\infty} \mathbb{P}(E_k^c) = \prod_{k=n}^{\infty} (1 - \mathbb{P}(E_k)) \leq \prod_{k=n}^{\infty} \exp(-\mathbb{P}(E_k)).$$

$$= \exp\left(-\sum_{k=n}^{\infty} \mathbb{P}(E_k)\right) = 0 \implies \mathbb{P}\left(\bigcap_{k=n}^{\infty} E_k^c\right) = 0 \implies \mathbb{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} E_k\right) = 1$$

Alternatively, we can see the same by taking negative the logarithm of both sides to get:

$$\mathbb{P}\left(\bigcap_{k=n}^{\infty} E_k^c\right) = \prod_{k=n}^{\infty} (1 - \mathbb{P}(E_k)) \implies -\log\left(\mathbb{P}\left(\bigcap_{k=n}^{\infty} E_k^c\right)\right) = -\log\left(\prod_{k=n}^{\infty} (1 - \mathbb{P}(E_k))\right)$$

$$-\log\left(\mathbb{P}\left(\bigcap_{k=n}^{\infty} E_k^c\right)\right) = -\log\left(\prod_{k=n}^{\infty} (1 - \mathbb{P}(E_k))\right) = -\sum_{k=n}^{\infty} \log(1 - \mathbb{P}(E_k)) = 0$$

$$\text{The reader should Note that } -\sum_{k=n}^{\infty} \log(1 - \mathbb{P}(E_k)) = 0 \impliedby \sum_{n=1}^{\infty} \mathbb{P}(E_n) = \infty.$$

$$-\sum_{k=n}^{\infty} \log(1 - \mathbb{P}(E_k)) = 0 \implies \mathbb{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} E_k\right) = 1.$$

∎

3.4 Proof of Strong Law of Large Numbers (Restricted Case) using The Markov Inequality and The Tonelli's Theorem.

Strong Law of Large numbers . Let $X_1, X_2, X_3, X_4, \ldots$ be a sequence of independent and idenically distributed random variables, then we have

$$\mathbb{P}\left(\lim_{n\to\infty}\left(\frac{X_1+X_2+X_3+\ldots+X_n}{n}\right)=\mathbb{E}(X_1)\right)=1$$

Note that Strong Law of Large Numbers remains valid without the assumption that $\mathbb{E}(X_1^4)<\infty$ (That is assuming that the fourth moment is finite), just assuming that $\mathbb{E}(|X_1|)<\infty$ (That is assuming that the first moment is finite). The proof for the general result is hard, but under the extra moment condition $\mathbb{E}(X_1^4)<\infty$ there is a nice proof.

Proof. We set $S_n = X_1+X_2+X_3+X_4+\ldots$. Now let's see a lemma.

Lemma 2.3.1. In The Strong Law of Large Numbers, there is a constant K such that for all $n \geq 0$ we have

$$\mathbb{E}((S_n - n\mathbb{E}(X_1))^4) \leq Kn^2.$$

Proof. Let $Z_k = X_k - \mathbb{E}(X_1)$ and $T_n = Z_1+Z_2+Z_3+\ldots+Z_n = S_n - n\mathbb{E}(X_1)$. We have

$$\mathbb{E}(T_n^4)=\mathbb{E}((S_n-n\mathbb{E}(X_1))^4)=\left(\left(\sum_{i=1}^{n} Z_i\right)^4\right)=n\mathbb{E}(Z_1^4)+3n(n-1)\mathbb{E}(Z_1^2Z_2^2)\leq Kn^2$$

by expanding the fourth power and noting that most terms vanish such as

$$\mathbb{E}(Z_1Z_2^3)=\mathbb{E}(Z_1)\mathbb{E}(Z_2^3)=0 \impliedby \mathbb{E}(XY)=\mathbb{E}(X)\mathbb{E}(Y).$$

K was chosen appropriately, say $K = 4\max\{\mathbb{E}(Z_1^4), (\mathbb{E}(Z_1^2))^2\}$. ■

Now by Lemma 2.3.1 we know that

$$\mathbb{E}\left(\left(\frac{S_n}{n}-\mathbb{E}(X_1)\right)^4\right)\leq \frac{k}{n^2}.$$

Now, by Tonelli's theorem we have,

$$\mathbb{E}\left(\sum_{n>0}\left(\frac{S_n}{n}-\mathbb{E}(X_1)\right)^4\right)=\sum_{n>0}\mathbb{E}\left(\left(\frac{S_n}{n}-\mathbb{E}(X_1)\right)^4\right)<\infty$$

$$\implies \sum_{n>0}\left(\frac{S_n}{n} - \mathbb{E}(X_1)\right)^4 < \infty$$

But if a series converges, the underlying sequence converges to zero, and so

$$\implies \left(\frac{S_n}{n} - \mathbb{E}(X_1)\right)^4 \to 0$$

$$\implies \mathbb{P}\left(\lim_{n\to\infty}\left(\frac{X_1 + X_2 + X_3 + \ldots + X_n}{n}\right) = \mathbb{E}(X_1)\right) = 1$$

■

3.5 Proof of Strong Law of Large Numbers (Restricted Case) using The Borel-Cantelli Lemmas

Proof. Note that we will Assume, without loss of generality, that the fixed $\mathbb{E}(X_1) = 0$. We can allow this just by the transformation of Random variables where if $\mathbb{E}(X_1)$ would be non-zero then we could always start with a new set of random varibles $W_k = X_k - \mathbb{E}(X_1)$, with $\mathbb{E}(W_i) = 0$. Now we will consider a scenario where the limit of Z_n as $n \to \infty$ is not equal to the $\mathbb{E}(X_1) = 0$: (Note that here $Z_n = (X_1 + X_2 + X_3 + \ldots + X_n)/n = S_n/n$).

$$\lim_{n\to\infty}(Z_n) \neq \mathbb{E}(X_1) = 0$$

$$\implies \lim_{n\to\infty}\left(\frac{X_1 + X_2 + X_3 + \ldots + X_n}{n}\right) = \lim_{n\to\infty}\left(\frac{S_n}{n}\right) \neq \mathbb{E}(X_1) = 0.$$

This would therefore mean that there exists some $\varepsilon > 0$ such that for infinitely many n we have

$$\left|\frac{X_1 + X_2 + X_3 + \ldots + X_n}{n}\right| = \left|\frac{S_n}{n}\right| > \varepsilon \implies |S_n| > n\varepsilon.$$

Thus we need to prove that the below equality which would imply the latter

$$\mathbb{P}\left(\lim_{n\to\infty}(|X_1 + X_2 + X_3 + \ldots + X_n|) > n\varepsilon\right) = \mathbb{P}\left(\lim_{n\to\infty}(|S_n|) > n\varepsilon\right) = 0$$

$$\implies \mathbb{P}\left(\lim_{n\to\infty}\left(\frac{X_1 + X_2 + X_3 + \ldots + X_n}{n}\right) = 0\right) = \mathbb{P}\left(\lim_{n\to\infty}\left(\frac{S_n}{n}\right) = 0\right) = 1.$$

We know that Markov Inequality States that: For a non-negative random variable $X \geq 0$, we have

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}, \text{ for all } a > 0.$$

$$\implies \mathbb{P}(X \geq a) = \mathbb{P}\left((|X|)^2 \geq a^2\right) = \mathbb{P}\left((|X|)^3 \geq a^3\right) = \mathbb{P}\left((|X|)^k \geq a^k\right) \leq \frac{(|X|)^k}{a^k}$$

$$\mathbb{P}\left((|X|)^k \geq a^k\right) \leq \frac{(|X|)^k}{a^k} \implies \mathbb{P}\left(|S_n| \geq n\varepsilon\right) \leq \frac{\mathbb{E}\left(S_n^4\right)}{n^4\varepsilon^4}.$$

$$\mathbb{E}\left(S_n^4\right) = \mathbb{E}\left(\left(\sum_{i=1}^{n} X_i\right)^4\right) = \mathbb{E}\left(\sum_{i,j,k,l \leq n} X_i X_j X_k X_l\right) \impliedby \text{ Multinomial theorem.}$$

Now expanding, we will get terms of the form $\mathbb{E}\left(X_i^3 X_j\right), \mathbb{E}\left(X_i^2 X_j X_k\right), \mathbb{E}\left(X_i X_j X_k X_l\right)$ with i, j, k, l all distinct. These terms are all equal to zero since we know that $\mathbb{E}(X_1) = 0$. Thus the nonzero terms in the above sum are

$$\mathbb{E}\left(X_i^4\right), \mathbb{E}\left(X_i^2 X_j^2\right) = \left(\mathbb{E}\left(X_i^2\right)\right)^2 \impliedby \mathbb{E}\left(X_i^2 X_j^2\right) = \mathbb{E}\left(X_i^2\right)\mathbb{E}\left(X_i^2\right) = \left(\mathbb{E}\left(X_i^2\right)\right)^2.$$

Now we know that there are n terms of the form $\mathbb{E}\left(X_i^4\right)$ in the sum and $3n(n-1)$ terms of the form $\mathbb{E}\left(X_i^2 X_j^2\right) = \left(\mathbb{E}\left(X_i^2\right)\right)^2$. Thus we have shown that

$$\mathbb{E}\left(S_n^4\right) = \mathbb{E}\left(\left(\sum_{i=1}^{n} X_i\right)^4\right) = n\mathbb{E}\left(X_i^4\right) + 3n(n-1)\left(\mathbb{E}\left(X_i^2\right)\right)^2 = n\mathbb{E}\left(X_i^4\right) + 3n(n-1)\sigma^4$$

Note that the last equality is implied by $\sigma^2 = \left(\mathbb{E}\left(X_i^2\right)\right) \implies \sigma^4 = \left(\mathbb{E}\left(X_i^2\right)\right)^2$.

$$\mathbb{E}\left(S_n^4\right) = \mathbb{E}\left(\left(\sum_{i=1}^{n} X_i\right)^4\right) = n\mathbb{E}\left(X_i^4\right) + 3n(n-1)\left(\mathbb{E}\left(X_i^2\right)\right)^2 = n\mathbb{E}\left(X_i^4\right) + 3n(n-1)\sigma^4$$

$$= n\mathbb{E}\left(X_i^4\right) + 3n^2\sigma^4 - 3n\sigma^4 = 3n^2\sigma^4 + n\left(\mathbb{E}\left(X_i^4\right) - 3\sigma^4\right) \leq n^2\left(3\sigma^2 + 1\right)$$

With the help of Markov Inequality we already know that $\mathbb{P}\left(|S_n| \geq n\varepsilon\right) \leq \dfrac{\mathbb{E}\left(S_n^4\right)}{n^4\varepsilon^4}$.

$$\implies \mathbb{P}\left(|S_n| \geq n\varepsilon\right) \leq \frac{\mathbb{E}\left(S_n^4\right)}{n^4\varepsilon^4} \leq \frac{n^2\left(3\sigma^2+1\right)}{n^4\varepsilon^4} = \frac{\left(3\sigma^2+1\right)}{n^2\varepsilon^4} \impliedby \mathbb{E}\left(S_n^4\right) \leq n^2\left(3\sigma^2+1\right).$$

$$\mathbb{P}\left(|S_n| \geq n\varepsilon\right) \leq \frac{\left(3\sigma^2+1\right)}{n^2\varepsilon^4} \implies \sum_{n \geq \alpha} \mathbb{P}\left(|S_n| \geq n\varepsilon\right) \leq \sum_{n \geq \alpha} \frac{\left(3\sigma^2+1\right)}{n^2\varepsilon^4}.$$

Note that here α is the first n such that $\mathbb{E}\left(S_n^4\right) \leq n^2\left(3\sigma^2+1\right)$ holds. Since we are neglecting a finite set of terms in the sum, this cannot affect the convergence or divergence of the infinite series.) Thus by the Borel-Cantelli lemma we have

$$\mathbb{P}\left(\lim_{n\to\infty}\left(|X_1 + X_2 + X_3 + \ldots + X_n|\right) > n\varepsilon\right) = \mathbb{P}\left(\lim_{n\to\infty}\left(|S_n|\right) > n\varepsilon\right) = 0$$

$$\implies \mathbb{P}\left(\lim_{n\to\infty}\left(\frac{X_1+X_2+X_3+\ldots+X_n}{n}\right)=\mathbb{E}(X_1)\right)=1.$$

Thus we have Proved the Strong Law of Large Numbers. A monument to ingenuity and persistence! ■

4 Difference Between Strong Law of Large Numbers and Weak Law of Large Numbers

Given that there is both a Weak Law of Large Numbers and Strong Law of Large Numbers, eventually it raises very important questions such as, How does the Weak Law differ from the Strong Law? and also Are there any situations in which one version of the Law of Large Numbers is better than the other one?
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 $\mathbb{E}(X_1)$. Thus there is a possibility that $(-\mathbb{E}(X_1)) > \varepsilon$ happens a large number of times albeit at infrequent intervals. With Strong Law, it is almost certain that $(-\mathbb{E}(X_1)) > \varepsilon$ will not occur that is the probability is 1.

5 Limitations of Law of Large Numbers

The average of the results obtained from a large number of trials may fail to converge in some cases. For instance, the average of n results taken from the Cauchy distribution or some Pareto distributions ($\alpha < 1$) will not converge as n becomes larger; the reason is heavy tails. The Cauchy distribution and the Pareto distribution represent two cases: the Cauchy distribution does not have an expectation, whereas the expectation of the Pareto distribution ($\alpha < 1$) is infinite.

6 Applications of The Law of Large Numbers

Like many other great results in the fields of probability and statistics, the Law of Large Numbers has many useful applications to a variety of fields. In this section we willexplore how to apply the Law of Large Numbers by means of general examples as well as discussing Monte Carlo Methods. To begin to think about how to apply the Law of Large Numbers let us consider a very basic example.

Example 1. A coin is flipped 9999 times with the coin landing on heads 4998 times and tails 5001 times. What is the probability that the next coin flip will land on heads? This is a trick question. The probability that the coin will land on heads is 0.5 as it always is. We know by the Law of Large Numbers that as the number of trials increases we get closer and closer to the mean, however this does not guarantee that any fixed flip will land on heads. This is an important idea as it utterly disproves what is called the Gambler's Fallacy; the belief that in an event such as this, the absence of a coin landing on heads several (or many) consecutive flips implies that a flip landing on heads is imminent.

Example 2. Another good example of the Law of large Numbers is the Monte Carlo method and 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. Monte Carlo methods are mainly used in three categories of problem namely: Optimization problem, Integration of numerals and draws generation from a probability distribution.

Example 3. Another good example of the Law of large Numbers is the 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.

Example 4. Insurance companies rely on the law of large numbers to help estimate the value and frequency of future claims they will pay to policyholders. However, the theoretical benefits from the law of large numbers do not always hold up in the real world. The law of large numbers states that as a company grows, it becomes more difficult to sustain its previous growth rates.

Example 5. In the field of cryptography there is a heavy reliance on probability when trying to decode messages encoded using specific methods. The English alphabet has a very specific frequency distribution for each individual letter. The law of large numbers implies that the longer the encoded message, the higher the probability the frequency distribution of each letter approaches the expected value. Similarly, if one wishes to know the probability that a specific word appears in a message, or piece of writing, the law of large numbers implies that this probability increases significantly the larger the length of the message. The applications of the Law of Large Numbers are not only limited to the ones discussed in this paper. There are many other interesting concepts that could be explored in regards to the Law of Large Numbers.