RANDOM VARIABLES

RANDOM VARIABLES

- Outcomes of a random experiment can be numerical or non-numerical
  - → Even if it's non-numerical, there is often a numerical aspect we are interested in
    - ↳ Example:
      1) Toss a coin 10 times : X = number of heads obtained
      2) Roll a dice until 6 is obtained : Y = number of rolls required
- Let X denote the **NUMERICAL** outcome of interest of a random experiment
  - → X = random variable
- Value to be observed for a random variable cannot be determined in advance
  - → We simply assume we know all possible values of the random variable
- Can be seen as the **SAMPLE SPACE** of the random experiment
- Types:
  - **DISCRETE** random variable
    - → Number of possible value is '**COUNTABLE**'
    - → Usually associated with 'counting' experiment
    - → Example:

      X = number of heads obtained when a coin is tossed 2 times

      X = 0, 1, 2
  - **CONTINUOUS** random variable = '**MEASURABLE**'
    - → Possible values can change continuously in a certain interval
    - → Usually associated with 'measurement' experiment
    - → Example:

      X = number of hours a cat sleeps a day

      Possible values of X : (0, 24)

NOTE : We are interested in some events related to a random variable and the probabilities that they will occur

→ Example :

What is the probability of obtaining more than 5 heads when a coin is tossed 10 times : $P(X \geq 5)$?

→ Knowing the **PROBABILITY DISTRIBUTION** of X will allow us to assign probability to any event related to X

DISCRETE PROBABILITY DISTRIBUTION

- When X is a discrete random variable
- Probability distribution of X is usually given in the form of :

• **A TABLE**

x	x_1	x_2	x_3	x_4	...
$P(X = x_i)$	p_1	p_2	p_3	p_4	...

• An **EQUATION** → Called **PROBABILITY MASS FUNCTION**

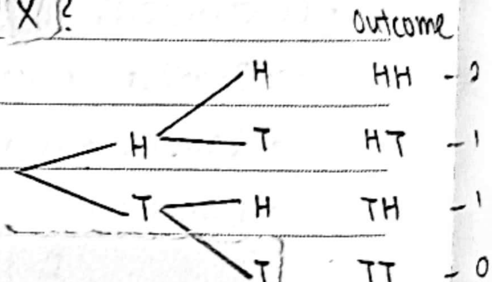
$$P(X = x_i) = f(x_i),\ i = 1, 2, 3, 4, \ldots$$

↳ Example :

A coin is tossed twice. X = number of heads obtained

What is the probability distribution of X ?

Possible value of X : X = 0, 1, 2



x	0	1	2
$P(X = x)$	0.25	0.5	0.25

NOTE : Probability distribution (Discrete) satisfies :

1) $0 \leq P(X=x) \leq 1$

2) $\sum P(X=x) = 1$

- **CUMULATIVE DISTRIBUTION FUNCTION**

→ Probability distribution can be represented with cumulative distribution function, denoted by F

$$F(x) = P(X \leq x), \text{ for any } x$$

→ Example :

Using the data from before, construct a cumulative distribution table for X.

x	< 0	$0 \leq x < 1$	$1 \leq x < 2$	$x \geq 2$
$F(x)$	0	0.25	0.75	1

- **EXPECTED VALUE** → denoted by μ_x or $E(X)$

→ AKA **MEAN** value

→ Formula :

$$E(X) = \sum x P(X=x)$$

→ Example :

Find the expected value of the previous data.

$$E(X) = 0(0.25) + 1(0.5) + 2(0.25)$$
$$= 1$$

- **VARIANCE** → denoted by σ_x^2 or $Var(X)$
  - → Measure of dispersion / spread of the values of X around its mean
  - → Formula :

1) $Var(X) = E(X-\mu)^2$

$$= \sum (x-\mu)^2 P(X=x)$$

2) $Var(X) = E(X^2) - \mu^2$

$$= \sum x^2 P(X=x) - (E(X))^2$$

$$= \sum x^2 P(X=x) - \left(\sum x P(X=x)\right)^2$$

→ Example :

Find the variance of the previous data.

1) $Var(X) = E(X-\mu)^2$

$$= (0-1)^2(0.25) + (1-1)^2(0.5) + (2-1)^2(0.25)$$

$$= 0.5$$

2) $Var(X) = E(X^2) - \mu^2$

$$= 0^2(0.25) + 1^2(0.5) + 2^2(0.25) - 1^2$$

$$= 0.5$$

NOTE : STANDARD DEVIATION (SD) → denoted by σ_x or $SD(X)$

→ Formula : $SD(X) = \sqrt{Var(X)}$

→ Example :

Find the SD of the previous data.

$$SD(X) = \sqrt{0.5}$$

$$= 0.7071$$

SOME PROPERTIES OF EXPECTED VALUE AND VARIANCE

1) $E(a) = a$, $Var(a) = 0$

2) $E(a + bX) = E(a) + E(bX)$
$= a + bE(X)$

3) $Var(a + bX) = Var(a) + Var(bX)$
$= 0 + b^2 Var(X)$
$= b^2 Var(X)$

4) $SD(a + bX) = SD(a) + SD(bX)$
$= 0 + |b| SD(X)$
$= |b| SD(X)$

EXAMPLE

X is a discrete random variable with probability distribution

$$P(X = x) = \begin{cases} k & \text{for } x = 2, 4, 6 \\ 2kx & \text{for } x = 1, 3, 5 \end{cases}$$

Find :

a) the value of k

$P(X=1) + P(X=2) + P(X=3) + P(X=4) + P(X=5) + P(X=6) = 1$

$2k + k + 6k + k + 10k + k = 1$

$21k = 1$

$k = \frac{1}{21}$

b) $E(X)$ and $Var(X)$

$E(X) = \sum x P(X = x)$

$= 1\left(\frac{2}{21}\right) + 2\left(\frac{1}{21}\right) + 3\left(\frac{6}{21}\right) + 4\left(\frac{1}{21}\right) + 5\left(\frac{10}{21}\right) + 6\left(\frac{1}{21}\right)$

$= \frac{82}{21}$

$$Var(X) = E(X^2) - \mu^2$$
$$= \sum x^2 P(X=x) - \left(\frac{82}{21}\right)^2$$
$$= 1^2\left(\frac{2}{21}\right) + 2^2\left(\frac{1}{21}\right) + 3^2\left(\frac{6}{21}\right) + 4^2\left(\frac{1}{21}\right) + 5^2\left(\frac{10}{21}\right) + 6^2\left(\frac{1}{21}\right) - \frac{6724}{441}$$
$$= \frac{878}{441}$$

c) $P(X > E(X))$

$$P(X > E(X)) = P(X=2) + P(X=3) + P(X=4) + P(X=5) + P(X=6)$$
$$= \frac{1}{21} + \frac{6}{21} + \frac{1}{21} + \frac{10}{21} + \frac{1}{21}$$
$$= \frac{19}{21}$$

d) $Var(3X)$

$$Var(3X) = 3^2 Var(X)$$
$$= 9\left(\frac{878}{441}\right)$$
$$= \frac{878}{49}$$

e) $E(4X+1)$

$$E(4X+1) = 4E(X) + 1$$
$$= 4\left(\frac{82}{21}\right) + 1$$
$$= \frac{349}{21}$$

SPECIAL PROBABILITY DISTRIBUTION

DISCRETE PROBABILITY DISTRIBUTION

- BINOMIAL DISTRIBUTION
- POISSON DISTRIBUTION

BINOMIAL DISTRIBUTION

- Binomial trial (AKA Bernoulli Trial)
  → A random experiment with only **TWO** possible outcomes:
  - Success (1) → P(success) = p
  - Failure (0) → P(failure) = $1 - p$ (q)
- Binomial experiment
  → Perform n independent binomial trials
  ↳ Probability of success, p, remains the **SAME** in each trial
  ↳ X = number of successes in n trials
  → X is said to have a binomial distribution with parameters n and p, denoted by $X \sim \text{Binomial}(n, p)$
  → Probability distribution is given by:

$$P(X = r) = {}^{n}C_{r}\, p^{r} q^{n-r} \quad *$$

→ Example:

1) 5% of adults in a population are colour-blind. A sample of 10 adults are selected. Let X = number of colour-blind adults, find:

a) The probability distribution of X

$$X \sim \text{Binomial}(10, 0.05)$$

* Formula is available on formula sheet, no. 11

b) The probability that:

i) No adults are colour-blind.

$$P(X=0) = {}^{10}C_0 (0.05)^0 (0.95)^{10}$$
$$= 0.5987$$

ii) Exactly one adult is colour-blind

$$P(X=1) = {}^{10}C_1 (0.05)^1 (0.95)^9$$
$$= 0.3151$$

2) 40% of students wear glasses. In a sample of 10 students, find the probability that:

a) Exactly 3 students wear glasses.

$$P(X=3) = {}^{10}C_3 (0.4)^3 (0.6)^7$$
$$= 0.2150$$

b) More than 7 students wear glasses

$$P(X>7) = P(X=8) + P(X=9) + P(X=10)$$
$$= {}^{10}C_8 (0.4)^8 (0.6)^2 + {}^{10}C_9 (0.4)^9 (0.6)^1 + {}^{10}C_{10} (0.4)^{10} (0.6)^0$$
$$= 0.0123$$

c) At least 1 student wear glasses

$$P(X>0) = 1 - P(X=0)$$
$$= 1 - {}^{10}C_0 (0.4)^0 (0.6)^{10}$$
$$= 0.9940$$

* Formula are available on formula sheet, no. 12 and 13, respectively

---

- Mean and Variance of Binomial
  - → **MEAN** : $E(X) = np$ *
  - → **VARIANCE** : $Var(X) = npq$
  - → **STANDARD DEVIATION** : $SD(X) = \sqrt{Var(X)} = \sqrt{npq}$ *
  - → Example :

Find the mean, variance and SD of the colour-blind adults.

Mean : $E(X) = 10(0.05) = 0.5$

Variance : $Var(X) = 10(0.05)(0.95) = 0.475$

SD : $SD(X) = \sqrt{0.475} = 0.6892$

POISSON DISTRIBUTION

- When X is a **DISCRETE** random variable which represents the number of times a random event occurs in a **INTERVAL OF TIME/SPACE**
  - → Example :
    - Number of accidents in a day
    - Number of typing mistakes in a page
    - Number of bacteria in 1 ml of lake water
- Probability distribution is given by :

$$P(X = x) = \frac{e^{-\lambda} \lambda^{x}}{x!} \quad *$$

- If X has a Poisson distribution, it is denoted by $X \sim P_0(\lambda)$

NOTE : λ is the **MEAN** number of times an event occurs in an interval of time/space

---

* Formula is available on formula sheet, no. 14

* Formula are available on formula sheet, no. 15 and 16, respectively

- Mean and Variance of Poisson
  - → MEAN : $E(X) = \lambda$ *
  - → VARIANCE : $Var(X) = \lambda$ *
  - → STANDARD DEVIATION : $SD(X) = \sqrt{Var(X)} = \sqrt{\lambda}$

- Example :

1) The mean number of bacteria per ml of liquid is 2. Assume that the number of bacteria follows a Poisson distribution, find :

a) The probability distribution of X.

$$X \sim P_o(2)$$

b) The probability that :

i) There's no bacteria

$$P(X=0) = \frac{e^{-2}2^0}{0!} = 0.1353$$

ii) There's less than 2 bacteria

$$P(X<2) = P(X=0) + P(X=1)$$
$$= 0.1353 + \frac{e^{-2}2^1}{1!}$$
$$= 0.4060$$

2) The number of goals a football team score followed a Poisson distribution with mean λ. If the number of zero goals has a probability of 0.2231, find the mean λ.

$$P(X=0) = \frac{e^{-\lambda}\lambda^0}{0!}$$
$$0.2231 = e^{-\lambda}$$
$$\ln(0.2231) = -\lambda$$

$$-\lambda = -1.500$$
$$\lambda = 1.5$$

3) The mean number of people in a library in 30 minutes is 4. Find the probability that:

a) There is no one in the library in 3 minutes

$$\lambda = 4 \times \frac{3}{30} = 0.4$$

$$P(X=0) = \frac{e^{-0.4}\, 0.4^{0}}{0!} = 0.6703$$

b) There is at least 2 person in the library in 1.5 minutes

$$\lambda = 4 \times \frac{1.5}{30} = 0.2$$

$$P(X \geq 2) = 1 - \left(P(X=0) + P(X=1)\right)$$

$$= 1 - \left(\frac{e^{-0.2}\, 0.2^{0}}{0!} + \frac{e^{-0.2}\, 0.2^{1}}{1!}\right)$$

$$= 0.01752$$

c) There is at most 2 person in the library in 6 minutes

$$\lambda = 4 \times \frac{6}{30} = 0.8$$

$$P(X \leq 2) = P(X=0) + P(X=1) + P(X=2)$$

$$= \frac{e^{-0.8}\, 0.8^{0}}{0!} + \frac{e^{-0.8}\, 0.8^{1}}{1!} + \frac{e^{-0.8}\, 0.8^{2}}{2!}$$

$$= 0.9526$$

NOTE: If the mean for x time/space given is different from the time/space in the question, just change it accordingly using:

$$\lambda = \text{Mean given} \times \frac{\text{Time/Space in question}}{\text{Original time/space}}$$

Example: Mean of people in one room = 3

Mean of people in two room = $3 \times \frac{2}{1}$

= 6

CONTINUOUS PROBABILITY DISTRIBUTION

- NORMAL DISTRIBUTION
- STANDARD NORMAL DISTRIBUTION

NORMAL DISTRIBUTION

- A **CONTINUOUS** random variable X is said to have a normal distribution with parameters μ and σ, denoted by $X \sim N(\mu, \sigma^2)$ if its probability density function is given by:

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \quad -\infty < x < \infty$$

- Characteristics:
  1) Takes value from $-\infty$ to ∞
  2) **BELL** shaped
  3) **SYMMETRICAL** about the mean

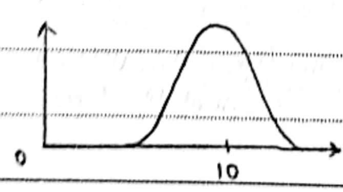
NOTE: μ is the **MEAN** of X, it determines the 'location' of X

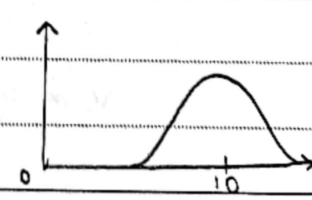
→ The maximum point of the bell shaped graph is μ

σ is the **SD** of X, it determines the 'shape' of X

→ The greater σ is, the flatter the graph is

Example: $X \sim N(10, 2)$ $\quad$ $X \sim N(10, 3)$





* Formula is available in formula sheet, no. 17

STANDARD NORMAL DISTRIBUTION

- A normal distribution with mean 0 and SD 1 is denoted by $Z \sim N(0,1)$
- **STANDARDIZATION** of X

↳ Any normal distribution, $X \sim N(\mu, \sigma)$, can be **TRANSFORMED** into a standard normal:

$$\frac{X-\mu}{\sigma} \sim N(0,1)$$

NOTE: $Z = \frac{X-\mu}{\sigma}$ *

NOTE: To find **PROBABILITY** of any **NORMAL** random variable

→ Carry out standardization first

$$P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right)$$
$$= P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right)$$

→ Then look up the probability from the normal distribution table *

- Example:

1) $X \sim N(10, 2)$, find:

a) $P(X < 15)$

$$P(X < 15) = P\left(Z < \frac{15-10}{2}\right)$$
$$= P(Z < 2.5)$$
$$= 0.9938$$

* Normal distribution table is provided along with formula sheet

b) $P(X > 8)$

$$P(X > 8) = P(Z > \frac{8-10}{2})$$
$$= P(Z > -1)$$
$$= P(Z < 1)$$
$$= 0.8413$$

c) $P(8 < X < 12)$

$$P(8 < X < 12) = P(\frac{8-10}{2} < Z < \frac{12-10}{2})$$
$$= P(-1 < Z < 1)$$
$$= P(0 < Z < 1) + P(0 < Z < 1)$$
$$= 2(0.8413 - 0.5)$$
$$= 0.6826$$

2) $X \sim N(68, 16)$, find:

a) $P(X < 58)$

$$P(X < 58) = P(Z < \frac{58-68}{16})$$
$$= P(Z < -0.625)$$
$$= P(Z > 0.63)$$
$$= 1 - P(Z < 0.63)$$
$$= 1 - 0.7357$$
$$= 0.2643$$

NOTE : Important things to remember :

1) if there are 3 or more decimals , **ROUND** to nearest 2 decimals

2) it must be $z <$ whatever number in order to count

↳ Negative number must change to positive = Change the sign too

↳ If its $z >$ positive number , make it into $1 - (z < \text{number})$

3) If there are two $<$ signs , split it into two $(0 < z < \text{number})$

4) If question is $(z < m)$ = number , find the number in the table

↳ If there's no exact 'z' value for the number , take the closer one that is bigger than the number

5) If it's the top 10% , it will be $Z_{0.1}$, if it's bottom 10% , it will be $-Z_{0.1}$

NORMAL DISTRIBUTION AS APPROXIMATION TO BINOMIAL DISTRIBUTION

- A binomial distribution with parameters n and p can be **APPROXIMATE** by a normal distribution $X \sim N(np, \sqrt{npq})$
  - → where $\mu = np$, $\sigma = \sqrt{npq}$
- Can only be used if $np \geq 5$ or $nq \geq 5$ or $n \geq 30$
- When binomial distribution is approximated by normal distribution, discrete distribution is changed to continuous distribution
  = Must be corrected with **CONTINUITY CORRECTION**

CONTINUITY CORRECTION

- Binomial ⟶ Normal

$P(X = a)$	$P(a-0.5 < X < a+0.5)$
$P(X \geq a)$	$P(X > a-0.5)$
$P(X > a)$	$P(X > a+0.5)$
$P(X \leq a)$	$P(X < a+0.5)$
$P(X < a)$	$P(X < a-0.5)$
$P(a \leq X \leq b)$	$P(a-0.5 < X < b+0.5)$
$P(a < X < b)$	$P(a+0.5 < X < b-0.5)$

- Example :

A regular tetrahedral shaped dice with 1, 2, 3 and 4 is tossed 200 times. Find the probability of obtaining the digit 4 :

a) More than 60 times

$X \sim Binomial(200, 0.25)$

since $np = 200 \times 0.25 = 50 > 5$, normal distribution is used as approximation

$\sqrt{npq} = \sqrt{(200)(0.25)(0.75)} = 6.12$

$X \sim N(50, 6.12)$

$$\begin{aligned} P(X > 60) &= P(X > 60.5) \\ &= P\left(Z > \frac{60.5 - 50}{6.12}\right) \\ &= P(Z > 1.72) \\ &= 1 - P(Z < 1.72) \\ &= 1 - 0.9573 \\ &= 0.0427 \end{aligned}$$

b) At most 50 times

$$\begin{aligned} P(X \leq 50) &= P(X < 50.5) \\ &= P\left(Z < \frac{50.5 - 50}{6.12}\right) \\ &= P(Z < 0.08) \\ &= 0.5319 \end{aligned}$$

c) Exactly 50 times

$$P(X = 50) = P(49.5 < X < 50.5)$$
$$= P\left(\frac{49.5-50}{6.12} < z < \frac{50.5-50}{6.12}\right)$$
$$= P(-0.08 < z < 0.08)$$
$$= P(0 < z < 0.08) + P(0 < z < 0.08)$$
$$= 2(0.5319 - 0.5)$$
$$= 0.0638$$

d) More than 30 times but less than 70 times

$$P(30 < X < 70) = P(30.5 < x < 69.5)$$
$$= P\left(\frac{30.5-50}{6.12} < x < \frac{69.5-50}{6.12}\right)$$
$$= P(-3.19 < x < 3.19)$$
$$= P(0 < x < 3.19) + P(0 < x < 3.19)$$
$$= 2(0.9993 - 0.5)$$
$$= 0.9986$$

e) At least 55 times

$$P(X \geq 55) = P(X > 54.5)$$
$$= P\left(z > \frac{54.5-50}{6.12}\right)$$
$$= P(z > 0.74)$$
$$= 0.7704$$