Chapter 2
Random Variable

L1- Discrete random variable

L2- Continuous random variable

DISCRETE RANDOM VARIABLES

UNIVERSITI TEKNOLOGI PETRONAS

Learning Objectives:

At the end of the lecture, you will be able to :

- describe types of random variables

- calculate their probability distribution and their cumulative distribution

Random Variable:

A numerical variable whose measured value can change from is one outcome of random experiment.

An uppercase letter (say X) is used to denote a random variable.

After the experiment is conducted, the measured value is denoted by a lowercase letter, say $x = 10$.

Probability distribution / distribution of a random variable X:

-description of the set of probabilities associated with the possible values of X.

Probability mass function:

- describe the probability distribution of a discrete random variable

Probability density function:

- describe the probability distribution of a continuous random variable

Examples of random variables:

The **number of scratches** on a surface. Integer values ranging from zero to about 5 are possible values.
$X = \{ 0, 1, 2, 3, 4, 5\}$

The **time taken to complete an examination.** Possible values are 15 minutes to over 3 hours.
$X = \{ 15 \leq x \leq 180 \}$

DISCRETE RANDOM VARIABLE

X is a discrete random variable if:

- The set x of values of X is finite or countable.

- The Probability Mass Function (pmf) of X is a set of probability values p_i assigned to each of the values of x_i

1. $f(x_i) = P(X = x_i)\ , 0 \le P(x) \le 1$ for each value of x
2. $\sum P(x) = \sum P(X = x) = 1$

Example 1: The sample space for a machine breakdown problem is

$$S = \{ \text{electrical, mechanical, misuse} \}$$

and each of these failures is associated with a repair cost of about RM200, RM350 and RM50 respectively.

Identify the random variable giving reasons for your answer.

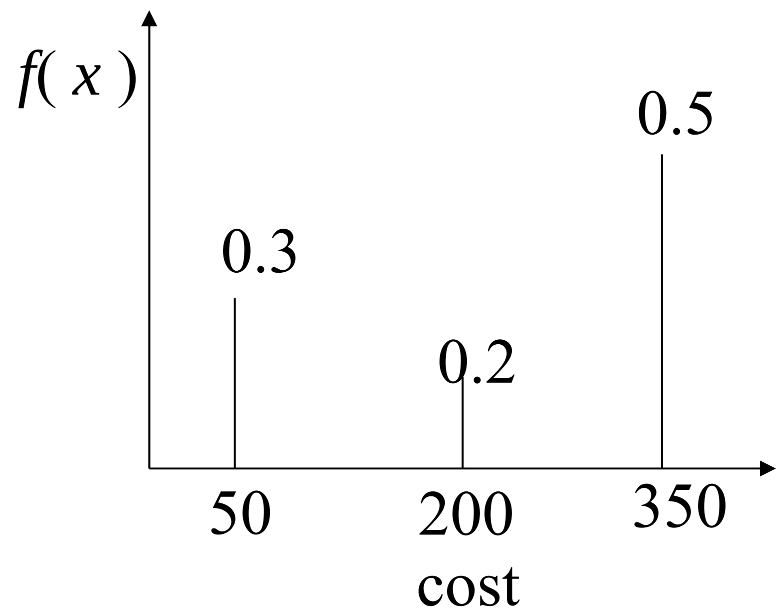
Example 2: The analysis of the surface of semi conductor wafer records the number of particles of contamination that exceed a certain size. Identify the possible random variable and its values.

Probability mass function may typically be given in tabular or graphical form

If from Example 1 that P(cost $_{\text{Misuse}}$ =50)=0.3, P(cost $_{\text{Electrical}}$ = 200) = 0.2 and P(cost $_{\text{Mechanical}}$ = 350) = 0.5. The probability mass function is given either

$X = x$	**50**	**200**	**350**
$f(x) = P(X = x)$	**0.3**	**0.2**	**0.5**

tabular form



line graph

MEAN AND VARIANCE :- Discrete R.V

We can summarize probability distribution by its mean and variance.

Mean or expected value is

$$\boxed{\mu = E(X) = \sum_{i=1}^{n} x_i f(x_i)}$$

Variance of X is given as

$$\boxed{\sigma^2 = V(X) = E(X-\mu)^2 = \sum_{i=1}^{n} x_i^2 f(x_i) - \mu^2 = E(X^2) - (E(X))^2}$$

Standard deviation of X is σ .

From Example 1 that P(cost $_{\text{Misuse}}$ =50)=0.3, P(cost $_{\text{Electrical}}$ = 200) = 0.2 and P(cost $_{\text{Mechanical}}$ = 350) = 0.5. The probability mass function is given either

$X = x$	50	200	350
$f(x) = P(X = x)$	0.3	0.2	0.5

Find

i. the mean of *X*;
ii. The variance and standard deviation of *X*.
iii. Probability X greater than 200.

Find

i. the mean of *X*;

ii. the variance and standard deviation of *X*;

iii. probability X greater than 200.

$X = x$	**50**	**200**	**350**	**Total**
$f(x) = P(X=x)$	0.3	0.2	0.5	1
$Xf(x)$	15	40	175	230
$X^2f(x)$	750	8000	61250	70000

i. The mean of $X = E(X) = \sum Xf(X) = 230.$

ii. The variance of $X = E(X^2) - (E(X))^2$ = 70,000-52,900 = 17,100

and standard deviation of $X = sqrt(var(X)) = 130.77$

iii. Probability *X* greater than 200 = $P(X>200)$ = 0.5

CUMULATIVE DISTRIBUTION FUNCTION

The cumulative distribution F(x) of a discrete random variable X with probability mass function $f(x)$ is

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i)$$

The cumulative distribution of F(x) is an increasing step function with steps at the values taken by the random variable.

The height of the steps are probabilities of taking these values.

From **Example 1 (machine breakdowns)** : The probability distribution is

$X = x$	50	200	350
$f(x) = P(X = x)$	0.3	0.2	0.5

The following cumulative distribution is obtained

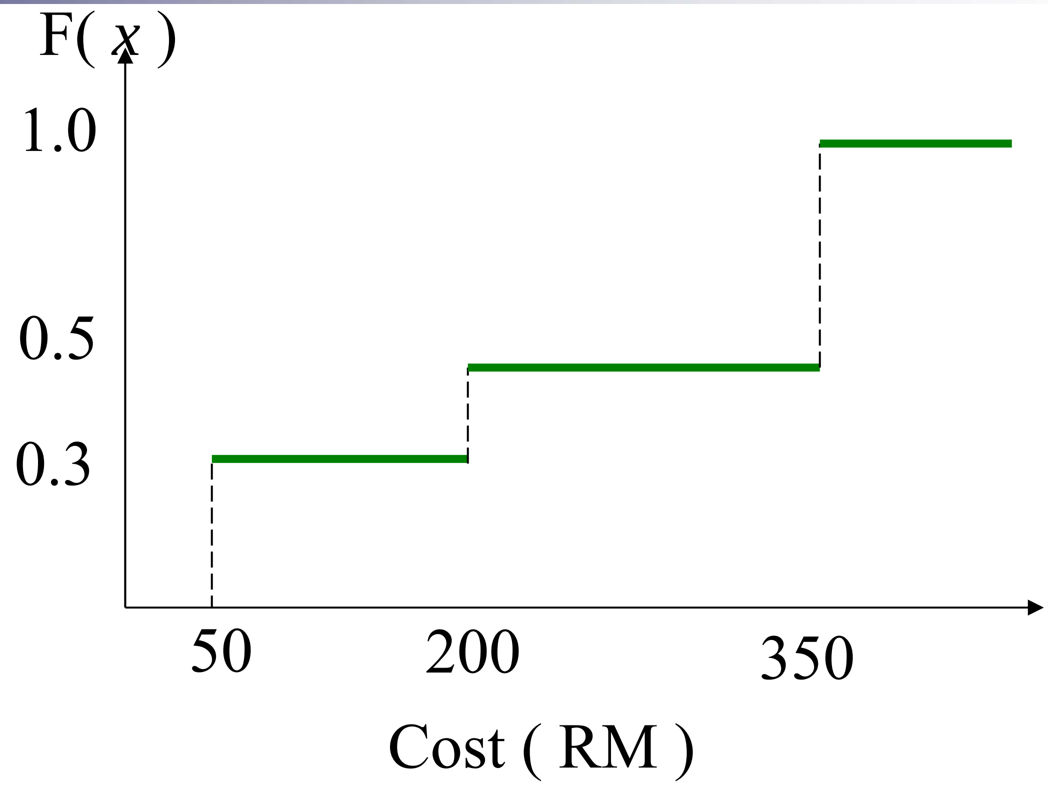
$$-\infty < x < 50 \quad \Rightarrow \quad F(x) \quad = \quad P(\text{cost} \le x\ ) \quad = \quad 0$$

$$50 \le x < 200 \quad \Rightarrow \quad F(x) \quad = \quad P(\text{cost} \le x\ ) \quad = \quad 0.3$$

$$200 \le x < 350 \quad \Rightarrow \quad F(x) \quad = \quad P(\text{cost} \le x\ ) \quad = \quad 0.3 + 0.2 = 0.5$$

$$350 \le x < \infty \quad \Rightarrow \quad F(x) \quad = \quad P(\text{cost} \le x\ ) \quad = \quad 0.3+0.2+0.5 = 1.0$$

Graph of F(x)



$$F(x) = P(X \le x) = \begin{cases} 0 & for \quad x < 50 \\ 0.3 & for \quad 50 \le x < 200 \\ 0.5 & for \quad 200 \le x < 350 \\ 1 & for \quad x \ge 350 \end{cases}$$

Example 2:

Given the pmf :

$X=x$	0	1	2	3
$P(X=x)$	0.15	0.25	k	0.35

Find,

i. the value of *k* that result in a valid probability distribution.
ii. the expected value of *X*.
iii. the variance and the standard deviation of *X*.
iv. the probability that *X* greater than or equal to 1?
v. the CDF of *X* .

X	0	1	2	3	Total
P(X=x)	0.15	0.25	0.25	0.35	1
X.P(X=x)	0	0.25	0.5	1.05	1.8
X^2.P(X=x)	0	0.25	1	3.15	4.4
		Var(x)	1.16		
		sd(x)	1.077033		

$$F(x) = P(X \leq x) = \begin{cases} 0 & for \quad x < 0 \\ 0.15 & for \quad 0 \leq x < 1 \\ 0.4 & for \quad 1 \leq x < 2 \\ 0.65 & for \quad 2 \leq x < 3 \\ 1 & for \quad x \geq 3 \end{cases}$$

Exercise :

Let X denote the number of bars of service on your cell phone whenever you are at an intersection with the following probabilities:

x	0	1	2	3	4	5
$P(X=x)$	0.05	0.15	0.20	k	0.15	0.1

Determine the following:

(i) the value of k,

(ii) Mean and variance,

(iii) $P(X < 2)$

(iv) $P(X > 2.5)$

(v) cumulative distribution function, $F(x)$

Chapter 2
Random Variable

L2- Continuous random variable

CONTINUOUS RANDOM VARIABLES

Learning Objectives:

At the end of the lecture, you will be able to :

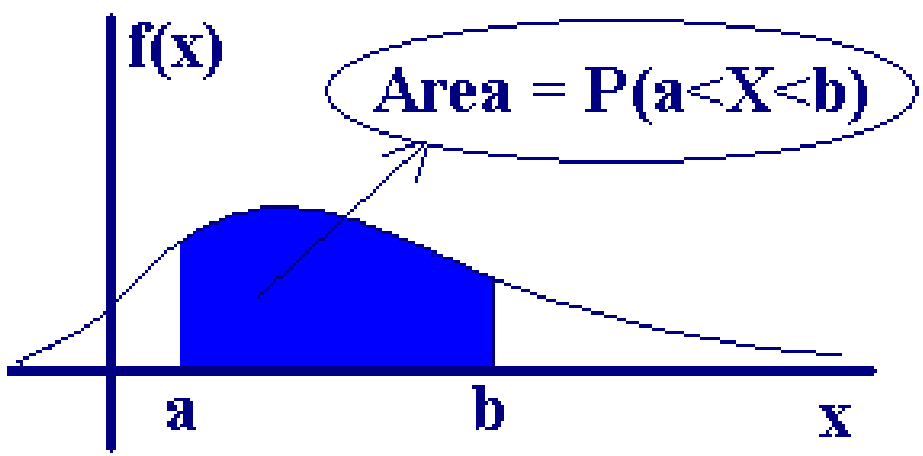
- describe continuous random variables

- calculate their probability distribution and their cumulative distribution

CONTINUOUS RANDOM VARIABLE

• A CONTINUOUS random variable is a random variable with an interval of real numbers for its range.

▪ The probability density function $f(x)$ of a continuous r.v. is used to determine probabilities from areas as follows:



$$P(a < X < b) = \int_a^b f(x)\,dx$$

CONTINUOUS RANDOM VARIABLE

- The properties of the probability density function $f(x)$ are:

$$i)\ f(x) \geq 0$$

$$ii)\ \int_{-\infty}^{\infty} f(x)dx = 1$$

- If X is continuous r.v., for any x_1 and x_2,

$$P(x_1 \leq X \leq x_2) = P(x_1 < X \leq x_2) = P(x_1 \leq X < x_2) = P(x_1 < X < x_2)$$

CUMULATIVE DISTRIBUTION FUNCTION

The cumulative distribution F(x) of a continuous random variable X with probability density function $f(x)$ is

$$\boxed{F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt \qquad \text{for} \qquad -\infty < x < \infty}$$

MEAN AND VARIANCE :- Continuous R.V

Mean or expected value is

$$\boxed{E(X) = \mu_X = \mu = \int_{-\infty}^{+\infty} x\, f(x)\, dx}$$

Variance of X is given as

$$\boxed{\sigma^2 = V(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2}$$

Standard deviation of X is σ .

Example 1

Suppose that *X* is a continuous random variable having the probability density function

$$f(\mathrm{x}) = \begin{cases} k\mathrm{x}^2 & for \quad -1 < \mathrm{x} < 1 \\ 0, & elsewhere \end{cases}$$

Find,

i. the value of constant *k ;*
ii. *P(-0.5<X<0.5)*;
iii. *x* such that P(*X* > *x*) = 0.5;
iv. the mean and the variance of *X;*
v. the CDF of *X.*

Given probability density function:

$$f(\mathrm{x}) = \begin{cases} k\mathrm{x}^2 & for \quad -1 < \mathrm{x} < 1 \\ 0, & elsewhere \end{cases}$$

1. Find the value of constant *k*:

$$\int f(\mathrm{x}) = \int_{-1}^{1} k\mathrm{x}^2 dx = \left.\frac{kx^3}{3}\right|_{-1}^{1} = \frac{2k}{3} = 1 \Rightarrow k = \frac{3}{2}$$

Given probability density function:

$$f(\mathrm{x}) = \begin{cases} \dfrac{3}{2}\mathrm{x}^2 & \textit{for} \quad -1 < \mathrm{x} < 1 \\ 0, & \textit{elsewhere} \end{cases}$$

2. $$p(-0.5 < x < 0.5) = \int_{-0.5}^{0.5} f(x)dx = \int_{-0.5}^{0.5} \frac{3}{2}x^2 dx = \frac{1}{8}$$

3. $$P(X > x) = 0.5 \Rightarrow \int_x^1 \frac{3}{2}x^2 dx = \left.\frac{x^3}{2}\right|_x^1 = \frac{1}{2} - \frac{x^3}{2} = \frac{1}{2}$$

$$\Rightarrow x = 0$$

Given probability density function:

$$f(\mathrm{x}) = \begin{cases} \dfrac{3}{2}\mathrm{x}^2 & for \quad -1 < \mathrm{x} < 1 \\ 0, & elsewhere \end{cases}$$

4. $$E(x) = \int_{-1}^{1} xf(x)dx = \int_{-1}^{1} x\frac{3}{2}x^2 dx = \left.\frac{3x^4}{8}\right|_{-1}^{1} = 0.$$

$$E(x^2) = \int_{-1}^{1} x^2 f(x)dx = \int_{-1}^{1} x^2 \frac{3}{2}x^2 dx = \left.\frac{3x^5}{10}\right|_{-1}^{1} = \frac{6}{10} = 0.6$$

$$\Rightarrow V(x) = E(x^2) - (E(x))^2 = 0.6 - 0 = 0.6$$

$$\Rightarrow sd(x) = \sqrt{0.6} = 0.775$$

Given probability density function:

$$f(\mathrm{x}) = \begin{cases} \dfrac{3}{2}\mathrm{x}^2 & \mathit{for} \quad -1 < \mathrm{x} < 1 \\ 0, & \mathit{elsewhere} \end{cases}$$

5. CDF(x) :

$$x < -1, \quad F(x) = 0$$

$$-1 \le x < 1, \quad F(x) = \int_{-1}^{x} \frac{3}{2} x^2 dx = \left.\frac{x^3}{2}\right|_{-1}^{x} = \frac{x^3}{2} + \frac{1}{2}$$

$$x \ge 1, \quad F(x) = \int_{-1}^{1} \frac{3}{2} x^2 dx + \int_{1}^{x} 0\, dx = \left.\frac{x^3}{2}\right|_{-1}^{1} = 1$$

$\therefore$ CDF(x) is :

$$F(x) = \begin{cases} 0 & , x < -1 \\ \dfrac{x^3}{2} + \dfrac{1}{2} & , -1 \le x < 1 \\ 1 & , x \ge 1 \end{cases}$$

Exercise

The density function is

$$f(x) = \begin{cases} k(1 - x^4), & for \quad 0 < x < 1 \\ 0, & elsewhere \end{cases}$$

Find

- the value of constant *k*;
- *P(0.25 < X < 0.5)*
- the mean of *X*; and
- the variance of *X*.
- the *cdf, F(x);*

Exercise

A commuter travels into town by KTM train and then has to catch a bus from station to the office. The time, *X* minutes, that the commuter has to wait for the bus can be modelled by the probability density function:

$$f(x) = \begin{cases} k & , 0 \le x \le 4 \\ 0.5k(6-x) & , 4 < x \le 6 \\ 0 & , \text{otherwise} \end{cases}$$

where *k* is a constant.

i. Find the value of *k*.

ii. Sketch the graph of *F*(*x*) versus *X.*

iv. Using *F*(*x*)*,* show that the probability the commuter has to wait longer than 5.5 minutes for the bus is 0.0125.

v. Find the cumulative distribution function of *X*.