STATISTICAL ANALYSIS

DISCRETE RANDOM VARIABLES (I)

Contents include:

- Random Variables
- Types of Data
- Discrete Random Variable Table

www.mathifyhsc.com.au

- What is a Random Variable?

A random variable is a variable which represents the outcome of a trial, an experiment or event. It is a specific number that varies with each different trial and is usually represented by X.

For example, let's say I throw two dice:

If I let X be the random variable equal to the sum of the two dice after I throw them, then:

$$P(getting\ a\ sum\ of\ 12)\ is\ expressed\ as\ P(X = 12)$$

Where:

$$P(X = 12) = \frac{1}{36}$$

Example 1: I have a bag of 10 marbles where 4 are red and 6 are blue. If I take out 2 marbles without replacement and let $X = number\ of\ red\ marbles$, then calculate $P(X = 2)$

Solution:

Since $X =$ number of red marbles:

$$\begin{aligned} P(X = 2) &= \frac{4}{10} \times \frac{3}{9} \\ &= \frac{12}{90} \\ &= \frac{2}{15} \end{aligned}$$

- Types of Data

There are two types of data which we can consider:

1. **Discrete Data:** Data which can only take fixed whole number values and may be listed, so written down in the form x_1, x_2, x_3 ... Examples include number of patients in a hospital, number of apples on a tree
2. **Continuous Data:** Data which can take on an infinite number of outcomes between two intervals and can't be written in a list. Examples include weight and height

In summary, discrete random variables only have a countable number of outcomes; we can only have 100 or 101, etc. number of patients, we can't have 100.5 patients.

Continuous random variables on the other hand have an uncountable number of outcomes; our height may be 180.01cm, or 180.0013 cm, or 180.0000098 cm, etc. As you can see, there is a continuous, infinite interval for which the exact height of someone may be, it's just that conventionally we may not be used to this idea since we estimate to the nearest cm.

- Discrete Random Variable Table

Let's say we conduct an experiment where each event is a discrete result, for example, rolling a dice where the outcome can only be either $1, 2, 3, 4, 5, or\ 6$. We may express each event and the probability of each event in a table like the following:

x	1	2	3	4	5	6
$P(X = x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Note that $P(X = x)$ represents the probability of an event, x occurring.

Example 2: Using the discrete probability table provided above, calculate:

i. $P(X = 3)$
ii. $P(X \geq 4)$
iii. $P(X < 3)$
iv. $P(2 \leq X < 5)$

Solutions:

i. Looking at the table:

$$P(X = 3) = \frac{1}{6}$$

ii. Notice how for the discrete probability distribution, 4, 5 and 6 are the only integers that can fit within the interval $X \geq 4$

$$\begin{aligned} \therefore P(X \geq 4) &= P(X = 4) + P(X = 5) + P(X = 6) \\ &= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} \\ &= \frac{3}{6} \\ &= \frac{1}{2} \end{aligned}$$

iii. Notice how for the discrete probability distribution, 1 and 2 are the only integers that can fit within the interval $X < 2$

$$\begin{aligned} \therefore P(X < 3) &= P(X = 1) + P(X = 2) \\ &= \frac{1}{6} + \frac{1}{6} \\ &= \frac{2}{6} \\ &= \frac{1}{3} \end{aligned}$$

iv. Notice how for the discrete probability distribution, 2, 3 and 4 are the only integers that can fit within the interval $2 \leq X < 5$

$$\therefore P(2 \leq X < 5) = P(X = 2) + P(X = 3) + P(X = 4)$$

$$= \frac{1}{6} + \frac{1}{6} + \frac{1}{6}$$
$$= \frac{3}{6}$$
$$= \frac{1}{2}$$

- The Sum of probabilities

Recall that if something has a probability of 1, then it is certain. This therefore means that the sum of all the probabilities of possible events must equal to 1. In other words:

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

This is helpful when attempting to **calculate an unknown probability** within a discrete probability table!

Example 3: Using the following table, calculate the value of a:

x	0	1	2	3	4	5
$P(X = x)$	$2a$	$3a$	a	$7a$	$4a$	$3a$

Solution:

Since the sum of all probabilities is equal to 1:

$$\therefore 2a + 3a + a + 7a + 4a + 3a = 1$$
$$20a = 1$$
$$\therefore a = \frac{1}{20}$$

