

# Contents

<b>1</b>	<b>5. Random Variables</b>	<b>2</b>
1.1	Random Experiments . . . . .	2
1.1.1	Random Variables . . . . .	2
1.1.2	Random Variables . . . . .	2
1.1.3	Random Variables . . . . .	3
1.2	Random Variables . . . . .	3
1.3	Random Variables . . . . .	3
1.3.1	Types of Probability Distribution . . . . .	3
1.4	Probability Distributions . . . . .	4
1.5	Random Variables . . . . .	4
1.6	Introduction to Distributions . . . . .	4
1.7	Introduction to Distributions . . . . .	6
1.8	Joint Probability Tables . . . . .	7
1.8.1	Probability Mass Function . . . . .	7
1.9	Discrete Probability Distributions . . . . .	7
3		

# Chapter 1

## 5. Random Variables

### 1.1 Random Experiments

- 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**

#### 1.1.1 Random Variables

There are two types of random variable - discrete and continuous. The distinction between both types will be important later on in the course.

##### Examples

- A coin is tossed ten times. The random variable  $X$  is the number of tails that are noted.  $X$  can only take the values  $\{0, 1, \dots, 10\}$ , so  $X$  is a discrete random variable.
- A light bulb is burned until it burns out. The random variable  $Y$  is its lifetime in hours.  $Y$  can take any positive real value, so  $Y$  is a continuous random variable.

#### 1.1.2 Random Variables

1. Random Variables
2. Expected Values of RVs

### 1.1.3 Random Variables

- The outcome of an experiment need not be a number, for example, the outcome when a coin is tossed can be ‘heads’ or ‘tails’.
- However, we often want to represent outcomes as numbers.
- A **random variable** is a function that associates a unique numerical value with every outcome of an experiment.
- The value of the random variable will vary from trial to trial as the experiment is repeated.
- Numeric values can be assigned to outcomes that are not usually considered numeric.
- For example, we could assign a ‘head’ a value of 0, and a ‘tail’ a value of 1, or vice versa.

## 1.2 Random Variables

A random variable is defined as a numerical event whose value is determined by a chance process. When probability values are assigned to all possible numerical values of a random variable  $X$ , either by a listing or by a mathematical function, the result is a probability distribution.

The sum of the probabilities for all the possible numerical outcomes must equal 1.0. Individual probability values may be denoted by the symbol  $f(x)$ , which indicates that a mathematical function is involved, by  $P(x = X)$ , which recognizes that the random variable can have various specific values, or simply by  $P(X)$ .

## 1.3 Random Variables

1. Random Variables
2. Expected Values of RVs

Hence, the random variable  $X$  should be thought of as the unknown numerical result of an experiment to be carried out ( $X$  is described by a distribution).

A realisation, denoted by a small letter, is a known result of an experiment already carried out.

**X** : random variable name (e.g Height)

**x** : realisation (e.g. 1.82 metres)

### 1.3.1 Types of Probability Distribution

- A probability distribution is a mathematical approach to quantifying uncertainty.
- There are two main classes of probability distributions: Discrete and continuous.
  - Discrete distributions describe variables that take on discrete values only (typically the positive integers),
  - Continuous distributions describe variables that can take on arbitrary values in a continuum (typically the real numbers).

## 1.4 Probability Distributions

Probability Distributions will be covered in detail in this course as part of the R component (i.e. weeks 8-13).

It is worth bearing in mind that all of the material that will be covered in the R section of the course can just as easily be implemented using MATLAB. In fact substantial use is made of these commands in real world applications, particularly in Finance and Engineering. As such we will briefly look at some.

- Suppose we have a set of **n** items.
- From that set, we create a subset of **k** items.
- The **order** in which items are selected is recorded. (The ordering of selected items is very important.)
- The total number of **ordered subsets** of **k** items chosen from a set of **n** items is

$$\frac{n!}{n - k!}$$

## Joint probability tables

A joint probability table is a table in which all possible events (or outcomes) for one variable are listed as row headings, all possible events for a second variable are listed as column headings, and the value entered in each cell of the table is the probability of each joint occurrence.

Often, the probabilities in such a table are based on observed frequencies of occurrence for the various joint events. The table of joint-occurrence frequencies which can serve as the basis for constructing a joint probability table is called a contingency table.

## 1.5 Random Variables

1. Random Variables
2. Expected Values of RVs

Hence, the random variable  $X$  should be thought of as the unknown numerical result of an experiment to be carried out ( $X$  is described by a distribution).

A realisation, denoted by a small letter, is a known result of an experiment already carried out.

**X** : random variable name (e.g Height)

**x** : realisation (e.g. 1.82 metres)

## 1.6 Introduction to Distributions

### Random Variables

- A random variable is a variable whose value is determined by the outcome of a random phenomenon. The statistical techniques we've learned so far deal with variables, not events, so we need to define a variable in order to analyze a random phenomenon.

- A discrete random variable has a finite number of possible values, while a continuous random variable can take all values in a range of numbers.
- The probability distribution of a random variable tells us the possible values of the variable and how probabilities are assigned to those values.
- The probability distribution of a discrete random variable is typically described by a list of the values and their probabilities. Each probability is a number between 0 and 1, and the sum of the probabilities must be equal to 1.
- The probability distribution of a continuous random variable is typically described by a density curve.
- The curve is defined so that the probability of any event is equal to the area under the curve for the values that make up the event, and the total area under the curve is equal to 1.
- One example of a continuous probability distribution is the normal distribution.
- We use the term parameter to refer to a number that describes some characteristic of a population.
- We rarely know the true parameters of a population, and instead estimate them with statistics.
- Statistics are numbers that we can calculate purely from a sample.
- The value of a statistic is random, and will depend on the specific observations included in the sample.
- The law of large numbers states that if we draw a bunch of numbers from a population with mean  $\mu$ , we can expect the mean of the numbers  $\bar{y}$  to be closer to  $\mu$  as we increase the number we draw. This means that we can estimate the mean of a population by taking the average of a set of observations drawn from that population.

### Discrete random variables

- For a discrete random variable observed values can occur only at isolated points along a scale of values.
- For a six sided dice, the only possible observed values are 1, 2, 3, 4, 5 and 6. It is not possible to observe values such as 5.732.
- Therefore, it is possible that all numerical values for the variable can be listed in a table with accompanying probabilities.
- There are several standard probability distributions that can serve as models for a wide variety of discrete random variables involved in business applications.
- The standard models described in this course are the binomial, hypergeometric, and Poisson probability distributions.
- For a continuous random variable all possible fractional values of the variable cannot be listed, and therefore the probabilities that are determined by a mathematical function are portrayed graphically by a probability density function, or probability curve.

**Random Variables** Remark: We will call random variables R.V.s for short. We associate a number with each outcome of the experiment and the particular numerical value of the random variable depends on the outcome.

For example, consider the experiment of picking an employee at random from an office. We are interested in the number of years work experience the employee has.

The outcome of the experiment could be one year, two years and so on. The random variable is the number of experience the employee has and the value of the random variable can change every time we repeat the experiment.

We can repeat the experiment of picking an employee at random many times and count the number of times that the outcome was one year, two years and so on. Recall that the probability  $p = r/n$  where

- $n$  is the number of time we carry out an experiment (i.e. number of trials)
- $r$  is the number of times we get the result we are interested in (i.e. number of successes)

We can therefore associate probabilities with the various possible values that the random variable might take.

The values of the random variable and the associated probabilities form a probability distribution.

Years (R.V.)	1	2	3	4	5
Probability	0.2	0.3	0.3	0.1	0.1

- The value of the random variable is unknown before we carry out the experiment.
- But by using the probability distribution in the table above, we can say that the probability that an employee selected at random will have one years experience is 0.2 or 20%.

$$P(X = 1) = 0.20).$$

- Although experiments have outcomes that are naturally described using numbers, other do not. For example, an employee might be asked if they enjoy their work. The random variable is enjoyment of work, and the outcomes are yes and no.
- We can arbitrarily assign the value zero to an employee who says no and one to an employee who says yes.
- In this way, the random variable still provides a numerical description of the outcome of the experiment. Many methods of statistical analysis assume that data collected follows a known probability distribution. We will later examine the most commonly used probability distributions.

## 1.7 Introduction to Distributions

### Random Variables

- A random variable is a variable whose value is determined by the outcome of a random phenomenon. The statistical techniques we've learned so far deal with variables, not events, so we need to define a variable in order to analyze a random phenomenon.
- A discrete random variable has a finite number of possible values, while a continuous random variable can take all values in a range of numbers.

## 1.8 Joint Probability Tables

Find  $E[X|Y = 2]$

	X=0	X=1	X=2	
Y=1	0.15	0.2	0.25	$P(Y=1) = 0.6$
Y=2	0.05	0.15	0.20	$P(Y=2) = 0.4$
	$P(X=0) = 0.2$	$P(X=1) = 0.35$	$P(X=2)=0.45$	

**Solution**

$$\frac{(0 \times 0.05) + (1 \times 0.15) + (2 \times 0.2)}{0.4} = \frac{0.55}{0.4}$$

$$E[X|Y = 2] = 1.375$$

### 1.8.1 Probability Mass Function

- a probability mass function (pmf) is a function that gives the probability that a discrete random variable is exactly equal to some value.
- The probability mass function is often the primary means of defining a discrete probability distribution

## 1.9 Discrete Probability Distributions

- Over the next set of lectures, we are now going to look at two important discrete probability distributions
- The first is the *binomial* probability distribution.
- The second is the Poisson probability distribution.
- In R, calculations are performed using the `binom` family of functions and `pois` family of functions respectively.