bookboon.com

Mathematics for Computer Scientists

Gareth J. Janacek; Mark Lemmon Close

Download free books at
bookboon.com

Gareth J. Janacek & Mark Lemmon Close

Mathematics for Computer Scientists

Download free eBooks at bookboon.com

Mathematics for Computer Scientists
© 2011 Gareth J. Janacek, Mark Lemmon Close & Ventus Publishing ApS
ISBN 978-87-7681-426-7

Download free eBooks at bookboon.com

Contents

	Introduction	5
1	Numbers	6
2	The statement calculus and logic	20
3	Mathematical Induction	35
4	Sets	39
5	Counting	49
6	Functions	56
7	Sequences	73
8	Calculus	83
9	Algebra: Matrices, Vectors etc.	98
10	Probability	119
11	Looking at Data	146

Discover the truth at www.deloitte.ca/careers

Deloitte.

© Deloitte & Touche LLP and affiliated entities.

Download free eBooks at bookboon.com

Click on the ad to read more

Introduction

The aim of this book is to present some the basic mathematics that is needed by computer scientists. The reader is not expected to be a mathematician and we hope will find what follows useful.

Just a word of warning. Unless you are one of the irritating minority mathematics is hard. You cannot just read a mathematics book like a novel. The combination of the compression made by the symbols used and the precision of the argument makes this impossible. It takes time and effort to decipher the mathematics and understand the meaning.

It is a little like programming, it takes time to understand a lot of code and you never understand how to write code by just reading a manual - you have to do it! Mathematics is exactly the same, you need to do it.

Download free eBooks at bookboon.com

Chapter 1

Numbers

Defendit numerus: There is safety in numbers

We begin by talking about numbers. This may seen rather elementary but is does set the scene and introduce a lot of notation. In addition much of what follows is important in computing.

1.0.1 Integers

We begin by assuming you are familiar with the *integers*

$$1,2,3,4,\ldots,101,102,\ \ldots,n,\ldots,2^{32582657}-1,\ldots,$$

sometime called the whole numbers. These are just the numbers we use for counting. To these integers we add the *zero*, 0, defined as

$$0+\ \text{any integer } n = 0+n = n+0 = n$$

Once we have the integers and zero mathematicians create negative integers by defining $(-n)$ as:
the number which when added to n gives zero, so $n+(-n) = (-n)+n = 0$.

Eventually we get fed up with writing $n+(-n) = 0$ and write this as $n-n = 0$. We have now got the positive and negative integers $\{\ldots,-3,-2,-1,0,1,2,3,4,\ldots\}$

You are probably used to arithmetic with integers which follows simple rules. To be on the safe side we itemize them, so for integers a and b

1. $a+b = b+a$
2. $a\times b = b\times a$ or $ab = ba$
3. $-a\times b = -ab$

Download free eBooks at bookboon.com

4. $(-a) \times (-b) = ab$

5. To save space we write a^k as a shorthand for a multiplied by itself k times. So $3^4 = 3 \times 3 \times 3 \times 3$ and $2^{10} = 1024$. Note $a^n \times a^m = a^{n+m}$

6. Do note that n^0=1.

Factors and Primes

Many integers are products of smaller integers, for example $2 \times 3 \times 7 = 42$. Here 2, 3 and 7 are called the *factors* of 42 and the splitting of 42 into the individual components is known as *factorization*. This can be a difficult exercise for large integers, indeed it is so difficult that it is the basis of some methods in cryptography.

Of course not all integers have factors and those that do not, such as

$$3, 5, 7, 11, 13, \ldots, 2^{216091} - 1, \ldots$$

are known as *primes*. Primes have long fascinated mathematicians and others see

`http://primes.utm.edu/`,

and there is a considerable industry looking for primes and fast ways of factorizing integers.

To get much further we need to consider division, which for integers can be tricky since we may have a result which is not an integer. Division may give rise to a *remainder*, for example

$$9 = 2 \times 4 + 1.$$

and so if we try to divide 9 by 4 we have a remainder of 1 .

In general for any integers a and b

$$b = k \times a + r$$

where r is the *remainder*. If r is zero then we say a *divides* b written $a \mid b$. A single vertical bar is used to denote *divisibility*. For example $2 \mid 128$, $7 \mid 49$ but 3 does not divide 4, symbolically $3 \nmid 4$.

Aside

To find the factors of an integer we can just attempt division by primes i.e. $2, 3, 5, 7, 11, 19, \ldots$. If it is divisible by k then k is a factor and we try again. When we cannot divide by k we take the next prime and continue until we are left with a prime. So for example:

1. 2394/2=1197 can't divide by 2 again so try 3

Download free eBooks at bookboon.com

2. 1197/3=399
3. 399/3 = 133 can't divide by 3 again so try 7 (not divisible by 5)
4. 133/7 = 19 which is prime so 2394 $=2 \times 3 \times 3 \times 7 \times 19$

Modular arithmetic

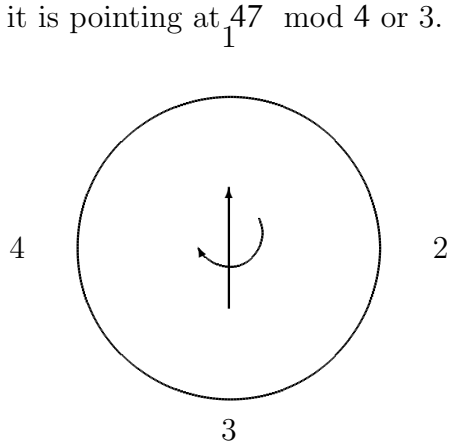
The `mod` operator you meet in computer languages simply gives the remainder after division. For example,

1. $25 \mod 4 = 1$ because $25 \div 4 = 6$ remainder 1.
2. $19 \mod 5 = 4$ since $19 = 3 \times 5 + 4$.
3. $24 \mod 5 = 4$.
4. $99 \mod 11 = 0$.

There are some complications when negative numbers are used, but we will ignore them. We also point out that you will often see these results written in a slightly different way i.e. $24 = 4 \mod 5$ or $21 = 0 \mod 7$. which just means $24 \mod 5 = 4$ and $21 \mod 7 = 0$

Modular arithmetic is sometimes called clock arithmetic. Suppose we take a 24 hour clock so 9 in the morning is 09.00 and 9 in the evening is 21.00. If I start a journey at 07.00 and it takes 25 hours then I will arrive at 08.00. We can think of this as 7+25 = 32 and 32 mod 24 = 8. All we are doing is starting at 7 and going around the (25 hour) clock face until we get to 8. I have always thought this is a complex example so take a simpler version.

Four people sit around a table and we label their positions 1 to 4. We have a pointer point to position 1 which we spin. Suppose it spins 11 and three quarters or 47 quarters. The it is pointing at $47 \mod 4$ or 3.



Download free eBooks at bookboon.com

The Euclidean algorithm

Algorithms which are schemes for computing and we cannot resist putting one in at this point. The Euclidean algorithm for finding the gcd is one of the oldest algorithms known, it appeared in Euclid's Elements around 300 BC. It gives a way of finding the greatest common divisor (gcd) of two numbers. That is the largest number which will divide them both.

Our aim is to find a a way of finding the greatest common divisor, $\gcd(a, b)$ of two integers a and b.
Suppose a is an integer smaller than b.

1. Then to find the greatest common factor between a and b, divide b by a. If the remainder is zero, then b is a multiple of a and we are done.

2. If not, divide the divisor a by the remainder.

Continue this process, dividing the last divisor by the last remainder, until the remainder is zero. The last non-zero remainder is then the greatest common factor of the integers a and b.

Download free eBooks at bookboon.com
Click on the ad to read more

The algorithm is illustrated by the following example. Consider 72 and 246. We have the following 4 steps:

1. $246 = 3 \times 72 + 30$ or $246 \mod 72 = 30$
2. $72 = 2 \times 30 + 12$ or $72 \mod 30 = 12$
3. $30 = 2 \times 12 + 6$ or $30 \mod 12 = 6$
4. $12 = 2 \times 6 + 0$

so the gcd is 6.

There are several websites that offer Java applications using this algorithm, we give a Python function

```
def gcd(a,b):
        """ the euclidean algorithm """
        if b == 0:
                return a
        else:
                return gcd(b, (a%b))
```

Those of you who would like to see a direct application of some these ideas to computing should look at the section on random numbers

1.0.2 Rationals and Reals

Of course life would be hard if we only had integers and it is a short step to the rationals or fractions. By a rational number we mean a number that can be written as P/Q where P and Q are integers. Examples are

$$\frac{1}{2} \qquad \frac{3}{4} \qquad \frac{7}{11} \qquad \frac{7}{6}$$

These numbers arise in an obvious way, you can imagine a ruler divided into 'iths' and then we can measure a length in 'iths'. Mathematicians, of course, have more complicated definitions based on modular arithmetic . They would argue that for every integer n, excluding zero, there is an inverse, written $1/n$ which has the property that

$$n \times \frac{1}{n} = \frac{1}{n} \times n = 1$$

Of course multiplying $1/n$ by m gives a fraction m/n. These are often called *rational numbers*.

We can manage with the simple idea of fractions.

Download free eBooks at bookboon.com

One problem we encounter is that there are numbers which are neither integers or rationals but something else. The Greeks were surprised and confused when it was demonstrated that $\sqrt{2}$ could not be written exactly as a fraction. Technically there are no integer values P and Q such that $P/Q = \sqrt{2}$.

From our point of view we will not need to delve much further into the details, especially as we can get good enough approximation using fractions. For example $22/7$ is a reasonable approximation for π while $355/113$ is better. You will find people refer to the *real numbers*, sometimes written $\mathbb{R}$, by which they mean all the numbers we have discussed to date.

Notation

As you will have realized by now there is a good deal of notation and we list some of the symbols and functions you may meet.

- If x is *less than* y then we write $x < y$. If there is a possibility that they might be equal then $x \leq y$. Of course we can write these the other way around. So $y > x$ or $y \geq x$. Obviously we can also say y is greater than x or greater than or equal to x

- The *floor function* of a real number x, denoted by $\lfloor x \rfloor$ or $\texttt{floor}(x)$, is a function that returns the largest integer less than or equal to x. So $\lfloor 2.7 \rfloor = 2$ and $\lfloor -3.6 \rfloor = -4$. The function `floor` in Java and Python performs this operation. There is an obvious(?) connection to $\bmod$ since $b \bmod a$ can be written $b - \texttt{floor}(b \div a) \times a$. So $25 \bmod 4 = 25 - \lfloor 25/4 \rfloor \times 4 = 25 - 6 \times 4 = 1$

Download free eBooks at bookboon.com

- A less used function is the *ceiling function*, written $\lceil x \rceil$ or `ceil`(x) or `ceiling`(x), is the function that returns the smallest integer *not less than* x. Hence $\lceil 2.7 \rceil = 3$.
- The *modulus* of x written $| x |$ is just x when $x \geq 0$ and $-x$ when $x < 0$. So $| 2 |= 2$ and $| -6 |= 6$. The famous result about the modulus is that for any x and y
$$| x + y | \leq | x | + | y |$$
- We met a^b when we discussed integers and in the same way we can have x^y when x and y are not integers. We discuss this in detail when we meet the exponential function. Note however
  - a^0=1 for all $a \neq 1$
  - $0^b = 0$ for all values of b including zero.

1.0.3 Number Systems

We are so used to working in a decimal system we forget that it is a recent invention and was a revolutionary idea. It is time we looked carefully at how we represent numbers. We normally use the decimal system so 3459 is shorthand for 3 x 1000 + 4 x 100 + 5 x 10 + 9. The *position* of the digit is vital as it enables us to distinguish between 30 and 3. The decimal system is a positional numeral system; it has positions for units, tens, hundreds and so on. The position of each digit implies the multiplier (a power of ten) to be used with that digit and each position has a value ten times that of the position to its right.

Notice we may save space by writing 1000 as 10^3 the 3 denoting the number of zeros. So 100000 = 10^5. If the superscript is *negative* then we mean a fraction e.g $10^3 = 1/1000$. Perhaps the cleverest part of the positional system was the addition of the decimal point allowing us to include decimal fractions. Thus 123.456 is equivalent to

$$1 \times 100 + 2 \times 10 + 3 + \text{ numbers after the point } + 4 \times 1/10 + 5 \times 1/100 + 6 \times 1/1000$$

Multiplier	...	10^2	10^1	10^0	.	10^{-1}	10^{-2}	10^{-3}	...
digits	...	1	2	3	.	4	5	6	...
					↑				
					decimal point				

However there is no real reason why we should use powers of 10, or base 10. The Babylonians use base 60 and base 12 was very common during the middle ages in Europe. Today the common number systems are

Download free eBooks at bookboon.com

- Decimal number system: symbols 0-9; base 10
- Binary number system:symbols symbols 0,1; base 2
- Hexadecimal number system:symbols 0-9,A-F; base 16
  here A $\equiv$ 10 , B $\equiv$ 11 , C $\equiv$ 12 , D $\equiv$13 E $\equiv$ 14 , F$\equiv$ 15.
- Octal number system: symbols 0-7; base 8

Binary

In the binary scale we express numbers in powers of 2 rather than the 10s of the decimal scale. For some numbers this is easy so, if recall $2^0 = 1$,

Decimal number		in powers of 2	power of 2 3	2	1	0	Binary number
8	=	2^3	1	0	0	0	1000
7	=	$2^2 + 2^1 + 2^0$	0	1	1	1	111
6	=	$2^2 + 2^1$	0	1	1	0	110
5	=	$2^2 + 2^0$	0	1	0	1	101
4	=	2^2	0	1	0	0	100
3	=	$2^1 + 2^0$	0	0	1	1	11
2	=	2^1	0	0	1	0	10
1	=	2^0	0	0	0	1	1

As in decimal we write this with the position of the digit representing the power, the first place after the decimal being the 2^0 position the next the 2^1 and so on. To convert a decimal number to binary we can use our `mod` operator.

As an example consider 88 in decimal or 88_{10}. We would like to write it as a binary. We take the number and successively divide `mod` 2. See below

Step number n	x_n	$\lfloor x_n/2 \rfloor$	x_n mod 2
0	88	44	0
1	44	22	0
2	22	11	0
3	11	5	1
4	5	2	1
5	2	1	0
6	1	0	1

Writing the last column *in reverse*, that is from the bottom up, we have 1011000 which is the binary for of 88, i.e.$88_{10} = 1011000_2$.

Download free eBooks at bookboon.com

Binary decimals are less common but quite possible, thus 101.1011 is just $2^2+2^0+2^{-1}+2^{-3}+2^{-4}$ which is, after some calculation 5.6875. We have see how to turn the integer part of a decimal number into a binary number and we can do the same with a *decimal* fraction. Consider 0.6875. As before we draw up a table

Step number n	x_n	$x_n \times 2$	$\lfloor x_n \times 2 \rfloor$
0	0.6875	1.375	1
1	0.375	0.75	0
2	0.75	1.5	1
3	0.5	1	1

giving *reading down* $0.6875_{10} = 1011_2$

Beware it is possible to get into a non-ending cycle when we have a non terminating decimal. For example 0.4.

Step number n	x_n	$x_n 2$	$\lfloor x_n \times 2 \rfloor$	
0	0.4	0.8	0	
1	0.8	1.6	1	
2	0.6	1.2	1	
3	0.2	0.4	0	
4	0.4	0.8	0	$\leftarrow$ here we repeat
3	0.8	1.6	1	
4	...		...	

so $0.4_{10} = 0.0110011001100\ldots_2$

Download free eBooks at bookboon.com

Click on the ad to read more

- Addition in binary
  - 0+0 = 0
  - 0+1 = 1
  - 1+1 = 10 so we carry 1 and leave a zero
  - 1+1+1 = 1+(1+0)=1+10=11 .

  We can write this in very much the same way as for a decimal addition

$$\begin{array}{cccccccl} & 1 & 1 & 0 & 1 & 0 & 1 & \\ + & 1 & 0 & 1 & 1 & 1 & 0 & \\ \hline 1 & 1 & 0 & 0 & 0 & 1 & 1 & \text{Sum} \\ & \uparrow & & & \uparrow & & & \end{array}$$

  the right hand uparrow show where we carry a 1. The left hand one shows where we have $1 + 1 + 1$ so we carry a 1 and have a 1 left over

- To subtract

$$\begin{array}{cccccccl} & 1 & 1 & 0 & 1 & 0 & 1 & \\ - & 1 & 0 & 1 & 1 & 1 & 0 & \\ \hline & 0 & 0 & 0 & 1 & 1 & 1 & \text{difference} \end{array}$$

Multiplication in decimal

$$\begin{array}{ccccccccr} & & 1 & 2 & 5 & 6 & 7 & 8 & \text{Multiplicand} \\ \times & & & & & 3 & 8 & 7 & \text{Multiplier} \\ \hline & & 8 & 7 & 9 & 7 & 4 & 6 & \text{times 7} \\ 1 & 0 & 0 & 5 & 4 & 2 & 4 & & \text{Shift left one and times 8} \\ 3 & 7 & 7 & 0 & 3 & 4 & & & \text{Shift left two and times 3} \\ \hline 4 & 8 & 6 & 3 & 7 & 3 & 8 & 6 & \text{Add to get product} \end{array}$$

Multiplication in binary

$$\begin{array}{ccccccccr} & & 1 & 0 & 0 & 1 & 1 & 1 & 0 & \text{Multiplicand} \\ \times & & & & & & 1 & 0 & 1 & \text{Multiplier} \\ \hline & & 1 & 0 & 0 & 1 & 1 & 1 & 0 & \text{times 1} \\ & 0 & 0 & 0 & 0 & 0 & 0 & 0 & & \text{Shift left one and times 0} \\ 1 & 0 & 0 & 1 & 1 & 1 & 0 & & & \text{Shift left two and times 1} \\ \hline 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & \text{Add to get the product} \end{array}$$

As you can see multiplication in binary is easy.

Download free eBooks at bookboon.com

Octal

Base 8 or octal does not bring any new problems. We use the symbols 0, 1, 2,. . . ,7 and the position denotes the power of 8. So 12_8 is $1 \times 8 + 2 = 10$ in decimal, while 3021_8 is

$$3 \times 8^3 + 0 \times 8^2 + 2 \times 8 + 1 \times 8^0 = 1536 + 16 + 1 = 1553$$

in decimal. Obviously we do not need the symbol for 9 as $9_{10} = 8 + 1 = 11_8$ in octal. To convert a decimal number to octal we can use our `mod` operator as we did in the binary case.

As an example consider 1553 in decimal or 1553_{10}. We would like to write it as an octal number. We take the number and successively divide `mod` 8. See below

Step number n	x_n	$\lfloor x_n/8 \rfloor$	$x_n \mod 8$
0	1553	194	1
1	194	24	2
2	24	3	0
3	3	0	3

Writing the last column *in reverse* we have 3021 which is the octal number we require since

$$3 \times 8^3 + 0 \times 8^2 + 2 \times 8 + 1 \times 8^0 = 1553$$

There is a simple link between octal and binary if we notice that

$7 = 2^2 + 2^1 + 2^0 = 111_2$	$3 = 2^1 + 2^0 = 11_2$
$6 = 2^2 + 2^1 = 110_2$	$2 = 2^1 = 10_2$
$5 = 2^2 + 2^0 = 101_2$	$1 = +2^1 = 1_2$
$4 = 2^2 = 100_2$	$0 = 0_2$

You might like to check that 1553 is

11000010001

in binary.

Separating this into blocks of 3 gives

11 000 010 001

If we use our table to write the digit corresponding to each binary block of 3 we have

3 0 2 1

Download free eBooks at bookboon.com

which is our octal representation!

As in the binary case we can also have octal fractions, for example 0.3012_8. This is a way of representing

$$3 \times 1/8^1 + 0 \times 1/8^2 + 1 \times 1/8^3 + 2 \times 1/8^4$$

To convert 0.3012_8 to decimal we proceed as for the binary case only here we use 8 rather that 2 to give

Step number n	x_n	$8 \times x_n$	$\lfloor 8x_n \rfloor$
0	0.3012	2.4096	2
1	0.4096	3.2768	3
2	0.2768	2.2144	2
3	0.2144	1.7152	1
4	0.7152	5.72165	5
5	0.7216	5.7728	5
6	0.7728	6.1824	6
7	0.1824	1.4592	1
8	0.4592	3.6736	3
9	0.6736	5.3888	5
10	0.3888	3.1104	3
11.	0.1104002	0.8832016	0
12	0.8832016	7.0656128	7
13	0.06561279	0.52490234	0
14	0.5249023	4.1992188	4
15	0.1992188	1.5937500	1
16	0.59375	4.75000	4
17	0.75	6.00	6
18	0	0	

giving *reading down* $0.3012_8 = 0.232155613530704146_{10}$

hexadecimal

Base 16 is more complicated because we need more symbols. We have the integers 0 to 9 and we also use A $\equiv$ 10 , B $\equiv$ 11 , C $\equiv$ 12 , D $\equiv$13 E $\equiv$ 14 , F$\equiv$ 15.

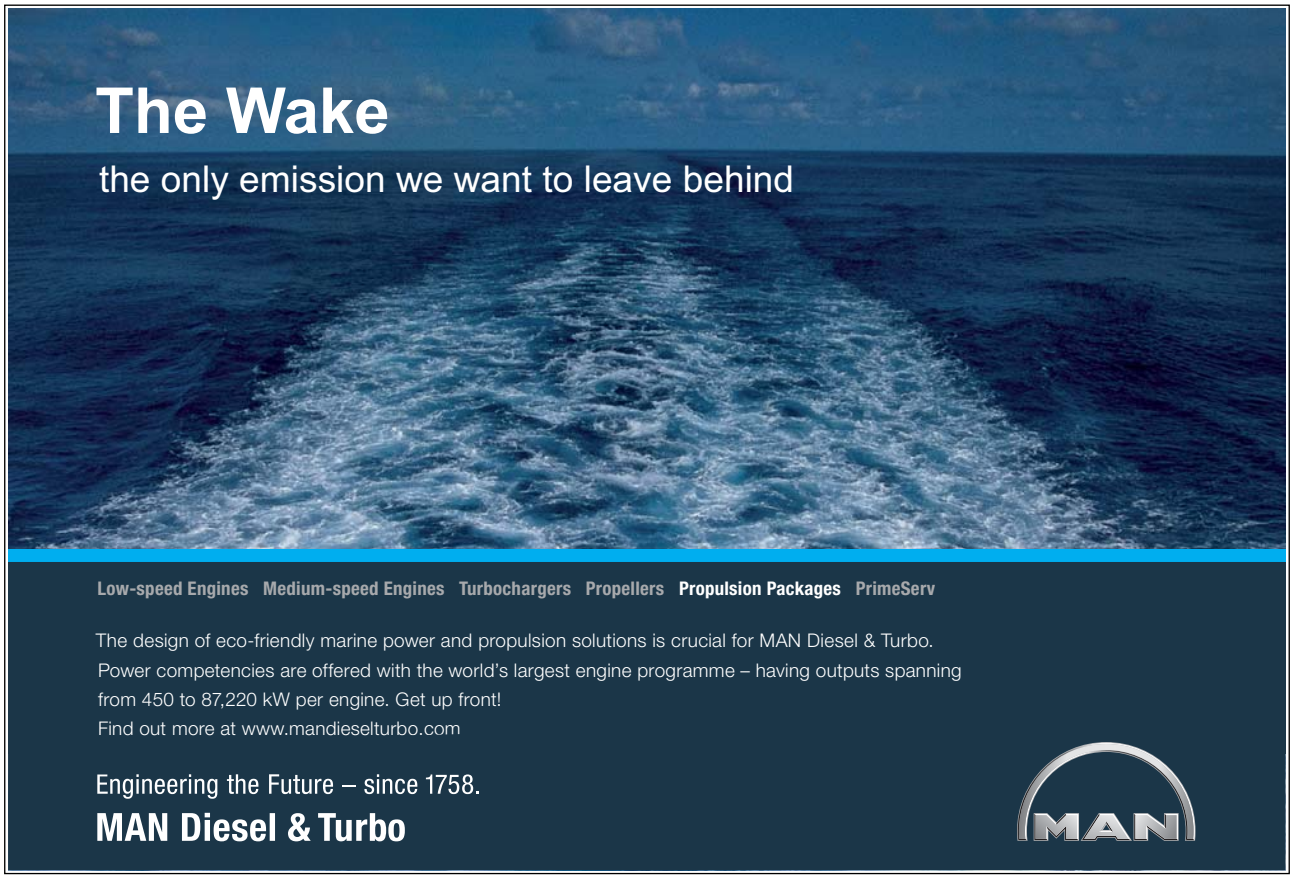
So 123_{16} is $1 \times 16^2 + 2 \times 16^1 + 3$ in decimal and $A2E_{16}$ is $10 \times 16^2 + 2 \times 16^1 + 14$ in decimal. The good thing about hex is that each of the symbols corresponds to a 4 digit binary sequence (if we allow leading zeros). This means we can easily translate from hex to binary as below $01011110101101010010 2_2 =$ 0101 1110 1011 0101 0010

$$= 5 \quad E \quad B \quad 5 \quad 2_{16} = 5EB52_{16}$$

Download free eBooks at bookboon.com

exercises

1. Factorize
   (a) 3096
   (b) 1234
   (c) $2^4 - 1$
2. It was thought that $2^p - 1$ was prime when p is a prime. Shown that this is not true when $p = 11$
3. Find the gcd for 3096 and 1234.
4. Write the following decimal numbers in binary
   (a) 256_{10}
   (b) $2^4 - 1$
   (c) 549
   (d) 12.34



Download free eBooks at bookboon.com
Click on the ad to read more

5. Convert the following binary numbers into decimal numbers and explain your answers.
   (a) 101.001_2
   (b) 101111_2
   (c) 0.10101_2
   (d) 11.0001_2
   (e) 1001_2
   (f) 0.11_2
6. Convert the following decimal numbers into binary numbers and explain your answers.
   (a) 50_{10}
   (b) 70_{10}
   (c) 64_{10}
   (d) 39.56_{10}
   (e) 20.625_{10}
   (f) 13.11_{10} (8 significant digits)
7. Add the following numbers in binary and explain your answers.
   (a) $111_2 + 111_2$
   (b) $1110_2 + 11_2$
   (c) $11101_2 + 11001_2$
8. Multiply the following numbers in binary and explain your answers.
   (a) $1110_2 \times 11_2$
   (b) $111_2 \times 101_2$

Download free eBooks at bookboon.com

Chapter 2

The statement calculus and logic

"Contrariwise," continued Tweedledee, "if it was so, it might be; and if it were so, it would be; but as it isn't, it ain't. That's logic.
Lewis Carroll

You will have encountered several languages - your native language or the one in which we are currently communicating(English) and other natural languages such as Spanish, German etc. You may also have encountered programming languages like Python or C. You have certainly met some mathematics if you have got this far.

A language in which we describe another language is called a *metalanguage*. For almost all of mathematics, the metalanguage is English with some extra notation.

In computing we need to define, and use, languages and formal notation so it is essential that we have a clear and precise metalanguage. We begin by looking at some English expressions which we could use in computing. Most sentences in English can be thought of as a series of statements combined using connectives such as "and", "or", "if . . . then . . ."

For example the sentence "if it is raining and I go outside then I get wet" is constructed from the three simple statements:

1. "It is raining."
2. "I go outside."
3. "I get wet."

Whether the original sentence is true or not depends upon the truth or not of these three simple statements.

If a statement is true we shall say that its logical value is true, and if it is false, its logical value is false. As a shorthand we shall use the letter T for true and F for false.

Download free eBooks at bookboon.com

We will build compound statements from simple statements like "it is raining", "it is sunny" by connecting them with *and* and *or* In order to make things shorter and we hope more readable, we introduce symbolic notation.

1. Negation will be denoted by $\neg$.
2. "and" by $\wedge$.
3. "or" by $\vee$.

We now look at these connectives in a little more detail.

Negation $\neg$

The negation of a statement is false when the statement is true and is true if the statement is false. So a statement and its negation always have different truth values. For example "It is hot" and "It is not hot." In logic you need to be quite clear about meanings so the negation of,

"All computer scientists are men"

is

"Some computer scientists are men"

NOT

"No computer scientists are men."

The first and third statement are both false!

Download free eBooks at bookboon.com

In symbolic terms if p is a statement, say " it is raining" , then $\neg p$ is its negation. That is $\neg p$ is the statement "it is not raining". We summarize the truth or otherwise of the statements in a *truth table*, see table 2.1.

p	$\neg p$
T	F
F	T

Table 2.1: Truth table for negation ($\neg$)

In the truth table 2.1 the first row reads in plain English - "If p is true then $\neg p$ is false" and row two "If p is false then $\neg p$ is true'.

Conjunction $\wedge$

Similarly, if p and q are statements, then $p \wedge q$ is read as "p and q" . This (confusingly) is called the *conjunction* of p and q.

So if p is the statement " it is green" while q is the statement " it is an apple" then

$p \wedge q$ is the statement "It is green and it is an apple "

We often write this in the shorter form:
If p=" it is green" and q = " it is an apple" then $p \wedge q$ = "It is green and it is an apple "

Clearly this statement is true only when both p and q are true. If either of them is false then the compound statement is false. It will be helpful if we have a precise definition of $\wedge$ and we can get one using a truth table.

p	q	$p \wedge q$
T	T	T
T	F	F
F	T	F
F	F	F

Table 2.2: The truth table for $\wedge$

From table 2 we see that if p and q are both true then $p \wedge q$ is also true. If p is true and q is false then $p \wedge q$ is false.

Download free eBooks at bookboon.com

Disjunction $\vee$

Suppose we now look at "or". In logic we use $p \vee q$ as a symbolic way of writing p or q. The truth table in this case is given in table 2.3 This version of "or" , which

p	q	$p \vee q$
T	T	T
T	F	T
F	T	T
F	F	F

Table 2.3: The truth table for $\vee$

is the common one used in logic is sometimes known as the "inclusive or" because we can have $p \vee q$ true if either one of p and q is true or if *both* are true.

You could of course define the exclusive or , say $\not\equiv$ as having the truth table in 2.4

p	q	$p \not\equiv q$
T	T	F
T	F	T
F	T	T
F	F	F

Table 2.4: The truth table for $\not\equiv$

The Conditional $\Rightarrow$

A rather more interesting connective is "implies" as in p "implies" q. This can be written many ways, for example

- p implies q
- If p then q
- q if p
- p is a sufficient condition for q

I am sure you can think of other variants. We shall use the symbolic form $p \Rightarrow q$ and the truth table for our definition is given in table 2.5.

Download free eBooks at bookboon.com

p	q	$p \Rightarrow q$
T	T	T
T	F	F
F	T	T
F	F	T

Table 2.5: The truth table for $\Rightarrow$

We sometimes call p the *hypothesis* and q the *consequence* or conclusion. Many people find it confusing when they read that " p only if q" is the same as "If p then q". Notice that " p only if q" says that p cannot be true when q is not true, in other words the statement is false if p is true but q is false. When p is false q may be true or false.

You need to be aware that " q only if p" is *NOT* a way of expressing " $p \Rightarrow q$. We see this by checking the truth values. The truth value in line 3 of table 2.5 is the critical difference.

You might like to check that " $\neg p \vee q$ is equivalent to $p \Rightarrow q$, see the table below

p	$\neg p$	q	$\neg p \vee q$
T	F	T	T
T	F	F	F
F	T	T	T
F	T	F	T

Table 2.6: The truth table for $\Rightarrow$

Notice that our definition of implication is rather broader than the usual usage.

Download free eBooks at bookboon.com

Typically you might say

"if the sun shines today we will have a barbecue" .

The hypothesis and the conclusion are linked in some sensible way and the statement is true unless it is sunny and we do not have a barbecue. By contrast the statement

"If the sun shines today 19 is prime"

is true from the definition of an implication because the conclusion is always true no matter if it is sunny or not. If we consider

"if the sun shines today 8 is prime"

The statement is obviously false if today is sunny because 8 is never prime. However the whole statement is true when the sun does not shine today even though 8 is never prime. Of course we are unlikely to make statements like these in real life.

The Biconditional $\iff$

Suppose p and q are two statements. Then the statement "p if and only if q" is called the *biconditional* and denoted by $p \iff q$ or iff. Yes there are two f's! It is true only when p and q have the same logical values, i.e., when either both are true or both are false.

You may also meet the equivalent

- p iff q
- p is necessary and sufficient for q

The truth table is shown in figure 2.7. For example we might say

p	q	$p \iff q$
T	T	T
T	F	F
F	T	F
F	F	T

Table 2.7: The truth table for $\iff$

You can go to the match if and only if you buy a ticket.

This sort of construction is not very common in ordinary language and it is often hard to decide whether a biconditional is implied in ordinary speech. In mathematics or computing you need to be clear if you are dealing with implication $p \Rightarrow q$ or the biconditional $p \iff q$

Download free eBooks at bookboon.com

Converse, contrapositive and inverse

Propositional logic has lots of terminology. So If $p \Rightarrow q$ then

- $q \Rightarrow p$ is the converse.
- $\neg q \Rightarrow \neg p$ is the contrapositive.
- $\neg p \Rightarrow \neg q$ is the inverse.

Truth tables

It is probably obvious that we aim to use logic to help us in checking arguments. We hope to be able to translate from English to symbols. Thus if p is "John learns to cook" and q is " John will find a job" then $p \Rightarrow q$ represents . "If John learns to cook" and then John will find a job" In problems like these the truth table, while cumbersome can be very helpful in giving a mechanical means of checking the truth values of arguments.

To construct tables for compound statements such as $p \vee \neg q \Rightarrow (p \wedge q)$ we need to think about the order we work out the truth values of symbols. The table 2.8 gives the order of precedence.

Precedence	1(Highest)	2	3	4	5(Lowest)
Operator	$\neg$	$\wedge$	$\vee$	$\Rightarrow$	$\Longleftrightarrow$

Table 2.8: Operator precedence

So we negate first, then and etc. As in algebra we also use brackets to indicate that we evaluate the terms in brackets first. Thus for $(p \vee q) \wedge r$ we evaluate the term in brackets $(p \vee q)$ first. Thus

	p	q	$(p \vee q)$	$\neg p$	$(p \vee q) \vee \neg p$
	T	T	T	F	T
	T	F	T	F	T
	F	T	T	T	T
	F	F	F	T	T
precidence	-	-	1	2	3

The vital point about logical statements and about truth tables is :
Two symbolic statements are equivalent if they have the same truth table.
and two statements $p1$ and $p2$ are equivalent, we will write $p1 \Longleftrightarrow p2$.

Download free eBooks at bookboon.com

Thus, for example, the statements $(p \vee q) \wedge \neg p$ and $\neg p \wedge q$ are equivalent. We can deduce this from the truth tables, see table 2.9

p	q	$p \vee q$	p	$(p \vee q) \wedge \neg p$
T	T	T	F	F
T	F	T	F	F
F	T	T	T	T
F	F	F	T	F

p	$\neg p$	q	$\neg p \wedge q$
T	F	T	F
T	F	F	F
F	T	T	T
F	T	F	F

Table 2.9: The truth tables for $(p \vee q) \wedge \neg p$ and$(\neg p \wedge q)$

The reader can use truth table to verify the following equivalences.

1. $\neg(p \vee q) \iff \neg p \wedge \neg q$
2. $\neg(p \wedge q) \iff \neg p \vee \neg q$

One can avoid writing truth tables in table 2.9 and verify the first equivalence as follows:
$p \vee q$ is false only when both p and q are false. Therefore $\neg(p \vee q)$ is true only when both p and q are false. Similarly, $\neg p \wedge \neg q$ is true only when both $\neg p$ and $\neg q$ are true, which is when p and q are false. This proves the equivalence.

Exercise

Construct truth tables for

1. $\neg(p \wedge q)$
2. $\neg(p \vee q) \wedge \neg(q \vee p)$
3. $(p \Rightarrow q) \wedge (q \Rightarrow r) \Rightarrow (p \Rightarrow r)$
4. $(p \vee q \Rightarrow r) \wedge (r \Rightarrow s)$
5. $(p \vee q \Rightarrow r) \wedge (r \Rightarrow s) \Rightarrow (p \Rightarrow r)$

Download free eBooks at bookboon.com

Arguments

We now look briefly at logical arguments and begin with some definitions. Definition:

- A statement that is always true is called a *tautology.*
- A statement that is always false is called a *contradiction.*

So a statement is

1. A tautology if its truth table has no value F.
2. A contradiction if its truth table has no value T.

Notice you may find some writers who say that a formula (in the statement calculus we have just described) is *valid* rather than use the term tautology. The symbol $\vDash$ A is often used as a shorthand for "A is a tautology" or " A is valid".

Examples

1. The statement $p \vee \neg p$ is a tautology, while the statement $p \wedge \neg p$ is a contradiction.
2. The statement $((p \vee q) \wedge p) \iff p$ is a tautology.
3. Two statements $p1$ and $p2$ are *equivalent* when $p1 \iff p2$ is a tautology, and so $p1 \equiv p2$ when $p1 \iff p2$ is a tautology.

Definition 1: Given two statements $p1$ and $p2$ we say that $p1$ *implies* $p2$ if $p1 \Rightarrow p2$ is a tautology.

In everyday life we often encounter situations where we make conclusions based on evidence. In a courtroom the fate of the accused may depend the defence proving that the opposing side's arguments are not valid. A typical task in theoretical sciences is to logically come to conclusions given premises. That is to provide principles for reasoning.

A scientist might say

"if all the premises are true then we have the following conclusion."

Thus they would assert that the conditional

"if all the premises are true then we have the following conclusion"

is a tautology, or that the premises imply his/her conclusion. If his/her reasoning is correct we say that his argument is valid.

Download free eBooks at bookboon.com

Definition 2: A conditional of the form

(a conjunction of statements) implies c

where c is a statement, is called an *argument*. Symbolically

$$p_1, p_2, \ldots, p_m \Rightarrow c$$

The statements in the conjunction on the left side of the conditional are called *premises*, while c is called the *conclusion*.

An argument is valid if it is a tautology, that is, if the premises imply the conclusion (every line of the truth table is T), otherwise it is invalid. So we might have a sequence of premises $p_1, p_2, p_3, \ldots, p_m$ for which c is a valid consequence, symbolically

$$p_1, p_2, p_3, \ldots, p_m \models c$$

You should note that

1. A conjunction of several statements is true only when all the statements are true.
2. A conditional is false only when the antecedent (the left hand side) is true and the consequent (the right hand side) is false.
3. Therefore, an argument is invalid only when there is a situation where all the premises are true, but the conclusion is false. If such a situation cannot occur, the argument is valid.

Exercise s:

1. Is the following argument valid?

   > All birds are mammals and the platypus is a bird. Therefore, the platypus is a mammal.

   Note the premises may be wrong but we are interested in the argument.

2. Sketch how you might show that the statements below below imply that "It rained". Beware this is a big truth table so you are probably best to ensure you understand the method.

   > If it does not rain or if it is not foggy then the regatta will be held and the lifeboat demonstration will go on. If the regatta is held then the trophy will be awarded.

   and

Download free eBooks at bookboon.com

the trophy was not awarded.

3. Show that the following argument is valid.

   Blodwin works hard. If Blodwin works hard then she is a dull girl. If Blodwin is a dull girl she will not get the job therefore Blodwin will not get the job.

So far we have used truth tables only to determine the validity of arguments that are given in symbolic form. However, we can do the same with other arguments by first rewriting them in symbolic form. This is illustrated in the following example.

Either I shall go home or stay and have a drink. I shall not go home. Therefore I stay and have a drink.

Suppose p= I shall go home and q = I shall stay and have a drink. The argument is $\neg p \Rightarrow q$.

p	$\neg p$	q	$\neg p \Rightarrow q$
T	F	T	T
T	F	F	F
F	T	T	T
F	T	F	F

Table 2.10: The truth table for $\Rightarrow$

From the truth table table 2.10 we have a F and so the argument is not valid is , we do not have a tautology. We summarize the process of determining the validity of arguments as follows.

Download free eBooks at bookboon.com

2.0.4 Analyzing Arguments Using Truth Tables

- Step 1: Translate the premises and the conclusion into symbolic form.
- Step 2: Write the truth table for the premises and the conclusion.
- Step 3: Determine if there is a row in which all the premises are true and the conclusion is false. If yes, the argument is invalid, otherwise it is valid.

However truth table can become unwieldy if we have several premises. Consider the following

$$p, r, (p \wedge q) \rightarrow \neg r \vDash \neg q$$

Given we have p, q and r we need 8 rows (2^3) in our table 2.11 as we need all combinations of p, q and r. If we examine line 3 in table 2.11 we can see that when $p, r, (p \wedge q) \rightarrow \neg r$ are all true (we can ignore q) then the result $\neg q$ is true and we have a tautology.

p	q	r	$p \wedge q \Rightarrow \neg r$	$\neg q$	
T	T	T	F	F	
T	T	F	T	F	
T	F	T	T	T	$\leftarrow$
T	F	F	T	T	
F	T	T	T	T	
F	T	F	T	F	
F	F	T	T	T	
F	F	F	T	T	

Table 2.11: Truth table with p, q and r

Now suppose we have p, q, r, s and t. Our table will have $2^5 = 32$ rows.
Take as an example :
If I go to my first class tomorrow , then I must get up early, and if I go to the dance tonight, I will stay up late. If I stay up late and get up early, then I will be forced to exist on only five hours sleep. I cannot exist on five hours of sleep. Therefore I must either miss my fist class tomorrow or not go to the dance.

- Let p be " I go to my first class tomorrow"
- Let q be " I must get up early"
- Let r be " I go to the dance "
- Let s be " I stay up late ".

Download free eBooks at bookboon.com

- Let t be "I can exist on five hours sleep".

The premises are

$$(p \Rightarrow q) \wedge (r \Rightarrow t), s \wedge q \Rightarrow t, \neg t$$

and the conclusion is $\neg p \vee \neg r$. We will prove that $\neg p \vee \neg r$ is a valid consequence of the premises.

Of course we could write out a truth table, however we can try to be cunning.

1. Take the consequence $\neg p \vee \neg r$ and assume that it is FALSE.
2. Then both p and r must be TRUE.
3. The first premise $(p \Rightarrow q) \wedge (r \Rightarrow t)$ implies that q and t are true.
4. So t is true and the last premise is $\neg t$ is assumed TRUE so we have a contradiction.
5. Thus our premise is valid.

I think you might agree that this is a good deal shorter than using truth tables!.

Exercises

Show that

1. $\models (p \Rightarrow q) \Rightarrow ((q \Rightarrow r)) \Rightarrow (p \Rightarrow r))$
2. $\models p \Rightarrow (\neg q \Rightarrow \neg p) \Rightarrow q)$

We add some tables of tautologies which enable us to eliminate conditionals and biconditionals.

1. $\models p \Rightarrow q \iff \neg p \vee q$
2. $\models p \Rightarrow q \iff \neg(p \vee \neg q)$
3. $\models p \vee q \iff \neg p \rightarrow q$
4. $\models p \vee q \iff \neg(p \Rightarrow \neg q)$
5. $\models p \vee q \iff \neg p \rightarrow q$

Download free eBooks at bookboon.com

6. $\models p \vee q \iff \neg p \rightarrow q$

7. $\models p \wedge q \iff \neg(p \Rightarrow \neg q)$

8. $\models p \wedge q \iff \neg(\neg p \vee \neg q)$

9. $\models (p \iff q) \iff (p \Rightarrow q) \wedge (q \Rightarrow p)$

Normal forms

A statement is in *disjunctive normal form* (DNF) if it is a disjunction i.e. a sequence of $\vee$'s consisting of one or more *disjuncts*. Each disjuncts is a conjunction, $\wedge$, of one or more literals (i.e., statement letters and negations of statement letters. For example

1. p

2. $(p \wedge q) \vee (p \wedge \neg r)$

3. $(p \wedge q \wedge \neg r) \vee (p \wedge \neg q)$

4. $p \vee (q \wedge r)$

However $\neg(p \vee q)$ is not a disjunctive normal form($\neg$ is the outermost operator) nor is $p \vee (q \wedge (r \vee s)$ as a $\vee$ is inside a $\wedge$. Converting a formula to DNF involves using logical equivalences, such as the double negative elimination, De Morgan's laws, and the distributive law. All logical formulas can be converted into disjunctive normal form but conversion to DNF can lead to an explosion in the size of of the expression.

A formula is in *conjunctive normal form* (CNF) if it is a conjunction of clauses, where a clause is a disjunction of literals. Essentially we have the same form as a DNF but we use $\wedge$ rather than $\vee$. As a normal form, it is useful (as is the DNF) in theorem proving.

We leave with some ideas which are both important and common in mathematics.

Download free eBooks at bookboon.com

2.0.5 Contradiction and consistency

We say a contradiction is a formula that always takes the value F, for example $p \wedge \neg p$. Then a set of statements $p_1, p_2, \ldots, p_n$ is *inconsistent* if a contradiction can be drawn as a valid consequence of this set.

$p_1, p_2, \ldots, p_n \models q \wedge \neg q$ for some formula b if a contradiction can be derived as a valid consequence of $p_1, p_2, \ldots, p_n \models q$ and $\neg q$

Mathematics is full of proofs by contradiction or Reductio ad absurdum (Latin for "reduction to the absurd"). For example

There are infinitely many prime numbers.

Assume to the contrary that there are only finitely many prime numbers, and all of them are listed as follows: $n_1, n_2 \ldots, p_m$. Consider the number

$$q = n_1 \times n_2 \times \ldots \times p_m + 1$$

Then the number q is either prime or composite. If we divided any of the listed primes n_i into q, there would result a remainder of 1 for each $i = 1, 2, ..., m$ Thus, q cannot be composite. We conclude that q is a prime number, not among the primes listed above, contradicting our assumption that all primes are in the list $n_1, n_2 \ldots, n_m$. Thus there are and infinite number of primes.

there is no smallest rational number greater than 0

Remember that a ration can be written as the ratio of two integers p/q say.

Assume $n_0 = p/q$ is the smallest rational bigger that zero. Consider $n_0/2$. It is clear that $n_0/2 < n_0$ and n_0 is rational. Thus we have a contradiction and can assume that there is no smallest rational number greater than 0.

Download free eBooks at bookboon.com

Chapter 3

Mathematical Induction

I have hardly ever known a mathematician who was capable of reasoning.

Plato (427 BC - 347 BC), The Republic

The integers , $1, 2, 3, 4, \ldots$ are also known as the natural numbers and *Mathematical induction* is a technique for proving a theorem, or a formula, that is asserted about every natural number. Suppose for example we believe

$$1 + 2 + 3 + ... + n = n(n + 1)/2$$

that is the sum of consecutive numbers from 1 to n is given by the formula on the right. We want to prove that this will be true for *all* n. As a start we can test the formula for any given number, say n = 3:

$$1 + 2 + 3 = 3 \times 4/2 = 6$$

It is also true for $n = 4$

$$1 + 2 + 3 + 4 = 4 \times 5/2 = 10$$

But how are we to prove this rule for every value of n? The method of proof we now describe is called the principle of mathematical induction. The idea is simple. Suppose we have some statement that is true for a particular natural number n and we want to prove that it is true for every value of n from $1, 2, 3, \ldots$ If all the following are true

1. When a statement is true for some natural number n, say k.
2. When it is also true for its successor, $k + 1$.
3. The statement is true for some value n, usually $n = 1$.

Download free eBooks at bookboon.com

then the statement is true for every natural number n.

This is because, when the statement is true for $n = 1$, then according to 2, it will also be true for 2. But that implies it will be true for 3; which implies it will be true for 4. And so on. Hence it will be true for every natural number and thus is true for all n.

To prove a result by induction, then, we must prove parts 1, 2 and 3 above. The hypothesis of step 1

"The statement is true for $n = k$"

is called the induction assumption, or the induction hypothesis. It is what we assume when we prove a theorem by induction.

Example

Prove that the sum of the first n natural numbers is given by this formula:

$$S_n = 1 + 2 + 3 + ... + n = n(n + 1)/2$$

We will call this statement S_n, because it depends on n. Now we do steps 1 and 2 above.

1. First, we will assume that the statement is true for $n = k$ that is, we will assume that S_k is true so

   $$S_k = 1 + 2 + 3 + ... + k = k(k + 1)/2$$

   Note this is the induction assumption.

2. Assuming this, we must prove that $S_{(k+1)}$ is also true. That is, we need to show:

   $$S_{(k+1)} = 1 + 2 + 3 + ... + (k + 1) = (k + 1)(k + 2)/2$$

   To do that, we will simply add the next term $(k + 1)$ to both sides of the induction assumption,

   $$S_{(k+1)} = S_{(k+1)} + (k + 1) =$$

   $$1 + 2 + 3 + \ldots + (k + 1) = k(k + 1)/2 + (k + 1) = (k + 1)(k + 2)/2$$

   This is line 2, which is we wanted to show.

3. Next, we must show that the statement is true for $n = 1$. We have $S(1) = 1 = 1 \times 2/2$. The formula therefore is true for $n = 1$.

We have now fulfilled both conditions of the principle of mathematical induction. S_n is therefore true for every natural number.

Download free eBooks at bookboon.com

Example

We prove that $8^n - 3^n$ is divisible by 5 for all $n \in \mathbb{N}$. The proof is by mathematical induction.

1. Assume the result holds for $n = k$, that is $8^k - 3^k \mod 5 = 0$. Then $8^{k+1} - 3^{k+1} = 8 \times 8^k - 3 \times 3^k$.

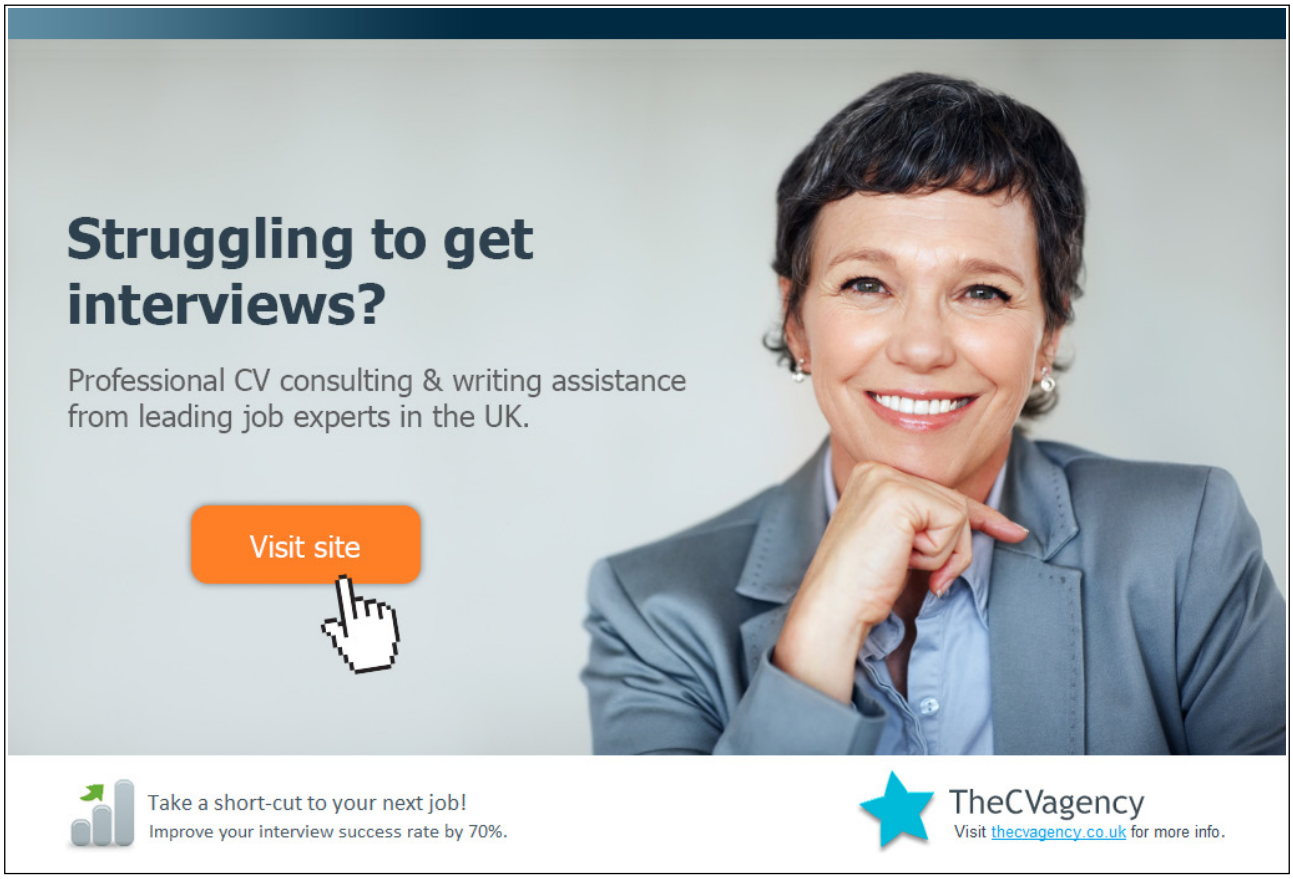
2. Now the clever step

   $$8^{k+1} - 3^{k+1} = 8 \times 8^k - 3 \times 3^k = 3 \times 8^k - 3 \times 3^k + 5 \times 8^k = 3 \times (8^k - 3^k) + 5 \times 8^k$$

   But $8^k - 3^k$ is divisible by 5 (by the induction hypothesis) and $5 \times 8k$ is obviously a multiple of 5. Therefore it follows that $(8^{k+1} - 3^{k+1})$ is divisible by 5. Hence, the result holds for $n = k + 1$.

3. The result holds for $n = 1$ because $8 - 3 = 5$ and so is divisible by 5.

So we have shown that the result holds for all n - by induction.



Download free eBooks at bookboon.com
Click on the ad to read more

Another Example

We prove this rule of exponents:

$$(ab)^n = a^n b^n, \text{ for every natural number } n.$$

Call this statement $S(n)$ and assume that it is true when $n = k$; that is, we assume $S(k) = (ab)^k = a^k b^k$ is true.

We must now prove that $S(k+1)$ is true, that is

$$S(k+1) = (ab)^{k+1} = a^{k+1} b^{k+1}$$

Simply by multiplying both sides of line (3) by ab gives :

$$(ab)^k ab = a^k b^k ab = a^k ab^k b$$

since the order of factors does not matter,

$$(ab)^k ab = a^{k+1} b^{k+1}.$$

Which is what we wanted to show. So, we have shown that if the theorem is true for any specific natural number k, then it is also true for its successor, $k+1$. Next, we must show that the theorem is true for $n = 1$ which is trivial since $(ab)^1 = ab = a^1 b^1$.

This theorem is therefore true for every natural number n.

Exercises

In each of the following $0 \leq n$ is an integer

1. Prove that $n^2 + n$ is even.
2. Prove that $\sum_{i=1}^{n} n^2 = n(n+1)(2n+1)/6$.
3. Prove that $1 + 4 + 7 + \ldots + (3n-2) = n(3n-1)/2$.
4. Prove that $n! \geq 2^n$ when $n > 1$

Download free eBooks at bookboon.com

Chapter 4

Sets

Philosophers have not found it easy to sort out sets . . .

D. M. Armstrong,

It is useful to have a way of describing a collection of "things" and the mathematical name for such a collection is a *set*. So the collection of colours {Red,Blue, Green } is a set we might call A and write as A={Red, Blue, Green } . Other examples are

1. $\{1, 3, 7, 14\}$
2. $\{1, 2, 3, 5, 7, 11 \ldots\}$ the set of all prime numbers.
3. { Matthew, Mark, Luke, John}
4. {k : k is an integer and k is divisible by 4} here the contents are defined by a rule.
5. { All songs available on iTunes} again the contents are defined by a rule.

We do not care about the order of the elements of a set so {1, 2, 3} *is the same as* {3,2,1}.

Of course we may want to do things with sets and there is a whole mathematical language attached as you might expect. For example you will often see the statement a *belongs* to the set A written as $a \in \mathsf{A}$. The symbol $\notin$ is, of course, the converse i.e. *does not belong to.*

So

- Mark $\in$ {Matthew, Mark, Luke, John}
- Abergail $\notin$ {Matthew, Mark, Luke, John} .

Download free eBooks at bookboon.com

- 7∈ {1,2,3,4,5,6,7}

There are some sets that have special symbols because they are used a lot. Examples are

1. The set with nothing in it, called the *empty set* is written as $\emptyset$.
2. $\mathbb{N} = \{1, 2, 3, \ldots\}$ the set of natural numbers.
3. $\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$ the integers.
4. $\mathbb{Q}$ = the set of fractions.
5. $\mathbb{R}$ = the set of real numbers.
6. The set that contains everything is called the *universal* set written S, U or $\emptyset$.

Finally we will write $\bar{A}$ when we mean the set of things which are not in A.

Subsets

It is probably obvious that some set are "bigger" than others, for example {A,B,C,D,E} and {B,C,D}. We formalize this idea by defining *subsets*.

If the set B contains all the elements in the set A together with some others then we write $A \subset B$. We say that A is a *subset* of B. So
{Matthew, Mark, Luke, John} ⊂ {Matthew, Mark, Luke, John, Thomas }

We can of course write this the other way around, so $A \subset B$ is the same as $B \supset A$.

1. Formally for $A \subset B$ we say if $a \in A$ then $a \in B$ or

$$a \in A \Rightarrow a \in B$$

2. If B is a subset but might possibly be the same as A then we use A⊆B.
3. We will use $A = B$ to mean A contains *exactly the same* things as B. Note that if $A \subseteq B$ and $B \subseteq A$ then $A = B$.

   In our logical symbolism we have

$$(A \subseteq B) \wedge (B \subseteq A) \Rightarrow A = B.$$

Download free eBooks at bookboon.com

The *power set* of A, written, $P(A)$, or 2^A, is the set of all subsets of A. So if A = { Matthew, Mark, Luke } then P(A) is the set with eight elements
{ Matthew, Mark, Luke }
{ Matthew, Mark }
{ Matthew, Luke }
{ Mark, Luke }
{ Matthew }
{ Mark }
{ Luke }
$\emptyset$
The number of elements in a set A is called the *cardinality* of A and written $\|A\|$. So if A = { Matthew, Mark, Luke, John} then $\|A\|$=4.

Venn Diagrams and Manipulating Sets

We intend to manipulate sets and it helps to introduce Venn diagrams to illustrate what we are up to. We can think of the universal set S as a rectangle and a set, say A as the interior of the circle drawn in S, see figure 4.1 The speckled area is

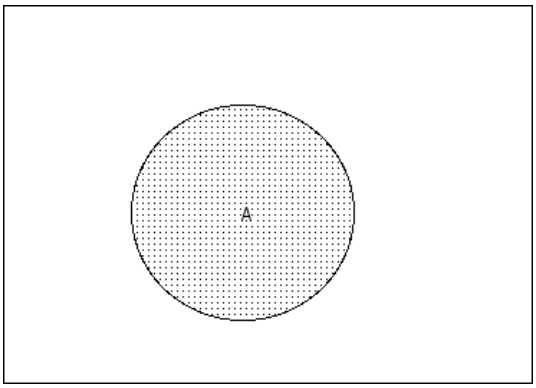


Figure 4.1: Venn diagram of set A and universal set S

A while the remainder of the area of the rectangle is $\bar{A}$. We see immediately that A together with $\bar{A}$ make up S

Download free eBooks at bookboon.com

Intersection

We can write the set of items that belong to both the set A and the set B as $A \cap B$. Formally $(x \in A) \wedge (x \in B) \Rightarrow (x \in A \cap B)$.

We call this the *intersection* of A and B or, less formally, A *and* B. In terms of the Venn diagram in figure 4.2 the two circles represent A and B while the overlap (in black) is the intersection. As examples

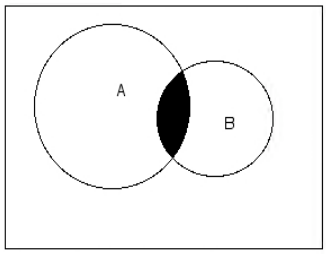


Figure 4.2: Venn diagram of $A \cap B$

1. {1,2,3,4} $\cap$ { 3,4,5,6,7} ={ 3,4}. Notice $3 \in$ { 3,4} while $1 \notin$ { 3,4}.
2. {1,2,3,4} $\cap$ { 13,14,15,16,27} =$\emptyset$.
3. {Abergail, Ann, Blodwin, Bronwin, Clair,}$\cap$ { Abergail, Bronwin, Gareth, Ian} = {Abergail, Bronwin, }.
4. In figure 4.2 we see $A \cap \bar{A} = \emptyset$ so A and $\bar{A}$ have nothing in common.
5. $A \cap B \subset B$ and $A \cap B \subset A$

Union:

We can write the set of items that belong to the set A or the set B or to both as A $\cup$ B. Formally $(x \in A) \vee (x \in B) \Rightarrow x \in (A \cup B)$.

We call this the *union* of A and B or, less formally, A *or* B. The corresponding diagram is 4.3 Here the speckled area represents $A \cup B$

Download free eBooks at bookboon.com

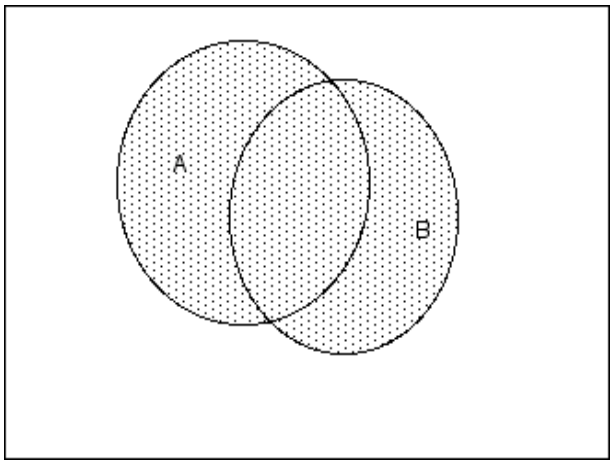


Figure 4.3: Venn diagram of set $A \cup B$ (speckled) and universal set S

As examples we have

1. {1,2,3,4} $\bigcup$ { 3,4,5,6,7} ={ 1,2,3,4,5,6,7}
2. { Blue,Green} $\bigcup$ { Red,Green} ={ Red,Blue , Green}
3. In figure 4.2 we see $A \cup \bar{A} = S$ so A and $\bar{A}$ together make up S.
4. If $A \subset B$ then $A \cup B \subset B$

We can now use our basic definitions to get some results.

bookboon.com

Join our sales team! We are a dynamic global business experiencing rapid growth

Are you looking for a new challenging position that requires self-motivation and allows you to work independently?

Bookboon is the world's largest eBook publisher with a very strong presence in USA.

Established in 1988, Bookboon is currently operating in more than 30 countries.

We have an impressive client base of over 100 major accounts. Bookboon is looking for dedicated sales people in USA to be part of our ongoing success.

As a part of our worldwide sales team, you will be rewarded with a competitive sales and target structure and uncapped earning potential. We are looking for sales people with experience in selling solutions towards CEOs, HR, Marketing or sales departments.

Please send your CV to jobs@bookboon.com

Apply Now!

Download free eBooks at bookboon.com

1. A= $\bar{\bar{A}}$ The set $\bar{A}$ consists of all the elements of S (the universal set) which do not belong to A. So $\bar{\bar{A}}$ is the set of elements that do not belong to $\bar{A}$, or the elements of S which do not belong to $\bar{A}$. That is the elements that belong to A.

   Or suppose $a \in \bar{\bar{A}} \Rightarrow a \notin \bar{A} \Rightarrow a \in A$

2. $\overline{(A \cap B)} = \bar{A} \cup \bar{B}$
   We have $a \in \overline{(A \cap B)} \Rightarrow a \notin (A \cap B) \Rightarrow (a \notin A) \vee (a \notin B) \Rightarrow (a \in \bar{A}) \vee (a \in \bar{B}) \Rightarrow a \in \bar{A} \cup \bar{B}$

There is a table of useful results in table 4.1. Notice each rule in the left column has a dual rule in the right. This dual has the $\cup$ symbol replace by $\cap$

$A \cup A = A$	$A \cap A = A$
$(A \cup B) \cup C = A \cup (B \cup C)$	$(A \cap B) \cap C = A \cap (B \cap C)$
$A \cup B = B \cup A$	$A \cap B = B \cap A$
$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
$A \cup \emptyset = A$	$A \cap S = A$
$A \cup S = S$	$A \cap \emptyset = \emptyset$
$A \cup \bar{A} = S$	$A \cap \bar{A} = \emptyset$
$\overline{(A \cup B)} = \bar{A} \cap \bar{B}$	$\overline{(A \cap B)} = \bar{A} \cup \bar{B}$

Table 4.1: Rules for set operations

Cartesian Product

Suppose we have two sets A and B. We define the Cartesian Product $P = A \times B$ to be the set of ordered pairs (a, b) where $a \in A$ and $b \in B$. Or

$$P = \{(a, b) : (a \in A) \wedge (b \in B)\}.$$

The pair (a, b) is ordered in the sense that the first term (a) comes from the set A in $A \times B$. The obvious example and hence the name comes from the geometry of the plane. We usually write (x, y) to denote the coordinates of a point on the plane. This is an ordered pair! If we take real values x and y with $x \in \mathbb{R}$ and $y \in \mathbb{R}$ then the Cartesian product is $\mathbb{R} \times \mathbb{R}$

1. Suppose $A = \{a, b\}$ and $B = \{1, 2\}$ then $A \times B = \{(a, 1), (a, 2), (b, 1), (b, 2)\}$.

2. We can extend to 3 or more sets so $A \times B \times C$ is the set of ordered triples (a, b, c).

Download free eBooks at bookboon.com

4.0.6 Relations and functions

Given two sets A and B and the product $A \times B$ we define a *relation* between A and B as a subset R of $A \times B$. We say that $a \in A$ and $b \in B$ are related if $(a, b) \in R$, more commonly written aRb. This is a quite obscure definition unless we look at the rule giving the subset.

Take the simple example of $A = \{1, 2, 3, 4, 5, 6\}$ and $B = \{1, 2, 3, 4, 5, 6\}$ then $A \times B$ is the array of pairs below - a set of 36 pairs.

$$\left\{ \begin{array}{cccccc} (1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\ (3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\ (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6) \end{array} \right\}$$

A *relation* R is the subset $\{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\}$ or the set $\{(i, j) : i = j\}$. Other example are

1. $R = \{(i, j) : i + j = 8\} = \{(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)\}$

2. $R = \{(i, j) : i = 2j\} = \{(2, 1), (4, 2), (6, 3)\}$

3. $R = \{i < j\} = \left\{ \begin{array}{cccccc} (1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ & & (2,3) & (2,4) & (2,5) & (2,6) \\ & & & (3,4) & (3,5) & (3,6) \\ & & & & (4,5) & (4,6) \\ & & & & & (5,6) \end{array} \right\}$

As you can see we can think of the relation R as a rule connecting elements of A to elements of B. The relation aRb between 2 sets A and B can be represented as in figure 4.4

For example

1. if A={ one, two, three, four, five} and $B = \{1, 2, 3, 4, 5\}$ we can define R as the set of pairs {(word, number of letters)} eg. {(one, 3), (two, 3), (three,5), . . .}

2. If A={ 2,4,8,16,32} and B={1,2,3,4,5} then we might define R as the set { (2,1),(4,2),(8,3),(16,4),(32,5)}

1. The *domain* of relation $\{(x, y)\}$ is the set of all the first numbers of the ordered pairs. In other words, the domain is all of the x-values.

2. The *range* of relation $\{(x, y)\}$ is the set of the second numbers in each pair, or the y-values.

Download free eBooks at bookboon.com

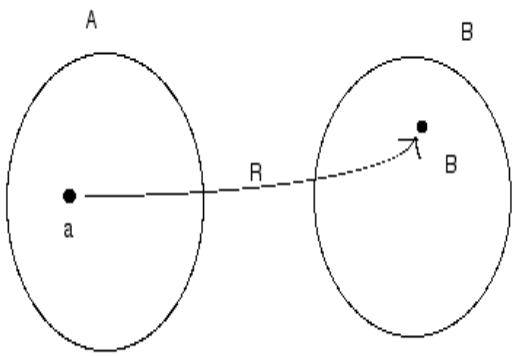


Figure 4.4: The relation R between 2 sets A and B

There are all kinds of names for special types of relations. Some of them are

1. *reflexive:* for all $x \in X$ it follows that xRx. For example, "greater than or equal to" is a reflexive relation but "greater than" is not.

2. *symmetric:* for all x and y in X it follows that if xRy then yRx. "Is a blood relative of" is a symmetric relation, because x is a blood relative of y if and only if y is a blood relative of x.

3. *antisymmetric:* for all x and y in X it follows that if xRy and yRx then $x = y$. "Greater than or equal to" is an antisymmetric relation, because if $x \geq y$ and $y \geq x$, then $x = y$.

Download free eBooks at bookboon.com

Click on the ad to read more

4. *asymmetric:* for all x and y in X it follows that if xRy then not yRx. "Greater than" is an asymmetric relation, because $\text{if} x > y$ then $y \not> x$.

5. *transitive:* for all x, y and z in X it follows that if xRy and yRz then xRz. "Is an ancestor of" is a transitive relation, because if x is an ancestor of y and y is an ancestor of z, then x is an ancestor of z.

6. *Euclidean:* for all x, y and z in X it follows that if xRy and xRz, then yRz.

7. A relation which is reflexive, symmetric and transitive is called an *equivalence relation.*

You can now speculate as the name "Relational Database".

exercises

1. If $A - B$ is the set of elements x that satisfy $x \in A$ and $x \notin B$ draw a Venn diagram for $A - B$

2. Prove that for sets A, B and C

   (a) If $A \subseteq B$ and $B \subseteq C$ then $A \subseteq C$
   (b) If $A \subseteq B$ and $B \subset C$ then $A \subset C$
   (c) If $A \subset B$ and $B \subseteq C$ then $A \subset C$
   (d) If $A \subset B$ and $B \subset C$ then $A \subset C$

3. Recall that $\mathbb{Z} = \{0, 1, 2, 3, 4, \ldots\}$ and we define the following sets

   (a) $A = \{x \in \mathbb{Z} : \text{ for some integer } y > 0, x = 2y\}$
   (b) $B = \{x \in \mathbb{Z} : \text{ for some integer } y > 0, x = 2y - 1\}$
   (c) $A = \{x \in \mathbb{Z} : \text{ for some integer } x < 10\}$

   Describe $\bar{A}, (A \bar{\cup} B), \bar{C}, A - \bar{C}, \text{ and} C - (A \cup B)$

4. Show that for all sets A, B and C

$$(A \cap B) \cup C = A \cap (B \cup C)$$

   iff $C \subseteq A$

5. What is the cardinalty of $\{\{1, 2\}, \{3\}, 1\}$.

6. Give the domain and the range of each of the following relations. Draw the graph in each case.

Download free eBooks at bookboon.com

(a) $\{(x, y) \in \mathbb{R} \times \mathbb{R}\} \mid x^2 + 4y^2 = 1\}$

(b) $\{(x, y) \in \mathbb{R} \times \mathbb{R}\} \mid x^2 = y^2\}$

(c) $\{(x, y) \in \mathbb{R} \times \mathbb{R}\} \mid 0 \leq y, y \leq x$ and $x + 1y \leq 1\}$

7. Define the relation $\rhd$ between the ordered pairs $\{(x, y)$ and (u, v) where $x, y, v, v \in \mathbb{Z}\}$ where $(x, y) \rhd (u, v)$ means $xv = yu$. Show that $\rhd$ is an equivalence relation.

TURN TO THE EXPERTS FOR
SUBSCRIPTION CONSULTANCY

Subscrybe is one of the leading companies in Europe when it comes to innovation and business development within subscription businesses.

We innovate new subscription business models or improve existing ones. We do business reviews of existing subscription businesses and we develope acquisition and retention strategies.

Learn more at linkedin.com/company/subscrybe or contact
Managing Director Morten Suhr Hansen at mha@subscrybe.dk

SUBSCRYBE - to the future

Download free eBooks at bookboon.com

Click on the ad to read more

Chapter 5

Counting

There are three types of people in this world: Those who can count, and those who can't.

Counting seem quite simple but this is quite deceptive, especially when we have complicated system. If you do not believe me have a look at the probability section. To make like a little simpler we lay down some rules.

Sets

If we have two sets A and B the number of item in the sets (the cardinality) is written $\|A\|$ and $\|B\|$. Then we can show that

$$\|A \cup B\| = \|A\| + \|B\| - \|A \cap B\|$$

This is fairly easy to see if you use a Venn diagram. For 3 sets

$$\|A \cup B\| = \|A\| + \|B\| + \|C\| - \|A \cap B\| - \|B \cap C\| - \|A \cap C\| + \|A \cap B \cap C\|$$

Example

Let S be the set of all outcomes when two dice (one blue ; one green) are thrown. Let A be the subset of outcomes in which both dice are odd, and let B be the subset of outcomes in which both dice are even. We write C for the set of outcomes when the two dice have the same number showing.

How many elements are there in the following sets?

It is useful to have the set S set out as below

Download free eBooks at bookboon.com

1 1	1 2	1 3	1 4	1 5	1 6
2 1	2 2	2 3	2 4	2 5	2 6
3 1	3 2	3 3	3 4	3 5	3 6
4 1	4 2	4 3	4 4	4 5	4 6
5 1	5 2	5 3	5 4	5 5	5 6
6 1	6 2	6 3	6 4	6 5	6 6

then we have

1. $\|A\| = 9$
2. $\|B\| = 9$
3. $\|C\| = 6$
4. $\|A \cap B\|$=0
5. $\|A \cup B\| = 18$
6. $\|A \cap C\| = \|(1,1),(3,3),(5,5)\| = 3$
7. $\|A \cup C\| = \|A\| + \|C\| - \|A \cap C\| = 9 + 6 - 3 = 12$

Chains of actions

If we have to perform two actions in sequence and the first can be done m ways while the second can be done in n there will be mn possibilities in total.

- Suppose we wish to pick 2 people from 9. The first can be picked in 9 ways the second in 8 giving $9 \times 8 = 72$ possibilities in total.
- If we roll a die and then toss a coin there are $6 \times 2 = 12$ possibilities.

Download free eBooks at bookboon.com

This extends to several successive actions. Thus

1. If we roll a die 3 times then there are $6 \times 6 = 216$ possibilities.
2. If we toss a coin 7 times there are $2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 = 2^7 = 128$ possibilities.
3. My bicycle lock has 4 rotors each with 10 digits. That gives $10 \times 10 \times 10 \times 10 = 10^4$ combinations.
4. Suppose you have to provide an 8 character password for a credit card company. They say that you can use a to z (case is ignored) and 0 to 1 but there must be at least one number and at least one letter.

   there are 26 letters and 10 numbers so you can make 8^{36} possible passwords. Of these there are 8^{10} which are all numbers and 8^{26} which are all letters. This gives $8^{36} - 8^{26} - 8^{10} = 3.245 \times 10^{32}$ allowable passwords.

Permutations

Suppose I have n distinct items and I want to arrange them in a line. I can do this in

$$n \times (n-1) \times (n-2) \times (n-3) \times \cdots \times 3 \times 2 \times 1$$

We compute this product so often it has a special symbol $n!$. However to avoid problems we *define*

$$1! = 0 \text{ and } 0! = 1$$

So $3! = 3 \times 2 \times 1 = 6$ while $5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$

If we look at the characters in (1D4Y) there are $4! = 24$ possible distinct arrangements.

Sometimes we do not have all distinct items. We might have n item of which r are identical then there are $n!/r!$ different possible arrangements. So WALLY can be arranged in $5!/2! = 60$ ways.

It is simpler to just state a rule in the more general case:

Suppose we have n objects and

- there are n_1 of type 1.
- there are n_2 of type 2.
- $\cdots\cdots$
- there are n_k of type k.

Download free eBooks at bookboon.com

The total number of items in n, so $n = n_1 + n_2 + \cdots n_k$ then there are

$$\frac{n!}{n_1!n_2!n_3!\cdots n_k!}$$

possible arrangements.

Suppose we have 3 white, 4 red and 4 black balls. They can be arranged in a row in

$$\frac{11!}{3!4!4!} = 11550$$

possible ways while the letters in WALLY can be arranged in

$$\frac{5!}{2!1!1!1!} = 60 \text{ ways}$$

Combinations

The number of ways of picking k items from a group of size n is written $\binom{n}{k}$ or (for the traditionalists) ${}_nC_k$. The definition is

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

So the number of ways of picking 5 students from a group of 19 is

$$\binom{19}{5} = \frac{19!}{5!14!} = \frac{19 \times 18 \times 17 \times 16}{4 \times 3 \times 2 \times 2}$$

Download free eBooks at bookboon.com

Click on the ad to read more

Examples

1. Suppose you want to win the lottery. There are 49 numbers and you can pick 6. This can be done in

$$\frac{49!}{6!43!} = 13983816 \text{ ways}$$

so your chances of a win are 1/13983816.

2. How many ways can you pick 5 correct numbers in the lottery. There are $\binom{6}{5}$ ways to pick the 5 correct numbers and 49-6=43 ways of picking the remaining number. This gives 6×43 ways.

3. When we pick 3 correct numbers there are $\binom{6}{3}$ ways of picking the winning numbers and $\binom{43}{3}$ ways of picking the losing ones. This gives $\binom{6}{3} \times \binom{43}{3} = 20 \times 12341 = 246820$ ways in all.

5.0.7 Binomial Expansions

Now we have combinations we can examine a very useful result known as the binomial expansion. To start we can show that

$$(a+b)^2 = a^2 + 2ab + b^2$$

and

$$(a+b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$$

In general we can prove that for an integer $n > 0$

$$(a+b)^n = a^n + \binom{n}{1}a^{n-1}b + \binom{n}{2}a^{n-2}b^2 + \cdots + \binom{n}{n-2}a^2b^{n-2} + \binom{n}{n-1}ab^{n-1} + b^n$$

or

$$(a+b)^n = \sum_{i=0}^{n} \binom{n}{i} a^{n-i}b^i.$$

This can be done by induction, but there isis a page or so of algebra!

For example

$$(2+x)^5 = 2^5 + \binom{5}{1}2^4x + \binom{5}{2}2^3x^2 + \binom{5}{3}2^2x^2 + \binom{5}{4}2x^4 + x^5$$

or

$$(2+x)^5 = 2^5 + 5 \times 2^4x + 10 \times 2^3x^2 + 10 \times 2^2x^2 + 5 \times 2x^4 + x^5$$

Download free eBooks at bookboon.com

Suppose you were given $\left(3x + 5/x^3\right)^8$ and you wanted the term in the expansion which *did not* have an x. From the above the general term is

$$\binom{8}{i}(3x^3)^{8-i}(5/x^3)^i.$$

The x terms cancel when $8 - i = 3i$ or $i = 2$. Then the term is

$$\binom{8}{2}(3x^3)^6(5/x^3)^2 = \binom{8}{2}3^6 5^2$$

We can do something similar for non-integral n as follows:

$$(1+x)^n = 1+nx+\frac{n(n-1)}{1.2}x^2+\frac{n(n-1)(n-2)}{1.2.3}+\cdots+\frac{n(n-1)(n-2)\cdots(n-k+1)}{1.2.3\cdots k}x^k+\cdots$$

but this is only true when $|x| < 1$.

Thus $(1 + x)^{1/2} = 1 + \left(\frac{1}{2}\right)x^{1/2} + \left(\frac{1}{2}\right)\left(-\frac{1}{2}\right)x^{-1/2} + \left(\frac{1}{2}\right)\left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right)x^{-3/2} + \ldots$

Examples

1. Suppose we look at sports scholarships awarded by American universities. A total of 147,000 scholarships were earned in 2001. Out of the 5,500 scholarships for athletics, 1500 were earned by women. Women earned 75,000 scholarships in total. How many men earned scholarships in athletics?

2. In clinical trials of the suntan lotion, Delta Sun, 100 test subjects experienced third degree burns or nausea (or both). Of these, a total of 35 people experienced third degree burns, and 25 experienced both third degree burns and nausea. How many subjects experienced nausea?

3. A total of 1055 0 MSc degrees were earned in 2002. Out of the 41 MSc degrees in music and music therapy, 5 were earned by men. Men earned 650 MSc degrees. How many women earned MSc degrees in fields other than music and music therapy?

4. A survey of 200 credit card customers revealed that 98 of them have a Visa account, 113 of them have a Master Card, 62 of them have a Visa account and a American Express, 36 of them have a Master Card account and an American Express, 47 of them have only a Master Card account, 32 have a Visa account and a Master Card account and an American Express. Assume that every customer has at least one of the services. The number of customers who have only have a Visa card is?

Download free eBooks at bookboon.com

5. So for example from the New York Times According to a New York Times report on the 16 top-performing restaurant chains

   (a) 11 serve breakfast.

   (b) 11 serve beer.

   (c) 10 have full table service i.e. they server alcohol and all meals.

   All 16 offered at least one of these services. A total of 5 were classified as "family chains," meaning that they serve breakfast, but do not serve alcohol. Further a total of five serve breakfast and have full table service, while none serve breakfast, beer, and also have full table service. We ask

   (a) (How many serve beer and breakfast?

   (b) How many serve beer but not breakfast?

   (c) How many serve breakfast, but neither have full table service, nor serve beer?

   (d) How many serve beer and have full table service?

6. When $| x |< 1$ then show that

   - $1/(1-x) = 1 + x + x^2 + x^3 + x^4 + \cdots + x^n + \cdots$
   - $1/(1-x)^{1/2} = 1 + (1/2)x + \frac{(1/2)(-1/2)x^2}{1.2} + \frac{(1/2)(-1/2)(-3/2)x^3}{1.2.3} + x^4 + \cdots$
   - $1/(1-x)^2 = 1 + 2x + 3x^2 + 4x^3 + 5x^4 + \ldots + nx^{n-1} + \ldots$

7. Expand $(1+2x)^7$

8. Which is the coefficient of the term without an x in $(x + 2/x)^{11}$.

9. Find an approximation for $(0.95)^{11}$.

10. Find the first 3 terms of the expansion of $(1+x)^{1/4}$.

Download free eBooks at bookboon.com

Chapter 6

Functions

Mathematicians are like Frenchmen: whatever you say to them they translate into their own language and forthwith it is something entirely different.

Johann Wolfgang von Goethe

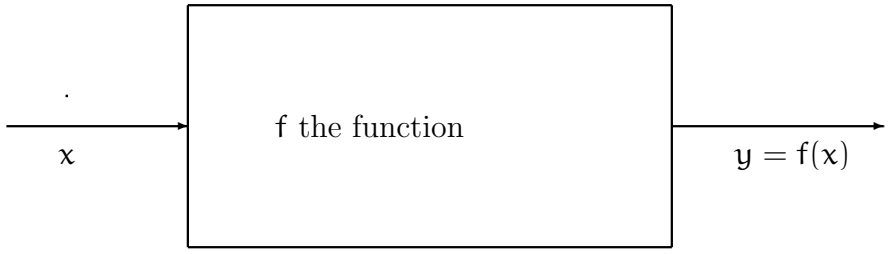
One of the most fundamental (and useful) ideas in mathematics is that of a *function*. As a preliminary definition suppose we have two sets X and $Y$and we also have a rule which assigns to every $x \in X$ a UNIQUE value $y \in Y$. We will call the rule f and say that for each x there is a $y = f(x)$ in the set Y. This is a very wide definition and one that is very similar to that of a relation , the critical point is that for each a there is a *unique value* y. A common way of writing functions is

$$f : X \to Y$$

which illustrates that we have two sets X and Y together with a rule f giving values in Y for values in X. We can think of the pairs (x, y) or more clearly $(x, f(x))$. This set of pairs is the *graph* of the function

In what follows we show how functions arise from the idea of relations and come up with some of the main definitions. You need to keep in mind the simple idea a function is a rule that takes in x values and produces y values. It is probably enough to visualize f as a device which when given an x value produces a y.

Download free eBooks at bookboon.com



Download free eBooks at bookboon.com

Click on the ad to read more

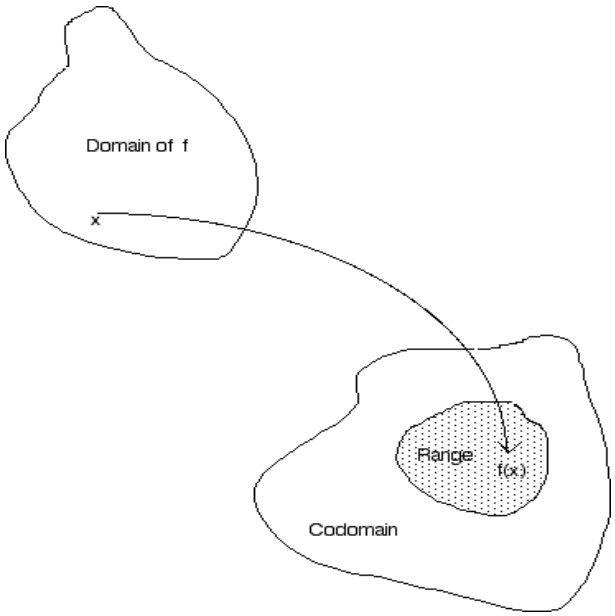


Figure 6.1: Function f

Clearly if you think of f as a machine we need to take care about what we are allowed to put in, x, and have a good idea of the range of what comes out, y. It is these technical issues we look at next.

The set X is called the domain of the function f and Y is *codomain*. We are normally more interested in the set of values $\{ f(x) : x \in X\}$. This is the *range* R sometimes called the *image* of the function. See figure 6.1

Examples

We can have

$$f : X \to Y$$

where

1. $f(x) = 2^x$ where $X = \{x : 0 \leq x < \infty\}$ and $Y = \{y : 0 \leq x < \infty\}$
2. $f(x) = \sqrt{x}$ where $X = \{x : 0 \leq x < \infty\}$ and $Y = \{y : 0 \leq y < \infty\}$
3. $f(x) = \sin^{-1}(x)$ where $X = \{x : -\pi/2 \leq x < pi/2\}$ and $Y = \{-1 \leq y \leq 1\}$

If we think of the possibilities we have

- There may be some points in Y (the codomain) which cannot be reached by function f. If we take all the points in X and apply f we get a set

Download free eBooks at bookboon.com

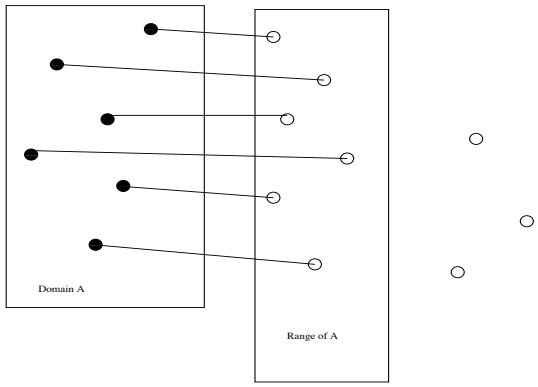


Figure 6.2: An onto function

$R = \{f(x) : x \in X\}$ which is the *range* of the function f. Notice R is a subset of Y i.e.$R \subset Y$.

- *Surjections* (or *onto* functions) have the property that for every y in the codomain there is an x in the domain such that $f(x) = y$. If you look at 6.1 you can see that in this case the codomain is bigger than the range of the function. See figure 6.2 If the range and codomain are the same then out function is a surjection. This means every y has a corresponding x for which $y = f(x)$
- Another important kind of function is the injection (or one-to-one function), which have the property that if $x_1 = x_2$ then y_1 must equal y_2. See figure 6.3
- Lastly we call functions *bijections*, when they are are both one-to-one and onto.

A more straightforward example is as follows. Suppose we define

$$f : X \to Y$$

where $f(x) = 2^x$ and $X = \{x : 0 \leq x < \infty\}$ and $Y = \{y : -\infty \leq x < \infty\}$. The range of the function is $R = \{y : 0 \leq x < \infty\}$ while the codomain Y has negative values which we cannot reach using our function.

Composition of functions

The *composition* of two or more functions uses the output of one function, say f, as the input of another, say g. The functions $f : X \to Y$ and $g : Y \to Z$ can be

Download free eBooks at bookboon.com

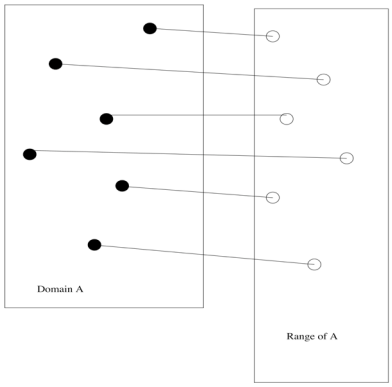


Figure 6.3: An 1 to 1 function

composed by applying f to an argument x to obtain $y = f(x)$ and then applying g to y to obtain $z = g(y)$. See figure 6.4. The composite function formed in this way from f and g can be written $g(f(x))$ or $g \circ f$. This last form can be a bit dangerous as the order can be different in different subjects. Using composition we can construct complex functions from simple ones, which is the point of the exercise.

One interesting function, given f, would be the function g for which x=g(f(x)). In other words g is the *inverse function*. Not all functions have inverses, in fact there is an inverse g written f^{-1} if and only if f is bijective. In this case $x = f^{-1}(f(x)) = f(f^{-1}(x))$.

The arrows and blob diagrams are not the usual way we draw functions. You will recall that the technical description of $f : X \to Y$ is the set of values $(x, f(x))$ Suppose we take the reals $\mathbb{R}$ so our function takes real values and gives us a new set of reals, say $f(x) = x^3$ we take x values , compute $y = f(x)$ for these values and plot them as in figure 6.6. Plotting functions is a vital skill, you know very little about a function until you have drawn the graph. It need not be very accurate, mathematicians often talk about *sketching* a function. By this they mean a drawing which is not completely accurate but which illustrates the main characteristics of the function,

Now we might reasonably does every sensible looking function have an inverse? An example consider $f(x) = x^2$ which is plotted in figure 6.8. There is now problem in the definition of f for all real values of x, that is the domain is $\mathbb{R}$ and the codomain $\mathbb{R}$. However if we examine the inverse we have a problem.

if we take y=4, this may arise from x=2 or x=-2. So there is not an $f^{-1} = y^{-1/2}$! If we change the domain we can get around this. Suppose we define $\mathbb{R}^+ = \{x :$

Download free eBooks at bookboon.com

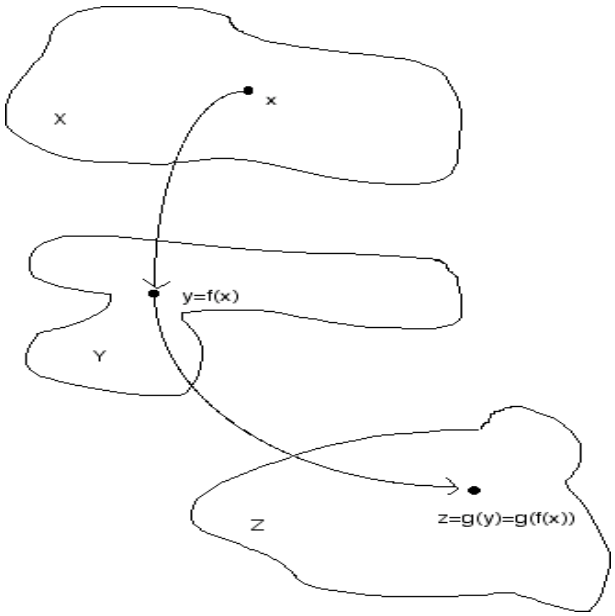


Figure 6.4: Composition of two functions f and g

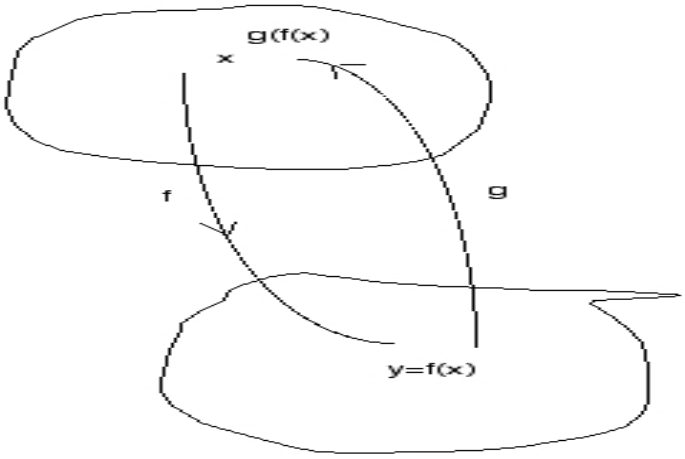


Figure 6.5: The inverse f and $g = f^{-1}$

Examples

1. Suppose $f(x) = x^2$ and $g(y) = 1/y$ then $g(f(x)) = 1/x^2$. We of course have to take care about the definition if the range and the domain to avoid $x = 0$

2. When $f(x) = x^2$ and $g(x) = x^{1/2}$ g is the inverse function when f is defined on the positive reals.

Download free eBooks at bookboon.com

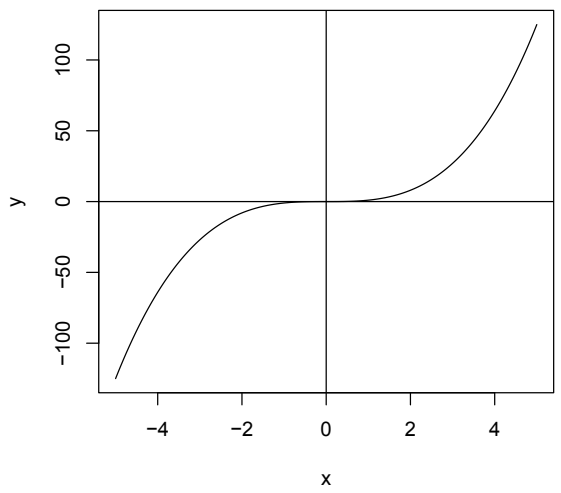


Figure 6.6: Plot of $f(x) = x^3$

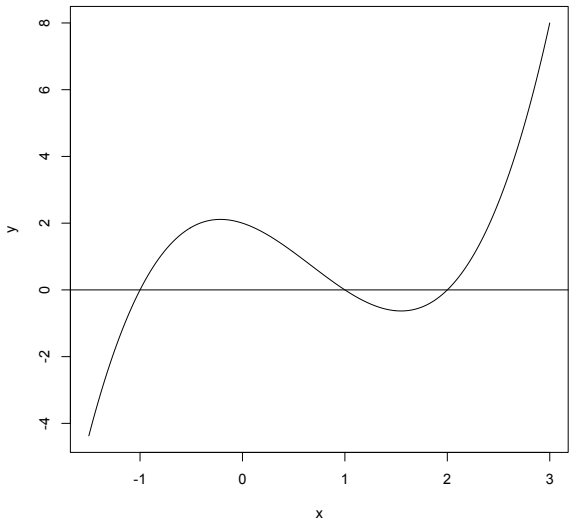


Figure 6.7: Plot of $f(x) = x^3 - 2x^2 - x + 2$

Download free eBooks at bookboon.com

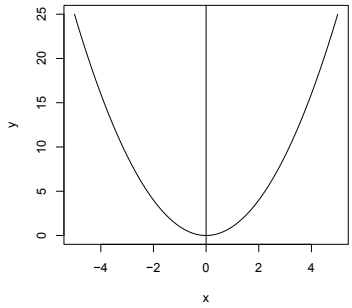


Figure 6.8: Plot of $f(x) = x^2$

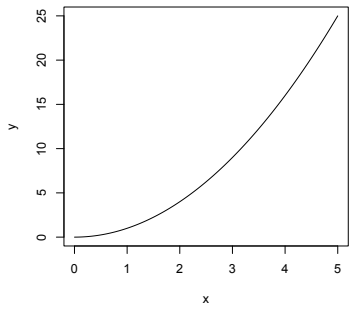


Figure 6.9: Plot of $f(x) = x^2$

$0 \leq x < \infty\}$ and consider $f(x) = x^2$ defined on $\mathbb{R}^+$ i.e.

$$\mathbb{R}^+ : f \to \mathbb{R}^+$$

In this case we do not have the problem of negative values of x. Every value of y arises from a unique x.

Download free eBooks at bookboon.com

Exercises

For the following pairs evaluate $g(f(x))$ and $f(g(x))$.

1. $f(x) = 1/x, \ g(x) = x^2$
2. $f(x) = 3 + 4x, \ g(x) = 2x - 5$
3. $f(x) = x + 1, \ g(x) = x - 1$

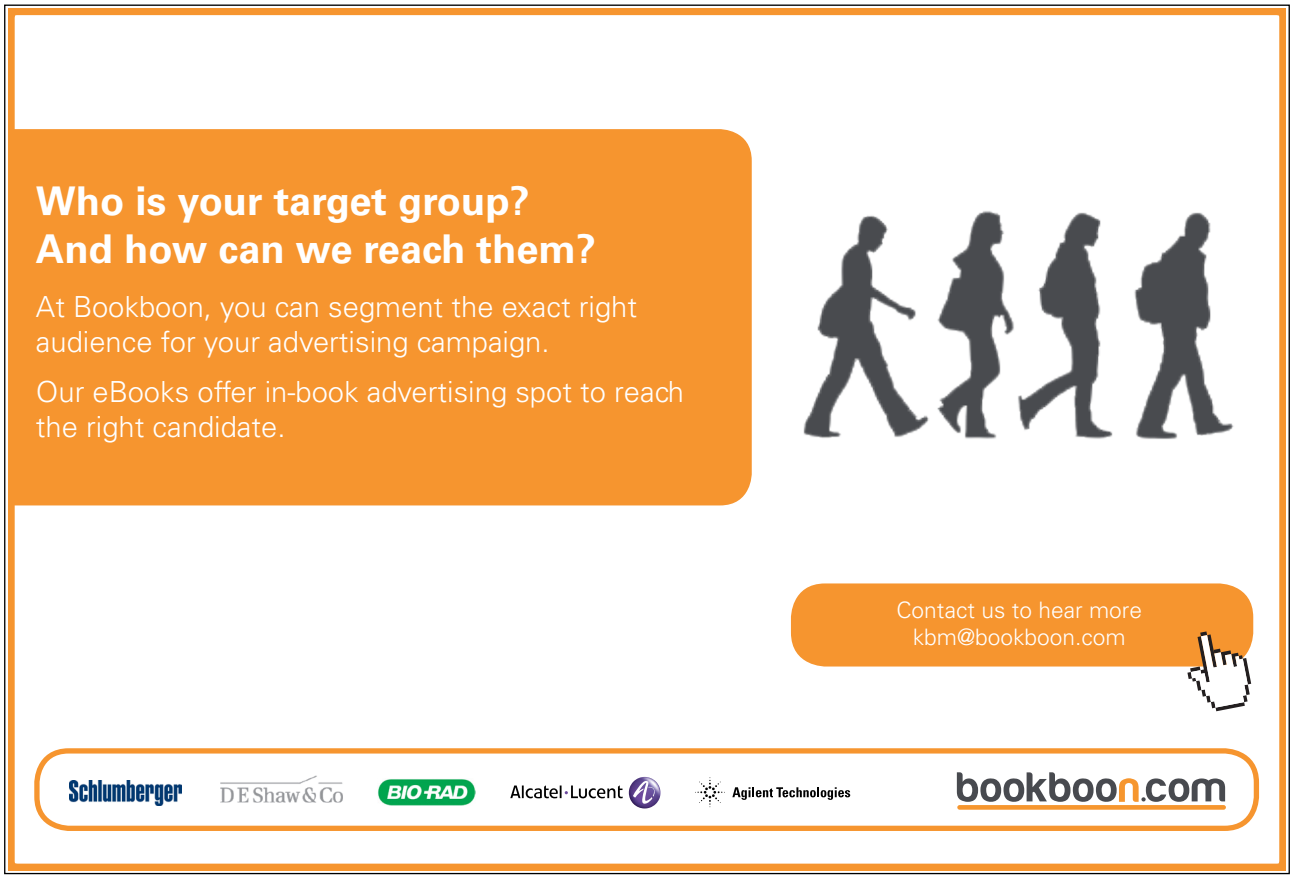
6.0.8 Important functions

Over time we have come to see that some functions crop up again and again in applications. This seems a good point to look at some of these.

polynomials

We call functions like $f(x) = a_p x^p + a_{p-1}x^{p-1} + \ldots + a_1 x + a_0$ polynomials and these usually have a domain consisting of the reals. In out example the coefficients $a_0, a_1, \ldots, a_p$ are numbers and our polynomial is said to have order p. Examples are

1. $f(x) = x + 2$
2. $f(x) = x^3 - x^2 + x + 2$
3. $f(x) = x^{17} - 11$
4. $f(x) = x^2 - 3x + 2$



Download free eBooks at bookboon.com
Click on the ad to read more

Zeros

Very often we need to know for what values of x for which $f(x) = a_p x^p + a_{p-1}x^{p-1} + \ldots + a_1 x + a_0 = 0$ is zero. The values are called the zeros or the roots of the polynomial. We can prove that a polynomial of degree p has at most p roots which helps a little. The simplest to way to find zeros is to factorize the polynomial so if

$$f(x) = x^3 - 6 * x^2 + 11x - 6 = (x-1)(x-2)(x-3)$$

so $f(x) = 0$ when $x = 1, 2, 3$.

Factorization is (as for integers) rather difficult. The best strategy is to try and guess one zero, say x=a and then divide the polynomial by (x-a). We then repeat. Polynomial division is just like long division. So to divide $x^3 - 6x^2 + 11x - 6$ by $x - 1$:

write out the sum

$$\begin{array}{r|l} x-1 & x^3 - 6x^2 + 11x - 6 \end{array}$$

find the power of x to multiply $x - 1$

$$\begin{array}{r|l} x-1 & x^2 \\ \hline & x^3 - 6x^2 + 11x - 6 \end{array}$$

multiply $x - 1$ by x^2 as shown.

$$\begin{array}{r|l} x-1 & x^2 \\ \hline & x^3 - 6x^2 + 11x - 6 \\ & -x^3 \quad + x^2 \\ \hline \end{array}$$

subtract as shown.

$$\begin{array}{r|l} x-1 & x^2 \\ \hline & x^3 - 6x^2 + 11x - 6 \\ & -x^3 \quad + x^2 \\ \hline & -5x^2 + 11x \end{array}$$

find a multiplier to multiply $x - 1$ to get a $-5x^2$

$$\begin{array}{r|l} x-1 & x^2 \quad -5x \\ \hline & x^3 - 6x^2 + 11x - 6 \\ & -x^3 \quad + x^2 \\ \hline & -5x^2 + 11x \end{array}$$

multiply $x - 1$ and subtract as shown

$$\begin{array}{r|l} x-1 & x^2 \quad -5x \\ \hline & x^3 - 6x^2 + 11x - 6 \\ & -x^3 \quad + x^2 \\ \hline & -5x^2 + 11x \\ & 5x^2 \quad -5x \\ \hline & 6x - 6 \end{array}$$

Download free eBooks at bookboon.com

$$
\begin{array}{r|rrrr}
 & & x^2 & -5x & \\
\hline
x-1 & x^3 & -6x^2 & +11x & -6 \\
 & -x^3 & +x^2 & & \\
\hline
 & & -5x^2 & +11x & \\
 & & 5x^2 & -5x & \\
\hline
 & & & 6x & -6
\end{array}
$$

find a multiplier to multiply $x-1$ to get a $6x$

$$
\begin{array}{r|rrrr}
 & & x^2 & -5x & +6 \\
\hline
x-1 & x^3 & -6x^2 & +11x & -6 \\
 & -x^3 & +x^2 & & \\
\hline
 & & -5x^2 & +11x & \\
 & & 5x^2 & -5x & \\
\hline
 & & & 6x & -6 \\
 & & & -6x & +6 \\
\hline
 & & & & 0
\end{array}
$$

nothing left so we stop!

The answer is $x^2 - 5x + 6$. If there is something left then it is the remainder.

Hence

$$
\begin{array}{r|rrr}
 & & x & -3 \\
\hline
x-2 & x^2 & -5x & +6 \\
 & -x^2 & +2x & \\
\hline
 & & -3x & +6 \\
 & & 3x & -6 \\
\hline
 & & & 0
\end{array}
$$

The answer is $x^2 - 5x + 6 = (x-2)(x-3)$.

However suppose we try

$$
\begin{array}{r|rrr}
 & & x & -4 \\
\hline
x-1 & x^2 & -5x & +6 \\
 & -x^2 & +x & \\
\hline
 & & -4x & +6 \\
 & & 4x & -4 \\
\hline
 & & & 2
\end{array}
$$

We have a remainder and the answer is $x^2 - 5x + 6 = (x-1)(x-4) + 2$.

Exercises

Factorize

1. $2x^3 - x^2 - 7x + 6$
2. $2x^3 - 3 * x^2 - 5x + 6$

Download free eBooks at bookboon.com

The power function

Suppose we take values x from the reals and consider the function $P(x) = x^a$ for some value a. We can suppose that a is also real. So we have

$$\mathcal{R} : P \to \mathcal{R}$$

An example might be $P(x) = x^2$ or $P(x) = x^{1.5}$. In the second case we clearly have to redefine the domain. Can you see why? The properties of the power function

1. $x^a \times x^b = x^{a+b}$
2. $x^0 = 1$

Logarithms

We know that we can write powers of numbers, so

$$10^0 = 1 \quad 10^1 = 2 \qquad 10^2 = 100 \qquad 10^2 = 1000 \qquad \ldots$$

and $10^{0.5} = 3.162278\ldots$.

Now consider the backwards problem:

> Given y can we find an x such that $y = 10^x$.

In other words if we define the power function $y = P(x) = 10^x$ for $x \in \mathcal{R}$, as above, then what is the inverse of this $P^{-1}(y)$? It may help to look at figure 6.10. We have plotted dotted lines from (1.5,0) to the curve. Going from x vertically to the curve and then to the y axis gives the power value $P(x) = y$. The reverse path from y to x is the logarithm.

Download free eBooks at bookboon.com

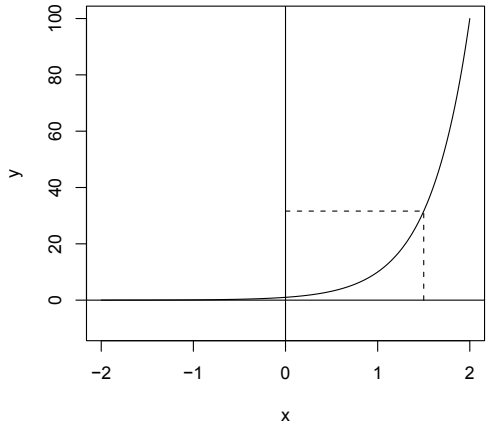


Figure 6.10: Plot of $f(x) = 10^x$

The inverse of $p(x)$ is call the logarithm or log and is written $\log_{10}(x)$. So

$$\log_{10}(1) = 0 \quad \log_{10}(10) = 1 \quad \log_{10}(100) = 0 \quad \log_{1000}(1) = \ldots$$

Often we are lazy and drop the 10 and just write $\log(x)$

Because we know that log is the inverse of the power function we have some useful rules

1. $\log(u) + \log(v) = \log(uv)$
2. $\log(u^v) = v\log(uv)$
3. $\log(u - \log(v) = \log\left(\frac{u}{v}\right)$
4. $-\log(u) = \log\left(\frac{1}{u}\right)$

Of course we did not have to choose 10 in our definitions. We could have choose 2, like many engineers, or any positive number a say. We then write $y = \log_a(x)$ to indicate the number y which satisfies $x = a^y$. The $\log_a(x)$ is called the log of x to *base* a.

For reasons which will (we hope) become apparent mathematicians like to use *natural logs* which have a base $e = 2.718282\ldots$. because they are used so often rather than write $\log_e(x)$ you will often see them written as $\ln(x)$ or just as $\log(x)$. All logs satisfy the rules set out in the list 6.0.8. We shall be lazy and just use logarithms to base e.

Download free eBooks at bookboon.com

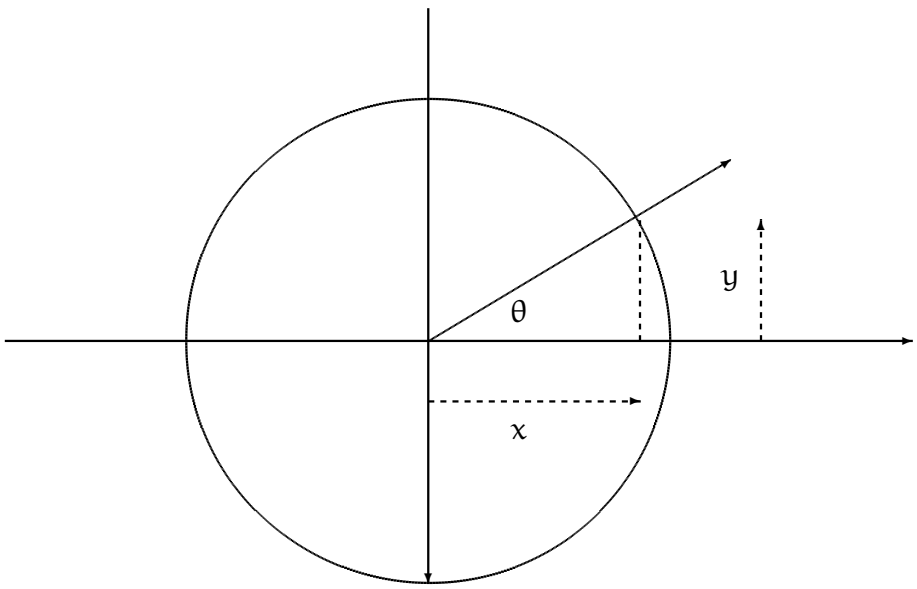
We can of course express logs in one base as logs in another. Suppose $x = a^{\log_a(x)} = b^{\log_b(x)}$ then taking logs gives

$$\log_a(x) = \log_a(b)\log_b(x)$$

Sometime it is natural to express powers as base 2 for example $y = P(x) = 2^x$. Mathematicians often use the number e so the power definition is $y = e^x$ which you will often see written as $y = \exp(x)$ since e^x is called the *exponential function*.

6.1 Functions and angular measure

We look briefly at the measurement of angles. Angular measure has been important from the very beginning of human history both in astronomy and navigation. Consider a circle with the angle θ made with the x axis as shown. Unlike maps in mathematics the reference line is *not* North but along the x axis and if we rotate anti-clockwise we sweep out an angle θ. The angle is traditionally measured in degrees, minutes and seconds. We will stick to degrees for the moment.



If we sweep anti-clockwise through 360 degrees we sweep out a circle. 180 degrees is a half circle and $720 = 3 \times 360$ two circles. Rotations in a clockwise direction are assumed to be negative degrees, so $-90^o = 270^o$

To complicate things a little we can also measure the angle in an equivalent way by measuring the length of the arc we make out on the circle as we sweep through the angle θ. Suppose this is s. For a circle of radius 1 s is a measure of the angle, although in different units called *radians*. So one circle is 2π radians and 90° is $\pi/2$ radians. We convert from degrees to radians as follows

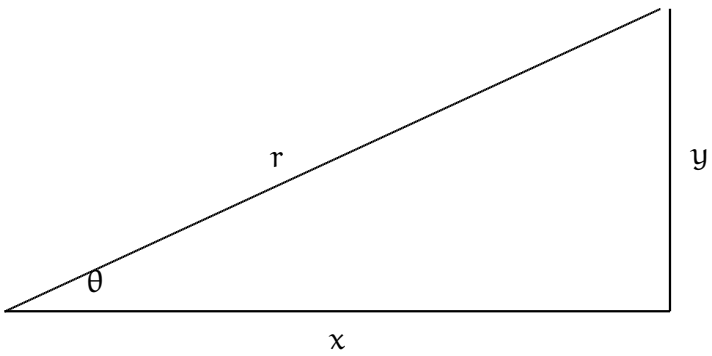
Download free eBooks at bookboon.com

degrees	radians
θ	$2\pi\theta/360$
$360s/(2\pi)$	s

If you look at most "scientific calculators" you will see a button for switching from degrees to radians and vice versa.

The trigonometric functions

Of course we can measure angles in other ways. Suppose we look at the angle θ in the diagram. The ratio of the y and x values is related to the angle. Roman surveyors would often choose and angle by fixing the x value and the y value. As you can. imagine, five steps and then 3 steps vertically gives the same angle no matter where you are



Thus from the diagram θ is related to y/x. In fact we define y/x to be the tangent of θ written as $\tan\theta = y/x$. The inverse function is $\tan^{-1}\theta = y/x$ or sometimes $\arctan\theta = y/x$ The reader might like to examine our triable and see why the tangent of 90° does not exist. We provide a plot of the tangent from 0 to just under 90 degrees in figure 6.11. If we keep the definition on the domain $0 \leq \theta < 90$ as is (relatively) simple. While the domain is easily extended we leave this to those of you will interests in this direction.

Of course we do not have to use tangents, although they are probably the most practical in applications. Alternative are to use the ratio y/r the height y divided by the radius of the circle r. This is called the sine function and written $\sin\theta = y/x$.

In a similar we we could use the cosine written $\cos\theta = x/r$. Both of these functions are plotted in figure 6.12 There are lots of links between these functions,

Download free eBooks at bookboon.com

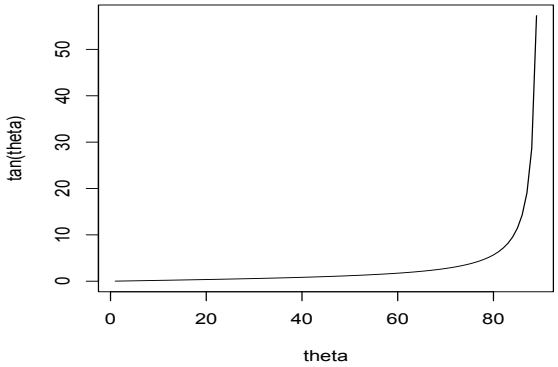


Figure 6.11: $\tan x$

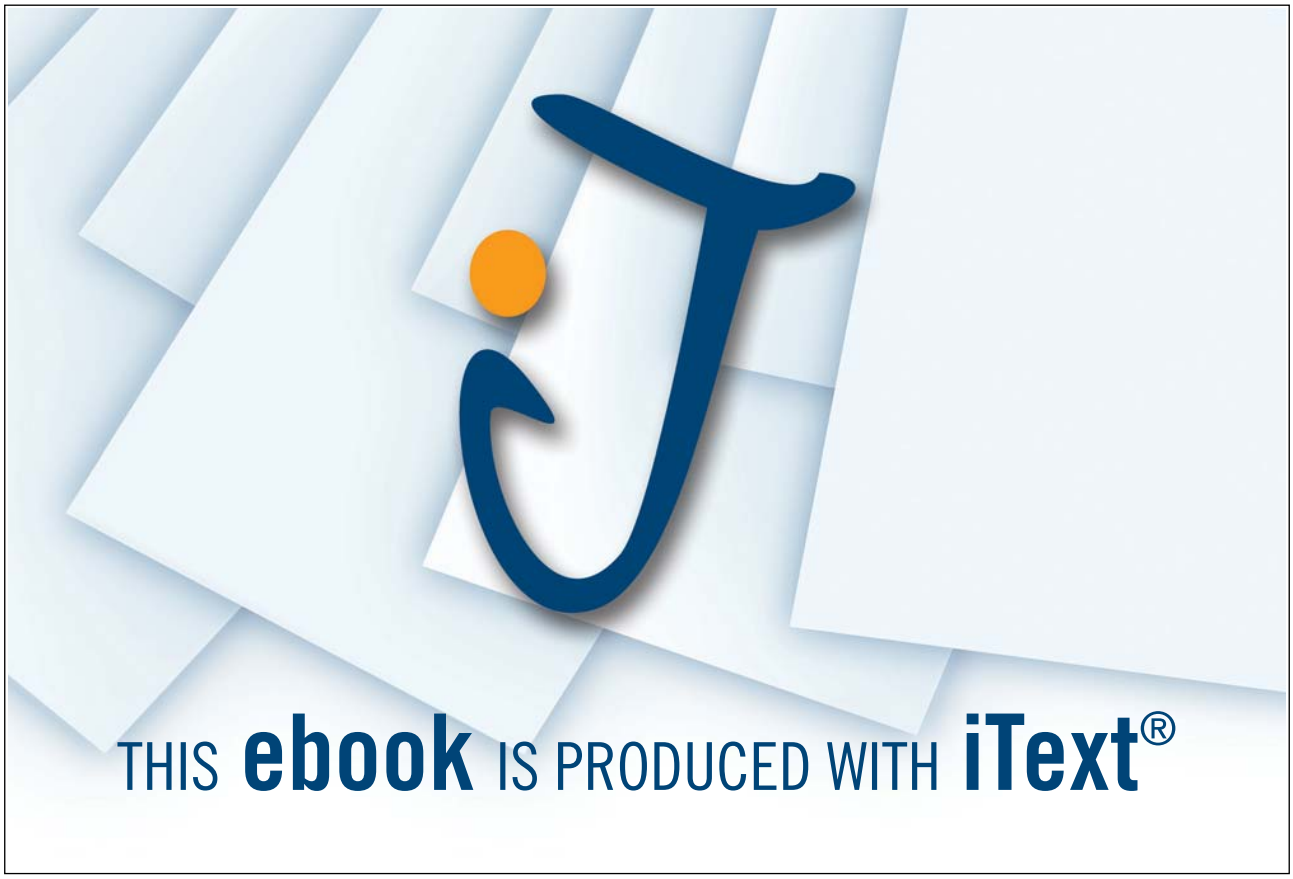
for example

$$\tan\theta = \frac{\sin\theta}{\cos\theta}$$

This can be deduced quite simple from the definitions. Try it yourself!

The trigonometric functions are periodic in that if we plot them over a large part of the axis they repeat as in figure 6.13

Out next step is the study of the shapes of functions which brings us to Calculus.



Download free eBooks at bookboon.com

Click on the ad to read more

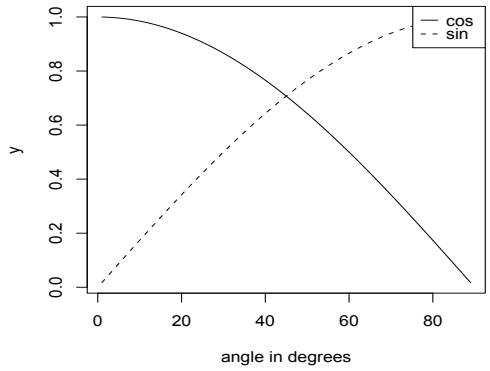


Figure 6.12: $\tan x$

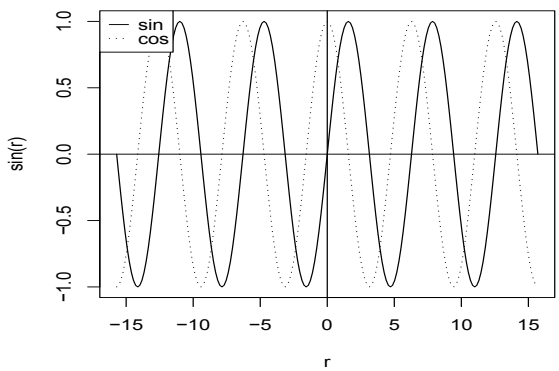


Figure 6.13: Plot of sin and cos

Download free eBooks at bookboon.com

Chapter 7

Sequences

Reason's last step is the recognition that there are an infinite number of things which are beyond it. Pascal

We write a sequence $a_1, a_2, a_3, \cdots, a_n, \cdots$ as $\{a_n\}$ and our interest is normally whether the sequence tends to a limit A written

- $a_n \to A$ as $n \to \infty$.
- or $\lim_{n\to\infty} a_n = A$

However there are many interesting sequences where limits are not the main interest. For example the Fibonacci sequence. In Fibonacci's Liber Abaci (1202) poses the following problem

> How Many Pairs of Rabbits Are Created by One Pair in One Year:
> A certain man had one pair of rabbits together in a certain enclosed place, and one wishes to know how many are created from the pair in one year when it is the nature of them in a single month to bear another pair, and in the second month those born to bear also.

The resulting sequence is

$$1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, \ldots$$

and each term is the sum of the previous two terms. An interesting aside is that the nth Fibonacci number $F(n)$ can we written as

$$F(n) = [\phi^n - (1-\phi)^n]/\sqrt{5} \text{ where } \phi = (1+\sqrt{5})/2 \simeq 1.618\ldots$$

which is a surprise since $F(n)$ is an integer and the formula contains $\sqrt{5}$. For lots more on sequences see

http://www.research.att.com/ njas/sequences/

Download free eBooks at bookboon.com

7.0.1 Limits of sequences

We turn our attention to the behaviour of sequences such as $\{a_n\}$ as n becomes very large.

1. A sequence may approach a finite value A. We say that it tends to a limit, so for example we write

$$1, \left(\frac{1}{2}\right), \left(\frac{1}{2}\right)^2, \left(\frac{1}{2}\right)^3 \ldots \left(\frac{1}{2}\right)^n, \ldots$$

or

$$1.0000 \quad 0.5000 \quad 0.2500 \quad 0.1250 \quad 0.0625 \quad 0.0312 \quad 0.0156 \quad 0.0078 \quad 0.0039 \quad 0.0020 \ldots$$

as

$$\left\{\left(\frac{1}{2}\right)^n\right\}$$

and we shall see that

$$\left\{\left(\frac{1}{2}\right)^n\right\} \to 0 \text{ as } n \to \infty$$

2. If a sequence does not converge it may go to $\pm\infty$, that is keep increasing or decreasing.

$$1 \quad 2 \quad 4 \quad 8 \quad 16 \quad 32 \quad 64 \quad 128 \quad 256 \quad 512 \quad 1024 \ldots$$

Informally $\{2^n\} \to \infty$ as $\to \infty$.

3. A sequence may just oscillate

$$1 \quad -1 \quad 1 \quad -1 \quad 1 \quad -1 \quad 1 \quad -1 \quad 1 \quad -1 \quad 1$$

Limit

We need a definition of a limit and after 2000 years of trying we use :
$\{a_n\} \to A$ as $\to \infty$ if and only if, given any number ϵ there is an N such that for $n \geq N \quad |a_n - A| < \epsilon$.

In essence I give you a guarantee that I can get as close as you wish to a limit (if it exists) for all members of the sequence with sufficiently large N, that is after N all the values of the sequence satisfy $\mid a_n - A \mid < \epsilon$. The idea is that if there is a limit then if you give me some tolerance, here ϵ, I can guarantee that for some point in the sequence all the terms beyond that *all* lie within ϵ of the limit.

Download free eBooks at bookboon.com

Examples

- $\{\frac{1}{n}\} \to 0$.
- $\{x^n\} \to 0$ for $|x| < 1$.
- We argue as follows:
  Suppose you give me a (small) value for ϵ. I can then choose a value N where $N > 1/\epsilon$. We can do this as, for $|x| < 1$

  $$\ldots |x|^4 < |x|^3 < |x|^2 < |x|$$

  It then follows that as $N > 1/\epsilon$ then $\epsilon > 1/N$. But if $n > N$ then $1/n < 1/N$ so we can say:
  if we choose $N > 1/\epsilon$ the when $N > n$ $\quad |1/n - 0| < \epsilon$ and so $1/n \to 0$
- We argue as follows:
  Suppose you give me a (small) value for ϵ. I can then choose a value N where $|x|^N < \epsilon$. Or $N \log |x| < \log \epsilon$. Rearranging

  $$N > \frac{\log \epsilon}{\log |x|} \text{ beware the signs!}$$

  But if $\log |x| < 1$ then

  $$\log |x|^2 < \log |x|, \log |x|^3 < \log |x|^2, \ldots \log |x|^n < \log |x|^{n+1}$$

  So we choose $N > \log \epsilon / \log |x|$ then when $N > n$ $\quad |x^n| = |x^n - 0| < \epsilon$ and so $|x^n| \to 0$

HIT YOUR EMPLOYEE RETENTION TARGETS

We help talent and learning & development teams hit their employee retention & development targets by improving the quality and focus of managers' coaching conversations.

Start improving employee retention & performance now.
Get your FREE reports and analysis on 10 of your staff today.

GET MY REPORTS

Download free eBooks at bookboon.com

Click on the ad to read more

Rules

Manipulating expressions like $\mid a_n - a \mid$ can be tricky so it is easier to develop some rules. Using these is very much easier as we shall see.

If $\{a_n\}$ and $\{b_n\}$ are two sequences and $\{a_n\} \to A$ while $\{b_n\} \to B$ then

- $\{a_n \pm b_n\} \to A \pm B$
- $\{a_n b_n\} \to AB$
- $\{a_n/b_n\} \to A/B$ provided B is nonzero as are the $\{b_n\}$.
- For a constant c we have $\{ca_n\} \to cA$

also

- If $\{a_n\} \to \pm\infty$ then $\{1/a_n\} \to 0$
- If $\{a_n\} \to \pm\infty$ while $\{b_n\} \to B$ (finite B) then$\{a_n + b_n\} \to \pm\infty$
- If $\{a_n\} \to \infty$ while $\{b_n\} \to B$ (finite B) then$\{a_n b_n\} \to \pm\infty$ depending on the sign of B.

We can look at rational functions as follows

1. $$\left\{\frac{n+1}{n+13}\right\} = \left\{\frac{1+1/n}{1+13/n}\right\} \to \left\{\frac{1+0}{1+0}\right\} \to 1$$
2. $$\left\{\frac{n^2-3n+11}{n^4+13n^2-n+43}\right\} = \left\{\frac{1/n^2-3/n^3+11/n^4}{1+13/n^2-1/n^3+43/n^4}\right\} \to \left\{\frac{0-0+0}{1+0-0+0}\right\} \to 0/1 \to 0$$
3. $$\left\{\frac{n+1}{n+13x^n}\right\} = \left\{\frac{1+1/n}{1+13x^n/n}\right\} \to 1/1 \to 1 \qquad \mid x \mid < 1$$
4. $$\left\{\frac{3^n+1}{4^n+13}\right\} = \left\{\frac{(3/4)^n+1/4^n}{1+13(1/4)^n}\right\} \to 0/1 = 0 \to 1$$

Subsequence

A subsequence of a sequence $\{a_n\}$ is an infinite succession of its terms picked out in any way. Note that if the original series converges to A so does any subsequence. If $a_{n+1} \geq a_n$ we say the subsequence is increasing while if $a_{n+1} \leq a_n$ we say the subsequence is decreasing. Increasing or decreasing sequences are sometimes called monatonic.

Download free eBooks at bookboon.com

Bounded

If an increasing sequence is bounded above then it must converge to a limit. Similarly If an decreasing sequence is bounded below then it must converge to a limit.

7.1 Series

A series is the sum of terms of a sequence written

$$u_1 + u_2 + u_3 + \cdots + u_N = \sum_{i=1}^{N} u_i$$

We use capital sigma (Σ) for sums and by

$$\sum_{i=a}^{b} u_i$$

we mean the sum of terms like u_i for i taking the values a to b. Of course there are many series we sum, for example we have met the Binomial series and we have the following useful results.

Download free eBooks at bookboon.com
Click on the ad to read more

- $1+2+3+4+\cdots+N=\sum_{i=1}^{N}=N(N+1)/2$
- $1^2+2^2+3^2+4^2+\cdots+N^2=\sum_{i=1}^{N}i^2=N(2N+1)(N+1)/6$
- $1^3+2^3+3^3+4^3+\cdots+N^3=\sum_{i=1}^{N}i^3=[N(N+1)/2]^2$
- $1\frac{1}{2}+\frac{1}{2}\frac{1}{3}+\cdots+\frac{1}{N}\frac{1}{N+1}=\sum_{i=1}^{N}\left(\frac{1}{i(i+1)}\right)=1-\frac{1}{N+1}$
- $1+x+x^2+x^3+\cdots+x^N=\sum_{i=0}^{N}x^i=\left(1-x^{N+1}\right)/(1-x)$

7.1.1 Infinite series

If we want the sum of the infinite series $\sum_{i=1}^{\infty}u_i$ -if such a thing exists - we need to be clear we mean. *Assume that all the terms in the series are non-negative, that is* $0\leq u_i$. Consider the partial sums

$$S_1=u_1$$

$$S_2=u_1+u_2$$

$$S_3=u_1+u_2+u_3$$

$$S_4=u_1+u_2+u_3+u_4$$

$$\cdots$$

$$S_N=u_1+u_2+u_3+\cdots+u_N$$

$$\cdots$$

If the sequence $\{S_n\}$ converges to a limit S then we say that the series $\sum_{i=1}^{\infty}u_i$ is convergent and the sum is S. Otherwise we say the series diverges or is divergent.

Examples

- $\sum_{n=1}^{\infty}\frac{1}{n}$ is divergent.
  We can argue: Let

$$S_4=1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}=1+\frac{1}{2}+\left(\frac{1}{3}+\frac{1}{4}\right)>1+\frac{1}{2}+\frac{1}{2}>2$$

  and

$$S_8=1+\frac{1}{2}+\left(\frac{1}{3}+\frac{1}{4}\right)+\left(\frac{1}{5}+\frac{1}{6}+\frac{1}{7}+\frac{1}{8}\right)>1+\frac{1}{2}+\frac{1}{2}+\frac{1}{2}>3/2$$

Download free eBooks at bookboon.com

$$S_{16} = 1+\frac{1}{2}+\left(\frac{1}{3}+\frac{1}{4}\right)+\left(\frac{1}{5}+\frac{1}{6}+\frac{1}{7}+\frac{1}{8}\right)+\left(\frac{1}{9}+\frac{1}{10}+\frac{1}{11}+\frac{1}{12}+\frac{1}{13}+\frac{1}{14}+\frac{1}{15}+\frac{1}{16}\right)$$

$$> 1+\frac{1}{2}+\frac{1}{2}+\frac{1}{2}+\frac{1}{2} > 6/2$$

In general we can (with care show)

$$S_{2^k} > \frac{k}{2}+1$$

So we can make the partial sums of 2^k terms as large as we like and they are increasing and unbounded. Thus the series must be divergent.

This has an important consequence if $u_n \to 0$ **it does not mean that the sum is convergent**. It may be but it may not be!

- $\sum_{n=0}^{\infty} x^n$ is convergent for $|x| < 1$ and the sum is $1/(1-x)$. When $|x| > 1$ the series is divergent.

  We can argue that

  $$\sum_{n=0}^{N} x^n = \frac{1-x^{N-1}}{1-x} \to 1/(1-x)$$

  and since we have an explicit form for the sum the result follows.

- $\sum_{n=1}^{\infty} \frac{1}{n(n+1)}$ converges and the sum is 1 since

  $$\sum_{n=1}^{N} \frac{1}{n(n+1)} - \sum_{n=1}^{N}\left(\frac{1}{n}-\frac{1}{(n+1)}\right) = 1-\frac{1}{N+1}$$

- $\sum_{n=1}^{\infty} \frac{1}{n^\alpha}$ is divergent for $\alpha \geq 1$ and convergent otherwise.

Some Rules for series of positive terms

- If $\sum_{n=1}^{\infty} u_n$ and $\sum_{n=1}^{\infty} v_n$ are both convergent with sums S and T then $\sum_{n=1}^{\infty}(u_n \pm v_n)$ converges to $S \pm T$.

- If $\sum_{n=1}^{\infty} u_n$ converges then adding or subtracting a finite number of terms does not affect convergence, it will however affect the sum.

Download free eBooks at bookboon.com

- If u_n does not converge to zero then $\sum_{n=1}^{\infty} u_n$ does not converge.
- The comparison test: If $\sum_{n=1}^{\infty} u_n$ and $\sum_{n=1}^{\infty} v_n$ are two series of positive terms and if $\{u_n/v_n\}$ tends to a non zero finite limit R then the series either both converge or both diverge.
- The Ratio test: If $\sum_{n=1}^{\infty} u_n$ is a series of positive terms and suppose $\{u_{n+1}/u_n\} \to L$ then
  - If $L < 1$ the series converges.
  - If $L > 1$ the series diverges.
  - If $L = 1$ the question is unresolved.
- The integral test: Suppose we have $\sum_{n=1}^{\infty} u_n$ and $f(n) = u_n$ for some function f which satisfies

  1. $f(x)$ is decreasing as x increases.
  2. $f(x) > 0$ for $x \geq 1$

  Then

  1. $0 < \sum_{n=1}^{N} u_n - \int_1^{N+1} f(x)dx < f(1)$
  2. The sum converges if the integral $\int_1^{\infty} f(x)dx$ is finite and diverges if $\int_1^{\infty} f(x)dx$ is infinite.

Absolute Convergence

We say that $\sum_{n=1}^{\infty} u_n$ is *absolutely convergent* if $\sum_{n=1}^{\infty} \mid u_n \mid$ converges. If $\sum_{n=1}^{\infty} |u_n|$ does not converge but $\sum_{n=1}^{\infty} u_n$ does then we say the series is *conditionally convergent*. The nice thing about absolutely convergent series is we can rearrange the terms without affecting the convergence or the sum.

Alternating sign test

On simple test for non conditionally convergent series is the alternating sign test. Suppose we have a decreasing sequence of positive terms $\{u_n\}$ and let

$$S = u_1 - u_2 + u_3 - u_4 + \ldots + (-1)^n u_n \ldots$$

Then S converges. For example

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} \ldots$$

Download free eBooks at bookboon.com

Power series

A series of the form

$$S = a_0 + a_1x + a_1x + a_2x^2 + a_3x^3 \ldots = \sum_{n=0}^{\infty} a_n x^n$$

is called a *power series* Many power series only converge for values of x which satisfy $\mid x \mid < R$ for some value R. This value is called the *radius of convergence.* We can usually rind R using the ratio test, for example

$$S = 1 + \left(\frac{x}{5}\right) + \left(\frac{x}{5}\right)^2 + \left(\frac{x}{5}\right)^3 + \left(\frac{x}{5}\right)^4 \ldots$$

Then

$$\mid u_{n+1}/u_n \mid = \left| \left(\frac{x}{5}\right)^{n+1} / \left(\frac{x}{5}\right)^n \right| = \left| \left(\frac{x}{5}\right) \right|$$

for this to be less than 1 we need $\mid x \mid < 5$ You can then check $x \pm 5$ separately.

Exercises

1. Write down the first five terms of each of the sequences defined below
   (a) $a_n = 1 - (0.2)^n$
   (b) $a_n = 1 - (-0.2)^n$
   (c) $a_n = (n^2 + 1)/(n + 1)$.
   (d) $a_n = 3/a_{n-1} \qquad a_1 = -1$

2. Graph the sequences in question 1.

3. Decide which of the following sequences converges and find the limit if it exists.
   (a) $2 - (0.2)^n$
   (b) $2 - (-0.2)^n$
   (c) $(n + 1)/(n^2 + 1)$
   (d) $(4 + n)/(3n - 2)$
   (e) $(4 + n)$
   (f) $(n^2 - n + 2)/(5n^2 + 4n + 1)$
   (g) $2^n - \left(-\frac{1}{2}\right)$

Download free eBooks at bookboon.com

4. How large must n be for $(1/3)^n$ to be less that
   (a) 0.01
   (b) 10^{-6}
5. Find a number N such that $n^2/2^n \leq 0.001$ if $n > N$.
6. Suppose $a_n = x^{1/n} \qquad x > 1$
   (a) Show that the sequence is decreasing.
   (b) Show that the sequence is bounded below.
   (c) Is the sequence convergent?
7. Show that
$$1 + 3 + 5 + \cdots + (2N - 1) = N^2$$
8. Find
$$\sum_{n=1}^{N} \frac{1}{(n+1)(n+2)}$$
9. Decide which of the following sums are convergent.
   (a) $\sum_{n=1}^{\infty} 1/(2n-1)$
   (b) $\sum_{n=1}^{\infty} 2/(n^2+3)$
   (c) $\sum_{n=1}^{\infty} 1/\sqrt{2n-1}$

Download free eBooks at bookboon.com

Chapter 8

Calculus

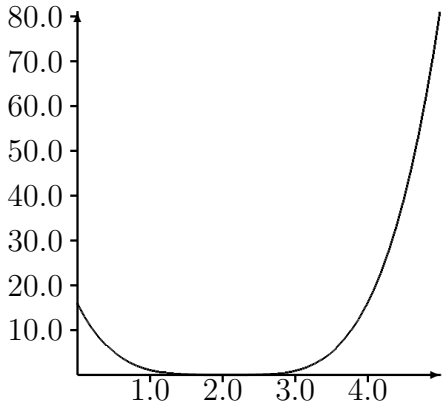
I'm very good at integral and differential calculus, I know the scientific names of beings animalculous; In short, in matters vegetable, animal, and mineral, I am the very model of a modern Major-General.
The Pirates of Penzance. Act 1.

We have looked at limits of sequences, now I want to look at limits of functions. Suppose we have a function $f(x)$ defined on an interval $a \leq x \leq b$. I have a sequence $x_1, x_2, \cdots, x_n$ which tends to a limit x_0. Can I say that the sequence $f(x_1), f(x_2, \ldots, f(x_n)$ tends to ℓ and what do I mean? We normally define the limit as follows:

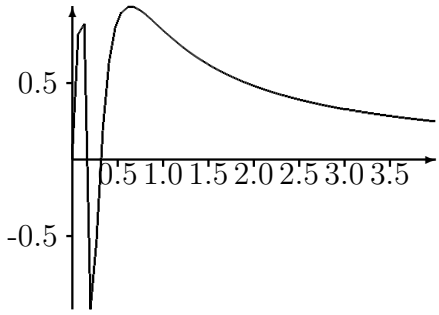
> We say that $f(x) \to f(x_0)$ as $x \to x_0$ if for any $\epsilon > 0$ there is a value $\delta > 0$ such that $\mid x - x_0 \mid < \delta \Rightarrow \mid f(x) - \ell \mid < \epsilon$

This is in the same spirit as our previous definition for sequences. We can be as close as we wish to the limiting value ℓ.

For example $(x-2)^4 \to 0$ as $x \to 2$. If you given me an $0 < \epsilon < 1$ then if $\mid x - 2 \mid \leq \delta$ we know $\mid (x-2)^4 - 0 \mid \leq \delta^4$. So provided $\delta \leq \epsilon$ we have a limit as $x \to 0$!



Download free eBooks at bookboon.com

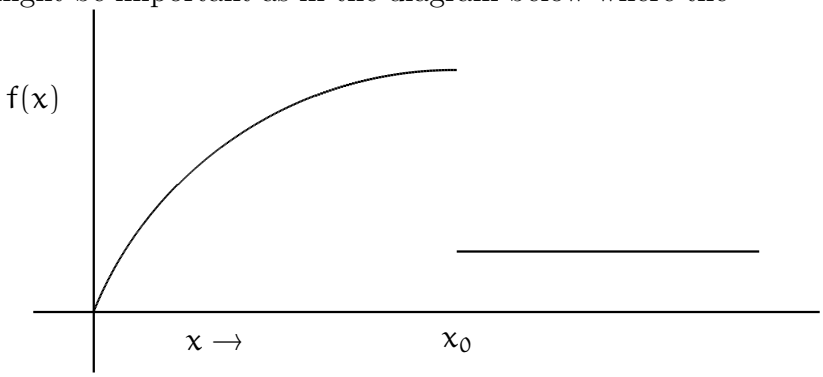


In the second case we plot $\sin(1/x)$. This starts to oscillate faster and faster as it approaches zero and (it is not quite simple to show) does not have a limit.

Download free eBooks at bookboon.com
Click on the ad to read more

8.0.2 Continuity and Differentiability

We did not specify which direction we used to approach the limiting value, from above or from below. This might be important as in the diagram below where the

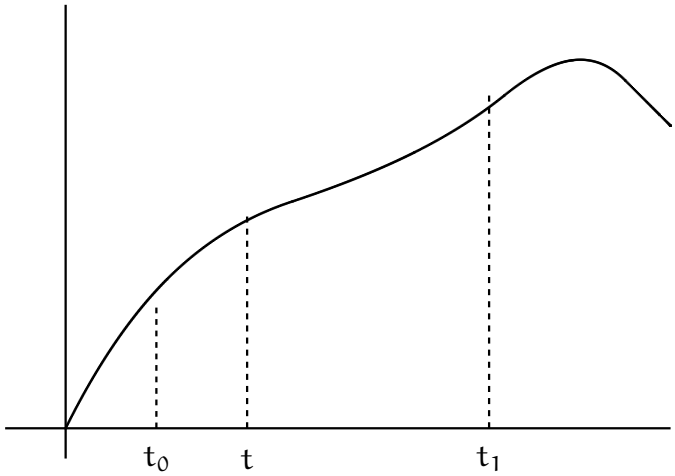


function has a *jump* at x_0.

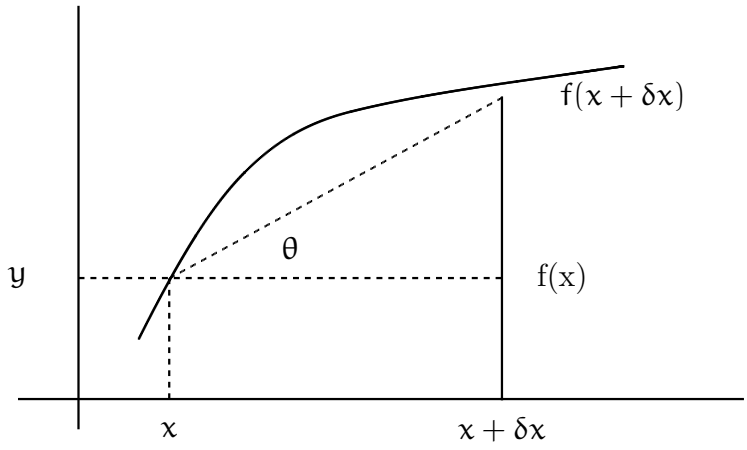
We like *continuous* functions, these are functions where $f(x) \to f(x_0)$ as $x \to x_0$ from above and below. You can think of these as functions you can draw without lifting your pencil off the page. Continuous functions have lots of nice properties.

If we have a continuous function we might reasonably look at the *slope* of the curve at any point. This may have a real physical meaning. So suppose we have the track of a car. We might plot the distance it travels, East say, against time.

If the difference between the distance at times t_0 and t_1 is D then $D/(t_1 - t_0)$ gives the approximate speed. This is just the procedure followed by average speed cameras on roads! However what we have observed is an average speed. If we want an estimate of speed at a particular time t we need t_0 and t_1 to approach t.



Download free eBooks at bookboon.com



If we take the times to be t and $t+\delta t$, where δt means a small extra bit of t, then we want

$$\frac{f(t+\delta t)-f(t)}{(t+\delta t-t)}$$

as δt becomes small or more explicitly

$$\frac{f(t+\delta t)-f(t)}{\delta t} \text{ as } t \to 0$$

This limit gives the *derivative* which is the slope of the curve $f(t)$ at the point t and is written $f'(t)$ or

$$\frac{df}{dx} = \lim_{\delta t \to 0} \frac{f(t+\delta t)-f(t)}{\delta t} \tag{8.1}$$

Suppose we take $y = f(t) = 3-4t$, a line with constant negative slope. Using the equation 8.1 we have

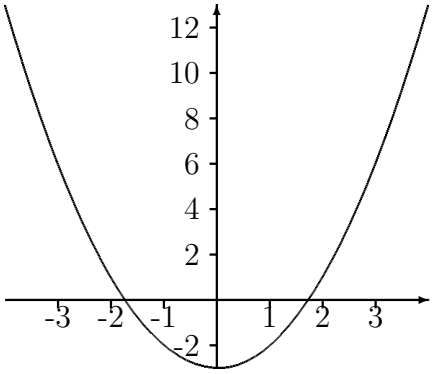
$$\frac{df}{dx} = \lim_{\delta t \to 0} \frac{3-4(t+\delta t)-3+4t}{\delta t} = \frac{-4\delta t}{\delta t} = -4$$

If we now have $y = x^2 - 3$ we have, writing x for t

$$\frac{df}{dx} = \lim_{\delta t \to 0} \frac{(x+\delta x)^2-3-x^2+3}{\delta x} = \frac{x^2+2x\delta x+(\delta x)^2-3-x^2+3}{\delta x} = \frac{2x\delta x+(\delta x)^2}{\delta x} = 2x+\delta x =$$

So at x=2 the slope is zero while when x is negative the slope is down and then is upwards when x is greater that zero. You might find it useful to consider the plot. Note that if we take a point on a curve and draw a straight line whose slope is $f'(x)$ this line is known as the *tangent* at x.

Download free eBooks at bookboon.com



Of course life is too short for working out the derivatives dy/dx like this from first principles so we tend to use rules (derived from first principles).

1. $\frac{d}{dx}[af(x)] = a\frac{df}{dx}$ where a is a constant.
2. $\frac{d}{dx}[f(x) + g(x)] = \frac{df}{dx}\frac{dg}{dx}$
3. $\frac{d}{dx}[f(x)g(x)] = f(x)\frac{dg}{dx} + g(x)\frac{df}{dx}$
4. $\frac{d}{dx}\frac{1}{f(x)} = -\frac{1}{f^2(x)}\frac{df}{dx}$
5. $\frac{d}{dx}[x^n] = nx^{n-1}$ when $n \neq 0$ and zero otherwise.
6. $\frac{df(g(x))}{dx} = f'(g(x))g'(x)$ using $'$ for the derivative.

Download free eBooks at bookboon.com

This set of rules makes like very easy, so

$$\frac{d}{dx}(3x^2 - 11x + 59) = 3 \times 2x - 11$$

$$\frac{d}{dx}\frac{1}{(3x^2 - 11x + 59)} = -\frac{6x - 11}{(3x^2 - 11x + 59)^2}$$

$$\frac{d}{dx}(3x^2 - 11x + 59)(x - 1) = (6x - 11)(x - 1) + (3x^2 - 11x + 59)(1)$$

Example

Suppose we would like to show that $\sin x \leq x$ for $0 \leq x \leq \pi/2$. We know that when $x = 0 \qquad x = \sin x = 0$. But

$$\frac{dx}{dx} = 1 \text{ and } \frac{d \sin x}{dx} = \cos x$$

Since $\cos x \leq 1$ in the interval it implies that $\sin x$ grows more slowly than x and the result follows.

Once we move away from polynomials life gets a little more complex. In reality you need to know the derivative to be able to proceed so you need a list such as in table 8.1. Note that the derivative of $\exp(x)$ is just $\exp(x)$. So for example

Table 8.1: Table of derivatives: all logs are base e and a is a constant

Function	Derivative
$\exp(ax)$	$a \exp(ax)$
a^x	$a^x \log(a)$
$\log(ax)$	$\frac{1}{x}$
x^x	$x^x(1 + \log x)$
$\sin(ax)$	$a \cos(x)$
$\cos(ax)$	$-a \sin(x)$
$\tan(ax)$	$\frac{a}{\cos^2(x)}$

- If $y = \exp(-x^2)$ then $\frac{d \exp(-x^2)}{dx} = \exp(-x^2)(-2x)$
- If $y = \log(3x^2 - 4x + 1)$ then $\frac{d \log(3x^2 - 4x + 1)}{dx} = \frac{6x - 4}{(3x^2 - 4x + 1)}$

It is important to remember that the formulas only work for logarithms to base e and trigonometric functions, sin, cos etc expressed in *radians.*

Download free eBooks at bookboon.com

higher derivatives

Since $\frac{dy}{dx}$ is a function we might wish to differentiate it again to get $\frac{d\left[\frac{dy}{dx}\right]}{dx}$ called the second derivative and written $\frac{d^2y}{dx^2}$. If we differentiate 4 times we write $\frac{d^4y}{dx^4}$ and in general

$$\frac{d^ny}{dx^n} \qquad n = 2, 3, 4, \ldots$$

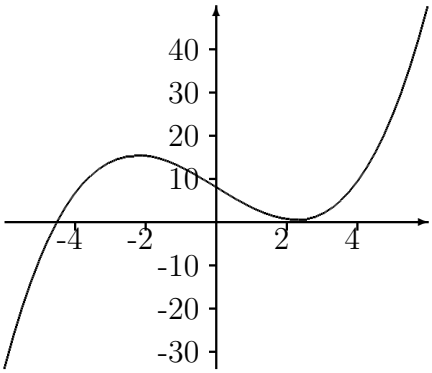
So if $y = \log(x)$ we have

$$\frac{dy}{dx} = \frac{1}{x} \quad \frac{d^2y}{dx^2} = -\frac{1}{x^2} \quad \frac{d^3y}{dx^3} = \frac{2}{x^3} \quad \frac{d^4y}{dx^4} = -\frac{6}{x^4} \ldots$$

Maxima and minima

One common use for the derivative is to find the maximum or minimum of a function. It is easy to see that if we have a maximum or minimum of a function then the derivative is zero. Consider $y = \frac{1}{3}x^3 + \frac{1}{2}x^2 - 6x + 8$

$$f(x) = \tfrac{1}{3}x^3 + \tfrac{1}{2}x^2 - 6x + 8$$



We compute $\frac{df}{dx} = x^2 + x - 6$ which is zero when $x^2 + x - 6 = (x+3)(x-2) = 0$ or $x = -3$ and $x = 2$ and from the plot it we see that we have found the turning points of the function. These are the local maxima and minima.

However when we step back and look at the whole picture it is possible to we have a stationary point i.e. $\frac{df}{dx} = 0$ which is not a turning point and hence we need a local max or minimum rule:

Download free eBooks at bookboon.com

$\frac{dy}{dx} = 0$	
$\frac{dy}{dx} < 0$ for $x < x_0$ $\frac{dy}{dx} > 0$ for $x > x_0$	$\frac{dy}{dx} > 0$ for $x < x_0$ $\frac{dy}{dx} < 0$ for $x > x_0$
x_0 is a minimum	x_0 is a maximum
$\frac{d^2y}{dx^2} > 0$	$\frac{d^2y}{dx^2} < 0$

1. The function $f(x) = \frac{1}{3}x^3 + \frac{1}{2}x^2 - 6x + 8$ has derivative $\frac{dy}{dx} = x^2 + x - 6$ so at $x = 2$ we have $\frac{dy}{dx} = 0$. When $x < 2$ the derivative is negative while when $x > 2$ it is positive so we have a minimum.

2. Or perhaps simpler $\frac{d^2y}{dx^2} = 2x + 1 > 0$ at $x = 2$ so we have a minimum.

3. When $x = -3$ again $\frac{dy}{dx} = 0$. For $x < -3 \frac{dy}{dx} > 0$ while when $x > -3 \frac{dy}{dx} < 0$ implying a maximum.

4. Again for simplicity $\frac{d^2y}{dx^2} = 2x + 1 < 0$ at $x = -3$ hence we have a maximum.

Download free eBooks at bookboon.com

Example

Suppose we make steel cans. If the form of the can is a cylinder of height h and radius r the volume of the can is $V = \pi r^2 h$ and the area of the steel used is $A = 2\pi rh + 2\pi r^2$.

We want the volume to be 64cc. and hence $V = \pi r^2 h = 64$ which gives $h = 64/(\pi r^2)$. The area is therefore $A = 2\pi rh + 2\pi r^2 = 128/r^2 + 2\pi r^2$

To minimize the area we compute

$$\frac{dA}{dr} = -128/r^2 + 4\pi r$$

which is zero when $4\pi r^3 = 128$ giving $r \simeq 2.17$ and $h = 64/(\pi r^2) \simeq 4.34$.

To check that this is a minimum

$$\frac{d^2A}{dr^2} = 256/r^3 + 4\pi$$

which is positive when r is positive so we have a minimum.

The Taylor Expansion

We leave you with one useful approximation. If we have a function $f(x)$ then we have

$$f(x+a) = f(x) + a\frac{df}{dx} + \frac{a^2}{2!}\frac{df^2}{dx^2} + \ldots + \frac{a^n}{n!}\frac{df^n}{dx^n} + \ldots$$

When a is small and we evaluate the derivatives at x. For example if we take $\sin x$ the derivatives are $\cos x$, $-sinx$, $-cosx$, $sinx$, ... So at $x = 0$ since $\sin 0 = 0$ and $\cos 0 = 1$

$$\sin(a) = a - \frac{a^3}{3!} + \frac{a^5}{5!} - \frac{a^7}{7!} - \ldots$$

8.0.3 Newton-Raphson method

We now examine a method, known as the Newton-Raphson method, that makes use of the derivative of the function to find a zero of that function. Suppose we have reason to believe that there is a zero of $f(x)$ near the pointx_0. The Taylor expansion for $f(x)$ about x_0 can be written as:

$$f(x) = f(x_0) + (x - x_0)f'(x_0) + \frac{1}{2!}(x - x_0^2 f''(x_0) + \ldots$$

If we drop the terms of this expansion beyond the first order term we have

$$f(x) = f(x_0) + (x - x_0)f'(x_0)$$

Download free eBooks at bookboon.com

Now set $f(x) = 0$ to find the next approximation, x_1, to the zero of $f(x)$, we find:

$$f(x_1) = f(x_0) + (x_1 - x_0)f'(x_0) = 0$$

or

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$$

This provides us with an *iteration scheme* which may well converge on the zero of $f(x)$, under appropriate conditions.

example

Suppose we want the cube root of 2 or the value of x for which $f(x) = x^3 - 2 = 0$. Here $f'(x) = 3x^2$ so

$$x_1 = x_0 - \frac{x_0^3 - 2}{3x_0^2}$$

Starting with $x_0 = 1$ we have $x_1 = 1.333333$ and using this value for x_0 we get $x_1 = 1.263889$. The steps are laid out below

Step	Estimate
0	1
1	1.333333
2	1.263889
3	1.259933
4	1.259921

Or suppose $f(x) = \sin x - \cos x$ then $f'(x) = \cos x + \sin x$ and so

$$x_1 = x_0 - \frac{\sin x_0 - \cos x_0}{\cos x_0 + \sin x_0}$$

then starting with $x_0 = 1$ we have

Step	Estimate
1	1
2	0.7820419
3	0.7853982
4	0.7853982
5	0.7853982

To examine the conditions under which this iteration converges, we consider the iteration function

$$g(x) = x - \frac{f(x)}{f'(x)}$$

Download free eBooks at bookboon.com

whose derivative is:

$$g'(x) = 1 - \frac{(f'(x))^2 - f(x)f''(x)}{(f'(x))^2} = \frac{f(x)f''(x)}{(f'(x))^2}$$

At the actual zero, $f(x) = 0$, so that as long as $f'(x) = 0$, we have $g'(x) = 0$ at the zero of $f(x)$. In addition we would like the iteration function to get smaller, that is $| g'(x) | < 1$. We conclude that the Newton-Raphson method converges in the interval where.

$$\frac{f(x)f''(x)}{(f'(x))^2} < 1$$

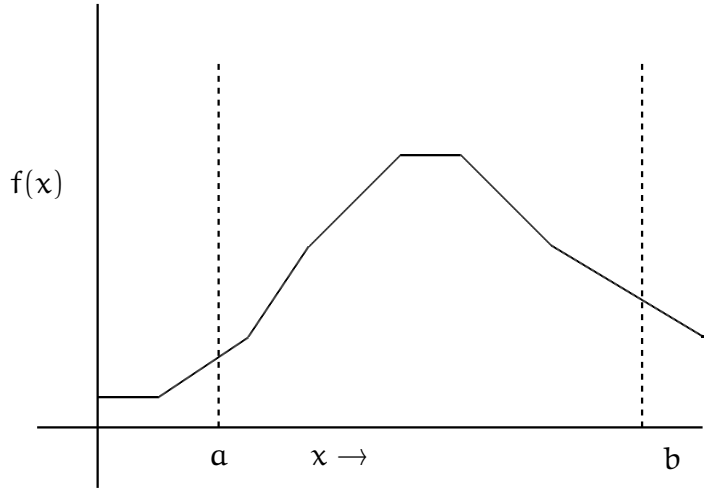
Step	Estimate
0	1
1	1.333333
2	1.263889
3	1.259933
4	1.259921

Download free eBooks at bookboon.com

Click on the ad to read more

8.0.4 Integrals and Integration

Many important problems can be reduced to finding the area under a curve between two points a and b



The obvious idea is to split the area into small rectangles and sum the area of these. So if we take the rectangle between x_j and x_{j+1} this has a height of $f(x_j)$ and an area of $f(x_j)(x_{j+1} - x_j)$. If we add all such rectangles this gives an gives an approximation to the area. We do better when the width of the rectangles gets small so if we choose all the widths as δ our approximation is

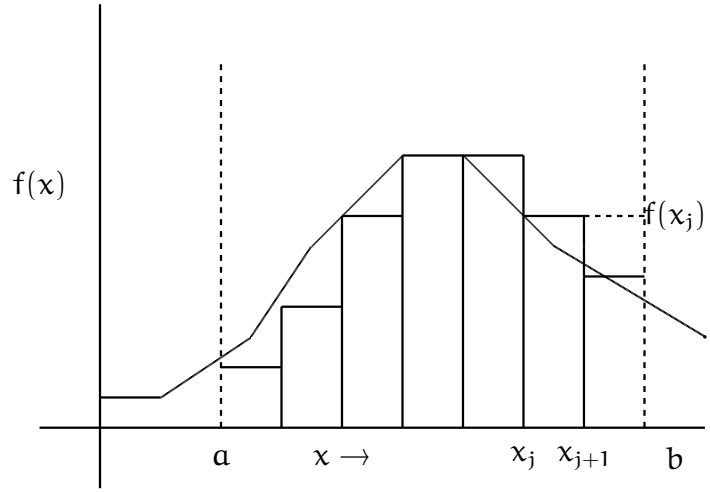
$$\sum f(x_j)\delta x \text{ for } a = x_1, x_2, \ldots, x_n = b$$

When we shrink δx to zero we have the area we need and write

$$\int_a^b f(x)\,dx$$

The $\int$ sign was originally a capital S, for sum.

Download free eBooks at bookboon.com



We avoid technicalities and define the *definite integral* of a function $f(x)$ between a and b as

$$\int_a^b f(x)dx$$

which is the area under the curve, see figure 8.1 Using the idea of areas we have

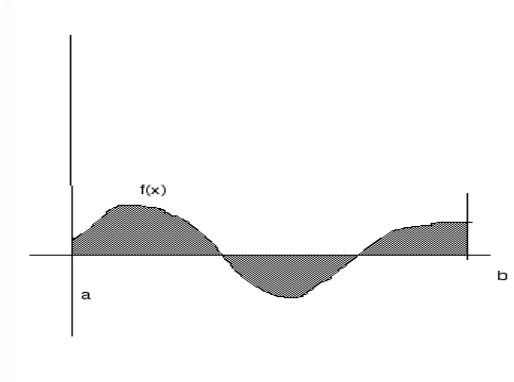


Figure 8.1: Areas under $f(x)$

some rule for integrals

1. If $a \leq c \leq b$ then $\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx$
2. For a constant c $\int_a^b cf(x)dx = c\int_a^b f(x)dx$
3. For two functions $f(x)$ and $g(x)$ $\int_a^b c\left(f(x) + g(x)\right)dx = \int_a^b f(x)dx + \int_a^b g(x)dx$

Download free eBooks at bookboon.com

Perhaps the most important result about integration is the fundamental theorem of calculus. It is easy to follow, if not to prove. Suppose we have a function $f(x)$ and we define

$$F(x) = \int_a^x f(t)dt$$

Then

$$\frac{dF(x)}{dx} = \frac{d}{dx}\left(\int_a^x f(t)dt\right) = f(x)$$

In other words integration is rather like the reverse of differentiation. We need to be a bit careful so define $F(x)$ as the *primitive* of $f(x)$ if

$$\frac{dF(x)}{dx} = f(x)$$

So $\log x$ is a primitive for $1/x$ as is $\log x + 23$. The primitive is normally called the indefinite integral $\int f(x)dx$ of $f(x)$ and is defined up to a constant, so $\int f(x)dx = F(x) + constant$

If the limits of the integration exist, say a and b then we have the definite integral

$$\int_a^b f(x)dx = F(b) - F(a) \tag{8.2}$$

We can of course spend time looking at functions which differentiate to what we want. Normally however we use tables (or our memory) So

$f(x)$	$F(x) = \int f(x)dx$
x^n $(n \neq -1)$	$x^n/(n+1)$
$1/x$	$\log x$
$\exp(ax)$	$\exp(ax)/a$
$\log x$	$x \log x - x$
a^x	$a^x/\log a$
$\sin(ax)$	$-cos(ax)/a$
$\cos(ax)$	$\sin(ax)/a$
$1/\sqrt{a^2 - x^2}$	$\sin^{-1}(x/a)$ $(-1 < x < a)$
$1/(a^2 - x^2)$	$\tan^{-1}(x/a)$

Example

1. $\int x^2 dx = x^3/3 + constant$
2. $\int_{-2}^{3} x^2 dx = \left[x^3/3\right]_{-2}^{3} = (3)^3/3 - (-2)^3/3 = (27+8)/3$
3. $\int_1^{10} = 1/x dx = [\log x]_1^{10} = \log 10 - \log 1 = 2.30\ldots - 0$
4. $\int_0^{1/2} dx/\sqrt{1-x^2} = [\sin_1(x)]_1^{10} = \sin^{-1}(1/2) - \sin_{-1}(0) = \pi/6$

Download free eBooks at bookboon.com

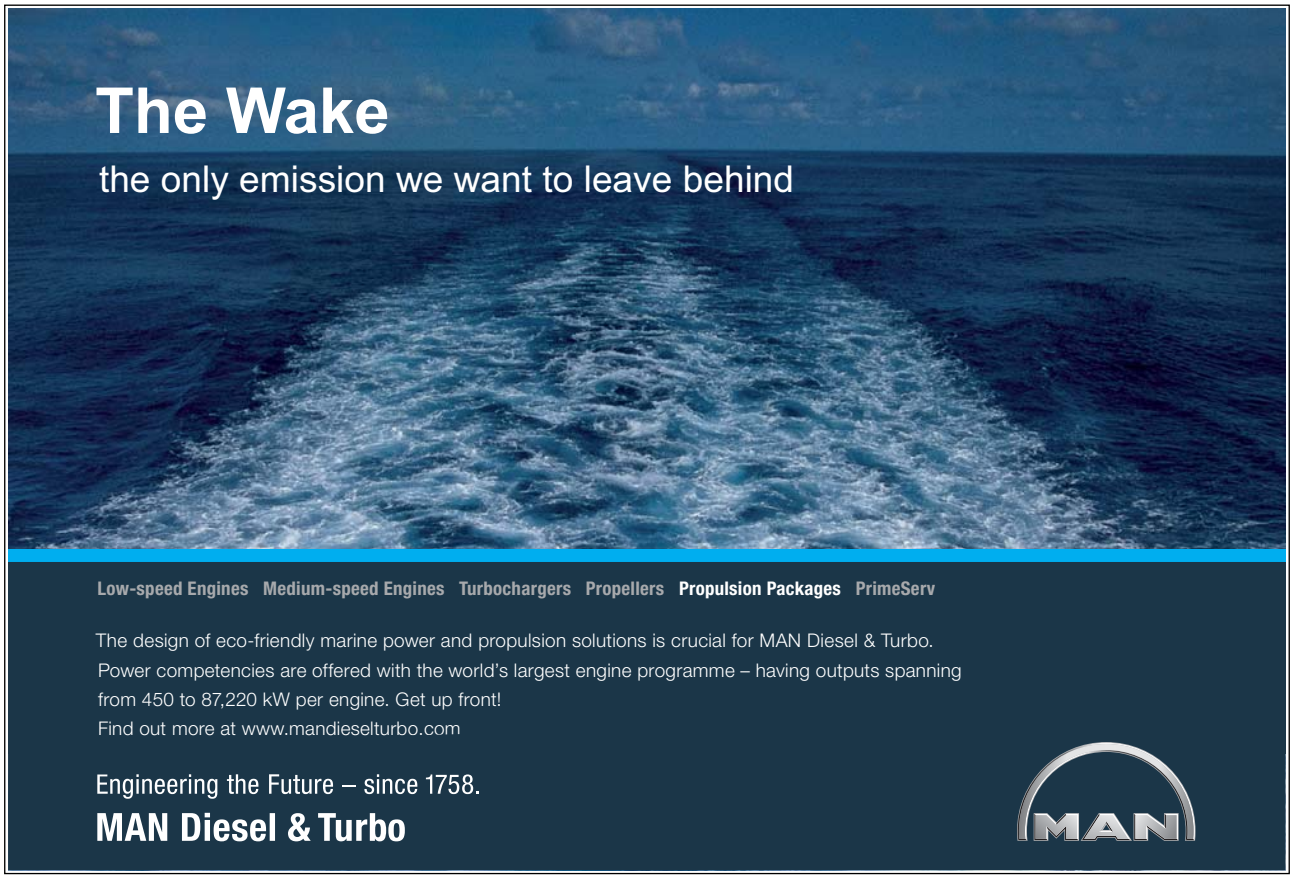
Exercises

Evaluate the following integrals and check your solutions by differentiating.

1. $\int x^3 dx$
2. $\int 1/x^2 dx$
3. $\int (25 + x^2)^{-1} dx$

Evaluate

1. $\int_3^7 \log x dx$
2. $\int_1^2 x^{-3/2} dx$
3. $\int_a^{2a} (a^2 + x^2)^{-1} dx$



Download free eBooks at bookboon.com

Click on the ad to read more

Chapter 9

Algebra: Matrices, Vectors etc.

The human mind has never invented a labor-saving machine equal to algebra

Author Unknown

We now meet the ideas of *matrices* and *vectors*. While they may seem rather odd at first they are vital for studies in almost all subjects. The easiest way to see the power of the idea is to consider simultaneous equations. Suppose we have the set of equations

$$\begin{aligned} 3x - 5y &= 12 \\ x + 5y &= 24 \end{aligned}$$

We can find the solution $x = 9 \quad y = 3$ in several ways . For example if we add the second equation to the first we have equations

$$\begin{aligned} 4x &= 36 \\ x + 5y &= 24. \end{aligned}$$

Thus $x = 9$ and substituting 9 in the second equation gives $9 + 5y = 24$ or $5y = 15$ giving $y = 3$. Many mathematical models result in sets of simultaneous equations, like these except much more complex which need to be solved, or perhaps just to be examined. To do this more easily the matrix was invented. The essence of the set of equations

$$\begin{aligned} 3x - 5y &= 12 \\ x + 5y &= 24 \end{aligned}$$

is captured in the array or matrix of coefficients $\begin{pmatrix} 3 & -5 \\ 1 & 5 \end{pmatrix}$ or the *augmented matrix* $\begin{pmatrix} 3 & -5 & 12 \\ 1 & 5 & 0 \end{pmatrix}$ These arrays of numbers are called matrices. To save

Download free eBooks at bookboon.com

space we often give matrices names in boldface, for example

$$\mathbf{A} = \begin{pmatrix} 3 & -5 & 12 \\ 1 & 5 & 24 \\ 6 & 8 & -3 \\ 11 & 0 & 0 \end{pmatrix}$$

or

$$\mathbf{X} = \begin{pmatrix} 3 & -5 & 12 & 0 \\ 1 & 5 & 24 & 0 \end{pmatrix}.$$

We define an $r \times c$ matrix as a rectangular array of numbers with r rows and c columns, for example $\mathbf{A}$ above is a 4×3 matrix while $\mathbf{X}$ is 2×4 is

A matrix with just one column is called a *column vector* while one with just one row is a *row vector*, for example a column vector

$$\mathbf{a} = \begin{pmatrix} 3 \\ 1 \\ 6 \end{pmatrix}$$

and a row vector

$$\mathbf{b} = \begin{pmatrix} 3 & -5 & 12 & -19 \end{pmatrix}$$

We use matrices in ways which keep our links with systems of equations. Before looking at the arithmetic of matrices we see how we can use them to come up with a general method of solving equations.

9.0.5 Equation Solving.

If you were to look at ways people use to solve equations you would be able do deduce some simple rules.

1. Equations can be multipled by a non-zero constant
2. Equations can be interchanged
3. Equations can be added or subtracted to other equations

If equations are manipulated following these rules they may look different but *they have the same solutions* as when you started. We can solve equations by writing the coefficients in the *augmented matrix* form and manipulating as follows

1. rows of the matrix may be interchanged
2. rows of the matrix may be multiplied by a nonzero constant.

Download free eBooks at bookboon.com

3. rows can be added (or subtracted) to (from) other rows

Our aim is to reduce the matrix to what is known as *row echelon form.* This means that:

- the leading non zero term in each row is a one.
- Also the leading 1 in the first row lies to the left of that in the second row and so on. More precisely the leading 1 in any row lies to the left of the leading ones in all the rows below it.

For example

$$\begin{pmatrix} 1 & -5 & 12 \\ 0 & 0 & 24 \\ 0 & 0 & 0 \end{pmatrix} \text{ or } \begin{pmatrix} 1 & -5 & 12 \\ 0 & 1 & 24 \\ 0 & 0 & 1 \end{pmatrix} \text{ or } \begin{pmatrix} 1 & 1 & 12 & -19 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

The reason for this will become apparent when we do it.
Lets try it out: We start with the equations

$$\begin{aligned} 2x + y + 2z &= 10 \\ x - 2y + 3z &= 2 \\ -x + y + z &= 0 \end{aligned}$$

in this case the coefficients are

$$\begin{pmatrix} 2 & 1 & 2 & 10 \\ 1 & -2 & 3 & 2 \\ -1 & 1 & 1 & 0 \end{pmatrix}$$

We are allowed to manipulate rows, these are *row operations*, to try and get to the row echelon form. Thus we have

1. Add row 2 to row 3 to get $\begin{pmatrix} 2 & 1 & 2 & 10 \\ 1 & -2 & 3 & 2 \\ 0 & -1 & 4 & 2 \end{pmatrix}$
2. Subtract row 2 from row 1 to get $\begin{pmatrix} 1 & 3 & -1 & 8 \\ 1 & -2 & 3 & 2 \\ 0 & -1 & 4 & 2 \end{pmatrix}$
3. Subtract row 1 from row 2 $\begin{pmatrix} 1 & 3 & -1 & 8 \\ 0 & -5 & 4 & -6 \\ 0 & -1 & 4 & 2 \end{pmatrix}$

Download free eBooks at bookboon.com

4. Tidy to get $\begin{pmatrix} 1 & 3 & -1 & 8 \\ 0 & 5 & -4 & 6 \\ 0 & 1 & -4 & -2 \end{pmatrix}$

5. Subtract 5 times row 3 from row 2 to get $\begin{pmatrix} 1 & 3 & -1 & 8 \\ 0 & 0 & 16 & 16 \\ 0 & 1 & -4 & -2 \end{pmatrix}$

6. Interchange rows 2 and 3 $\begin{pmatrix} 1 & 3 & -1 & 8 \\ 0 & 1 & -4 & -2 \\ 0 & 0 & 16 & 16 \end{pmatrix}$

7. Tidy $\begin{pmatrix} 1 & 3 & -1 & 8 \\ 0 & 1 & -4 & -2 \\ 0 & 0 & 1 & 1 \end{pmatrix}$

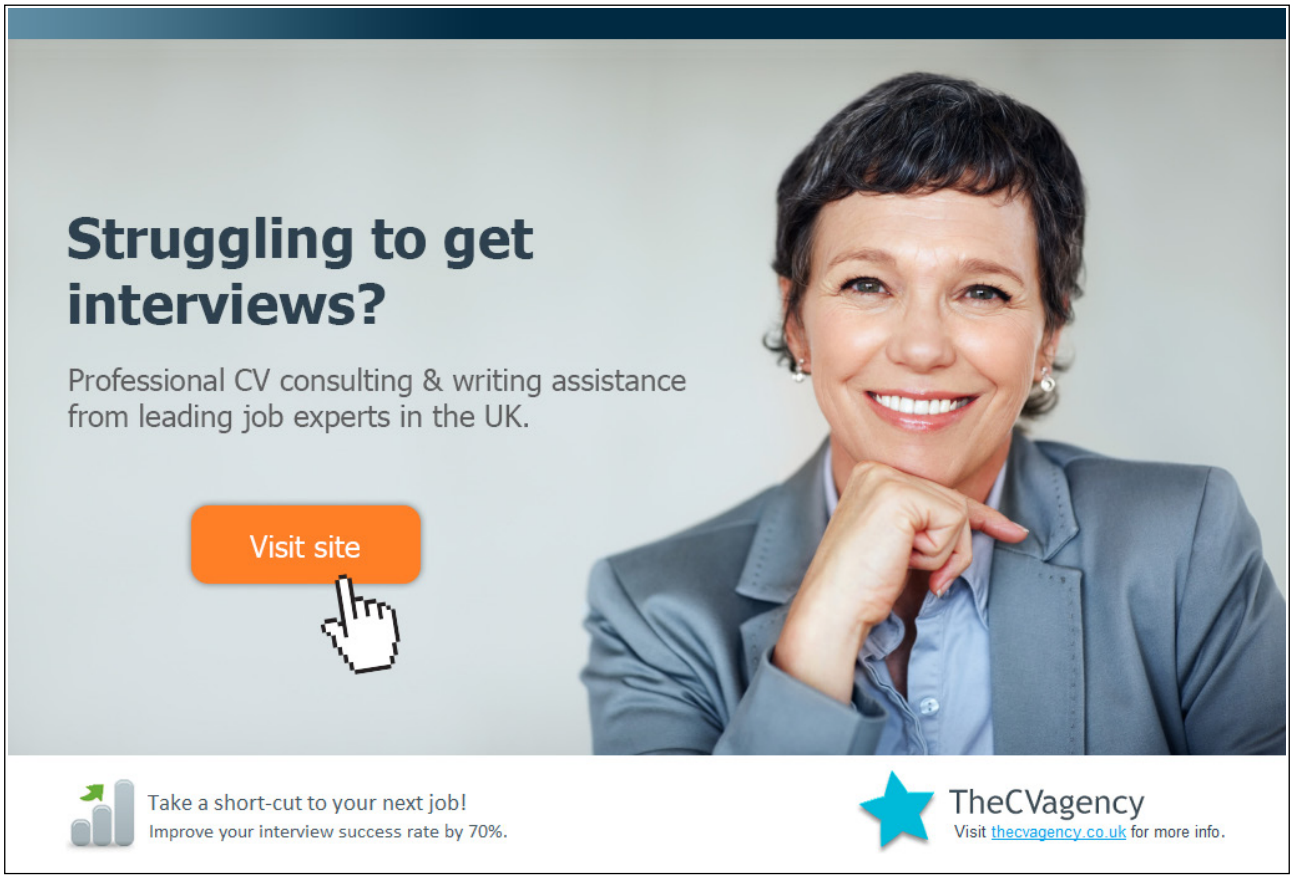
This seems more like row echelon .

This last matrix corresponds to the set of equations

$$\begin{aligned} x + 3y - z &= 8 \\ y - 4z &= -2 \\ z &= 1 \end{aligned}$$

These are much easier to solve! Here

$$z = 1 \quad y = 2 \quad x = 3.$$



Download free eBooks at bookboon.com
Click on the ad to read more

It is often nicer to go a bit further and get rid of as much of the upper triangle as possible. Clearly the leading 1 in each row can be used to get zeros in the column above it. The resulting matrix is called *reduced row echelon form* of the original matrix. Here we get

$$\begin{pmatrix} 1 & 3 & -1 & 8 \\ 0 & 1 & -4 & -2 \\ 0 & 0 & 1 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 3 & 0 & 9 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

It does really matter a great deal to us which we use since we are only interested in solutions.

Lets look at another example

$$\begin{aligned} 6x + 3y + 6z &= 9 \\ x + 2y &= 16 \\ 4x + 5y + 1z &= 18 \end{aligned}$$

The augmented form is

$$\begin{pmatrix} 6 & 3 & 6 & 9 \\ 1 & 2 & 0 & 6 \\ 4 & 5 & 1 & 18 \end{pmatrix}$$

We have

$$\begin{pmatrix} 6 & 3 & 6 & 9 \\ 1 & 2 & 0 & 6 \\ 4 & 5 & 1 & 1 \end{pmatrix} \to \begin{pmatrix} 0 & -9 & 6 & -27 \\ 1 & 2 & 0 & 6 \\ 0 & -3 & 1 & -6 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 0 & 6 \\ 0 & 3 & -2 & 9 \\ 0 & 3 & -1 & -6 \end{pmatrix} \to \ldots \to \begin{pmatrix} 1 & 2 & 0 & 6 \\ 0 & 1 & -2/3 & 3 \\ 0 & 0 & 1 & -3 \end{pmatrix}$$

Some steps have been concatenated!

What can go wrong

In reality nothing much can go wrong but we need to examine a couple of cases where the results we obtain require some thought.

1. Suppose we end up with *a row of zeros.* This is no problem, except when the number of non-zero rows is less that the number of variables. This just means there is not an unique solution e.g

$$\begin{aligned} x + 2y - z &= 0 \\ x + z &= 3 \\ 2x + 2y &= 3 \end{aligned}$$

Download free eBooks at bookboon.com

We have

$$\begin{pmatrix} 1 & 2 & -1 & 0 \\ 1 & 0 & 1 & 3 \\ 2 & 2 & 0 & 3 \end{pmatrix} \to \cdots \to \begin{pmatrix} 1 & 2 & -1 & 0 \\ 0 & 1 & -2 & -3 \\ 0 & 0 & 0 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 1 & -6 \\ 0 & 1 & -1 & -3/2 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

This corresponds to

$$\begin{aligned} x + y + z &= -69 \\ y - z &= -3/2 \end{aligned}$$

Now there is a solution for these equations but it is not the explicit unique type we have been dealing with up to now. If z is known, say z_0 then it follows $x = 3 - z_0$ and $y = (2z_0 - 3)/2$. We have a solution for every z_0 value. Technically there are an *infinite number of solutions.* It is obvious if you think about it that if you have fewer equations than variables (unknowns) then you will not have a simple solution.

If we have 2 rows all zero then we have to give a value to two variables, if 3 then 3 variables and so on.

2. No Solution
   Of course your equations may not have a solution in that they are contradictory, for example:

$$x = 1 \quad y = 3 \quad x = -2 \quad z = 16$$

We recognize the equations are contradictory (have no solutions at all) in the following way. *If we have a row of which is all zero except for the very last element then the equations have no solution.* For example:
Suppose we have the equations

$$\begin{aligned} x - 2y - 3z &= 1 \\ 2x + cy + 6z &= 6 \\ -x + 3y \quad +(c-3)z &= 0 \end{aligned}$$

where c is some constant. We proceed to row echelon

$$\begin{pmatrix} 1 & -2 & 3 & 1 \\ 2 & c & 6 & 6 \\ -1 & 3 & c-3 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & -2 & 3 & 2 \\ 0 & c+6 & 0 & 4 \\ 0 & 1 & c & 0 \end{pmatrix}$$

Download free eBooks at bookboon.com

Before we go further what happens if $c = -6$? The middle row of our matrix corresponds to 0=4 which is nonsense. Thus the original equation set does not have a solution when $c = -6$

However we will just carry on

$$\begin{pmatrix} 1 & -2 & 3 & 1 \\ 2 & c & 6 & 6 \\ -1 & 3 & c-3 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & -2 & 3 & 1 \\ 0 & c+6 & 2c & 6 \\ 0 & 1 & c & 1 \end{pmatrix} \to \begin{pmatrix} 1 & -2 & 3 & 1 \\ 0 & 0 & 2c-c(c+6) & 4 \\ 0 & 1 & c & 0 \end{pmatrix}$$

$$\to \begin{pmatrix} 1 & -2 & 3 & 1 \\ 0 & 1 & c & 0 \\ 0 & 0 & 2c-c(c+6) & 6 \end{pmatrix} \to \begin{pmatrix} 1 & -2 & 3 & 1 \\ 0 & 1 & c & 0 \\ 0 & 0 & -4c-c^2 & 1 \end{pmatrix}$$

Now if $-4c - c^2 = 0$, that is $c = 0$, or $c = -4$ our last equation is $0 = 1$ which is clearly nonsense! This means that the original equations had no solution.

You may feel that this is a bit of a sledge hammer to crack a nut, but there is a real purpose to our exercise. If you move away from the trivial cases then the scheme we have outlined above is the best approach. It is also the technique use in the computer programs available for equation solving. In addition the shape of the reduced row echelon form tell us a lot about matrices. Often we have a system of equations where we have some parameters e.g. using our techniques above we can find the range of values, or perhaps the values themselves when solutions are possible.

The row elimination ideas we have outlined are known as *Gaussian elimination* in numerical circles. The algorithms which bear tis name, while very much slicker are based on these simple ideas.

Download free eBooks at bookboon.com

Exercises

1. Solve

   (a)

$$\begin{aligned} 2x + 3y &= 7 \\ 5x - y &= 9 \end{aligned}$$

   (b)

$$\begin{aligned} x + 3y + 3z &= 1 \\ 2x + 5y + 7z &= 1 \\ -2x - 4y - 5z &= 1 \end{aligned}$$

   (c)

$$\begin{aligned} v - w - x - y - z &= 1 \\ 2v - w + 3x + 4z &= 2 \\ 2v - 2w + 2x + y + z &= 1 \\ v + x + 2y + z &= 0 \end{aligned}$$

   (d)

$$\begin{aligned} w + 2x - 3y - 4z &= 6 \\ w + 3x + y - 2z &= 4 \\ 2w + 5x - 2y - 5z &= 10 \end{aligned}$$

2. Consider the equations

$$\begin{aligned} v - w - x - y - z &= 1 \\ 2v - w + 3x + 4z &= 2 \\ 2v - 2w + 2x + y + z &= 1 \\ v + x + 2y + z &= c \end{aligned}$$

   For what values of c do these equations have a unique solution? Are there any values of c for which there is no solution?

Download free eBooks at bookboon.com

9.0.6 More on Matrices

If we have an $n \times m$ matrix $\mathbf{A}$ we need some way of referring to a particular element. It is common to refer to the (ij)th element meaning the element in row i and column j. We think of the matrix as having the form

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1,n-1} & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2,n-1} & a_{2n} \\ a_{31} & a_{32} & \cdots & a_{3,n-1} & a_{3n} \\ \text{cdots} & \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{m,n-1} & a_{m,n} \end{pmatrix}$$

If we have a typical ijth element we sometimes write

$$\mathbf{A} = (a_{ij})$$

The *unit matrix* is an $n \times n$ matrix with ones on the diagonal and zeros elsewhere, usually written $\mathbf{I}$ for example

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ or } \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

So $\mathbf{A}$ is a unit matrix if

1. It is square.
2. The elements a_{ij} satisfy $a_{ii} = 1$ for all i and $a_{ij} = 0$ for all $i \neq j$

9.0.7 Addition and Subtraction

We can add or subtract matrices *that have the same dimensions* by just adding or subtracting the corresponding elements. For example

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{pmatrix}$$

and

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} - \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11} - b_{11} & a_{12} - b_{12} \\ a_{21} - b_{21} & a_{22} - b_{22} \end{pmatrix}$$

when $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 4 & 0 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} -4 & -3 \\ -2 & -1 \\ 0 & 5 \end{pmatrix}$ then $\mathbf{A} + \mathbf{B} = \begin{pmatrix} -3 & -1 \\ 1 & 3 \\ 4 & 1 \end{pmatrix}$

while $\mathbf{A} - \mathbf{B} = \begin{pmatrix} 5 & 5 \\ 5 & 5 \\ 4 & 5 \end{pmatrix}$

Download free eBooks at bookboon.com

Multiplication by a scalar (number)

We can multiply a matrix $\mathbf{A}$ by a number s to give $s\mathbf{A}$ which is the matrix whose elements are those of $\mathbf{A}$ multiplied by s, so if

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1,n-1} & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2,n-1} & a_{2n} \\ a_{31} & a_{32} & \cdots & a_{3,n-1} & a_{3n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{m,n-1} & a_{m,n} \end{pmatrix}$$

then

$$s\mathbf{A} = \begin{pmatrix} sa_{11} & sa_{12} & \cdots & sa_{1,n-1} & sa_{1n} \\ sa_{21} & sa_{22} & \ldots & sa_{2,n-1} & sa_{2n} \\ sa_{31} & sa_{32} & \cdots & sa_{3,n-1} & sa_{3n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ sa_{m1} & sa_{m2} & \cdots & sa_{m,n-1} & sa_{m,n} \end{pmatrix}$$

We use the term *scalar* for quantities that are not vectors.

Transpose of a matrix

If we take a matrix $\mathbf{A}$ and write the columns as rows then the new matrix is called the transpose $\mathbf{A}$ written $\mathbf{A}^{\mathsf{T}}$ or $\mathbf{A}'$

Thus if $\mathbf{A} = \begin{pmatrix} 1 & 2 & 4 \\ 11 & 12 & 0 \end{pmatrix}$ then $\mathbf{A}^{\mathsf{T}} = \begin{pmatrix} 1 & 11 \\ 2 & 12 \\ 4 & 5 \end{pmatrix}$ Notice that $(\mathbf{A}^{\mathsf{T}})^{\mathsf{T}} = \mathbf{A}$.

Any matrix that satisfies

$$\mathbf{A} = \mathbf{A}^{\mathsf{T}}$$

is said to be *symmetric*. If

$$\mathbf{A} = -\mathbf{A}^{\mathsf{T}}$$

then it is anti-symmetric.

Multiplication of Matrices

This is a rather more complicated topic. We define multiplication in a rather complex way so that we keep a connection with systems of equations. Suppose $\mathbf{A}$ is an $n \times p$ matrix and $\mathbf{B}$ is a $p \times m$ matrix. Then the (ij)th element of $\mathbf{AB}$ is

$$\sum_{k=1}^{p} a_{ik}b_{kj} = a_{i1}b_{1j} + a_{i2}b_{2j} + a_{i3}b_{3j} + a_{i4}b_{4j} + \ldots + a_{ip-1}b_{p-1j} + a_{ip}b_{pj}$$

Download free eBooks at bookboon.com

Note that $\mathbf{AB}$ is an $n \times m$ matrix. One way of thinking of this is to notice that the (ij)th element of the product matrix is made up by multiplying elements in the ith row of the first matrix by the corresponding elements in the jth column of the second matrix. The products are then summed.

examples

$$\begin{pmatrix} 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 7 \\ 6 \\ 4 \end{pmatrix} = 1 \times 7 + 2 \times 6 + 3 \times 4 = 31$$

$$\begin{pmatrix} 7 \\ 6 \\ 4 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \end{pmatrix} = \begin{pmatrix} 7 & 14 & 21 \\ 6 & 12 & 18 \\ 4 & 8 & 12 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 2 \\ 4 & 12 \end{pmatrix} \begin{pmatrix} 1 & 4 \\ 2 & 9 \end{pmatrix} = \begin{pmatrix} 5 & 22 \\ 28 & 124 \end{pmatrix}$$

Some consequences are

- You can only multiply matrices if they have the right dimensions.
- In general $\mathbf{AB} \neq \mathbf{BA}$
- $\mathbf{AI} = \mathbf{A}$
- $\mathbf{IA} = \mathbf{A}$ but $\mathbf{I}$ has different dimensions to that above
- $\mathbf{A0} = \mathbf{0}$
- $\mathbf{0A} = \mathbf{0}$ but $\mathbf{0}$, a matrix of zeros, has different dimensions to that above

bookboon.com

Join our sales team! We are a dynamic global business experiencing rapid growth

Are you looking for a new challenging position that requires self-motivation and allows you to work independently?

Bookboon is the world's largest eBook publisher with a very strong presence in USA.

Established in 1988, Bookboon is currently operating in more than 30 countries.

We have an impressive client base of over 100 major accounts. Bookboon is looking for dedicated sales people in USA to be part of our ongoing success.

As a part of our worldwide sales team, you will be rewarded with a competitive sales and target structure and uncapped earning potential. We are looking for sales people with experience in selling solutions towards CEOs, HR, Marketing or sales departments.

Please send your CV to jobs@bookboon.com

Apply Now!

Download free eBooks at bookboon.com

As we said the reason for this strange idea is so that it ties in with linear equations, thus if

$$\begin{aligned} x + 2y &= u \\ 4x + 9y &= v \end{aligned}$$

and

$$\begin{aligned} v + 4y &= 3 \\ 2v - y &= 0 \end{aligned}$$

these can be written in matrix form

$$\mathbf{A}\boldsymbol{x} = \begin{pmatrix} 1 & 2 \\ 4 & 9 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} u \\ v \end{pmatrix} = \mathbf{u}$$

and

$$\mathbf{B}\mathbf{u} = \begin{pmatrix} 1 & 4 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 3 \\ 0 \end{pmatrix}$$

So we can write both e can write systems of equations as one matrix equation

$$\mathbf{B}\mathbf{A}\boldsymbol{x} = \begin{pmatrix} 3 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 4 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 4 & 9 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 17 & 38 \\ -2 & -5 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 3 \\ 0 \end{pmatrix}$$

This is exactly the same set of equations we would have had if we had eliminated u and v without any matrices.

Inverses

So we have a whole set of algebraic operations we can use to play with matrices, except we have not defined division since if we can multiply then why not divide?

For a (non-zero) number z we can define the inverse z^{-1} which satisfies

$$zz^{-1} = z^{-1}z = 1.$$

In the same way we say that the matrix $\mathbf{A}$ has an inverse $\mathbf{A}^{-1}$. if there is a matrix $\mathbf{A}^{-1}$ which satisfies

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}.$$

Beware not all matrices have inverses! Those that do are said to be *non-singular* otherwise a matrix which does not have an inverse said to be is *singular.* If you

Download free eBooks at bookboon.com

think about it you will see that only square matrices can have inverses. Suppose $\mathbf{A}$ is an $n \times n$ matrix and $\mathbf{B}$ is another $n \times n$ matrix. If

$$\mathbf{AB} = \mathbf{BA} = \mathbf{I}$$

where I is an $n \times n$ unit matrix then $\mathbf{B}$ is the inverse of $\mathbf{A}$. Notice $\mathbf{A}$ must be square *but* not all square matrices have inverses.

We can of course find the inverse by solving equations. For example

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

So

$$\begin{pmatrix} ae + bg & af + bh \\ ce + dg & cf + dh \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

we then solve the four equations .

$$\begin{aligned} ae + bg &= 1 \\ af + bh &= 0 \\ ce + dg &= 0 \\ cf + dh &= 1 \end{aligned}$$

Not a very promising approach. However we can use the row-echelon ideas to get an inverse. All we do is take a matrix $\mathbf{A}$ and paste next to it a unit matrix $\mathbf{I}$. Write this augmented matrix as $\mathbf{B} = (\mathbf{AI})$.

We row reduce $\mathbf{B}$ to *reduced row echelon form.* The position of the original $\mathbf{I}$ is the inverse. For example suppose $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 4 & 9 \end{pmatrix}$ then

$$\mathbf{B} = (\mathbf{AI}) \begin{pmatrix} 1 & 2 & 1 & 0 \\ 4 & 9 & 0 & 1 \end{pmatrix}$$

We get using row operations

$$\begin{pmatrix} 1 & 2 & 1 & 0 \\ 4 & 9 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 1 & 0 \\ 0 & 1 & -4 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 9 & 2 \\ 0 & 1 & -4 & 1 \end{pmatrix}$$

and the inverse is $\mathbf{A}^{-1} = \begin{pmatrix} 9 & 2 \\ -4 & 1 \end{pmatrix}$

Of course we check

$$\begin{pmatrix} 1 & 2 \\ 4 & 9 \end{pmatrix} \begin{pmatrix} 9 & 2 \\ -4 & 1 \end{pmatrix} = \begin{pmatrix} 9 & 2 \\ -4 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 4 & 9 \end{pmatrix}$$

Download free eBooks at bookboon.com

What can go wrong?

1. If you manage to convert the left hand matrix $\mathbf{A}$ to a unit matrix $\mathbf{I}$ then you have succeeded.

2. Sometimes as you manipulate the augmented matrix $\mathbf{B}$ you introduce a row of zeros into the position where you placed $\mathbf{A}$. In this case you can stop as *there is no solution*.

Consider $\mathbf{A} = \begin{pmatrix} 6 & 3 & 6 \\ 1 & 2 & 0 \\ 4 & 5 & 0 \end{pmatrix}$. The augmented matrix is $\mathbf{B} = \begin{pmatrix} 6 & 3 & 6 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 & 1 & 0 \\ 4 & 5 & 1 & 0 & 0 & 0 \end{pmatrix}$
Now using row operations we have

$$\begin{pmatrix} 6 & 3 & 6 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 & 1 & 0 \\ 4 & 5 & 1 & 0 & 0 & 0 \end{pmatrix} \to \begin{pmatrix} 0 & -9 & 6 & 1 & -6 & 0 \\ 1 & 2 & 0 & 0 & 1 & 0 \\ 0 & -3 & 1 & 0 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 0 & 0 & 3 & 1 & -6 & 0 \\ 1 & 2 & 0 & 0 & 1 & 0 \\ 0 & -3 & 1 & 0 & -4 & 1 \end{pmatrix}$$

$$\to \begin{pmatrix} 1 & 2 & 0 & 0 & 1 & 0 \\ 0 & 1 & -1/3 & 0 & -4/3 & -1/3 \\ 0 & 0 & 1 & 1/3 & 2 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & -2/9 & -3 & 4/3 \\ 0 & 1 & 0 & 1/9 & 2 & -1/3 \\ 0 & 0 & 1 & 1/3 & 2 & -1 \end{pmatrix}$$

giving us our inverse

$$\begin{pmatrix} -2/9 & -3 & 4/3 \\ 1/9 & 2 & -1/3 \\ 1/3 & 2 & 0 \end{pmatrix}$$

Consider now $\mathbf{A} = \begin{pmatrix} 6 & 3 & 6 \\ 1 & 2 & 0 \\ 4 & 8 & 0 \end{pmatrix}$. The augmented matrix is $\mathbf{B} = \begin{pmatrix} 6 & 3 & 6 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 & 1 & 0 \\ 4 & 8 & 0 & 0 & 1 & 0 \end{pmatrix}$
Now using row operations we have

$$\begin{pmatrix} 6 & 3 & 6 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 & 1 & 0 \\ 4 & 8 & 0 & 0 & 0 & 0 \end{pmatrix} \to \begin{pmatrix} 0 & -9 & 6 & 1 & -6 & 0 \\ 1 & 2 & 0 & 0 & 1 & 0 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & 0 & 0 & 0 \end{pmatrix}$$

Given the zeros we know there is no inverse!

Of course we can think of solving equations using inverse matrices. It is almost always better to use row operations on the augmented matrix but we can proceed as follows. If we have the equations

$$\begin{aligned} 6x + 3y + 6z &= 9 \\ x + 2y &= 6 \\ 4x + 5y + z &= 18 \end{aligned}$$

Download free eBooks at bookboon.com

this can be written as

$$\begin{pmatrix} 6 & 3 & 6 \\ 1 & 2 & 0 \\ 4 & 5 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 9 \\ 6 \\ 18 \end{pmatrix}$$

so

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 6 & 3 & 6 \\ 1 & 2 & 0 \\ 4 & 5 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 9 \\ 6 \\ 18 \end{pmatrix} = \begin{pmatrix} -2/9 & -3 & 4/3 \\ 1/9 & 2 & -2/3 \\ 1/3 & 2 & 1 \end{pmatrix} \begin{pmatrix} 9 \\ 6 \\ 18 \end{pmatrix}$$

In general if

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

then

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$$

provided $\mathbf{A}^{-1}$ exists.

Download free eBooks at bookboon.com
Click on the ad to read more

Summary

1. The transpose of $\mathbf{A}$ written $\mathbf{A}^{\mathsf{T}}$ is the matrix made by writing the rows of $\mathbf{A}$ as columns in $\mathbf{A}^{\mathsf{T}}$.
2. $\mathbf{A}$ is symmetric if $\mathbf{A} = \mathbf{A}^{\mathsf{T}}$
3. The zero matrix is the $n \times m$ array of zeros e.g $\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$
4. The unit matrix $\mathbf{I}$ (of order n) is the $n \times m$ matrix with 1's on the diagonal and zeros elsewhere e.g. $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$
5. The matrix $\mathbf{A}$ has an inverse $\mathbf{B}$ iff $\mathbf{AB} = \mathbf{BA} = \mathbf{I}$. $\mathbf{B}$ is written $\mathbf{A}^{-1}$.
6. A matrix which has an inverse is said to be non-singular.
7. Do remember that except in special cases $\mathbf{AB} \neq \mathbf{BA}$

Exercises

1. Given $\mathbf{A} = \begin{pmatrix} 1 & -1 & 1 \\ -3 & 2 & -1 \\ -2 & 1 & 0 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 1 & 2 & 0 \end{pmatrix}$ compute $\mathbf{AB}$ and $\mathbf{BA}$
2. Show that $\begin{pmatrix} 0 & -2 & 3 \\ 2 & 0 & 4 \\ -3 & -4 & 3 \end{pmatrix}$ is skew symmetric.
3. If $\mathbf{A} = \begin{pmatrix} 2 & -2 & 4 \\ -1 & 3 & 4 \\ 1 & -2 & 0 \end{pmatrix}$ show that $\mathbf{A} = \mathbf{A}^2$
4. Show that $\mathbf{AB}^{\mathsf{T}} = \mathbf{B}^{\mathsf{T}}\mathbf{A}^{\mathsf{T}}$
5. Show that the inverse of $\mathbf{AB}$ is $\mathbf{B}^{-1}\mathbf{A}^{-1}$
6. Find the inverse of $\begin{pmatrix} 2 & 3 & 1 \\ 1 & 2 & 3 \\ 3 & 2 & 2 \end{pmatrix}$ and $\begin{pmatrix} 2 & 4 & 3 & 2 \\ 3 & 6 & 5 & 2 \\ 2 & 5 & 2 & -3 \\ 4 & 5 & 14 & 0 \end{pmatrix}$

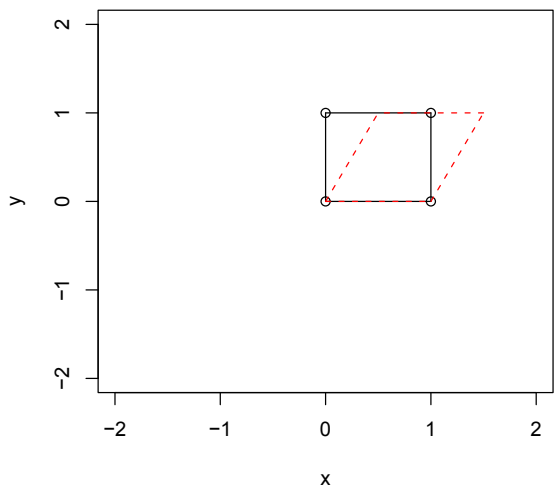
Download free eBooks at bookboon.com

Geometry

We write the point (x, y) in the plane as the vector $\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}$. If $\mathbf{A}$ is a 2×2 matrix $\mathbf{Ax}$ transforms $\mathbf{x}$ into a new point. Suppose $\mathbf{A} = \begin{pmatrix} 1 & 1/2 \\ 0 & 1 \end{pmatrix}$ Then

1. $\mathbf{A}\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$
2. $\mathbf{A}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$
3. $\mathbf{A}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1/2 \\ 1 \end{pmatrix}$
4. $\mathbf{A}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3/2 \\ 1 \end{pmatrix}$

If we plot the 4 points (0,0),(0,1),(1,1),(0,1) and their transforms we get



9.1 Determinants

Consider the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$. We can show that this has an inverse $\begin{pmatrix} e & f \\ g & d \end{pmatrix}$ when $\nabla = ad - bc \neq 0$, see 9.0.7. The quantity ∇ is called the *determinant* of the

Download free eBooks at bookboon.com

matrix $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and is written $\begin{vmatrix} a & b \\ c & d \end{vmatrix}$ or $\det(\mathbf{A})$. Similarly $\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$ has an inverse when

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix} \neq 0$$

The general definition of a determinant of an $n \times n$ matrix $\mathbf{A}$ is as follows.

1. If $n = 1$ then $\det(\mathbf{A}) = a_{11}$
2. if $n > 1$ Let M_{ij} be the *determinant* of the $(n-1) \times (n-1)$ matrix obtained from $\mathbf{A}$ by deleting row i and column j. M_{ij} is called a *minor*.

Then

$$\det(\mathbf{A}) = a_{11}M_{11} - a_{12}M_{12} + a_{13}M_{13} - a_{14}M_{14} + \ldots (-1)^{n+1}a_{1n}M_{1n} = \sum_{j=1}^{n} (-1)^{j+1} a_{1j} M_{1j}$$

Determinants are pretty nasty but we are fortunate as we really only need them for $n = 1, 2$ or 3.

9.2 Properties of the Determinant

1. Any matrix $\mathbf{A}$ and its transpose $\mathbf{A}^T$ have the same determinant, i.e. $\det(\mathbf{A}) = \det(\mathbf{A}^T)$.

   Note: This is useful since it implies that whenever we use rows, a similar behavior will result if we use columns. In particular we will see how row elementary operations are helpful in finding the determinant.

2. The determinant of a triangular matrix is the product of the entries on the diagonal, that is $\begin{pmatrix} a & b & c \\ 0 & e & f \\ 0 & 0 & i \end{pmatrix} = aei$

3. If we interchange two rows, the determinant of the new matrix is the opposite sign of the old one, that is

$$\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = - \begin{pmatrix} d & e & f \\ a & b & c \\ g & h & i \end{pmatrix}$$

Download free eBooks at bookboon.com

4. If we multiply one row by a constant, the determinant of the new matrix is the determinant of the old one multiplied by the constant, that is $\begin{pmatrix} a & b & c \\ d & e & f \\ \lambda g & \lambda h & \lambda i \end{pmatrix} = \lambda \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$ In particular, if all the entries in one row are zero, then the determinant is zero.

5. If we add one row to another one multiplied by a constant, the determinant of the new matrix is the same as the old one, that is $\begin{pmatrix} a & b & c \\ d & e & f \\ \lambda a + g & \lambda b + h & \lambda c + i \end{pmatrix} = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$

   Note that whenever you want to replace a row by something (through elementary operations), do not multiply the row itself by a constant. Otherwise, it is easy to make errors, see property 4

6. $\det(\mathbf{AB})=\det(\mathbf{A})\det(\mathbf{B})$

7. $\mathbf{A}$ is invertible if and only if $\det(\mathbf{A}) \neq 0$. Note in that case $\det(\mathbf{A}^{-1})=1/\det(\mathbf{A})$

While determinants can be useful in geometry and theory they are complex and quite difficult to handle. Our last result is for completeness and links matrix inverses with determinants.

Recall that the $n \times n$ matrix $\mathbf{A}$ does not have an inverse when $\det(\mathbf{A})=0$. However the connection between determinants and matrices is more complex. Suppose we define a new matrix, the adjoint of $\mathbf{A}$ say $\mathrm{adj}(\mathbf{A})$ as

$$\mathrm{adj}\mathbf{A} = \left((-1)^{i+1}M_{ij}\right)^T = \begin{pmatrix} M_{11} & -M_{12} & \cdots & (-1)^{n+1}M_{1,n} \\ -M_{21} & M_{22} & \cdots & (-1)^{n+2}M_{2,n} \\ \cdots & \cdots & \cdots & \cdots \\ (-1)^{n+1}M_{n1} & (-1)^{n+2}M_{n2} & \cdots & (-1)^{2n}M_{nn} \end{pmatrix}^T$$

Here the M_{ij} are just the minors defined above.

So if $\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 5 \\ 1 & 5 & \cdots \end{pmatrix}$ then $\mathrm{adj}(\mathbf{A}) = \begin{pmatrix} 11 & -7 & 2 \\ -9 & 9 & -3 \\ 1 & -2 & 1 \end{pmatrix}^T = \begin{pmatrix} 11 & -9 & 1 \\ -7 & 9 & -2 \\ 2 & -3 & 1 \end{pmatrix}$

Why is anyone interested in the adjoint? The main reason is

$$\mathbf{A}^{-1} = \frac{\mathrm{adj}\mathbf{A}}{\det(\mathbf{A})}$$

Of course you would have to have a very special reason to compute an inverse this way.

Download free eBooks at bookboon.com

9.2.1 Cramer's Rule

Suppose we have the set of equations

$$\begin{aligned} a_1x + b_1y + c_1z &= d_1 \\ a_2x + b_2y + c_2z &= d_2 \\ a_3x + b_3y + c_3z &= d_3 \end{aligned}$$

and let $D = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}$

Then Cramer' s rule states that

$$x = \frac{1}{D}\begin{vmatrix} d_1 & b_1 & c_1 \\ d_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}$$

$$y = \frac{1}{D}\begin{vmatrix} a_1 & d_1 & c_1 \\ a_2 & d_2 & c_2 \\ a_3 & d_3 & c_3 \end{vmatrix}$$

$$z = \frac{1}{D}\begin{vmatrix} a_1 & b_1 & d_1 \\ a_2 & b_2 & d_2 \\ a_3 & b_3 & d_3 \end{vmatrix}$$

There is even a more general case. Suppose we have

$$\mathbf{A}\boldsymbol{x} = \mathbf{d}$$

where $\mathbf{x}^T = (x_1, x_2, \ldots, x_n)$ and $\mathbf{d}^T = (d_1, d_2, \ldots, d_n)$. Let $D = \det(\mathbf{A})$.

Then

$$x_k = \frac{1}{D}\begin{pmatrix} a_{11} & \cdots & a_{1(k-1)} & d_1 & a_{1(k+1)} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n1} & \cdots & a_{n(k-1)} & d_n & a_{n(k+1)} & \cdots & a_{nn} \end{pmatrix}$$

While this is a nice formula you would have to be mad to use it to solve equations since the best way of evaluating big determinants is by row reduction, and this gives solutions directly.

Exercises

1. Evaluate $\begin{vmatrix} 2 & 4 \\ 3 & 6 \end{vmatrix}$

Download free eBooks at bookboon.com

2. Evaluate $\begin{vmatrix} 2 & 4 & 3 \\ 3 & 6 & 5 \\ 2 & 5 & 2 \end{vmatrix}$

3. Evaluate $\begin{vmatrix} x & 1 & 2 \\ x^2 & 2x+1 & x^3 \\ 0 & 3x-2 & 2 \end{vmatrix}$

4. If $\mathbf{A} = \begin{pmatrix} a & b & 0 & 0 \\ c & d & 0 & 0 \\ 0 & 0 & e & f \\ 0 & 0 & g & 14 \end{pmatrix}$ show that

$$\det(\mathbf{A}) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} \begin{vmatrix} e & f \\ g & d \end{vmatrix}$$

TURN TO THE EXPERTS FOR
SUBSCRIPTION CONSULTANCY

Subscrybe is one of the leading companies in Europe when it comes to innovation and business development within subscription businesses.

We innovate new subscription business models or improve existing ones. We do business reviews of existing subscription businesses and we develope acquisition and retention strategies.

Learn more at linkedin.com/company/subscrybe or contact
Managing Director Morten Suhr Hansen at mha@subscrybe.dk

SUBSCRYBE - to the future

Download free eBooks at bookboon.com

Click on the ad to read more

Chapter 10

Probability

Probability theory is nothing but common sense reduced to calculation.

Pierre Simon Laplace

In what follows we are going to cover the basics of probability. The ideas are reasonably straightforward, however as it involves counting it is very easy to make mistakes - as we shall see.

Suppose we perform an experiment whose outcome is not perfectly predictable e.g. roll a die or toss a coin. Imagine we make a list of all possible outcomes, call this list S the *sample space*. So

- If we toss a coin S consists of {Head, Tail}, we write S = {Head, Tail},
- If we roll a die S={ 1,2,3,4,5,6}
- If a princess kisses a frog then we have two possibilities
  S={ we get a prince, we get an embarrassed frog}
- When we roll two dice then S is the set of pairs

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

An *event* A is a collection of outcomes of interest, for example rolling two dice and getting a double. In this case the event A is defined as

Download free eBooks at bookboon.com

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

Suppose that the event B is that the sum is less that 4 when we roll two dice, then

$$B=\{ (1,1),(1,2),(2,1)\} .$$

If two events A and B have no elements in common then we say they are *mutually exclusive*. For example let A be the event {At least one 6} that is

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

Since A and B have no elements in common they are mutually exclusive. Define the event C as

$$C=\{ (2,3),(5,7)\}$$

Then A and C are also mutually exclusive. If D={sum exceeds 10} then A and D are **not** mutually exclusive! *Check this yourself.*

Combining events

- It is handy to have a symbol for not A, we use $\sim A$ but we are not very picky and not A is acceptable.
- The event A and B, often written $A \cap B$ is the set of outcomes which belong both to A and to B.
- The event A or B, often written $A \cup B$ is the set of outcomes which belong either to A or to B or to both.

You will recognise the notation from the earlier discussion on sets.

Suppose S={0,1,2,3,4,5,6,7,8,9} then if we define A={1,3,5,7,9} and B={4,5,7} we have

- $A \cap B = A$ and $B = \{5,7\}$
- While $A \cup B = A$ or $B = \{1,3,4,5,7,9\}$
- $\sim B$=not $B = 1,2,3,4,8,9$

Download free eBooks at bookboon.com

10.0.2 Probability - the rules

Now to each event we are going to assign a measure (in some way) called the probability. We will write the probability of an event A as $P[A]$. We will set out some rules for probabilities, the main ones are as follows:

1. $0 \leq P[A] \leq 1$.
2. $P[S] = 1$
3. For mutually exclusive events A and B $P[A \text{ or } B] = P[A] + P[B]$

We will add a few extra rules

(i) For mutually exclusive events A_1 and A_2 and $A_3 \cdots A_n$ then

$$P[A_1 \cup A_2 \cup A_3 \cdots \cup A_n \cdots] = P[A_1] + P[A_2] + P[A_3] + \cdots + P[A_n] + \cdots$$

or written differently

$$P[A_1 \text{ or } A_2 \text{ or } A_3 \cdots \text{ or } A_n \cdots] = P[A_1] + P[A_2] + P[A_3] + \cdots + P[A_n] + \cdots$$

(ii) For an event A

$$P[\text{ not } A] = 1 - P[A]$$

(iii) For events A and B

$$P[A \text{ or } B] = P[A] + P[B] - P[A \text{ and } B]$$

All this is a bit fiddley but is not really very hard. If you were not too confused at this point you will have noticed that we do not have a way of getting the probabilities. This is a difficult point except in the case we are going to discuss.

10.0.3 Equally likely events

Suppose that every outcome of an experiment is equally likely. Then we can show from the rules above for any event A

$$P[A] = \frac{\text{the number of outcomes in A}}{\text{the number of possible outcomes}}$$

Download free eBooks at bookboon.com

This means we can do some calculations.

examples

1. Suppose that the outcomes

   - that a baby is a girl
   - that a baby is a boy

   are equally likely. Then as there are two possible outcomes we have P[girl]=1/2=P[boy].

2. Suppose now a family has 3 children, the possibilities are

   BB BG GB GG

   and so P[one boy and one girl]= 2/4=1/2 while P[two girls]=1/4

3. The famous statistician R A Fisher had seven daughters. If you count the possible sequences BBBBBBB to GGGGGGG you will find that there are $2^7 = 128$. Only one sequence is all girl so the probability of this event is 1/128.

4. A pair of dice is thrown. What is the probability of getting totals of 7 and 11. Suppose now we throw the two dice twice. What is the probability of getting a total of 11 and 7 in this case?

5. We draw 2 balls from an urn containing 6 white and 5 black, WHat is the probability that we get one white and one black ball?

As you can see we really need some help in counting.

ExercisesS

1. A poker hand consists of 5 cards drawn from a pack of 52. What is the probability that a hand is a *straight*, that is 5 cards in numerical order, but not all of the same suit.

2. What is the probability that a poker hand is a *full house*, that is a triple and a pair.

3. A and B flip a coin in turn. The first to get a head wins. Find the sample space. What is the probability that A wins?

Download free eBooks at bookboon.com

4. The game of craps is played as follows: A player rolls two dice. If the sum is a 2, 3 or 12 he loses. If the sum is a seven or an 11 he wins. Otherwise the player rolls the dice until he gets his initial score, in which case he wins or gets a 7 in which case he loses. What is the probability of winning?

5. A man has n keys, one of which will open his door. He tries keys at random, discarding those that don't work until he opens the door. What is the probability that he is successful on the kth try.

6. **The birthday problem** How many people should be in a room to make the probability of two or more having the same birthday more than 0.5? This is quite difficult and a simpler approach is to consider the probability that no two people have the same birthday.

   It is often a useful dodge in probability to look at P[not A] when P[A] is hard.

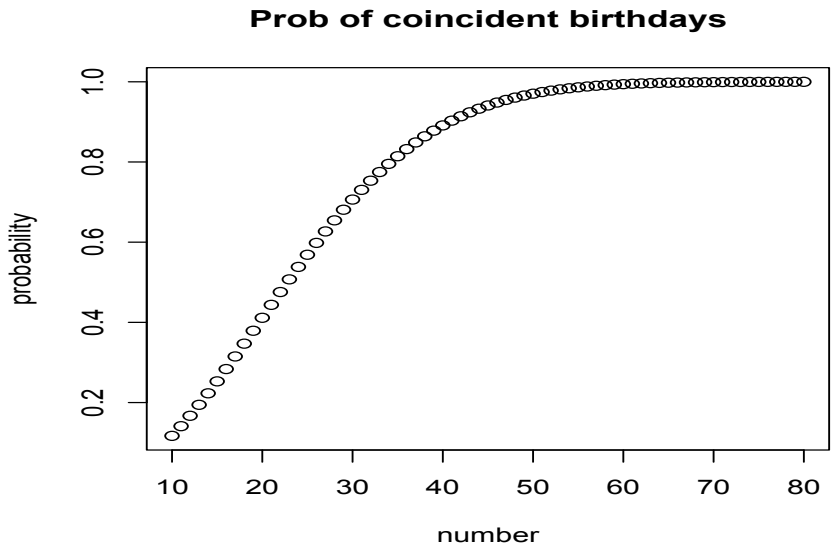
   So P[no coincidences] =

$$\frac{365 \times 364 \times 363 \times \cdots \times (365 - n + 1)}{365 \times 365 \times \cdots \times 365}$$

$$= 1 \times (1 - 364/365) \times (1 - 364/365) \times (1 - 363/365) \times \cdots \times (1 - (365 - n + 1)/365)$$

Number	Probability
15	0.74709868
16	0.71639599
17	0.68499233
18	0.65308858
19	0.62088147
20	0.58856162
21	0.55631166
22	0.52430469
23	0.49270277
24	0.46165574
25	0.43130030

Download free eBooks at bookboon.com



10.0.4 Conditional Probability

Sometime it is natural to talk of the probability of an event A **given** some other event has occurred. We write the probability of A given B as $P[A \mid B]$ and define it as

$$P[A \mid B] = \frac{P[A \cap B]}{P[B]}$$

Remember this is a fancy way of writing

$$P[A \mid B] = \frac{P[A \text{ and } B]}{P[B]}$$

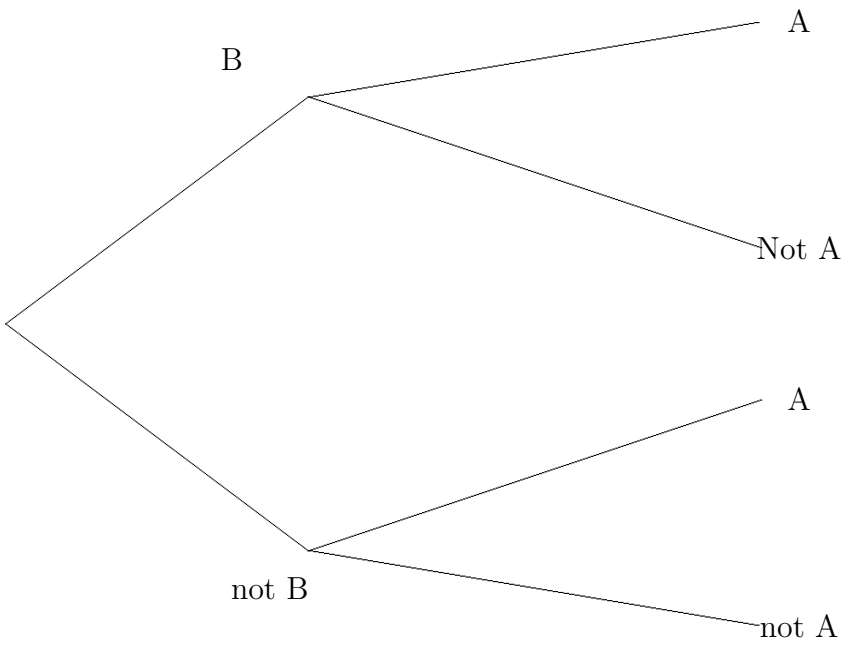
While conditional probabilities can have interesting philosophical implications they also allow one to do calculations. Thus

$$P[A] = P[A \mid B]P[B] + P[A \mid \sim B]P[\sim B]$$

or more generally if $B_1, B_2, \cdots$ are the only possibilities so $\sum_{i=1}^{n} B_i = 1$ then

$$P[A] = \sum_{i=1}^{n} P[A \mid B_i]P[B_i]$$

Download free eBooks at bookboon.com



Download free eBooks at bookboon.com

Click on the ad to read more

Examples

1. Consider the table below.

	Employed	Unemployed	Total
Male	460	40	500
Female	140	260	400
Total	600	300	900

Then

- P[Male] = 500/900
- P[Male and Unemployed] =40/900
- P[Unemployed — Male] =40/500 = P[Unemployed and Male] /P[Male] = 40/900 ÷ 500/900 =40/500.

2. Suppose we buy widgets from 3 suppliers A,B and C. They supply all production and the number of defective items per batch as well as their share of our supply is given below.

	A	B	C
	Supplier		
Proportion supplied	0.60	0.30	0.10
Proportion defective	0.03	0.05	0.07

What proportion of widgets are defective?
We know

- P[defective|A] = 0.03
- P[defective|B]=0.03
- P[defective|C]=0.07

so using the formula we have
P[defective]=P[defective|A]×P[A]+P[defective|B]×P[B]+P[defective|C]×P[C]
So P[defective] = $0.03 \times 0.6 + 0.03 \times 0.3 + 0.07 \times 0.1 = 0.034$

Download free eBooks at bookboon.com

10.0.5 Bayes

We also have **Bayes Theorem**

$$P[A|B] = \frac{P[B|A]P[A]}{P[B]} \tag{10.1}$$

or

$$P[A|B] \propto P[B|A]P[A] \tag{10.2}$$

Here $\propto$ means equal to but multiplied by a constant.

You will often find that you can compute $P[A \mid B]$ when really you want $P[B \mid A]$. Bayes theorem gives you the means for turning one into the other.

Examples

1. Take the data in the example 2 above. We know that $P[\text{defective} \mid A]$=0.03 and we found that $P[\text{defective}]$=0.034. Then suppose we pick up a defective component and ask what is the probability that it come from A. Thus we need $P[A \mid \text{defective}]$.

   We can use Bayes to give

   $$P[A \mid \text{defective}] = P[\text{defective} \mid A]P[A]/P[\text{defective}]$$

   $$= 0.03 \times 0.6/0.34 = 9/17 = 0.529.$$

2. Suppose that the probability that a person has a disease $P[D] = 0.01$. A test is available which is correct 90% of the time. If we use Y to denote that the test is positive and $\sim Y$ negative we mean

   $$P[Y|D] = P[\sim Y| \sim D] = 0.9$$

   Now the probability of a yes is

   $$P[Y] = P[Y|D]P[D] + P[Y| \sim D]P[\sim D] = 0.9 \times 0.01 + 0.1 \times 0.99 = 0.108.$$

   The more interesting case is

   $$P[D|Y] = \frac{P[Y|D]P[D]}{P[Y]} = 0.009/0.108 = 0.0833$$

Download free eBooks at bookboon.com

Exercises

1. An insurance broker believes that a quarter of drivers are accident prone. What is more the probability of an accident prone driver making a claim is 1/3 while for a non accident prone drive the probability is 1/5. What is the probability of a claim? On his way home the broker sees that one of his customers has driven his car into a tree. What is the probability that this customer is accident prone?

2. An urn contains 4 red and 6 green balls. One ball is drawn at random and it's colour observed. It is then returned to the urn and 3 new balls of the same colour are added to the urn, which now contains 13 balls. A second ball is now drawn from the urn.

   (a) What is the probability that the first ball drawn was green?

   (b) What is the probability of getting a red ball *given* the first ball drawn was green

   (c) What is the probability of getting a green ball in the second draw?

3. Sometime used by unscrupulous students of probability -
   We have 3 cards. The first card has two red sides, the second two black sides. The remaining card has one black and one red side. Otherwise the cards are identical.

   The three cards are mixed in a hat and one card is selected at random an placed on a table. If the exposed side is red what is the probability that the hidden side is black?

Download free eBooks at bookboon.com

Click on the ad to read more

Independence

If $P[A|B] = P[A]$ then we say A and B are independent. This is usually written in the equivalent form

$$P[A \cap B] = P[A]P[B]$$

Independent is very useful and plays a central role in statistics.

10.0.6 Random Variables and distributions

If we conduct and experiment and see an outcome we almost always code the outcome in same way, say H,T for head and tail or even 0,1. The coding is known as a random variable, usually written as a capital such as X. If we toss a coin we can say that the outcome is X. The actual values may be head, head, tail, giving the sequence of values of X as H, H, T, . . We use random variables when we have probability distributions, that is lists of possible outcomes and probabilities, such as in the table

k	0	1	2	3
$P[X = k]$	0.1	0.3	0.5	0.1

We point out that the sum of the probabilities must be one, that is $\sum_{k=0}^{3} P[X = k]$ We define the *cumulative distribution function* (c.d.f.) $F(x)$ as the cumulative sum of the probabilities

$$F(x) = \sum_{x=0}^{k} P[X = k]$$

So in the example above

k	0	1	2	3
$P[X = k]$	0.1	0.3	0.5	0.1
$F(x)$	0.1	0.4	0.9	1.0

It is more usual to give a formula for a random variable, for example

$$P[x = k] = 0.3 \times 0.7^{x-1} \qquad x = 1, 2, 3, \cdots$$

As the formula is commonly shorter you can see why.

Download free eBooks at bookboon.com

10.1 Expectation

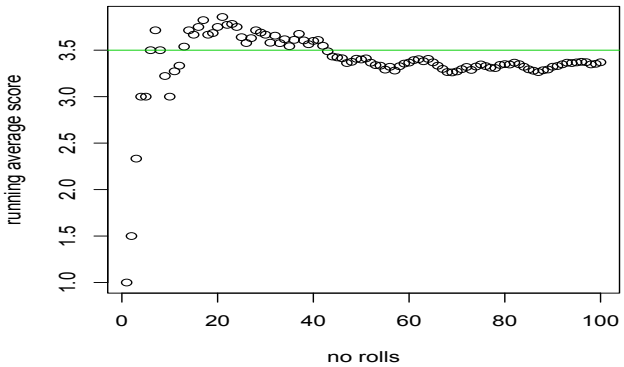
We can also view probability from the point of view of what happens in the long run. Given a random variable X define the expected value of X written $E[X]$ as

$$E[X] = \sum_{\text{all}x} xP[X = x]$$

The expected value can be regarded as the *long run average*. So if we roll a fail die and the outcome is X then $P[X = i] = 1/6 \qquad i = 1, 2, \cdots, 6]$ and so

$$E[X] = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + \cdots + 6 \times \frac{1}{6} = 3.5$$

You can be sure that if you roll a die you will never get 3.5, *however* if you rolled a die and kept an average of the score you will find that this will approach 3.5, see the plot below

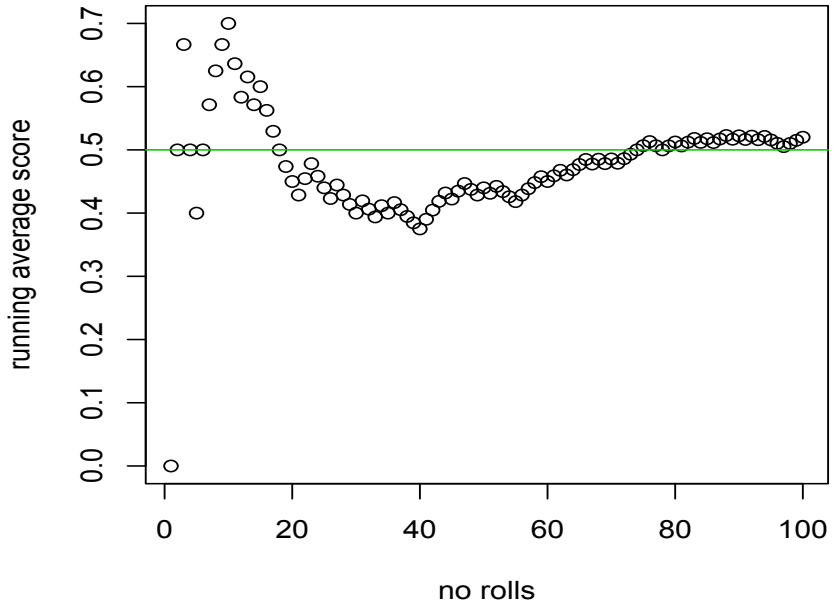


For a coin we have Head and Tail. Suppose we count head as 1 and tail as zero, then

$$P[X = 1] = 1/2 \text{ and } P[X = 0] = 1/2$$

and so $E[X] = 1 \times \frac{1}{2} + 0 \times \frac{1}{2} = \frac{1}{2}$. A similar experiment gives the following

Download free eBooks at bookboon.com



10.1.1 Moments

Some important expected values in statistics are the *moments*

$$\mu_r = E[X^r] \qquad r = 1, 2, \ldots$$

since we can usually estimate these while probabilities are much more difficult. You will have met the

- *mean* $\mu = E[X]$
- The *variance* $\sigma^2 = E[(X - \mu)^2]$.
- The parameter σ is known as the *standard deviation.*

The central moments are defined as

$$\mu_r = E[(X - \mu)^r] \qquad r = 1, 2, \ldots$$

The third and fourth moments $E[(X-\mu)^3]$,$E[(X-\mu)^4]$ are less commonly used. We can prove an interesting link between the mean μ and the variance σ^2. The result s known as Chebyshev's inequality

$$P[|X - \mu| > \epsilon] \leq \left[\frac{\sigma}{\epsilon}\right]^2 \tag{10.3}$$

This tells us that departure from the mean have small probability when σ is small.

Download free eBooks at bookboon.com

10.1.2 Some Discrete Probability Distributions

We shall run through some of the most common and important discrete probability distributions.

The Discrete Uniform distribution

Suppose X can take one values $1, 2, \cdots, n$ with equal probability, that is

$$P[X = k] = \frac{1}{n} \qquad k = 1, 2, \cdots, n \tag{10.4}$$

- The mean is $E[X] = \frac{n+1}{2}$
- the variance is $\text{var}(X) = \frac{1}{3}n^3 + \frac{1}{4}n^2 + \frac{2}{3}n - \frac{1}{4}$

For example a die is thrown, the distribution of the score X is uniform on the integer 1 to 6.

The Binomial distribution

Suppose we have a series if trials each of which has two outcomes, success S and failure F. We assume that the probability of success, p, is constant, so for every trial

$$P[\text{ Success}] = p \text{ and } P[\text{ failure }] = 1 - p$$

e the probability of X successes in n trails is given by

$$P[X = k] = \binom{n}{k} p^k (1-p)^{n-k} \qquad k = 0, 1, 2, \cdots n \tag{10.5}$$

- The mean is $E[X] = np$
- the variance is $\text{var}(X) = np(1-p)$

The probability that a person will survive a serious blood disease is 0.4. If 15 people have the disease the number of survivors X has a Binomial B(15,0.4) distribution.

- $P[X = 3] = \binom{15}{3}(0.4)^3(0.6)^{12}$
- $P[X \leq 8] = \sum_{x=0}^{8} \binom{15}{x}(0.4)^x(0.6)^{15-x}$
- $P[3 \leq X \leq 8] = P[X \leq 8] - P[X \leq 2] = \sum_{x=2}^{8} \binom{15}{x}(0.4)^x(0.6)^{15-x}$

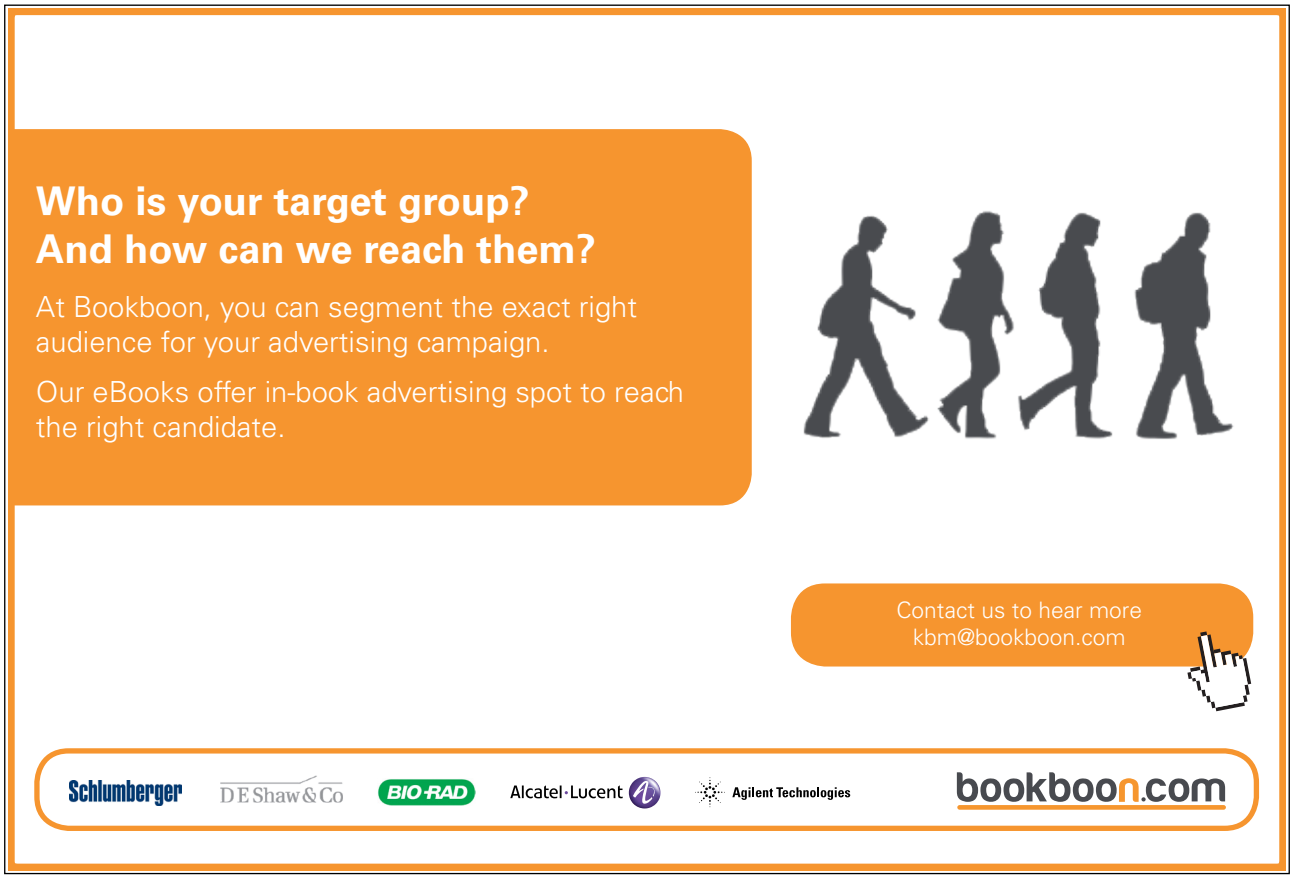
Download free eBooks at bookboon.com

Applying expectation using the Binomial

A more interesting use is:
Suppose we wish to test whether N people have a disease. It would seem that the only way to do this is to take a blood test, which will require N blood tests. Suppose we try the following:

1. We pool the blood of $k < N$ people.
2. If the combined sample is negative we have k people without the disease.
3. If the pooled test is positive we then test all k people individually, resulting in $k + 1$ tests in all.
4. Repeat until everyone is diagnosed

What does this save us?



Download free eBooks at bookboon.com
Click on the ad to read more

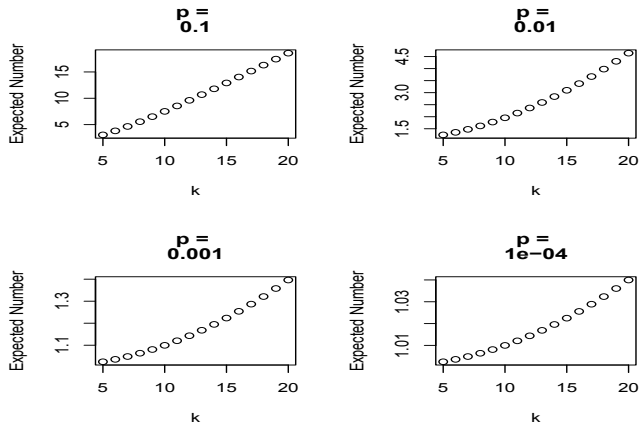
Assume the probability of a person having the disease is p and that we have a Binomial distribution for the number with the disease. Then for a group of k

1. $P[\text{ just 1 test}] = (1-p)^k$
2. $P[\text{ k+1 tests}] = 1 - P[\text{ just 1 test}] = 1-(1-p)^k$

So the expected number of tests is

$$E[\text{ no. of tests}] = (1-p)^k + (k+1)\left(1-(1-p)^k\right) = k\left(1-(1-p)^k\right)+1$$

This does give a considerable saving in the number of tests, see the diagram below



The Hypergeometric distribution

Suppose we have N items and D of these are defective. I take a sample of size n from these items, then the probability that this sample contains k defectives is

$$P[X=k] = \frac{\binom{N-D}{n-k}\binom{D}{k}}{\binom{N}{n}} \qquad k = 0,1,2,\cdots n \tag{10.6}$$

- The mean is $E[X] = n\frac{D}{N}$
- the variance is $\text{var}(X) = \frac{(N-n)}{(N-1)} n \frac{D}{N}\left(1-\frac{D}{N}\right)$

While situations involving the Hypergeometric are common ii common practice to approximate with the Binomial when N is large compared to D. We set $p = D/N$ and sue

$$P[X=k] = \binom{n}{k} p^k (1-p)^{N-k} \qquad k = 0,1,2,\cdots n$$

Download free eBooks at bookboon.com

The Poisson distribution

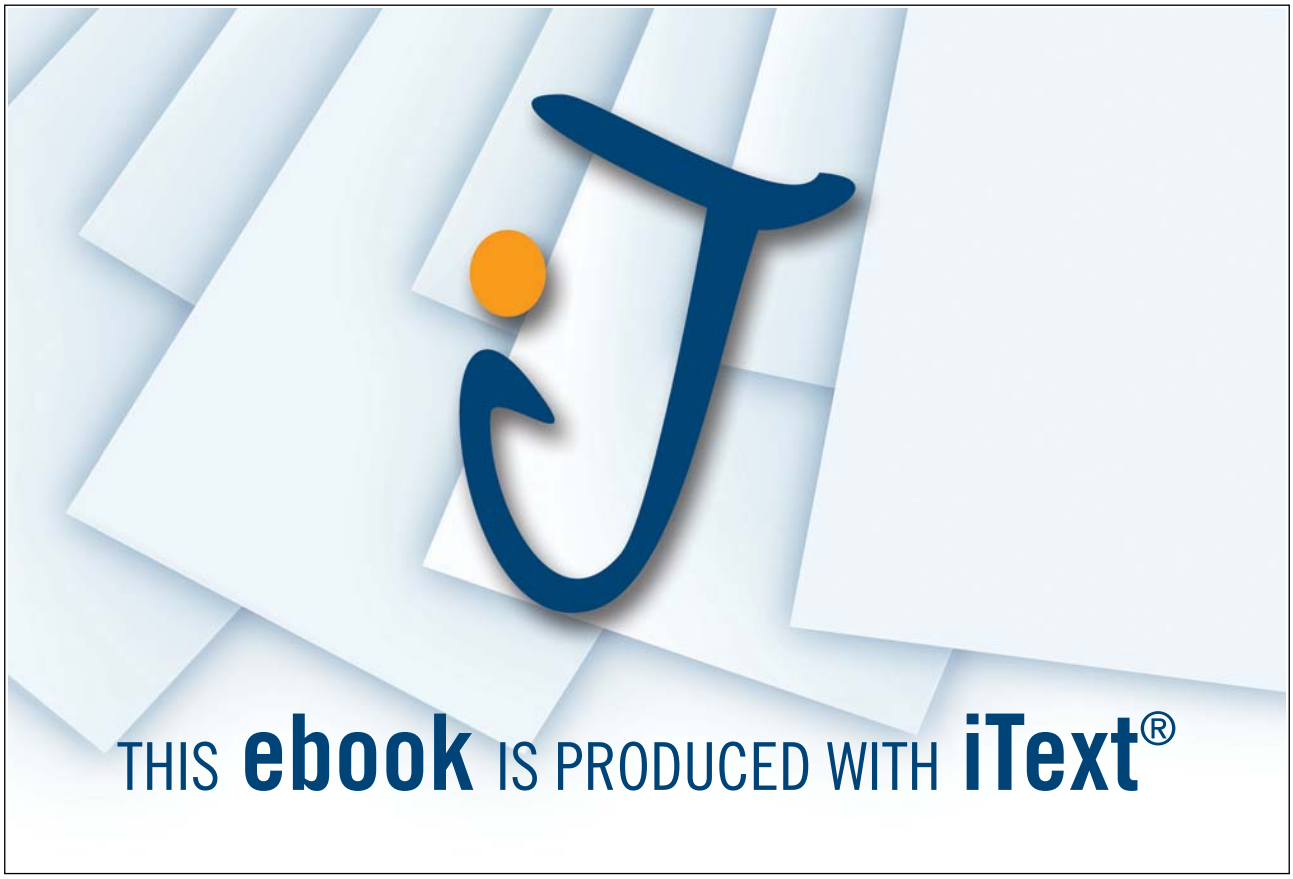
Suppose events occur at random

$$P[X = k] = \frac{\lambda^k e^{-\lambda}}{k!} \qquad k = 1, 2, \cdots, n \tag{10.7}$$

- The mean is $E[X] = \lambda$
- the variance is $\mathrm{var}(X) = \lambda$

The average number of oil tankers arriving per day at a port is 10. The facilities at the port can handle at most 15 arrivals in a day. What is the probability that the port will not be able to handle all the arrivals in a day? The variable X is Poisson $\lambda = 10$ so

$$P[X \geq 16] = \sum_{x=16}^{\infty} \frac{10^x}{x!} \exp(-10) = 1 - \sum_{x=0}^{15} \frac{10^x}{x!} \exp(-10) = 1 - 0.9513$$



Download free eBooks at bookboon.com

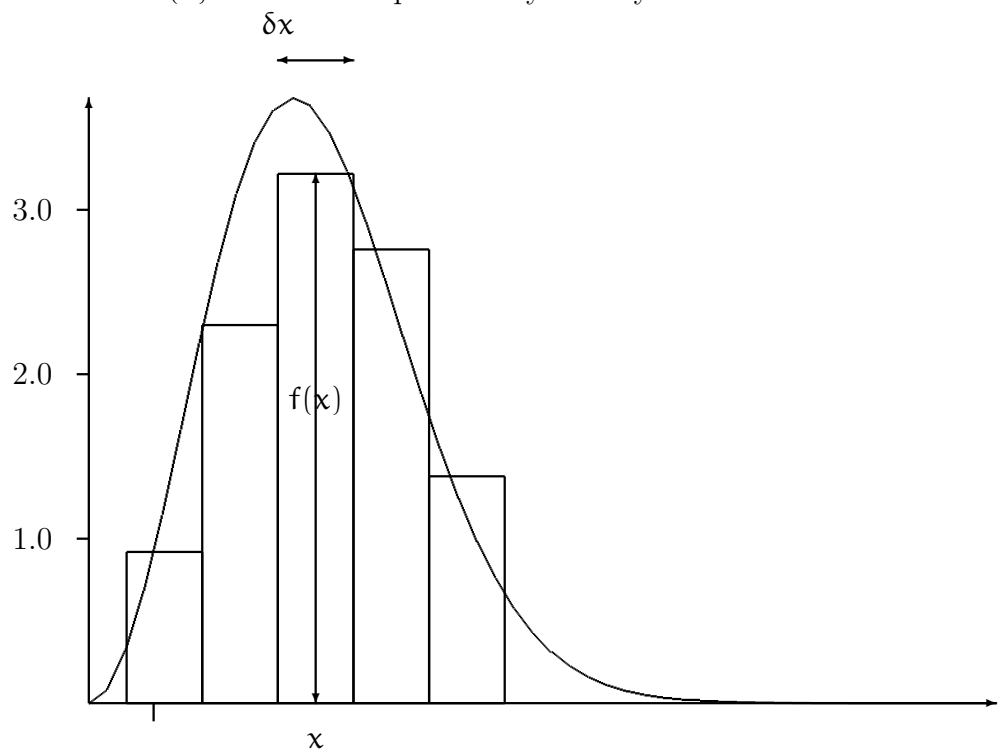
Click on the ad to read more

10.1.3 Continuous variables

All the cases we have considered so far have been where X takes discrete values. This does not have to be true - we can imagine X taking a continuous set of values. SInce we have though of a probability at X=k we might think of the probability of X being in some small interval $x, x + \delta x$ This probability will be

$$P[x \leq X \leq x + \delta x] = f(x)\delta x$$

The function $f(x)$ is called the probability density function.



The probability, as can be seen from the sketch is made up of boxes, and if we add these together we get a probability.

Personally I find it simpler to think of the cumulative distribution function $F(x)$ which is defined as $P[X \leq x] = F(x)$. This is just a probability and is what you find in tables. We relate this to the density function by

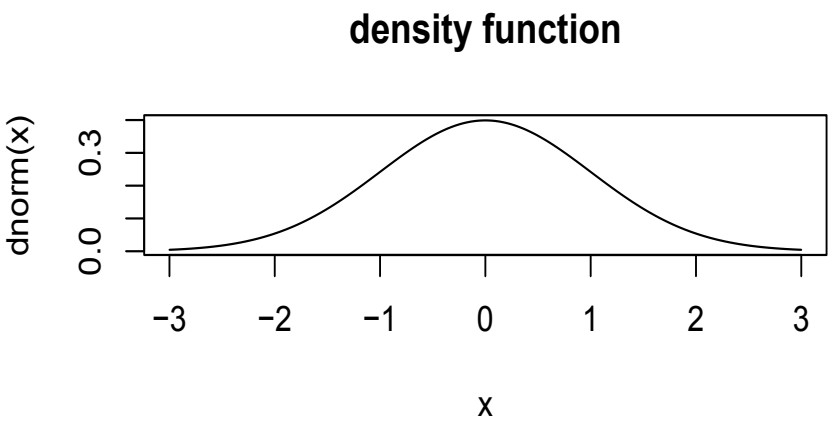
$$F(x) = \int_{-\infty}^{x} f(t)dt$$

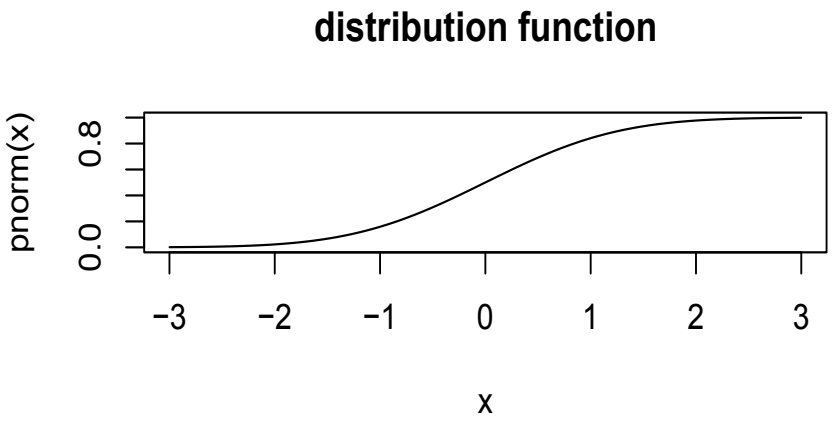
It is then not difficult to show that

$$P[a \leq X \leq b] = \int_{a}^{b} f(t)dt$$

Typical shapes are

Download free eBooks at bookboon.com





HIT YOUR EMPLOYEE RETENTION TARGETS

We help talent and learning & development teams hit their employee retention & development targets by improving the quality and focus of managers' coaching conversations.

Start improving employee retention & performance now.
Get your FREE reports and analysis on 10 of your staff today.

GET MY REPORTS

Download free eBooks at bookboon.com

Click on the ad to read more

10.1.4 Some Continuous Probability Distributions

Uniform Distribution

Here X is uniformly distributed on a range, say (a, b) so

$$f(x) = \frac{1}{b-a} \tag{10.8}$$

It follows that $F(x) = P[X \leq x] = \frac{x-a}{b-a}$ and $P[c \leq X \leq d] = \frac{d-c}{b-a}$

- The mean is $E[X] = \frac{a+b}{2}$
- the variance is $\text{var}(X) = \frac{1}{12}(b-a)^2$

This is a useful model for a random choice in he interval from$a$ to b.

Exponential Distribution

Here X is distributed on the range $(0, \infty)$ and

$$f(x) = \lambda \exp(-\lambda x) \tag{10.9}$$

where λ is a constant. It follows that $F(x) = P[X \leq x] = 1 - \exp(\lambda x)$ and $P[c \leq X \leq d] = \exp(\lambda c)1 - \exp(\lambda d)$

- The mean is $E[X] = \frac{1}{\lambda}$
- the variance is $\text{var}(X) = \frac{1}{\lambda^2}$

Normal Distribution

Here X is distributed on the range $(-\infty, \infty)$ and

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \tag{10.10}$$

where μ and σ are constants.

- The mean is $E[X] = \mu$
- the variance is $\text{var}(X) = \sigma^2$

The normal distribution crops up all over the place the problem is that there is no simple way of working out the probabilities. They can be computed but you either need the algorithm or tables.

Download free eBooks at bookboon.com

Normal Computation

Suppose X has a Normal distribution with mean μ and Variance σ^2, often written $N(\mu, \sigma^2)$. We can show that X is related to a *Standard Normal* variable z, that is z is *N(0,1)* by

$$z = \frac{X - \mu}{\sigma} \tag{10.11}$$

And of course we have the reverse

$$X = \mu + \sigma \times z \tag{10.12}$$

Now the standard normal is what is given in the tables do we convert our problem into a standard one.

1. Suppose X is $N(100, 9^2)$ Then
   (a) $P[X \leq 70] = P\left[z = \frac{X-100}{9} \leq \frac{70-100}{9} \leq -3.33\right] = 0.004$
   (b) $P[X \leq 95] = P\left[z = \frac{X-100}{9} \leq \frac{95-100}{9} \leq -5/9\right] = 0.2893$
   (c) $P[X \geq 109] = 1 - P\left[z = \frac{X-100}{9} \geq \frac{109-100}{9} \geq 1\right] = 1 - 0.2893$
   (d) $P[70 \leq X \leq 109] = P[X \geq 109] - P[X \leq 70] = 0.7017 - 0.004$

   Suppose we wish to find the value a so that $P[X \leq a] = 0.95$. Then

   $$P[X \geq a] = P[z = \frac{X - 100}{9} \geq z = \frac{a - 100}{9}] = 0.9$$

   From tables $\frac{a-100}{9} = 1.645$ and so $a = 100 + 1.645 \times 9$

Another example

Suppose we know

1. $P[X < 2] = 0.05$
2. $P[X > 14] = 0.25$

So we have

$$P[X < 2] = P[z = (X - \mu)/\sigma < (2 - \mu)/\sigma] = 0.05$$

and so from tables $(2 - \mu)/\sigma = -1.645$

We also have $P[X > 14] = 0.25$ or

$$P[X < 14] = P[z < (X - \mu)/\sigma < (X - \mu)/\sigma] = 1 - 0.25 = 0.975$$

Hence $(14 - \mu)/\sigma = 1.96$ We have a pair of equations

Download free eBooks at bookboon.com

1. $2 - \mu = -\sigma \times 1.645$
2. $14 - \mu = \sigma \times 1.96$

Solving gives

$$(14 - \mu) - (2 - \mu) = 12 = 0.315\sigma$$

or $\sigma = 3.32871$ and so $\mu = 7.475728$

The Normal approximation to the Binomial

A Binomial variable X which is $B(n, p)$ can be approximated by a Normal variable Y, mean np, variance $np(1-p)$. This can be very useful as the Binomial tables provided are not very extensive. *This is known as the Normal approximation to the Binomial.*

In this case

$$z = (Y - np)/\sqrt{(np(1-p))}$$

is standard Normal.

Example

Suppose X is number of 6's in 40 rolls of a die. Let Y be $N(\frac{40}{6}, \frac{40}{6}\frac{5}{6})$. Then

$$P[X < 5] \simeq P[Y < 5] = P[z < \frac{5 - 20/3}{\sqrt{50/9}}] = \Phi(-0.7071068) = 0.2398$$

You can refine this approximation but we will settle for this at the moment.

Exercises

1. A die is rolled, what is the probability that
   (a) The outcome is even.
   (b) The outcome is a prime.
   (c) The outcome exceeds 2.
   (d) The outcome is -1.
   (e) The outcome is less than 12.
2. Two dice are rolled. What is the probability that
   (a) The sum of the upturned faces is 7?
   (b) The score on one die is exactly twice the score on the other.

Download free eBooks at bookboon.com

(c) You throw a double, that is the dice each have the same score.

3. Suppose we toss a coin 3 times. Find the probability distribution of

   (a) X=the number of tails.

   (b) Y = the number of runs. Here a run is a string of heads or tails. So for HTT Y=2.

4. The student population in the Maths department at the University of San Diego was made up as follows

   - 10% were from California
   - 6% were of Spanish origin
   - 2% were from California and of Spanish origin.

   If a student from the class was to be drawn at random what is the probability that they are

   (a) From California or of Spanish origin.

   (b) Neither from California nor of Spanish origin.

   (c) Of Spanish origin but not from California

Download free eBooks at bookboon.com
Click on the ad to read more

5. For two events A and B the following probabilities are known

$$P[A] = 0.52 \qquad P[B] = 0.36 \qquad P[A \cup B] = 0.68$$

Determine the probabilities

(a) $P[A \cap B]$

(b) $P[\sim A]$

(c) $P[\sim B]$

6. A hospital trust classifies a group of middle aged men according to body weight and the incidence of hypertension. The results are given in the table.

	Overweight	Normal Weight	Underweight	Total
Hypertensive	0.10	0.08	0.02	0.20
Not Hypertensive	0.15	0.45	0.20	0.80
Total	0.25	0.53	0.22	1.00

(a) What is the probability that a person selected at random from this group will have hypertension?

(b) A person selected at random from this group is found to be overweight, what is the probability that this person is also hypertensive?

(c) Find $P[\text{hypertensive} \cup \text{Underweight}]$

(d) Find $P[\text{hypertensive} \cup \text{Not Underweight}]$

7. Two cards are drawn from an ordinary deck of 52 cards. What is the probability of drawing

(a) Two aces.

(b) The two black aces.

(c) Two cards from the court cards K,Q,J

8. Five cards are drawn from a deck of cards. What is the chance that

(a) Four cards are aces

(b) Four cards are the same i.e. 4 10's, 4 9'2 etc.

(c) All the cards are of the same suit.

(d) All the card are of the same suit and are in sequence.

Download free eBooks at bookboon.com

9. A student of statistics was told that there was a chance of 1 in a million that there was a bomb on an aircraft. The reasoned that there would be a one in 10^{12} chance of being two bombs on a plane. He thus decided that he should take a bomb with him (defused - he was not stupid) to reduce the odds of an explosion.

   Assuming no security problems is this a sensible strategy?

10. There are four tickets numbered 1,2,3,4. A two digit number is formed by drawing a ticket at random from the four and a second from the remaining three. So if the tickets were 4 and 1 the resulting number would be 41. What is the probability that

    (a) The resulting number is even.

    (b) The resulting number exceeds 20

    (c) The resulting number is between 22 and 30.

11. Three production lines contribute to the total pool of parts used by a company.

    - Line 1 contributes 20% and 15% of items are defective.
    - Line 2 contributes 50% and 5% of items are defective.
    - Line 3 contributes 30% and 6% of items are defective.

    (a) What percentage of items in the pool are defective?

    (b) Suppose an item was selected at random and found to be defective, what is the probability that it came from line 1?

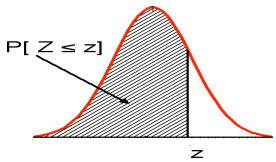
    (c) Suppose an item was selected at random and found not to be defective, what is the probability that it came from line 1?

Download free eBooks at bookboon.com

10.2 The Normal distribution

This table gives the cumulative probabilities for the standard normal distribution, that is

$$P[Z \leq z] = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)dx$$



This is the shaded area in the figure.

z	0.00	-0.01	-0.02	-0.03	-0.04	-0.05	-0.06	-0.7	-0.08	-0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.0	0.5000	-	-	-	-	-	-	-	-	-

Download free eBooks at bookboon.com

This table gives the cumulative probabilities for the standard normal distribution, that is

$$P[Z \leq z] = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)\,dx$$

This is the shaded area in the figure.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Download free eBooks at bookboon.com

Chapter 11

Looking at Data

It is very much more difficult to handle data rather than to construct nice probability arguments. We begin by considering the problems of handling data. The first questions are the provenance of the data.

- Is it reliable?
- Who collected it?
- Is it what it is said to be?
- Is it a sample and from what population?

Such questions are important because *if the data is wrong no amount of statistical theory will make it better.* Collecting your own data is the best as you should know what is going on. Almost all statistical theory is based on the assumption that the observations are independent and in consequence there is a large body of methodology on sampling and data collection.

11.1 Looking at data

Once you have the data what is he next step? If it is presented as a table (do read the description) it may well be worth reordering the table and normalising the entries. Simplifying and rounding can be very effective, especially in reports. After gathering data, it pays to look at the data in as many ways as possible. Any unusual or interesting patterns in the data should be flagged for further investigation.

The Histogram

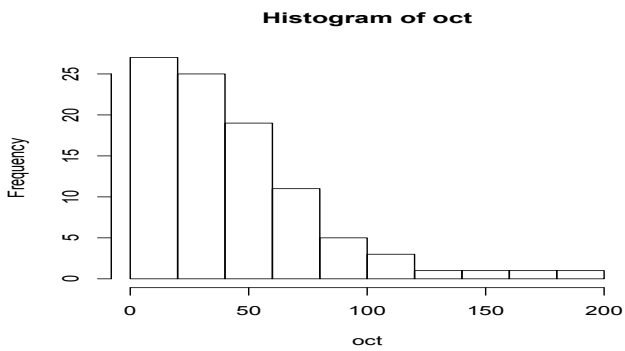
Anyone who does not draw a picture of their data deserves all the problems that they will undoubtedly encounter. The basic picture is the histogram. For the histogram we split the range of the data into intervals and count the number of observations in each

Download free eBooks at bookboon.com

interval. We then construct a diagram made up of rectangles erected on each interval. The *area* of the rectangle being proportional to the count.

110	190	55	65	43	15	40	32
11	44	76	23	28	12	15	57
19	63	70	12	17	33	49	16
150	29	18	21	60	43	23	36
22	11	26	29	82	6	21	64
84	73	54	44	82	16	95	29
30	27	85	35	5	22	52	19
18	175	10	20	29	16	16	20
17	6	47	130	115	37	50	17
41	61	116	55	67	26	51	9
50	73	43	80	52	17	22	28
8	27	32	75	10	45		

Table 11.1: Dorsal lengths of octapods



11.1.1 Summary Statistics

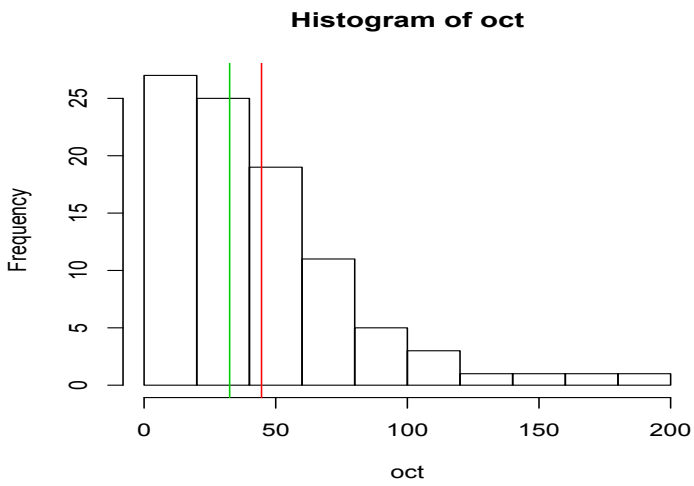
Location This is often called the "measure of central tendency" in our textbooks, or the "centre" of the dataset in other sources. Common measures of location are the mean and median. Less common measures are the mode and the truncated mean. Given observations $x_1, x_2, \ldots, x_n$

- The sample mean is just $\frac{1}{n}\sum_{i=1}^{n} x_i$ written $\bar{x}$. For the Octopods it is 44.67021.
- The median is the middle value, we arrange the observations in order and if n is odd pick the middle one. If n is even then we take the average of the two middle values. For the Octopods it is 32.5

Download free eBooks at bookboon.com

- A truncated mean is the mean of a data set where some large or small (or both) observations have been deleted.

As you might expect the median is much less influenced by outliers - it is a robust estimate.



Example

The Australian Bureau of Meteorology collects data on rainfall across Australia. Given below is the mean monthly rainfall in Broken Hill as well as the median monthly rainfall.

Average Monthly Rainfall in Broken Hill (in millimeters) 1900 to 1990

Month	Mean	Median
Jan	23	9
Feb	24	10
Mar	18	9
Apr	19	9
May	22	13
Jun	22	15
Jul	17	15
Aug	19	17
Sep	20	12
Oct	25	15
Nov	19	10
Dec	20	7

Download free eBooks at bookboon.com

(a) Note that the median monthly rainfall is January is much smaller than the mean monthly rainfall. What does this imply about the shape of the distribution of the rainfall data for the month of January?

(b) Which measure of central tendency, the mean or the median, is more appropriate for describing rainfall in Broken Hill? Justify your answer using knowledge of mean and median.

(c) Use the above table to calculate the total yearly rainfall for Broken Hill.

(d) In the north of Australia, the wet season occurs from November to April. Broken Hill, in central Australia, is occasionally drenched by a northern storm during these months. These storms tend drop a large amount of rain in a comparatively short time. How does the table reflect this fact?

Spread This is the amount of variation in the data. Common measures of spread are the sample variance, standard deviation and the interquartile range. Less common is the range. The traditional measure is the *sample variance*

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

and the square root of the sample variance known as the *standard deviation*

$$s^2 = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

For the octopods s=36.06159. Alternatives are:

The range This is defined as
range = largest data value - smallest data value
this is obviously not very robust and hence is not often used which is a shame.

Interquartile Range The interquartile range Q3-Q1, while simple in concept, has caused much grief to introductory statistics teachers since different respectable sources define it in different respectable ways! First we find the lower quartile Q1, this is the $k = (n/4)$th of the ordered observations. If k is not an integer we take the integer part of k plus 1 otherwise we take $k+1$. The upper quartile Q3 is obtained by counting down from the upper end of the ordered sample. This is a good robust measure of spread. For the Octopods Q3-Q1= 59.25 -19.00 = 40.25.

Download free eBooks at bookboon.com

Shape

The shape of a dataset is commonly categorized as symmetric, left-skewed, right-skewed or bi-modal. The shape is an important factor informing the decisions on the best measure of location and spread. There are several summary measures. The sample third moment

$$\kappa_3 = \frac{1}{ns^3}\sum_{i=1}^{n}(x_i - \bar{x})^3$$

measures skewness-it is zero for a symmetric distribution. The fourth moment

$$\kappa_4 = \frac{1}{ns^4}\sum_{i=1}^{n}(x_i - \bar{x})^3$$

gives a flat top measure. It is 3 for a normal variable!

Outliers

Outliers are data values that lie away from the general cluster of other data values. Each outlier needs to be examined to determine if it represents a possible value from the population being studied, in which case it should be retained, or if it is non-representative (or an error) in which case it can be excluded. It may be that an outlier is the most important feature of a dataset. It is said that the ozone hole above the South Pole had been detected by a satellite years before it was detected by ground-based observations, but the values were tossed out by a computer program because they were smaller than were thought possible.

Clustering

Clustering implies that the data tends to bunch up around certain values.

Granularity

Granularity implies that only certain discrete values are allowed, e.g. a company may only pay salaries in multiples of £1,000. A dotplot shows granularity as stacks of dots separated by gaps. Data that is discrete often shows granularity because of its discreteness. Continuous data can show granularity if the data is rounded.

11.1.2 Diagrams

There is much to be said for drawing pictures. It is hard to imagine a data set where a histogram is not useful. If your computer program does not draw pictures then replace it! I rather like to smooth the histogram to get an idea of the shape of the p.d.f.

Note however we need to take care even with the humble histogram! Ideally a histogram should show the shape of the distribution of the data. For some datasets but

Download free eBooks at bookboon.com

the choice of bin width can have a profound effect on how the histogram displays the data.

Stem and Leaf charts

If you are in a computer-free environment a stem-and-leaf plot can be a quick an effective way of drawing up such a chart. Consider the data below

27	28	29	30	31	32	33	34	35	36
37	38	39	40	41	42	43	44	45	46
47	48	49	50	51	52	53	54	55	56

```
stem   leaves        freq  cum freq

   2   789             3       3
   3   0123456789      10      13
   4   0123456789      10      23
   5   0123456          7      30
```

Such a stem and leaf chart is valuable in giving an approximate histogram and giving the basis for some interesting data summaries. As you can see it is fairly easy to find the median, range etc. from the stem and leaf chart.

Dotplots

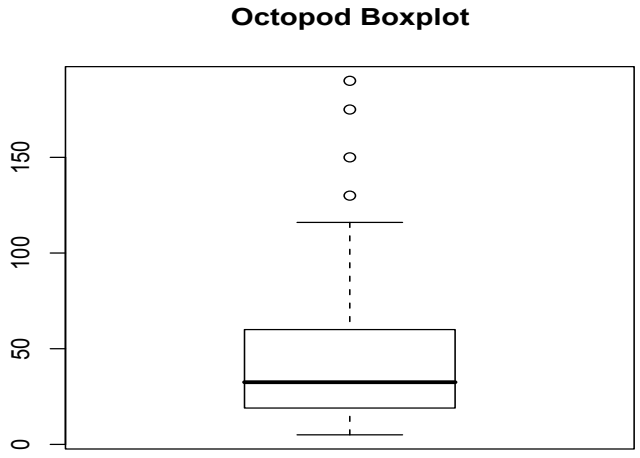
A traditional dotplot resembles a stemplot lying on its back, with dots replacing the values on the leaves. It does a good job of displaying the shape, location and spread of the distribution, as well as showing evidence of clusters, granularity and outliers. And for smallish datasets a dotplot is easy to construct, so the dotplot is a particularly valuable tool for the statistics student who is working without technology.

Box-Plots

Another useful picture is the box plot. Here we mark the quartiles Q1 Q2 on an axis and draw a box whose ends are at these points. The ends of the vertical lines or "whiskers" indicate the minimum and maximum data values, unless outliers are present in which case the whiskers extend to a maximum of 1.5 times the inter-quartile range. The points outside the ends of the whiskers are outliers or suspected outliers. can be very useful, especially when making comparisons.

One drawback of boxplots is that they tend to emphasize the tails of a distribution, which are the least certain points in the data set. They also hide many of the details of the distribution. Displaying a histogram in conjunction with the boxplot helps. Both are important tools for exploratory data analysis.

Download free eBooks at bookboon.com



11.2 Scatter Diagram

A common diagram is the scatter diagram where we plot x values against y values. We illustrate the ideas with two examples.

Breast cancer

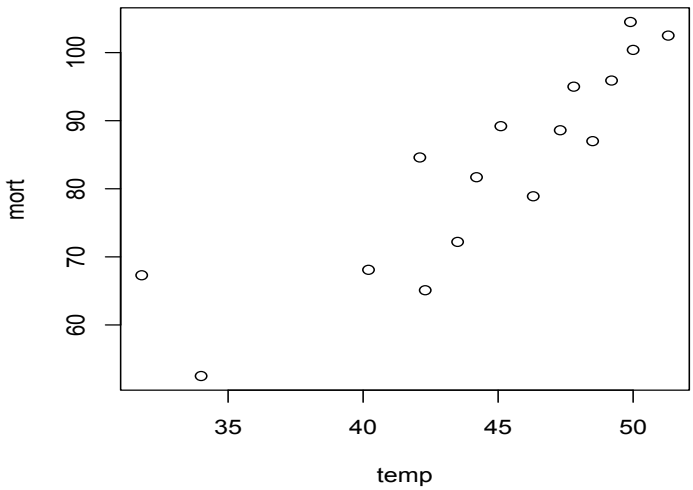
In a 1965 report, Lea discussed the relationship between mean annual temperature and the mortality rate for a type of breast cancer in women. The subjects were residents of certain regions of Great Britain, Norway, and Sweden. A simple regression of mortality index on temperature shows a strong positive relationship between the two variables.

Data

Data contains the mean annual temperature (in degrees F) and Mortality Index for neoplasms of the female breast. Data were taken from certain regions of Great Britain, Norway, and Sweden. Number of cases: 16 Variable Names

1. Mortality: Mortality index for neoplasms of the female breast

2. Temperature: Mean annual temperature (in degrees F) The Data:

Download free eBooks at bookboon.com

Mortality	Temperature
102.5	51.3
104.5	49.9
100.4	50
95.9	49.2
87	48.5
95	47.8
88.6	47.3
89.2	45.1
78.9	46.3
84.6	42.1
81.7	44.2
72.2	43.5
65.1	42.3
68.1	40.2
67.3	31.8
52.5	34