Introductory Probability and the Central Limit Theorem

Vlad Krokhmal

07/29/2011

Abstract

In this paper I introduce and explain the axioms of probability and basic set theory, and I explore the motivation behind random variables. I build upon these concepts towards an introduction to the limit theorems, specifically the Weak Law of Large Numbers and the Central Limit theorem. I prove these two theorems in detail and provide a brief illustration of their application.

1 Basics of Probability

Consider an experiment with a variable outcome. Further, assume you know all possible outcomes of the experiment. The set of all possible outcomes of the experiment is called the sample space and is denoted by S. A collection of outcomes within this sample space is called an event and is denoted by E. We can think of an event as a subset of the set of all the possible outcomes.

Often, we are interested in the interaction between the events of a sample space. One such interaction is the union of events, denoted by the union operator $\cup$. The union of a set of events will occur if any of the events in the union occur. Thus, the union of the events A,B and C, i.e. $A \cup B \cup C$, will occur if either event A, B or C occur. Another interaction is the intersection of events, denoted by the intersection operator $\cap$. The intersection of a set of events will occur if all of the events in the intersection occur. Thus, the intersection of the events A,B, and C, i.e. $A \cap B \cap C$, will occur if the events A, B and C all occur.

Knowing these operations, we can define some interactions between events.

Definition 1 (Mutually Exclusive)**.** Two sets A and B are called mutually exclusive if their intersection is empty:

$$A \cap B = \emptyset$$

Definition 2 (Independent)**.** An event E is said to be independent of an event F if

$$P(E \cap F) = P(E) \cdot P(F)$$

Next, the three axioms of probability begin to relate set theory to probabilistic measurements. I use P(E) to represent the probability of some event E and P(S) to represent the probability of the entire sample space.

Axioms of Probability.

1. $0 \leq P(E) \leq 1$
2. $P(S) = 1$
3. For any sequence of mutually exclusive events $E_1, E_2, \ldots$:

$$P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i).$$

These definitions and axioms explain the underpinnings of basic probabilistic calculations. With these basics we can advance to more intricate probabilistic questions.

2 Random Variables

We are often interested in considering multiple outcomes of an experiment. For instance, we might be interested in the number of odd results from rolling three dice. In this example, we would be interested in multiple outcomes: the probability of the first die being odd, the probability of the second die being odd and the probability of the third die being odd. To work with multiple outcomes, we create a random variable.

Definition 3 (Random Variable)**.** A random variable is a function X that assigns a rule of correspondence for every point ξ in the sample space S (called the domain) a unique real value $X(\xi)$.

The rule of correspondence is given either by a probability mass function or the probability density function, depending on the type of random variable considered.

Definition 4 (Probability Mass Function)**.** For a random variable that can take on at most a countable number of possible values, a probability mass function p(a) is defined by

$$p(a) = P(X = a).$$

Definition 5 (Probability Density Function)**.** For a random variable X that is continuously defined, a probability density function f(x) is defined such that for a subset B $\in \mathbb{R}$

$$P(X \in B) = \int_B f(x)dx.$$

We might also be interested in the probability that the random variable is less than some value. For such cases we define the distribution function.

Definition 6 (Distribution Function)**.** For a random variable X, the distribution function F is defined by

$$F(x) = P(X \leq x).$$

For continuous random variables, the above equation can be represented as

$$F(x) = \int_{-\infty}^{\infty} f(t)dt$$

where f(t) is a probability density function.

Hence, we can see that the derivative of the distribution function yields the probability density function.

In the following example, I will illustrate the application of the random variable in the case mentioned in the beginning of this section.

Example 1. We let X denote the number of odd dice that turn up odd. Thus, X is a random variable that takes on one of the values 0,12,3 with the ollowing probabilities:
$P(X = 0) = 1/2 \cdot 1/2 \cdot 1/2 = 1/8$ $\quad P(X = 1) = \binom{3}{1} \cdot 1/2 \cdot 1/2 \cdot 1/2 = 3/8$
$P(X = 3) = 1/2 \cdot 1/2 \cdot 1/2 = 1/8$ $\quad P(X = 2) = \binom{3}{1} \cdot 1/2 \cdot 1/2 \cdot 1/2 = 3/8$
Thus, we can see that X is a discrete function, i.e. a random variable.

Certain probability distributions interest us more than others because of their qualities. The normal random variable has such a distribution. This distribution's peculiar qualities will make it the subject of the Central Limit theorem.

Definition 7 (Normal Random Variable)**.** X is a normal random variable with parameters μ and σ^2 if the density of X is given by

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

Whenever $\mu = 0$ and $\sigma^2 = 1$ we get a simplified equation:

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

We can see that f(x) is indeed a distribution function since integrating it from $-\infty$ to ∞ gives 1 and hence the sample space has probability 1, as required by the second axiom of probability.

Having defined the random variable, we are now interested in its properties. We can discribe the whole distribution of probabilities through two qualities of a random variable: its average value and spread. These terms are called expected value and variance, respectively.

Definition 8 (Expected Value or Mean)**.** If X is a discrete random variable having the probability mass function p(x), the expected value, denoted by E[X], is defined as

$$\text{E[X]} = \textstyle\sum_{x:p(x)>0} xp(x)$$

If X is a continuous random variable having the probability density function f(x), the expected value is defined as

$$\text{E[X]} = \textstyle\int_{-\infty}^{\infty} xf(x)\,dx.$$

Definition 9 (Variance)**.** If X is a random variable with mean μ ,where $\mu = E[X]$, then the variance of X, denoted by Var(X), is defined as

$$\text{Var(X)} = \text{E[(X - }\mu)^2].$$

Now I will introduce a few properties of the expected value and mean that will appear later in the paper. First I will show that expected value is linear and apply this result to transformations of variance.

Lemma 2.1 (Transformations of Mean and Variance)**.** If a and b are constants, then

$$E[aX+b] = aE[X]+b$$

and

$$\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X)$$

Proof.

$$\begin{aligned}
E[aX+b] &= \sum_{x:p(x)>0} (ax+b)p(x) && (1)\\
&= a\sum_{x:p(x)>0} xp(x) + b\sum_{x:p(x)>0} p(x) && (2)\\
&= aE[X]+b && (3)
\end{aligned}$$

and when $\mu = E[X]$,

$$\begin{aligned}
\mathrm{Var}(aX+b) &= E[(aX+b-a\mu-b)^2] && (4)\\
&= E[a^2(X-\mu)^2] && (5)\\
&= a^2E[(X-\mu)^2] && (6)\\
&= a^2\mathrm{Var}(X). && (7)
\end{aligned}$$

□

In the above proof, Step 4 simply applies the equality from Step 3. All other steps are easy algebraic manipulations. Next, I prove a lemma concerning transformations of the original random variable.

Lemma 2.2. If X is a discrete random variable that takes on one of the values x_i, $i \geq 1$, with respective probabilities $p(x_i)$, then for any real-valued function g

$$E[g(X)] = \sum_i g(x_i)p(x_i)$$

Proof. We start by grouping together all the terms having the same value of $g(x_i)$. Thus, let y_j, $j \geq 1$ represent the different values of $g(x_i)$, $i \geq 1$. Then:

$$\begin{aligned}
\sum_i g(x_i)p(x_i) &= \sum_j y_j \sum_{i:g(x_i)=y_j} p(x_i) && (1)\\
&= \sum y_j P(g(X)=y_j) && (2)\\
&= E[g(X)] && (3)
\end{aligned}$$

□

Next, I will demonstrate relevant properties of expected value and variance for joint random variables.

Lemma 2.3. If X and Y are random variables with finite expected value, then

$$E[X+Y] = E[X] + E[Y]$$

Proof. Let the sample space of $X = x_1, x_2, ...$ and the sample space of $Y = y_1, y_2, ...$ Then, we can write the random variable X+Y as a result of applying a function g(x,y) = x + y to the joint random variable (X,Y).

$$\begin{aligned}
E[X+Y] &= \sum_j \sum_k (x_j + y_k) P(X = x_j, Y = y_k) && (1)\\
&= \sum_j \sum_k x_j P(X = x_j, Y = y_k) + \sum_j \sum_k y_k P(X = x_j, Y = y_k) && (2)\\
&= \sum_j x_j P(X = x_j) + \sum_k y_k P(Y = y_k) && (3)\\
&= E[X] + E[Y] && (4)
\end{aligned}$$

□

Here, step 1 follows by Lemma 2.2. Step 3 follows since $\sum_k P(X = x_j, Y = y_k) = P(X = x_j)$. Next, I show a property of mean and variance when considering independent random variables.

Lemma 2.4. If X and Y are independent random variables, then

$$E[X \cdot Y] = E[X]E[Y]$$

and

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

Proof. Let the sample space of $X = x_1, x_2, ...$ and the sample space of $Y = y_1, y_2, ...$

$$\begin{aligned}
E[X \cdot Y] &= \sum_j \sum_k x_j y_k P(X = x_j) P(Y = y_k) && (1)\\
&= \left(\sum_j x_j P(X = x_j) \right) \left(\sum_k y_k P(Y = y_k) \right) && (2)\\
&= E[X]E[Y] && (3)
\end{aligned}$$

Next, let E[X] = a and E[Y] = b.

$$\begin{aligned}
\mathrm{Var}(X+Y) &= E[(X+Y)^2] - (a+b)^2 && (4)\\
&= E[X^2] + 2E[XY] + E[Y^2] - a^2 - 2ab - b^2 && (5)\\
&= E[X^2] - a^2 + E[Y^2] - b^2 && (6)\\
&= \mathrm{Var}(X) + \mathrm{Var}(Y) && (7)
\end{aligned}$$

□

Here, step 1 follows from the definition of independence, $P(X = x_j, Y = y_k) = P(X = x_j)P(Y = y_k)$. Step 6 follows from the definition of independence, $E[XY] = E[X]E[Y] = ab$.

We can extend the concept of the mean to the situation where we are dealing with multiple random variables. In this case, we calculate the sample mean.

Definition 10 (Sample Mean)**.** Let $X_1, ..., X_n$ be independent and identically distributed random variables having distribution function F and expected value μ. Such a sequence constitutes a sample from the distribution F. Given a sample, we define the sample mean, $\widehat{X}$, as:

$$\widehat{X} = \sum_{i=1}^{n} \frac{X_i}{n}$$

Furthermore,

$$E[\widehat{X}] = \frac{1}{n} \sum_{i=1}^{n} E[X_i]$$

So we now know how to take the expected value of a random variable, but let's say we were interested in the expected value of the square of the random variable, or the cube, or so on. This brings us to the concept of the moments of a random variable.

Definition 11 (Moments of a Random Variable)**.** The k-th moment of a random variable X is $E[X^k] \quad \forall k \in \mathbb{N}$.

To make the computation of the moments of a random variable easier, we define a special Moment Generating Function.

Definition 12 (Moment Generating Function)**.** The moment generating function M(t) of a random variable X is defined for all real values of t by

$$M(t) = E[e^{tX}] = \begin{cases} \sum_x e^{tX} p(x) & \text{if X is discrete with mass function } p(x), \\ \int_{-\infty}^{\infty} e^{tX} f(x) dx & \text{if X is continuous with density } f(x). \end{cases}$$

The moment generating function (MGF) has a few interesting properties which we will need to keep in mind throughout the paper. First, for independent random variables X and Y, the MGF satisfies $M_{X+Y}(t) = E[e^{t(X+Y)}] = E[e^{tX} \cdot e^{tY}] = M_x(t) \cdot M_y(t)$. Second, all moments of a random variable can be obtained by differentiating the MGF and evaluating the derivative at 0. More precisely,

$$M^{(k)}(0) = E[X^k]$$

Although I won't prove this in general, it can be easily seen through induction as shown in the following demonstration.

Demonstration 1.

$$M'(t) = \frac{d}{dt} E[e^{tX}] = E\Big[\frac{d}{dt}(e^{tX})\Big] = E[Xe^{tX}]$$

and when we evaluate M'(t) at t= 0, we get:

$$M'(0) = E[X \cdot e^0] = E[X]$$

which is the first moment of the random variable X.

The use of MGFs, and in particular of the MGF of the standard normal distribution, will be key to the proof of the Central Limit Theorem. Hence, I compute the MGF of the standard normal distribution below.

Example 2. Let Z be a unit normal random vaariable with mean 0 and variance 1.

$$\begin{aligned}
M_Z(t) &= E[e^{tZ}] && (1)\\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{tx}e^{-x^2/2}dx && (2)\\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \exp\Big(-\frac{x^2-2tx}{2}\Big)dx && (3)\\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \exp\Big(-\frac{(x-t)^2}{2}+\frac{t^2}{2}\Big)dx && (4)\\
&= e^{t^2/2}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-(x-t)^2/2}dx && (5)\\
&= e^{t^2/2} && (6)
\end{aligned}$$

Step 2 is accomplished by simply inserting a normal random variable for Z and taking the expected value. Step 3 uses the rule $e^{x+y} = e^x e^y$ to combine the exponents. Step 4 uses the complete the square technique. Step 5 pulls out $e^{t^2/2}$ from the integrand since that term is simply a constant. Step 6 uses the fact that we are integrating a probability distribution function previously shown to be equal to 1.

3 Preparatory Results

Determining probabilities of certain outcomes of a random variable becomes more complicated when we are given only the mean or both the mean and the variance. However, while exact probabilities are impossible to find, bounds on the probability can be derived. One such bound is given by Markov's inequality.

Lemma 3.1. Markov's Inequality
If X is a random variable that takes only nonnegative values, then for any value $a > 0$,

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

Proof. Define a new random variable I such that

$$I = \begin{cases} 1 & \text{if } X \geq a, \\ 0 & \text{otherwise.} \end{cases}$$

Case 1: If I = 1 then X $\geq$ a; therefore I $\leq \frac{X}{a}$ since X $\geq$ 0 and a > 0.
Case 2: If I = 0 then X < a but $\frac{X}{a} \geq 0$ since X $\geq$ 0 and a > 0.
Thus, the inequality I $\leq \frac{X}{a}$ holds in all cases.

Next, we take the expected value of both sides, i.e. E[I] and E[$\frac{X}{a}$], s.t.

$$\text{E[I]} = \textstyle\int_{-\infty}^{\infty} p(x)\cdot \text{I} \qquad\qquad \text{E}[\frac{X}{a}] = \textstyle\int_{-\infty}^{\infty} p(x)\cdot \frac{X}{a}$$

Clearly E[I] $\leq$ E[$\frac{X}{a}$] since I $\leq \frac{X}{a}$

Furthermore, E[I] = P(X $\geq$ a) since E[I] is just the sum the probabilities where X $\geq$ a.
Hence, P$(X \geq a) \leq \frac{E[X]}{a}$. This proves Lemma 3.1. $\square$

Thus, using Markov's inequality, we can create a bound on the probability of a certain outcome given only the distribution's mean. Furthermore, this result is integral to deriving other bounds on the probability distribution of a random variable, such as the bound given by Chebyshev's inequality.

Lemma 3.2. Chebyshev's Inequality
If X is a finite random variable with finite mean μ and variance σ^2, then for any value $k > 0$,

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

Proof. From Markov's inequality we know that

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

where X is any random variable that takes on only nonnegative values and $a > 0$.
We can apply Markov' inequality to the random variable $(X - \mu)^2$, which satisfies the necessary conditions. We get:

$$P((X - \mu)^2 \geq k^2) \leq \frac{E[(X - \mu)^2]}{k^2}.$$

Next, notice that $(X - \mu^2) \geq k^2 \iff |X - \mu| \geq k$. Hence, whenever one of these inequalities is true so is the other inequality. In other words, the probability of either inequality being true is the same. Thus, we can swap one inequality for the other to get:

$$P(|X - \mu| \geq k) \leq \frac{E[(X - \mu)^2]}{k^2}.$$

Lastly, by definition of variance, $E[(X - \mu)^2] = \sigma^2$, so we get:

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

This proves Lemma 3.2. $\square$

In the next section of the paper, Chebyshev's inequality will be used to prove the weak law of large numbers that states the conditions under which the average of a sequence of random variables converges to the expected average. This result will rely primarily on Chebyshev's inequality by allowing the random variable X to be a sequence of random variables.

4 Weak Law of Large Numbers and the Central Limit Theorem

Theorem 4.1. The Weak Law of Large Numbers
Let X_1, X_2, ... be a sequence of independent and identically distributed random variables, each having finite mean $E[X_i] = \mu$ and variance σ^2. Then, for any $\epsilon > 0$,

$$P\left(\left|\frac{X_1 + ... + X_n}{n} - \mu\right| \geq \epsilon\right) \to 0 \qquad \text{as} \qquad n \to \infty$$

Proof. We see that,

$$E\left[\frac{X_1 + ... + X_n}{n}\right] = \frac{1}{n} \cdot E[X_1] + ... + E[X_n] = \frac{n\mu}{n} = \mu.$$

Furthermore,

$$\text{Var}\left(\frac{X_1 + ... + X_n}{n}\right) = \text{Var}\left(\frac{X_1}{n}\right) + ... + \text{Var}\left(\frac{X_n}{n}\right)$$

We can see that $\text{Var}(\frac{X_1}{n}) = \text{E}[(\frac{X_1 - \mu}{n})^2] = (\frac{1}{n^2}) \cdot \text{E}[(X_1 - \mu)^2]$, and therefore we get:

$$\text{Var}\left(\frac{X_1 + ... + X_n}{n}\right) = \overbrace{\frac{\sigma^2}{n^2} + ... + \frac{\sigma^2}{n^2}}^{\text{n times}} = \frac{\sigma^2}{n}$$

Now we treat $\left(\frac{X_1 + ... + X_n}{n}\right)$ as a new random variable X. The new random variable X clearly satisfies the conditions for Chebyshev's Inequality. Hence, we apply the lemma to get the following:

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

As n $\to \infty$, it then follows that

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

This proves Theorem 3.1. □

The Weak Law of Large Numbers demonstrates that given a large aggregate of identical random variables, the average of the results obtained will approach the sample mean.

Next, I will prove a restricted case of the Central Limit theorem that deals only with a standard normal random variable. This theorem is concerned with determining the conditions under which the sum of a large number of random variables has a probability distribution that is approximately normal. The following Lemma is integral to the proof of the Central Limit theorem. This is a technical result and will not be proven in this paper. However, a proof of this can be found in *Probability and Random Processes* [1].

Lemma 4.1. Let $Z_1, Z_2, ...$ be a sequence of random variables having distribution functions F_{Z_n} and moment generating functions M_{Z_n} s.t. n ≥ 1. Furthermore, let Z be a random variable having distribution function F_Z and moment generating functions M_Z. If $M_{Z_n}(t) \to M_Z$ for all t, then $F_{Z_n}(t) \to F_Z(t)$ for all t at which $F_Z(t)$ is continuous.

To see the relevance of this Lemma, let's set $Z_n = \sum_{i=1}^{n} \frac{X_i}{\sqrt{n}}$ where X_i are independent and identically distributed random variables and let Z be a normal random variable. Then, if we show that the MGF of $\sum_{i=1}^{n} \frac{X_i}{\sqrt{n}}$ approaches the MGF of Z (which we previously calculated to be $e^{t^2/2}$) as $n \to \infty$, we simultaneously show that the probability distribution of $\sum_{i=1}^{n} \frac{X_i}{\sqrt{n}}$ approaches the normal distribution as $n \to \infty$. This is the method we will use to prove the Central Limit theorem.

Theorem 4.2. The Central Limit Theorem
Let $X_1, X_2, ...$ be a sequence of independent and identically distributed random variables each having mean μ and variance σ^2. Then the distribution of $\frac{X_1+...+X_n-n\mu}{\sigma\sqrt{n}}$ tends to the standard normal as $n \to \infty$. That is, for $-\infty < a < \infty$,

$$P\Big(\frac{X_1 + ... + X_n - n\mu}{\sigma\sqrt{n}} \leq a\Big) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2} dx \qquad as \quad n \to \infty$$

Proof. We begin the proof with the assumption that $\mu = 0$, $\sigma^2 = 1$ and that the MGF of the X_i exists and is finite.
We already know the MGF of a normal random variable, but we still need to compute the MGF of the sequence of random variables we are interested in: $\sum_{i=0}^{n} X_i/\sqrt{n}$.
By definition of MGF, we can see that $M\Big(\frac{t}{\sqrt{n}}\Big) = E\Big[\exp\Big(\frac{tX_i}{\sqrt{n}}\Big)\Big]$. However, we are interested in the MGF of $E\Big[exp\Big(t\sum_{i=1}^{n} \frac{X_i}{\sqrt{n}}\Big)\Big]$. Here is how we find the MGF:

$$E\Bigg[exp\Bigg(t\sum_{i=1}^{n} \frac{X_i}{\sqrt{n}}\Bigg)\Bigg] = E\Bigg[\exp\Bigg(\sum_{i=1}^{n} X_i \cdot \frac{t}{\sqrt{n}}\Bigg)\Bigg] \tag{1}$$

$$= E\Bigg[\prod_{i=1}^{n} \exp\Bigg(X_i \cdot \frac{t}{\sqrt{n}}\Bigg)\Bigg] \tag{2}$$

$$= \prod_{i=1}^{n} E\Bigg[\exp\Bigg(X_i \cdot \frac{t}{\sqrt{n}}\Bigg)\Bigg] \tag{3}$$

$$= \prod_{i=1}^{n} M\Big(\frac{t}{\sqrt{n}}\Big) \tag{4}$$

$$= \Bigg[M\Big(\frac{t}{\sqrt{n}}\Big)\Bigg]^n \tag{5}$$

Step 1 is accomplished by simple distribution. Step 2 uses the rule $e^{x+y} = e^x \cdot e^y$. Step 3 relies on the fact that the X_is are independent and therefore the E and the $\prod$ operators are interchangeable. In step 4 I simply substitute a previously calculated identity. And in step 5 I simply rewrite the equation into a more accessible format.
Now we define L(t) =log M(t) and evaluate $L(0), L'(0), L''(0)$.

$$L(0) = \log\, M(0) = \log\, E[e^{0 \cdot X_i}] = \log E[1] = \log\, 1 = 0 \tag{1}$$

$$L'(0) = \frac{M'(0)}{M(0)} = M'(0) = E[Xe^{0 \cdot X_i}] = E[X] = \mu = 0 \tag{2}$$

$$L''(0) = \frac{M(0)M''(0) - [M'(0)]^2}{[M(0)]^2} = \frac{1 \cdot E[X^2] - 0^2}{1^2} = E[X^2] = \sigma^2 = 1 \tag{3}$$

Now we are ready to prove the Central Limit theorem by showing that $\left[M\left(\frac{t}{\sqrt{n}}\right)\right]^n \to e^{t^2/2}$as $n \to \infty$. By taking the log of both sides, we can see that this is equivalent to showing $nL(t/\sqrt{n}) \to t^2/2$ as $n \to \infty$. Hence, we compute:

$$\begin{aligned}
\lim_{n\to\infty} \frac{L(t/\sqrt{n})}{n^{-1}} &= \lim_{n\to\infty} \frac{-L'(t/\sqrt{n})n^{-3/2}t}{-2n^{-2}} && (1)\\
&= \lim_{n\to\infty} \frac{-L'(t/\sqrt{n})t}{-2n^{-1/2}} && (2)\\
&= \lim_{n\to\infty} \frac{-L''(t/\sqrt{n})n^{-3/2}t^2}{-2n^{-3/2}} && (3)\\
&= \lim_{n\to\infty} L''\left(\frac{t}{\sqrt{n}}\right)\frac{t^2}{2} && (4)\\
&= \frac{t^2}{2} && (5)
\end{aligned}$$

Here, step 1 was accomplished by L'Hopital's rule since both the top and the bottom of the original fraction equaled 0. Step 2 simply reduces the fraction. Step 3 is again accomplished by L'Hopital's rule since again both the top and the bottom of the reduced fraction equal 0. Step 4 reduces the equation. Step 5 uses the previously calculated value of L"(0) to give us the final result. Having shown this, we can now apply Lemma 3.3. to prove the Central Limit theorem for the case where $\mu = 0$ and $\sigma^2 = 1$. □

I will briefly illustrate the theorem with a simple application from investment. Consider an investor who chose a diversified portfolio with 100 stocks. We assume that possible yields of each stock are identically distributed(although in reality such an assumption would be difficult to satisfy). Then, using the Central Limit thorem, he can model the returns of this portfolio using the normal distribution.

Acknowledgements

I would like to express my gratitude to the Mathematics Research Experience for Undergraduates program for providing me with the resources and inspiration for writing this paper. In particular, I would like to thank Peter May, the director of the program, for organizing such an amazing summer experience. Furthermore, I would like to thank my mentor, Yan Zhang, for helping me with difficult concepts in probability and teaching me to write in mathematical style.

References

[1] Probability and Random Processes by Venkatarama Krishnan. 2006 Edition.

[2] A First Course in Probability by Sheldon Ross. Sixth Edition.