## Weak Law of Large number:

- Statement $\Rightarrow$ Let $x_1, x_2, \ldots x_n$ be a sequence of random variables and $\mu_1, \mu_2 \ldots \mu_n$ be their respective expectations and let,

$$B_n = Var(x_1 + x_2 + \cdots + x_n) < \alpha$$

Then, $$P\left\{ \left| \frac{x_1 + x_2 + \cdots + x_n}{n} - \frac{\mu_1 + \mu_2 + \cdots + \mu_n}{n} \right| < \epsilon \right\} > 1 - \eta$$

for all $n > n_0$ where $\epsilon$ and $\eta$ are arbitrary small positive numbers provided $\lim \frac{B_n}{n^2} \to 0$ as $n \to \alpha$

**Note:** If $x_1, x_2 \ldots x_n$ be a sequence of independent random and identically distributed (i.i.d) random variables then $\mu_1 = \mu_2 = \ldots \mu_n = \mu$ (say) and,

$$B_n = Var(x_1 + x_2 + \cdots + x_n)$$
$$= \sum_{i=1}^{n} Var(x_i) = n\sigma_1^2 \quad \text{If } Var(x_i) = \sigma_1^2$$

Then the statement can be written as follows:

Let $x_1, x_2 \ldots x_n$ be a sequence of independent and identically distributed random variables each having finite mean $E[x_i] = \mu$, then for any $\epsilon > 0$,

$$P\left\{ \left| \frac{x_1 + x_2 + \cdots + x_n}{n} - \mu \right| \geq \epsilon \right\} \to 0 \quad \text{as } n \to \alpha$$

**Proof of first statement** $\Rightarrow$

Using chebychev's inequality to the random variable $\left( \frac{x_1 + x_2 + \cdots + x_n}{n} \right)$ we get for any $\epsilon > 0$,

$$P\left\{ \left| \frac{x_1 + x_2 + \cdots + x_n}{n} - E\left( \frac{x_1 + x_2 + \cdots + x_n}{n} \right) \right| < \epsilon \right\} \geq 1 - \frac{B_n}{n^2 \epsilon^2}$$

Since $B_n = var(X_1 + X_2 + \cdots + X_n)$

$$\frac{B_n}{n^2} = \frac{1}{n^2} var(X_1 + X_2 + \cdots + X_n)$$

$$= var\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right)$$

$$\therefore P\left\{\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \frac{E(X_1) + E(X_2) + \cdots + E(X_n)}{n}\right| < \epsilon\right\} \geq 1 - \frac{B_n}{n^2\epsilon^2}$$

$$\Rightarrow P\left\{\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \frac{\mu_1 + \mu_2 + \cdots + \mu_n}{n}\right| < \epsilon\right\} \geq 1 - \frac{B_n}{n^2\epsilon^2}$$

As $n \to \alpha$, $\frac{B_n}{n^2\epsilon^2} \to 0$, from the definition of limit we can say that, for any chosen $\eta > 0$ (arbitrarily small) there exist a positive $n_0 > 0$ such that, for all $n > n_0$,

$$\frac{B_n}{n^2\epsilon^2} < \eta.$$

Hence $-\frac{B_n}{n^2\epsilon^2} > -\eta$

$$\Rightarrow P\left\{\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \frac{\mu_1 + \mu_2 + \cdots + \mu_n}{n}\right| < \epsilon\right\} \geq 1 - \eta$$

for all $n > n_0$, where $\epsilon$ and $\eta$ are arbitrary small positive numbers provided, $\lim_{n \to \alpha} \frac{B_n}{n^2} \to 0$

~~so~~ So, far nothing is assumed about the behaviours of $B_n$ for ~~identifying~~ indefinitely increasing value of n. Since $\epsilon$ is arbitrary, we assume $\frac{B_n}{n^2\epsilon^2} \to 0$ as $n$ become indefinitely large. Thus having chosen two arbitrary small positive numbers $\epsilon$ and $\eta$, number $n_0$ can be found so that the inequality $\frac{B_n}{n^2\epsilon^2} < \eta$ will hold for $n > n_0$. Consequently, we shall have,

$$P\left\{\left|\frac{X_1+X_2+\cdots+X_n}{n} - \frac{\mu_1+\mu_2+\cdots+\mu_n}{n}\right| < \epsilon\right\} \geq 1-\eta.$$

for, all $n > n_0(\epsilon, \eta)$.

<u>Note:</u> Proof of the second statement is directly follows from the first one by replacing,

$$E\left(\frac{X_1+X_2+\cdots+X_n}{n}\right) = \frac{\mu_1+\mu_2+\cdots+\mu_n}{n}$$

$$= \mu$$

and $B_n = Var(X_1+X_2+\cdots+X_n)$

$= n\sigma^2$ [since $x_i$'s are idedependent $Cov(x_i, x_j) = 0 \ \forall \ i \neq j$]

since $x_i$'s are identical, $Var(x_i) = Var(x_j) \ \forall \ i \neq j$ ]

<u>Central Limit Theorem (CLT) :</u>

Statement ⇒ If $X_i\ (i=1,2,\ldots n)$ be independent random variables such that $E(X_i) = \mu_i$ and $Var(X_i) = \sigma_i^2$, then under certain general conditions, the random variable $S_n = X_1 + \cdots + X_n$ is asymptotically normal with mean $\mu$ and s.d $\sigma$ where,

$$\mu = \sum_{i=1}^{n} \mu_i \quad , \quad \sigma = \sum_{i=1}^{n} \sigma_i^2$$

☑ De Moivre's Laplace Theorem : (This is also a CLT)

$$\text{If } x_i = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with Probability } q \end{cases}$$

then the distribution of the random variable $S_n = x_1 + x_2 + \cdots + x_n$, where $x_i$'s are independent, is asymptotically normal as $n \to \infty$.

<u>Proof</u> : We prove this theorem using moment generating function.
Let $M_{x_i}(t)$ be the moment generating function of $x_i$.

Then,

$$M_{x_i}(t) = E(e^{tx_i})$$
$$= e^{t.1}.p + e^{t.0}.q$$
$$= p.e^t + q$$
$$= q + p.e^t \longrightarrow (1)$$ [this is true for all $i = 1, 2 .. n$]

Thus

$$M_{S_n}(t) = M_{x_1+x_2+\cdots+x_n}(t)$$
$$= M_{x_1}(t).M_{x_2}(t) \cdots M_{x_n}(t)$$
$$= (q + pe^t)^n \longrightarrow (2)$$ [Since $M_{x_1+x_2}(t) = M_{x_1}(t).M_{x_2}(t)$]

(eqn 2)
This is the moment generating function of Binomial random variable with parameter $n, p$. From the uniqueness of M.G.F of a distribution, we can write,

$S_n \sim B(n, p)$

Thus $E(S_n) = \mu = np$

and $Var(S_n) = \sigma = npq$

Let $Z = \frac{S_n - \mu}{\sigma}$ is a standardize random variable.

If we can show that $Z$ is a standard normal variable $N(0,1)$, then $S_n$ will a ~~ra~~ normal random variable with parameter $\mu, \sigma$.

Let us find the m.g.f of $Z$, i.e $M_Z(t)$.

$$M_Z(t) = E(e^{Zt})$$

$$= E\left(e^{\frac{S_n - \mu}{\sigma} t}\right)$$

$$= E\left(e^{S_n t/\sigma} \cdot e^{-\mu t/\sigma}\right)$$

$$= e^{-\mu t/\sigma} \cdot E\left(e^{S_n t/\sigma}\right)$$

$$= e^{-\mu t/\sigma} \cdot M_{S_n}(t/\sigma)$$

$$= e^{-\mu t/\sigma} \cdot (q + p e^{t/\sigma})^n \qquad \left[M_{S_n}(t) = (q + pe^t)^n\right]$$

$$= e^{-\frac{npt}{\sqrt{npq}}} \left(q + p \cdot e^{\frac{t}{\sqrt{npq}}}\right)^n$$

$$= \left[q \cdot e^{-pt/\sqrt{npq}} + p \cdot e^{\frac{t - pt}{\sqrt{npq}}}\right]^n$$

$$= \left[q \cdot e^{-\frac{pt}{\sqrt{npq}}} + p \cdot e^{\frac{qt}{\sqrt{npq}}}\right]^n$$

$$= \left[q\left\{1 - \frac{pt}{\sqrt{npq}} + \frac{1}{2!}\left(\frac{pt}{\sqrt{npq}}\right)^2 - \frac{1}{3!}\left(\frac{pt}{\sqrt{npq}}\right)^3 + \cdots\right\}\right.$$
$$\left. + p\left\{1 + \frac{qt}{\sqrt{npq}} + \frac{1}{2!}\left(\frac{qt}{\sqrt{npq}}\right)^2 + \frac{1}{3!}\left(\frac{qt}{\sqrt{npq}}\right)^3 + \cdots\right\}\right]^n$$

$$= \left[(p+q) - \cancel{\frac{pqt}{\sqrt{npq}}} + \cancel{\frac{pqt}{\sqrt{npq}}} + \frac{1}{2!}\left(\frac{t}{\sqrt{npq}}\right)^2 (qp^2 + pq^2)\right.$$
$$\left. + \frac{1}{3!}\left(\frac{t}{\sqrt{npq}}\right)^3 (pq^3 - qp^3) + \cdots\right]^n$$

$$= \left[ 1 + \frac{1}{2!} \frac{t^2}{npq} \cdot pq(p+q) + \frac{1}{3!} \left( \frac{t}{\sqrt{npq}} \right)^3 \cdot (pq^3 - qp^3) \right.$$

$$\left. + \frac{1}{4!} \left( \frac{t}{\sqrt{npq}} \right)^4 \cdot (pq^4 + qp^4) + \cdots \right]^n$$

$$= \left[ 1 + \frac{t^2}{2n} + O(n^{-3/2}) \right]^n$$

where $O(n^{-3/2})$ represents terms involving $n^{3/2}$ and higher powers of $n$ in the denominator.

Proceeding to the limit as $n \to \infty$, we get.

$$\lim_{n \to \infty} M_Z(t) = \lim_{n \to \infty} \left[ 1 + \frac{t^2}{2n} + O(n^{-3/2}) \right]^n$$

$$= \lim_{n \to \infty} \left[ 1 + \frac{t^2}{2n} \right]^n$$

$$= e^{t^2/2}$$

which is the M.G.F. of a standard normal variate. Hence by uniqueness theorem of M.G.F,

$$Z = \frac{S_n - \mu}{\sigma}$$ is asymptotically $N(0,1)$.

Hence $S_n = X_1 + X_2 + \cdots + X_n$ is asymptotically $N(\mu, \sigma^2)$ as $n \to \infty$.

☑ Lindeberg-Levy Theorem: (This is also a CLT).

Statement:

If $x_1, x_2 \ldots x_n$ are independently and identically distributed random variables with $\left.\begin{aligned} E(X_i) &= \mu_1 \\ V(X_i) &= \sigma_1^2 \end{aligned}\right\}; i = 1, 2, \ldots n$

then the sum $S_n = X_1 + X_2 + \cdots + X_n$ is asymptotically normal with mean $\mu = n\mu_1$ and variance $\sigma^2 = n\sigma_1^2$.

[ Here we make the following assumption

i> The variables are independent and identically distributed.

ii> $E(X_i^2)$ exists for all $i = 1, 2, \ldots$ ]

Proof: Let $M_1(t)$ denote the M.G.F of each of the deviation $(X_i - \mu_1)$ and $M(t)$ denote the M.G.F of the standard variate

$$Z = \frac{S_n - \mu}{\sigma}$$

$\mu_1'$ of the deviation $(X_i - \mu_1)$ is $E(X_i - \mu_1)$

$$= E(X_i) - E(\mu_1)$$
$$= \mu_1 - \mu_1$$
$$= 0$$

$\mu_2'$ of the deviation $(X_i - \mu_1)$ is $E(X_i - \mu_1)^2$

$$= E(X_i^2 - 2\mu_1 X_i + \mu_1^2)$$
$$= E(X_i^2) - 2\mu_1 E(X_i) + \mu_1^2$$
$$= E(X_i^2) - 2\mu_1^2 + \mu_1^2$$
$$= E(X_i^2) - \mu_1^2$$
$$= \sigma_1^2$$

Now we know that,

$$M_1(t) = 1 + \mu_1' t + \mu_2' \frac{t^2}{2!} + \mu_3' \frac{t^3}{3!} + \cdots$$

$$= 1 + \sigma_1^2 \frac{t^2}{2!} + O(t^3)$$

where $O(t^3)$ denotes the terms with $t^3$ and higher powers of $t$.

We have,

$$Z = \frac{S_n - \mu}{\sigma}$$

$$= \frac{(x_1 + x_2 + \cdots + x_n) - n\mu_1}{\sigma}$$

$$= \frac{\sum_{i=1}^{n}(x_i - \mu_1)}{\sigma}$$

$$= \sum_{i=1}^{n}\left(\frac{x_i - \mu_1}{\sigma}\right) \quad \left[\text{since } x_i\text{'s are independent}\right]$$

Now

$$M_Z(t) = M_{\sum_{i=1}^{n}\left(\frac{x_i - \mu_i}{\sigma}\right)}(t)$$

$$= M_{\sum_{i=1}^{n}(x_i - \mu_i)}(t/\sigma)$$

$$= \prod_{i=1}^{n} M_{(x_i - \mu_1)}(t/\sigma)$$

$$= \{M_1(t/\sigma)\}^n$$

$$= \left[M_1\left(\frac{t}{\sqrt{n}\,\sigma_1}\right)\right]^n$$

$$= \left[1 + \sigma_1^2 \cdot \frac{t^2}{n\sigma_1^2\, 2!} + O\left(n^{-3/2}\right)\right]^n$$

For every fixed $t$, the term $O(n^{-3/2}) \to 0$ as $n \to \infty$

Therefore as $n \to \infty$, we get,

$$\lim_{n\to\infty} M_Z(t) = \lim_{n\to\infty} \left[1 + \frac{t^2}{2n} + O(n^{-3/2})\right]^n$$

$$= \lim_{n\to\infty} \left[1 + \frac{t^2/2}{n}\right]^n$$

$$= e^{t^2/2}$$

which is M.G.F of standard normal variate $N(0,1)$. Hence by uniqueness of M.G.F's, $Z = \frac{S_n - \mu}{\sigma}$ is asymptotically $N(0,1)$ or $S_n$ is asymptotically $N(\mu, \sigma^2)$, where $\mu = n\mu_1$ and $\sigma^2 = n\sigma_1^2$.

Note: The C.L.T can be stated in another forms as follows →

i) If $X_1, X_2 \ldots X_n$ are i.i.d, ~~th~~ with mean $\mu_1$ and variance $\sigma_1^2$ and, $S_n = X_1 + X_2 + \cdots + X_n$, then for $-\infty < a < b < \infty$,

$$\lim_{n\to\infty} P\left[a \leq \frac{S_n - n\mu_1}{\sigma_1\sqrt{n}} \leq b\right] = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$$

$$\text{or, } \lim_{n\to\infty} P\left[a < \frac{S_n - E(S_n)}{\sqrt{var(S_n)}} \leq b\right] = \Phi(b) - \Phi(a)$$

$$= \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx.$$