7.0.0 Limit Theorems

In this chapter, we will discuss limit theorems and convergence modes for random variables. Limit theorems are among the most fundamental results in probability theory. We will discuss two important limit theorems in Section 7.1: *the law of large numbers (LLN)* and *the central limit theorem (CLT)*. We will also talk about the importance of these theorems as applied in practice. In Section 7.2, we will discuss the convergence of sequences of random variables.

7.1.0 Limit Theorems

In this section, we will discuss two important theorems in probability, *the law of large numbers (LLN)* and *the central limit theorem (CLT)*. The LLN basically states that the average of a large number of i.i.d. random variables converges to the expected value. The CLT states that, under some conditions, the sum of a large number of random variables has an approximately normal distribution.

7.1.1 Law of Large Numbers

The **law of large numbers** has a very central role in probability and statistics. It states that if you repeat an experiment independently a large number of times and average the result, what you obtain should be close to the expected value. There are two main versions of the law of large numbers. They are called the **weak** and **strong** laws of the large numbers. The difference between them is mostly theoretical. In this section, we state and prove the weak law of large numbers (WLLN). The strong law of large numbers is discussed in Section 7.2. Before discussing the WLLN, let us define the *sample mean*.

Definition 7.1. For i.i.d. random variables $X_1, X_2, \ldots, X_n$, the **sample mean**, denoted by $\overline{X}$, is defined as

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}.$$

Another common notation for the sample mean is M_n. If the X_i's have CDF $F_X(x)$, we might show the sample mean by $M_n(X)$ to indicate the distribution of the X_i's.

Note that since the X_i's are random variables, the sample mean, $\overline{X} = M_n(X)$, is also a random variable. In particular, we have

$$
\begin{aligned}
E[\overline{X}] &= \frac{EX_1 + EX_2 + \ldots + EX_n}{n} &&\text{(by linearity of expectation)} \\
&= \frac{nEX}{n} &&\text{(since } EX_i = EX) \\
&= EX.
\end{aligned}
$$

Also, the variance of $\overline{X}$ is given by

$$\begin{aligned}
\mathrm{Var}(\overline{X}) &= \frac{\mathrm{Var}(X_1 + X_2 + \ldots + X_n)}{n^2} && (\text{since } \mathrm{Var}(aX) = a^2 \mathrm{Var}(X)) \\
&= \frac{\mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \ldots + \mathrm{Var}(X_n)}{n^2} && (\text{since the } X_i\text{'s are independent}) \\
&= \frac{n\mathrm{Var}(X)}{n^2} && (\text{since } \mathrm{Var}(X_i) = \mathrm{Var}(X)) \\
&= \frac{\mathrm{Var}(X)}{n}.
\end{aligned}$$

Now let us state and prove the **weak law of large numbers (WLLN).**

The weak law of large numbers (WLLN)

Let X_1, X_2 , ... , X_n be i.i.d. random variables with a finite expected value $EX_i = \mu < \infty$. Then, for any $\epsilon > 0$,

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

Proof

The proof of the weak law of large number is easier if we assume $\mathrm{Var}(X) = \sigma^2$ is finite. In this case we can use Chebyshev's inequality to write

$$\begin{aligned}
P(|\overline{X} - \mu| \geq \epsilon) &\leq \frac{\mathrm{Var}(\overline{X})}{\epsilon^2} \\
&= \frac{\mathrm{Var}(X)}{n\epsilon^2},
\end{aligned}$$

which goes to zero as $n \to \infty$.

7.1.2 Central Limit Theorem

The **central limit theorem (CLT)** is one of the most important results in probability theory. It states that, under certain conditions, the sum of a large number of random variables is approximately normal. Here, we state a version of the CLT that applies to i.i.d. random variables. Suppose that X_1, X_2 , ... , X_n are i.i.d. random variables with expected values $EX_i = \mu < \infty$ and variance $\mathrm{Var}(X_i) = \sigma^2 < \infty$. Then as we saw above, the sample mean $\overline{X} = \frac{X_1+X_2+...+X_n}{n}$ has mean $E\overline{X} = \mu$ and variance $\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}$. Thus, the normalized random variable

$$Z_n = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{X_1 + X_2+...+X_n - n\mu}{\sqrt{n}\sigma}$$

has mean $EZ_n = 0$ and variance $\mathrm{Var}(Z_n) = 1$. The central limit theorem states that the CDF of Z_n converges to the standard normal CDF.

The Central Limit Theorem (CLT)

Let $X_1, X_2,..., X_n$ be i.i.d. random variables with expected value $EX_i = \mu < \infty$ and variance $0 < \mathrm{Var}(X_i) = \sigma^2 < \infty$. Then, the random variable

$$Z_n = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{X_1 + X_2+...+X_n - n\mu}{\sqrt{n}\sigma}$$

converges in distribution to the standard normal random variable as n goes to infinity, that is

$$\lim_{n\to\infty} P(Z_n \leq x) = \Phi(x), \qquad \text{for all } x \in \mathbb{R},$$

where $\Phi(x)$ is the standard normal CDF.

An interesting thing about the CLT is that it does not matter what the distribution of the X_i's is. The X_i's can be discrete, continuous, or mixed random variables. To get a

feeling for the CLT, let us look at some examples. Let's assume that X_i's are $Bernoulli(p)$. Then $EX_i = p$, $\text{Var}(X_i) = p(1-p)$. Also, $Y_n = X_1 + X_2 + \ldots + X_n$ has $Binomial(n, p)$ distribution. Thus,

$$Z_n = \frac{Y_n - np}{\sqrt{np(1-p)}},$$

where $Y_n \sim Binomial(n, p)$. Figure 7.1 shows the PMF of Z_n for different values of n. As you see, the shape of the PMF gets closer to a normal PDF curve as n increases. Here, Z_n is a discrete random variable, so mathematically speaking it has a PMF not a PDF. That is why the CLT states that the CDF (not the PDF) of Z_n converges to the standard normal CDF. Nevertheless, since PMF and PDF are conceptually similar, the figure is useful in visualizing the convergence to normal distribution.

PMF of Z_1

$$Z_1 = \frac{X_1 - p}{\sqrt{p(1-p)}}$$

Assumptions:
- $X_1, X_2 \ldots$ are iid Bernoulli(p).
- $Z_n = \dfrac{X_1 + X_2 + \ldots + X_n - np}{\sqrt{np(1-p)}}$.

We choose $p = \frac{1}{3}$.

PMF of Z_2

$$Z_2 = \frac{X_1 + X_2 - 2p}{\sqrt{2p(1-p)}}$$

PMF of Z_3

$$Z_3 = \frac{X_1 + X_2 + X_3 - 3p}{\sqrt{3p(1-p)}}$$

PMF of Z_{30}

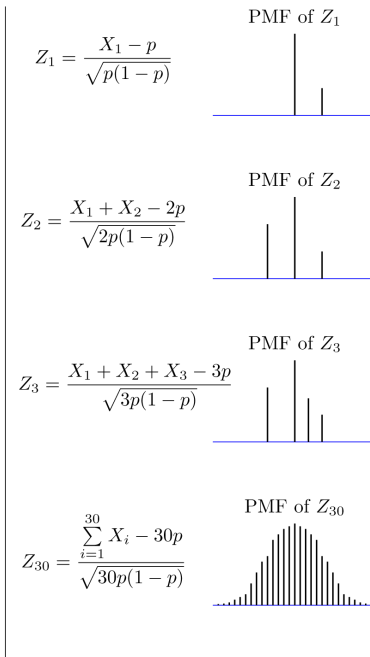$$Z_{30} = \frac{\sum\limits_{i=1}^{30} X_i - 30p}{\sqrt{30p(1-p)}}$$

Fig.7.1 - Z_n is the normalized sum of n independent $Bernoulli(p)$ random variables. The shape of its PMF, $P_{Z_n}(z)$, resembles the normal curve as n increases.

As another example, let's assume that X_i's are $Uniform(0,1)$. Then $EX_i = \frac{1}{2}$, $\mathrm{Var}(X_i) = \frac{1}{12}$. In this case,

$$Z_n = \frac{X_1 + X_2 + \ldots + X_n - \frac{n}{2}}{\sqrt{n/12}}.$$

Figure 7.2 shows the PDF of Z_n for different values of n. As you see, the shape of the PDF gets closer to the normal PDF as n increases.

$$Z_1 = \frac{X_1 - \frac{1}{2}}{\sqrt{\frac{1}{12}}}$$

PDF of Z_1

Assumptions:
- $X_1, X_2 \ldots$ are iid Uniform(0,1).
- $Z_n = \frac{X_1 + X_2 + \ldots + X_n - \frac{n}{2}}{\sqrt{\frac{n}{12}}}.$

$$Z_2 = \frac{X_1 + X_2 - 1}{\sqrt{\frac{2}{12}}}$$

PDF of Z_2

$$Z_3 = \frac{X_1 + X_2 + X_3 - \frac{2}{3}}{\sqrt{\frac{3}{12}}}$$

PDF of Z_3

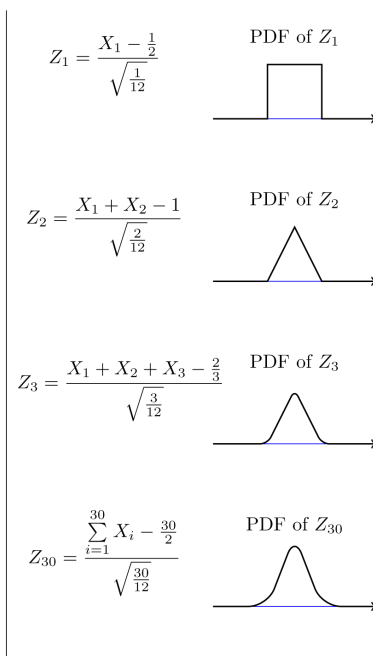$$Z_{30} = \frac{\sum_{i=1}^{30} X_i - \frac{30}{2}}{\sqrt{\frac{30}{12}}}$$

PDF of Z_{30}

Fig. 7.2 - Z_n is the normalized sum of n independent $Uniform(0,1)$ random variables. The shape of its PDF, $f_{Z_n}(z)$, gets closer to the normal curve as n increases.

We could have directly looked at $Y_n = X_1 + X_2 + \ldots + X_n$, so why do we normalize it first and say that the normalized version (Z_n) becomes approximately normal? This is because $EY_n = nEX_i$ and $\mathrm{Var}(Y_n) = n\sigma^2$ go to infinity as n goes to infinity. We normalize Y_n in order to have a finite mean and variance ($EZ_n = 0$, $\mathrm{Var}(Z_n) = 1$). Nevertheless, for any fixed n, the CDF of Z_n is obtained by scaling and shifting the CDF of Y_n. Thus, the two CDFs have similar shapes.

The importance of the central limit theorem stems from the fact that, in many real applications, a certain random variable of interest is a sum of a large number of independent random variables. In these situations, we are often able to use the CLT to justify using the normal distribution. Examples of such random variables are found in almost every discipline. Here are a few:

- Laboratory measurement errors are usually modeled by normal random variables.

- In communication and signal processing, Gaussian noise is the most frequently used model for noise.

- In finance, the percentage changes in the prices of some assets are sometimes modeled by normal random variables.

- When we do random sampling from a population to obtain statistical knowledge about the population, we often model the resulting quantity as a normal random variable.

The CLT is also very useful in the sense that it can simplify our computations significantly. If you have a problem in which you are interested in a sum of one thousand i.i.d. random variables, it might be extremely difficult, if not impossible, to find the distribution of the sum by direct calculation. Using the CLT we can immediately write the distribution, if we know the mean and variance of the X_i's.

Another question that comes to mind is how large n should be so that we can use the normal approximation. The answer generally depends on the distribution of the X_is. Nevertheless, as a rule of thumb it is often stated that if n is larger than or equal to 30, then the normal approximation is very good.

Let's summarize how we use the CLT to solve problems:

<div style="border:1px solid;">

How to Apply The Central Limit Theorem (CLT)

Here are the steps that we need in order to apply the CLT:

1. Write the random variable of interest, Y, as the sum of n i.i.d. random variable X_i's:

$$Y = X_1 + X_2 + \ldots + X_n.$$

2. Find EY and $\text{Var}(Y)$ by noting that

$$EY = n\mu, \qquad \text{Var}(Y) = n\sigma^2,$$

where $\mu = EX_i$ and $\sigma^2 = \text{Var}(X_i)$.

3. According to the CLT, conclude that $\frac{Y-EY}{\sqrt{\text{Var}(Y)}} = \frac{Y-n\mu}{\sqrt{n}\sigma}$ is approximately standard normal; thus, to find $P(y_1 \leq Y \leq y_2)$, we can write

$$P(y_1 \leq Y \leq y_2) = P\left(\frac{y_1 - n\mu}{\sqrt{n}\sigma} \leq \frac{Y - n\mu}{\sqrt{n}\sigma} \leq \frac{y_2 - n\mu}{\sqrt{n}\sigma}\right)$$

$$\approx \Phi\left(\frac{y_2 - n\mu}{\sqrt{n}\sigma}\right) - \Phi\left(\frac{y_1 - n\mu}{\sqrt{n}\sigma}\right).$$

</div>

Let us look at some examples to see how we can use the central limit theorem.

Example 7.1

A bank teller serves customers standing in the queue one by one. Suppose that the service time X_i for customer i has mean $EX_i = 2$ (minutes) and $\text{Var}(X_i) = 1$. We assume that service times for different bank customers are independent. Let Y be the total time the bank teller spends serving 50 customers. Find $P(90 < Y < 110)$.

Solution

Example 7.2

In a communication system each data packet consists of 1000 bits. Due to the noise, each bit may be received in error with probability 0.1. It is assumed bit errors occur independently. Find the probability that there are more than 120 errors in a certain data packet.

Continuity Correction:

Let us assume that $Y \sim Binomial(n = 20, p = \frac{1}{2})$, and suppose that we are interested in $P(8 \leq Y \leq 10)$. We know that a $Binomial(n = 20, p = \frac{1}{2})$ can be written as the sum of n i.i.d. $Bernoulli(p)$ random variables:

$$Y = X_1 + X_2 + \ldots + X_n.$$

Since $X_i \sim Bernoulli(p = \frac{1}{2})$, we have

$$EX_i = \mu = p = \frac{1}{2}, \qquad \text{Var}(X_i) = \sigma^2 = p(1 - p) = \frac{1}{4}.$$

Thus, we may want to apply the CLT to write

$$
\begin{aligned}
P(8 \leq Y \leq 10) &= P\left(\frac{8 - n\mu}{\sqrt{n}\sigma} < \frac{Y - n\mu}{\sqrt{n}\sigma} < \frac{10 - n\mu}{\sqrt{n}\sigma} \right) \\
&= P\left(\frac{8 - 10}{\sqrt{5}} < \frac{Y - n\mu}{\sqrt{n}\sigma} < \frac{10 - 10}{\sqrt{5}} \right) \\
&\approx \Phi(0) - \Phi\left(\frac{-2}{\sqrt{5}} \right) \\
&= 0.3145
\end{aligned}
$$

Since, here, $n = 20$ is relatively small, we can actually find $P(8 \leq Y \leq 10)$ accurately. We have

$$
\begin{aligned}
P(8 \leq Y \leq 10) &= \sum_{k=8}^{10} \binom{n}{k} p^k (1 - p)^{n-k} \\
&= \left[\binom{20}{8} + \binom{20}{9} + \binom{20}{10} \right] \left(\frac{1}{2} \right)^{20} \\
&= 0.4565
\end{aligned}
$$

We notice that our approximation is not so good. Part of the error is due to the fact that Y is a discrete random variable and we are using a continuous distribution to find $P(8 \leq Y \leq 10)$. Here is a trick to get a better approximation, called **continuity correction**. Since Y can only take integer values, we can write

$$
\begin{aligned}
P(8 \leq Y \leq 10) &= P(7.5 < Y < 10.5) \\
&= P\left(\frac{7.5 - n\mu}{\sqrt{n}\sigma} < \frac{Y - n\mu}{\sqrt{n}\sigma} < \frac{10.5 - n\mu}{\sqrt{n}\sigma} \right) \\
&= P\left(\frac{7.5 - 10}{\sqrt{5}} < \frac{Y - n\mu}{\sqrt{n}\sigma} < \frac{10.5 - 10}{\sqrt{5}} \right) \\
&\approx \Phi\left(\frac{0.5}{\sqrt{5}} \right) - \Phi\left(\frac{-2.5}{\sqrt{5}} \right) \\
&= 0.4567
\end{aligned}
$$

As we see, using continuity correction, our approximation improved significantly. The continuity correction is particularly useful when we would like to find $P(y_1 \leq Y \leq y_2)$, where Y is binomial and y_1 and y_2 are close to each other.

Continuity Correction for Discrete Random Variables

Let $X_1, X_2, \cdots, X_n$ be independent discrete random variables and let

$$
Y = X_1 + X_2 + \cdots + X_n.
$$

Suppose that we are interested in finding $P(A) = P(l \leq Y \leq u)$ using the CLT, where l and u are integers. Since Y is an integer-valued random variable, we can write

$$
P(A) = P(l - \frac{1}{2} \leq Y \leq u + \frac{1}{2}).
$$

It turns out that the above expression sometimes provides a better approximation for $P(A)$ when applying the CLT. This is called the continuity correction and it is particularly useful when X_i's are Bernoulli (i.e., Y is binomial).