

# Introduction to Probability

There are many situations in everyday life where the outcome is not known with certainty. For example; applying for a job or sitting an examination. We use words like "Chance", "the odds", "likelihood" etc but the most effective way of dealing with uncertainty is based on the concept of probability.

Probability can be thought of as a number which measures the chance or likelihood that a particular event will occur.

An example of the use of probability is in decision making. Decision making usually involves uncertainty. For example, should we invest in a company if there is a chance it will fail?

Should we start production of a product even though there is a likelihood that the raw materials will arrive on time in poor? Having a number which measures the chances of these events occurring helps us to make a decision.

Why are we interested in probability in this module? Many statistical methods use the idea of a probability distribution for this data.

We have already looked at relative frequency distribution in Section 2.

Probability distributions are based on the same concepts as relative frequency distributions. They are used to calculate probabilities of different values occurring in the data collected.

We will examine probability distributions in more detail in Section 4. First we need to learn about the basic concepts of probability.

# Random experiment

- **Sample Space**,  $S$ . For a given experiment the sample space,  $S$ , is the set of all possible outcomes.
- **Event**,  $E$ . This is a subset of  $S$ . If an event  $E$  occurs, the outcome of the experiment is contained in  $E$ .

Probability concerns itself with random phenomena or probability experiments. These experiments are all different in nature, and can concern things as diverse as rolling dice or flipping coins. The common thread that runs throughout these probability experiments is that there are observable outcomes. If we collect all of the possible outcomes together, then this forms a set that is known as the sample space.

In this set theory formulation of probability the sample space for a problem corresponds to an important set. Since the sample space contains every outcome that is possible, it forms a setting of everything that we can consider. So the sample space becomes the universal set in use for a particular probability experiment.

A probability distribution is a table of values showing the probabilities of various outcomes of an experiment.

For example, if a coin is tossed three times, the number of heads obtained can be 0, 1, 2 or 3. The probabilities of each of these possibilities can be tabulated as shown:

Number of Heads	0	1	2	3
Probability	$1/8$	$3/8$	$3/8$	$1/8$

A discrete variable is a variable which can only take a countable number of values. In this example, the number of heads can only take 4 values (0, 1, 2, 3) and so the variable is discrete. The variable is said to be random if the sum of the probabilities is one.

# Common Sample Spaces

Sample spaces abound and are infinite in number. But there are a few that are frequently used for examples in introductory statistics. Below are the experiments and their corresponding sample spaces:



# Common Sample Spaces

- For the experiment of flipping a coin, the sample space is Heads, Tails and has two elements.
- For the experiment of flipping two coins, the sample space is (Heads, Heads), (Heads, Tails), (Tails, Heads), (Tails, Tails) and has four elements.
- For the experiment of flipping three coins, the sample space is (Heads, Heads, Heads), (Heads, Heads, Tails), (Heads, Tails, Heads), (Heads, Tails, Tails), (Tails, Heads, Heads), (Tails, Heads, Tails), (Tails, Tails, Heads), (Tails, Tails, Tails) and has eight elements.
- For the experiment of flipping  $n$  coins, where  $n$  is a positive whole number, the sample space consists of  $2^n$  elements. There are a total of  $C(n, k)$  ways to obtain  $k$  heads and  $n - k$  tails for each number  $k$  from 0 to  $n$ .

# Forming Other Sample Spaces

These are the basic sample spaces. Others are out there for different experiments. It is also possible to combine several of the above experiments. When this is done, we end up with a sample space that is the Cartesian product of our individual sample spaces. We can also use a tree diagram to form these sample spaces.

# Probability

Two dice A and B are cast. What is the probability of getting

- 1 a total of 2 or a total of 6?
- 2 a total greater than 9?
- 3 a total which is three times as great as other possible totals?

# Probability

	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>
<b>1</b>	2	3	4	5	6	7
<b>2</b>	3	4	5	6	7	8
<b>3</b>	4	5	6	7	8	9
<b>4</b>	5	6	7	8	9	10
<b>5</b>	6	7	8	9	10	11
<b>6</b>	7	8	9	10	11	12

# Probability

	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>
<b>1</b>	2	3	4	5	6	7
<b>2</b>	3	4	5	6	7	8
<b>3</b>	4	5	6	7	8	9
<b>4</b>	5	6	7	8	9	10
<b>5</b>	6	7	8	9	10	11
<b>6</b>	7	8	9	10	11	12

# Probability

Part 1) A total of 2 or 6

$$E_1 = \{(1, 1), (1, 5), (5, 1), (4, 2), (2, 4), (3, 3)\}$$

$$P(E_1) = \frac{6}{36}$$

# Probability

Part 2) A total greater than 9

$$E_2 = \{(4, 6), (5, 5), (6, 4), (6, 5), (6, 6), (3, 3)\}$$

$$P(E_2) = \frac{6}{36}$$

# Probability

Part 3) A total which is three times as great as other possible totals.

These totals are 6, 9 and 12.

$$E_3 = \{(1, 5), (2, 4), (4, 2), (1, 5), (3, 3), (6, 3), (4, 5), (5, 4), (6, 6)\}$$

$$P(E_3) = \frac{10}{36}$$



# What is a contingency table?

A contingency table is essentially a display format used to analyse and record the relationship between two or more categorical variables. It is the categorical equivalent of the scatterplot used to analyse the relationship between two continuous variables.

# Combining Probabilities

Events rarely occur in isolation. Usually we are interested in a combination or compound of events; for example

- The probability that two sections of a factory will be understaffed on the same day
- The probability of having a car accident today, given that you have had a car accident in the last five years.

We will look at two laws of probability for combining events

- The Addition Law
- The multiplication Law

# Conditional Probability

The conditional probability of an event is the probability that an event A occurs given that another event B has already occurred. This type of probability is calculated by restricting the sample space that were working with to only the set B.

The formula for conditional probability can be rewritten using some basic algebra. Instead of the formula:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

# Histograms

A histogram is constructed from a frequency table. The intervals are shown on the X-axis and the number of scores in each interval is represented by the height of a rectangle located above the interval. A histogram of the response times from the dataset Target RT is shown below.

# Cumulative Distribution Function

The cumulative distribution function (c.d.f.) of a discrete random variable  $X$  is the function  $F(t)$  which tells you the probability that  $X$  is less than or equal to  $t$ . So if  $X$  has p.d.f.  $P(X = x)$ , we have:

$$F(t) = P(X \leq t)$$

In other words, for each value that  $X$  can be which is less than or equal to  $t$ , work out the probability that  $X$  is that value and add up all such results.

## Example

In the above example where the die is thrown repeatedly, let's work out  $P(X \leq t)$  for some values of  $t$ .

$P(X \leq 1)$  is the probability that the number of throws until we get a 6 is less than or equal to 1. So it is either 0 or 1.

- $P(X = 0) = 0$
- $P(X = 1) = 1/6$ .
- Hence  $P(X \leq 1) = 1/6$

Similarly,  $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$   
 $= 0 + 1/6 + 5/36 = 11/36$

- Combinations
- Permutations
- Permutations with constraints



# Permutations of subsets

The number of permutations of subsets of  $k$  elements selected from a set of  $n$  different elements is

$$P(n, r) = \frac{n!}{(n - k)!}$$

# Combinations of subsets

The number of combinations that can be selected from  $n$  items is

$$\binom{n}{k} = \frac{n!}{k! \times (n-k)!}$$

- A random variable is a numerical description of the outcome of an experiment.
- Random variables can be classified as discrete or continuous, depending on the numerical values they may take.
- A random variable that may assume any numerical value in an interval or collection of intervals is called a continuous random variable.