Central Limit theorem :-

Statement :- $\{X_n\}_{n\geq 1}$ i.i.d. sequence of random variables.

Suppose $E[X_n] = 0 \;\forall n$ & $Var(X_n) = \underset{<\infty}{\sigma^2} \;\forall n$

Then $\dfrac{S_n}{\sigma\sqrt{n}} \xrightarrow{d} X \sim N(0,1)$, where $S_n = \sum_{i=1}^{n} X_i$.

$= n\overline{X}_n$, where $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

(Alternatively, $\dfrac{\sqrt{n}\,\overline{X}_n}{\sigma} \xrightarrow{d} X \sim N(0,1)$).

Outline of the proof :- Try to show that $M_{\frac{\sqrt{n}\,\overline{X}_n}{\sigma}}(t) \longrightarrow$ the MGF of $N(0,1)$ random variable for all t.

Remark:-

i) Don't need to assume $E(X_n) = 0 \; \forall n$.

If $E(X_n) = \mu$, then consider $\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} X \sim N(0,1)$

ii) Here the existence of the second moment is crucial.

For that reason, you cannot apply CLT for t_1 or t_2 distribution.

A few results:- i) $X_n \xrightarrow{P} X \implies X_n \xrightarrow{d} X$.

However, ~~at~~ otherway may NOT be true.

But, if $X_n \xrightarrow{d} c \implies X_n \xrightarrow{P} c$.

is a constant

Slutsky's theorem:-

i) $X_n \xrightarrow{d} X$ & $Y_n \xrightarrow{P} c$
constant

Then 1) $X_n + Y_n \xrightarrow{d} X + c$.

2) $X_n Y_n \xrightarrow{d} cX$.

3) $\dfrac{X_n}{Y_n} \xrightarrow{d} \dfrac{X}{c}$ if $c \neq 0$.

Example:-

$\{X_n\}_{n\geq 1} \sim t_n$. Claim $X_n \xrightarrow{d} X \sim N(0,1)$.

$$X_n \sim t_n = \frac{X}{\sqrt{Y/n}}, \text{ where } X \sim N(0,1),\ Y \sim \chi^2_n,\ X \perp\!\!\!\perp Y.$$

Since $Y \sim \chi^2_n$, we can write $Y = Z_1^2 + \cdots + Z_n^2$, where $Z_i \sim N(0,1)$.
$\forall\, i = 1, \ldots, n$.
$Z_i \perp\!\!\!\perp Z_j\ \forall\, i \neq j$.

So, $\dfrac{Y}{n} = \dfrac{1}{n}\sum_{i=1}^{n} Z_i^2 \xrightarrow[\text{WLLN}]{P} E(Z_1^2) = 1$. Hence, $\sqrt{\dfrac{Y}{n}} \xrightarrow{P} 1$. Now, using (3) of Slutsky's theorem, $\dfrac{X}{\sqrt{Y/n}} \xrightarrow{d} X \sim N(0,1)$

Result:-

$\{X_n\}_{n \geq 1}$ i.i.d. seq. of random variables.

$E(X_n) = \mu \quad \forall n, \quad \& \quad Var(X_n) = \sigma^2 < \infty \quad \forall n.$

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \sim N(0,1) \iff \sqrt{n}(\bar{X}_n - \mu) \sim N(0, \sigma^2).$$

Delta-theorem states that

$$\sqrt{n}\left[g(\bar{X}_n) - g(\mu)\right] \sim N\left(0, \sigma^2 \{g'(\mu)\}^2\right).$$

(We have to assume that the third derivative of g is bounded). (Second ??). (Try to prove by yourself).

Example! — Let $X_1, \ldots, X_n$ i.i.d. $\exp(\theta)$.

$\hookrightarrow$ mean $= \theta$.

$E(X_1) = \theta$ & $Var(X_1) = 2\theta^2$ (check!!).

Hence, by CLT, $$\frac{\sqrt{n}\,(\overline{X}_n - \theta)}{\sqrt{2}\,\theta} \xrightarrow{d} N(0,1).$$

Suppose, we want to know the asymptotic dist^n of

$$\sqrt{n}\left(\frac{1}{\overline{X}_n} - \frac{1}{\theta}\right).$$

Take $g(x) = \frac{1}{x}$, then $g'(x) = -\frac{1}{x^2}$, hence by Delta theorem

$$\sqrt{n}\left(\frac{1}{\overline{X}_n} - \frac{1}{\theta}\right) \sim N\left(0,\ 2\theta^2 \times \frac{1}{\theta^4}\right)$$
$$\sim N\left(0, \frac{2}{\theta^2}\right).$$

Section 8:-

Descriptive Statistics

- Visualization tools / graphical device
- Numerical Methods.

Statistics:- This is a science of collecting, analyzing & interpretating data.

It is a "Mathematical Science" to analyse the data.

A few terminologies:-

Population:- A group of interest.

Sample:– The subset selected from the population.

Parameter:– A "fixed" but unknown quantity.

Numerical Methods:–

i) Measure of central tendency.

Suppose Data $\mathcal{X} = \{x_1, \ldots, x_n\}$.

i) Sample mean $= \frac{1}{n}\sum_{i=1}^{n} x_i \doteq \overline{x_n}$

Remark 1:– $\overline{x}_n$ is NOT robust against the outliers.

Remark 2:– $\overline{x}_n$ is equivariant under ~~at~~ arbitrary affine transformation.

$\mathcal{X} = \{x_1, \dots, x_n\}$

$\cancel{x}\ \mathcal{X}' = \{ax_1 + b, \quad \dots \quad , \quad ax_n + b\}.$

$\bar{x}_n(\mathcal{X}') = a\,\bar{x}_n(\mathcal{X}) + b.$

ii) Sample median

$\mathcal{X} = \{x_1, \dots, x_n\}$

$$\tilde{x}_{med} = x_{\left(\left[\frac{n}{2}\right]+1\right)} \quad \text{if } n \text{ is odd.}$$

$$= \frac{x_{\left(\left[\frac{n}{2}\right]\right)} + x_{\left(\left[\frac{n}{2}\right]+1\right)}}{2} \quad \text{if } n \text{ is even.}$$

Remark 1:- Sample median is robust against the outliers

Remark 2:- Sample median is also equivariant under arbitrary affine transformation.

Remark 3:- Sample median is NOT uniquely defined for multivariate data.

3) **α- trimmed mean:-**

$\mathcal{X} = \{x_1, \dots, x_n\}$

$$\tilde{x}_{\alpha, \text{Mean}} = \frac{1}{n - 2[n\alpha]} \sum_{i=[n\alpha]+1}^{n-[n\alpha]} x_{(i)}.$$

As $\alpha \to 0$, $\tilde{x}_{\alpha, \text{Mean}} \to \overline{x}_n$

As $\alpha \to \frac{1}{2}$, $\tilde{x}_{\alpha, \text{Mean}} \to \tilde{x}_{\text{Med}}$.