Spring 2022

A NEW APPLICATION OF THE CENTRAL LIMIT THEOREM

Kenneth Winters
Southeastern University

A NEW APPLICATION OF THE CENTRAL LIMIT THEOREM

by

Kenneth Winters

Submitted to the School of Honors Committee

in partial fulfillment

of the requirements for University Honors Scholars

Southeastern University

2021

Dedication

This thesis is dedicated to my family and my girlfriend. I probably would not have gotten this far in my academic career if it was not for the support of my family and girlfriend. I would like to thank my mother in particular for homeschooling me all the way through high school and continuing to push me to grow as an academic and a person well into my collegiate career.

Acknowledgments

I want to thank Dr. Berhane Ghaim for his tremendous help, not only with my thesis, but throughout my whole undergraduate career. He encouraged me when I needed it and was always there to answer my questions. I would not have had the boldness to find or attempt to prove an original theorem without Dr. Ghaim. I would also like to thank Mr. Cody Tessler for giving me my first rigorous introduction to probability. I do not think I would have been able to complete this thesis or enjoy probability theory as much as I do without Mr. Tessler.

Abstract

This paper discusses the Central Limit Theorem (CLT) and its applications. The paper gives an introduction to what the CLT is and how it can be applied to real life. Additionally, the paper gives a conceptual understanding of the theorem through various examples and visuals. The paper discusses the applications of the CLT in fields such as computer science, psychology, and political science. The author then suggests a new mathematical theorem as an application of the CLT and provides a proof of the theorem. The new theorem relates to expected value and probabilities of random variables and provides a link between the two using the CLT.

KEY WORDS: (Central Limit Theorem, Expected Value, Probability Theory, Random Variable, Mathematical Statistics)

Table of Contents

Introduction

Probability and Statistics is a fairly new development in the world of mathematics, being only a few centuries old, however, it has already made a huge impact on the daily lives of nearly everyone around the world. Probability and Statistics is arguably one of the most practical fields of mathematics and at the highlight of statistics is the Central Limit Theorem (CLT). The CLT is a theorem relating many distributions of all different kinds back to one, easily understood and symmetric distribution. As discussed later in the paper, the applications of the CLT are far reaching. We see the CLT appearing in fields such as psychology, computer science, and political science. This paper serves a few purposes. First, we would like to give a thorough introduction to the CLT through technical definitions, examples and visuals. Second, we want to discuss a few of the many applications of the CLT in an attempt to exemplify the importance and use of the theorem today as well as pointing out where the theorem still needs some work and more research. Finally, we offer a new application of the CLT by suggesting and proving an original theorem. This theorem is a new one in the field of probability and statistics. While the theorem presented in this paper is a fairly intuitive one, it may be able to provide some interesting insight and applications in the field of statistics but also in real life. The theorem we offer discusses expected value and probabilities of random variables and more specifically, how the two of those things relate to each other. The theorem is proven using the CLT. In addition to serving as a mathematical application of the CLT, my new theorem will also provide several practical applications. The most easily understood application of this theorem is in the field of business. It can be a practical theorem for companies who engage in targeted advertising. We will also discuss limitations of my theorem and what it does versus what it does not mean. It is my goal that this paper will inform readers and provide them with a new understanding and appreciation of the CLT and its applications. Additionally, I would like for my theorem to serve as a tool to be used for further theorems and applications in various fields beyond that which is presented in this paper.

Literature Review

Although the Central Limit Theorem may not impress the average high school statistics class, it has certainly been noticed by mathematicians and impacts numerous fields of study and work. The CLT is one of the most important theorems in all of statistics and probability. We will attempt to give a thorough definition and explanation of the CLT in this literature review as well as provide many examples of how the theorem can be applied to real life. Before discussing the CLT, it is important to note and define the Normal Distribution. The distribution is called Normal because it appears in many seemingly random places in nature. The less colloquial name for the distribution is the Gaussian Distribution. The Probability Density Function (PDF) for the Normal Distribution is defined as the following:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} - \infty < x < \infty$$

Where μ represents the population mean and σ is population standard deviation. The Normal Distribution is often described as a bell-shaped curve because of the smooth bell-like shape it holds.

The CLT can be traced back to French mathematician, Abraham de Moivre (Dassau 135). De Moivre noticed a pattern when working with Bernoulli distributions. For the reader who may not be familiar a Bernoulli distribution, we can think of a Bernoulli trial as something like a coin flip. We define the “flip” to be either a success or failure with a certain probability of success p. In the case of a fair coin, our probability of success is .5 and our probability of failure is also .5. When we take many Bernoulli trials, say 10 coin flips, we get what is known as a Binomial distribution. De Moivre was working with Binomial distributions when he noticed that the trials tended toward Normal. In fact, the more trials he used, the closer the distribution tended toward the Normal distribution. He determined that as sample size n got large, the Binomial distribution could be estimated by the Normal distribution.

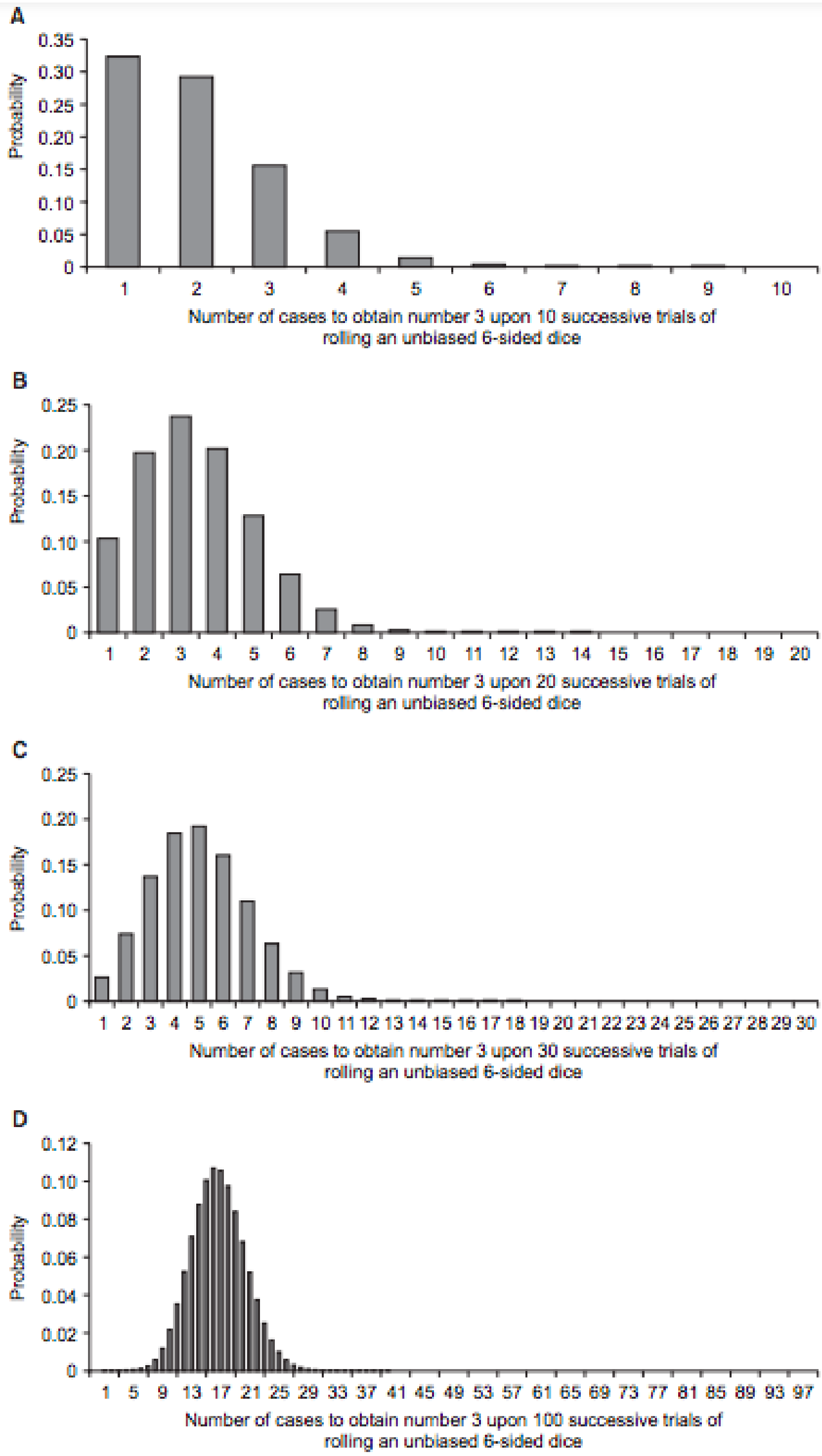

Fig. 2. The probability density function of a binomial distribution with a probability parameter of $\frac{1}{6}$ (i.e., the probability of rolling the number 3 in each trial), based on to the number of trials.

A visual representation of the Binomial distribution converging to the Gaussian distribution can be seen in this example by Kwak (147). In the figure above, Kwak shows how the PMF of a Binomial distribution gradually looks more like the bell-shaped curve of a Gaussian distribution as n gets large. This phenomenon, as it turns out, is not unique to the Binomial distribution. This fact is true of many different types of distributions. The specifications under which a distribution would converge to Gaussian were largely unknown until 1922 when Jarl Lindeberg presented his version of the CLT (Le Cam 79). While his CLT went through a few variations and tweaks, the most common version of the CLT and the one to which we will refer throughout this paper can be summarized by the following:

Let $X_1, \ldots, X_n$ be identically independent random variables with mean 0 and variance of 1 and let $S = \sum_{i=1}^{n} X_i$. Then,

$$\text{the} \lim_{n\to\infty} P(S \le x) = \phi(x)$$

where we define $\phi(x)$ to be the cumulative normalized Gaussian distribution

$$\phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

As previously stated, this is not the only form of the CLT. Many improvements have been made so that fewer stipulations apply, such as not always needing identical distributions. This, however, is a good enough definition for our purposes. Though this definition is relatively math heavy, all it means is that independent and identical distributions tend toward the Normal distribution. The larger the sample size, the closer the distributions become to a Normal distribution, and, in fact, as our sample size becomes infinitely large, the distribution is exactly the Normal distribution.

Most distributions follow the CLT, but not every one does. Lindeberg found a way to show which distributions qualify for the CLT. It is called the Lindeberg Condition. Goldstein, a PhD mathematician, who is well researched in mathematical statistics and has written many papers on the Central Limit Theorem, provides a simplified version of Lindeberg Condition in page 5 of his paper. The Lindeberg Condition essentially reads:

$$\forall\, \epsilon > 0 \ \lim_{n\to\infty} L_{n,\epsilon} = 0 \ \text{ Where } L_{n,\epsilon} = \sum_{i=1}^{n} E\{X_{i,n}^2 \mathbf{1} \left|X_{i,n}\right| \geq \epsilon\}$$

Goldstein explains, essentially, the variances of each distribution need to be small in comparison to the variance of the sum as a whole. While “small” is a very vague and somewhat subjective term, the Lindeberg Condition shown above provides exact specifications under which a distribution converges to Normal.

Mathematicians have been able to prove the Central Limit Theorem a variety of ways. Many of these proofs are difficult to understand and utilize high level mathematics, but one of the more interesting methods of proof is by using moments of a PMF (Probability Mass Function). We will not include an entire proof of the CLT by any method, but we will give a couple brief summaries. The proof by moments utilizes the MGFs (Moment Generating Functions) of certain random variables. The MGF of a random variable is a certain function that tells us information about the mathematical moments of the random variable. It can be shown that the MGF of the sum of any independent and identically distributed random variables matches exactly with the MGF of a Normal distribution as n approaches infinity. This, therein, shows that the sum of random variables gets closer and closer to Normal as n gets larger. While this proof is nice because it is relatively simple and easy to understand, it is also somewhat limited. It only works when the random variables are independent and identically distributed. Additionally, proof by moments does not really tell us anything about how quickly a distribution converges to Normal. A more recent method of proving the Central Limit Theorem is known as Stein's Method (Rinott and Rotar 15). While this method is more difficult to understand than the one using MGFs, it offers advantages the MGFs cannot. One of the major appeals of Stein's Method is that it allows for the CLT to occur for dependent random variables under specific conditions (Goldstein and Reinert 935). Additionally, Stein's Method can be used to show the CLT to be true for non-identically distributed random variables. Stein's Method can be shown via a process known as Zero Bias Transformation. Stein's method and Zero Bias Transformation have been used to prove the CLT, but they also have applications in many related areas and have greatly impacted the fields of probability and mathematical statistics in general.

While the CLT is interesting enough on its own to be studied simply as a mathematical theorem, its main significance comes from its versatility. The CLT applies to many real-life situations. It would be a disservice to mathematical statistics if we did not discuss the many uses of the CLT. We will start with one of the oldest: gambling. It may come as a surprise to some, but probability theory was created for the use of gamblers. Gamblers wanted to know likelihoods of various scenarios in gambling games, and therefore, sought the help of mathematicians. Now, hundreds of years later, casinos still use the fundamentals of probability theory to ensure their large profits. Dr. Singh, an expert in mathematical game optimization, writes about the use of the CLT in casinos. In his paper, he utilizes the CLT to show how profits can be maximized. He specifically studies slot machines and the game Baccarat and uses the CLT to show intriguing results on how casinos make money.

Of course, calling poker an application is a bit of a stretch, as it is highly theoretical and the situations it presents would rarely occur naturally in everyday life. So, the reader would be justified in asking if there are any more applicable uses of the CLT. The answer is a resounding yes.

Many people are familiar with 20th century mathematician Alan Turing. He was most famously known as an Englishman who worked on breaking German codes in World War II. What many people do not know about him, is that while he was attending college, he was largely invested in the CLT. He researched and provided a few proofs for the CLT. So, later when he worked as a code breaker, he used high level mathematical statistics (Zabell 492). While it is unknown whether the CLT was specifically used to help Turing break codes during WWII, I think it is safe to say his research on the CLT helped influence him and hone his statistical skills needed to be a code breaker. Even Turing did not specifically use the CLT for code breaking, it is known that the CLT has a huge role in computer science nowadays.

Random numbers are extraordinarily important for many computer programs. Unfortunately, generating random numbers on a computer is rather difficult. A common way of creating random numbers on a computer program is through something called a Gaussian Random Number Generator (GRNG). GRNGs provide random numbers utilizing Gaussian distributions through various methods. In his paper about GRNGs, Dr. Thomas, a professor of electrical engineering, explains different ways that GRNGs can be created. He provides a thorough explanation of the nature of GRNGs and how they can be produced. The different ways GRNG can be programmed are not really the point of this paper, so we will not get into them; however, it is important to note how fundamental GRNGs are to the world of programming. GRNGs can often have high hardware costs to run, though. So, an efficient way to maintain the efficacy of GRNGs and keep hardware costs down is through the use of the CLT. This, of course, is because nearly any distribution can be transformed into a Gaussian one through the use of the CLT. So, hardware costs are kept much lower by using Uniform distributions as opposed to Gaussian ones, and they are transformed to Gaussian ones because of the CLT (Malik 60).

If we take a step out of the world of computer science and into the realm of psychology, we will find more use for our CLT. Psychology is a statistically heavy field. Psychologists rely on statistical concepts and have studied statistics, to some degree, in their training. It should not be surprising, then, that the CLT appears within the field of psychology. Mathematical psychologist Dr. Massaro wrote an interesting psychological paper about information theory. The paper is littered with mathematical statistics, including a use of the CLT (610). This paper is one specific case where the CLT is employed; however, due to the ever-present nature of statistics within the field of psychology, I suspect the CLT is a familial theorem to many calculations within psychology.

The final example we will discuss is one most readers will recognize: political science. Adults in the United States that have been around for one or more presidential elections are familiar with pollsters making predictions about the election. So, it is unlikely to come as a shock that the process of polling and making predictions involves the CLT. In recent years, pollsters received quite a bit of flack for their inaccuracies, so before talking about the CLT, we will address these errors, so the readers do not doubt the soundness of the CLT. Most of the issues are not due to mathematical miscalculations, but rather polling methods, statistical survey methods, and somehow introducing a bias into the equations such that the results of the mathematics do not reflect reality. Political polling, though, is a fairly intricate and statistically complex process. Stanford professor of political science and statistics, Dr. Jackman, discusses the details of political polling and suggests pooling several political polls to create a more accurate poll. In his paper, he mentions how the CLT is beneficial when pooling the political polls and how it can help estimate the error of such a pool (501).

It should be fairly obvious by now why the CLT is an important theorem that can be applied to many areas of life. It would be deceptive, however, if we did not mention the failures of the CLT. The CLT is often misused in the field of actuarial science. Actuarial science is one of the biggest applications of mathematical statistics in our current world. Strangely enough, the CLT is often used in actuarial science when, indeed, it does not apply. Researcher of actuarial science and risk management, Dr. Brockett, outlines the exact details of when this occurs and why it is wrong. In summary, many actuaries assume that specific losses can be estimated by the CLT when n is large enough. Dr. Brockett, however, explains why this is not the case. Though the CLT is effective at estimating many different distributions, this is not an applicable case (2).

My final note is less about the shortcomings of the CLT and more of a description of when the CLT can be used and how it can still be improved. In his paper "Bounds on the Constant of the Mean Central Limit Theorem," Dr. Goldstein discusses the Berry-Esseen Theorem. This theorem is helpful in estimating how quickly a distribution converges to Normal. Although some distributions can be estimated by the Normal distribution, they may not be useful estimations if they converge to Normal very slowly. In this, Goldstein shows how the Berry-Esseen Theorem is helpful with knowing when the estimation is a good one (1672). Recently, much of the work on the CLT has to do with dependent distributions. Mathematicians are still not entirely sure how and when dependent distributions tend toward Normal. Stein came closer to solving this problem when he created Stein's Method, but more work surrounding a dependent CLT is still needed.

The CLT is truly an incredible theorem. What started out as a pattern of Binomial distributions has blossomed into one of the most powerful theorems in all of mathematical statistics. The CLT has been refined through the years by Lindeberg and his condition, followed by Stein's method. Not only has the actual mathematics improved, but many applications have been found for this theorem to impact people in their everyday lives. From gambling to psychology to computer science, this theorem works its way into nearly any field. One exception is actuarial science. The theorem is not quite a finished issue. Work is still required to discover how distributions tend toward Normal if they are dependent in addition to other ways this theorem can be applied to everyday life.

Methodology

In the thesis so far, we have discussed the history, definition and applications of the Central Limit Theorem (CLT). For my original work, I plan on proving a new theorem that is an application of the CLT. I plan on proving this by showing the theorem to be true for a normally distributed random variable and then working backward from the CLT to show that it applies generally to independently and identically distributed (iid) random variables. After proving my theorem, we will discuss why this new theorem is important and how it will have applications in real life. Additionally, we will discuss what my theorem does and does not mean. In order to prove the theorem, I have researched and provided sources to any definitions or prerequisite theorems that are used in my paper. My sources for definitions and theorems will be found mainly in probability and statistics and mathematics textbooks. Additional sources are found on JSTOR and ProQuest online databases for mathematical sciences.

Proof of Theorem

We have previously discussed the CLT and some of its applications, but I would like to propose a new theorem to serve as an application of the CLT. We will state the theorem, then prove it and discuss what this new theorem means and how it can be applied to life. The theorem is the following:

Let X and Y be two identically and independently distributed (iid) random variables with finite variances where $E(X) > E(Y)$. Then, $\Pr(X > Y) > 0.5$.

I would like to make a note about this theorem. When we say that X and Y are iid random variables, we are working under the assumption that the variable converges to Normal when summed over many times. This is not necessarily true of all iid random variables, but it is true for almost all of them and it is true for all the intents of this paper. So, when we refer to an iid in this paper now on, know that we are referring to an iid random variable that converges to Normal when summed over many times. Before we can prove this theorem, we need to provide a couple technical definitions and explanations. Firstly, we will define what expected value is. We define expected value as the weighted average of a certain random variable. In our theorem, we will only be working with continuous random variables. We define expected value of a random variable X to be the following: $E(X) = \int xf(x)dx \;\; \forall x \varepsilon \mathbb{R}$

We will use a theorem provided in Casella and Berger, in order to help define what it means for a random variable to be independently and identically distributed. Specifically, we will use their theorem to show how we can understand identical distribution. Their theorem says, that the following two statements are equivalent.

a. The random variables X and Y are identically distributed.

b. $F_X(x) = F_Y(x)$ for every x.

What this essentially means is that the two distributions have the same probability at every x value in their domain. It is generally agreed that part b of the preceding theorem is the requirement for variables X and Y to be considered equal random variables. Therefore, saying that two distributions are identically distributed and that they are equal are equivalent statements. This will be important for us later on in our proof. We will also define two random variables X and Y to be independent if $\Pr(X = x, Y = y) = \Pr(X = x) * \Pr(Y = y)$.

To prove our theorem, we will start by first examining a collection of random variables that occur many times. Define the collection of random variables X and Y as the following:

$X = \lim_{n\to\infty} x_1 + x_2 + \cdots + x_n$ and $Y = \lim_{n\to\infty} y_1 + y_2 + \cdots + y_n$

By the CLT as stated earlier in the paper, X and Y approach a Normal distribution. It is important that we note a difference here in the expected value and variance. In our originally stated version of the CLT, our iid random variables had a mean of 0 and variance of 1. This is not necessary for the random variables to converge to Normal, but it does change the mean and variance of the convergent distribution. We can again appeal to Casella and Berger and use a portion of their theorem 2.2.5 to show that:

$E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \ldots + E(X_n)$.

In other words, the expected value of a sum is the sum of the expected values. Since we are given that each $E(X_i) > E(Y_i)$, this implies $E(X) = \lim_{n\to\infty} E(X_1) + E(X_2) + \cdots + E(X_n) >$

$\lim_{n\to\infty} E(Y_1) + E(Y_2) + \cdots + E(Y_n) = E(Y)$

So, we can say that $E(X) > E(Y)$. Now, we want to know the $\Pr(X > Y)$. This, of course, is the same as $\Pr(X - Y > 0)$. Let's define a new random variable $Z = X - Y$. Since X and Y are

Normally distributed, $X - Y$ is also Normally distributed. We can easily find the mean of Z by using the above theorem. $E(Z) = E(X - Y) = E(X) - E(Y)$. When a random variable is Normally distributed, the mean of that distribution is the same as its expected value. This means $\mu_Z = E(X) - E(Y)$. We have shown previously that the $E(X) > E(Y)$. It logically follows then, that $\mu_Z = E(X) - E(Y) > 0$. We will not include the proof here, but it can be shown that the Normal distribution is symmetric about the mean. This means that exactly half of the distribution is less than the mean and half of the distribution is greater than the mean. Another way of saying this is to say: For any given Normally distributed random variable X, $\Pr(X < \mu_X) = \Pr(X > \mu_X) = 0.5$. With all of this out of the way, we can say $\Pr(Z > 0) > \Pr(Z > \mu_Z) = 0.5$. Therefore, $\Pr(Z > 0) > 0.5$. Remember that we have shown about that $\Pr(Z > 0) = \Pr(X > Y)$. So, we have successfully shown our theorem to be true whenever X and Y are Normally distributed and or occur a large number of times. Now, we will show this theorem to be true for any iid random variables X_i, Y_i regardless of whether they occur many times.

We have shown that for large n in X and Y, $\Pr(X > Y) > 0.5$. Equivalently, we could say the following: $\lim_{n\to\infty} \Pr(X_1 + X_2 + \ldots + X_n > Y_1 + Y_2 + \cdots + Y_n) > 0.5$. We can rewrite this using summation notation to say: $\lim_{n\to\infty} \Pr\left(\sum_{i=1}^{n} X_i > \sum_{i=1}^{n} Y_i\right) > 0.5$. We will now use our above theorem from Casella and Berger to manipulate this equation. Remember that if two random variables are identically distributed then they are also equal. Here we have X_i and Y_i which are identically distributed. Now, we must be careful to qualify my statement here. We do not mean that $X_i = Y_i \ \forall i \varepsilon \mathbb{N}$, but rather we mean to say that $X_i = X_j \ \forall i, j \varepsilon \mathbb{N}$ where $i \neq j$. Similarly, $Y_i = Y_j \ \forall i, j \varepsilon \mathbb{N}$. So, in this, we mean that X and Y are not identical to each other, but rather that each individual X_i and Y_i are identical to themselves. With that out of the way, we can now justify our next step. $X_1 = X_2 = \cdots = X_n$ and $Y_1 = Y_2 = \cdots = Y_n$. Since each individual X_i and Y_i are equal, then it logically follows that $\lim_{n\to\infty} \Pr\left(\sum_{i=1}^{n} X_i > \sum_{i=1}^{n} Y_i\right) > 0.5$ is equivalent to $\lim_{n\to\infty} \Pr\left(nX_i > nY_i\right) > 0.5$. We can make this a little bit simpler by writing $\Pr\left(\lim_{n\to\infty} nX_i > \lim_{n\to\infty} nY_i\right) > 0.5$. Since this is an inequality, we can divide both sides of the inequality by n. This produces the following result. $\Pr(X_i > Y_i) > 0.5$. This may produce some uneasiness for the reader who might've noticed that dividing by n is dividing by a limit. But this is easily justified using L'Hôpital's rule. Note: $\Pr\left(\lim_{n\to\infty} \frac{n}{n} X_i > \frac{n}{n} Y_i\right) > 0.5$ is a scenario in which $\frac{\infty}{\infty}$ is produced. But invoking L'Hôpital's rule quickly verifies our above answer of $\Pr(X_i > Y_i) > 0.5$. This is the result we have been trying to show, however. Remember that X_i and Y_i are simply iid random variables. We were using the subscripts to show that the random variables do not have to be large sums of variables but can be any given iid random variable. Therefore, we have successfully shown that for identically and independently distributed (iid) random variables X and Y where $E(X) > E(Y), \Pr(X > Y) > 0.5$.

Q.E.D.

With this method of proof already under our belt, it can quickly be shown that the converse of our theorem is true using a similar proof. Our resulting theorem is the following: Let X and Y be two identically and independently distributed (iid) random variables with finite variances where $\Pr(X > Y) > 0.5$. Then, $E(X) > E(Y)$.

Our stipulations for X and Y converging to Normal still apply in this case. Much like what we did above, we will start with X and Y occurring a large number of times. So, we define the collection of random variables X and Y as the following:

$X = \lim_{n\to\infty} X_1 + X_2 + \cdots + X_n$ and $Y = \lim_{n\to\infty} Y_1 + Y_2 + \cdots + Y_n$

We know due to the CLT, X and Y converge to Normal. $\Pr(X > Y) > 0.5$ is equivalent to $\Pr(X - Y >) > 0.5$. Let us define a new random variable $Z = X - Y$. Our given information now states $\Pr(Z > 0) > 0.5$. We know that since X and Y are Normally distributed, Z is also Normally distributed. Using the same argument as above, we know that $\mu_Z = E(Z) > 0$ due to the symmetry of the Normal distribution. $E(Z) > 0$ implies $E(X - Y) > 0$ which, in turn, implies that $E(X) > E(Y)$. We are not quite finished though because this is only true if X and Y are Normal. We can apply it generally, though, using a similar argument to what we used above. Using our definition of X and Y, we know that $E(X) > E(Y)$ is the same as $\lim_{n\to\infty} E(X_1 + X_2 + \cdots + X_n) > \lim_{n\to\infty} E(Y_1 + Y_2 + \cdots + Y_n)$ Which is equivalent to $\lim_{n\to\infty} E(X_1) + E(X_2) + \cdots + E(X_n) > \lim_{n\to\infty} E(Y_1) + E(Y_2) + \cdots + E(Y_n)$. Since we know X and Y are iid, $E(X_1) = E(X_2) = \cdots = E(X_n)$ and $E(Y_1) = E(Y_2) = \cdots = E(Y_n)$. So $\lim_{n\to\infty} E(X_1 + X_2 + \cdots + X_n) > \lim_{n\to\infty} E(Y_1 + Y_2 + \cdots + Y_n)$, can be replaced with $\lim_{n\to\infty} \sum_{i=1}^{n} E(X_i) > E(Y_i)$. This in turn implies that $\lim_{n\to\infty} nE(X_i) > nE(Y_i)$. From here we derive our conclusion that $E(X) > E(Y)$ for any given iid random variables X and Y that meet our conditions of convergence.

Q.E.D.

Discussion of Theorem

We have successfully shown this theorem to be true but have not discussed exactly what it means or what the implications of the theorem are. I would like to start by explaining what this theorem means in a practical sense. Imagine we are at a casino and you have an option of playing two different games: game A or game B. We do not know what kind of distributions these games follow but we do know they follow some kind of iid distribution. We know that, on average, game A has expected winnings of 10 dollars and game B has expected winnings of 5 dollars. Even though we do not know specifically how these games are distributed, we can use our theorem to determine the best game to play. We know that the expected winnings of A are greater than B. So, we can use our theorem to know that you are more likely to win a bigger amount of money in A than in B for any given round of the games. So, game A would always be a better choice. We can also consider a class with various grades. Say, for example, that two school's grades follow iid random distributions with known means. If we know that school A has an average grade of 85 for its graduating class and we know that school B has an average grade of 80 for its graduating class, then we would know it is more likely that any given student of school A would have a higher grade than any given student in school B. I believe this theorem could be particularly useful for data mining companies or companies that focus on targeted advertising. The theorem shows how probabilistic information can be known about an individual by only knowing information about the group to which the individual belongs. Many companies already use this fact to their advantage without the mathematical proof of the preceding theorem, but I believe the theorem could enhance the advertising processes already in place. If a company knows that a specific demographic of people generally buys their product more than another demographic, then any given individual of the first group would be more likely to buy the company's product than any given individual of the second group. This, of course, assumes that these individuals follow some sort of iid distribution. In that case, it would make the most sense to target members of the first demographic with advertising as they would be the most likely to buy said product. As previously stated, many companies already employ this train of thought, but now it is justified by a mathematical theorem.

These are only a few examples of how this theorem could be used, but it is easy to see how the theorem could be used in many more circumstances. After seeing the theorem in use, one can quickly verify in their own mind it matches up with intuition. The theorem produces results that are to be expected, however, now we know definitively that the expected result is indeed the mathematically true one under the given circumstances.

I would like to also briefly discuss what this theorem does not imply and how this theorem could be poorly applied to real life. This theorem works due to the random sampling of random variables and the fact that these random variables are iid. So, this theorem will always hold true if you start with an iid distribution and work towards finding probabilities of a random observation of that distribution. It does not work, however, to start with an observation, and then retroactively apply a distribution to that observation. This usurps the inherent randomness of the random variable and is a poor use of statistics. Let me explain this further with an example. Suppose that a company is wanting to target certain demographics for their product. They know that the number of products that men and women buy are iid and that women on average buy more of the product than men do. Then, it would be a correct application of my theorem for the company to target a random woman over a random man because the woman is more likely to buy the product than the man. However, it would be incorrect for a company to find a woman and a man and retroactively say that the woman is more likely to buy the product than the man because women on average buy more of the product than men. Finding an observation and then choosing a distribution that it belongs to is an incorrect application of this theorem. Let me elaborate further on why this does not work by continuing with my previous example. Let us say that the company does know that women buy more of the product on average than men. However let us also that the company knows that people under the age of 20 years old and people over the age of 20 years old similarly follow iid distributions where the expected number of products bought by under 20 year olds is greater than the expected number of products bought by over 20 year olds. Now, it would be correct for the company to target a random person under 20 years old as opposed to someone over 20 years because the younger person is more likely to buy the product than the older person. Similarly, the company should target a random woman over a random man. However, let us say that the company finds a 19-year-old man and a 30-year-old woman. The company cannot use the theorem to justify targeting either person. In order to do this, the company would have to decide which distribution to apply to each person. Is the young person more likely to buy the product because they are part of the young distribution as opposed to the other person who is a part of the old distribution? Or, should the company target the woman because she is a part of the woman distribution as opposed to the other person who is a part of the man distribution. The problem with this scenario is that we have started with an individual and attempted to apply a large-scale distribution to that individual as opposed to starting with the distribution and randomly choosing an individual from that distribution. When we start with the individual, we ruin many of the aspects that make this theorem work. The distributions that we are working with are indeed iid, however the people are not iid. So, if we start with a person/observation and attempt to retroactively fit them into a distribution, we are ruining the randomness of the sampling and several other aspects of this theorem that are imperative to what makes this theorem work.

So, we have shown that this theorem can be a powerful tool with several real-life applications, however the theorem is only as powerful in the real world as it is applied correctly. I would like to think that this theorem may have applications, not only in real life, but also in the field of theoretical probability and statistics. Much more research on this theorem could be done to find out, not only if this theorem can be applied to other areas of mathematics, but also if this theorem holds true for certain distributions that do not converge to the Normal Distribution.

Conclusion

The CLT is a powerful theorem with many applications both mathematical and practical. We have examined the CLT from many standpoints and determined its origin, uses and mathematical importance. Because the theorem is so important to the field of statistics, it can be found in many different fields, many of which may seem totally unrelated to the theorem. The CLT has come a long way from when it was first discovered. Its practicality has been widely expanded throughout its life. Additionally, we found a new use for the CLT. Through method of rigorous proof, we showed how the CLT can be used to discover a new theorem relating probabilities with expected values. This new theorem provides a means for determining important information based off limited knowledge. Using this theorem, we can make mathematically informed decisions about what is the most likely to occur without the luxury of exact distribution knowledge. This theorem may have applications for casinos, schools and more. The most pertinent of applications of this theorem is likely in the area of targeted advertising for businesses. We must be careful, however when applying this theorem to real life as there are certainly incorrect ways to use the theorem. We must additionally understand the limitations of this new theorem, not only in how it is applied, but also in its inherent boundaries. This theorem only works, as shown, with iid random variables. It would be an interesting area of further study to determine if this theorem could hold true for other types of distributions as well. My theorem is not the only application of the CLT. Many other theorems have already been discovered due to the CLT and it is my hope that many more will continue to be found and positively impact the field of statistics as well as countless other fields. Despite the progress and long history of the CLT, it is not a completely closed matter. Much research is still required in order to find the applicability of the CLT in dependent distributions. It is my hope, however, that this paper has brought light to what has already been done with the CLT, how it is used in life, and what still needs to be done. Additionally, I hope that my new theorem, has provided insight and further understanding in the field of probability and statistics, as well as in other areas of research.

Works Cited

Brockett, Patrick L. "On the Misuse of the Central Limit Theorem in Some Risk Calculations: Introduction." *Journal of Risk and Insurance* 50, 4, 1983, 727. ProQuest. https://seu.idm.oclc.org/login?url=https://www.proquest.com/docview/235112046?accountid=43912

Casella, George, and Roger Berger. *Statistical Inference*. The Wadsworth Group, 2002.

Dassau, Kemo J., and J. E. Hasbun. "Demonstrating The Central Limit Theorem Using Matlab." *Georgia Journal of Science* 62, 3, 2004, 133-40. ProQuest, https://seu.idm.oclc.org/login?url=https://www.proquest.com/docview/230548909?accountid=43912.

Goldstein, Larry. “A Probabilistic Proof of the Lindeberg-Feller Central Limit Theorem.” *The American Mathematical Monthly*, vol. 116, no. 1, 2009, pp. 45–60. JSTOR, www.jstor.org/stable/27642663.

Goldstein, Larry. “Bounds on The Constant in The Mean Central Limit Theorem.” *The Annals of Probability*, vol. 38, no. 4, Institute of Mathematical Statistics, 2010, pp. 1672–89.

Goldstein, Larry, and Gesine Reinert. “Stein's Method and the Zero Bias Transformation with Application to Simple Random Sampling.” *The Annals of Applied Probability*, vol. 7, no. 4, 1997, pp. 935–952. JSTOR, www.jstor.org/stable/2245253.

Jackman, Simon. "Pooling the Polls Over an Election Campaign." *Australian Journal of Political Science* 40, 4, (2005), 499-517. ProQuest. https://seu.idm.oclc.org/login?url=https://www.proquest.com/docview/204411509?accountid=43912

Kwak, Sang Gyu, and Jong Hae Kim. "Central Limit Theorem: The Cornerstone of Modern Statistics." *Korean Journal of Anesthesiology*, vol. 70, no. 2, Apr. 2017, pp. 144–156. EBSCOhost, doi:10.4097/kjae.2017.70.2.144. Accessed 18 Sept. 2020.

Le Cam, L. "The Central Limit Theorem Around 1935." *Statistical Science,* vol. 1, no. 1, 1986, pp. 78–91. *JSTOR*, www.jstor.org/stable/2245503.

Malik, J.S., et al. "Generating High Tail Accuracy Gaussian Random Numbers in Hardware Using Central Limit Theorem," *2011 IEEE/IFIP 19th International Conference on VLSI and System-on-Chip*, 2011, pp. 60-65, doi: 10.1109/VLSISoC.2011.6081630.

Massaro, Dominic W. "Resurrecting Information Theory." *The American Journal of Psychology* 108, 4, 1995, 609. ProQuest. 10.2307/1423078.

Rinott, Yosef, and Vladimir Rotar. "Normal Approximations by Stein's Method." *Decisions in Economics and Finance* 23.1 (2000): 15-29. ProQuest, http://dx.doi.org/10.1007/s102030050003

Singh, Ashok, et al. "Casino Games and the Central Limit Theorem1." *UNLV Gaming Research & Review Journal* 17, 2, 2018, 45-61. ProQuest, https://seu.idm.oclc.org/login?url=https://www.proquest.com/docview/1494797744?accountid=43912.

Thomas, David B., et al. "Gaussian Random Number Generators." *ACM Computing Surveys*, vol. 39, no. 4, Nov. 2007, pp. 11–es. *2007*, ACM Digital Library, doi:10.1145/1287620.1287622.

Zabell, S. L. "Alan Turing and the Central Limit Theorem." *The American Mathematical Monthly*, vol. 102, no. 6, 1995, pp. 483–494. JSTOR, www.jstor.org/stable/2974762.