Chapter 4
Random Sample
Central Limit Theorem

Learning Outcomes:

At the end of the lesson student should be able to

- Describe the terms random sample, statistics and sampling distribution.
- Understand the concept of central limit theorem.
- Use central limit theorem to approximate sampling distribution of the sample mean.

DEFINITIONS

- **Parameters**: a quantity of interest that is a property of an unknown probability distribution. Example it may be the mean or variance of the probability distribution.
- Parameters are unknown but could be estimated.

Example:

Let p be the probability that a machine breakdown is due to operator misuse.

p depends on the probability distribution that represents the causes of the machine breakdown.

p is a parameter and unknown but can be estimated from records of machine breakdown.

DEFINITIONS

- A ***statistic*** is a property of a sample from the population. A function of a set of data observations. Example are the sample mean or sample variance.
- Statistics can be calculated from a set of observed data.
- Statistics can be used to estimate unknown parameters.

Example:

$$\overline{X} = f(X_1, \ldots X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$g\left(X_1, \ldots X_n\right) = \sum_{i=1}^{n} X_i^{\;2}$$

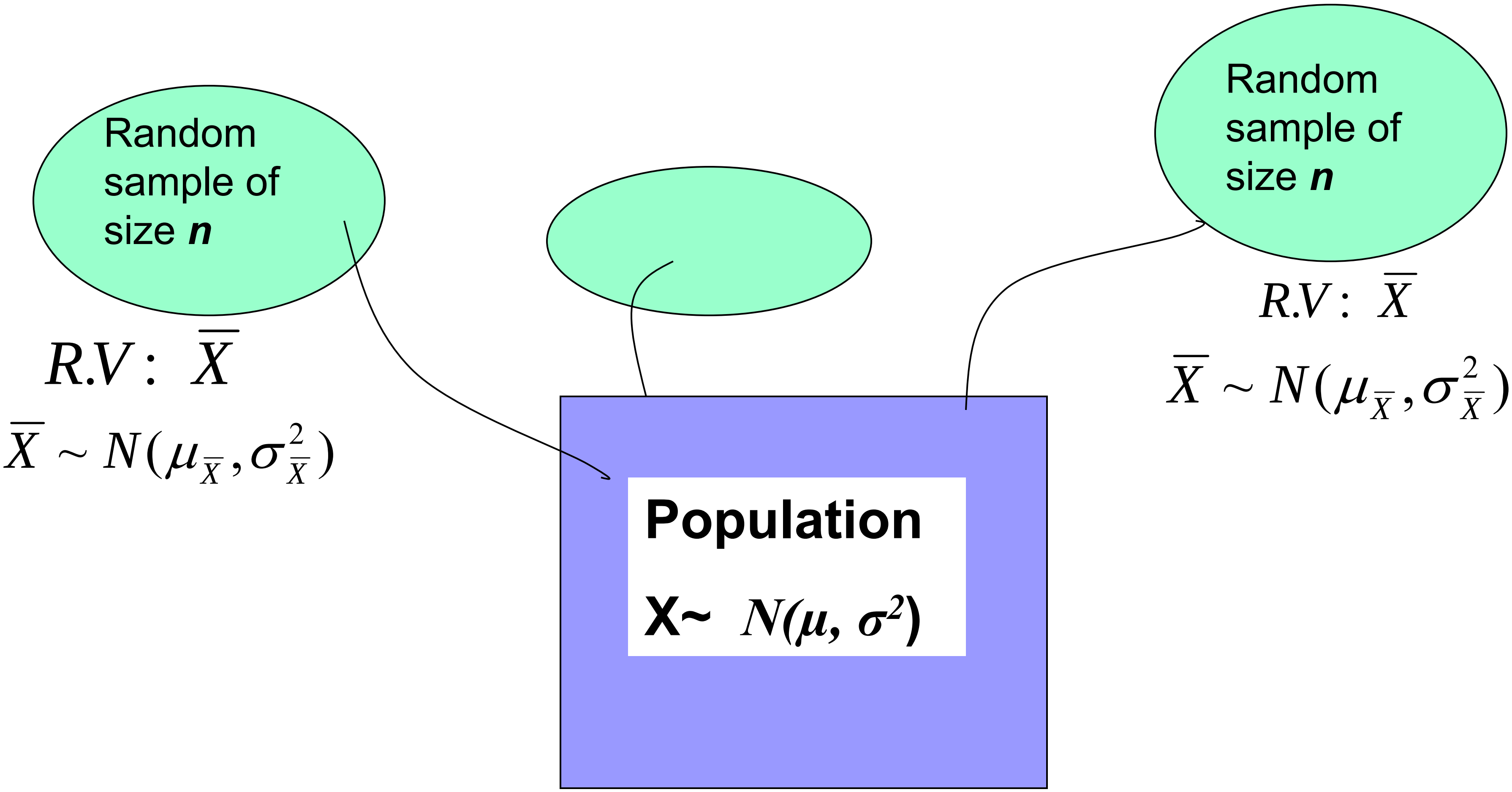
DEFINITIONS

- **Random Sample**: Independent random variables $X_1, X_2, \ldots, X_n$ with the same distribution and each has the same chance of being selected.

- **Sampling Distribution**: the probability distribution of a statistic.

  Example: $\overline{X} \sim N(\mu_{\overline{X}}, \sigma^2_{\overline{X}})$

Random Sample; Sampling Distribution



Distribution of Sample Mean

- Let $X_1, X_2, \ldots, X_n$ be a random sample from a population with mean μ and variance σ^2

- Definition and Properties (sampling distribution of $\overline{X}$):

$$\boxed{\overline{\mathbf{X}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i}$$

$$\boxed{E\left(\overline{X}\right) = \mu \ , \ V\left(\overline{X}\right) = \frac{\sigma^2}{n}}$$

Central Limit Theorem

- Let $X_1, X_2, \ldots, X_n$ be a random sample from a population with mean μ and variance σ^2

$$\boxed{\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i}$$

$$\boxed{E(\bar{X}) = \mu \;,\; V(\bar{X}) = \frac{\sigma^2}{n}}$$

- If n is sufficiently large

$$\boxed{\bar{X} \approx N(\mu, \frac{\sigma^2}{n})}$$

Apply safely
if $n \geq 30$

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim \mathcal{N}(0,1),$$

UNIVERSITI TEKNOLOGI PETRONAS

Example 1

At chemical engineering department, UTP, the mean age of the students is 20.6 years old, and the variance is 20 years. A random sample of 80 students is drawn from 250 students. What is the probability that the average age of these students is greater than 22 years old?

Solution

Example 1

At chemical engineering department, UTP, the mean age of the students is 20.6 years old, and the variance is 20 years. A random sample of 80 students is drawn from 250 students. What is the probability that the average age of these students is greater than 22 years old?

Solution

The mean of $X = E(X) = 20.6$ and the variance of $X = V(X) = 20$

For $n = 80$, the mean of $\overline{X} = E(\overline{X}) = 20.6$ and $V(\overline{X}) = \dfrac{\sigma^2}{n} = \dfrac{20}{80} = 0.25$

Hence, $\overline{X} \sim N(20.6,\ 0.25)$

So, $P(\overline{X} > 22) = P(Z > \dfrac{22 - 20.6}{\sqrt{0.25}}) = P(Z > 2.8) = 1 - P(Z \leq 2.8)$

$$= 1 - \Phi(2.8) = 1 - 0.9974 = 0.0026$$

Example 2: Let X denote the number of flaws in a 1 in length of copper wire. The pmf of X is given in the following table

$X = x$	0	1	2	3
$P(X = x)$	0.48	0.39	0.12	0.01

100 wires are sampled from this population. What is the probability that the average number of flaws per wire in this sample is less than 0.5?

Solution: $\mu = 0.66, \quad \sigma^2 = 0.5244, \quad n = 100$

Let X_1 ….. X_{100} denote the number of flaws in the 100 wires sampled from the population.

$$\boxed{\bar{X} \approx N(\mu, \frac{\sigma^2}{n})}$$

From CLT $\bar{X} \approx N(0.66, 0.005244)$

$$P(\bar{X} < 0.5) = P(Z < \frac{0.5 - 0.66}{0.7242/\sqrt{100}})$$

$$= P(Z < -2.21) = 0.0136$$

UNIVERSITI TEKNOLOGI PETRONAS

Example 3:

The flexural strength (in MPa) of certain concrete beams is $X \sim N(8, 2.25)$. Find the probability that the sample mean of strength of 36 concrete beams will belong to (7.55, 8.75)

Solution:

$$\mu = 8, \quad \sigma^2 = 2.25$$

$$n = 36, \quad \sigma = 1.5$$

$$P(7.55 < \bar{X} < 8.75) = P\left(\frac{7.55 - 8}{1.5/6} < \frac{\bar{X} - 8}{1.5/6} < \frac{8.75 - 8}{1.5/6} \right)$$

$$= P\left(-1.8 < \frac{\bar{X} - 8}{1.5/6} < 3 \right) = P(-1.8 < Z < 3)$$

$$= \Phi(3) - \Phi(-1.8) = 0.9987 - 0.0359 = 0.9628$$

Exercise:

At a large university, the mean age of the students is 22.3 years, and the standard deviation is 4 years. A random sample of 64 students is drawn. What is the probability that the average age of these students is greater than 23 years?

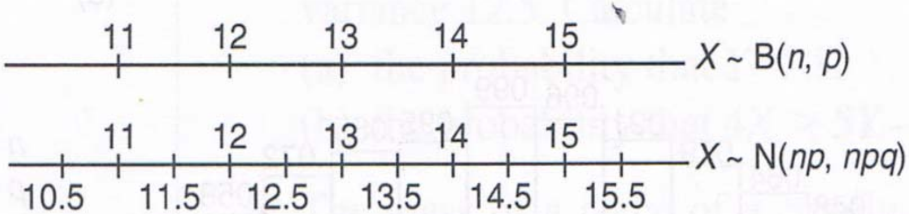
Normal Approximation

The binomial and Poisson distributions are discrete random variables, whereas the normal distribution is continuous.
The Normal approximation can be used to find the probability of *X* for Binomial or Poisson by add the continuity correction value 0.5.

The continuity correction, ± 0.5 for probability of ***X*** is depend on the inequality sign, $<,\ \leq,\ >,\ \geq$

For example ***P(X < a) = P(X - 0.5 < a - 0.5***) and for

$$P(X \leq a) = P(X + 0.5 \leq a + 0.5)$$



▲ Figure 66

Hence, we can approximate the probailities as follows:

Discrete B(n, p)	**Continuous N(np, npq)**
$P(X = 12)$	$P(11.5 < X < 12.5)$
$P(X < 12)$	$P(X < 11.5)$
$P(X \leqslant 12)$	$P(X < 12.5)$
$P(X > 12)$	$P(X > 12.5)$
$P(X \geqslant 12)$	$P(X > 11.5)$
$P(11 \leqslant X \leqslant 14)$	$P(10.5 < X < 14.5)$
$P(11 < X < 14)$	$P(11.5 < X < 13.5)$
$P(11 < X \leqslant 14)$	$P(11.5 < X < 14.5)$
$P(11 \leqslant X < 14)$	$P(10.5 < X < 13.5)$

Normal approximation to Binomial

A random variable ***X ~ Bin(n , p)*** and ***n*** is large and ***p*** is small such that ***np > 5***, than ***X*** can be calculated approximately using the Normal distribution.

It means that the random variable ***X*** will be normally distributed with mean ***μ = np*** and variance, $\sigma^2 = np(1-p)$ so, ***X ~ N(np, npq)***.

Example :
Suppose in experiment of tossing a fair coin for 200 times. What is the probability of getting between 90 and 110 heads?

Solution:

Let X be the random variable representing the number of heads thrown.

X ~ Bin (200, 0.5)

Since ***n*** is large and ***np > 5***, then we can use normal approximation to find the probability. It mean that now, X is normally distributed with mean ***np =100*** and variance ***50***.

So, ***X ~ N (100, 50)*.**

Hence,

$$P(90 \le X \le 110) = P\left(\frac{(90-0.5)-100}{\sqrt{50}} \le Z \le \frac{(110+0.5)-100}{\sqrt{50}}\right)$$

$$= P\left(\frac{89.5-100}{7.07} \le Z \le \frac{110.5-100}{7.07}\right) = P(-1.48 \le Z \le 1.48)$$

$$= \Phi(1.48) - \Phi(-1.48) = 0.9306 - 0.0694 = 0.8612$$

Normal approximation to Poisson

The normal distribution can also be used to approximate the Poisson distribution for large values of λ (the mean of the Poisson distribution).

If we have a random variable ***X ~ Poisson (λ)*** and λ is large than ***X*** can be calculated approximately using the Normal distribution. It means that the random variable ***X*** will be normally distributed with mean $\mu = \lambda$ and variance, $\sigma^2 = \lambda$ so, $X \sim N(\lambda, \lambda)$

Example:

A car hire firm has 20 cars to hire. The number of demands for a car is hired per day is a Poisson distribution with mean of 5. Calculate the probability that at most ten cars will be hired in one day.

Solution:

Let a random variable ***X*** denotes the number of demands for a car. The given mean value is 5. By the Poisson distribution

$$P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!} \quad , x = 0, 1, 2, 3, 4, ..., 20$$

Since λ is large, then the probability can be calculated using a normal approximation with mean $\lambda=5$ and variance is also $\lambda=5$.
so, $\boldsymbol{X \sim N(5, 5)}$.

Hence,

$$P(X \le 10) = P\left(\frac{(X + 0.5) - 5}{\sqrt{5}} \le \frac{(10 + 0.5) - 5}{\sqrt{5}} \right)$$

$$= P(Z \le 2.46) = \Phi(2.46) = 0.9931$$