DISCRETE RANDOM VARIABLES

When performing a sampling procedure, a number of different outcomes are expected. For example, when rolling a normal die a large number of times, you can expect to observe some of each of the values from $\{1,2,3,4,5,6\}$. The outcome can vary between rolls. X, the observed outcome, is called a **random variable**. In particular, X is a **discrete random variable**, because the list of possible outcomes is countable. Discrete random variables are often associated with number or size.

On the other hand, if the list of possible outcomes is not countable, the variable is a **continuous random variable.** For example, when measuring the heights of a sample of people, although you might expect the measurements to fall within a range, say 140 cm to 190 cm, each individual value is dependent only of the degree of accuracy of the measuring instrument. Continuous random variables, to be covered at a later stage, are often associated with height, mass and time.

NOTATION:

Capital letters (e.g. X and Y) are used for random variables, and their corresponding lower-case letters (e.g. x and y) are used for the values that the random variable takes. Subscripts distinguish the various possible values of X. For example, we denote $P(X = x_2)$ the probability that the random variable X takes the value x_2.

Discrete or Continuous random variables?

- the number of wristbands that the next student walking in the corridor has:
- the number of leaves on each tree in the playground:
- the weight of a blue whale:
- the time it takes to fry an egg:

Example: Consider the experiment where three coins are tossed. Let X be the random variable which stands for the number of Heads obtained. There are 8 possible outcomes:

Outcome observed	HHH	HHT	HTH	THH	HTT	THT	TTH	TTT
Value of X	3	2	2	2	1	1	1	0

So the random variable X can take values from $\{0,1,2,3\}$. Assuming that each of the observed outcomes is equally likely, which is a reasonable assumption using fair coins, a probability table can be created for the variable X.

x	0	1	2	3
$P(X = x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

The sum of all these probabilities is equal to 1, i.e.:

$$P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = 1$$

More generally, if a random variable X would take values from $\{x_1, x_2, x_3, \dots, x_n\}$, we would have:

$$P(X = x_1) + P(X = x_2) + P(X = x_3) + \cdots + P(X = x_n) = 1$$

To shorten this expression, the symbol $\sum$ is used, meaning "the sum of"; so the expression above simplifies as:

$$\sum_{i=1}^{n} P(X = x_i) = 1$$

In this notation, $i = 1$ (underneath the $\sum$ symbol) means that the indexes start at 1, whereas the n above the $\sum$ symbol means that the indexes finish at n.

So in the example of the experiment above with the tossing of 3 coins, that would be noted:

$$\sum_{i=0}^{3} P(X = x_i) = 1$$

Example 3

Does the following table represent a discrete probability distribution?

x	1	3	5	7
$P(X=x)$	$\frac{1}{5}$	$\frac{1}{10}$	$\frac{3}{10}$	$\frac{2}{5}$

Solution

Are all the probabilities between 0 and 1, inclusive; i.e. is $0 \le P(X) \le 1$? Yes.

Do the probabilities add up to 1?

$$\begin{aligned}\sum P(X=x) &= \frac{1}{5} + \frac{1}{10} + \frac{3}{10} + \frac{2}{5} = \frac{2}{10} + \frac{1}{10} + \frac{3}{10} + \frac{4}{10} \\ &= \frac{10}{10} \\ &= 1\end{aligned}$$

Both conditions have been met, so the table of data represents a discrete probability distribution.

Example 4

The table of data below represents a discrete probability distribution.

x	3	4	5	6	7
$P(X=x)$	0.14	k	0.36	0.21	0.13

Find the value of k.

Solution

Add up the given probabilities: $0.14 + k + 0.36 + 0.21 + 0.13 = 0.84 + k$

As the sum of probabilities $= 1$, then: $0.84 + k = 1$

Solve for k: $k = 1 - 0.84$

$k = 0.16$