INTRODUCTION

Mathematics Refresher: Part 0

Prof. John Polak
Centre for Transport Studies
Department of Civil and Environmental Engineering
Imperial College London
London SW7 2AZ

September 2005

INTRODUCTION

The Masters of Science courses in Transport are quantitatively oriented courses that require students to have some familiarity with certain basic mathematical concepts and notation. The aim of the Mathematics Refresher is to briefly review this material in order to help your preparation for the rest of the course.

These notes accompany the Mathematics Refresher, which will be presented in five parts, during the first two weeks of the course.

BACKGROUND

Transport is a quantitative discipline – it uses quantitative data, evidence and arguments to draw conclusions about transport problems and their solution. Mathematical reasoning and argument is therefore an important part of the work of many transport professionals and therefore is an important part of the course.

However, the course is *not* a course in mathematics or statistics and it certainly does *not* require that you are a mathematical wizard. Students on the course come from a very wide range of (mathematical) backgrounds and many of the most successful students come from 'softer' backgrounds

All of you will need to have *some* familiarity with key concepts and methods covered in the maths refresher during the core modules (T1-T6 + BM1-4 or SD1-2) but you will not necessarily require a detailed understanding of all the material.

Those of you who want to study certain subjects (e.g., transport demand modelling, traffic flow theory) in greater mathematical detail can do so in specialist option modules in the spring term. Conversely, there are plenty of specialist option modules for those of you who want to study less mathematical aspects of transport

SCOPE

The maths refresher aims to cover a number of topics, which are important for other parts of the courses.

- The idea of a variable, operators that can be applied to a variable and how variables are combined and manipulated in mathematical expressions and equations

- The idea that a variable can be regarded as a function of one or more other variables

- The interpretation and use of the concepts of the derivative and the integral of a function

- The concept of probability and the notion of a random variable and the probability distribution of a random variable

ADVICE

If you find you are having difficulties with the mathematical elements of the course, either in the maths refresher or later on, then here is some good advice.

- DON'T panic or over-react but also DON'T do nothing…

- If you don't understand a particular part of a lecture then *ask* the *relevant* lecturer (either in class, during the break, after class or make an appointment – by email, don't just turn up and expect to be seen) to go over it again.

- Don't be embarrassed to ask questions in lectures – we won't mind and in all probability other students will be grateful that you did.

- Work thoroughly and consistently through the class exercises and coursework – think about forming small work groups to help this (we encourage students to work together in this way *but* submitted coursework must be your own work)

BOOKS

There are number of useful books that cover the material in the maths refresher. Our current recommendation for background reading to accompany the maths refresher is:

Croft, A. and Davison, R. (2003) *Foundation Mathematics*, Prentice Hall, Harlow.

The course reading list contains details on a number of other, more advanced, texts.

NUMBERS, FACTORIALS, VECTORS AND MATRICES, SERIES AND LIMITS

Mathematics Refresher: Part 1

Prof. John Polak
Centre for Transport Studies
Department of Civil and Environmental Engineering
Imperial College London
London SW7 2AZ

September 2005

NUMBERS, COMBINATIONS OF NUMBERS AND FACTORIALS

The Number Line

The concept of number is perhaps the most fundamental in mathematics. One of the easiest ways of envisaging numbers is by means of a number line. The simplest such line is shown below. It is defined by an arbitrarily chosen point called zero and a series of equally spaced points in each direction representing the other *positive and negative integers.* Zero is regarded as both positive and negative. Any other real number x is represented as a point distant x from the origin (the unit of distance being the distance between adjacent integer points).

-6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6

Comparison Operators

The following are common relationships between two real numbers x and y

$$x = y \qquad x \neq y \qquad x < y \qquad x \leq y \qquad x \geq y \qquad x > y$$

If $x < y$, the set of numbers z such that $x < z < y$ is called the *open interval* (x,y). The set of numbers such that $x \leq z \leq y$ is called the *closed interval* $[x,y]$.

Simple Arithmetic Operators

Two numbers x and y can be combined to form a third number as follows:

$$x + y \qquad x - y \qquad xy \qquad \text{and} \qquad x/y \;\; (\text{provided } y \neq 0)$$

Numbers of the form m/n where m and n are integers are called *rational numbers*. Most real numbers are not rational numbers, but any real number can be approximated arbitrarily closely by a rational number.

Exponential Operations

A number x multiplied by itself n-1 times where n is an integer ≥ 1, is *raised to the power n*, written x^n; this definition extends to $n \leq 0$ when multiplication –1 times is interpreted as division.

Any real number y such that $y^n = x$ where n is a positive integer, is called the n^{th} root of x. We write $y = x^{1/n}$. If n is even, there are two equal and opposite real n^{th} roots of x; if n is odd there is only one. Roots of rational numbers are usually not rational.

$x^{1/2}$ is called the square root of x and is often written as $\sqrt{x}$. The n^{th} root of x is sometimes written as $\sqrt[n]{x}$

The result of raising $x^{1/n}$ to the power m is written $x^{m/n}$, thus defining x^r for all rational r, and this definition can be extended to define x^y for all real y. Powers combine as follows:

$x^y x^z = x^{y+z}$ $x^y/x^z = x^{y-z}$ $(x^y)^z = x^{xz}$

Unitary Operators

The magnitude of a number x regardless of its sign is called its modulus, $|x|$ i.e.,

$|x| = x$ if $x > 0$
$|x| = -x$ if $x \leq 0$

Precedence of Operators

Algebraic expressions often denote several pairwise combinations of numbers. Such expressions are unambiguous only because of the following strict convention regarding the order in which operations are carried out in combining adjacent numbers:

1. raising to a power
2. multiplication
3. division
4. addition and subtraction

Any departure from this convention is indicated by the use of brackets. The expression inside a pair of brackets is reduced to a single number before being combined with an adjacent number.

Factorials

In many areas of mathematics there arises a need to think in terms of rearrangements of a set of objects. The number of ways in which n different objects can be rearranged in a row is

n x (n-1) x (n-2) x …x 3 x 2 x 1.

This quantity is called *factorial n* and is denoted by n!

Summations and Products

A common requirement is to work with quantities that are the summation or product of other quantities and we often use a special notation to simplify expressions involving such sums or products. Consider a set of variables $\{x_1, x_2, ...x_n\}$ then we use the symbol Σ and Π as a shorthand for summation and multiplication respectively, as follows:

$$x_1 + x_2 + ... + x_n = \sum_{i=1}^{n} x_i$$

and

$$x_1 \times x_2 \times ... \times x_n = \prod_{i=1}^{n} x_i$$

These conventions are used extensively in the course.

VECTORS AND MATRICES

It is often very useful to be able to manipulate numbers as structured groups, rather than individually. It turns out that not only can this significantly simplify notation and manipulation, it also enables us to derive a whole range of interesting and useful results that might not otherwise be available. The two most common structures are the *matrix* and the *vector*. Vectors can be regarded as special cases of matrices.

A matrix is a rectangular array of numbers. A typical matrix will have I rows and J columns ($I,J \geq 1$) and is written as follows:

$$\boldsymbol{B} = \begin{pmatrix} b_{11} & b_{12} & . & . & b_{1J} \\ b_{21} & b_{22} & . & . & b_{2J} \\ . & . & . & . & . \\ . & . & . & . & . \\ b_{I1} & b_{I2} & . & . & b_{IJ} \end{pmatrix} = (b_{\mathrm{ij}})$$

The matrix (λb_{ij}) where λ is a real number is denoted $\lambda\boldsymbol{B}$.

An I x J matrix $\boldsymbol{D} = (d_{ij})$ is called a *diagonal matrix* if $d_{ij} = 0$ unless $i=j$

The J x J *unit matrix* is the J x J diagonal matrix with each diagonal element equal to 1. It is usually denoted by I_J.

A vector is a special case of a matrix in which there is either only one row or only one column, and it is important to be clear which.

For example the vector $\boldsymbol{a}$

$$\boldsymbol{a} = (a_1,a_2,\ldots,a_M)$$

is a 1 x M matrix and is called a *row vector*, whereas the vector $\boldsymbol{c}$

$$\boldsymbol{c} = \begin{pmatrix} c_1 \\ c_2 \\ . \\ . \\ c_M \end{pmatrix}$$

is an M x 1 matrix and is called a *column vector.*

There is no clear convention about this, so care is needed when using vectors in reading and writing. Whichever way round is chosen in a particular piece of mathematics, it is best to be consistent, i.e. either define all vectors as row vectors or column vectors.

Matrix Transposition

If we interchange the rows and columns of the vector ***B*** above, we obtain a *J* x *I* matrix called the transpose $\boldsymbol{B}^{\mathrm{T}}$ of ***B*** (the transpose is also sometime written **B′**).

$$\boldsymbol{B}^{\mathrm{T}} = \begin{pmatrix} b_{11} & b_{21} & . & . & b_{I1} \\ b_{12} & b_{22} & . & . & b_{I2} \\ . & . & . & . & . \\ . & . & . & . & . \\ b_{1J} & b_{2J} & . & . & b_{IJ} \end{pmatrix} = (b^{T}_{\mathrm{ji}})$$

where, $b^{T}_{\mathrm{ji}} = b_{\mathrm{ij}}$ $(1 \le i \le I,\ 1 \le j \le J)$.

Notice that if ***a*** is a row vector then $\boldsymbol{a}^{\mathrm{T}}$ is a column vector and vice versa.

A square matrix ***S*** such that $\boldsymbol{S}^{\mathrm{T}} = \boldsymbol{S}$ is called *symmetric*.

Matrix Addition and Subtraction

The operations of addition and subtraction generalise in a straightforward manner to matrices.

If $P = (p_{ij})$ and $Q = (q_{ij})$ are two *I* x *J* matrices then $P+Q$ is the *I* x *J* matrix $(p_{ij} + q_{ij})$ and $P\text{-}Q$ is the *I* x *J* matrix $(p_{ij} - q_{ij})$.

Matrix Multiplication

It is also possible to generalise the operation of multiplication to matrices, although it is not so straightforward. If $U = (u_{ij})$ is an *I* x *J* matrix and $V = (v_{ij})$ is an *J* x *K* matrix, then the matrix product ***UV*** is the *I* x *K* matrix (w_{ik}) where

$$w_{ik} = \sum_{j-1}^{J} u_{ij} v_{jk}$$

Notice that the product ***VU*** does not exist at all unless *I*=*K* and even then $\boldsymbol{UV} \neq \boldsymbol{VU}$ (in general).

It follows from the above definitions that

$$(\lambda\boldsymbol{P})^{\mathrm{T}} = \lambda\boldsymbol{P}^{\mathrm{T}}$$

$$(\boldsymbol{P}+\boldsymbol{Q})^{\mathrm{T}} = \boldsymbol{P}^{\mathrm{T}} + \boldsymbol{Q}^{\mathrm{T}}$$

$$\lambda(\boldsymbol{P}+\boldsymbol{Q}) = \lambda\boldsymbol{P} + \lambda\boldsymbol{Q}$$

$$(\lambda\boldsymbol{U})\boldsymbol{V} = \boldsymbol{U}(\lambda\boldsymbol{V}) = \lambda(\boldsymbol{UV})$$

and $(\boldsymbol{UV})^{\mathrm{T}} = \boldsymbol{V}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}}$

The Inverse of a Square Matrix

The notion of the inverse of a matrix is extremely important since it arises in an enormous range of different mathematical contexts including the solution of systems of linear equations, which you will study later in this refresher course.
If $\boldsymbol{P}$ is an J x J matrix, then there may or may not exist another J x J matrix $\boldsymbol{P}^{-1}$ such that

$$\boldsymbol{P}\boldsymbol{P}^{-1} = \boldsymbol{P}^{-1}\boldsymbol{P} = I_J$$

If so, then $\boldsymbol{P}^{-1}$ is called the inverse of $\boldsymbol{P}$ and $\boldsymbol{P}$ is said to be non-singular (or invertable).

SUM OF SERIES

A series is a set of numbers (a_1, a_2, a_3,…,a_n,…,a_N) in which there is a particular relationship between successive terms. This relationship may be arithmetic or geometric or a combination of both.

There are many contexts in which we are interested in computing the value of the sum of the terms in a series. When a_n can be expressed as a simple function of n, it is sometimes possible to also express the sum $S_N = \sum_{n=1}^{N} a_n$ as a simple function of N. For example,

If $a_n = n$ then $S_N = \frac{1}{2}N(N+1)$

If $a_n = x^n$ then $S_N = \frac{x(1-x^N)}{(1-x)}$, provided $x \neq 1$

If $a_n \geq 0$ for all n and instead of stopping after N terms we keep on summing indefinitely, then either the sum increases indefinitely and the series is said to *diverge* or the sum approaches some finite limit called the sum to *infinity of the series*, denoted by $S_\infty = \sum_{n=1}^{\infty} a_n$ and the series is said to *converge*. In the second example above ($a_n = x^n$), it can be shown that the series diverges if $x > 1$ but converges if $0 < x < 1$ with $S_\infty = x/(1-x)$.

LIMITS AS $N \to \infty$

An expression in N can do one of the following as $N \to \infty$.

tend to ∞	(e.g. N)
tend to - ∞	(e.g. $-N$)
oscillate infinitely	(e.g. $(-1)^N N$)
oscillate between finite limits	(e.g. $(-1)^N$)
tend to a finite limit	(e.g. 1 + 1/N)

Acknowledgement
These notes draw upon material in Richard Allsop's Mathematics Checklist document. Any errors present here are, of course, the responsibility of the author.

LINEAR EQUATIONS, CARTESIAN COORDINATES, ANGLES AND TRIGONOMETRIC FUNCTIONS

Mathematics Refresher: Part 2

Prof. BG Heydecker
Centre for Transport Studies
Department of Civil and Environmental Engineering
University College London
London WC1E 6BT

August 2005

Linear equations

Equations are used to determine one or more unknown quantities from knowledge of other quantities.

Linear equations contain only known quantities and known multiples of unknown quantities. They do not contain powers, roots, products or any more complicated expressions in the unknown quantities, nor do they contain ratios with unknown quantities in the denominator.

If there is only one unknown x , the linear equation for x takes the form

$$ax = b \ .$$

If a ... 0 , this has just one solution: $x = b/a$

If $a = 0$, then either b ... 0 and there is no solution
or $b = 0$ and any x is a solution.

The case $a = 0$ is trivial, but has non-trivial counterparts in the case where there are several unknowns
$x_1 , x_2 , \ldots , x_n$.

Let $\mathbf{x} = (x_1 , x_2 , \ldots , x_n)$.

In general there will also be several equations. Let there be m equations, viz

$$\mathbf{A} \, \mathbf{x} = \mathbf{b}$$

where $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{b}$ is a column vector of m elements.

If $m > 1$ the equations may be inconsistent, in which case there is no solution. This can be checked by examining $\mathbf{A}$ and $\mathbf{b}$. In the case that they are consistent, if:

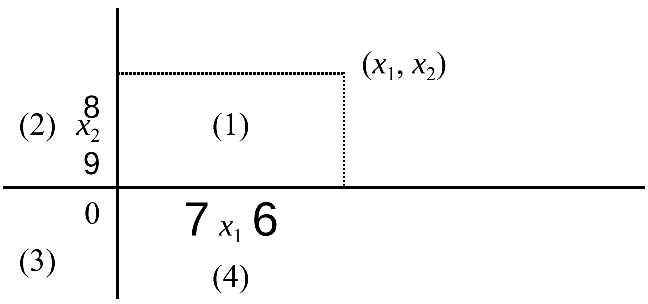
$m < n$ then the equations have a whole range of solutions,

$m = n$ then they have a unique solution: $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ if $\mathbf{A}^{-1}$ exists, and a whole range of solutions if $\mathbf{A}^{-1}$ does not exist, or

$m > n$ then they have a unique solution if some n of them have a matrix whose inverse exists, and a whole range of solutions if not.

A whole range of solutions exists if and only if non-zero solutions of $\mathbf{A} \, \mathbf{x} = \mathbf{0}$ exist, and the latter together with any one solution of $\mathbf{A} \, \mathbf{x}^{\mathrm{T}} = \mathbf{b}^{\mathrm{T}}$ determine all such solutions.

Cartesian co-ordinates



Given an origin O and the directions and scales of two perpendicular axes in the plane, every point in the plane corresponds to a pair of real numbers (x_1, x_2) as shown, and vice versa.

The orientation of the axes shown is the conventional one, and the quadrants are then numbered 1, 2, 3 and 4 as shown.

The set of pairs satisfying an equation of the form

$$a_1 x_1 + a_2 x_2 = 0$$

is the line through both the origin and the point $(-a_2, a_1)$, which has slope $-a_1/a_2$, unless $a_2 = 0$ in which case it is the x_2 axis. If the slope is positive, the line lies in quadrants 1 and 3, if negative in quadrants 2 and 4. If the slope is small the line is near to the x_1 axis $(x_2 = 0)$, and if large, near to the x_2 axis $(x_1 = 0)$.

The set of pairs satisfying

$$a_1 x_1 + a_2 x_2 = b$$

is a line parallel to this, passing at a distance $|b| / (a_1^2 + a_2^2)^{1/2}$ from the origin.

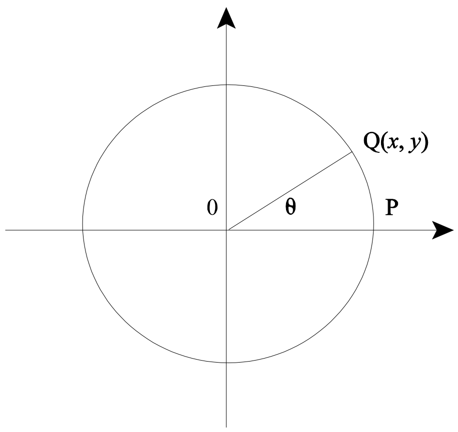
Cartesian co-ordinates can also be used in 3 or more dimensions.

In 3 dimensions, the conventional orientation of the x_3 axis is out of this side of the paper in the above diagram.

More than 3 dimensions cannot be envisaged physically, but are dealt with algebraically in just the same way as 1, 2 and 3 dimensions.

Angles and trigonometric functions

Consider a circle of unit radius centred on the origin in the (x,y) plane.



The area of the circle is denoted by π, which is approximately 3.1416 . The circumference is 2π . Let P be the point where the circle meets the positive part of the x axis.

For a typical point Q on the circumference, with co-ordinates (x, y), the angle θ between OQ and OP is normally measured anti-clockwise as shown.

When measured in units called *radians*, θ = arc PQ .

Angles can also be measured in degrees (π radians = 180E) but this is not recommended. Angles must always be measured in radians when functions of them are being differentiated or integrated.

The following *trigonometrical functions* are commonly used, in which the signs of x and y must be respected.

$\sin\theta = y$ $\quad\cos\theta = x$ $\quad\tan\theta = y/x$ $\quad$ ($\tan\theta$ is the slope of the line OQ)

$\mathrm{cosec}\,\theta = 1/y$ $\quad\sec\theta = 1/x$ $\quad\cot\theta = x/y$

These are not mutually independent. Given $\sin\theta$, all the others can be calculated using the fact that $x^2 + y^2 = 1$.

There are hundreds of formulae connecting these functions, but the following are the most important ones.

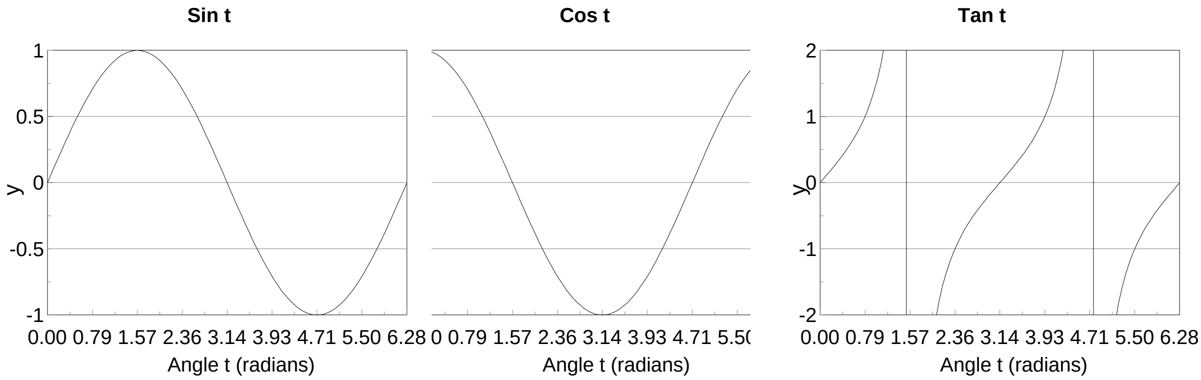
$\cos\theta = \sin(\pi/2 - \theta)$ $\qquad \cos^2\theta + \sin^2\theta = 1$

$\tan\theta = \sin\theta / \cos\theta$

$\sin(\theta+\varphi) = \sin\theta\cos\varphi + \cos\theta\sin\varphi$

$\cos(\theta+\varphi) = \cos\theta\cos\varphi - \sin\theta\sin\varphi$

The graphs of $\sin\theta$, $\cos\theta$ and $\tan\theta$ over the range $(0, 2\pi)$ are as follows. By definition, they repeat themselves over all other intervals $(2n\pi, 2(n+1)\pi)$, n integer.



The following limits are important.

$(\sin\theta)/\theta$ 6 1 $\quad$ as θ 6 0 $\qquad (\tan\theta)/\theta$ 6 1 $\quad$ as θ 6 0 (radians)

(note that these limits would be an awkward number if θ were measured in degrees)

$\tan\theta$ 6 4 $\quad$ as θ 6 $(2n+1)\pi/2$ from below

$\tan\theta$ 6 -4 $\quad$ as θ 6 $(2n+1)\pi/2$ from above $\quad$ n integer

We sometimes need to know the angle that gives rise to a certain value of a trigonometric function. We
denote this in respect of sin, cos, tan respectively as arcsin, arccos and arctan, or as $\sin^{-1}$, $\cos^{-1}$,

and $\tan^{-1}$ respectively. Thus, for example,

$\arcsin[\sin(\theta)] = \theta$, and

$\cos^{-1}(x) = \arccos(x)$.

Note that there is no angle θ for which $\sin(\theta) = 2$, so that there is no value corresponding to arcsin(2).

Acknowledgement

These notes draw upon material in Richard Allsop's Mathematics Checklist document. Any errors present here are, of course, the responsibility of the author.

FUNCTIONS, QUADRATIC AND POLYNOMIAL EQUATIONS, DIFFERENTIATION, AND TAYLOR'S AND MacLAURIN'S SERIES

Mathematics Refresher: Part 3

Prof. BG Heydecker
Centre for Transport Studies
Department of Civil and Environmental Engineering
University College London
London WC1E 6BT

August 2005

Functions

A mathematical ***function*** is a mapping from one set of values to another that maps each value in the first set to exactly one in the second. Some examples of explicit functions $y = f(x)$ from a set of values of the argument $\boldsymbol{x}$ to a set of values of $\boldsymbol{y}$ are

$$y = \alpha + \beta x$$

$$y = \sin x$$

$$y = \log_e x$$

$$y = \exp(x) = e^x$$

$$y = \Theta(x) \equiv \begin{cases} 0 & (x \le 0) \\ 1 & (x > 0) \end{cases}$$

Other functions are implicit rather than explicit (*eg* the value of y that satisfies $y + e^y = x$) whilst other relations have multiple values (*eg* $y = \sqrt{x}$, which has two values whenever $x \neq 0$).

The set of values for which a function is defined is called the ***domain*** of that function, and the set of values to which it maps the elements of the domain is called its ***range***. For each of the example functions given above, the domain and range are given in Table 1.

Table 1: Domain and range of certain functions

Function $f(x)$	Domain	Range
$\alpha + \beta x$	$\mathbb{R}$ (all real numbers)	$\mathbb{R}$ if $\beta \neq 0$ $\{\alpha\}$ if $\beta = 0$
Sin x	$\mathbb{R}$	[-1, 1]
$\log_e x$	$(0, \infty)$	$\mathbb{R}$
e^x	$\mathbb{R}$	$(0, \infty)$
$\Theta(x)$	$\mathbb{R}$	{0, 1}

We often sketch functions as graphs in Cartesian co-ordinates (x, y): because of the requirement that each value of the domain is mapped to exactly one in the range, a graph of this kind meets any line that is parallel to the y axis at most once. For example, the graphs of the exponential, logarithm and sine functions are shown in Figure 1.

In some cases, a variable $\boldsymbol{y}$ is associated with a function that depends on several variables $x_1, x_2, \ldots x_n$, . This can be expressed as

$$y = f(x_1, x_2, \ldots x_n)$$

Graphing a function of this kind is often impractical, though the case of $n = 2$ can be represented by a surface over the relevant part of the (x_1, x_2) plane.

A function $f(x)$ can be used to identify a certain value of x which satisfies the equation $f(x) = 0$; a value of this kind is known as a ***root*** of the equation or a ***zero*** of the function. For example, we might wish to know the value of x for which $\alpha + \beta x = 0$; this can be identified as a zero of the function $f(x) = \alpha + \beta x$. Provided that $\beta \neq 0$, this is seen to be $x = -\alpha/\beta$. If $\beta = 0$, then the solution depends on α : if $\alpha = 0$, then any x is a root of the equation, whilst if $\alpha \neq 0$, there is no root.

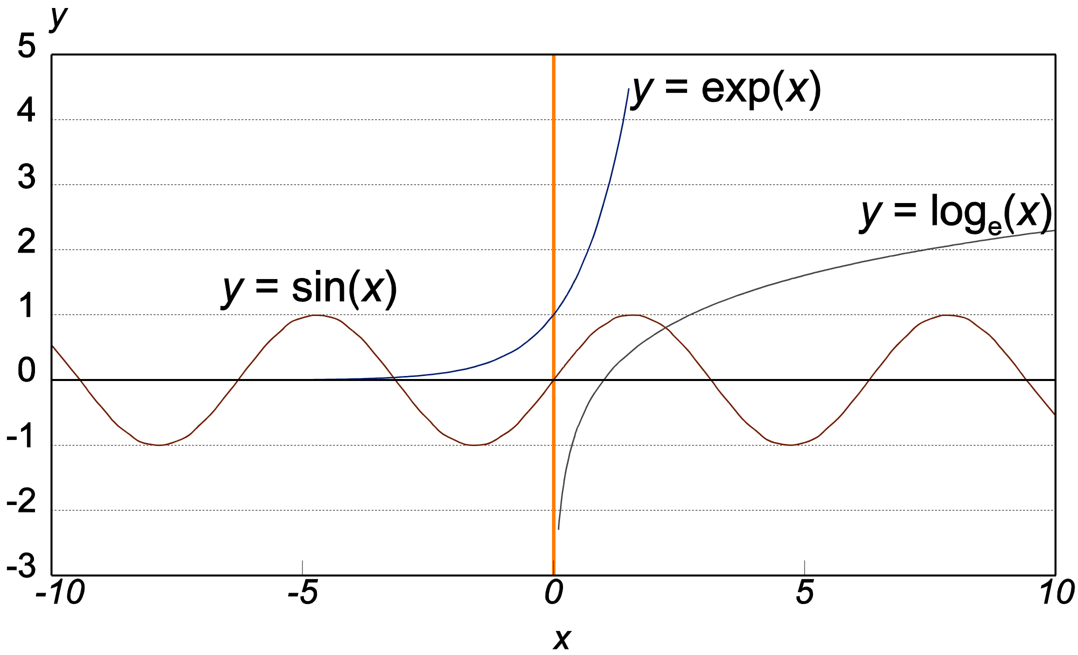


Figure 1: Graphs of some example functions

Quadratic and polynomial equations

We have seen linear functions of the form $f(x) = \alpha + \beta x$. The inclusion of a term in x^2 extends this to the *quadratic* function which has the form $f(x) = \alpha + \beta x + \gamma x^2$. More generally, a kind of function that arises quite often is one of the form

$$f(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \ldots + \alpha_n x^n = \sum_{k=0}^{n} \alpha_k x^k$$

where by convention, $\alpha_n \neq 0$. This is called a ***polynomial*** function of degree ***n***, and gives rise to the corresponding polynomial equation of degree ***n***: $f(x) = 0$. Polynomial functions are always continuous and have domain $\mathbb{R}$; if the degree ***n*** is odd, then they have range $\mathbb{R}$ as well. The linear function $f(x) = \alpha + \beta x$ is a polynomial of degree 1 and the ***quadratic*** function $f(x) = ax^2 + bx + c$ is a polynomial of degree 2.

The graph of a quadratic function is always a parabola with axis parallel to the y axis.

The number of roots that a quadratic equation has depends on the value of the discriminant $b^2 - 4ac$: if this discriminant is strictly positive, then the equation has two distinct roots; if the discriminant is equal to zero, then the equation has one root which corresponds to two coincidental solutions, and if the discriminant is negative, then the equation has no real roots at all. These cases are illustrated in Figure 2.

The values of the roots of a quadratic equation can be expressed in closed form as

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

provided that the discriminant is positive; where two distinct roots exist, they arise from the alternative signs of the square root.

A polynomial equation of degree 3 is called a ***cubic*** and always has at least one real root; it can have one, two or three distinct ones depending on the values of the coefficients α_i, $(0 \le i \le 3)$.

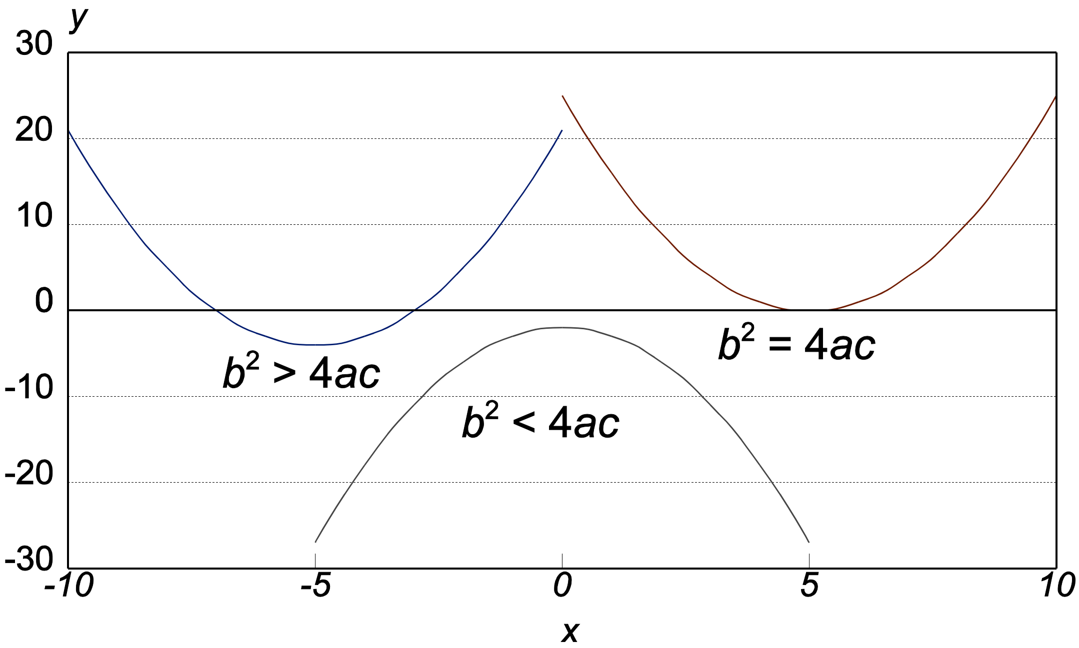


Figure 2: Graphs of some quadratic functions

In general, a polynomial equation of degree n can have up to n distinct real roots but no more than that, and it has at least one if n is odd and in which case the range of the polynomial function is $\mathbb{R}$.

A polynomial equation of degree 3 or 4 can be solved by means of formulae, but these are complicated. If the degree n is greater than 4, then solution by formulae is not always possible. If a root x_0 is known, then the polynomial $p_n(x)$ of degree n can be expressed in the form

$$p_n(x) = (x - x_0)\, p_{n-1}(x)$$

where $p_{n-1}(x)$ is a polynomial of degree $n-1$. This result, which uses the fact that $a \times b = 0 \Rightarrow a = 0$ or $b = 0$, allows us to use knowledge of a root to decrease the degree of polynomial equation that we have to solve.

We usually use numerical methods to solve for roots by successive approximation whenever $n \geq 3$.

Differentiation

For a function $y = f(x)$ we are often interested in how fast the value of y changes as x changes. For example, if y is the position of a vehicle at time x, then the rate of change of y with respect to x is the instantaneous velocity of the vehicle. This rate of change is another function of x and is called the ***derivative*** of y with respect to x. It is often written as

$$f'(x) = \frac{dy}{dx}$$

We can approximate the derivative of y with respect to x for a function $y = f(x)$ at a certain value x_0 by taking the ratio of the finite differences $f(x_0 + h) - f(x_0)$ and h for $h \neq 0$. This is illustrated in Figure 3.

For a function such as the one graphed in Figure 3, and indeed for most of the functions that we shall encounter, the limit of the gradient

$$g(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

exists and is the same in each of the cases that $h \to 0$ from above (*ie* using positive values of h) and from below (*ie* using negative ones). This limiting value is the derivative $f'(x)$ of the function $f(x)$ at x_0. According to convenience and emphasis that is desired in the context of use, we write this as

$$f'(x_0) = \left.\frac{dy}{dx}\right]_{x_0}$$

Given a function $f(x)$, the calculation of $f'(x)$ is called ***differentiation***. If $f(x)$ is given as an expression in x, it is not usually difficult to find $f'(x)$ as an expression in x using the results in Table 2.

Differentiation can be repeated, so that for example once we have obtained the velocity of a vehicle as a function of time by differentiating its position, we can obtain its acceleration by differentiating the velocity function. We write

$$\frac{d}{dx}\left(\frac{dy}{dx}\right) = \frac{d^2y}{dx^2} = f''(x) = f^{(2)}(x)$$

and so on for higher derivatives.

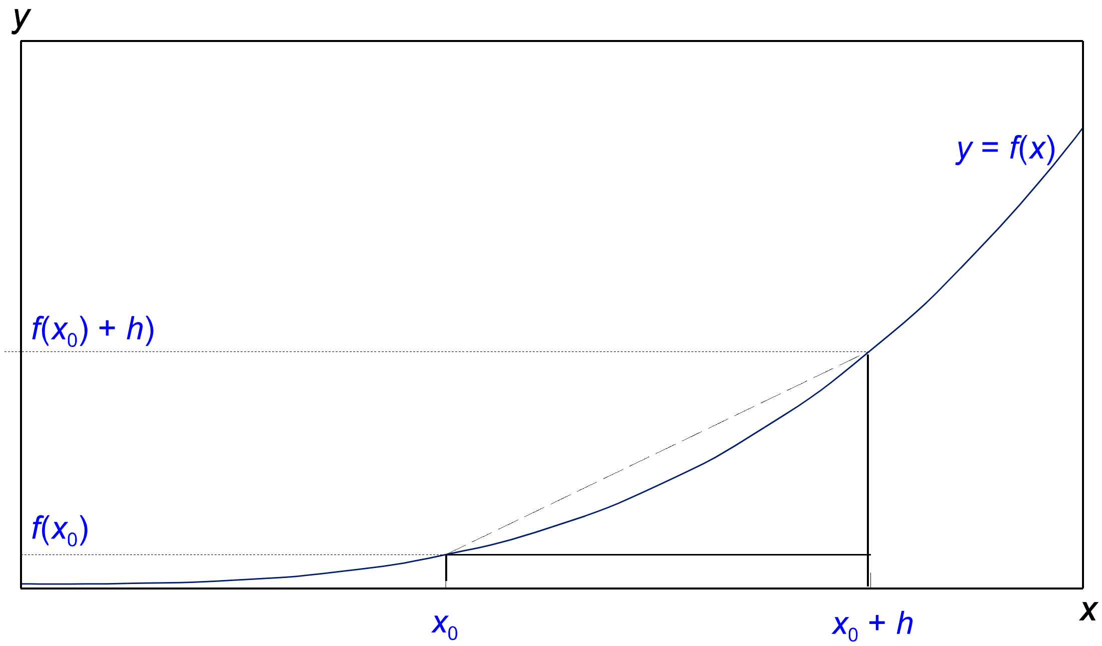


Figure 3: A function and its approximate gradient at x_0

Table 2: Some functions and their derivatives

Function *f(x)*	Notes and restrictions	Derivative *f′(x)*
k	k is a constant	0
x^n	n constant and $x \neq 0$ if $n<1$	$n\,x^{n-1}$
$\sin x$		$\cos x$
$\cos x$		$-\sin x$
$\tan x$	$x \neq (2n+1)\,\pi/2$, n integer	$\sec^2 x$
e^x		e^x
$\log_e(x)$		$1/x$
$k\,u(x)$	k constant	$k\,u'(x)$
$u(x) + v(x)$		$u'(x) + v'(x)$
$u(x)\,v(x)$	Product rule	$u'(x)\,v(x) + u(x)\,v'(x)$
$u(x)\,/\,v(x)$	Quotient rule	$[u'(x)\,v(x) - u(x)\,v'(x)]\,/\,[v(x)]^2$
$u[v(x)]$	Chain rule	$u'[v(x)]\,v'(x)$

An important use of differentiation is to find local maximum and minimum values of functions. For example, the function shown in Figure 4 has local maxima at x_1 and x_4 , and a local minimum at x_2 .

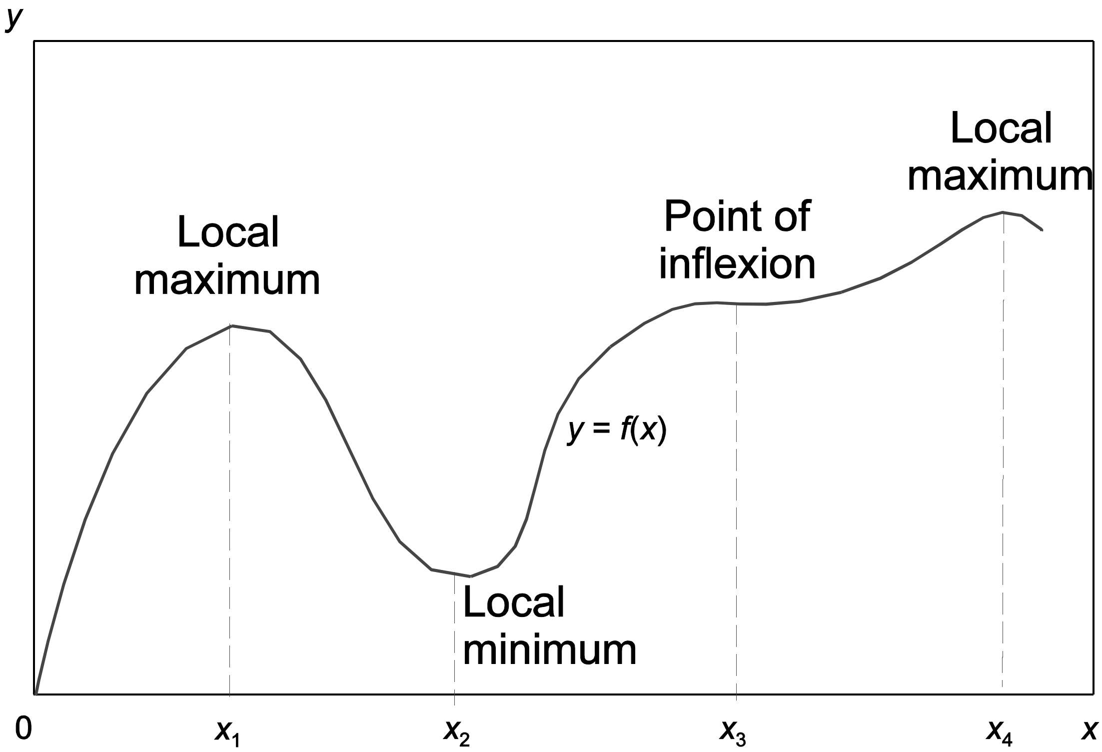


Figure 4: A function with stationary points at $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, $\boldsymbol{x}_3$ and $\boldsymbol{x}_4$.

The common feature of such points is that the derivative $f'(x)$ is zero there, so that $f(x)$ is stationary in the sense that variations in the value of x have only small effect on the value of $f(x)$. The first step in finding local maxima and minima is to solve the equation $f'(x) = 0$. However, not all the roots of this equation are either maxima or minima: an example of another kind of root is x_3 in Figure 4, which is an example of a ***point of horizontal inflexion***.

Having solved $f'(x) = 0$, we need to find which of its roots are maxima, which are minima, and which are points of inflexion. We do this by evaluating successively higher order derivatives $f^{(n)}(x)$ at each of these roots until we find one that is non-zero: the order of this derivative and its sign then determine the nature of the stationary point. Often the second derivative is non-zero, in which case we need look no further. The possibilities for this are identified in Table 3.

Table 3: Nature of $f(x_0)$ according to the sign of the derivative of least order n that is non-zero, and the parity of that order.

$f'(x_0)$	Parity of n	$f^{(n)}(x_0)$	Nature of $f(x_0)$
0	even	< 0	x_0 is a local maximum of $f(x)$
0	even	> 0	x_0 is a local minimum of $f(x)$
0	odd	$\neq 0$	x_0 is a point of inflexion of $f(x)$

For most of the functions that we shall be using, this approach will identify all maxima and minima that lie strictly within their domains. However, we note that where this domain has a finite boundary, the values on the boundary may be extremal (*ie* minimal or maximal) ones: this occurs with the function graphed in Figure 4 which is minimised at the left-hand boundary but is not stationary there. In the case that the function is not defined on the boundary of a finite range (such as logarithm), values taken by the function near the boundary can be more extreme than those at any stationary point. Similarly, if the range of a function is infinite (such as $f(x) = \alpha + \beta x$, for $\beta \neq 0$), values of the function can become progressively more extreme as x becomes either more positive or more negative. In these cases, the functions do not attain their extrema unless constraints are imposed on x.

A function of more than one variable can be differentiated with respect to each variable separately: this is known as ***partial differentiation***. The notation used for this is illustrated in terms of a function $f(x, y, z)$ of the 3 argument variables x, y and z. We write

$$\frac{\partial f}{\partial x} = f_x = \lim_{h \to 0} \frac{f(x+h, y, z) - f(x, y, z)}{h}$$

$$\frac{\partial f}{\partial y} = f_y = \lim_{h \to 0} \frac{f(x, y+h, z) - f(x, y, z)}{h}$$

$$\frac{\partial f}{\partial z} = f_z = \lim_{h \to 0} \frac{f(x, y, z+h) - f(x, y, z)}{h}$$

The calculus for this partial differentiation is identical to that for ordinary differentiation: the key distinction is in the interpretation and use of the resulting function.

Second and higher order partial derivatives are defined as in the following examples

$$\frac{\partial f}{\partial x}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial x^2} = f_{xx}(x, y, z)$$

$$\frac{\partial f}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial y \, \partial x} = f_{xy}(x, y, z)$$

For the kinds of functions that we shall encounter, the mixed partial derivatives are independent of the order of differentiation, so that for example

$$\frac{\partial^2 f}{\partial y \, \partial x} = \frac{\partial^2 f}{\partial x \, \partial y}$$

Partial differentiation is used to identify the sensitivity of a function to changes in one of the arguments whilst the others remain unchanged. In practice, this is not always possible because of interrelationships between the argument variables. For example, we could consider flow of traffic past a point as a function of traffic speed, density and time; the partial derivatives of this function represent the sensitivity to unilateral changes in each of these variables and do not incorporate information such as relationships between speed and density. The total derivative of a function can be found using its partial derivatives together with information about the relationships between the arguments according to the formula

$$\frac{df}{dx} = f_x + f_y \frac{dy}{dx} + f_z \frac{dz}{dx}$$

Taylor's and MacLaurin's series

In some cases we have or can find the value of a function ***f*(*x*)** and its derivatives at a certain point $x = x_0$ and wish to estimate values of the function in the neighbourhood. Provided that the function is sufficiently smooth, this can be done by means of ***Taylor's series***:

$$f(x_0+h) = f(x_0) + h f'(x_0) + \frac{h^2}{2!} f''(x_0) + \ldots + \frac{h^n}{n!} f^{(n)}(x_0+\theta h)$$

$$= \sum_{i=0}^{n-1} \frac{h^i}{i!} f^{(i)}(x_0) + \frac{h^n}{n!} f^{(n)}(x_0+\theta h) \qquad (\text{where } 0 \le \theta \le 1)$$

This is most useful when h is small enough for successive terms to become smaller rapidly so that a good approximation can be obtained with a few terms: in some cases just 2 terms can be adequate. If bounds are available for the value of the derivative $f^{(n)}(x)$ in the interval $[x_0 , x_0+h]$, then Taylor's series can be used to provide a polynomial approximation of specified accuracy using only information about the function at the single point $x = x_0$ by evaluating the final term there.

The special case of Taylor's series in which $x_0 = 0$ is called ***MacLaurin's series*** which, upon replacing h by x, can be written

$$f(x) = f(0) + x f'(0) + \frac{x^2}{2!} f''(0) + \ldots + \frac{x^n}{n!} f^{(n)}(\theta x)$$

This provides a method of expressing certain functions as power series. For example, the exponential function can be expanded as

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots$$

It also provides the basis for ***Newton's method*** for the iterative solution of equations **$f(x) = 0$**. This takes an initial estimate x_0 of a root and calculates a sequence of approximations $x_{n+1} = x_n - f(x_n)/f'(x_n)$. In the case that the function $f(x)$ is linear, then x_1 is a root of the equation whatever the choice of initial estimate x_0 ; in other cases the sequence x_i will yield diminishing values of $|f(x_i)|$ provided that the second derivative $f''(x)$ is sufficiently small in the neighbourhood of the root and the initial estimate x_0 is sufficiently good in the context of the values of the second derivative. This approach can be used to help find maxima and minima of a function by solving for zeros of its derivative.

Acknowledgement
These notes draw upon material in Richard Allsop's Mathematics Checklist document. Any errors present here are, of course, the responsibility of the author.

INTEGRATION, LOGARITHMIC AND EXPONENTIAL FUNCTIONS, DIFFERENTIAL EQUATIONS

Mathematics Refresher: Part 4

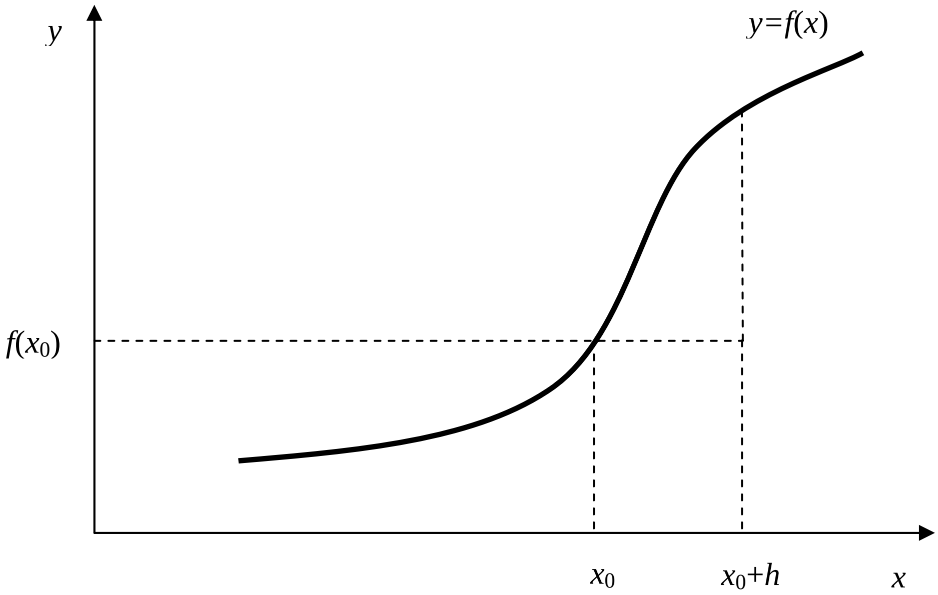
Prof. John Polak
Centre for Transport Studies
Department of Civil and Environmental Engineering
Imperial College London
London SW7 2AZ

September 2005

INTEGRATION

Consider now the reverse of the question that was answered by differentiation. Suppose we know that an unknown function, $F(x)$ say, has a rate of change with respect to x which is a know function $f(x)$, and we wish to find $F(x)$ (e.g., we know the speed of a vehicle at each instant in a give period, and we want to work out how

Suppose we draw the graph of $f(x)$



This tells us that at any point x_0 the rate of change of $F(x)$ is $f(x_0)$, so that for sufficiently small h, we have:

$$\frac{F(x_0+h)-F(x_0)}{h} \cong f(x_0)$$

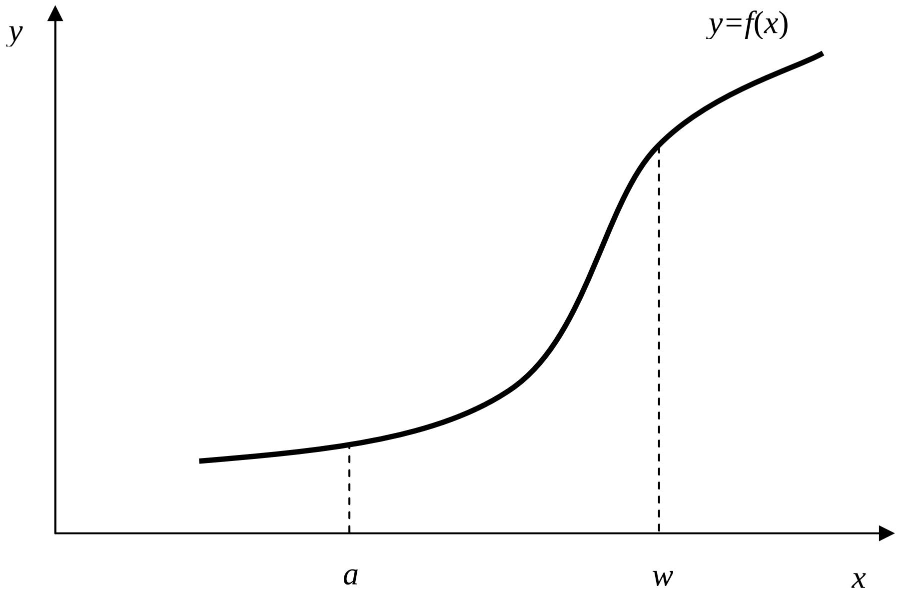
i.e., $$F(x_0+h) \cong F(x_0)+hf(x_0)$$

where the symbol $\cong$ means “approximately equal to”.

In the case shown in the diagram, if $F(x_0)$ is in some sense the area to the left of x_0 between the curve $y=f(x)$ and the x axis, then the error in this approximation is represented by the shared triangle, and $hf(x_0)$ is represented by the rectangle below it.

For sufficiently small h, the triangle becomes negligible compared to the rectangle.

This shows that the shaded area in the diagram below represents $F(w)$-$F(a)$, and the rate of change of this with respect to w is $f(w)$.



Thus, if we know the value of $F(x)$ at some point which we can use as the point a, then the area under the curve between a and w tells us the value of $F(w)$ for any $w > a$.

We write the area as $\int_a^w f(x)dx$, so that:

$$F(w) - F(a) = \int_a^w f(x)dx$$

The same applies with a and w interchanged, provided that we count as negative the area covered as we move to the left along the x axis. It also applies if the curve $y=f(x)$ crosses the x axis, provided that we count areas below the x axis as negative.

Thus quite generally (for the kinds of functions we shall be considering), whatever the relative positions of a and b and whatever the sign of $f(x)$, we have:

$$F(b) - F(a) = \int_a^b f(x)dx$$

where $F(x)$ is any function whose rate of change with respect to x is $f(x)$. The quantity $\int_a^b f(x)dx$ is know as the *definite integral* of $f(x)$ over the interval [a,b] and is also written $\left[F(x)\right]_a^b$, if we have an expression for $F(x)$ in terms of x.

To find an expression for $F(x)$ given an expression for $f(x)$, we have to perform the reverse of differentiation (if we can) e.g., $ax^{a-1} \to x^a$, $\cos(x) \to \sin(x)$ etc. We also have to add on an arbitrary constant because for any constant $k, \dfrac{d(F(x)+k)}{dx} = \dfrac{dF(x)}{dx} = F'(x)$, so that knowledge of $F'(x)$ does not enable us to distinguish between functions $F(x)$ which differ only by a constant.

This process is written, for example

$$\int ax^{a-1}dx = x^a + k$$

or more generally

$$\int f(x)dx = F(x) + k$$

where the quantity $\int f(x)dx$ is called the *indefinite integral* of $f(x)$.

It is in this sense that we can think of integration as being the inverse of differentiation. A list of functions and their derivatives can therefore be used in reverse to help to find indefinite integrals.

Moreover, the properties of sums, products and quotients of derivatives discussed in Maths Refreshers 3 can be used to derive a number of useful properties of indefinite integrals. For example the linearity of differentiation implies that,

$$\int \lambda f(x)dx = \lambda \int f(x)dx \quad \text{for any constant } \lambda$$

and

$$\int [f(x) + g(x)]dx = \int f(x)dx + \int g(x)dx$$

Moreover, if $\int u(x)dx = U(x) + k$ and $\int v(x)dx = V(x) + k$ then it follows from the product rule of differentiation that:

$$\int u(v(x))v'(x)dx = U(v(x)) + k$$

and

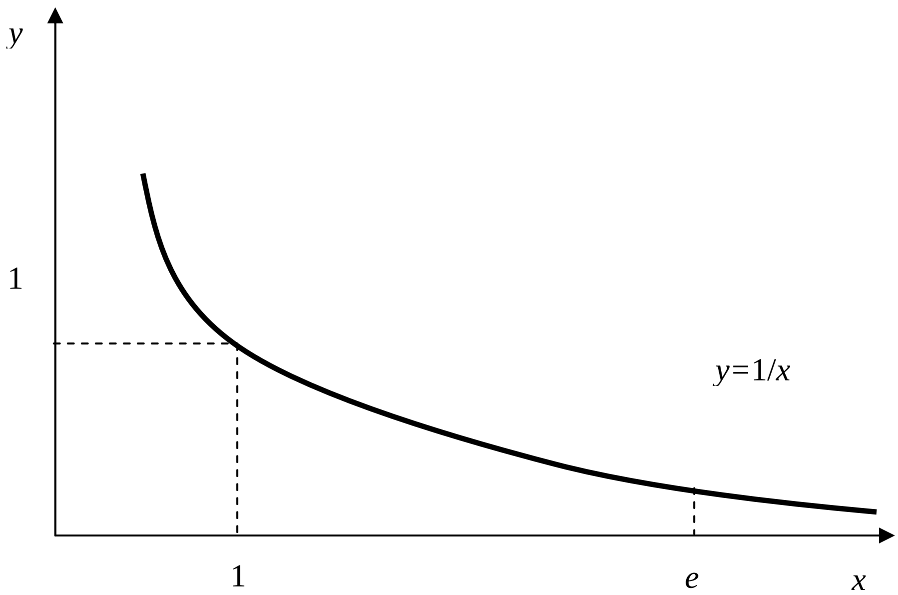
$$\int U(x)v(x)dx = U(x)V(x) - \int u(x)V(x)dx$$

This is sometimes called the formula for *integration by parts*.

There are very many methods of evaluating an indefinite integral, note however, that by no means all expression $f(x)$, even quite-simple looking ones, have indefinite integrals that are similarly simple expressions. For those that do not, however, we can evaluate definite integrals numerically.

LOGARITHMIC AND EXPONENTIAL FUNCTIONS

The simplest function that has no indefinite integral in term of the functions of x as have so far considered is $f(x) = 1/x$



In such cases, mathematicians often use the integral as the definition of a new function. In this case the result of doing so is extremely useful. It is easy to shown that for u and $v > 0$

$$\int_1^{uv} \frac{1}{x}dx = \int_1^{u} \frac{1}{x}dx + \int_1^{v} \frac{1}{x}dx$$

Thus, the integral $\int_1^{t} \frac{1}{x}dx$ has the property of a logarithm. For this reason it is called ln(t) (the *natural logarithm* of t) and the above equation becomes

$$\ln(uv) = \ln(u) + \ln(v)$$

By definition, and as for any logarithm, ln(1)=0, $\ln(x) \to \infty$ as $x \to \infty$ and $\ln(x) \to -\infty$ as $x \to 0$.

The number whose natural logarithm is 1 is of fundamental importance. It is denoted by e and is approximately 2.7183. In the diagram above, it is the point on the x axis such that the shaded area is 1.

It is easy to see that for all rational r, $\ln(e^r) = r$.

For all real x we can define e^x to be the number w such that ln(w) =x. The function e^x is called the *exponential function*. Then, as we expect of a power, $e^x e^y = e^{x+y}$, for all x and y.

Moreover, $e^{\ln(x)}$ is the number y such that ln(y)=ln(x), i.e., x. Thus for any real $x > 0$, x= $e^{\ln(x)}$.

This enables us to fill in the gap in in our definition of powers by defining x^y to be $e^{y\ln(x)}$ for all real y and all $x > 0$

We often write exp(x) for e^x (the alternative notations can be used interchangably). This is particularly useful when writing $e^{f(x)}$ when $f(x)$ is already a complicated expression.

The functions ln(x) and exp(x) are both easy to differentiate. By definition $\frac{d\ln(x)}{dx} = \frac{1}{x}$ and it is easy to show that $\frac{d\exp(x)}{dx} = \exp(x)$

Thus exp(x) is the function of x that remains unaltered under differentiation.

From the fact that $\frac{du(v(x))}{dx} = u'(v(x))v'(x)$ (chain rule), it follows that $\frac{d(e^{kx})}{dx} = ke^{x}$.

Incidentally, it can also be shown from the figure above that $\underset{N\to\infty}{Lim}\left(1+\frac{1}{N}\right)^{N} = e$ (but this is not a result we will use in the MSc courses).

A few of the most useful standard indefinite integrals are given in the table below (where n is a constant and k is the constant of integration). Many mathematics textbooks contain extensive lists of know indefinite integrals.

$f(x)$	$\int f(x)dx$
$x^n \quad (n \neq -1)$	$\frac{x^{n+1}}{n+1} + k$
$x^{-1} = \frac{1}{x}$	$\ln(\lvert x \rvert) + k$
$\mathrm{e}^{nx} \quad (n \neq 0)$	$\frac{\mathrm{e}^{nx}}{n} + k$
$\sin(nx) \quad (n \neq 0)$	$\frac{-\cos(nx)}{n} + k$
$\cos(nx) \quad (n \neq 0)$	$\frac{\sin(nx)}{n} + k$

DIFFERENTIAL EQUATIONS

Integration was introduced as a procedure for finding a function $F(x)$ such that $F'(x)$ is a know function $f(x)$. This can also be described as solving for y as a function of x the following equation

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

This is a simple example of a differential equation, that is an equation in x, y and various derivatives of y with respect to x (e.g., $\frac{dy}{dx}, \frac{d^2y}{dx^2}$ and so on). Only the simplest of such differentia equations can be solved analytically to yield a neat algebraic expression for y as a function of x. Most differential equations must be solved numerically.

However, the above differential equation can be solved if the corresponding indefinite integral $\int f(x)dx$ can be found. The solution involves an arbitrary constant called the *constant of integration*, and this is true of any differential equation involving $\frac{dy}{dx}$ but no higher order derivative. After solution, the appropriate value of the constant can be found if the value y_0 of y corresponding to just one value x_0 of x is know. The solution of a differential equation involving derivatives up to the n^{th} order involves n constants of integration.

Another very simple form of differential equation that can be solved similarly is:

$$\frac{dy}{dx} + f(x)y = g(x)$$

This equation has the general solution (not as complicated as it looks)

$$y = \exp\left(-\int f(x)dx\right)\left[k + \int g(x)\exp\left(\int f(x)dx\right)dx\right]$$

where k is the constant of integration and the various indefinite integral are evaluated without introducing further constants. This expression is naturally most useful if the two indefinite integrals involved can be found.

Acknowledgement

These notes draw upon material in Richard Allsop's Mathematics Checklist document. Any errors present here are, of course, the responsibility of the author.

PROBABILITY AND STATISTICS

Mathematics Refresher: Part 5

Prof. John Polak
Centre for Transport Studies
Department of Civil and Environmental Engineering
Imperial College London
London SW7 2AZ

September 2005

PROBABILITY

Many of the problems that arise in transport studies involve uncertain events (e.g., whether a particular traveller will be involved in a traffic accident during a given observation period). The laws of probability provide us with the basic concepts to enable us to coherently describe uncertain events (and uncertain information) and provide the foundation for a wide range of practical tools (called statistical models) that help us draw inferences and make decisions about such uncertain events.

Definition of Probability

Consider an event *A*, which may or may not occur (or a statement *A* that may or may not be true) under the circumstances *Z*.

The probability of *A* happening (or being true) in the circumstances *Z* is denoted by P(*A* | *Z*) (sometimes this is written Pr(*A* | *Z*)). Suppose that *Z* can be repeated many times (*N*) over and a number of trials are made. Let the number of times that event *A* occurs in these *N* trials be N_A, then for finite *N* , the probability P(*A* | *Z*) is estimated by N_A/N and in the limit as $N \to \infty$, we have the definition:

$$\mathrm{P}(A \mid Z) = \lim_{N \to \infty} \left(\frac{N_A}{N} \right)$$

If the circumstances *Z* cannot be repeated, then a particular person's (subjective) estimate of P(*A* | *Z*) can nevertheless still be expressed in terms of a fair bet. If for example, the person regarded as fair a bet in which he wins £1 if *A* happens and loses £x if it does not, then his estimate of P(*A* | *Z*) is $x/(1+x)$. If the person is consistent in a certain mathematical sense, the probabilities estimated in both these ways satisfy the following axioms.

Axioms of Probability

Suppose that under the circumstances *Z*, the set of all possible events that might occur is denoted by *S*.

1. The probability of any particular event *A* in *S* occurring lies between zero and unity. i.e., $0 \le \mathrm{P}(A \mid Z) \le 1$ for all *A* in the set *S* and all *Z*

2. Each even that occurs belongs to *S* i.e., P(*S* | *Z*) =1

3. If *A* and *B* are two events in *S* that cannot both happen in circumstances *Z* (i.e., *A* and *B* are *mutually exclusive events*) then P(*A* or *B* | *Z*) = P(*A* | *Z*) + P(*B* | *Z*)

Although these axioms are very simple, they allow us to build up sophisticated and useful tools of analysis.

Some Basic Properties of Probabilities

1. For any event *A* in S, the probability that *A* does not occur is given by:
   P(*A* does not happen | *Z*) = 1 – P(*A* | *Z*)

2. For mutually exclusive events A_1, $A_2,\ldots,A_n$ all of which belong to S
   $$\mathrm{P}(A_1 \text{ or } A_2 \text{ or } ,\ldots, \text{or } A_n \mid Z) = \sum_{i=1}^{n} \mathrm{P}(A_i \mid Z)$$

3. For any two events A and B in S that are not mutually exclusive
   P(A or B | Z) = P(A | Z) + P(B | Z) – P(A and B | Z)

The term P(A and B | Z) is called the *joint probability* of the event A and B occurring.

Conditional Probability

Often it is required to calculate the probability of an event B occurring, given that an event A occurs. This is written P(B | Z and A) and is called a *conditional probability*, since it is the probability of B conditional on A. We can show that

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

Similarly

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

and hence

$$\mathrm{P}(A \text{ and } B \mid Z) = \mathrm{P}(A \mid Z)\mathrm{P}(B \mid Z \text{ and } A) = \mathrm{P}(B \mid Z)\mathrm{P}(A \mid Z \text{ and } B)$$

It is very important to understand the difference between joint probability and conditional probability.

Marginal Probability

If events $A_1,A_2,\ldots,A_n$ are mutually exclusive and exhaustive in circumstances Z, then for any other event B we can extend the concept of conditional probability as follows:

$$\mathrm{P}(B \mid Z) = \sum_{i=1}^{n} \mathrm{P}(B \mid Z \text{ and } A_i)\mathrm{P}(A_i \mid Z)$$

In these circumstances P(B | Z) is called the *marginal probability* of B. It is the probability that B occurs, taking into account the influence of all the possible events A_i.

Independent Events

If P(B | Z and A) = P(B | Z and (A doesn’t happen)) (i.e., the probability of the event B happening is completely unaffected by whether or not the event A happens) then the events A and B are said to be *independent* and then

$$\mathrm{P}(B \mid Z \text{ and } A) = \mathrm{P}(B \mid Z \text{ and } (A \text{ doesn't happen})) = \mathrm{P}(B \mid Z)$$

and hence from above

$$\mathrm{P}(A \text{ and } B \mid Z) = \mathrm{P}(A \mid Z)\mathrm{P}(B \mid Z)$$

Independence is a useful property since it greatly simplifies the calculation of joint probabilities.

Likelihood

When event *A* is known to have happened but the circumstances *Z* are uncertain, then P(*A* | *Z*) is called the *likelihood* of the circumstances *Z* in the light of the event *A*. The likelihood plays an important role in more advanced statistical modelling.

Bayes Theorem

This states that for any events $A_1, A_2, \ldots, A_n$ and any other event *B* such that $\mathrm{P}(B \mid Z) \neq 0$

$$\mathrm{P}(A_i \mid Z \text{ and } B) \propto \mathrm{P}(B \mid Z \text{ and } A_i)\mathrm{P}(A_i \mid Z)$$

that is:

$$(\text{probability of } A_i \text{ posterior to } B) \propto (\text{likelihood of } A_i \text{ in the light of } B) \times (\text{probability of } A_i \text{ prior to } B)$$

STATISTICS

Statistical analysis and modelling ("statistics" for short) is concerned with applying the ideas of probability to the practical problems associated with analysing data. There are a number of key concepts involved in statistics.

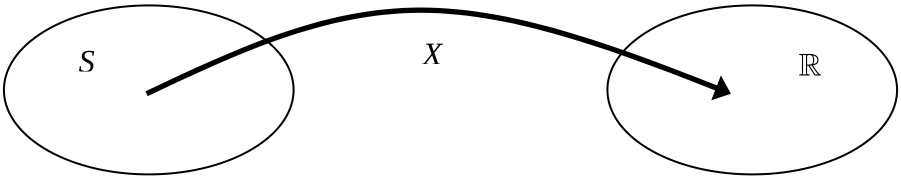
Random Variable

The key idea in making the transition from probability to statistics is the idea of characterising data using *random variables*.

A random variable *X* is a function with the following properties:

1. *X* is defined over the set *S* of all possible events that can occur in the circumstances *Z* and maps each element of *S* into a point in the set of real numbers $\mathbb{R}$.

2. For every real number *x*, the probability P(*X* = *x*) that *X* assumes the value *x* in a trial is well behaved. Likewise for every interval $[x_1,x_2]$, the probability $\mathrm{P}(X \in [x_1, x_2])$ that *X* assumes any value in the interval $[x_1,x_2]$ is well behaved.



Distribution of a Random Variable

The probabilities P define what is called the *probability distribution* (sometimes shortened to just the *distribution*) of the random variable *X*. This is given by the *probability distribution function* $F(x)$ (sometimes shortened to just the *distribution function*), which is defined by:

$$F(x) = \mathrm{P}(X \le x)$$

Although these definitions of random variable and distribution function are very general, only a rather limited number of distributions occur frequently in typical applications. The properties and use of a number of these common distributions will be introduced in the module T2.

An important classification of random variables and their distribution functions is according to whether they are discrete or continuous.

Discrete Random Variables

A random variable *X* and its distribution $F(x)$ is discrete if *X* can assume only finitely (or more correctly, countably infinitely) many values. Corresponding to each possible value of *X*, x_1, x_2,… and so on are positive probabilities p_1, p_2,… and so on, whereas the probability of any interval not containing a possible value of *X* is zero. Discrete random variables arise in circumstances where the trial naturally gives rise to an outcome that is a classification (e.g., a traveller chooses either bus or car as the mode of travel to work) or a count (e.g., an engineer records the number of accidents per unit time at a road junction).

The probabilities define the *probability density function* $f(x)$ (sometimes shortened to just the *density function*) by:

$$f(x) = \begin{cases} p_i & \text{if } x = x_i \quad (i = 1, 2, ...) \\ 0 & \text{otherwise} \end{cases}$$

For a discrete variable, the density function is related to the distribution function by summation:

$$F(x) = \sum_{x_i \le x} f(x_i) = \sum_{x_i \le x} p_i$$

The second axiom of probability requires that the summation of the density function over all possible values of *X* is unity:

$$\sum_{\text{all possible } x_i} f(x_i) = 1$$

The probability $\mathrm{P}(a \le X \le b)$ that the random variable *X* assumes some value in the range $a \le X \le b$ is therefore given by

$$\mathrm{P}(a \le X \le b) = F(b) - F(a) = \sum_{a \le x_i \le b} p_i$$

Continuous Random Variables

Continuous random variables can assume any real value within a certain range (which may be infinite). Therefore we cannot enumerate each possible value of *X*. Continuous random variables arise in circumstances in which the trial naturally gives rise to a continuous measured quantity such as the speed of a traffic stream over a section of road or the number of miles operated by a bus company under a particular contract.

The distribution function of a continuous random variable is related to its density function by integration

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

From which it follows that $F'(x) = f(x)$.

The second axiom of probability requires, in the case of a continuous random variable, that the integral of the density function over the real line is unity:

$$\int_{-\infty}^{\infty} f(t)dt = 1$$

(Note that in practice many continuous distributions have a more limited range, in which case their density function is zero outside that range, so the normalisation above is general in nature).

Furthermore, in the case of a continuous random variable, the probability $\mathrm{P}(a \le X \le b)$ that the random variable *X* assumes some value in the range $a \le X \le b$ is given by

$$\mathrm{P}(a \le X \le b) = F(b) - F(a) = \int_{a}^{b} f(t)dt$$

Role of Probability Distributions

In module T2 (and elsewhere in the course), we will use random variables and probability distributions (and methods derived from these concepts) to analyse data from a wide range of different transport applications. Although the details will differ from case to case, there will be a common basic approach. This involves

1. Conceiving of the data we collect in the real world as being represented by one or more random variables.
2. Selecting appropriate probability distributions to describe these random variables, in the light of our knowledge of the real world setting.
3. Manipulating the resulting probability distributions in order to draw conclusions about the probability of certain events occurring or certain assertions regarding the random variables as being true.
4. Translating these conclusions about the random variables and their distributions back into conclusions about the original real world setting.

For example, we might want to find out whether a bus lane has affected the travel time of buses along a particular corridor. Using the ideas from this lectures, we could approach this problem by imagining that the travel time of buses along the corridor is a random variable (called t say) and that without the bus lane its distribution function is $F_1(t)$ whereas with the bus lane its distribution is $F_2(t)$. If we make measurements of the travel time with and without the bus lane, then we want to know whether the data provide us with sufficient evidence to claim that $F_2(t)$ is a different function than $F_1(t)$.

Answering this type of question requires that we have ways of easily characterising probability distributions and of making inferences about them. These issues are taken up in more detail in the module T2.

Acknowledgement
These notes draw upon material in Richard Allsop's Mathematics Checklist document. Any errors present here are, of course, the responsibility of the author.