

# 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:

- 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  $n$ .
- For the experiment consisting of rolling a single six-sided die, the sample



- For the experiment of rolling two six-sided dice, the sample space consists of the set of the 36 possible pairings of the numbers 1, 2, 3, 4, 5 and 6.
- For the experiment of rolling three six-sided dice, the sample space consists of the set of the 216 possible triples of the numbers 1, 2, 3, 4, 5 and 6.
- For an experiment of drawing from a standard deck of cards, the sample space is the set that lists all 52 cards in a deck. For this example the sample space could only consider certain features of the cards, such as rank or suit.

## 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

- Probability theory is the mathematical study of randomness. A probability model of a random experiment is defined by assigning probabilities to all the different outcomes.
- Probability is a numerical measure of the likelihood that an event will occur. Thus, probabilities can be used as measures of degree of uncertainty associated with outcomes of an experiment. Probability values are always assigned on a scale from 0 to 1.
- A probability of 0 means that the event is impossible, while a probability near 0 means that it is highly unlikely to occur.
- Similarly an event with probability 1 is certain to occur, whereas an event with a probability near to 1 is very likely to occur.

# Experiments and Outcomes

- In the study of probability any process of observation is referred to as an experiment.
- The results of an experiment (or other situation involving uncertainty) are called the outcomes of the experiment.
- An experiment is called a random experiment if the outcome can not be predicted.
- Typical examples of a random experiment are
  - a role of a die,
  - a toss of a coin,
  - drawing a card from a deck.

If the experiment is yet to be performed we refer to possible outcomes or possibilities for short. If the experiment has been performed, we refer to realized outcomes or realizations.

# Sample Spaces and Events

- The set of all possible outcomes of a probability experiment is called a *sample space*, which is usually denoted by  $S$ .
- The sample space is an exhaustive list of all the possible outcomes of an experiment. We call individual elements of this list *sample points*.
- Each possible outcome is represented by one and only one sample point in the sample space.

# Sample Spaces: Examples

For each of the following experiments, write out the sample space.

- Experiment: Rolling a die once
  - Sample space  $S = \{1, 2, 3, 4, 5, 6\}$
- Experiment: Tossing a coin
  - Sample space  $S = \{Heads, Tails\}$
- Experiment: Measuring a randomly selected persons height (cms)
  - Sample space  $S =$  The set of all possible real numbers.

# Events

- An event is a specific outcome, or any collection of outcomes of an experiment.
- Formally, any subset of the sample space is an event.
- Any event which consists of a single outcome in the sample space is called an *elementary* or *simple event*.
- Events which consist of more than one outcome are called *compound events*.
- For example, an elementary event associated with the die example could be the “die shows 3”.
- An compound event associated with the die example could be the “die shows an even number”.

# The Complement Event

- The complement of an event  $A$  is the set of all outcomes in the sample space that are not included in the outcomes of event  $A$ .
- We call the complement event of  $A$  as  $A^c$ .
- The complement event of a die throw resulting in an even number is the die throwing an odd number.
- Question: if there is a 40% chance of a randomly selected student being male, what is the probability of the selected student being female?



# Set Theory : Union and Intersection

Set theory is used to represent relationships among events.

## **Union of two events:**

The union of events  $A$  and  $B$  is the event containing all the sample points belonging to  $A$  or  $B$  or both. This is denoted  $A \cup B$ , (pronounce as “A union B”).

## **Intersection of two events:**

The intersection of events  $A$  and  $B$  is the event containing all the sample points common to both  $A$  and  $B$ . This is denoted  $A \cap B$ , (pronounce as “A intersection B”).

# More Set Theory

In general, if  $A$  and  $B$  are two events in the sample space  $S$ , then

- $A \subseteq B$  ( $A$  is a subset of  $B$ ) = ‘if  $A$  occurs, so does  $B$ ’
- $\emptyset$  (the empty set) = an impossible event
- $S$  (the sample space) = an event that is certain to occur

# Examples of Events

Consider the experiment of rolling a die once. From before, the sample space is given as  $S = \{1, 2, 3, 4, 5, 6\}$ . The following are examples of possible events.

- $A = \text{score} < 4 = \{1, 2, 3\}$ .
- $B = \text{'score is even'} = \{2, 4, 6\}$ .
- $C = \text{'score is 7'} = \emptyset$
- $A \cup B = \text{'the score is } < 4 \text{ or even or both'} = \{1, 2, 3, 4, 6\}$
- $A \cap B = \text{'the score is } < 4 \text{ and even} = \{2\}$
- $A^c = \text{'event A does not occur'} = \{4, 5, 6\}$

# Sample Spaces

A complete list of all possible outcomes of a random experiment is called sample space or possibility space and is denoted by  $S$ .

A sample space is a set or collection of outcome of a particular random experiment.

For example, imagine a dart board. You are trying to find the probability of getting a bullseye. The dart board is the sample space. The probability of a dart hitting the dart board is 1.0. For another example, imagine rolling a six sided die. The sample space is 1, 2, 3, 4, 5, 6.

# Sample Spaces

The following list consists of sample spaces of examples of random experiments and their respective outcomes.

The tossing of a coin, sample space is Heads, Tails

The roll of a die, sample space is 1, 2, 3, 4, 5, 6

The selection of a numbered ball (1-50) in an urn, sample space is 1, 2, 3, 4, 5, ..., 50

# Sample Spaces

Percentage of calls dropped due to errors over a particular time period, sample space is  $\{2\%, 14\%, 23\%, \dots\}$

The time difference between two messages arriving at a message centre, sample space is  $0, \dots, \text{infinity}$

The time difference between two different voice calls over a particular network, sample space is  $0, \dots, \text{infinity}$