CENTRAL LIMIT THEOREM

DHRUV RAGHAVAN

In this paper, we will explore discrete probability spaces, stating and proving some preliminary results concerning probability distributions and random variables. We then introduce Moment Generating Functions, a crucial tool in proving the Central Limit Theorem. Next, we will prove Markov's Inequality and Chebyshev's Inequality, leading to the proof of the Weak Law of Large Numbers, another fundamental theorem in probability. Lastly, we use these ideas to prove the Central Limit Theorem, and conclude with some computational examples.

1. Discrete Probability spaces

Definition 1.1. A *(discrete) probability space* is a pair $(\Omega, \mathbb{P})$, where Ω is a set and $\mathbb{P} : \Omega \to [0, 1]$ is a function such that $\sum_{\omega \in \Omega} \mathbb{P}(\omega) = 1$.

Alternatively, if we list out $\Omega = (\omega_1, \dots, \omega_n)$ where Ω is finite, then we can represent $\mathbb{P}$ as a vector $v = (v_1, \dots, v_n)$, where $v_i = \mathbb{P}(\omega_i)$. Each $v_i \in [0, 1]$, and the entries sum to 1.

If $E \subseteq \Omega$ is an event, then the probability of E is

$$\mathbb{P}(E) = \sum_{\omega \in E} \mathbb{P}(\omega).$$

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 define a random variable.

Definition 1.2. Suppose we have a discrete probability space $(\Omega, \mathbb{P})$ and a set E. A *random variable* is a function $X : \Omega \to E$. Given a subset $S \subseteq E$, we can ask for the probability that X is in S. That is, $\mathbb{P}(X \in S)$, by which we mean more formally

$$\mathbb{P}(X \in S) = \sum_{\omega : X(\omega) \in S} \mathbb{P}(\omega).$$

We say that n events $A_1, \dots, A_n$ are independent if, for any subset $S \subseteq \{1, 2, \dots, n\}$, we have

$$\mathbb{P}\left(\bigcap_{i \in S} A_i\right) = \prod_{i \in S} \mathbb{P}(A_i).$$

Events that are not independent are said to be *dependent*.

Definition 1.3. 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

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

Date: October 2020.

Definition 1.4. For a random variable X, the distribution function F is defined by

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

Definition 1.5. Let $F : \mathbb{R} \to \mathbb{R}$ be a function with the following properties:

- F is monotonically increasing, i.e. if $a \leq b$, then $F(a) \leq F(b)$.
- $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to +\infty} F(x) = 1$.
- F is continuous on the right, i.e. for all $a \in \mathbb{R}$, $\lim_{x \to a^+} F(x) = F(a)$.

Such a function is called a *cumulative density function*, or CDF. Then we can use F to define a probability on $\mathbb{R}$ by specifying that $\mathbb{P}(x \leq a) = F(a)$.

When the CDF is differentiable, its derivative $f(x) = F'(x)$ is called the *probability density function* or PDF. When the PDF exists, it determines the CDF, so it suffices to describe the distribution in terms of the PDF. An important distribution is the *normal distribution*, which is a continuous probability distribution that is symmetrical on both sides of the mean, The area under this bell-curve represents probability and sums to one. It the most important probability distribution, and comes up quite often in statistics.

Definition 1.6. 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 axioms of probability. We are now interested in the properties of random variables, namely their average and spread.

Definition 1.7. If X is a discrete random variable having the probability mass function $p(x)$, the expected value, denoted by $\mathbb{E}[X]$, is defined as

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

In other words, the expected value is a weighted average of all possible values of X, where the weights are the probabilities that X takes those values.

Example. If X is the number showing on a fair, standard die, then

$$\mathbb{E}[X] = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6 = \boxed{3.5}.$$

Definition 1.8. If X is a random variable with mean μ, where $\mu = \mathbb{E}[X]$, then the variance of X, denoted by $Var(X)$, is defined as

$$Var(X) = \mathbb{E}\left[(X - \mu)^2\right]$$

. The standard deviation of X is $\sqrt{Var(X)}$.

Variance helps us determine the expected deviation from μ.

Example. Going back to our dice example, the expected value $\mu = 3.5$, so its variance is

$$Var(x) = \frac{1}{6}(1-3.5)^2 + \frac{1}{6}(2-3.5)^2 + \frac{1}{6}(3-3.5)^2 + \frac{1}{6}(4-2.5)^2 + \frac{1}{6}(5-3.5)^2 + \frac{1}{6}(6-3.5)^2$$

$$= \frac{1}{6}\left(\frac{25}{4} + \frac{9}{4} + \frac{1}{4} + \frac{1}{4} + \frac{9}{4} + \frac{25}{4}\right) = \boxed{\frac{35}{12}}.$$

Now we will prove a few basic properties of expected value and variance.

Lemma 1.9. *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*

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

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:

$$\sum_i g(x_i)\, p(x_i) = \sum_j y_j \sum_{i:g(x_i)=y_j} p(x_i)$$

$$= \sum y_j P\left(g(X) = y_j\right) = \mathbb{E}[g(X)].$$

■

Lemma 1.10. *(Linearity of Expectation) Let X and Y be two random variables on a sample space Ω, and let $c \in \mathbb{R}$. Then we have*

$$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y) \qquad \textit{and} \qquad \mathbb{E}[aX+b] = a\mathbb{E}[X] + b.$$

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

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

Here, step 1 follows by lemma 1.9. Step 3 follows since $\sum_k P(X = x_j, Y = y_k) = P(X = x_j)$ As for the second part of the lemma,

$$\mathbb{E}[aX+b] = \sum_{x:p(x)>0} (ax+b)p(x) = a \sum_{x:p(x)>0} xp(x) + b \sum_{x:p(x)>0} p(x) = a\mathbb{E}[X] + b.$$

■

Linearity of Expectation is quite useful when solving expected value problems.

Example. Suppose we flip a fair coin n times. What is the expected number of heads? To do this, we let X be the number of heads in n flips. Now, break up X into a sum of random variables. For $1 \leq i \leq n$, let $X_i = 1$ if the i^{th} flip is heads, and $X_i = 0$ if the i^{th} flip is tails. Then we get $X = X_1 + X_2 + \cdots + X_n$. For each i, we have

$$\mathbb{E}(X_i) = 1 \cdot \mathbb{P}(X_i = 1) + 0 \cdot \mathbb{P}(X_i = 0) = \frac{1}{2}.$$

Thus we have $\mathbb{E}(X) = \frac{1}{2} \cdot n = \frac{n}{2}$.

We can apply this to get a similar lemma concerning variance.

Lemma 1.11. *If a and b are constants, then*

$$Var(aX + b) = a^2 Var(X).$$

Proof. Using linearity of expectation and letting $\mu = \mathbb{E}[X]$, we have

$$Var(aX + b) = \mathbb{E}[(aX + b - a\mu - b)^2] = \mathbb{E}[a^2(X - \mu)^2] = a^2\mathbb{E}[(X - \mu)^2] = a^2 Var(X).$$

■

Lastly, we have property of expected value and variance for independent random variables.

Lemma 1.12. *If X and Y are independent random variables, then*

$$\mathbb{E}[X \cdot Y] = \mathbb{E}[X]\mathbb{E}[Y]$$

and

$$Var(X + Y) = Var(X) + Var(Y)$$

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

$$\mathbb{E}[X \cdot Y] = \sum_j \sum_k x_j y_k P(X = x_j) P(Y = y_k) = \left(\sum_j x_j P(X = x_j)\right)\left(\sum_k y_k P(Y = y_k)\right) = \mathbb{E}[X]\mathbb{E}[Y]$$

Next, let $\mathbb{E}[X] = a$ and $\mathbb{E}[Y] = b$.

$$\begin{aligned} Var(X + Y) &= \mathbb{E}\left[(X + Y)^2\right] - (a + b)^2 \\ &= \mathbb{E}\left[X^2\right] + 2\mathbb{E}[XY] + \mathbb{E}\left[Y^2\right] - a^2 - 2ab - b^2 \\ &= \mathbb{E}\left[X^2\right] - a^2 + \mathbb{E}\left[Y^2\right] - b^2 \\ &= Var(X) + Var(Y) \end{aligned}$$

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 other definition of independence,

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] = ab$$

. ■

2. Moment Generating Functions

So far, we have encountered two important attributes of a random variable, namely the expected value and variance. However, they do not contain all the available information about the density function. For example, suppose X and Y are random variables with distributions

$$p_x = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 0 & \frac{1}{4} & \frac{1}{2} & 0 & 0 & \frac{1}{4} \end{pmatrix},$$
$$p_y = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{4} & 0 & 0 & \frac{1}{2} & \frac{1}{4} & 0 \end{pmatrix}$$

. Then we have $\mathbb{E}[X] = \mathbb{E}[Y] = \frac{7}{2}$ and $V(X) = V(Y) = \frac{9}{4}$, while p_x and p_y are clearly different density functions. What else must we know to determine p? This brings us to the moments of a random variable.

Definition 2.1. The k^{th} moment of a random variable X is $\mathbb{E}[X^k] = \sum_{j=1}^{\infty}(x_j)^k p(x_j)$, provided the sum converges.

We define the Moment Generating Function to compute the moments of a random variable.

Definition 2.2. The Moment Generating Function (MGF) $M(t)$ of a discrete random variable X is defined for all real t by

$$M(t) = \mathbb{E}[e^{tX}] = \sum_x e^{tX} p(x)$$

where $p(x)$ is the mass function.

If we differentiate $M(t)$ n times and then set $t = 0$, we have

$$\left.\frac{d^n}{dt^n} g(t)\right|_{t=0} = g^{(n)}(0)$$
$$= \left.\sum_{k=n}^{\infty} \frac{k!\mu_k t^{k-n}}{(k-n)!k!}\right|_{t=0} = \mu_n$$

It is easy to calculate the moment generating function for simple examples.

Example. (Grinstead and Snell) Suppose X has range $\{1, 2, 3, \ldots, n\}$ and $p_X(j) = 1/n$ for $1 \leq j \leq n$ (uniform distribution). Then

$$g(t) = \sum_{j=1}^{n} \frac{1}{n} e^{tj} = \frac{1}{n}\left(e^t + e^{2t} + \cdots + e^{nt}\right) = \frac{e^t\left(e^{nt} - 1\right)}{n\left(e^t - 1\right)}$$

Using the previous equation, we see that

$$\mu_1 = g'(0) = \frac{1}{n}(1 + 2 + 3 + \cdots + n) = \frac{n+1}{2}\mu_2$$
$$= g''(0) = \frac{1}{n}\left(1 + 4 + 9 + \cdots + n^2\right) = \frac{(n+1)(2n+1)}{6},$$

and that $\mu = \mu_1 = \frac{n+1}{2}$ and $\sigma^2 = \mu_2 - \mu_1^2 = \frac{n^2-1}{12}$.

Moment Generating Functions will be crucial in the upcoming proof of the Central Limit Theorem. Thus, we compute the MGF of the Normal/Gaussian distribution.

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

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

Step 2 follows from substituting a normal random variable for Z and taking the expected value, and step 6 uses the fact that the distribution function is equal to one.

3. Probability Bounds

As determining probabilities of certain outcomes of a random variable become more difficult only with mean and variance, exact probabilities are impossible to find, but bounds can be calculated. One of these 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 $a > 0$,*

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

Proof. We have

$$\begin{aligned}
\mathbb{E}[X] &= \int_0^{\infty} xP(x)dx \\
&= \int_0^{u} xP(x)dx + \int_a^{\infty} xP(x)dx.
\end{aligned}$$

Since $P(X)$ and X are nonnegative, we have

$$\begin{aligned}
\mathbb{E}[X] &= \int_0^{u} xP(x)dx + \int_a^{\infty} xP(x)dx \\
&\geq \int_a^{\infty} xP(x)dx \\
&\geq \int_a^{\infty} aP(x)dx \\
&= a\int_a^{\infty} P(x)dx \\
&= aP(x \geq a)
\end{aligned}.$$

■

With Markov's Inequality, we can create a bound on the probability given only the mean. It will help us prove Chebyshev's Inequality, leading us closer to the Law of Large Numbers.

Lemma 3.2. *(Chebyshev's Inequality) If X is a discrete random variable with finite mean and variance, for any $a > 0$,*

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

Proof. Let $Y = (X - \mathbb{E}[X])^2$. Then Y is a nonnegative random variable with expected value $\mathbb{E}[Y] = Var(X)$. By Markov's inequality,

$$P\left(Y \geq a^2\right) \leq \frac{\mathbb{E}[Y]}{a} = \frac{Var(X)}{a^2}.$$

Notice that the event $Y \geq a^2$ is the same as $|X - \mathbb{E}[X]| \geq a$, so we can conclude that

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

■

Theorem 3.3. *(The Weak Law of Large Numbers) Let $X_1, X_2, \ldots$ 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 $\epsilon > 0$,*

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

Proof. We see that

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

Furthermore,

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

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

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

Let $\left(\frac{X_1 + \ldots + X_n}{n}\right)$ be a new random variable X. It clearly satisfies the conditions for Chebyshev's Inequality. Hence, we apply the lemma to get

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

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

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

■

The Law of Large Numbers essentially states that as a sample size grows, its mean gets closer to the average of the whole population

4. Central Limit Theorem

In essence, the theorem says that as the sample size expands, the distribution of the mean among multiple samples will come closer and closer to the Gaussian distribution, commonly known as a bell curve. What's astonishing is that no matter the initial distribution, the normal distribution arises. That means that we can use the CLT to examine things like income, height, weight, and political forecasts. Investors of all types rely on the CLT to analyze stock returns, construct portfolios, and manage risk. One just needs 30 random samples for the CLT to hold, and one can have as many pieces of data as one wants. Furthermore, previously-selected stocks must be swapped out with different names, to help eliminate bias. Although this concept was first developed by Abraham de Moivre in 1733, it wasn't formally named until 1930, when noted Hungarian mathematician George Polya officially dubbed it the Central Limit Theorem. Note that the LLN is different from the CLT, as the LLN relies on the size of a single sample, while the CLT relies on the number of samples. Furthermore, the CLT shows the the distribution of the difference between the sample means and the value, rather than showing that the sample means converge to a single value. Now, we will prove it. First, note the following lemma.

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} such that $n \geq 1$. 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.*

The proof of this lemma is left as an exercise for the reader. To see its importance, let $Z_n = \sum_{i=1}^{n} \frac{X_i}{\sqrt{n}}$ where X_i are i.i.d random variables, and let Z be a normal random variable. Then, if we show that the MGF of Z_n approaches the MGF of Z (Which we previously showed is $e^{t^2/2}$) as $n \to \infty$. we are also showing that the probability distribution of Z_n approaches the normal distribution as $n \to \infty$. Consequently, we are going to prove the Central Limit Theorem as such.

Theorem 4.2. *(Central Limit Theorem) Let $S_n = X_1 + X_2 + ... + X_n$ be the sum of n i.i.d random variables with expected value μ and variance σ^2. Then, for $-\infty < a < \infty$,*

$$\lim_{n\to\infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq a\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2} dx.$$

Proof. (Krokhmal) 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 $M\left(\frac{t}{\sqrt{n}}\right) = \mathbb{E}\left[\exp\left(\frac{tX_i}{\sqrt{n}}\right)\right]$. However, we are interested in the MGF of $\mathbb{E}\left[\exp\left(t\sum_{i=1}^{n}\frac{X_i}{\sqrt{n}}\right)\right]$. Here is how we find it:

$$\begin{aligned}
\mathbb{E}\left[\exp\left(t\sum_{i=1}^{n}\frac{X_i}{\sqrt{n}}\right)\right] &= \mathbb{E}\left[\exp\left(\sum_{i=1}^{n}X_i\cdot\frac{t}{\sqrt{n}}\right)\right] \\
&= \mathbb{E}\left[\prod_{i=1}^{n}\exp\left(X_i\cdot\frac{t}{\sqrt{n}}\right)\right] \\
&= \prod_{i=1}^{n}\mathbb{E}\left[\exp\left(X_i\cdot\frac{t}{\sqrt{n}}\right)\right] \\
&= \prod_{i=1}^{n}M\left(\frac{t}{\sqrt{n}}\right) \\
&= \left[M\left(\frac{t}{\sqrt{n}}\right)\right]^n
\end{aligned}$$

Step 3 relies on the fact that the X_i s are independent and therefore the $\mathbb{E}$ and the $\prod$ operators are interchangeable. Now we define L(t) = log M(t) and evaluate $L(0), L'(0), L''(0)$. We have

$$\begin{aligned}
&L(0) = \log M(0) = \log\mathbb{E}\left[e^{0\cdot X_i}\right] = \log\mathbb{E}[1] = \log 1 = 0 \\
&L'(0) = \tfrac{M'(0)}{M(0)} = M'(0) = \mathbb{E}\left[Xe^{0\cdot X_i}\right] = \mathbb{E}[X] = \mu = 0
\end{aligned}$$

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

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

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

Here, step 1 and 3 are justified by L'Hopital's rule since both the top and the bottom of the original fraction equaled 0. Having shown this, we can now apply the lemma to prove the Central limit theorem for the case where $\mu = 0$ and $\sigma^2 = 1$.

■

5. Examples and Applications

In this section, we will delve into the usefulness of the theorem, discussing how to use it using sample problems.

Example. A die is rolled 420 times. What is the probability that the sum of the rolls lies between 1400 and 1550? The sum is a random variable

$$S_{420} = X_1 + X_2 + \cdots + X_{420}$$

where each X_j has distribution

$$m_X = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{pmatrix}$$

We have seen that $\mu = E(X) = 7/2$ and $\sigma^2 = V(X) = 35/12$. Thus, $E(S_{420}) = 420 \cdot 7/2 = 1470, \sigma^2(S_{420}) = 420 \cdot 35/12 = 1225$, and $\sigma(S_{420}) = 35$. Therefore

$$\begin{aligned} P(1400 \leq S_{420} \leq 1550) &\approx P\left(\frac{1399.5 - 1470}{35} \leq S^*_{240} \leq \frac{1550.5 - 1470}{35}\right) \\ &= P(-2.01 \leq S^*_{420} \leq 2.30) \\ &\approx \text{NA}(-2.01, 2.30) = .9670 \end{aligned}$$

We'll now look at grading, which at a first glance, might seem completely unrelated to the normal distribution.

Example. A student's GPA is the average of his grades in 30 courses. The grades are based on 100 possible points and are recorded as integers. Assume that, in each course, the instructor makes an error in grading of k with probability $|p/k|$, where $k = \pm 1, \pm 2, \pm 3, \pm 4, \pm 5$. The probability of no error is then $1 - (137/30)p$. (The parameter p represents the inaccuracy of the instructor's grading.) Thus, in each course, there are two grades for the student, namely the "correct" grade and the recorded grade. So there are two average grades for the student, namely the average of the correct grades and the average of the recorded grades.

We wish to estimate the probability that these two average grades differ by less than .05 for a given student. We now assume that $p = 1/20$. We also assume that the total error is the sum S_{30} of 30 independent random variables each with distribution

$$m_X : \left\{ \begin{array}{ccccccccccc} -5 & -4 & -3 & -2 & -1 & 0 & 1 & 2 & 3 & 4 & 5 \\ \frac{1}{100} & \frac{1}{80} & \frac{1}{60} & \frac{1}{40} & \frac{1}{20} & \frac{463}{600} & \frac{1}{20} & \frac{1}{40} & \frac{1}{60} & \frac{1}{80} & \frac{1}{100} \end{array} \right\}$$

One can easily calculate that $E(X) = 0$ and $\sigma^2(X) = 1.5$. Then we have

$$\begin{aligned} P\left(-.05 \leq \frac{S_{30}}{30} \leq .05\right) &= P(-1.5 \leq S_{30} \leq 1.5) \\ &= P\left(\frac{-1.5}{\sqrt{30 \cdot 1.5}} \leq S^*_{30} \leq \frac{1.5}{\sqrt{30 \cdot 1.5}}\right) \\ &= P(-.224 \leq S^*_{30} \leq .224) \\ &\approx \text{NA}(-.224, .224) = .1772 \end{aligned}$$

This means that there is only a 17.7% chance that a given student's grade point average is accurate to within .05.

These examples involved discrete independent trials, but the CLT also holds for continous distributions with finite mean and variance.

Example. (Grinstead and Snell) Suppose a surveyor wants to measure a known distance, say of 1 mile. She knows that because of possible motion of her car, atmospheric distortions, and human error, any one measurement is apt to be slightly in error. She plans to make several measurements and take an average. She assumes that her measurements are independent random variables with a common distribution of mean $\mu = 1$ and standard deviation $\sigma = .0002$ (so, if the errors are approximately normally distributed, then her measurements are within 1 foot of the correct distance about 65% of the time). What can she say about the average?

She can say that if n is large, the average S_n/n has a density function that is approximately normal, with mean $\mu = 1$ mile, and standard deviation $\sigma = .0002/\sqrt{n}$ miles.

How many measurements should she make to be reasonably sure that her average lies within .0001 of the true value? Here's where we can use the Law of Large Numbers, which we previously proved. The Chebyshev inequality says

$$P\left(\left|\frac{S_n}{n} - \mu\right| \geq .0001\right) \leq \frac{(.0002)^2}{n\left(10^{-8}\right)} = \frac{4}{n}$$

so that we must have $n \geq 80$ before the probability that his error is less than .0001 exceeds .95.

We have already noticed that the estimate in the Chebyshev inequality is not always a good one, and here is a case in point. If we assume that n is large enough so that the density for S_n is approximately normal, then we have

$$\begin{aligned} P\left(\left|\frac{S_n}{n} - \mu\right| < .0001\right) &= P\left(-.5\sqrt{n} < S_n^* < +.5\sqrt{n}\right) \\ &\approx \frac{1}{\sqrt{2\pi}} \int_{-.5\sqrt{n}}^{+.5\sqrt{n}} e^{-x^2/2} dx \end{aligned}$$

and this last expression is greater than .95 if $.5\sqrt{n} \geq 2$. This says that it suffices to take $n = 16$ measurements for the same results. This second calculation is stronger, but depends on the assumption that $n = 16$ is large enough to establish the normal density as a good approximation to S_n^*, and hence to S_n. The Central limit Theorem here says nothing about how large n has to be. In most cases involving sums of independent random variables, a good rule of thumb is that for $n \geq 30$, the approximation is a good one. In the present case, if we assume that the errors are approximately normally distributed, then the approximation is probably fairly good even for $n = 16$.

Now suppose a different surveyor is measuring an unknown distance with the same instruments under the same conditions. He takes 36 measurements and averages them. How sure can he be that his measurement lies within .0002 of the true value? Again using the normal approximation, we get

$$\begin{aligned} P\left(\left|\frac{S_n}{n} - \mu\right| < .0002\right) &= P\left(\left|S_n^*\right| < .5\sqrt{n}\right) \\ &\approx \frac{2}{\sqrt{2\pi}} \int_{-3}^{3} e^{-x^2/2} dx \\ &\approx .997 \end{aligned}$$

This means that the surveyor can be 99.7 percent sure that his average is within .0002 of the true value. To improve his confidence, he can take more measurements, or require less accuracy, or improve the quality of his measurements (i.e., reduce the variance σ^2). In each case, the Central limit Theorem gives quantitative information about the confidence of a measurement process, assuming always that the normal approximation is valid.

Even if he doesn't know the mean or variance, he can make several measurements of a known distance and average them. As before, the average error is approximately normally distributed, but now with unknown mean and variance.

We have seen thus how the Central Limit Theorem has numerous applications, given a variety of parameters. Here are a few exercises for the reader.

Exercises

(1) A random walkers starts at 0 on the x-axis, and every second, they move 1 step to right or 1 step to the left with probability $\frac{1}{2}$. Estimate the probability that, after 400 steps, the walker is more than 16 steps from the start.
(2) The probabilities of error are probably too low in the example concerning GPAs. Find a more reasonable estimate for $m(x)$, and determine the probability that the GPA is accurate to within 0.05, 0.1, and 0.5.
(3) Prove Lemma 4.1.
(4) A surveying instrument makes an error of $-5, -1, 0, 1$, or 5 feet with equal probabilities when measuring the height of a 100 foot tower. Find the expected value and variance for the height obtained using this instrument once. Estimate the probability that in 41 independent measurements, the average of the measurements is between 99 and 101, inclusive.
(5) The price of a share of stock at the Markov Book Store is given by A_n on the nth day of the year. The differences $X_n = A_{n+1} - A_n$ are i.i.d. random variables with $\mu = 0$ and $\sigma^2 = \frac{1}{4}$. If $A_1 = 100$, estimate the probability that A_365 is greater than 100, greater than 110, and greater than 150.

6. References

[1] Introductory Probability and the Central Limit Theorem by Vlad Krokhmal, 7/29/2011.
[2] Introduction to Probability by Charles M. Grinstead and J. Laurie Snell. Second Revised Edition.