PROBABILITY AND STATISTICS (UCS401)

Lecture-27

(Sampling Distribution and The Central Limit Theorem)

Sampling Distributions and Theory of Estimation (Unit –V & VI)

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

Dr. Rajanish Rai
Assistant Professor
School of Mathematics
Thapar Institute of Engineering and Technology, Patiala

Sampling distribution and the Central Limit theorem

Rajnish Rai

Population : (N) A population Consist of the totality of the observations with which we are Concerned.

Note that : The number of observations in the population is defined to be the size of population. If there are 600 students in the school whom we classified according to blood type, we say that we have a population of size 600.

Examples : (i) The number on card in a deck.

(ii) The height of resident in a certain city.

(iii) The length of fish in a particular lake.

are examples of population with finite size.

Sample : (n) A sample is defined as a finite subset of a population.

Note that : (i) If the size of a sample is less than or equal to 30 $(n \leq 30)$ then sample is said to be small sample.

(ii) If the size of sample is greater than 30 (i.e, $n > 30$), the sample is said to be a large sample.

Sampling : The process of selecting a sample from a population is called sampling.

* A descriptive measure computed from the data of a sample is called a statistic.

* The statistics of a population are known as parameter.

Note that :

(i) If there are n observations say $x_1, x_2, x_3, \ldots, x_n$ in a sample, then mean of the sample is given by

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

and the variance of the sample is given by

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}.$$

(ii) Sampling where a member can be selected more than once is called sampling with replacement.

If N is the size of population, then the number of samples of size n

with replacement is N^n.

(iii) If a member of population cannot be chosen more than once, then sampling is known as sampling without replacement.

When drawing samples of size n from an infinite population of size N without replacement and ignoring the order in which the sample is drawn, the number of possible samples is given by $^{N}C_{n} = \frac{N!}{n!\,(N-n)!}$.

Standard error:

The standard deviation of sampling distribution of a statistic is called standard error. It is denoted by S.E.

If $x_1, x_2, x_3, \ldots, x_n$ (n-samples)

taken from normal population with mean μ and variance σ^2.

$$\bar{X} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$$

has a normal distribution with mean

$$\mu_{\bar{X}} = \frac{1}{n}(\underbrace{\mu + \mu + \mu + \cdots + \mu}_{n\text{ times}}) = \mu$$

and variance

$$\sigma^2_{\bar{X}} = Var(\bar{X})$$

$$= Var\left(\frac{1}{n}(x_1 + x_2 + \cdots + x_n)\right)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} Var(x_i)$$

$$= \frac{\sigma^2 \cdot n}{n^2} = \frac{\sigma^2}{n}$$

$$\sigma^2_{\bar{X}} = \frac{\sigma^2}{n} = Var(\bar{X}) \Rightarrow \boxed{S.E.(\bar{X}) = \frac{\sigma}{\sqrt{n}}}$$

* The standard error of the sampling mean $(\bar{x})$ is given by $\frac{\sigma}{\sqrt{n}}$,

Where σ is the standard deviation of the population and n is the sample size.

The standard error of the sample mean is given by

$$\boxed{S.E._{(\bar{x})} = \sigma_{(\bar{x})} = \frac{\sigma}{\sqrt{n}}, \quad \text{When } \sigma \text{ is known.}}$$

* When σ is not known, the standard error of mean is

$$S.E.(\bar{x}) = \frac{s}{\sqrt{n}},$$

Where s is standard deviation of sample.

* When the standard deviation is known, the standard error of difference of the sample mean $\overline{x_1}$ and $\overline{x_2}$ drawn from the same population is given by

$$S.E.\left(\overline{x_1}-\overline{x_2}\right) = \sqrt{\sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}.$$

Where n_1 and n_2 are the sizes of the two samples from the same population.

* The standard error of difference of the two sample means $\overline{x_1}$ and $\overline{x_2}$ drawn from different population with standard deviations σ_1 and σ_2, respectively is defined as:

$$S.E.\left(\overline{x_1}-\overline{x_2}\right) = \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}.$$

Note that: In general, the standard error of mean is used for large samples.

Note that : If we are sampling from a population with unknown distribution, either finite or infinite, the sampling distribution of $\bar{X}$ will still be approximately normal $\left(Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \right)$ with mean μ and variance σ^2/n, provided that the sample size is large.

* This amazing result is an immediate consequence of the following theorem, called the Central Limit Theorem.

The Central Limit Theorem

If $\overline{X}$ is the mean of a random sample of size n taken from a population with mean μ and finite variance σ^2, then the limiting form of the distribution of

$$Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$$

as $n \to \infty$, is the standard normal distribution, i.e. $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

$$Z \sim N(0,1).$$

Question: An electric firm manufactures light bulbs that have a length of life that is approximately normally distributed, with mean equal to 800 hours and a standard deviation of 40 hours. Find the probability that a random sample of 16 bulbs will have an average life of less than 775 hours.

Solution-: The sampling distribution of $\overline{X}$ will be approximately normal with $\mu_{\overline{X}} = 800 = \mu$

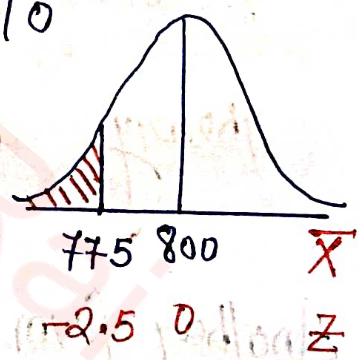
$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{40}{\sqrt{16}} = 10$$

$$\boxed{\sigma_{\overline{X}} = 10}$$

The standard normal variate for sample mean

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{\overline{X} - 800}{10}$$

The required probability



$$= P\left(\overline{X} < 775\right)$$

$$= P\left(\frac{\overline{X} - 800}{10} < \frac{775 - 800}{10}\right)$$

$$= P\left(Z < -2.5\right) = 0.0062$$

Note that: The Central limit theorem (CLT) is an important result in statistics, which states that the normal distribution is the limiting distribution to the sum of the independent random variables with finite variance as the number of random variables get indefinitely large.

Since many real processes yield distribution with finite variance, this theorem has a wide range of applicability in sampling theory and other area of statistics.

Another form of central Limit theorem:

If X_i $(i=1,2,3,\ldots,n)$ be independently distributed random variables such that

$$E(X_i) = \mu_i \quad \text{and} \quad Var(X_i) = \sigma_i^2$$

then as $n \to \infty$, the distribution of the sum of these random variables, namely

$$S_n = x_1 + x_2 + x_3 + \cdots + x_n$$

tends to the normal distributed with mean μ and variance σ^2, where

$$\mu = \sum_{i=1}^{n} \mu_i = \sum_{i=1}^{n} E(x_i)$$

$$\sigma^2 = \sum_{i=1}^{n} \sigma_i^2 = \sum_{i=1}^{n} var(x_i).$$

Question:- A coin is tossed 200 times. Find the approximate probability that the number of heads obtained is between 80 and 120. Given that $P(Z < -2.82) = 0.0024$ from normal-table.

Solution:- Let X: Number of heads

with $p = \frac{1}{2}$ & $q = \frac{1}{2}$.

Thus, required probability $= P(80 < X < 120)$.

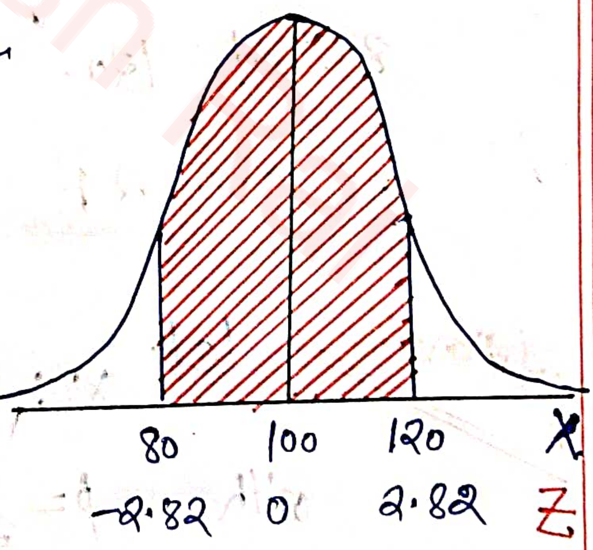
Since $n = 200$ is very large so we apply the central limit theorem.

By Binomial distribution approximation, we have

$\text{Mean} = np$	$\text{Variance } \sigma^2 = npq$
$= 200 \times \frac{1}{2}$	$\text{Variance} = 200 \times \frac{1}{2} \times \frac{1}{2}$
$\text{Mean} = 100$	$\text{Variance } \sigma^2 = 50$

Thus, $Z = \dfrac{X - \mu}{\sigma} = \dfrac{X - 100}{\sqrt{50}}$ —(i)

Hence, the required probability



$= P(80 < X < 120)$

$= P\left(\dfrac{80-100}{\sqrt{50}} < \dfrac{X-100}{\sqrt{50}} < \dfrac{120-100}{\sqrt{50}}\right)$

$= P\left(\dfrac{-20}{\sqrt{50}} < Z < \dfrac{20}{\sqrt{50}}\right)$

$= P(-2.82 < Z < 2.82) = 2P(-2.82 < Z < 0)$

—(2)

$$\because \quad P(Z < -2.82) = 0.0024$$

$$0.5 - P(-2.82 < Z < 0) = 0.0024$$

$$P(-2.82 < Z < 0) = 0.5 - 0.0024$$

$$P(-2.82 < Z < 0) = 0.4976.$$

Therefore, the required probability

$$P(80 < x < 120) = P(-2.82 < Z < 2.82)$$

$$= 2 \times P(-2.82 < Z < 0)$$

$$= 2 \times 0.4976$$

$$= 0.9952$$ Ans

Question: A random sample of size 100 is taken from a population whose mean is 60 and variance is 400. Using central limit theorem, with what probability can we assert that the mean of the sample will not differ from $\mu = 60$ by more than 4?
Given that $P(Z<-2) = 0.0228$.

Solution: Given that

$$n = 100, \quad \mu_i = 60, \quad \& \quad \sigma_i^2 = 400.$$

Thus, required probability

$$= P(|\overline{X} - 60| \leq 4).$$

The sample mean is

$$\overline{X} = \frac{X_1 + X_2 + X_3 + \cdots + X_{100}}{100}$$

$E(aX) = a\,E(X)$

$Var(aX) = a^2 Var(X)$

$$E(\overline{X}) = \frac{1}{100} \sum_{i=1}^{100} E(X_i)$$

$$= \frac{1}{100} \times 100 \times 60 = 60$$

$$\boxed{E(\overline{X}) = 60}$$

and

$$Var(\bar{X}) = Var\left(\frac{x_1+x_2+\dots+x_{100}}{100}\right)$$

$$= \frac{1}{(100)^2}\sum_{i=1}^{100} Var(x_i)$$

$$= \frac{1}{(100)^2}\, 100 \times 400 = 4$$

$$\boxed{Var(\bar{X}) = 4}$$

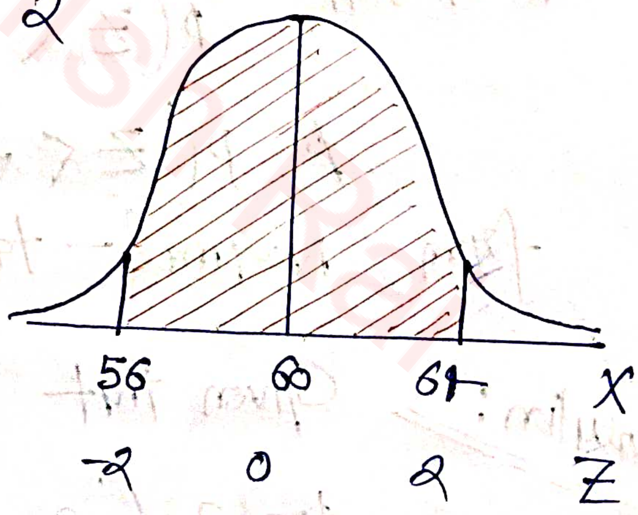
$$\boxed{\begin{aligned}\sqrt{Var(\bar{X})} &= \frac{\sigma}{\sqrt{n}} = S.E.(\bar{x}) \\ &= \frac{\sqrt{400}}{\sqrt{100}} \\ &= 2\end{aligned}}$$

Thus,

$$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = \frac{\bar{X}-60}{2}$$

Required probability

$$P\left(|\bar{X}-60| \le 4\right)$$



$$= P\left(-4 \le \bar{X}-60 \le 4\right)$$

$$= P\left(56 \le \bar{X} \le 64\right)$$

$$= P\left(\frac{56-60}{2} \le \frac{\bar{X}-60}{2} \le \frac{64-60}{2}\right).$$

$= P(-2 < Z < 2)$

$= 2 \times P(-2 < Z < 2)$

$= 2 \times 0.4772$

$= 0.9544$ Ans

Given

$P(Z < -2) = 0.0228$

$0.5 - P(-2 < Z < 0) = 0.0228$

$P(-2 < Z < 0) = 0.5 - 0.0228$

$P(-2 < Z < 0) = 0.4772$

Question: If $X_1, X_2, X_3, \ldots, X_n$ are Poisson Variables with parameter 2, use the central limit theorem to estimate $P(120 \leq S_n \leq 160)$

where;

$$S_n = X_1 + X_2 + X_3 + \cdots + X_n, \quad n = 75.$$

Given that

$$P(Z < 0.8165) = 0.7939$$

$$\& \quad P(Z < -2.45) = 0.0071.$$

from Normal-table.

Solution: Given that

$\lambda = 2$ (a poisson distribution parameter)

$\therefore E(X_i) = 2, \quad \& \quad Var(X_i) = 2.$

Thus, required probability

$$= P(120 \le S_n \le 160)$$

$$\therefore \quad E(S_n) = \sum_{i=1}^{75} E(x_i) = 75 \times 2$$

$$\boxed{E(S_n) = 150}$$

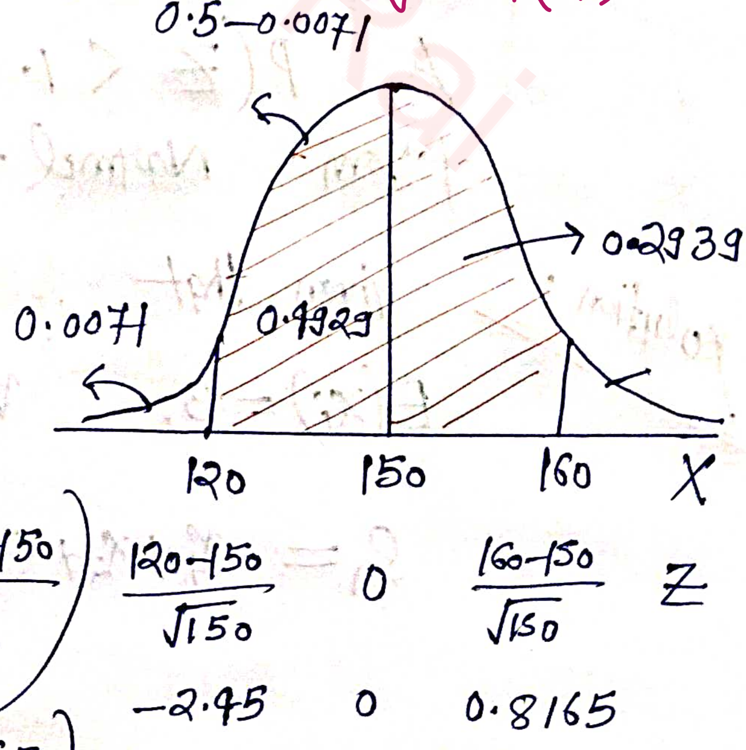
$$Var(S_n) = \sum_{i=1}^{75} Var(x_i)$$

$$= 75 \times 2 = 150$$

$$\boxed{Var(S_n) = 150}$$

Therefore $Z = \dfrac{S_n - 150}{\sqrt{150}} = \dfrac{S_n - E(S_n)}{\sqrt{Var(S_n)}}$

Required probability



$$P(120 \le S_n \le 160)$$

$$= P\left(\frac{120-150}{\sqrt{150}} \le \frac{S_n - 150}{\sqrt{150}} \le \frac{160-150}{\sqrt{150}}\right)$$

$$= P(-2.4495 \le Z \le 0.8165).$$

$$= P(-2.45 \leq Z \leq 0.8165)$$

$$= P(Z \leq 0.8165) - P(Z \leq -2.45)$$

$$= 0.7939 - 0.0071$$

$$= 0.7868$$ Ans

Question: Let $\{x_i\}$ be independent and identically distributed random variables with mean 3 and variance $1/2$. Use central limit theorem to estimate $P(340 \leq S_n \leq 370)$.

where $S_n = x_1 + x_2 + x_3 + \dots + x_n$ and $n = 120$.

Given that $P(Z < -2.58) = 0.0049$

& $P(Z < 1.29) = 0.9015$.

from Normal-table.

Solution: Given that

$E(x_i) = 3$, $Var(x_i) = 1/2$

$S_n = x_1 + x_2 + x_3 + \dots + x_n$;

$n = 120$

Thus, required probability

$$= P(340 \leq \underbrace{S_n}_{\text{mean} \quad \text{variance}} \leq 370)$$

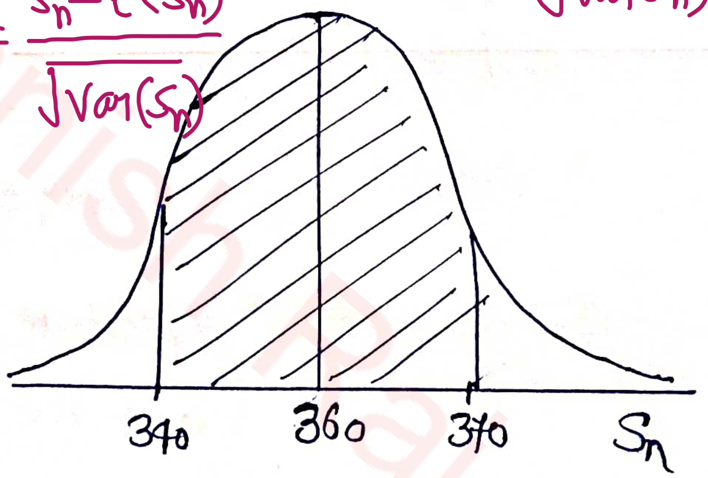
$$\therefore E(S_n) = \sum_{i=1}^{120} E(x_i) \qquad Var(S_n) = \sum_{i=1}^{120} Var(x_i)$$

$$= 120 \times 3 \qquad = 120 \times 1/2$$

$$\boxed{E(S_n) = 360} \qquad \boxed{Var(S_n) = 60}$$

$$Z = \frac{S_n - E(S_n)}{\sqrt{var(S_n)}}$$

$$\text{Thus, } Z = \frac{S_n - 360}{\sqrt{60}} = \frac{S_n - E(S_n)}{\sqrt{var(S_n)}}$$



$$\frac{340-360}{\sqrt{60}} \quad 0 \quad \frac{370-360}{\sqrt{50}} \quad Z$$

$$-2.582 \quad 0 \quad 1.291 \quad Z$$

The required probability

$$= P(340 \leq S_n \leq 370)$$

$$= P\left(\frac{340-360}{\sqrt{60}} \leq \frac{S_n - 360}{\sqrt{60}} \leq \frac{370-360}{\sqrt{60}}\right)$$

$$= P\left(\frac{-20}{\sqrt{60}} \leq Z \leq \frac{10}{\sqrt{60}}\right)$$

$$= P(-2.582 \leq Z \leq 1.291)$$

$$= P(-2.58 \leq Z \leq 1.29)$$

$$= P(Z \leq 1.29) - P(Z \leq -2.58)$$

$$= 0.9015 - 0.0049$$

$$= 0.8966$$