

---

## 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  $X_1, X_2, \dots, X_n$  are i.i.d. random variables with expected values  $EX_i = \mu < \infty$  and variance  $\text{Var}(X_i) = \sigma^2 < \infty$ . Then as we saw above, the sample mean  $\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$  has mean  $E\bar{X} = \mu$  and variance  $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$ . Thus, the normalized random variable

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

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

### The Central Limit Theorem (CLT)

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

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

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

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

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,  $Y_n = X_1 + X_2 + \dots + X_n$  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.

Assumptions:

- $X_1, X_2, \dots$  are iid  $Bernoulli(p)$ .
- $Z_n = \frac{X_1 + X_2 + \dots + X_n - np}{\sqrt{np(1-p)}}$ .

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

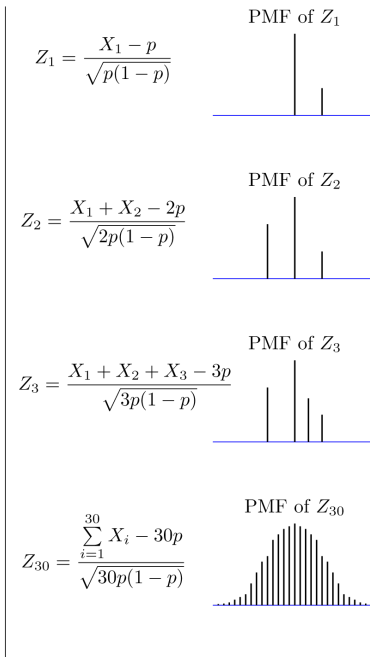


Fig.7.1 -  $Z_n$  is the normalized sum of  $n$  independent  $Bernoulli(p)$  random variables. The shape of its PMF,  $P_{Z_n}(z)$ , resembles the normal curve as  $n$  increases.

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

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

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.

Assumptions:

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

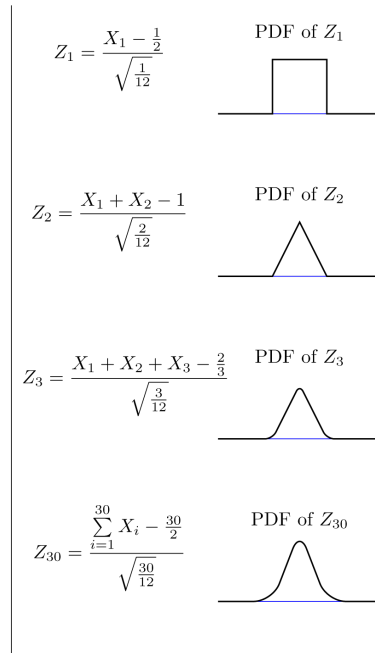


Fig. 7.2 -  $Z_n$  is the normalized sum of  $n$  independent  $\text{Uniform}(0,1)$  random variables. The shape of its PDF,  $f_{Z_n}(z)$ , gets closer to the normal curve as  $n$  increases.

We could have directly looked at  $Y_n = X_1 + X_2 + \dots + X_n$ , 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  $\text{Var}(Y_n) = n\sigma^2$  go to infinity as  $n$  goes to infinity. We normalize  $Y_n$  in order to have a finite mean and variance ( $EZ_n = 0$ ,  $\text{Var}(Z_n) = 1$ ). 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  $X_i$ 's. 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:

## 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:

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

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

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

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

$$\begin{aligned} P(y_1 \leq Y \leq y_2) &= P\left(\frac{y_1 - n\mu}{\sqrt{n}\sigma} \leq \frac{Y - n\mu}{\sqrt{n}\sigma} \leq \frac{y_2 - n\mu}{\sqrt{n}\sigma}\right) \\ &\approx \Phi\left(\frac{y_2 - n\mu}{\sqrt{n}\sigma}\right) - \Phi\left(\frac{y_1 - n\mu}{\sqrt{n}\sigma}\right). \end{aligned}$$

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

$$Y = X_1 + X_2 + \dots + X_n,$$

where  $n = 50$ ,  $EX_i = \mu = 2$ , and  $\text{Var}(X_i) = \sigma^2 = 1$ . Thus, we can write

$$\begin{aligned}
P(90 < Y \leq 110) &= P\left(\frac{90 - n\mu}{\sqrt{n}\sigma} < \frac{Y - n\mu}{\sqrt{n}\sigma} < \frac{110 - n\mu}{\sqrt{n}\sigma}\right) \\
&= P\left(\frac{90 - 100}{\sqrt{50}} < \frac{Y - n\mu}{\sqrt{n}\sigma} < \frac{110 - 100}{\sqrt{50}}\right) \\
&= P\left(-\sqrt{2} < \frac{Y - n\mu}{\sqrt{n}\sigma} < \sqrt{2}\right).
\end{aligned}$$

By the CLT,  $\frac{Y - n\mu}{\sqrt{n}\sigma}$  is approximately standard normal, so we can write

$$\begin{aligned}
P(90 < Y \leq 110) &\approx \Phi(\sqrt{2}) - \Phi(-\sqrt{2}) \\
&= 0.8427
\end{aligned}$$


---

### 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.

#### Solution

Let us define  $X_i$  as the indicator random variable for the  $i$ th bit in the packet. That is,  $X_i = 1$  if the  $i$ th bit is received in error, and  $X_i = 0$  otherwise. Then the  $X_i$ 's are i.i.d. and  $X_i \sim \text{Bernoulli}(p = 0.1)$ . If  $Y$  is the total number of bit errors in the packet, we have

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

Since  $X_i \sim \text{Bernoulli}(p = 0.1)$ , we have

$$EX_i = \mu = p = 0.1, \quad \text{Var}(X_i) = \sigma^2 = p(1 - p) = 0.09$$

Using the CLT, we have

$$\begin{aligned}
P(Y > 120) &= P\left(\frac{Y - n\mu}{\sqrt{n}\sigma} > \frac{120 - n\mu}{\sqrt{n}\sigma}\right) \\
&= P\left(\frac{Y - n\mu}{\sqrt{n}\sigma} > \frac{120 - 100}{\sqrt{90}}\right) \\
&\approx 1 - \Phi\left(\frac{20}{\sqrt{90}}\right) \\
&= 0.0175
\end{aligned}$$


---

### Continuity Correction:

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

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

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

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

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

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

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

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

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).