School of Computing, Engineering & Built Environment

Mathematics Summer School

Level 2 Entry – Engineering

&

Level 3 Entry – Computing

Probability & Probability Distributions

Table of Contents

1. Introduction
2. Probability
3. Basic rules of probability
4. Complementary events
5. Addition Law for Compound Events: A or B
6. Multiplication Law for Compound Events: A and B
7. Bayes Theorem
8. Tree Diagrams
9. Problems on Probability
10. Probability Distributions
11. Discrete Probability Distributions
12. The Expected value and Variance of a Discrete Probability Distribution
13. The Binomial Distribution
14. The Poisson Distribution
15. Continuous Probability Distributions
16. The Normal Distribution
17. Using Statistical Tables to Calculate Normal Probabilities
18. Worked Examples on the Normal Distribution
19. Problems on the Standard Normal Distribution
20. Problems on the Normal Distribution (General)

1. Introduction

In life there is no certainty about what will happen in the future but decisions still have to be taken. Therefore, decision processes must be able to deal with the problems of uncertainty. Uncertainty creates risk and this risk must be analysed. Both qualitative and quantitative techniques for modelling uncertainty exist. In many situations large amounts of numerical data is available which requires statistical techniques for analysis.

Quantitative modelling of uncertainty uses a branch of mathematics called **Probability**. This unit introduces the concepts of probability and probability distributions.

The application of statistical methods is very extensive and is used in all branches of Science and Technology, Industry, Business, Finance, Economics, Sociology, Psychology, Education, Medicine etc.

Examples of Scenarios Involving Uncertainty

In the following examples consider elements of uncertainty involved:

- A production scenario [costs and availability of raw materials, labour, production equipment; demand for product, etc.]
- Aircraft design [reliability of components, systems, safety issues, etc.]
- Financial Investment [interest rates, markets, economic issues, etc.]
- Development of a new drug [reliability, trials, etc.]

Statistics is sometimes described as the science of decision making under uncertainty and can be divided into two broad areas as follows:

Descriptive Statistics which includes the organisation of data, the graphical presentation of data (pie charts, bar charts, histograms, etc.) and the evaluation of appropriate summary statistics (measures of average e.g. the arithmetic mean and measures of spread e.g. standard deviation). Descriptive statistics are the first step in analysing data and always precedes inferential statistics but can be, depending on the type of study, the only way to analyse collected data.

Inferential Statistics which covers those statistical procedures used to help draw conclusions or inferences about a population on the basis of a sample of data collected from the population. Sampling is necessary because measuring every member of a population is time-consuming and expensive, impractical or impossible. Important areas inferential statistics include confidence intervals, hypothesis tests, regression analysis and experimental design. Underlying inferential statistics is the idea of probability and probability distributions.

Some Terminology
It is important, when dealing with data, to have an understanding of the terms used. Some are given below.

Random Variable
Data may come from a survey, a questionnaire or from an experiment. The ‘quantity’ being measured or observed is referred to as a **random variable** as it may vary from individual to individual, e.g. heights of university students. It is commonly denoted by a symbol such as x. There are two types of random variable: **qualitative** and **quantitative.**

Qualitative Variables
These variables cannot be determined numerically and are usually measured categorically. Examples include hair colour, blood group, marital status, gender.

Quantitative Variables
These variables can be determined numerically which allows comparison between values on the basis of magnitude. Examples include height of university students, daily temperature, daily £/Euro exchange rate.

Such variables can be further split into:

- **discrete** where the variable can take only values which differ among themselves by certain fixed amounts, usually integers (e.g. counts such as daily number of breakages of cups in a refectory)

- **continuous** where the variable can take any value within a given range (e.g. temperature at midday on a series of days).

Population
A **population** is the entire group of units, individuals or objects from which data concerning particular characteristics are collected or observed. Generally, it is impractical to observe the entire group, especially if it is large, and a representative part of the population called a sample is usually considered.

Sample
A **sample** is a representative collection of observations taken form the population. A **random sample** is a sample selected in such a way that each member of a population has an equal chance of being included in the sample.

2. Probability

The idea of probability is familiar from games of chance such as card games.

Definition of Probability

If an experiment is performed a large number of times (say, N), and the number of times that a particular event A occurs is counted (say, n), we define:

Probability of event A occurring, denoted by P (A) $= \lim_{N \to \infty} \left(\frac{n}{N} \right)$.

In other words, the probability that the event A occurs in any one performance of the experiment, is defined to be the limiting value of the proportion of times that A would occur if the number of repetitions of the experiment tended to infinity. For example, consider the experiment of tossing a fair coin. The set of possible outcomes for this experiment is {H, T}, the letter H representing the simple event "Heads", and T representing "Tails". According to this definition we would assign equal probability of 0.5 to each of these outcomes, because, in a large number of tosses, heads and tails turn up an approximately equal number of times.

If an experiment has n equally likely outcomes, r of which produce the event A then

$$\text{P}(A) = \frac{\text{number of outcomes which give } A}{\text{total number of outcomes}} = \frac{r}{n}$$

For example, consider the standard pack of 52 playing cards and an experiment defined by drawing a card from a well-shuffled pack. According to this definition, there are 52 possible outcomes to this experiment, and, if the shuffling is done thoroughly, it seems reasonable to regard them as equally likely.

Probability values may also be determined from sample data.

Examples

E1. Suppose that the probability of an error existing in a set of accounts is 0.15. This is equivalent to saying that 15% of all such accounts contain an error.

E2. Suppose that 45% of "small" businesses go bust within a year. This is equivalent to saying that a randomly chosen "small" business has a probability of 0.45 of going bust within a year.

E3. The probability of throwing a head when a fair coin is tossed is 0.5.

E4. A bag of sweets contains 5 Chews, 3 Mints and 2 Toffees. If a sweet is chosen at random from the bag, what is the probability that it is:

(a). a Chew; **(b)**. a Chew or a Toffee; **(c).** Not a Toffee?

Answers: **(a).** 0.5; **(b)**. 0.7; **(c).** 0.8.

E5. From a complete pack of playing cards (ace high) one card is picked at random. What is the probability of the card being:

(a). a red; **(b).** a black picture card (J,Q,K or A);
(c). lower than a 7 or higher than a 9?

Answers: **(a).** 26/52; **(b).** 8/52; **(c).** 40/52

E6. A factory manager inspects a batch of 1000 components and finds 75 of them faulty. On the basis of this data find the probability that a component chosen at random from the batch will be faulty.

Answer: 0.075

E7. Over the past 80 trading days on the London Stock Exchange, the closing DJIA index (Dow Jones Industrial Average) has fallen on 64 days, risen on 12 days and stayed the same on the remaining 4 days. The probability that it will fall on the next trading day is 64/80 = 0.8.

E8. A company receives regular deliveries of raw materials from a supplier. The supplies do not always arrive on time. Over the last 100 delivery days, supplies have been late on 13 occasions. The probability that the supplies will be on time on the next delivery day is 87/100 = 0.87.

Subjective Definition of Probability

There are situations where the above definitions seem to be appropriate and we must use a probability supplied by an ‘expert’. This is often the case in finance and economics.

Examples of Subjective Definition

1. What is the probability that Great Britain will adopt the *Euro* within the next 10 years?

2. What is the probability that a company’s stock will rise in value over the next year?

3. What is the probability that a particular football club will win the Champions League?

More Terminology

An **experiment** is an act or process of observation that leads to a single outcome that cannot be predicted with certainty.

A **sample point** is the most basic outcome of an experiment.

The set of all possible outcomes to a random experiment (all the sample points) is called the **sample space**.

Elements of the sample space are also called **simple events**. Simple events for the example of drawing a playing card from a well-shuffled pack would be "Ace of Clubs" or "Three of Hearts", etc.

Events consisting of more than one simple event are called **compound events**. For example, a compound event for this experiment would be "Drawing a Heart".

Consider the three compound events "Drawing an Ace", "Drawing a Heart", and "Drawing a Face Card" (i.e. one of King, Queen, Jack).

$$P\text{ (Ace)} = \frac{4}{52} = \frac{1}{13}$$

$$P\text{ (Heart)} = \frac{13}{52} = \frac{1}{4}$$

$$P\text{ (Face Card)} = \frac{12}{52} = \frac{3}{13}$$

Note
In these notes we use simple examples to illustrate the ideas discussed. Probabilities of events will be determined in different ways using sample points, relative frequency counting, laws of probability, two-way tables and probability trees. In any given situation, always use the quickest way!

3 Basic rules of probability

If $E_1, E_2, E_3, \ldots, E_n$ are the set of possible simple events which could occur from an experiment, then

(a). $0 \le P(E_i) \le 1$ **(b).** $\sum P(E_i) = 1$

or, in other words:

(a). All probabilities must be between 0 and 1.

An event that will never occur has probability 0.
An event that is certain to occur has probability 1.

(b). The sum of the probabilities of all possible simple events must equal 1.

Probability of an event using Sample Points
We can summarise the steps for calculating the probability of an event as follows:

1. Define the experiment.
2. List the sample points.
3. Assign probabilities to sample points.
4. Determine the collection of sample points contained in the event of interest.
5. Add up these sample probabilities to find the event probability.

4 Complementary events

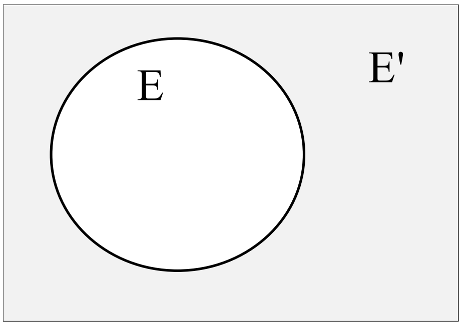
The complement of an event E is defined as the event that must take place if E does not occur. It is written E'. For example, if E = Student is male, E' = Student is female.

As either E or E' must take place we have, from Rule (b) above,

$$P(E) + P(E') = 1 \text{ or } P(E') = 1 - P(E).$$

This is useful because finding the probability of the complement of an event is sometimes easier than finding the probability of the event in question.

Diagrammatically this can be represented as follows: E'



5 Addition Law for Compound Events: A or B

Suppose A and B are two events in a given situation. If A and B cannot possibly occur simultaneously they are said to be *mutually exclusive* events.

For mutually exclusive events the probability that one event **or** the other occurs is given by:

$$P(A \text{ or } B) = P(A) + P(B)$$

Addition law for mutually exclusive events

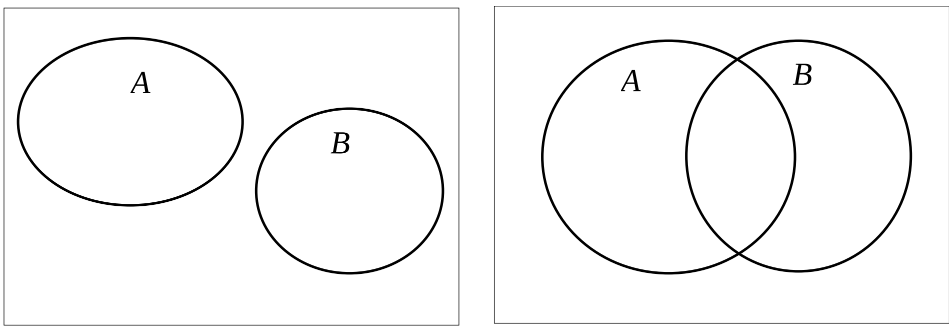
If two events are **not** mutually exclusive, i.e. they can take place at the same time, then the probability of at least one of the events occurring is given by:

$$\mathrm{P(A\,or\,B) = P(A) + P(B) - P(A\,and\,B)}$$

General addition law

NOTE: For mutually exclusive events P(A and B) = 0 so the general addition law "collapses" to the addition law for mutually exclusive events.

To represent this diagrammatically, consider the following:



Examples

1. What is the probability of throwing a 5 or a 6 in one roll of a six-sided die? (Ans: 1/3)

2. A bag contains a large number of balls coloured red, green or blue. If the probability of choosing a red ball from the bag is 0.425 and the probability of choosing a blue ball from the bag is 0.35, what is the probability of choosing either a red or a blue ball from the bag? (Ans: 0.755)

3. In example 2 above, if there are 200 balls in the bag, how many balls will be green? (Ans: 49)

4. A card is drawn from a standard pack. *F* is the event "face card is drawn" and *H* is the event "heart is drawn". What is the probability that a face card or a heart is drawn?
(Ans: 25/52)

6 Multiplication Law for Compound Events: A and B

If two events, A and B, are independent (i.e. the occurrence of the first event has no effect on whether or not the second event occurs) we have:

$$P(A \text{ and } B) = P(A) \times P(B)$$

Multiplication law for independent events

On the other hand if the probability of the second event occurring will be influenced by whether the first event occurs then we must introduce some new notation.

The probability that an event B will occur given that a related event A has already occurred is called the *conditional probability* of B given A and is denoted P(B|A).

We can then calculate the probability that both A and B will occur from:

$$P(A \text{ and } B) = P(A) \times P(B|A)$$

General multiplication law

(Note that if the events are independent, then $P(B|A) = P(B)$ and the above rules are equivalent)

Examples

1. The probability that a new small firm will survive for 2 years has been estimated at 0.22. Given that it survives for 2 years, the probability that it will have a turnover in excess of £250,000 per annum in a further 3 years is estimated at 0.44. Determine the probability that a new business starting now will have a turnover of more than £250,000 in 5 years.
   (Ans: 0.0968)

2. 42% of a population is aged 25 or over. Of such people, 75% have life insurance. Calculate the probability of a person selected at random being over 25 with no life insurance.
   (Ans: 0.105)

7 Bayes' Theorem

The general multiplication law in 1.7 is often rearranged to form what is known as Bayes' Theorem:

$$P(B \mid A) = \frac{P(A \text{ and } B)}{P(A)} = \frac{P(A \mid B).P(B)}{P(A)}$$

Often, P(B) is termed a **prior** probability – it is calculated without taking into account the influencing event A. P(B|A) on the other hand is termed a **posterior** probability- it gives a revised probability that B will occur.

Note: In the above formula P(A|B) and P(B|A) are very different probabilities, e.g. suppose A denotes "I use an umbrella" and B = "it rains".
P(A|B) is the probability that if it rains then I will use an umbrella (quite high) and P(B|A) is the probability that if I use an umbrella then it will rain.

The use of Bayes' Theorem is illustrated in the examples in the next section.

8 Tree Diagrams

A useful way of investigating probability problems is to use what are known as *tree diagrams*. Tree diagrams are a useful way of mapping out all possible outcomes for a given scenario. They are widely used in probability and are often referred to as probability trees. They are also used in decision analysis where they are referred to as decision trees. In the context of decision theory a complex series of choices are available with various different outcomes and we are looking for the bets of these under a given performance criterion such as maximising profit or minimising cost.

The use of a tree diagram is best illustrated with some examples.

Example 1

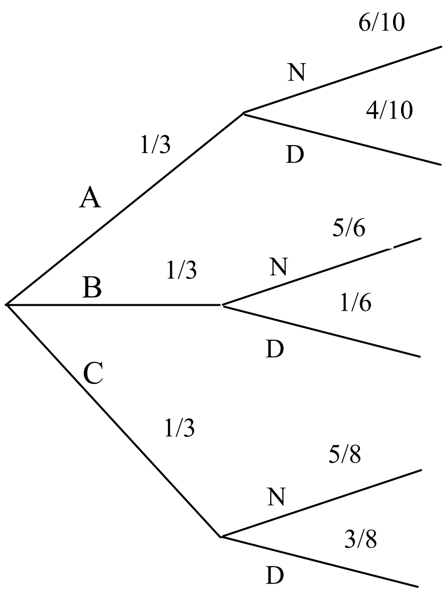
Suppose we are given three boxes, Box A contains 10 light bulbs, of which 4 are defective, Box B contains 6 light bulbs, of which 1 is defective and Box C contains 8 light bulbs, of which 3 are defective. We select a box at random and then draw a light bulb from that box at random. What is the probability that the bulb is defective?

Here we are performing two experiments:

a) Selecting a box at random
b) Selecting a bulb at random from the chosen box

If A, B and C denote the events choosing box A, B, or C respectively and D and N denote the events defective/non-defective bulb chosen, the two experiments can be represented on the diagram below.
We can compute the following probabilities and insert them onto the branches of the tree:

$P(A) = 1/3;$	$P(D \mid A) = 4/10;$	$P(N \mid A) = 6/10;$
$P(B) = 1/3;$	$P(D \mid B) = 1/6;$	$P(N \mid B) = 5/6;$
$P(C) = 1/3;$	$P(D \mid C) = 3/8;$	$P(N \mid C) = 5/8.$



To get the probability for a particular path of the tree (left to right) we multiply the corresponding probabilities on the branches of the path.

For example, the probability of selecting box A and then getting a defective bulb is:

$$P(A \text{ and } D) = P(A) \times P(D|A) = 1/3 * 4/10 = 4/30.$$

Since all the paths are mutually exclusive and there are three paths which lead to a defective bulb, to answer the original question we must add the probabilities for the three paths, i.e.,

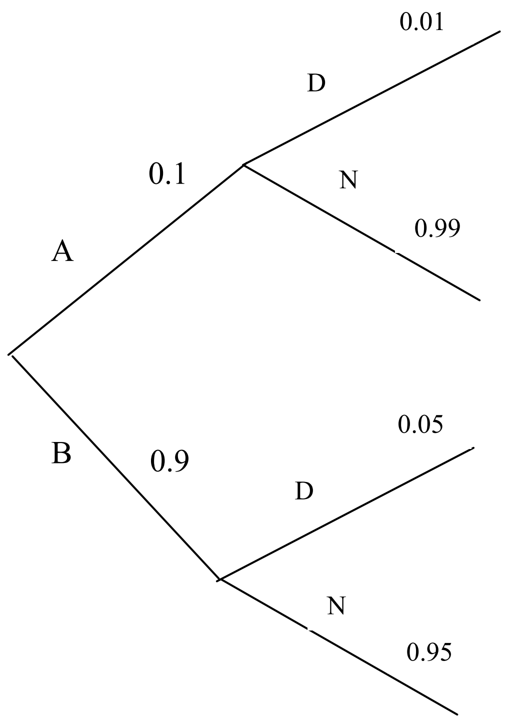
$$4/30 + 1/3*1/6 + 1/3*3/8 = 2/15 + 1/18 + 1/8 = 0.314.$$

Example 2

Machines A and B turn out respectively 10% and 90% of the total production of a certain type of article. The probability that machine A turns out a defective item is 0.01 and the probability that machine B turns out a defective item is 0.05.

(i) What is the probability that an article taken at random from the production line is defective?

(ii) What is the probability that an article taken at random from the production line was made by machine A, *given* that it is defective?



(i) P(D) = P(D and A) + P(D and B)

= P (D|A) P (A) + P (D|B) P (B)

= 0.01*0.1 + 0.05*0.9 = 0.046

(ii) $P(A|D) = \frac{P(A \text{ and } D)}{P(D)}$

$= \frac{0.01 * 0.1}{0.046}$

$= 0.022$

9 Problems on Probability

1. A ball is drawn at random from a box containing 10 red, 30 white, 20 blue and 15 orange balls. Determine the probability that it is

(a) orange or red (b) neither red nor blue (c) not blue

(d) white (e) red, white or blue.

(Ans: 1/3; 3/5; 11/15; 2/5; 4/5)

1. Assembly line items, on inspection, may be classified as passed, reparable or failed . The probability that an item is passed is13/20. The probability that an item is failed is 3/40.

(a) Calculate the probability that an item is reparable.

(b) If 1000 items are inspected, determine the expected number in each class.

(Ans: 11/40; 650, 75, 275)

2. Two firms out of five have a pension scheme, while one firm out of 10 has a sports club. It is known that one firm out of 20 have both. What proportion of firms have a pension scheme or a sports club?

(Ans: 9/20)

3. The probability that a family chosen at random from Britain has a combined income above £40,000 is 0.8. The probability that a family chosen at random from Britain has an income above £40,000 and a car is 0.6. Calculate the probability that a family chosen at random from Britain has a car given that it has a combined income above £40,000.

(Ans: 0.75)

4. Four out of 10 new cars are both foreign and need repair within 2 years. It is known that 7 out of 12 new cars are foreign. Calculate the probability of a new foreign car needing repair within 2 years.

(Ans: 24/35)

5. A foreman for an injection-moulding firm concedes that on 10% of his shifts, he forgets to shut off the injection machine on his line. This causes the machine to overheat, increasing the probability that a defective moulding will be produced during

the early morning run from 2% to 20%. What proportion of the mouldings from the early morning run may be expected to be defective?

(Ans: 5.6%)

6. In a bolt factory, machines A, B and C manufacture 25%, 35% and 40% respectively of the total output of bolts. Of their outputs, 5%, 4% and 2% respectively are defective. A bolt is chosen at random from the factory's output and found to be defective. What is the probability that it came from machine

(a) A (b) B (c) C?

(Ans: 25/69;28/69; 16/69)

8. A factory has a machine which measures the thickness of washers and classifies them as defective or satisfactory. The machine can make two type of error. On average, it classifies 1 out of every 100 defective washers as satisfactory and it classifies 3 out of every 100 satisfactory washers as defective.
5% of the washers produced are, in fact, defective.

Calculate

(a) the expected proportion of washers classified as satisfactory by the machine;
(b) the expected proportion of those washers classified by the machine as satisfactory which will, in fact, be defective.

(Ans: 0.922; 0.00054)

10 Probability Distributions

Now that we have explored the idea of probability, we can consider the concept of a **probability distribution**. In situations where the variable being studied is a random variable, then this can often be modelled by a probability distribution. Simply put, a probability distribution has two components: the collection of possible values that the variable can take, together with the probability that each of these values (or a subset of these values) occurs.

So, in **stochastic modelling**, a probability distribution is the equivalent of a **function** in **deterministic modelling** (in which there is no uncertainty). As we know, there are many different functions available for deterministic modelling and some of these are used more often than others e.g. linear functions, polynomial functions and exponential functions. Exactly the same situation prevails in stochastic modelling. Certain probability distributions are used more often than others. These include the Binomial Distribution, the Poisson Distribution, the Uniform Distribution, the Normal Distribution and the Negative Exponential Distribution.

At the beginning of this chapter we distinguished between discrete and continuous variables and we will consider their probability distributions separately, commencing with the probability distribution of a discrete variable.

11 Discrete Probability Distributions

A discrete random variable assumes each of its values with a certain probability, i.e. each possible value of the random variable has an associated probability. Let X be a discrete random variable and let each value of the random variable have an associated probability, denoted **p(x) = P(X = x)**, such that

X	x_1	x_2	.	.	.	x_n
p(x)	p_1	p_2	.	.	.	p_n

The function p(x) is called the **probability distribution** of the random variable X if the following conditions are satisfied:

(i) $p(x) \geq 0$ for all values x of X, (ii) $\sum_x p(x) = 1$.

p(x) is also referred to as the **probability function** or **probability mass function**.

Again we will use simple examples to illustrate the ideas involved.

Example 1 Throw a dice and record the number thrown. This is a random variable with the following probability distribution:

x	1	2	3	4	5	6
p(x)	1/6	1/6	1/6	1/6	1/6	1/6

Clearly, here it is easy to check that $p(x) \geq 0$ for all values x and $\Sigma\, p(x) = 1$, so the required conditions are satisfied.

Example 2 Consider an experiment that consists of tossing a fair coin twice, and let X equal the number of heads observed. We can list all the possible outcomes as shown in the table below.

Simple events	**Toss 1**	**Toss 2**	**$P(E_i)$**	**X**
E_1	H	H	1/4	2
E_2	H	T	1/4	1
E_3	T	H	1/4	1
E_4	T	T	1/4	0

This gives the following probability distribution.

X	**Simple events corresponding to values of X**	**P(X)**
0	E_4	1/4
1	E_2, E_3	1/2
2	E_1	1/4

Once again we see that the required conditions hold: $p(x) \geq 0$ for all values x and $\Sigma\, p(x) = 1$.

Note

1. In the examples above, probabilities have been computed by counting and relative frequency considerations. As we will see, often the probabilities of a random variable X are determined by means of a formula, e.g. $p(x) = x^2/10$. Individual probabilities for each X value can be found by substituting the value into the given expression and calculating the resulting value of the expression.

2. In some cases, it may be necessary to compute the probability that the random variable X will be greater than or equal to some real number x, i.e. $P(X \geq x)$. This represents the sum of the point probabilities from and including the value x of X onwards. This probability is defined as a **cumulative distribution**. For many probability distributions cumulative probabilities are given in statistical tables.

12 The Expected value and Variance of a Discrete Probability Distribution

Since the probability distribution for a random variable is a model for the relative frequency distribution of a population (think of organising a very large number of observations of discrete random variable into a relative frequency distribution-the probability distribution would closely approximate this), the analogous descriptive measures of the **mean** and **variance** are important concepts.

The **expected value** (or mean), denoted by E(X) or μ, of a discrete random variable X with probability function p(x) is defined as follows:

$$E(X) = \sum xp(x),$$

where the summation extends over all possible values x.

Note the similarity to the mean $\frac{1}{n}\sum fx$ (where $n = \sum f$) of a frequency distribution in which each value, x, is weighted by its frequency f. For a probability distribution, the probability p(x) replaces the observed frequency f.

Note

1. The concept has its roots in games of chance because gamblers want to know what they can "expect" to win over repeated plays of a game. In this sense the expected value means the average amount one stands to win (or lose) per play over a very long series of plays. This meaning also holds with regard to a random variable, i.e. the long-run average of a random variable is its expected value. The expected value can also be interpreted as a "one number" summary of a probability distribution, analogous to the interpretation of the mean of a frequency distribution.

2. The terminology used here of “expected” is a mathematical one, so a random variable may never actually equal its expected value. The expected value should be thought of as a measure of central tendency.

The **variance** of a random variable X, denoted by Var(X) or σ^2, is defined as the expected value of the quantity $(X-\mu)^2$ where μ is the mean of X, i.e.

$$Var(X) = E[(X-\mu)^2] = \sum (x-\mu)^2 p(x)$$

The **standard deviation** of X is the square root, σ, of the variance.

Note

1. As for frequency distributions of observed values, the standard deviation is a measure of the spread or dispersion of the values about the mean.

2. Note that random variables are usually denoted by capital letters e.g. X, Y, Z and <u>actual values</u> of the random variables by lower case letters x, y, z etc.
Also, the characteristics of a distribution (mean, standard deviation etc) of a random variable are denoted by Greek letters (μ, α etc).

Example
Consider an experiment that consists of tossing a fair coin three times, and let X equal the number of heads observed, describe the probability distribution of X and hence determine E(X) and Var(X) .

There are 8 possible outcomes here: HHH, HHT, HTH, THH, HTT, THT, TTH, TTT
This gives probability distribution as follows:

X	**Outcomes**	**Probability**
0	TTT	1/8
1	HTT, THT, TTH	3/8
2	HHT, HTH, THH	3/8
3	HHH	1/8
	Total probability:	**1.0**

The expected value of X is:

E(X) = 0(1/8) + 1(3/8) + 2(3/8) + 3(1/8) = 0 + 3/8 + 6/8 + 3/8 = 12/8 = 1.5.

The variance of X is:

$$
\begin{aligned}
Var(X) &= (0 - 1.5)^2(1/8) + (1 - 1.5)^2(3/8) + (2 - 1.5)^2(3/8) + (3 - 1.5)^2(1/8) \\
&= 2.25(1/8) + 0.25(3/8) + 0.25(3/8) + 2.25(1/8) \\
&= (2.25 + 0.75 + 0.75 + 2.25)/8 \\
&= 6/8 \\
&= 0.75
\end{aligned}
$$

There is an alternative way of computing Var(X). It can be shown that:

$Var(X) = \sum x^2 p(x) - \left[\sum xp(x)\right]^2$. Using this result we have the table below.

x	**P(x)**	**xP(x)**	**x^2P(x)**
0	1/8	0	0
1	3/8	3/8	3/8
2	3/8	6/8	12/8
3	1/8	3/8	9/8
Total	**1.0**	**1.5**	**3.0**

Var(X) $= 3.0 - 1.5^2 = 3.0 - 2.25 = 0.75$ [the same answer as before].

Many different discrete distributions exist and most have proved useful as models of a wide variety of practical problems. The choice of distribution to represent random phenomena should be based on knowledge of the phenomena coupled with verification of the selected distribution through the collection of relevant data. A probability distribution is usually characterised by one or more quantities called the **parameters** of the distribution.

Two important discrete probability distributions are the **Binomial Distribution** and the **Poisson Distribution**.

13 The Binomial Distribution

The **Binomial Distribution** is a very simple discrete probability distribution because it models a situation in which a single trial of some process or experiment can result in only one of two mutually exclusive outcomes the trial (called a **Bernoulli trial** after the mathematician Bernouilli). We have already met examples of this distribution in the earlier discussion on probability.

Example 1
If a coin is tossed there are two mutually exclusive outcomes, heads or tails.

Example 2
If a person is selected at random from the population there are two mutually exclusive outcomes, male or female.

Terminology
It is customary to denote the two outcomes of a Bernoulli trial by the terms "success" and "failure".
Also, if p denotes the probability of success then the probability of failure q satisfies

$$q = 1 - p$$

A Binomial Experiment

A sequence of Bernoulli trials forms a **Binomial Experiment** (or a Bernoulli Process) under the following conditions.

(a) The experiment consists of n identical trials.

(b) Each trial results in one of two outcomes. One of the outcomes is denoted as success and the other as failure.

(c) The probability of success on a single trial is equal to p and remains the same from trial to trial. The probability of failure q is such that $q = 1 - p$.

(d) The trials are independent.

Binomial Distribution

Consider a binomial experiment consisting of n independent trials, where the probability of success on any trial is p. The total number of successes over the n trials is a discrete random variable taking values from 0 to n. This random variable is said to have a **binomial distribution**.

As stated earlier, we have already met examples of the Binomial Distribution. Let us revisit the example in the previous section and examine it from a new perspective.

Example 3

Consider an experiment that consists of tossing a fair coin three times, and let X equal the number of heads observed, describe the probability distribution of X.
Recall that we listed all possible outcomes (we could have drawn a probability tree to do this) and then worked out the appropriate probabilities.

Key observation: for each possible value of X, each individual outcome which results in this value will have the same probability.

For example,

$P(X = 1) = P(HTT) + P(THT) + P(TTH) = pqq + qpq + qqp = 3pqq = 3pq^2 = 3(1/2)(1/2)^2 = 3/8$

In other words we can see that:

P(X = 1) = [number of ways X = 1 can occur]*[probability of any one way in which X = 1]

This observation holds in the general case and we can now present the general result.

General Formula for the Binomial Probability Distribution

If p is the probability of success and q the probability of failure in a Bernoulli trial, then the probability of exactly x successes in a sequence of n independent trials is given by

$$P(X = x) = {}^nC_x\, p^x\, q^{n-x}$$

The symbol ${}^nC_x$ denotes the number of x-combinations of n objects. This symbol arises from investigating how to count the number of ways in which an event can occur. We will look at a couple of examples involving counting before returning to the Binomial distribution.

A formula for ${}^nC_x$

$${}^nC_x = \frac{n(n-1)(n-2)......(n-x+1)}{1.2.3........x} = \frac{n!}{x!(n-x)!}, \text{ where } n! = n(n-1)(n-2)\ldots3.2.1$$

Note

Values for ${}^nC_x$ can be obtained using a calculator.

Example 4
Suppose that 24% of companies in a certain sector of the economy have announced plans to expand in the next year (and the other 76% will not). In a sample of twenty companies chosen at random drawn from this population, find the probability that the number of companies which have announced plans to expand in the next year will be

(i) precisely three, (ii) fewer than three, (iii) three or more, (iv) precisely five.

The values of n. p and q here are:

n = number in the sample = 20

p = probability a company, chosen at random, in a certain sector of the economy has announced plans to expand in the next year = 0.24

q = not p = 0.76

(i) $P(x = 3) = {}^{20}C_3 \, (0.24)^3 \, (0.76)^{17} = 0.1484$

(ii) $P(x \leq 2) = P(0) + P(1) + P(2)$

$$= (0.76)^{20} + {}^{20}C_1 \, (0.24) \, (0.76)^{19} + {}^{20}C_2 \, (0.24)^2 \, (0.76)^{18}$$

$$= 0.0041 + 0.0261 + 0.0783 = 0.1085$$

(iii) $P(x \geq 3) = 1 - P(x \leq 2) = 1 - 0.1085 = 0.8915$

(iv) $P(x = 5) = {}^{20}C_5 \, (0.24)^5 \, (0.76)^{15} = 0.2012$

Note
The above example illustrates the importance of having an efficient way of "counting". With a small number of possibilities to investigate, we can simply draw up a list or a tree, but this is not feasible with larger numbers.

Statistical Tables for the Binomial Distribution

In the above examples we have worked out various probabilities using the formula for the Binomial distribution. We can also make use of Murdoch and Barnes Statistical Tables.

Probabilities are tabled in the form of **Cummulative Binomial Probabilities** i.e. the tables give the probability of obtaining *r or more* successes in *n* independent trials.

Note that probabilities are tabulated for <u>selected values</u> of *p*, *n* and *r* only.

Binomial Distribution Summary

If X~B(n,p), then:

$$p(x)=P(X=x) = {}^{n}C_{x}p^{x}q^{n-x} \text{ , where } q = 1 - p, \text{ and } {}^{n}C_{x} = \frac{n!}{x!\,(n-x)!}$$

The mean: $\mu=E(X) = n.p$

The variance: $\sigma^2 = Var(X) = n.p.q$

14 The Poisson Distribution

Suppose that events are occurring at random points in time or space. The Poisson distribution is used to calculate the probability of having a specified number of occurrences of an event over a fixed interval of time or space. It provides a good model when the data count from an experiment is small i.e. the number of observations is rare during a given time period.

Examples
Some examples of random variables which usually follow the Poisson distribution are:

1. The number of misprints on a page of a book.
2. The number of people in a community living to 100 years of age.
3. The number of wrong telephone numbers dialled in a day.
4. The number of customers entering a shop on a given day.
5. The number of α-particles discharged in a fixed period of time from some radioactive source.

A random variable X satisfies the **Poisson Distribution** if

1. The mean number of occurrences of the event, m, over a fixed interval of time or space is a constant. This is called the **average characteristic**. (This implies that the number of occurrences of the event over an interval is proportional to the size of the interval.)

2. The occurrence of the event over any interval is independent of what happens in any other non-overlapping interval.

If X follows a Poisson Distribution with mean m, the probability of x events occurring is given by the formula:

$$P(X = x) = \frac{e^{-m}m^x}{x!}$$

Note

1. The Poisson distribution is tabulated, for selected values of m, in Murdoch and Barnes Statistical Tables as a cumulative distribution ie $P(X \geq x)$.

2. A Poisson random variable can take an infinite number of values. Since the sum of the probabilities of all the outcomes is 1 and if, for example, you require the probability of 2 or more events, you may obtain this from the identity $P(X \geq 2) = 1 - P(X = 0) - P(X = 1)$ or from Table 2 of Murdoch and Barnes.

3. For the Poisson distribution, mean = variance = m.

Poisson Distribution Summary

If X~P(m), then: $p(x)=P(X=x) = \frac{e^{-m}m^x}{x!}$,

Mean = variance: $\mu = \sigma^2 = m$

15 Continuous Probability Distributions

For a discrete random variable, probabilities are associated with **particular** individual values of the random variable and the sum of all probabilities is one. For a continuous random variable X, we do not have a formula which gives the probability of any particular value of X. The probability that a continuous random variable X assumes a specific value x is taken to be zero.

For a continuous random variable we deal with probabilities of **intervals** rather than probabilities of particular individual values.

The probability distribution of continuous random variable X is characterised by a function f(x) known as the **probability density function**. This function is not the same as the probability function in the discrete case. Since the probability that X is equal a specific value is zero, the probability density function does not represent the probability that X = x. Rather, it provides the means by which the probability of an **interval** can be determined.
The function f (x) whose graph is the limiting curve is the **probability density function** of the continuous random variable X, provided that the vertical scale is chosen in such a way that the total area under the curve is one.

Definition
Let X be a continuous random variable. If a function f (x) satisfies

1. $f(x) \geq 0, \ -\infty < x < \infty$

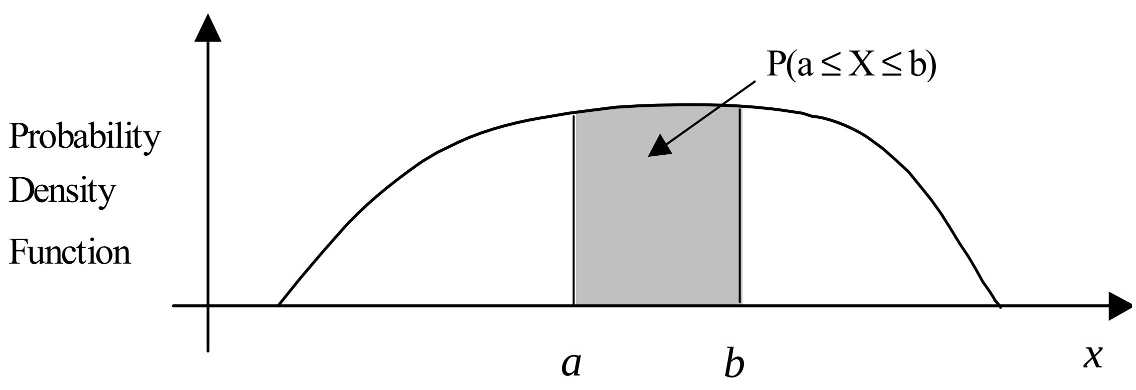
2. $\int_{-\infty}^{\infty} f(x)dx = 1$

and

3. $P(a \leq X \leq b) = \int_a^b f(x)dx$

for any a and b, then f (x) is the **probability density function** of X.

Note
The third condition indicates how the function is used:
The probability that X will assume *some* value in the interval [a, b] = the area under the curve between x = a and x = b. This is the shaded area as shown in the graph below.



In general the exact evaluation of areas requires us to use the probability density function f (x) and integral calculus. These calculations are time consuming and not straightforward. However, probability values can be obtained from statistical tables (just as for discrete probability distributions).

16 The Normal Distribution

The **Normal Distribution** is the most important and most widely used continuous probability distribution. It is the cornerstone of the application of statistical inference in analysis of data because the distributions of several important sample statistics tend towards a Normal distribution as the sample size increases.

Empirical studies have indicated that the Normal distribution provides an adequate approximation to the distributions of many physical variables. Specific examples include meteorological data, such as temperature and rainfall, measurements on living organisms, scores on aptitude tests, physical measurements of manufactured parts, weights of contents of food packages, volumes of liquids in bottles/cans, instrumentation errors and other deviations from established norms, and so on.
The graphical appearance of the Normal distribution is a symmetrical bell-shaped curve that extends without bound in both positive and negative directions:

The probability density function is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \qquad -\infty < x < \infty; \quad -\infty < \mu < \infty, \sigma > 0$$

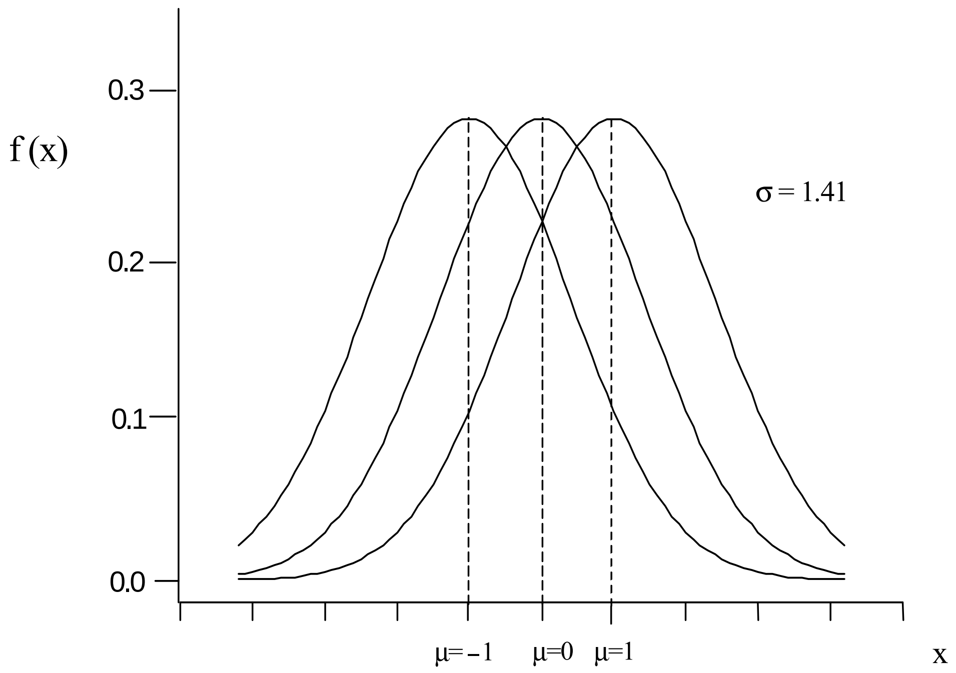
where μ and σ are **parameters**. These turn out to be the mean and standard deviation, respectively, of the distribution. As a shorthand notation, we write $X \sim N(\mu,\sigma^2)$.

The curve never actually reaches the horizontal axis buts gets close to it beyond about 3 standard deviations each side of the mean.

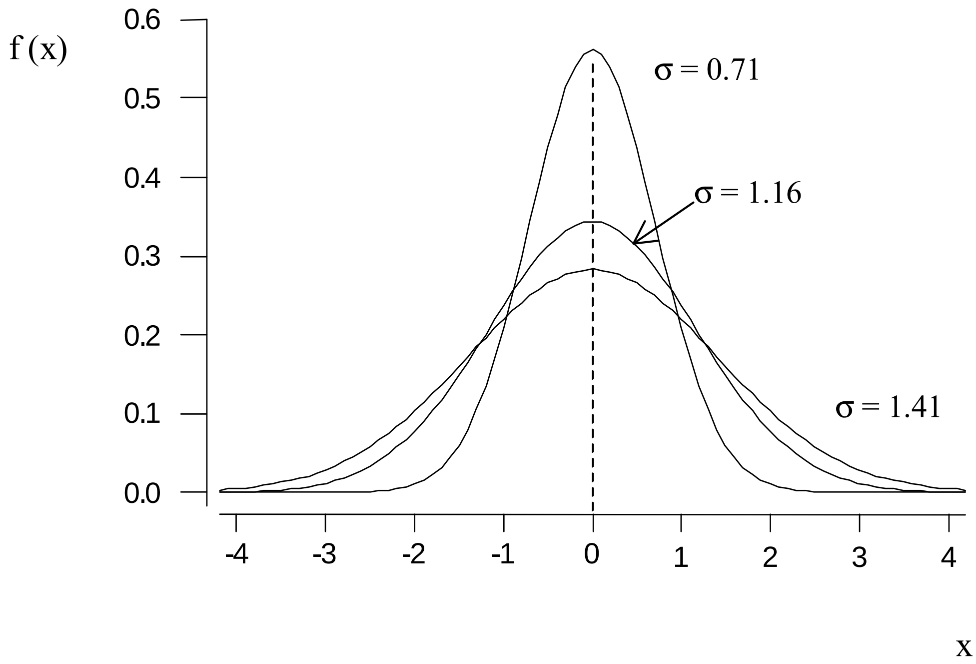
For any Normally distributed variable:

> 68.3% of all values will lie between $\mu - \sigma$ and $\mu + \sigma$ (i.e. $\mu \pm \sigma$)
> 95.45% of all values will lie within $\mu \pm 2\sigma$
> 99.73% of all values will lie within $\mu \pm 3\sigma$

The graphs below illustrate the effect of changing the values of μ and σ on the shape of the probability density function. Low variability ($\sigma = 0.71$) with respect to the mean gives a pointed bell-shaped curve with little spread. Variability of $\sigma = 1.41$ produces a flatter bell-shaped curve with a greater spread.



0.3
0.2
0.1
0.0
f (x)
σ = 1.41
μ= −1
μ=0
μ=1
x



0.6
0.5
0.4
0.3
0.2
0.1
0.0
f (x)
σ = 0.71
σ = 1.16
σ = 1.41
-4
-3
-2
-1
0
1
2
3
4
x

17 Using Statistical Tables to Calculate Normal Probabilities

Only one set of tabled values for the Normal distribution are available- this is for the S**tandard Normal variable**, which has mean = 0 and standard deviation = 1.

The calculation of Normal probabilities associated with a specified range of values of random variable X involves two steps:

Step1
Apply the transformation $Z = (X - \mu)/\sigma$ to change the basis of the probability calculation from X to Z, the standard Normal variable i.e. this expresses the probability calculation we need to carry out in terms of an equivalent probability calculation involving the Standard Normal variable (for which we have tabled values).

Step2
Use tables of probabilities for Z, together with the symmetry properties of the Normal Distribution to determine the required probability.

Tables of the Standard Normal distribution are widely available. The version we will use is from Murdoch and Barnes, Statistical Tables for Science, Engineering, Management and Business Studies, in which the tabulated value is the probability $P(Z > u)$ where u lies between 0 and 4, i.e. the tables provide the probability of Z exceeding a specific value u, sometimes referred to as the **right-hand tail** probability.

18 Worked Examples on the Normal Distribution

In the first set of examples, we learn how to use tables of the Standard Normal distribution.

Example 1 Find P $(z \geq 1)$.

Solution
From Tables, we find that P $(z \geq 1) = 0.1587$

Example 2 Find P $(z \leq -1)$.
Solution
By symmetry, P $(z \leq -1)$ = P $(z \geq 1) = 0.1587$.

Example 3 Find P $(-1 \leq z \leq 1)$.
Solution
$P(-1 \leq z \leq 1) = 1 - [P(z \leq -1) + P(z \geq 1)] = 1 - [P(z \geq 1)] + P(z \geq 1)]$
$=1 - 0.1587 - 0.1587 = 0.6826.$

Example 4 Find P $(-0.62 \leq z \leq 1.3)$.
Solution
$P(-0.62 \leq z \leq 1.3) = 1 - [P(z \leq -0.62) + P(z \geq 1.3)] = 1 - [P(z \geq 0.62) + P(z \geq 1.3)]$
$= 1 - [0.2676 + 0.0968] = 0.6356.$

In the next examples, we learn how to determine probabilities for Normal variables in general.

Example 5
The volume of water in commercially supplied fresh drinking water containers is approximately Normally distributed with mean 70 litres and standard deviation 0.75 litres. Estimate the proportion of containers likely to contain

(i) in excess of 70.9 litres, (ii) at most 68.2 litres, (iii) less than 70.5 litres.

Solution
Let X denote the volume of water in a container, in litres. Then $X \sim N(70, 0.75^2)$, i.e. $\mu = 70$, $\sigma = 0.75$ and $Z = (X - 70)/0.75$

(i) $X = 70.9$; $Z = (70.9 - 70)/0.75 = 1.20$. $P(X > 70.9) = P(Z > 1.20) = 0.1151$ or 11.51%

(ii) $X = 68.2$; $Z = -2.40$. $P(X < 68.2) = P(Z < -2.40) = 0.0082$ or 0.82%

(iii) $X = 70.5$; $Z = 0.67$

$P(X > 70.5) = 0.2514$; $P(X < 70.5) = 0.7486$ or 74.86%

19 Problems on the Standard Normal Distribution

In each of the following questions, Z has a Standard Normal Distribution.

Obtain:

1.	$\Pr(Z > 2.0)$	2.	$\Pr(Z > 2.94)$	3.	$\Pr(Z < 1.75)$
4.	$\Pr(Z < 0.33)$	5.	$\Pr(Z < -0.89)$	6.	$\Pr(Z < -1.96)$
7.	$\Pr(Z > -2.61)$	8.	$\Pr(Z > -0.1)$	9.	$\Pr(1.37 < Z < 1.95)$
10.	$\Pr(1 < Z < 2)$	11.	$\Pr(-1.55 < Z < 0.55)$		
12.	$\Pr(-3.1 < Z < -0.99)$	13.	$\Pr(-2.58 < Z < 2.58)$		
14.	$\Pr(-2.1 < Z < 1.31)$				

Find the value of u such that

15.	$\Pr(Z > u) = 0.0307$	16.	$\Pr(Z < u) = 0.01539$
17.	$\Pr(Z > u) = 0.2005$	18.	$\Pr(Z > u) = 0.8907$
19.	$\Pr(Z < u) = 0.9901$	20.	$\Pr(-u < Z < u) = 0.9010$

Answers

1.	0.02275	2.	0.00164	3.	0.9599	4.	0.6293
5.	0.1867	6.	0.0250	7.	0.99547	8.	0.5398
9.	0.0597	10.	0.13595	11.	0.2306	12.	0.16013
13.	0.99012	14.	0.88704	15.	1.87	16.	−2.16
17.	0.84	18.	−1.23	19.	2.33	20.	1.65

20 Problems on the Normal Distribution (General)

1. The weights of bags of potatoes delivered to a supermarket are approximately Normally distributed with mean 5 Kg and standard deviation 0.2 Kg. The potatoes are delivered in batches of 500 bags.

   (i) Calculate the probability that a randomly selected bag will weigh more than 5.5 Kg.
   (ii) Calculate the probability that a randomly selected bag will weigh between 4.6 and 5.5 Kg.
   (iii) What is the expected number of bags in a batch weighing less than 5.5 Kg?

2. A machine in a factory produces components whose lengths are approximately Normally distributed with mean 102mm and standard deviation 1mm.

   (i) Find the probability that if a component is selected at random and measured, its length will be:

   (a) less than 100mm; (b) greater that 104mm.

   (ii) If an output component is only accepted when its length lies in the range 100mm to 104mm, find the expected proportion of components that are accepted.

3. The daily revenue of a small restaurant is approximately Normally distributed with mean £530 and standard deviation £120. To be in profit on any day the restaurant must take in at least £350.

   (i) Calculate the probability that the restaurant will be in profit on any given day.

Answers

1. (i) 0.00621 (ii) 0.9710 (iii) 497 approximately

2. (i) (a) 0.02275; (b) 0.02275 (ii) 0.9545

3. (i) 0.9332