The Law of Large Numbers and its Applications

by
Kelly Sedor

A project submitted to the Department of Mathematical Sciences in conformity with the requirements for Math 4301 (Honours Seminar)

Lakehead University
Thunder Bay, Ontario, Canada

Abstract

This honours project discusses the Law of Large Numbers (LLN). The LLN is an extremely intuitive and applicable result in the field of probability and statistics. Essentially, the LLN states that in regards to statistical observations, as the number of trials increase, the sample mean gets increasingly close to the hypothetical mean.

In this project, a brief historical context will be outlined, basic probability theory concepts will be reviewed, the various versions of the theorem will be stated and one of these variations will be proved using two different methods. This project will continue by illustrating numerous situations in which the LLN can be applied in several different scientific and mathematical fields.

Acknowledgements

I would like to thank all of the friends, family and classmates that helped me reach this point. I would also especially like to thank Dr. Deli Li and Dr. Razvan Anisca for the support and encouragement they have given me over the past four years. I am very fortunate to have met so many wonderful people throughout my university career.

Contents

CHAPTER 1

Introduction

1. Historical Background of the Law of Large Numbers

Early in the sixteenth century, Italian mathematician Gerolamo Cardano (1501-1575) observed what would later become known as The Law of Large Numbers. He observed that in statistics the accuracy of observations tended to improve as the number of trials increased. However, he made no known attempt to prove this observation. It was not until over two-hundred years later that this conjecture was formally proved.

In the year 1713, Swiss mathematician Jacob Bernoulli published the first proof for what Cardano had observed centuries earlier. Bernoulli recognized the intuitive nature of the problem as well as its importance and spent twenty years formulating a complicated proof for the case of a binary random variable that was first published posthumously in his book, *Ars Conjectandi*. Bernoulli referred to this as his "Golden Theorem" but it quickly became known as "Bernoulli's Theorem". In his book, Bernoulli details the problem of drawing balls from an urn that contains both black and white balls. He describes trying to estimate the proportion of white balls to black if they are consecutively drawn from the urn and then replaced. Bernoulli states that when estimating the unknown proportion any degree of accuracy can be achieved through an appropriate number of trials. The official name of the theorem; "The Law of Large Numbers", was not coined until the year 1837 by French mathematician Simeon Denis Poisson.

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. Andrey Kolmogorov's Strong Law of Large Numbers which describes the behaviour of the variance of a random variable and Emile Borel's Law of Large Numbers which describes the convergence in probability of the proportion of an event occurring during a given trial, are examples of these variations of Bernoulli's Theorem.

2. Law of Large Numbers Today

In the present day, the Law of Large Numbers remains an important limit theorem that is used in a variety of fields including statistics, probability theory, and areas of economics and insurance. The LLN can be used to optimize sample sizes as well as approximate calculations that could otherwise be troublesome. As we will see throughout our paper, there are numerous situations in which the Law of Large Numbers is effective as well as others in which it is not.

The goal of this project is to introduce the various forms of The Law of Large Numbers as well as answer questions such as, "How powerful is the Law of Large Numbers?", "How can we prove the Law of Large Numbers?" and "How can the Law of Large Numbers be used?"

Our paper is structured as follows. In Chapter 2 we will introduce a groundwork for our theorems by providing key definitions and notation that will be used throughout the project. We will then move on to Chapter 3 which will state the various forms of the Law of Large Numbers. We will focus primarily on the Weak Law of Large Numbers as well as the Strong Law of Large Numbers. We will answer one of the above questions by using several different methods to prove The Weak Law of Large Numbers. In Chapter 4 we will address the last question by exploring a variety of applications for the Law of Large Numbers including approximations of sample sizes, Monte Carlo methods and more. We will conclude the project in Chapter 5 by providing topics for further discussion as well as important implications of the Law of Large Numbers.

CHAPTER 2

Preliminaries

Before we begin our discussion of the 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 chapter will outline the notation that will be seen throughout the remainder of the project. In order to best understand the content of the project, one should have an appropriate grasp on important concepts such as expected value or mean, variance, random variables, probability distributions and more.

1. Definitions

DEFINITION 2.1. A *population* is a group of objects about which inferences can be made.

DEFINITION 2.2. A *sample* is a subset of the population.

DEFINITION 2.3. A *random sample* of size n from the distribution of X is a collection of n independent random variables each with the same distribution as X.

DEFINITION 2.4. An *experiment* is a set of positive outcomes that can be repeated.

DEFINITION 2.5. A *random variable*, X, is a function that assigns to every outcome of an experiment, a real numerical value. If X can assume at most a finite or countably infinite number of possible values, X is said to be a *discrete* random variable. If X can assume any value in some interval, or intervals of real numbers and the probability that it assumes a specific given value is 0, then X is said to be a *continuous* random variable.

EXAMPLE 2.6. Say, for instance we are flipping a fair coin and are concerned with how many times the coin lands on heads. We can define the random variable X by

$$X = \begin{cases} 1, & \text{if the coin lands on heads} \\ 0, & \text{if the coin lands on tails} \end{cases}$$

This is an classic example of a random variable with a *Bernoulli Distribution*, which we will see again later in our discussion.

DEFINITION 2.7. The *probability distribution* of a discrete random variable X, is a function f that assigns a probability to each potential value of X. Additionally, f must satisfy the following necessary and sufficient conditions,

(1) $f(x) = P(X = x)$ for x real,

(2) $f(x) \geq 0$,
(3) $\sum_x f(x) = 1$.

DEFINITION 2.8. The *probability density function* for a continuous random variable X, denoted by $f(x)$, is a function such that

(1) $f(x) \geq 0$, for all x in $\mathbb{R}$,
(2) $\int_{-\infty}^{+\infty} f(x)dx = 1$,
(3) $P(a < X < b) = \int_a^b f(x)dx$ for all $a < b$.

DEFINITION 2.9. Let X be a discrete random variable with density $f(x)$, the *expected value* or *mean*, denoted $E(X)$ or μ, is

$$\mu = E(X) = \sum_x xf(x)$$

provided that $\sum_x xf(x)$ is finite

EXAMPLE 2.10. A drug is used to maintain a steady heart rate in patients who have suffered a mild heart attack. Let X denote the number of heartbeats per minute obtained per patient. Consider the hypothetical density given in the table below. What is the average heart rate obtained by all patients receiving the drug, that is what is $E(X)$?

Table 1.2

x	40	60	68	70	72	80	100
$f(x)$	0.01	0.04	0.05	0.8	0.05	0.04	0.01

$$\begin{aligned} E(X) &= \sum_x xf(x) \\ &= 40(.01) + 60(.04) + 68(.05) + ... + 100(.01) \\ &= 70 \end{aligned}$$

DEFINITION 2.11. Let X be a random variable with mean μ, the *variance* of X denoted $Var(X)$ or σ^2 is

$$\sigma^2 = Var(X) = E((X - \mu)^2) = E(X^2) - \mu^2 = E(X^2) - (E(X))^2.$$

DEFINITION 2.12. The *standard deviation* of a random variable X, denoted σ is the positive square root of the variance.

EXAMPLE 2.13. Let X and $f(x)$ be the same as that seen in Example 2.10, compute the variance and standard deviation for the given data.

$$\begin{aligned} E(X^2) &= \sum x^2 f(x) \\ &= \sum x^2 f(x) \\ &= (40)^2(.01) + (60)^2(.04)... + (80)^2(.04) + (100)^2(.01) = 4296.4 \end{aligned}$$

and by Example 2.10

$$\sigma^2 = E(X^2) - (E(X))^2 = 4296.4 - 70^2 = 26.4$$
$$\sigma = \sqrt{\sigma^2} = \sqrt{26.4} = 5.13809303147 \approx 5.14$$

Definition 2.14. Let $X_1, X_2, X_3, ..., X_n$ be a random sample from the distribution of X. The statistic $\sum_{i=1}^{n} \frac{X_i}{n}$ is called the *sample mean* and is denoted $\bar{X}_n$.

Definition 2.15. Random variables $X_1, X_2, X_3, ...X_n$ are said to be *independent* and *identically distributed* or *i.i.d.* if each random variable X_i has the same probability distribution as X_1, and the occurrence of one does not affect the probability of another.

Definition 2.16. A continuous random variable X is said to follow a *Poisson Distribution* with mean and variance λ if its density f is given by

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

for $x = 0, 1, 2, ...$ and $\lambda > 0$. We write $X \sim P_\lambda$.

Definition 2.17. A random variable Z is said to follow a *standard normal distribution* if it has the following cumulative distribution function

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{\frac{-s^2}{2}} ds.$$

We denote this distribution by $\Phi(z)$ *or* $Z \sim N(0, 1)$ for $-\infty < z < \infty$.

We will see these distributions later in our discussion. Finally, we will define several tools necessary for our proofs and discussion of the Weak Law of Large Numbers.

Definition 2.18. A *Characteristic Function* of a probability distribution is given by

$$\varphi_X(t) = E(e^{itX})$$

where $i = \sqrt{-1}$ and X is the variable with the given distribution.

Theorem 2.19. *(Chebyshev's Inequality) Let X be a random variable with mean μ and standard deviation σ. Then for any positive number k,*

$$P(\mid X - \mu \mid \leq k) \geq 1 - \frac{\sigma^2}{k^2}.$$

Chebyshev's Inequality is an important result in probability and will be especially useful in our proof of the Weak Law of Large Numbers. It is also important for us to know how to apply it.

Example 2.20. Let X be a random variable that follows a Poisson distribution with parameter $\lambda = 7$. Give a lower bound for $P(\mid X - \mu \mid \leq 4)$.

For a Poisson distribution with parameter 7 we have $\mu = \sigma^2 = 7$. Then

$$P(3 \leq X \leq 11) = P(\mid X - 7 \mid \leq 4) = P(\mid X - \mu \mid \leq 4) \geq 1 - \frac{\sigma^2}{4^2} = 1 - \frac{7}{16} = 0.4375.$$

Theorem 2.21. *(The Central Limit Theorem): Let $X_1, X_2, ..., X_n$ be a random sample of size n from a distribution with mean μ and variance σ^2. Then for large n, $\bar{X}$ is approximately normal with mean μ and variance $\frac{\sigma^2}{n}$. Furthermore for large n, the random variable $\frac{(\bar{X}-\mu)}{(\frac{\sigma}{\sqrt{n}})}$ is approximately standard normal.*

The Central Limit Theorem is closely related to the Law of Large numbers, and as such raises similar questions when discussing each theorem. "How large must n be in order to obtain a certain degree of accuracy?" We will address this question later through various examples.

2. Notation

For the majority of our discussion we will be using traditional notation from the field of probability, however we note several things.

Definition 2.22. The probability that a given random variable X takes on a value less than or equal to a particular x is denoted by $P(X \leq x)$.

Definition 2.23. We denote $X \longrightarrow_P k$ to mean that X *converges to k in probability.*

Definition 2.24. We denote $a_n = o(b_n)$ to mean that $\lim_{n\to\infty} \frac{a_n}{b_n} = 0$. (We will see this later in our proof of the Weak Law).

We now have all the necessary background to formally state both the Weak and Strong Law of Large Numbers as well as discuss various proofs for the general form of the Weak Law.

CHAPTER 3

The Law of Large Numbers

1. Theorems and Proofs

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.

THEOREM 3.1. *Let $X_1, X_2, ..., X_n$ be a sequence of i.i.d. random variables, each with mean $E(X_i) = \mu$ and standard deviation σ, we define $\bar{X}_n = \frac{X_1+X_2+...+X_n}{n}$. The Weak Law of Large Numbers (WLLN) states for all $\epsilon > 0$ then,*

$$\lim_{n\to\infty} P(\mid \bar{X}_n - \mu \mid > \epsilon) = 0.$$

PROOF. (Using Chebyshev's Inequality) Assume that $Var(X_i) = \sigma^2$ for all $i < \infty$. Since the $X_1, X_2, ...X_n$ are independent, there is no correlation between them. Thus

$$\begin{aligned}
Var(\bar{X}_n) &= Var\left(\frac{X_1 + X_2 + ... + X_n}{n}\right) \\
&= \frac{1}{n^2} Var(X_1 + X_2 + ... + X_n) \\
&= \frac{1}{n^2}(Var(X_1) + Var(X_2) + ... + Var(X_n)) \\
&= \frac{1}{n^2}(\underbrace{\sigma^2 + \sigma^2 + ... + \sigma^2}_{\text{n items}}) \\
&= \frac{n\sigma^2}{n^2} \\
&= \frac{\sigma^2}{n} \; for \; n > 1.
\end{aligned}$$

We note that the mean of each X_i in the sequence is also the mean of the sample average, i.e. $E(\bar{X}) = \mu$. We can now apply Chebyshev's Inequality on $\bar{X}_n$ to get, for all $\epsilon > 0$,

$$P(\mid \bar{X}_n - \mu \mid > \epsilon) \le \frac{\sigma^2}{n\epsilon^2}$$

$$
\begin{aligned}
P(|\bar{X}_n - \mu| > \epsilon) &= 1 - P(|\bar{X}_n - \mu| \leq \epsilon) \\
&\geq 1 - \left(1 - \frac{\sigma^2}{n\epsilon^2}\right) = \frac{\sigma^2}{n\epsilon^2}
\end{aligned}
$$

which $\longrightarrow 0 \; as \; n \to \infty$

□

Next we will prove the WLLN using characteristic functions.

PROOF. (Using Characteristic Functions) Let $X_1, X_2, ..., X_n, ...$ be a sequence i.i.d. random variables with $E \mid X_1 \mid < \infty$ and $\mu = E(X_1)$. Again, define $\bar{X}_n = \frac{X_1+X_2+...X_n}{n}, n \geq 1$. Let $\varphi(t) = E(e^{itX_1})$ with $t \in (-\infty, \infty)$. Then

$$
\begin{aligned}
\varphi\left(\frac{t}{n}\right) &= E\left(e^{i\frac{t}{n}X_i}\right) \\
&= E\left(1 + i\frac{t}{n}X_1 + o\left(\frac{1}{n}\right)\right) \\
&= E(1) + i\frac{t}{n}E(X_1) + o\left(\frac{1}{n}\right) \\
&= 1 + i\frac{t}{n}\mu + o\left(\frac{1}{n}\right) \text{ as } n \longrightarrow \infty. \qquad (1)
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\varphi_{\bar{X}_n}(t) &= E(e^{it\bar{X}_n}) \\
&= E(e^{it\sum_{j=i}^{n}\frac{X_i}{n}}) \\
&= E(\prod_{j=1}^{n} e^{i\frac{t}{n}X_j}) \\
&= \prod_{j=1}^{n} E(e^{i\frac{t}{n}X_1})(since\ X_1, X_2, ..., X_n\ are\ i.i.d.) \\
&= (\varphi(\frac{t}{n}))^n \\
&= (1 + \frac{it\mu}{n} + o(\frac{1}{n}))^n \quad by\ (1)
\end{aligned}
$$

Now recall that $(1 + \frac{1}{n})^n \longrightarrow e \implies (1 + \frac{a}{n} + o(\frac{1}{n}))^n \longrightarrow e^a$. So we have $\varphi_{\bar{X}_n}(t) \longrightarrow \varphi_D(t) = e^{it\mu}$ where $P(D = \mu) = 1$ and thus $\bar{X}_n \longrightarrow_P \mu$. □

To illustrate this concept we consider several basic examples.

EXAMPLE 3.2. (Die Rolling) Consider n rolls of a die. Let X_i be the outcome of the i^{th} roll. Then $S_n = X_1 + X_2 + ...X_n$ is the sum of the first n rolls. This is an independent Bernoulli trial with $E(X_i) = \frac{7}{2}$. Thus, by the LLN, for any $\epsilon > 0$,

$$P\left(\mid \frac{S_n}{n} - \frac{7}{2} \mid \geq \epsilon \right) \to 0$$

as $n \to \infty$. This can be restated as for any $\epsilon > 0$,

$$P\left(\mid \frac{S_n}{n} - \frac{7}{2} \mid < \epsilon \right) \to 1$$

as $n \to \infty$.

EXAMPLE 3.3. Let $\{X_n; n \geq 1\}$ be a sequence of i.i.d. Poisson random variables with parameter λ then we have

$$P(X_1 = k) = e^{-\lambda}\frac{\lambda^k}{k!},$$

for $k = 0, 1, 2, ...$ Thus $\mu = E(X_1) = \lambda$ and $Var(X_1) = \lambda$. Hence, by the WLLN $\bar{X_n} \longrightarrow_p \mu$.

Let us also consider an example using characteristic functions.

EXAMPLE 3.4. For a Bernoulli Sequence

$$\begin{aligned}
\varphi_n(t) = E(e^{it\bar{X}_n}) &= E(e^{i\frac{t}{n}\sum_{j=1}^{n} X_j}) \\
&= \prod_{j=1}^{n} E(e^{i\frac{t}{n}X_j}) \\
&= E((e^{i\frac{t}{n}X_1}))^n \\
&= (q + pe^{i\frac{t}{n}})^n \ (for\ parameter\ p\ of\ the\ distribution\ and\ q = 1 - p) \\
&= (q + p(1 + \frac{it}{n} + o(\frac{1}{n^2})))^n \\
&= (1 + \frac{ipt}{n} + o(\frac{1}{n^2}))^n \\
&\longrightarrow e^{ipt}\ as\ n \to \infty
\end{aligned}$$

Thus $\frac{S_n}{n} \longrightarrow_P p$.

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

THEOREM 3.5. *(The Strong Law of Large Numbers) Let $X_1, X_2, ..., X_n$ be a sequence of i.i.d. random variables, each with mean $E(X_i) = \mu$ and standard deviation σ then,*

$$P(\lim_{n\to\infty} \bar{X_n} = \mu) = 1.$$

The proof of the Strong Law of Large Numbers (SLLN) is much more complex than that of the WLLN as it is much more powerful. However something should be noted.

Remark 3.6. The proof for the both the Weak and Strong Laws for a continuous random variable are virtually identical to their respective proofs for the case of a discrete random variable.

There are several important variations of the Law of Large Numbers, one important one we will now define is known as Kolmogorov's Strong Law of Large Numbers, which derived its name from the mathematician who discovered it.

Theorem 3.7. *(Kolmogorov's Strong Law of Large Numbers) Let $X_1, X_2, ...X_n, ...$ be a sequence of independent random variables with $\mid E(X_n) \mid< \infty$ for $n \geq 1$. Then the SLLN holds if one of the following are satisfied:*

(1) *The random variables are identically distributed*
(2) $\forall n$, $Var(X_n) < \infty$ *and* $\sum_{n=1}^{\infty} \frac{Var(X_n)}{n^2} < \infty$

Like Bernoulli's SLLN, the proof of Kolmogorov's is very complex and as such will not be covered in our discussion.

2. The Weak Law Vs. The Strong Law

Given that there is both a Weak and Strong Law of Large Numbers, it raises very important questions such as, "How does the Weak Law differ from the Strong Law?" and "Are there situations in which one version of the Law of Large Numbers is better than the other?" We will address these questions in this section.

The Strong Law is considered to be "stronger" because it is equivalent to

$$\lim_{n\to\infty} \mid \frac{S_n}{n} - \mu \mid= 0$$

with probability 1 where as the Weak Law is equivalent to

$$\lim_{n\to\infty} P(\mid \frac{S_n}{n} - \mu \mid< \epsilon) = 1 \ \forall \ \epsilon > 0.$$

Clearly the convergence of the Strong Law is more powerful and as such, if a random variable abides by the SLLN, then the WLLN holds as well, that is, the Strong Law implies the Weak Law. However, the converse is not true, that is, the Weak Law does not imply the Strong Law. The following example will illustrate this very idea.

Example 3.8. Let $\left\{X_n; n \geq 1\right\}$ be a sequence of i.i.d. random variables with common density function given by

$$f(x) = \frac{C_\alpha}{(1 + x^2)ln(e+ \mid x \mid)^\alpha}$$

where $C_\alpha > 0$ such that $\int_{-\infty}^{\infty} f(x)dx = 1$ (e.g. $C_0 = \frac{1}{\pi}$). Then

$$E \mid X_1 \mid = \int_{-\infty}^{\infty} \mid x \mid \frac{C_\alpha}{(1+x^2)(ln(e+ \mid x \mid)^\alpha)} dx$$
$$= 2\int_0^{\infty} \frac{C_\alpha}{(1+x^2)(ln(e+x)^\alpha)} dx$$

$$\longrightarrow \begin{cases} < \infty, & \text{if } \alpha > 1 \\ = \infty, & \text{if } \alpha \leq 1 \end{cases}$$

Thus for $\alpha > 1$, $\frac{S_n}{n} \longrightarrow 0$ almost surely and for $\alpha \leq 1$ $\frac{S_n}{n} \notin SLLN$. Also,

$$nP(\mid X_1 \mid \geq n) = 2nC_\alpha \int_n^{\infty} \frac{dx}{(1+x^2)(ln(e+x))^\alpha}$$
$$\longrightarrow \begin{cases} 0, & \text{if } \alpha > 0 \\ \frac{2}{\pi}, & \text{if } \alpha = 0 \\ \infty, & \text{if } \alpha < 0 \end{cases}$$

and $E(X(\mid X \mid \leq n)) = 0$ for all n. Thus for $\alpha > 0$ we get $\frac{S_n}{n} \to_p 0$ and for $\alpha \leq 0$ we get $\frac{S_n}{n} \notin WLLN$ as the mean goes to ∞. Thus, we conclude that for this example for $0 < \alpha \leq 1$ we have $\frac{S_n}{n} \in WLLN$ but $\frac{S_n}{n} \notin SLLN$.

Although this is not always the case, there are more situations in which the Weak Law applies and the Strong Law does not. For example, if a function has a standard normal distribution, its density is given by $\Phi(z) = P(Z \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{\frac{-s^2}{2}} ds$. The WLLN holds for this function but the SLLN does not as $\frac{S_n}{n} \to 0$ in probability.

There also exist situations in which neither the Weak Law nor the Strong Law hold for a random variable. We will conclude this chapter with an example of such an occasion.

Example 3.9. Let X be a sequence of i..i.d. random variables with common density function f given by

$$f(n) = P(X = -n) = P(X = n) = \frac{3}{\pi^2 n^2}.$$

Let $\{X_n, n \geq 1\}$ be a sequence of i.i.d. random variables with common density f, then

$$E \mid X \mid = \sum_{n=1}^{\infty} n \frac{6}{\pi^2 n^2} = \frac{6}{\pi^2} \sum_{n=1}^{\infty} \frac{1}{n} = \infty.$$

Thus $E \mid X \mid$ does not exist and therefore this sequence of i.i.d. random variables $\frac{S_n}{n}$ does not abide by SLLN as it is required that $E \mid X \mid < \infty$. In general, for the i.i.d. case $\frac{S_n}{n}$ obeys the SLLN if and only if $E \mid X \mid < \infty$ and if $\frac{S_n}{n} \longrightarrow E(X)$ almost surely.

CHAPTER 4

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 will explore how to apply the Law of Large Numbers by means of general examples as well as discussing Monte Carlo Methods.

1. General Examples

To begin to think about how to apply the Law of Large Numbers let us consider a very basic example.

EXAMPLE 4.1. 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 LLN 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.

Now to continue our discussion of how to use the LLN we will work through a problem and then repeat the problem using the Central Limit Theorem and compare our answers.

EXAMPLE 4.2. Suppose someone gives you a coin and claims that this coin is biased; that it lands on heads only 48% of the time. You decide to test the coin for yourself. If you want to be 95% confident that this coin is indeed biased, how many times must you flip the coin?

Solution 1 (Using WLLN): Let X be the random variable such that $X = 1$ if the coin lands on heads and $X = 0$ for tails.

Thus $\mu = 0.48 = p$ and $\sigma^2 = p(1-p) = 0.48 \times 0.52 = 0.2496$. To test the coin we flip it n times and allow for a 2% error of precision, i.e. $\epsilon = 0.02$. This means we are testing the probability of the coin landing on heads being between $(0.46, 0.50)$. By the Law of Large Numbers, we want n such that

$$P[|\bar{X} - 0.48| > 0.02] \leq \frac{0.2496}{n(0.02)^2}$$

So for a 95% confidence interval we need

$$\frac{0.2496}{n(0.02)^2} = 0.05$$

Thus n should be

$$0.2496 \times 2500 \times 20 = 12,480.$$

That is quite a lot of coin flips! Now let's repeat the question using an alternative method.

Solution 2 (Using the Central Limit Theorem):

$$\begin{aligned}
P(\frac{S_n}{n}, 0.50) &= P\left(\frac{S_n - 0.48n}{n} < 0.02\right) \\
&= P\left(\frac{S_n - 0.48n}{\sqrt{n0.48 \times 0.52}} < \frac{0.02\sqrt{n}}{\sqrt{0.48 \times 0.52}}\right) \\
&\geq P\left(\frac{S_n - 0.48n}{\sqrt{0.2496n}} \leq 0.04\sqrt{n}\right) \\
&\approx \Phi(0.04\sqrt{n}) \geq 0.95
\end{aligned}$$

So we have,

$$\begin{aligned}
0.04\sqrt{n} &= 1.645 \\
&\implies \sqrt{n} \approx \frac{1.645}{0.04} = 41.125 \\
&\implies n \approx 1691.2625 \\
&\implies n = 1692.
\end{aligned}$$

Thus, the coin must be flipped 1692 times. As we can see, the Weak Law is not as powerful or accurate as the Central Limit Theorem however can still be used to a certain degree of accuracy. The next example demonstrates a common use of the WLLN: estimating error.

Example 4.3. A survey of 1500 people is conducted to determine whether they prefer Pepsi or Coke. The results show that 27% of people prefer Coke while the remaining 73% favour Pepsi. Estimate the Margin of error in the poll with a confidence of 90%.

Solution: Let

$$X_n = \begin{cases} 1 & \text{if } n^{th} \text{ person is in favour of Coke} \\ 0 & \text{otherwise} \end{cases}$$

for $n = 1, 2, ..., 1500$. Then $X_1, X_2, ...X_{1500}$ are i.i.d. Bernoulli random variables and $\hat{p} = \frac{S_{1500}}{1500} = 0.27$. So we let X_n be the Bernoulli random variable as described above with $P(X_1 = 1) = 0.27$. Then

$$\mu = E(X_1) = 0.27 \text{ and } \sigma^2 = 0.27 \times 0.73 = 0.1971.$$

Then by the LLN with $n = 1500$ we have

$$P\left(\left|\frac{X_1 + ... + X_{1500}}{1500} - 0.27\right| \geq \epsilon\right) \leq \frac{0.1971}{1500\epsilon^2}, \ \epsilon > 0.$$

So if we set $\frac{1}{10} = \frac{0.1971}{1500\epsilon^2}$, we get

$$\epsilon = \sqrt{\frac{0.1971 \times 10}{1500}} = 0.036$$

Thus we have that the margin of error is less than 4% with 90% confidence.

The Law of Large Numbers is heavily used in fields of finance and insurance to assess risks as well as predict economic behaviour. We illustrate this with in a very simplistic situation.

Example 4.4. Two new companies are being assessed. Company A has a total market value of \$60 Million and company B has a total market value of \$4 Million. Which company is more likely to increase its total market value by 50% within the next few years?

Solution: In order for Company A to increase its total market value by 50% it requires an increase of \$30 million whereas company B only requires a \$2 million increase. The Law of Large Numbers implies that it is much more likely for company B to expand by 50% than company A. This makes sense because if company A was just as likely to expand by 50% as company B, company A could quickly have a higher market value than the entire economy.

Similarly, insurance providers use the LLN in order to determine effective prices for their clients. By analyzing and recording the number accidents among men aged 23-25, they can ascertain with a high degree of accuracy, the probability of X amount of males aged 23 that will be the cause of an accident in any given year. This allows them to set an effective price for clients that fall into the range.

We may also consider an example of how to apply the LLN to gambling situations, where probabilities are extremely important

Example 4.5. A 1-dollar bet on craps has an expected winning of $E(X) = -.0141$. What does the Law of Large Numbers say about your winnings if you make a large number of 1-dollar bets at the craps table? Does the Law of Large Numbers guarantee your loses will be negligible? Does the Law of Large Numbers guarantee you will lose if n is large?

Solution: By the LLN your losses will not be small, they will average to 0.141 per game which would imply that your losses will, on average, become large for a large number of games. The Law of Large Numbers also does not guarantee you will lose for a large n. It says that for a large number n, it is very unlikely that you will lose. However for a fixed n finite, no assumption can be made whether you will win or lose.

We will continue our discussion of applications of the Law of Large Numbers by introducing an important method of approximation, namely Monte Carlo Methods.

2. Monte Carlo Methods

2.1. Introduction. The Monte Carlo Methods are a series of algorithms that use random sampling of numbers and data in order to approximate numerical results that would otherwise be difficult to achieve explicitly. Originally invented in the 1940's by Stanislaw Ulam, the Monte Carlo Methods have become an essential tool in many fields including: engineering, biology, computer graphics, applied statistics, finance and more. In this section we will outline the algorithm used in Monte Carlo methods for mathematical problems as well as several examples on how to apply them to integration. It is important to note, that in many instances, simulation software is needed to obtain the random sampling and complete the repeated calculations.

2.2. Algorithm for Integrals. The algorithm used in Monte Carlo Methods changes depending on the problem. For the sake of our discussion we will outline the procedure used when computing complicated integrals of a function $f(x)$

(1) Simulate uniform random variables $X_1, X_2, ...X_n$ on an interval $[a, b]$: This can be done with software or a statistical random number table. Using a random number table gives $U_1, U_2, ...U_n$ i.i.d. random variables on $[0, 1]$. We then let $X_i = a + (b - a)U_i \ for \ i = 1, 2, ..., n$. Then $X_1, X_2, ..., X_n$ are i.i.d. uniform random variables on $[a, b]$.
(2) Evaluate $f(X_1), f(X_2), ...f(X_n)$
(3) Average values by computing $(b-a)\frac{f(X_1)+f(X_2)+...+f(X_n)}{n}$ and then by the SLLN this converges to

$$(b-a)E(f(x_1)) = (b-a)\int_a^b f(x)\frac{1}{(b-a)}dx$$
$$= \int_a^b f(x)dx.$$

We will now apply this procedure to several integrals and introduce an algorithm to compute π.

2.3. Examples.

Example 4.6. Evaluate the integral where $g(x) = cos^2(x)\sqrt{x^3+1}$ on $[-1, 2]$. So we have,

$$\int_{-1}^{2} cos^2(x)\sqrt{x^3+1}\ dx$$

Clearly, this integral is quite difficult to compute by traditional means. However using the algorithm above as well as computational software we see

$$\int_{-1}^{2} g(x) = 0.905 \ for \ n = 25$$

and

$$\int_{-1}^{2} g(x) = 1.028 \ for \ n = 250$$

so by the Law of Large Numbers, as n increases our approximation becomes increasingly accurate. A more direct method of integration reveals that in fact

$$\int_{-1}^{2} cos^2(x)\sqrt{x^3+1} \ dx = 1.000194.$$

The next example cannot be evaluated directly and requires a numerical approximation to solve it.

Example 4.7.

$$\int_{0}^{1} \frac{e^{sin(x^3)}}{3(1+5x^8)}$$

This integral requires a slightly different method. Letting A denote the area under the function. We construct a random variable X such that $E(X) = A$ we proceed with the following algorithm

(1) Generate two random numbers S_1 and T_1 both in $[0,1]$. This is equivalent to generating a random point in the square $[0,1] \times [0,1]$.
(2) If $T_1 \leq f(S_1)$ set $X_1 = 1$ and if $T_1 > f(S_1)$ set $X_1 = 0$
(3) Repeat the previous steps to generate $X_2, X_3, ...X_n$.

Thus we have

$$P(X=1) = P(T \leq f(S)) = \frac{Area \ under \ the \ graph \ of \ f}{Area \ of \ [0,1] \times [0,1]} = A = \int_{0}^{1} f(x)$$

Following this procedure we are able to achieve an answer of $A = 0.3218$

It is easy to see that this is a very important application of the LLN. Integration is a fundamental tool used in a variety of fields of science and mathematics. To conclude this chapter of how the Law of Large Numbers can be used we consider the very interesting example of using Monte Carlo Methods and the LLN to approximate π.

2.4. Algorithm to Compute π. Similar to Example 4.7 a slightly different algorithm is required. As above we construct a square with side lengths 1. Additionally, we construct a circle of radius $\frac{1}{2}$ that is contained in the rectangle. Thus the square has area 1 and the circle has area $\frac{\pi}{4}$. Now to approximate π we generate a random point in the square. If the point is also in the circle, the point is accepted and if the point is outside the circle, it is rejected. By the LLN the amount of accepted numbers should approach $\frac{\pi}{4}$. The generalized algorithm is as follows:

(1) Generate two random numbers S_1, T_1 both in $[0,1]$ again, this is the same as generating a random point in $[0,1] \times [0,1]$

(2) If $S_1^2 + T_1^2 \leq 1$ then set $X_1 = 1$. If $S_1^2 + T_1^2 > 1$ set $X_1 = 0$
(3) Repeat the previous steps to generate $X_2, X_3, ..., X_n$

So we have,

$$P(X_1 = 1) = P(S_1^2 + T_1^2 \leq 1) = \frac{Area\ of\ the\ circle}{Area\ of\ the\ square} = \frac{\frac{\pi}{4}}{1} = \frac{\pi}{4}$$

and $P(X = 0) = 1 - \frac{\pi}{4}$. Also $E(X) = \mu = \frac{\pi}{4}$ and $Var(X) = \sigma^2 = 1 - \frac{\pi}{4}$.
Thus by the Law of Large Numbers and applying Chebyshev's Inequality we have

$$P\left(\left|\frac{X_1 + X_2 + ... + X_n}{n} - \frac{\pi}{4}\right| \geq \epsilon\right) \leq \frac{\frac{\pi}{4}(1 - \frac{\pi}{4})}{n\epsilon^2}, \epsilon > 0.$$

Now say we wish to compute π within $\pm\frac{1}{1000}$ meaning $\epsilon = \frac{1}{1000}$. This is equivalent to computing $\frac{\pi}{4}$ to within an accuracy of $\frac{1}{4000}$. Here we note, that although we do not know the value of $\frac{\pi}{4}(1 - \frac{\pi}{4})$ we can easily see that the equation $p(1-p)$ achieves its maximum of $\frac{1}{4}$ on$[0, 1]$ when $p = \frac{1}{2}$. So we have

$$P\left(\left|\frac{X_1 + X_2 + ... + X_n}{n} - \frac{\pi}{4}\right| \geq \frac{4}{100}\right) \leq \frac{4,000,000}{n},$$

meaning we need to compute the algorithm 80,000,000 times to make this probability 0.02.

CHAPTER 5

Further Discussion

The Law of Large Numbers is an extremely useful result with powerful implications that can be applied to a variety of disciplines including probability, statistics, finance and physical sciences. In this paper, we stated both the Weak and Strong Law of Large Numbers and proved the Weak Law using two separate methods. We continued by comparing the WLLN to the SLLN and discussing how to apply the LLN in instances of estimating sample sizes, estimating error and using the LLN and Monte Carlo Methods to approximate integrals.

The applications of the LLN 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 LLN. 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 LLN 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 LLN implies that this probability increases significantly the larger the length of the message. For example, the probability that the word "Math" appears in David Foster Wallace's magnum opus *Infinite Jest*, which contains approximately 484,000 words, is much higher than the probability that the same word appears in Haruki Murakami's 2013 novel *Colorless Tsukuru Tazaki and His Years of Pilgrimage*, which contains less than 50,000 words.

In many of our examples we discovered that the distribution of a given function converges in probability to a given number. This prompts an interesting question "At what rate does a given distribution converge to its expected value?" This is an interesting problem regarding the LLN to consider. This convergence rate could have some very important implications about the distribution of the function. What factors affect this convergence rate? Could these factors be modified to ensure a faster convergence? An additional thing to take note of is in Example 4.2 we found how many times a biased coin must be flipped in order to verify with high probability that the coin is indeed biased. We solved this problem using the LLN and using the Central Limit Theorem and got dramatically different results. This also raises a very important question, "Is there a way that the Law of Large Numbers can be used to find an *optimal* sample size?". Repeating trials can be very expensive, using both time and resources. As such, business would benefit very much from a smaller number of trials when testing such things.

Finally, we mention how to see the LLN in action. This could be as easy as flipping a coin! Additionally, there are many simulation methods that anyone can do. There are a variety of online websites and applications that simulate random events such as coin flips, dice rolls and much more and plot the results. As an aspiring educator, this could be an excellent demonstration of probability for young students to experience for themselves. Such electronic aids would allow students to discover on their own what happens when they modify the number of trials, giving the students a basic introduction to probability theory.

Bibliography

[1] S. Milton, J. Arnold, Introduction to Probability and Statistics: Principles and Applications for Engineering and the Computing Sciences, Fourth Edition (2002).

[2] C. Grinstead, L. Snell, Introduction to Probability, Second Edition (2003) 305–316.

[3] D. Li, Limit Theorems and the Laws of Large Numbers: The True Essence of Probability (2011).

[4] S. Pillai, Lecture 13 The Weak Law and the Strong Law of Large Numbers (2001).

[5] University of Massachusetts Amherst, The Law of Large Numbers and the Monte Carlo Method (2014). Web. `http://people.math.umass.edu/~lr7q/ps_files/teaching/math456/lecture17.pdf`.

[6] J. Watkins, Introduction to the Science of Statistics (2009) 153–165.