1. Introduction

Welcome to the series of E-learning modules on the concept of Weak Law of Large Numbers (Statement) with applications. In this module, we are going to cover the concept of law of large numbers, statement of weak law of large numbers, difference between weak and strong law of large numbers, its applications and results associated with it.

By the end of this session, you will be able to:

- Explain the law of large numbers
- Explain the statement of weak law of large numbers
- Explain the implications and applications of weak law of large numbers
- Explain the difference between weak and strong law of large numbers

In probability theory, the **law of large numbers** 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 trials should be close to the expected value, and will tend to become closer as more trials are performed.

For example, a single roll of a six-sided dice produces one of the numbers from 1 to 6, each with equal probability. Therefore, the expected value of a single dice roll is 1 plus 2 plus etc plus 6 by 6 is equal to 3 point five.

According to the law of large numbers, if a large number of six-sided dice are rolled, the average of their values (sometimes called the sample mean) is likely to be close to 3 point 5, with the accuracy increasing as more dice are rolled.
It follows from the law of large numbers that the empirical probability of success in a series of Bernoulli trials will converge to the theoretical probability.

For a Bernoulli random variable, the expected value is the theoretical probability of success and the average of n such variables (assuming they are independent and identically distributed) is precisely the relative frequency.

For example, tossing of a coin is a Bernoulli trial. When a fair coin is flipped once, the theoretical probability that the outcome will be heads is equal to half. Therefore, according to the law of large numbers, the proportion of heads in a "large" number of coin flips "should be" roughly half. In particular, the proportion of heads after n flips will almost converge to half as n approaches infinity.

Though the proportion of heads (and tails) approaches half, almost the absolute (nominal) difference in the number of heads and tails will become large as the number of flips becomes large. That is, the probability that the absolute difference is a small number approaches zero as the number of flips becomes large.

The Law of Large Number (LLN) is important because it "guarantees" stable long-term results for random events. For example, while a casino may lose money in a single spin of the roulette wheel, its earnings will tend towards a predictable percentage over a large number of spins. Any winning streak by a player will eventually be overcome by the parameters of the game.

It is important to remember that the LLN only applies (as the name indicates) when a *large number* of observations are considered. There is no principle that a small number of observations will converge to the expected value or that a streak of one value will immediately be "balanced" by the others.

Forms of Law of Large Numbers

Two different versions of the Law of Large Numbers are described, which are called the Strong Law of Large Numbers, and the Weak Law of Large Numbers. Both versions of the law state that – with virtual certainty – the sample average Xn bar converges to the expected value.

X n bar tends to mu for n tends to infinity
Where, X 1, X2, etc. is an infinite sequence of independent and identically distributed random variables with expected value, Expected value of X1 is equal to Expected value of X2 etc. is equal to mu. Expected value of modulus of Xj is finite for j is equal to 1,2 etc
An assumption of finite variance V of X1 is equal to V of X2 etc equal to sigma square which is finite is not necessary. Large or infinite variance will make the convergence slower, but the LLN holds anyway. This assumption is often used because it makes the proofs easier and shorter.

The difference between the strong and the weak version is concerned with the mode of convergence being asserted. Already we are familiar with two types of convergence of random variables and its interpretations namely Convergence in distribution and Convergence in probability.

2. Weak Law of Large Numbers

A theorem of importance usually known as Bernoulli's theorem was first published posthumously in the first part of the eighteenth century. French mathematician Simeon Poisson gave it the name “Law of Large Numbers”. According to Bernoulli, it took him twenty years to complete the theorem.

Later Poisson proved an analogous theorem at the beginning of the nineteenth century under more general conditions. The Russian mathematician Chebyshev discovered his proof in eighteen sixty-six using this inequality. Later Markov obtained a more general result using Tchebyscheff's reasoning.

In Nineteen twenty eight, Khinchin showed that for a sequence of independent and identically distributed random variables the Law of Large numbers holds if the expectations exist. The above works were concerned with what is known as Weak Law of Large Numbers.

If you toss a fair coin only twice, although "Heads" and "Tails" both have a probability of point five, you certainly would not be too surprised if the two tosses happened to produce both "Heads", or both "Tails", instead of producing exactly one "Head" and one "Tail".

However, if you now toss the same fair coin one thousand times, you certainly expect the number of "Heads" to be **very close** to five hundred.

The Weak Law of Large Numbers backs this intuition. This Law states that if a trial is reproduced a large number of times *n*, then it becomes exceedingly improbable that the average of the outcomes of these *n* trials will differ significantly from the expected value of one outcome as *n* grows without limit.

In more technical terms, the Weak Law of Large Numbers states that:

Suppose {Xi} is an infinite sequence of independent and identically distributed random variables with common mean mu and if we define Yn as the random variable equal to the mean of the first n, Xis, then for any epsilon, the probability for a realization of Yn to fall more than epsilon away from mu tends to 0 as n grows without limit.

No matter how small the epsilon is, all we have to do is to find the probability for the mean of the first n terms to differ it from the mean mu to keep it more than the epsilon but as small as you wish, so that we can make n large enough.

In the vocabulary of Estimation, the WLLN states that the sample mean is a consistent estimator of the population mean.

3. Tchebyscheff's Theorem of WLLN

Let {Xn} be a sequence of independent random variables such that Expected value of Xi is equal to mu i and variance of Xi is equal to sigma i square.
Let Bn is equal to Variance of (X1 plus X2 plus up to plus Xn) is finite.
Then,
Probability of modulus of (X1 plus X2 plus up to plus Xn by n minus mu 1 plus mu 2 plus up to plus mu n by n) less than epsilon is greater than or equal to 1 minus eta for all n greater than n not , where epsilon and eta are arbitrary small positive numbers provided Limit of summation sigma i square by n square as n tends to infinity is equal to limit of Bn by n square as n tends to infinity is equal to 0.

Assumed arithmetic mean of the expectations should be less than any given number however small it may be, provided the condition that number of variables can be taken sufficiently large and Limit of Bn by n square as n tends to infinity equal to 0 are fulfilled.

Remark:
1)WLLN can also be stated as follows:
Xn bar converges to mu bar in probability provided Bn by n square tends to zero as n tends to infinity, symbols having their usual meanings.
2) For the existence of the law, we assume the following conditions:
i) Expected value of Xi exists for all i
ii) Bn is equal to Variance of (X1 plus X2 plus up to plus Xn) exists and
iii) Bn by n square tends to zero as n tends to infinity

Condition (i) is necessary, without which the law cannot be stated. The conditions (ii) and (iii) need not be necessary. However, condition (iii) is a sufficient condition to prove the Weak Law of Law Numbers.

3) Corollary:
If the variables X1, X2, up to Xn are independently and identically distributed, then Expected value of Xi is equal mu and Variance of Xi is equal sigma square. Bn is equal to n sigma square.
Bn by n square is equal to n sigma square by n square tends to zero as n tends to infinity,
If sigma i square is equal to sigma square, then the condition of the theorem is automatically satisfied and we have the result Xn bar minus mu tends to zero in probability. Further, if the random variables {Xi} are independent and identically distributed with Expected value of Xi is equal to mu and Variance of Xi is equal to sigma square then,
Xn bar converges to mu in probability that is the sample mean converges stochastically to the population mean.

The standard terminology of Calculus would reformulate the Law as follows:

- No matter how small epsilon greater than 0,
- No matter how small delta greater than 0,

- There is a number N of (epsilon, delta) such that if *n* is greater than N of (epsilon, delta) then,

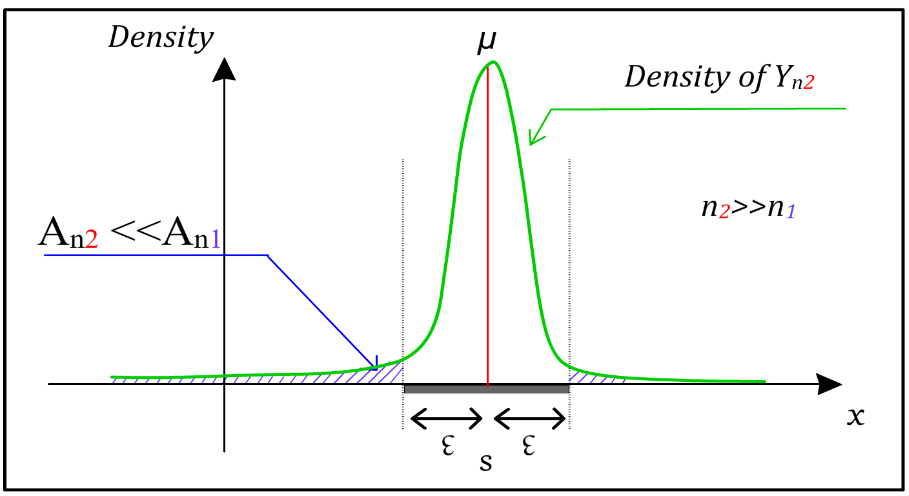
Probability of (modulus of (X1 plus X2 plus up to Xn by n minus mu) greater than epsilon is less than delta

Hence, the weak law of large numbers makes use of convergence in Probability and it states that the sample average converges in probability towards the expected value,
Xn bar converges in probability to mu, when n tends to infinity
That is to say that for any positive number epsilon,
Limit as n tends to infinity of Probability of (modulus of (Xn bar minus mu) greater than epsilon is equal to zero.

Interpreting this result, the weak law essentially states that for any non-zero margin specified, no matter how small, with a sufficiently large sample, there will be a very high probability that the average of the observations will be close to the expected value, that is, within the margin. Convergence in probability is also called weak convergence of random variables. This version is called the weak law because random variables may converge weakly (in probability) as above without converging strongly as in the case of Strong Law of Large Numbers.

Graphic interpretation of the Weak Law of Large Numbers

Figure 1



The WLLN receives a simple graphic interpretation. For each value of *n*, the random variable, Y_n has a probability distribution (that we here assume to be continuous) which is represented by the green curve as shown in the slide. The area under the curve is always 1.

Now, at a position 2 *Epsilon* long segment ***s*** on top of *mu*, denote A_n the area under the curve outside ***s***.
A_n is the probability for Y_n to be different from *mu* by more than *epsilon* (in absolute value).
The WLLN states that for a given epsilon, A_n tends to 0 as *N* grows without limit. In other words, outside of **s**, the "tails" of the distribution of Y_n become negligible when *n* tends to infinity.

4. Differences Between the Weak Law and the Strong Law

The weak law states that for a specified large n, the average Xn bar is likely to be near mu. Thus, it leaves open the possibility that modulus of (Xn bar minus mu) is greater than epsilon happens an infinite number of times, although at infrequent intervals.

The *strong law* shows that this strongly will not occur. In particular, it implies that with probability 1, we have for any epsilon greater than 0 the inequality modulus of (Xn bar minus mu) less than epsilon holds for all large enough *n*.
The proof of the Weak Law is easy when the Xi's have a finite variance. It is most often based on Chebyshev's Inequality.

The Laws of Large Numbers make statements about the convergence of Xn bar to mu. Both laws relate bounds on sample size, accuracy of approximation, and degree of condense. The Weak Laws deal with limits of probabilities involving Xn bar. The Strong Laws deal with probabilities involving limits of Xn bar.

5. Application and Limitation of Weak Law of Large Numbers

Application of Weak Law of Large Numbers
The coin-tossing paradigm is convenient example for the application of WLLN.
Consider a fair coin and for a given *n*, the set of all the possible outcomes of a sequence of *n* tosses of this coin. There are exactly 2 to the power n such sequences, and because of the fairness of the coin, they all have the same probability 2 to the power minus n of occurring when an actual series of *n* tosses is performed.

Suppose these sequences are numbered 1, 2, up to 2 to the power n, in an arbitrary way. For sequence Number i, denote mu i the average number of Heads per toss in the sequence Mu i is equal to 1 by n.

Number of Heads in the sequence take an arbitrary small positive real number epsilon, and count the number of sequences such that mu i departs from half by more than epsilon.

Observe that the proportion of these "deviant" sequences tends to 0 as n grows without limit. Therefore, the WLLN can easily be derived directly in the case of a fair coin tossing.

The interesting things about this derivation are:

- The concept of probability appears only when we state that all possible *n*-sequences of outcomes have the same probability 2 to the power minus n. The rest is just combinatory (as when we derived the probability mass function of the Binomial distribution).

- We never consider infinite sequences of tosses. The WLLN simply says that for a given *epsilon*, for any *n* larger than a certain **n of epsilon**, the proportion of deviant sequences in the (finite) set of (finite length) *n*-sequences is smaller than *epsilon.*

WLLN tells us that that ultimately, the numbers of Heads and Tails must be about equal. After this incredibly unlucky opening sequence, the following tosses must therefore produce mostly Tails for the game to return to a roughly balanced count of Heads and Tails. Hence, we are expecting the following tosses to generate mostly Tails.
In other words, an excess of Heads in an opening sequence must cause an increase of the probability of Tails for the ensuing tosses.

How could the balance be attained otherwise?

Of course, because the tosses are independent, there is no such thing as "probability adjustment". The WLLN does not care whether an opening sequence, however long, is balanced or not. All it says is that all you have to do is toss the coin long enough to increase the probability for the sequence to be just about balanced, but it says nothing about how long you have to wait for this to happen. Many gamblers went to their ruin for misinterpreting the WLLN.

Limitations of the Weak Law of Large Numbers
Mean and variance
The above expressions make an explicit reference to the common **mean** *mu* of the variables.
In addition, the WLLN will appear to be a consequence of Chebyshev inequality, which requires the variables to have a **variance**.
Therefore, the WLLN **seems** to apply only to variables that have at least a mean **and** a variance. In fact, the existence of the variance is **not** required.

The existence of the mean is of course always required. Therefore, for example, the WLLN does not apply to samples drawn from the Cauchy distribution, as this distribution has no mean. We already noticed that the distribution of the mean of a sample drawn from a Cauchy distribution does not depend on the sample size. Therefore, there is no "shrinking" of this distribution as the sample size increases.

Independence
We stated that the WLLN applies to independent or independent and identically distributed random variables. Assuming the independence of variables is so common in Statistics that we sometimes forget how strong a restriction this is.
There is some counter-example that deals with variables that are indeed identically distributed but that are not independent. A consequence of the breakdown of the independence hypothesis will be that the WLLN does not apply to this sequence of variables. It expects the variables to be at least independently distributed.

Why this Law is considered "weak”?
The term "Weak" refers to the way the sample mean converges to the distribution mean. At first sight, it may seem that there is no better way to converge than “convergence in probability".
However, it turns out that the sample mean converges to the distribution mean in a much "stronger" way than just "in probability”. This convergence is almost sure, and the Weak Law of Large Numbers is in fact superseded by a "Strong" Law of Large Numbers.

Here’s a summary of our learning in this session, where we understood:

- The concept of Law of Large Numbers
- The statement of Weak Law of Large Numbers
- The applications and implications of Weak Law of Large Numbers
- The difference between Weak and Strong Law of Large Numbers