

Central Limit Theorem (CLT)

Statement: Let $X_1, X_2, X_3, \dots$ be a sequence of independent and identically distributed random variables each with finite mean μ and variance σ^2 . Then the distribution of $Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$ tends to standard normal distribution as $n \rightarrow \infty$.

That is, $P(Z_n \leq z) \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{u^2}{2}} du$ as $n \rightarrow \infty$

Proof:

Consider Moment generating function of Z_n is

$$M_{Z_n}(t) = M_{\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}}(t)$$

$$= E\left[e^{\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \cdot t}\right] = e^{-\frac{\mu t}{\sigma/\sqrt{n}}} \cdot E\left[e^{\frac{\bar{X}_n}{\sigma/\sqrt{n}} t}\right]$$

$$= e^{-\frac{\sqrt{n} \mu t}{\sigma}} \cdot E\left(e^{t \frac{\sqrt{n} \bar{X}_n}{\sigma}}\right)$$

$$= e^{-\frac{\sqrt{n} \mu t}{\sigma}} \cdot M_{\bar{X}_n}\left(\frac{\sqrt{n} t}{\sigma}\right)$$

$$\left[\because M_X(t) = E(e^{tX}) \right]$$

$$\text{Now, } \bar{x}_n = \frac{x_1 + x_2 + \dots + x_n}{n}$$

(2)

$$\Rightarrow n\bar{x}_n = x_1 + x_2 + \dots + x_n$$

$\therefore$ All $x_1, x_2, \dots, x_n$ are iid random variables

$$\therefore M_{n\bar{x}_n}(t) = \left(M_x(t) \right)^n$$

$$\Rightarrow M_{n\bar{x}_n}\left(\frac{t}{\sigma\sqrt{n}}\right) = \left(M_x\left(\frac{t}{\sigma\sqrt{n}}\right) \right)^n \quad \text{--- (2)}$$

Here x have the same distribution with all x_i 's.

$$\text{Now, } \cancel{M_{\bar{x}_n}} M_{\bar{x}_n}\left(\frac{\sqrt{n}t}{\sigma}\right) = E\left(e^{\frac{t\sqrt{n}}{\sigma} \cdot \bar{x}_n}\right)$$

$$= E\left(e^{\frac{t\sqrt{n} \bar{x}_n}{\sigma} \cdot \frac{n}{n}}\right)$$

$$= E\left(e^{\frac{t}{\sigma\sqrt{n}} \cdot n\bar{x}_n}\right)$$

$$= M_{n\bar{x}_n}\left(\frac{t}{\sigma\sqrt{n}}\right) \quad \text{--- (3)}$$

using eqn (3) in eqn (1), we have

$$M_{Z_n}(t) = M_{\frac{\bar{x}_n - \mu}{\sigma/\sqrt{n}}}(t) = e^{-\frac{\sqrt{n}\mu t}{\sigma}} \cdot M_{n\bar{x}_n}\left(\frac{t}{\sigma\sqrt{n}}\right)$$

$$\Rightarrow M_{Z_n}(t) = e^{-\frac{\sqrt{n}\mu t}{\sigma}} \cdot \left(M_x\left(\frac{t}{\sigma\sqrt{n}}\right) \right)^n \quad \left[\text{using eqn (2)} \right]$$

Now taking logarithm to both sides, we have

(3)

$$\ln M_{2n}(t) = -\frac{\sqrt{n} \mu t}{\sigma} + n \ln M_x\left(\frac{t}{\sigma\sqrt{n}}\right)$$

$$\Rightarrow \ln M_{2n}(t) = -\frac{\sqrt{n} \mu t}{\sigma} + n \ln \left\{ E\left(e^{\frac{tX}{\sigma\sqrt{n}}}\right) \right\}$$

$$\Rightarrow \ln M_{2n}(t) = -\frac{\sqrt{n} \mu t}{\sigma} + n \ln \left\{ E\left[1 + \frac{tX}{\sigma\sqrt{n}} + \frac{1}{2!} \left(\frac{tX}{\sigma\sqrt{n}}\right)^2 + \frac{1}{3!} \left(\frac{tX}{\sigma\sqrt{n}}\right)^3 + \dots \right] \right\}$$

$$\left[\because \text{The expansion of } e^x \text{ is } e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \dots \right]$$

$$\therefore \ln M_{2n}(t) = -\frac{\sqrt{n} \mu t}{\sigma} + n \ln \left[E(1) + \frac{t}{\sigma\sqrt{n}} E(X) + \frac{t^2}{2n\sigma^2} E(X^2) + \frac{t^3}{6n^{3/2}\sigma^3} E(X^3) + \dots \right]$$

$$\Rightarrow \ln M_{2n}(t) = -\frac{\sqrt{n} \mu t}{\sigma} + n \ln \left[1 + \frac{t}{\sigma\sqrt{n}} \mu_1' + \frac{t^2}{2n\sigma^2} \mu_2' + \frac{t^3}{6n^{3/2}\sigma^3} \mu_3' + \dots \right]$$

where $\mu_1', \mu_2', \mu_3', \dots$ are the moments about origin. (4)

that is, $\mu_2' = E(X^2)$.

(4)

Now the expansion of $\ln(1+x)$ is

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots$$

using it in eqn (4), we have.

$$\begin{aligned} \ln M_{2n}(t) = & \frac{-\sqrt{n} \mu t}{\sigma} + n \left[\left(\mu_1' \frac{t}{\sigma\sqrt{n}} + \frac{\mu_2' t^2}{2n\sigma^2} + \frac{\mu_3' t^3}{6n^{3/2}\sigma^3} + \dots \right) \right. \\ & - \frac{1}{2} \left(\mu_1' \frac{t}{\sigma\sqrt{n}} + \frac{\mu_2' t^2}{2n\sigma^2} + \frac{\mu_3' t^3}{6n^{3/2}\sigma^3} + \dots \right)^2 \\ & \left. + \frac{1}{3} \left(\mu_1' \frac{t}{\sigma\sqrt{n}} + \frac{\mu_2' t^2}{2n\sigma^2} + \frac{\mu_3' t^3}{6n^{3/2}\sigma^3} + \dots \right)^3 + \dots \right] \end{aligned}$$

$$\begin{aligned} \Rightarrow \ln M_{2n}(t) = & \left(\frac{-\sqrt{n} \mu}{\sigma} + \frac{\sqrt{n} \mu_1'}{\sigma} \right) t + \left(\frac{\mu_2'}{2\sigma^2} - \frac{\mu_1'^2}{2\sigma^2} \right) t^2 \\ & + \left(\frac{\mu_3'}{6\sigma^3\sqrt{n}} - \frac{\mu_1' \mu_2'}{2\sigma^3\sqrt{n}} + \frac{\mu_1'^2}{3\sigma^3\sqrt{n}} \right) t^3 + \dots \end{aligned}$$

$$\therefore \mu_1' = E(X) = \mu$$

$$\text{and } \mu_2' - (\mu_1')^2 = E(X^2) - (E(X))^2 = V(X) = \sigma^2$$

$$\begin{aligned} \therefore \ln M_{2n}(t) = & 0 + \frac{\sigma^2}{2\sigma^2} t^2 + \left(\frac{\mu_3'}{6} - \frac{\mu_1' \mu_2'}{2} + \frac{\mu_1'^2}{3} \right) \frac{t^3}{\sigma^3\sqrt{n}} \\ & + \text{terms containing higher powers of } n^{-1/2}. \end{aligned}$$

Now taking limit $n \rightarrow \infty$ both sides, we have.

(5)

$$\lim_{n \rightarrow \infty} \ln M_{Z_n}(t) = \frac{t^2}{2} + 0 + 0 + \dots$$

$$\Rightarrow \lim_{n \rightarrow \infty} \ln M_{Z_n}(t) = \frac{t^2}{2}$$

$$\Rightarrow \lim_{n \rightarrow \infty} M_{Z_n}(t) = e^{\frac{t^2}{2}}$$

which is the mgf of standard normal distribution

$\therefore$ The Moment generating function is always unique.

$\therefore Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$ tends to standard normal
as $n \rightarrow \infty$.

Example:

The lifetime of a special type of battery is a random variable with mean 40 hours and standard deviation 20 hours. Assuming a stockpile of 25 such batteries, the lifetime of which are independent. Approximate the probability that over 1100 hours of use can be obtained.

Solⁿ let X_i denote the lifetime of the i th battery.
with $E(X_i) = \mu = 40$ & $\sigma = \text{S.D.}(X_i) = 20$

We would like to find $P(X_1 + X_2 + \dots + X_{25} \geq 1100)$.

Now we know that $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$ have standard normal distribution. (8)

$$\therefore P(X_1 + X_2 + \dots + X_{25} \geq 1100)$$

Here $n=25$, $\therefore \bar{X}_n = \frac{X_1 + X_2 + \dots + X_{25}}{25}$

$$\Rightarrow P\left(\frac{X_1 + X_2 + \dots + X_{25} - \mu}{\sigma/\sqrt{25}} \geq \frac{1100 - \mu}{\sigma/\sqrt{25}}\right) \quad \left[\text{Here } n=25\right]$$

$$= P\left(\frac{X_1 + X_2 + \dots + X_{25}}{25} \geq \frac{1100}{25}\right)$$

$$= P(\bar{X}_n \geq 44) = P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \geq \frac{44 - \mu}{\sigma/\sqrt{n}}\right)$$

$$= P\left(Z \geq \frac{44 - 40}{20/\sqrt{25}}\right) \quad \left[\because \text{Here } \mu = 40, \sigma = 20, n = 25\right]$$

$$= P\left(Z \geq \frac{4}{4}\right) = P(Z > 1)$$

$$= 1 - P(Z \leq 1) = 1 - \Phi(1) = 1 - 0.8413$$

$$= 0.1587$$

Q. (2) Consider a sequence of n independent random variables $X_1, X_2, \dots, X_n$ such that $P(X_i = 1) = \theta$, $P(X_i = 0) = 1 - \theta$; $i = 1, 2, 3, \dots, n$. Using CLT, show that the distribution of $X = \frac{\sum_{i=1}^n X_i}{n}$ tends to normal as n tends to infinity.

Solⁿ

(7)

By CLT, we know that

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \text{ tends to standard normal as } n \rightarrow \infty$$

Now given that $P(X_i = 1) = \theta$
& $P(X_i = 0) = 1 - \theta$.

$$\mu = E(X_i) = 1 \cdot \theta + 0 \cdot (1 - \theta) = \theta$$

$$E(X_i^2) = (1)^2 \cdot \theta + (0)^2 \cdot (1 - \theta) = \theta$$

$$\sigma^2 = V(X_i) = E(X_i^2) - (E(X_i))^2 = \theta - \theta^2 = \theta(1 - \theta)$$

$$\therefore \sigma = \sqrt{\theta(1 - \theta)}$$

$$\text{Now, given that } X = \frac{\sum_{i=1}^n X_i}{n}$$

$$\begin{aligned} \therefore E(X) &= \frac{1}{n} E\left(\sum_{i=1}^n X_i\right) = \frac{1}{n} \left\{ E(X_1) + E(X_2) + \dots + E(X_n) \right\} \\ &= \frac{1}{n} \{ \mu + \mu + \dots + \mu \} = \frac{n\mu}{n} = \mu = \theta \end{aligned}$$

$$\& V(X) = V\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} V\left(\sum_{i=1}^n X_i\right)$$

$$= \frac{1}{n^2} \{ V(X_1) + V(X_2) + \dots + V(X_n) \}$$

$$= \frac{1}{n^2} \{ \sigma^2 + \sigma^2 + \dots + \sigma^2 \} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} = \frac{\theta(1 - \theta)}{n}$$

$$\rightarrow \text{S.D. of } X = \frac{\sigma}{\sqrt{n}} = \frac{\sqrt{\theta(1 - \theta)}}{\sqrt{n}}$$

Now, using CLT

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \text{ tends to standard normal as } n \rightarrow \infty$$

$$\therefore \frac{X - \mu}{\sigma/\sqrt{n}} \text{ tends to standard normal as } n \rightarrow \infty$$

$$\Rightarrow X \text{ tends to normal as } n \rightarrow \infty.$$

Q. ③ Show that the probability that the number of heads in 400 throws of a fair coin lies between 180 and 220 is approximately $2\phi(2) - 1$.

Solⁿ let X be the number of heads in 400 throws then X follows binomial distribution with probability of success $\theta = \frac{1}{2}$.

Here $n = 400$

$$\therefore \mu = E(X) = n\theta = (400)\left(\frac{1}{2}\right) = 200$$

$$V(X) = n\theta(1-\theta) = (400)\left(\frac{1}{2}\right)\left(1-\frac{1}{2}\right) = 100$$

$$\sigma = \sqrt{V(X)} = \sqrt{100} = 10$$

$$\text{Now, } P(180 \leq X \leq 220) = P\left(\frac{180 - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{220 - \mu}{\sigma}\right)$$

$$= P\left(\frac{180 - 200}{10} \leq Z \leq \frac{220 - 200}{10}\right)$$

$$= P(-2 \leq Z \leq 2) = \cancel{P(-2 \leq Z \leq 2)} \left(\because Z \text{ is symmetric} \right)$$

(9)

$$= P(-2 \leq Z \leq 0) + P(0 \leq Z \leq 2)$$

$$= 2 P(0 \leq Z \leq 2) \quad \left(\begin{array}{l} \because Z \text{ is symmetric} \\ \therefore P(-2 \leq Z \leq 0) = P(0 \leq Z \leq 2) \end{array} \right)$$

$$= 2 \left\{ P(-\infty \leq Z \leq 2) - P(-\infty \leq Z \leq 0) \right\}$$

$$= 2 \left\{ \Phi(2) - \frac{1}{2} \right\}$$

$$= 2\Phi(2) - 1.$$

Now, the syllabus has been completed. Try all questions of all exercises and ask your doubts.