Maths Refresher

For The Short Course, Department of Infectious Disease Epidemiology, Imperial College London

Chris Wymant

September 13th 2015

Topics We'll Look At

Basics

Exponentials, Logarithms and Trigonometric Functions

Differentiation

What Functions Look Like

Integration

Discrete Probability Basics

Matrix Basics

Motivation

"As a matter of fact all epidemiology, concerned as it is with variation of disease from time to time or from place to place, must be considered mathematically, however many variables are implicated, if it is to be considered scientifically at all. *To say that a disease depends on certain factors is not to say much, until we can also form an estimate as to how largely each factor influences the whole result.* And the mathematical method of treatment is really nothing but the application of careful reasoning to the problems at hand."

Sir Ronald Ross, 1911. (My italics.)

Apologies to you and other lecturers for subsequent repetition of this quote...

A health warning: starting from the basics there's a lot to get through. Some of it will be hard for those who haven't done maths for a while. Not everyone will come out understanding everything, but it will definitely be helpful (and fun). Revisit the material during the following two weeks when you don't understand something mathematical in nature, and discuss with others on the course.

You can find some slightly more advanced material on calculus and probability for infectious disease modelling at http://www.imperial.ac.uk/people/c.wymant

Thanks to James Truscott with whom I wrote a course that fed into this one, and to Paul Parham whose idea it was to have this course.

Basics

A guiding principle in mathematics: starting from an equation that is correct,

- provided you do exactly the same thing to both sides of the equation (e.g. add something, multiply by something), and
- provided what you do makes sense (i.e. pretty much anything you might think of except dividing by zero),

you'll end up with something that's correct. The trick is figuring out what to do to both sides to get something *useful*.

e.g. to solve the equation $2 \times x + 1 = 7$ for x, subtract 1 from both sides and divide by 2 to see that $x = 3$.

Before We Get Started: Some Notation

Symbol	Meaning	Examples
$<$	less than	$1 < 2$ (true); $2 < 2$ (false); $3 < 2$ (false)
$>$	greater than	$1 > 2$ (false); $2 > 2$ (false); $3 > 2$ (true)
$\leq, \leqslant$	less than or equal to	$1 \leq 2$ (true); $2 \leq 2$ (true); $3 \leq 2$ (false)
$\geq, \geqslant$	greater than or equal to	$1 \geq 2$ (false); $2 \geq 2$ (true); $3 \geq 2$ (true)
$\ll$	much less than (subjective, context dependent). It's usually based on the relative difference, rather than the absolute difference.	$0.001 \ll 1$ (usually true); $9.001 \ll 10$ (usually false)
$\gg$	much greater than (subjective, context dependent)	$1 \gg 0.001$ (usually true); $10 \gg 9.001$ (usually false)

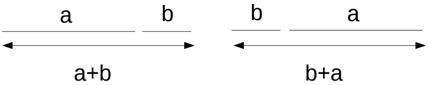
$\pm$	plus or minus	e.g. $\sqrt{4} = \pm 2$, i.e. $2^2 = 4$ and $(-2)^2 = 4$; e.g. death toll $= 120 \pm 10$, i.e. an indication of uncertainty
$\simeq, \approx$	approximately equal to (subjective, context dependent)	For a *total* budget for a national health emergency, $\$1 \approx \$2 \approx \$0$. For an average donation per person, the three values are very different!
$\propto$	proportional to. This means if one thing increases by a given factor, the other increases by that same factor (e.g. if one doubles, the other doubles). Equivalently, their ratio is constant. y is proportional to x means $y = mx$ for some constant c.	Without a multi-buy special offer (e.g. buy two get one free), the total price for multiple identically priced items is proportional to the number of items bought.

$\ldots$	Used to replace something the author hopes is obvious.	$1+2+3+\ldots+8 = 1+2+3+4+5+6+7+8$
$\sum$	'sum over', 'the sum of'	$\sum_{i=1}^{4} i = 1+2+3+4$
$\prod$	'multiply together', 'the product of'	$\prod_{i=3}^{5} i = 3\times 4\times 5$
!	factorial: the product of all numbers between 1 and the number in question, inclusive	$9! = 9\times 8\times 7\ldots\times 2\times 1$
%	per cent, literally 'one per hundred'	$f = 5\% \iff f = 0.05$; x decreased by 25% $\iff x_{\mathrm{new}} = 0.75\times x_{\mathrm{old}}$
$\implies$	implies. $A \implies B$ means 'B is true if A is true', or 'A is a sufficient condition for B'.	$x=2 \implies x^2=4$

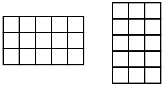
$\iff$	implies and is implied by. $A \iff B$ means ‘B is true if and only if A is true’, or ‘A is a necessary and sufficient condition for B’.	$x^2 = 4 \iff x = \pm 2$
$\therefore$	‘therefore’. Based on everything that has been said so far, we can draw the conclusion that follows.	$x < -2$, $y = -1$, $z > 3$, $\therefore x < y < z$. I think $\therefore$ I am.
$\mid$	‘given that’	The probability of wet pavements in the morning $\mid$ we’re in a drought $<$ the probability of wet pavements in the morning $\mid$ it rained in the night.
$\lvert \ldots \rvert$	the ‘modulus’ or ‘absolute value’ of something. $\lvert x \rvert = \begin{cases} x, & \text{if } x \geq 0 \\ -x, & \text{if } x < 0 \end{cases}$	$\lvert 2 \rvert = 2; \quad \lvert -2 \rvert = 2$

Basic basics: adding, multiplying, subtracting

$a + b = b + a$

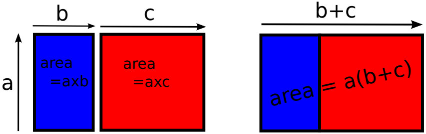


$a \times b = b \times a$ e.g. $3 \times 5 = 5 \times 3 = 15$



$a - b \neq b - a$ e.g. $1 - 2 = -1$, but $2 - 1 = 1$

$$a(b + c) = a \times b + a \times c$$



Basic basics: fractions I

It's best to avoid using the $\div$ symbol and write out fractions as one thing on top of another thing, for clarity.

For basic rules regarding fractions it helps to think of $\dfrac{a}{b}$ as the amount of cake one person gets when a cakes are divided equally between b people.

$\frac{a}{a} = 1$, because you have as many cakes as people: they get one each.

$\frac{a+b}{c} = \frac{a}{c} + \frac{b}{c}$ because if I divide $a + b$ cakes between c people, it makes no difference whether I put the a and b cakes together first and then divide them all up, or I divide the a cakes up first and then the b cakes up afterwards – each person gets the same amount of cake either way.

Basic basics: fractions II

$a \times \frac{b}{c} = \frac{a \times b}{c}$ because the left-hand side means sharing b cakes between c people then increasing by a factor a; the right-hand side means you had a times as much cake in the first place. It's the same. (Note that choosing b to be equal to 1, this rule tells us that $\frac{a}{c} = a \times \frac{1}{c}$.)

$\frac{a}{b} \div c = \frac{a}{b \times c}$ because the left-hand side means first sharing a cakes between b people, and then each person's share actually has to go to c people; the right-hand side means you had c times as people in the first place. It's the same.

Combining the last two rules: $\dfrac{a}{b} \times \dfrac{c}{d} = \dfrac{a \times c}{b \times d}$

If this is not all familiar & obvious you should practise working with fractions, e.g. at https://www.khanacademy.org/math/arithmetic/fractions

Operator Precedence

To evaluate $3 \times 2 + 2$ does one do the multiplication first (giving 8) or the addition first (giving 12)? Operators with *higher precedence* act first. Precedence is determined by *BIDMAS*:

- Brackets
- Indices (raising things to powers)
- Division & Multiplication
- Addition & Subtraction

NB addition and subtraction have the same precedence as each other, so between these things one evaluates from left to right. e.g. $3 - 2 + 1$ is 2, not 0, because the subtraction is done before the addition.
Likewise, division and multiplication have the same precedence.
Indices have higher precedence still: $2 + 3^2 \times 2 = 2 + 9 \times 2 = 2 + 18 = 20$.
Brackets have the highest precedence of all because their raison d'être is to allow manual specification of the intended order of evaluation, overriding the other rules. e.g. $(2 + 3)^2 = 5^2 = 25$.

Playing with Powers I

Raising a number to a power a means a factors of that number all multiplied together:

$x^2 = x \times x$

$x^a = \underbrace{x \times x \times \ldots x}_{a \text{ factors}}$

Raising a product to a power:

$$\begin{aligned}
(x \times y)^a &= \underbrace{(x \times y) \times (x \times y) \times \ldots (x \times y)}_{a \text{ factors of } (x \times y)} \\
&= \underbrace{x \times x \times \ldots x}_{a \text{ factors}} \times \underbrace{y \times y \times \ldots y}_{a \text{ factors}} \\
&= x^a \times y^a
\end{aligned}$$

Repeatedly applying this rule one can see that
$(x \times y \times z \times \ldots)^a = x^a \times y^a \times z^a \times \ldots$

Playing with Powers II

Multiplying two different powers of the same number:

$$
\begin{aligned}
x^a \times x^b &= \underbrace{x \times x \times \ldots x}_{a \text{ factors}} \times \underbrace{x \times x \times \ldots x}_{b \text{ factors}}, \\
&= \underbrace{x \times x \times \ldots x}_{(a+b) \text{ factors}} \\
&= x^{a+b}
\end{aligned}
$$

Playing with Powers III

Now, multiplying by a number then dividing by that same number does nothing. Therefore when multiplying by x a times and dividing by x b times, b of the a multiplications get cancelled out:

$$
\begin{aligned}
\frac{x^a}{x^b} &= \frac{\overbrace{x \times x \times \ldots x}^{a \text{ factors}}}{\underbrace{x \times x \times \ldots x}_{b \text{ factors}}}, \\
&= \frac{\overbrace{x \times x \times \ldots x}^{b \text{ factors}} \times \overbrace{x \times x \times \ldots x}^{(a-b) \text{ factors}}}{\underbrace{x \times x \times \ldots x}_{b \text{ factors}}}, \\
&= \underbrace{x \times x \times \ldots x}_{(a-b) \text{ factors}} \\
&= x^{a-b}
\end{aligned}
$$

Playing with Powers IV

We know that $x^a/x^b = x^{a-b}$. Choosing $b = a$ gives $x^a/x^a = x^0$; this should clearly equal 1, since any number divided by itself is 1. This tells us that any number raised to the power 0 gives 1:

$$x^0 = 1$$

We can define what is meant by a *negative* power using our understanding that $x^0 = 1$ and $x^a \times x^b = x^{a+b}$. Together, these imply that $x^a \times x^{-a} = x^0 = 1$, which means that x^{-a} is the *reciprocal* of x^a, i.e. 1 divided by it:

$$x^{-a} = \frac{1}{\underbrace{x \times x \times \ldots x}_{a \text{ factors}}}$$

Playing with Powers V

From the definition of x^a, x^1 is simply x. Then since $(x^a)^b = x^{a \times b}$, we know that $(x^{1/n})^n = x^1 = x$, which defines

$$x^{1/n} = \sqrt[n]{x}$$

as the number that gives x when multiplied together n times. For $n = 2$ we say the *square* root, for $n = 3$ the *cube* root, for greater n simply the nth root. For even n there are always two solutions – plus-or-minus the same thing – because the product of an even number of negative numbers is positive.
e.g. $4^{1/2} = \pm 2$ because $2 \times 2 = 4$ and $(-2) \times (-2) = 4$.
c.f. $8^{1/3} = 2$, because $2 \times 2 \times 2 = 8$ but $(-2) \times (-2) \times (-2) \neq 8$.

Scientific Notation a.k.a. Standard Form

A figure is in standard form when it is expressed in the following way: $a \times 10^b$,
where $1 \leq a < 10$ and b is an integer (a whole number). b is the number of positions one has to move the left-most non-zero digit to the right to get it on the left of the decimal point sign. e.g.

- $568,000,000 = 5.68 \times 10^8$
- $0.00014 = 1.4 \times 10^{-4}$

The following expressions are all in standard form and equal:
1 year $= 3.6525 \times 10^2$ days $= 8.766 \times 10^3$ hours $= 3.15576 \times 10^7$ s

Using the previous *playing with powers* rules, numbers in standard form are easily multiplied, divided and raised to powers. To add or subtract numbers in standard form it's easiest to take one of them *out* of standard form if necessary, so that they're both multiplied by the same power of ten.

Decimal Places and Significant Figures

How to round a digit: if it's followed by a 0, 1, 2, 3 or 4 we round down (i.e. leave it as is); if it's followed by a 5, 6, 7, 8 or 9 we round up (i.e. increase it by 1). To round the digit 9 upwards, set it to zero and increase the number to the left by 1.

Rounding to n decimal places: the nth digit after the decimal place is rounded. You should explicitly show the n decimal places even if the last ones are zero. Avoid rounding a figure twice: round the original.
e.g. $1.456 = 1.46$ (2 d.p.) $= 1.5$ (1 d.p.) $= 1$ (0 d.p.)
e.g. $2.9995 = 3.000$ (3 d.p.)

Rounding to n significant figures: the nth non-zero digit is rounded. This is useful because it allows you to keep an amount of precision relative to the value itself. e.g. rounding to the nearest whole number: the official population of Greater London in 2013 was $8,416,535$ (0 d.p.) and my height is $2\,\mathrm{m}$ (0 d.p.); c.f. 8.42 million (3 s.f.) and $1.73\,\mathrm{m}$ respectively.
NB for a number in standard form, rounding to n sig figs is the same as rounding to $(n-1)$ d.p.

Subscript Indices and Vectors

Sometimes subscripts are added to letters to distinguish between different instances of the same kind of quantity. e.g. if I is incidence, you might see I_{UK} and I_{France}.

Other times the subscript has a more precise role: it is an index (commonly i, j or k) that runs over certain values, allowing different numbers to be collected together into a single object. For example we might split a population into N groups that we label $1, 2, \ldots, N$, and call the incidence in the ith group I_i. In this way the incidences are collected into a single object – the vector $\mathbf{I} = (I_1, I_2, \ldots, I_N)$. Vector quantities are usually denoted by bold fonts in typeset text, and either wavy underlining $\underset{\sim}{x}$ or over-arrows $\vec{x}$ in handwritten text.

Multiplication Shorthand

For brevity we usually suppress the $\times$ symbol between two letters and also between a letter and a number (but not between two numbers). For example,
$a \times b = ab$,
$3 \times x = 3x$,
but $2 \times 3 \neq 23$.

Exceptions: a number of functions have a name with multiple letters: $\sin(x)$, $\cos(x)$, $\tan(x)$, $\log_a(x)$, $\ln(x)$, $\exp(x)$; this does not mean those letters are being multiplied together – it's just one function. Sometimes you'll see the brackets dropped with these functions: $\sin(x)$ as $\sin x$, and this doesn't mean the sine function multiplied by x, it still means x is the argument of the sine function.

Problem Session 1

Without using a calculator,

1. Express $\frac{3}{4} + \frac{2}{5}$ as a single fraction.
2. Evaluate
   2.1 $\Big((3+2)\times 3+2\Big) - \Big((3+2)\times(3+2)\Big)$
   2.2 $\frac{11^{11}}{11^{9}}$
   2.3 $(49)^{\frac{1}{2}}$
   2.4 $(-27)^{\frac{2}{3}}$ (hint: $x^{ab} = (x^a)^b = (x^b)^a$)
3. $a = 70$; $b = 3,000,000,000$; $c = 0.00035$.
   3.1 Quote each number to one significant figure, and also to one decimal place.
   3.2 Put each number in scientific notation.
   3.3 Calculate $(ab/c)^2$. (Hint: it's easier when a, b and c are already in scientific notation.)

Inequalities I

Inequalities are based on the following four symbols that we've already seen:

Symbol	Meaning	Examples
$<$	less than	$1 < 2$ (true); $2 < 2$ (false); $3 < 2$ (false)
$>$	greater than	$1 > 2$ (false); $2 > 2$ (false); $3 > 2$ (true)
$\leq, \leqslant$	less than or equal to	$1 \leq 2$ (true); $2 \leq 2$ (true); $3 \leq 2$ (false)
$\geq, \geqslant$	greater than or equal to	$1 \geq 2$ (false); $2 \geq 2$ (true); $3 \geq 2$ (true)

Inequalities II

Inequalities can be chained together to form composite statements. The arrows should all point in the same direction to avoid confusion.

$$1 < 2 < 10 \quad \text{true;}$$
$$1 < 3 < 2 < 5 \quad \text{false.}$$

Inequalities can be manipulated in a similar way to equalities, though sometimes more care is required. Adding to and subtracting from inequalities works just as with equalities:

$$a = b \implies a + c = b + c$$
$$a < b \implies a + c < b + c$$

Inequalities III

When it comes to multiplication, inequalities are trickier than equalities:

$$a = b \implies ac = bc$$

$$a < b \implies \begin{cases} ac < bc & \text{if } c > 0 \\ ac > bc & \text{if } c < 0 \end{cases}$$

In other words, you need to know whether the thing you're multiplying by is negative, and if so, you reverse the direction of the arrow. This is true because 'less than' does not mean 'closer to zero', it means 'less positive' which is equivalent to 'more negative', i.e. further to the left on the number line. An example is helpful: $2 < 3$ but $-2 > -3$.

The same is true for division: dividing by a negative number reverses the direction of the arrow.

Inequalities IV

$R_0 =$ the average number of new infections caused by a typical infected individual, in a totally susceptible population.
If $R_0 > 1$, the epidemic will grow; if $R_0 < 1$, it will shrink. Say a public health intervention prevents a fraction σ of transmissions, so that a typically infected individual now only infects $(1-\sigma)R_0$ other individuals on average in a totally susceptible population. We want σ to be sufficiently large that $(1-\sigma)R_0 < 1$. Rearranging this inequality by multiplying by $-1/R_0$ then adding 1:

$$\sigma > 1 - \frac{1}{R_0} = \frac{R_0 - 1}{R_0}$$

The right-hand side is less than 1, but gets closer to 1 as R_0 increases. This quantifies how it gets harder to control the spread of a disease the larger R_0 is.

Functions I

A function takes a number a gives you a number back.

We use functions to describe the dependence of one quantity on another.

The number you give it is called the function's *argument*.

A function can be described in words: for example "whatever number you give me, I give you it back plus two", or "whatever number you give me, I give you one back four times as big", but for brevity and clarity it's better to put a label on "whatever number you give me", i.e. the argument. Let's call it x. "whatever number you give me, I give you it back plus two" can be written more concisely as $x + 2$; the function for multiplication by 4 can be written as $4 \times x$ etc.

Functions II

If we call the number going into the function x, what do we call the function itself? Two notations are common:

- with x the independent variable – something we can choose freely – we say y is the dependent variable, depending on x.
- $f(x)$, read out loud as "f of x", meaning the function f is a function of x.

Sometimes the notations are mixed: you might see $y = y(x)$ or $y = f(x)$ written to mean y is a function of x.

You'll soon be seeing all sorts of variables, each with their own letter to represent their value; you will have to decide in each case what depends what on what, i.e. what is a function of what.

Simplifying Expressions

Preliminary nomenclature: an expression is a sum of terms; terms are things separated by $+$ signs in an expression. e.g. in the expression $\frac{a-b+c}{d}$, the terms are $\frac{a}{d}$, -$\frac{b}{d}$ and $\frac{c}{d}$.

The expression $2^3 - 2^2$ can be simplified: it equals $8 - 4 = 4$.
$x^3 - x^2$ cannot be simplified unless one picks one specific value of x, because the extent to which the second term cancels the first term depends on x.
$3x^2 - 2x^2$ can be simplified – it equals x^2 – because the two terms are the same function of x but with different *coefficients* (constant multipliers, 3 and -2 here).
By *collecting together like terms* one can simplify expressions and equations.

$$x + 7 - 2\sin x + \frac{1}{x} - 7 + \frac{3x}{2} + \frac{2}{3x} + \sin x = \frac{5x}{2} + \frac{5}{3x} - \sin x$$

Polynomials

A *polynomial* of a variable, x say, is an expression consisting of non-negative integer powers of x.

Recall that $a(b+c) = ab + ac$. Define a to be equal to $d+e$, and consider this equation again:

$$(d+e)(b+c) = b(d+e) + c(d+e) = bd + be + cd + ce$$

In words: each term in one bracket gets multiplied by each term in the other, then everything is added together; this way of saying it remains true when the brackets contain more than two terms. This tells us how we can multiply together polynomials.

$$\begin{aligned}
&(2 + x + 4x^2)(1 + 6x + 3x^2) \\
&\qquad = 2(1 + 6x + 3x^2) + x(1 + 6x + 3x^2) + 4x^2(1 + 6x + 3x^2) \\
&\qquad = 2 + 12x + 6x^2 \quad + x + 6x^2 + 3x^3 \quad + 4x^2 + 24x^3 + 12x^4 \\
&\qquad = 2 + 13x + 16x^2 + 27x^3 + 12x^4
\end{aligned}$$

Understanding Units I

The *dimension* of a quantity X, $[X]$, is the kind of units X is measured in.

The quantities '3 hours', 'seven days' and 'one decade' all have dimension time: [3 hours] = [seven days] = [one decade] = [time].

Some quantities are *dimensionless* – they consistent of a pure number (something on the number line e.g. $1, 2\ldots$) and no units. Examples include

- ratios of quantities that have the same dimension, e.g. the incidence in the UK / the incidence in France;
- relative changes in quantities: $(x_{\text{new}} - x_{\text{old}})/x_{\text{old}}$;
- products of quantities that have inverse dimensions, e.g. $2\ \text{day}^{-1} \times 1$ week.

Dimensionful quantities on the other hand require a unit to make sense, e.g. lengths, times.

Understanding Units II

Note that *per* literally means *divided by*; e.g. one per day $= 1/\text{day}$ $= 1\ \text{day}^{-1}$.

The dimension of the product of two quantities is equal to the product of the dimensions: $[XY] = [X][Y]$.
e.g. distance travelled = speed $\times$ travelling time.

- [distance travelled] = [length]
- [speed] = [length][time]$^{-1}$
- [travelling time] = [time]

Indeed, multiplying the dimensions of the latter two gives the dimension of the first, because the time dimension cancels.

Note that to see the cancellation of dimensions (and get the correct answer!) you might need to convert some units. e.g. sixty miles per hour $\times$ ten minutes is
$\frac{60\ \text{miles}}{1\ \text{hour}} \times 10\ \text{mins} = \frac{60\ \text{miles}}{60\ \text{mins}} \times 10\ \text{mins} = 10\ \text{miles}.$

Understanding Units III

The *units* on both sides of an equation do not have to be equal: 'one week = seven days' is fine. The *dimensions*, however, must be equal. Even stronger, every term in an equation must have the same dimension: if $a + b = c - d + e$, then a, b, c, d and e must all have the same dimension. This is because adding terms of different dimension is meaningless, e.g. 'one mile + one day'.

A tricky point: it's common in this field for people to give you equations in which they have forced a particular choice of units on you: e.g.

$$\begin{aligned}\text{number of pills taken} = {} & \text{number of pills taken in one day} \\ & \times \text{number of days during which pills are taken}\end{aligned}$$

Note that both quantities on the RHS are dimensionless. Better, but unfortunately not always done, is to say

$$\begin{aligned}\text{number of pills taken} = {} & \text{rate of taking pills} \\ & \times \text{time during which pills are taken}\end{aligned}$$

Here, the first quantity on the RHS has dimension $[\text{time}]^{-1}$ and the second [time].

Problem Session 2, I

1. x can be any number greater than or equal to -2, and less than 3: $-2 \leq x < 3$. What inequalities do
   1.1 $x - 3$, and
   1.2 $-2x$

   satisfy?
2. State the dimension, and give an example of an appropriate unit, for the following quantities:
   2.1 the mass of a particular drug required,
   2.2 the number of people infected with a given disease,
   2.3 the length of time between being infected and becoming infectious,
   2.4 the power to which we raise some quantity (e.g. a in x^a)
   2.5 an amount of money,
   2.6 a fractional increase in an amount of money,
   2.7 a rate of fractional increase in an amount of money (e.g. the interest rate on a bank account),
   2.8 the concentration of parasites in a patient's blood

Problem Session 2, II

1. A tiny parasite is found x times in one cm^2. Assuming constant number per unit area, how many would there be in an area that is a square, with each side 2 km long?
2. Fundraising for a vaccination project resulted in N people steadily donating at an average rate r_d over a time T. It costs P to buy an amount D of the vaccine. A single vaccination requires an amount d of the vaccine, and one (disposable) needle which costs p. In the project, n nurses each administer the vaccine at a rate r_v, and are each paid at a rate r_p. A fraction f of all money raised must be used indirectly (for administration etc.). How many people can this project vaccinate? Hint: considering the dimensions of all these quantities helps you to see which ones should be multiplied together.

Exponentials, Logarithms and Trigonometric Functions

Recap: Powers

Recall that

$$x^a = \underbrace{x \times x \times \ldots x}_{a \text{ factors}},$$

$$x^a \times x^b = x^{a+b},$$

$$x^{1/n} = \sqrt[n]{x},$$

where $x^{1/n}$ is the number that gives x when multiplied together n times. Together these rules tell us how to raise x to any positive power we want; e.g. $x^{7/3} = x^{2+1/3} = x^2 \times x^{1/3} = x^2 \times \sqrt[3]{x}$.

Recall also that

$$x^{-a} = \frac{1}{\underbrace{x \times x \times \ldots x}_{a \text{ factors}}},$$

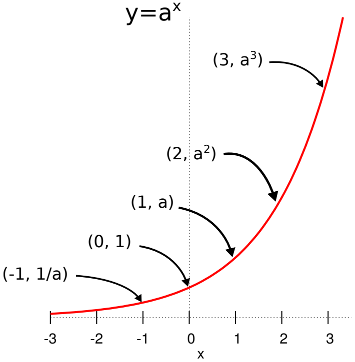
making it clear how to raise x to any negative power we want. e.g.

$$x^{-1.5} = \frac{1}{x^{1.5}} = \frac{1}{x \times \sqrt{x}} = \frac{\sqrt{x}}{x^2}$$

Recall also that $x^0 = 1$.

Exponential Functions

Now consider taking a fixed value for a, $a > 1$, and raising it to the power x, where we allow x to vary. This defines a function of x called *an exponential function*; x is said to increase *exponentially*. Each time we add 1 to x, a^x gets multiplied by a factor a.



Each choice of a defines a different function; the larger a is, the faster the function grows.

There is a fundamental mathematical constant called e, which has an infinite number of decimal places as π does; $e = 2.718$ (3 d.p.). Choosing $a = e$ gives *the* exponential function, $\exp x = e^x$, which has special properties that we'll see later.

A Prelude to Logarithms: Inverse Functions

If the function $f(x)$ maps x onto y (where any x can be chosen), the *inverse function* $f_{\mathrm{inv}}(x)$ maps y onto x. In other words, if you take the output of a function and put it into the inverse function, you get back what you started with. The inverse of the inverse is the original function.

e.g. $f(x) = x + c \iff f_{\mathrm{inv}}(x) = x - c$
e.g. $f(x) = cx \iff f_{\mathrm{inv}}(x) = x/c$
e.g. $f(x) = x^3 \iff f_{\mathrm{inv}}(x) = x^{1/3}$

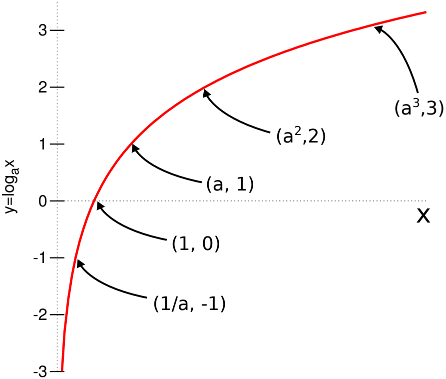
Logarithms I

The following two statements are equivalent:

$$a = b^c \qquad \text{and} \qquad c = \log_b a$$

$\log_b a$ is the power to which one needs to raise b in order to get a; from the first equation you can see that this is c.

$\log_a x$ is the inverse function of a^x; therefore $a^{\log_a x} = \log_a(a^x) = x$.



Whereas a^x grows very quickly (it gets multiplied by a factor a every time x increases by 1), $\log_a x$ grows very slowly: you have to multiply x by a factor a to get it to increase by 1.
e.g. $\log_{10}(10) = 1$,
$\log_{10}(100) = 2$,
$\log_{10}(1000) = 3$, etc.

Logarithms II

$$\text{Recall that} \quad x^a \times x^b = x^{a+b}.$$

$$\text{Take } \log_x \text{ of both sides:} \quad \begin{aligned} log_x(x^a \times x^b) &= \log_x(x^{a+b}) \\ &= a + b, \text{ by definition.} \end{aligned}$$

Define $c = x^a$ and $d = x^b$, i.e. $a = \log_x c$ and $b = \log_x d$. Then,

$$\log_x(cd) = \log_x c + \log_x d$$

i.e. the logarithm of a product equals the sum of the separate logs.
Analogously, because $x^a/x^b = x^{a-b}$ we find that

$$\log_x(c/d) = \log_x c - \log_x d$$

Using similar reasoning one can see that $(x^a)^b = x^{ab}$ implies

$$\log_a(x^n) = n \log_a x$$

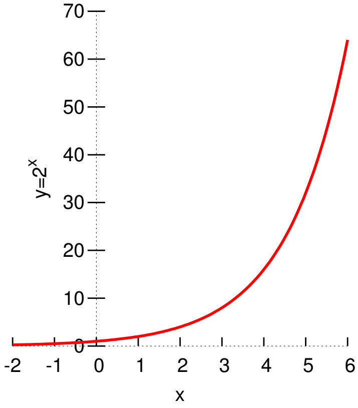
Also note that $x^0 = 1 \implies \log_x 1 = 0$
i.e. the logarithm of 1, with any base for the log, is zero.

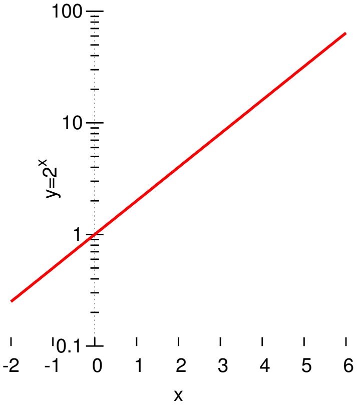
$\log_e(\ldots)$ is called the *natural* logarithm, abbreviated as $\ln(\ldots)$

Logarithmic Scales/Axes

Most commonly used in graphs are *linear axes*: a given distance between two points on the axis corresponds to a fixed *difference* between the values.

A *logarithmic* axis means that a given distance between two points on the axis corresponds to a fixed *multiplicative factor* between the two values. Exponential growth looks linear on a logarithmic scale.





Changing the Base for Exponentials and Logs

$$a^x = b^{(\log_b a)\, x}$$

e.g. choosing $b = e$: $\quad a^x = \exp((\ln a)\, x)$

Note that for a quantity increasing with t as e^{rt}, the constant r is called *the exponential growth rate*.

$$\log_a x = \frac{\log_b x}{\log_b a}$$

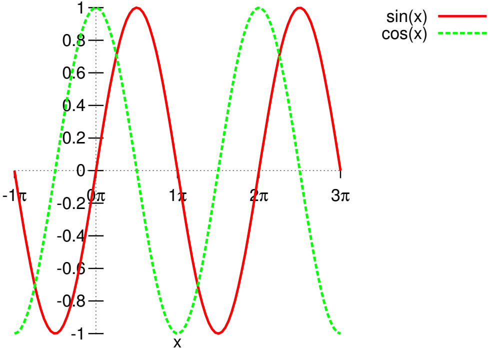
e.g. choosing $b = e$: $\quad \log_a x = \dfrac{\ln x}{\ln a}$

Problem Session 3

- Evaluate 2 raised to the following powers: $-2, -1, 0, 1, 2, 3, 4, 5$
- Solve for x: $3^x = 27, \quad 10^x = 0.001, \quad 100^x = 10$
- The incidence for a particular epidemic, I, grows exponentially with time t as $I(t) = 1 \text{ per day} \times 10^{t \times 2 \text{ per month}}$. What is
  - the incidence after one week (starting from $t = 0$)?
  - the time taken to reach an incidence of 1000 per day?
  - the doubling time (time taken to double)?
  - the exponential growth rate?

Trigonometric Functions: Need to Know

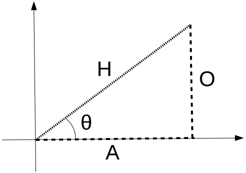
$\sin x$ and $\cos x$ are periodic functions, oscillating between -1 and $+1$ with period 2π. $\sin x$ is 0 when $x = 0, \pi, 2\pi, 3\pi, \ldots$; it equals 1 or -1, alternately, exactly half-way between these points where it is zero. $\cos x$ is the same but shifted along the x axis.



By modifying these functions (stretching & shifting horizontally & vertically – more on this later) we can describe things that vary periodically, in particular environmental factors that vary annually with the changing seasons.

Trigonometric Functions: Nice to Know

Firstly, radians: the angle at the middle of one whole circle is defined to be 360°, or alternatively, 2π radians. All other angles are measured relative to these, e.g. the angle at the middle of one quarter of a circle is 90° or $\frac{\pi}{2}$ radians. Radians are a more fundamental measure of angle for the following reason: for a fragment of a circle, dividing the length of the curved edge by the radius gives the angle at the point in radians. Doing this for a whole circle gives 2π radians because the circumference of a circle is 2π times the radius.

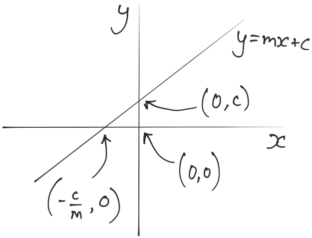


Take a right-angled triangle with a side of length A ('adjacent'), a side of length O ('opposite'), a hypotenuse of length H and an angle θ as shown. $\sin\theta$ is defined to be O/H, and $\cos\theta$ A/H.

In the triangle as drawn, $\theta < \frac{\pi}{2}$ radians; by rotating the line marked H past the positive y axis we can consider larger angles. By rotating it all the round once and then a bit more, we can consider angles larger than 2π radians; A, O and H will be the same as if we hadn't gone round once beforehand, which is why $\sin(x + 2\pi) = \sin(x)$.

Differentiation

A Prelude: The Anatomy of $y = mx + c$



$y = mx + c$ describes a straight line that passes through the points $(\frac{-c}{m}, 0)$ and $(0, c)$.

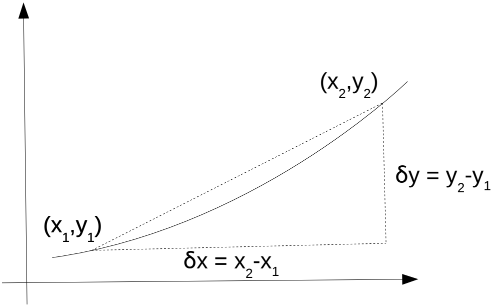
Let us take any particular value of x, x_1 say, and any larger value x_2. y changes from y_1 to y_2. The change in y relative to the change in x is the gradient of this line; it tells us how sensitively y depends on x.

NB Δ is used here (and most other places) to mean 'the change in', and so Δy is really just one quantity expressed using two symbols; it does not mean $\Delta \times y$.

$$\frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1} = \frac{(mx_2 + c) - (mx_1 + c)}{x_2 - x_1} = \frac{m(x_2 - x_1)}{x_2 - x_1} = m$$

This does not depend on x_1 or x_2: the gradient is constant.

The Gradient of a Curve?



The gradient at (x_1, y_1) is $\simeq \frac{\delta y}{\delta x}$.
(Δ means 'a change in'; δ means 'a small change in'.)
The smaller we make δx, the smaller δy becomes and the better the approximation gets. We cannot simply set $\delta x = 0$, because then $\delta y = 0$ too, and $\frac{0}{0}$ is undefined. Instead, we must consider the *limit* $\delta x \to 0$.

The definition of a derivative: $\quad \dfrac{dy}{dx} := \lim_{\delta x \to 0} \dfrac{y(x+\delta x) - y(x)}{\delta x}$

This is the slope / gradient / rate of change of y with respect to x, *defined at one specific point*. Note that unless we are are talking about a straight line, $y = mx + c$, the gradient changes; hence the need to define it at each point separately (letting $\delta x \to 0$).
Example: $y = x^2$

$$
\begin{aligned}
\frac{dy}{dx} &= \lim_{\delta x \to 0} \frac{(x+\delta x)^2 - x^2}{\delta x} \\
&= \lim_{\delta x \to 0} \frac{(x^2 + 2x\delta x + \delta x^2) - x^2}{\delta x} \\
&= \lim_{\delta x \to 0} (2x + \delta x) \\
&= 2x
\end{aligned}
$$

Not Just Geometry

Why are we so interested in finding the slopes of curves?

Differentiation properly establishes the concept of a *rate*: how much / how quickly one variable is changing with respect to another variable (often with respect to time). e.g. incidence is the rate of new infections.

If you find differentiation hard to understand, it is explained slowly in a video tutorial at

http://www.mathtutor.ac.uk/differentiation/differentiationfromfirstprinciples

The Derivatives of Some Common Functions

$f(x)$	$\frac{df}{dx}$
c	0
$mx + c$	m
x^2	$2x$
x^n	nx^{n-1}
e^x	e^x
e^{rx}	re^{rx}
$\sin(x)$	$\cos(x)$
$\sin(ax)$	$a\cos(ax)$
$\cos(x)$	$-\sin(x)$
$\cos(ax)$	$-a\sin(ax)$
$\ln(cx)$	$\frac{1}{x}$

Linearity I

The derivative of a sum is the sum of the derivatives:

$$
\begin{aligned}
\frac{d}{dx}\big(f(x)+g(x)\big) &= \lim_{\delta x \to 0}\left(\frac{f(x+\delta x)+g(x+\delta x)-f(x)-g(x)}{\delta x}\right) \\
&= \lim_{\delta x \to 0}\left(\frac{f(x+\delta x)-f(x)}{\delta x}+\frac{g(x+\delta x)-g(x)}{\delta x}\right) \\
&= \frac{df}{dx}+\frac{dg}{dx}
\end{aligned}
$$

The implication is that you can add two functions and then differentiate, or differentiate each of them and then add, and get the same result. This is what we mean when say that differentiation is a *linear* operation.

Linearity II

Now consider a new function which is some other function multiplied by a constant, A:

$$\begin{aligned}\frac{d(Af(x))}{dx} &= \lim_{\delta x \to 0} \frac{Af(x+\delta x) - Af(x)}{\delta x} \\ &= A \lim_{\delta x \to 0} \frac{f(x+\delta x) - f(x)}{\delta x} \\ &= A\frac{df}{dx}\end{aligned}$$

i.e. if you multiply a function by a constant A, the new function's gradient/slope is A times larger.

Problem Session 4

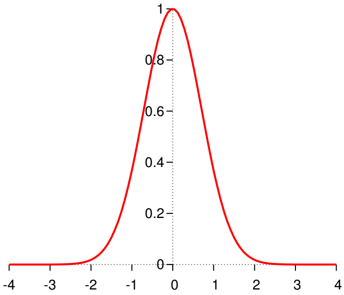
What is the rate of change of a given person's age, with respect to time?

What is the derivative of

- $\frac{3}{8}x^4 + 12$, with respect to x, evaluated at $x = 2$?
- $e^{2\ \text{day}^{-1} t}$ with respect to t?
- $10^\circ\text{C} + 3^\circ\text{C} \sin\left(\frac{2\pi t}{1\ \text{week}}\right)$ with respect to t?

One health centre can provide care for N people, and costs C to build. A new long-term government grant provides a constant source of funding, at a rate F, for building more health centres. What is the rate at which the number of people with access to health care centres increases?

Sketch the derivative of this function:



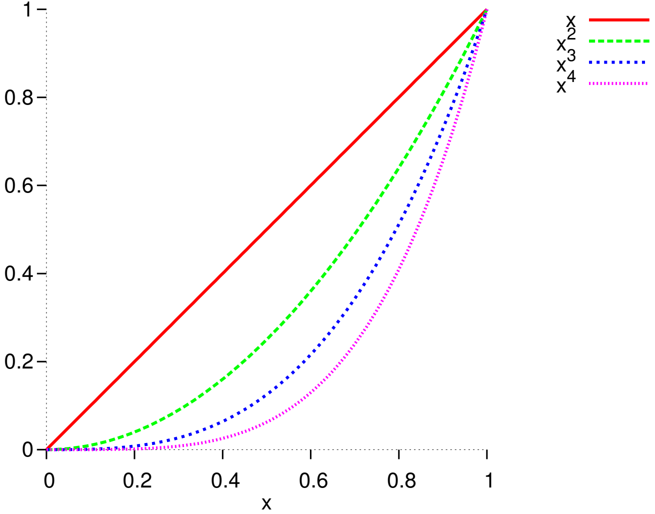
What Functions Look Like

Why Bother Plotting Functions?

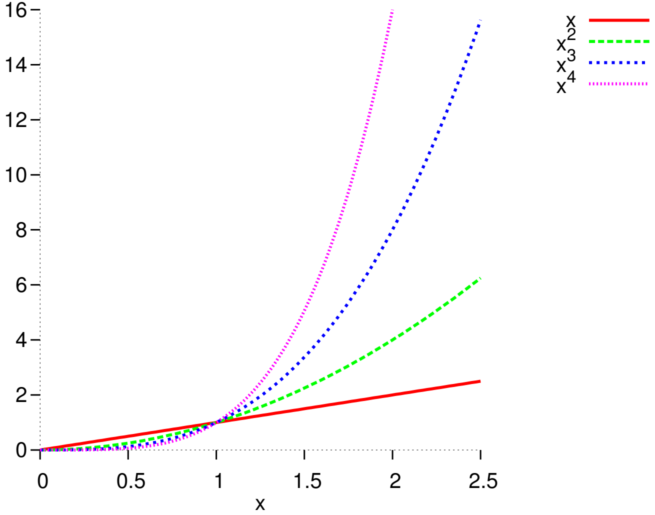
We use equations to specify how different quantities of interest depend on each other.
It's often not clear from the equation how the relationship behaves – where increases and decreases occur, how dramatic they are, etc.
Graphs show us this.

How to Plot Functions

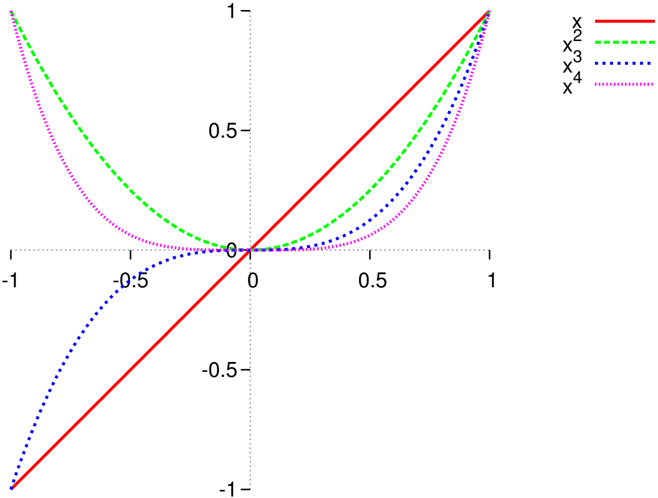
- Identify as many points as possible where you know the precise value of the function. It's sometimes easy to figure out the value of the function when $x = 0$, and/or the value(s) of x that make the function zero.
- Calculate the derivative (gradient) of the function.
  - The derivative $> 0 \iff$ the function is increasing.
  - The derivative $= 0 \iff$ the function is flat/horizontal.
  - The derivative $< 0 \iff$ the function is decreasing.
- Figure out what the function does as x becomes large and positive, and large and negative.
- Figure out if there are any 'singularities' (points where the function 'blows up' or 'diverges') – more on this shortly.



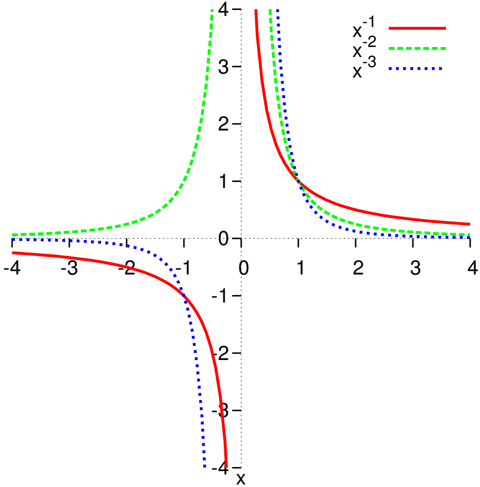
- 0 to any power is 0.
- 1 to any power is 1.
- Numbers between 0 and 1 get smaller the larger the power you raise them to.



Numbers larger than 1 get larger the larger the power you raise them to; x^n grows faster the larger n is.
Positive powers of x grow without bound as x gets larger.

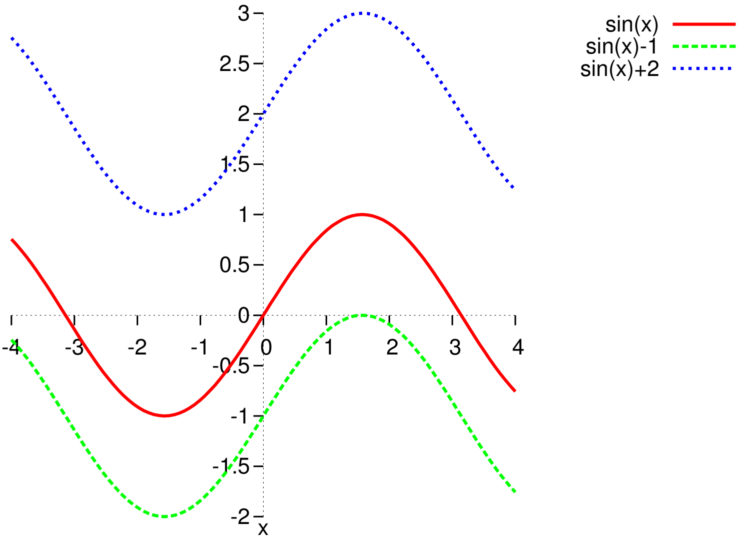


Negative numbers raised to an even power are positive; raised to an odd power they're negative.

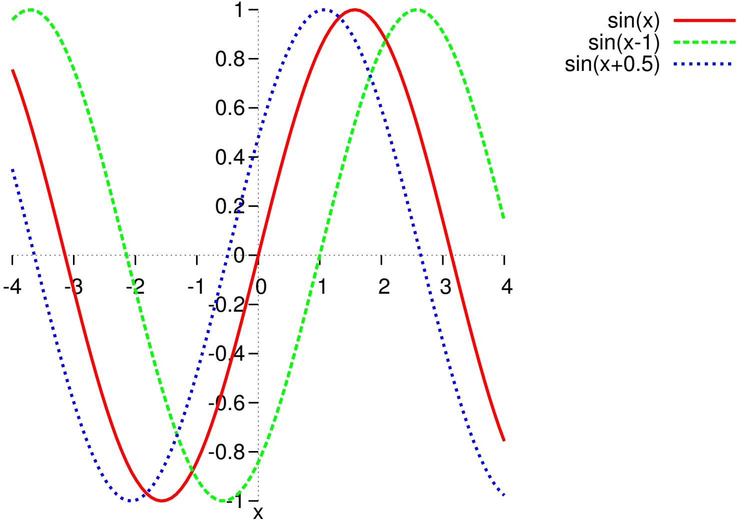


Dividing by an ever smaller number gives an ever larger result, so $\frac{1}{x}$ blows up as $x \to 0$, and $\frac{1}{x^2}$ blows up even faster. Dividing by an ever larger number gives something increasingly close to zero, so negative powers of x tend to zero as x becomes large and positive or large and negative.

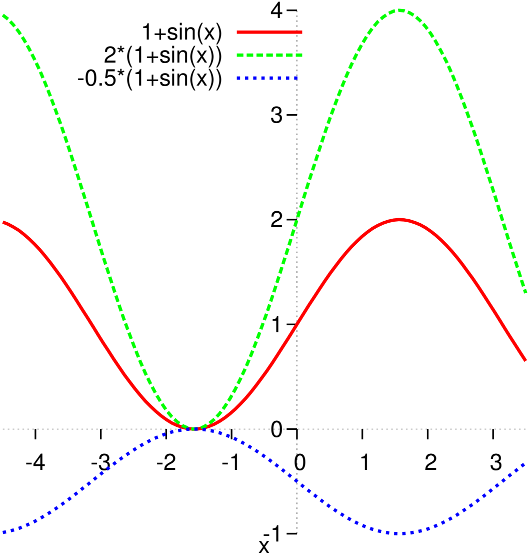
Again, negative numbers raised to an even power (such as -2 shown here) are positive.



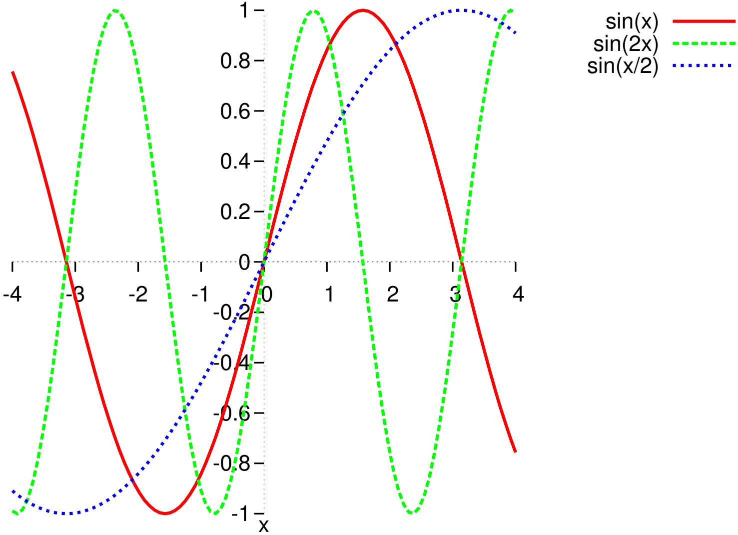
$f(x) + c$ is $f(x)$ shifted upwards by c.



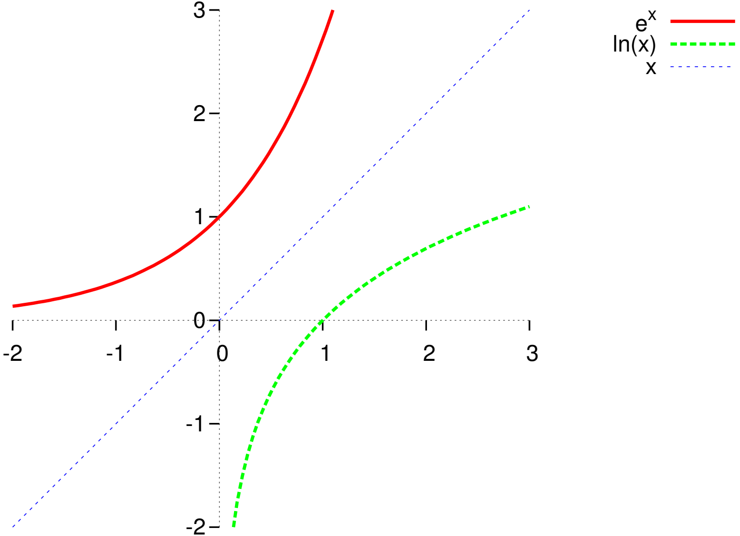
$f(x+c)$ is $f(x)$ shifted to the left by c.



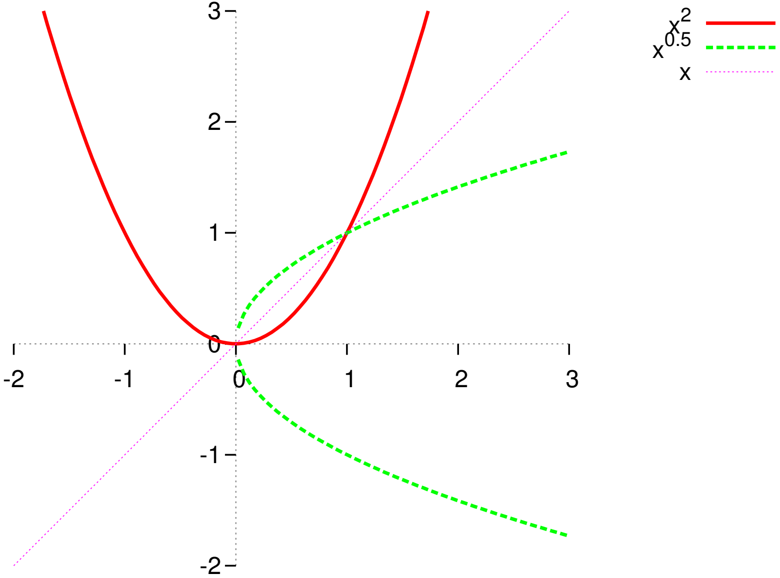
$A\,f(x)$ is $f(x)$ stretched vertically (away from the x axis) by a factor A. $-f(x)$ is the reflection in the x axis.



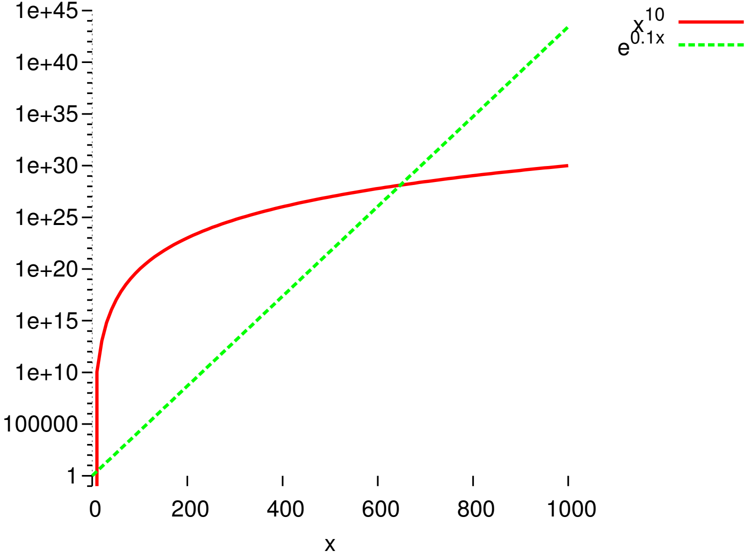
$f(Ax)$ is $f(x)$ squashed horizontally (towards the y axis) by a factor A. $f(-x)$ is the reflection in the y axis.



Where a function takes a value x and returns a value y, the inverse function would take a value y and return x. This means a function and its inverse are reflections of each other in the line $y = x$.



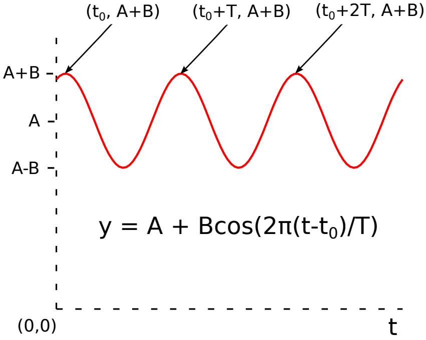
Another function and its inverse.



In the long run, exponential growth is always faster than power growth, no matter how large the power (i.e. n in x^n) and how small 'the exponential growth rate' (i.e. r in e^{rx}).

More General Sinusoidal Functions

- Let $f_0(t) = \cos(t)$.
- Let $f_1(t) = Bf_0(t)$: we stretch vertically by a factor B.
- Let $f_2(t) = f_1(2\pi t/T)$: we squash horizontally by a factor $\frac{2\pi}{T}$.
- Let $f_3(t) = A + f_2(t)$: we shift upwards by A.
- Finally let $y(t) = f_3(t - t_0)$: we shift to the right by t_0.



Property	$\cos t$	this $y(t)$
amplitude	1	B
period	2π	T
average	0	A
first maximum	$t = 0$	$t = t_0$

Problem Session 5

Sketch the following functions:

- $y = 3$
- $x = 2$
- $y = 3 + 3\sin(\pi x)$
- $x^3 - x$
- $\frac{1+x}{1-x}$
- $\frac{e^x}{1+e^x}$

Integration

Integration is the Reverse of Differentiation

For any given function $f(x)$, define a new function $A(x)$ as being the area under $f(x)$ (i.e. between the function and the x axis, with negative values of the function contributing negative area) up to the point x. The fundamental theorem of calculus states that obtaining $A(x)$ from $f(x)$, by definition 'integrating', is the reverse process to differentiation, i.e. $\frac{dA}{dx} = f(x)$.

http://en.wikipedia.org/wiki/Fundamental_theorem_of_calculus#Geometric_meaning

Because integration is the reverse of differentiation, it's what we need to go from equations for the derivatives of quantities (*differential equations*) to the quantities themselves.

Not Just Geometry

Why are we so interested in finding the area under a curve?

Integrating something which is a rate of change gives the total change. e.g. integrating the incidence between two time points gives the total number of new infections in that time window.

Definite VS Indefinite Integration

So, when we integrate a function $f(x)$ we get another function $A(x)$ that (a) gives the area under $f(x)$, and (b) we can differentiate to recover $f(x)$. Thinking about (a), the area under the graph but starting from where? Thinking about (b), we can add any constant and it remains true that differentiating it we recover $f(x)$. So 'the integral' of a function (or more precisely, the *indefinite integral*) is only defined up to an arbitrary additive constant. e.g. $\int x^n dx = \frac{1}{n+1}x^{n+1} + C$, with C arbitrary.

However the area under the graph between two specified points, a and b say, is well-defined: it's $A(b) - A(a)$, regardless of where the function $A(x)$ takes as its left-hand edge. This means the arbitrary constant C will cancel.
e.g. $\int_a^b x^n dx = [\frac{1}{n+1}x^{n+1}]_a^b = \frac{1}{n+1}b^{n+1} - \frac{1}{n+1}a^{n+1}$
is an example of a *definite integral*. The notation $[\ldots]_a^b$ means the difference between the brackets' contents at b and at a.

The Indefinite Integrals of Some Common Functions

$f(x)$	$\frac{df}{dx}$
$mx + c$	m
x^n	nx^{n-1}
e^{rx}	re^{rx}
$\sin(ax)$	$a\cos(ax)$
$\cos(ax)$	$-a\sin(ax)$
$\ln(cx)$	$\frac{1}{x}$

$\Longrightarrow$

$f(x)$	$\int f(x)dx$
m	$mx + C$
x^n	$\frac{1}{n+1}x^{n+1} + C$
e^{rx}	$\frac{1}{r}e^{rx} + C$
$\sin(ax)$	$\frac{-1}{a}\cos(ax) + C$
$\cos(ax)$	$\frac{1}{a}\sin(ax) + C$
$\frac{1}{x}$	$\ln x + C$

Knowing these, we can solve some differential equations. e.g.

$$\frac{dI}{dt} = 2 + 3\sin(2t) + e^{t/3} \implies I = 2t - \frac{3}{2}\cos(2t) + 3e^{t/3} + C$$

With a *boundary/initial condition* – one particular value of the function we are trying to solve for – we can fix the unknown constant C. e.g. If we know that $I = 1$ when $t = 0$:

$$I(0) = -\frac{3}{2}\cos(0) + 3e^0 + C = \frac{3}{2} + C = 1 \quad \therefore\ C = -\frac{1}{2}$$

Problem Session 6

What is the area of a right-angled triangle?

What is the area under

- $\sin x$ between 0 and $\pi/2$,
- e^{-x} between 0 and ∞?

Recall our previous epidemic in which incidence increased as $I(t) = 1 \text{ per day} \times 10^{t \times 2 \text{ per month}}$. What was the total number of new cases in the first month? (Hint: you know how to integrate *the* exponential funtion, i.e. with base e.)

Exponential Growth In Theory

Consider something which grows in proportion to how much of it there already is. e.g.

1. population size, neglecting competition for resources / density dependent mortality;
2. number of actively dividing cells (e.g. cancerous or bacterial) or viruses, with the same caveat as above;
3. number of people infected with a disease, if the probability of contact between two infecteds or an infected and a recovered can be neglected.

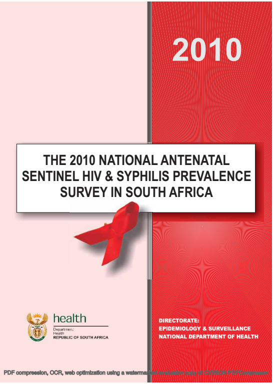
(NB 2 and 3 are really just sub-cases of 1.) Call that something N.

$$\frac{dN}{dt} \propto N \quad \text{or} \quad \frac{dN}{dt} = rN, \quad \text{for some } r$$

$$\text{Integrating:} \quad N \propto e^{rt} \quad \text{or} \quad N = Ae^{rt} \quad \text{or} \quad \ln(N) = rt + \ln(A),$$

for some A.

Exponential Growth In Practise I



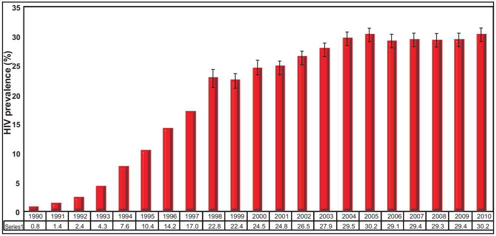
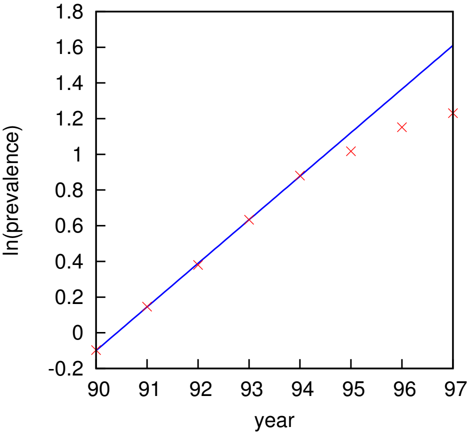


Figure 3: HIV prevalence trends among antenatal women, South Africa 1990 to 2010. The estimates from 2006 are based on a different sample to the previous years.

Exponential Growth In Practise II



$$\underbrace{\hphantom{xxxxxxxxxxxxxxx}}_{R^2 = 0.9998}$$

gradient $= (0.244 \pm 0.002)\ \text{year}^{-1}$

$\ln(\text{prevalence}) = 0.244\,\text{year}^{-1}\ t +$ const.

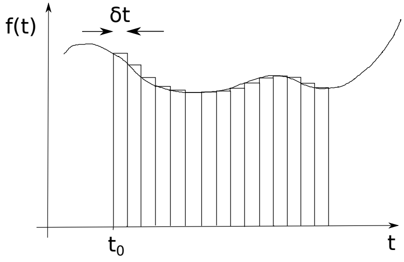
So for $1990 \leq t \leq 1994$,
prevalence $\propto e^{0.244\ \text{year}^{-1}\ t}$

Numerical Integration of Differential Equations

If $\frac{dy}{dt} = f(t)$ and you don't know how to integrate $f(t)$, you must resort to a method of numerical integration. The simplest is the *Euler method*. Let t refer to time for concreteness. We consider only a finite number of timepoints, starting at t_0 and incrementing by a step of size δt. At each timepoint we evaluate the derivative, and approximate it as staying constant over the following timestep, such that the change in y is given by that constant derivative $\times\ \delta t$. At the nth timestep:

$$\begin{aligned} y(t_0 + n\delta t) &= y(t_0 + (n-1)\delta t)\delta t + f(t_0 + (n-1)\delta t)\delta t \\ &= y(t_0 + (n-2)\delta t) + f(t_0 + (n-2)\delta t)\delta t + f(t_0 + (n-1)\delta t)\delta t \\ &= y(t_0) + \delta t \sum_{i=0}^{n-1} f(t_0 + i\delta t) \end{aligned}$$

Geometrically, this corresponds to integrating $f(t)$ by approximating the area under the curve using a series of narrow rectangles. The smaller δt is, the closer this area is to the true area.



The Need for Numerical Integration

If $\frac{dy}{dt} = f(t)$, you can often figure out how to integrate and solve for the function $y(t)$. However it's perfectly possible that the rate of change of y depends on y as well as t: $\frac{dy}{dt} = f(y, t)$. For all but the simplest function f this is much harder to solve, since we have a function of y but we don't know what y is. In models of infectious disease transmission, the situation is typically much worse: we don't have just a single function y that we are trying to solve for, but a collection of variables that are influencing each other. For example the simplest 'SIR' model, where S is the number of individuals susceptible to a given disease, I is the number infected and R is the number recovered, has the differential equations

$$\begin{aligned}
\frac{dS}{dt} &= -\frac{\beta SI}{N} \\
\frac{dI}{dt} &= +\frac{\beta SI}{N} - \gamma I \\
\frac{dR}{dt} &= +\gamma I
\end{aligned}$$

It is still not known how to integrate these equations: we must do so numerically to see how S, I and R change over time.

Discrete Probability Basics

What is Probability?

A meaningless definition, appealing to intuition: probability is a measure of how likely (i.e. how probable) something is.

Notation: $P(A)$ is

- the probability that A is true, or
- the probability that A will happen, or
- the probability of A,

depending on how exactly one phrases what it is that A represents.

The *frequentist* definition of probability, where N is the total number of trials / times an experiment is repeated, is

$$P(A) = \lim_{N \to \infty} (\text{the number of times } A \text{ occurs } / \, N)$$

$0 \le P(A) \le 1$ always: 0 means impossible, 1 means certain.

For probabilities, the percent symbol simply represents the number 0.01; e.g. $5\% = 0.05$ (see discussion in lecture 1 notes).

Independency, Mutual Exclusivity I

Two events being *independent* means they are not connected / do not affect each other. If A and B are possible outcomes from independent events, then $P(A \text{ and } B) = P(A)P(B)$.

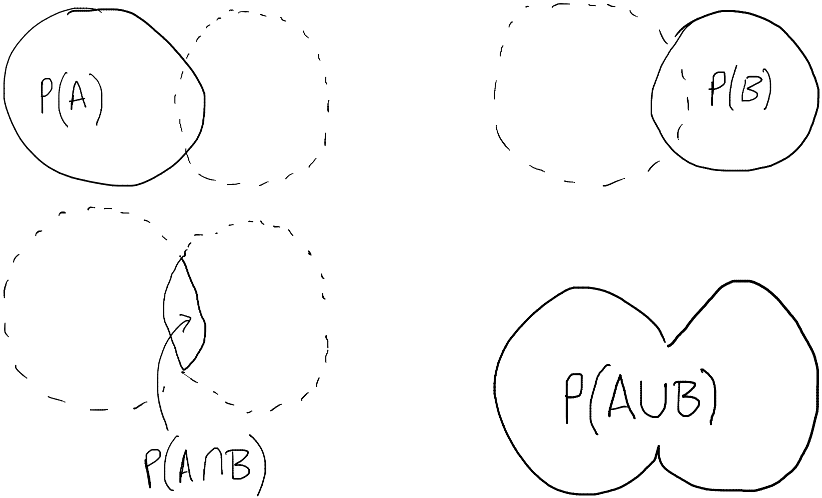
Two outcomes being *mutually exclusive* means that if one occurs, the other cannot: the probability of both is zero.
A and B are mutually exclusive $\iff P(A \text{ and } B) = 0$.

Avoiding double counting: $P(A)$ includes the probability $P(A \text{ and } B)$. $P(B)$ also includes the probability $P(A \text{ and } B)$.

$$\therefore\ P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B), \qquad \text{always.}$$

The subtracted term on the right-hand side vanishes if and only if A and B are mutually exclusive.

Venn Diagrams



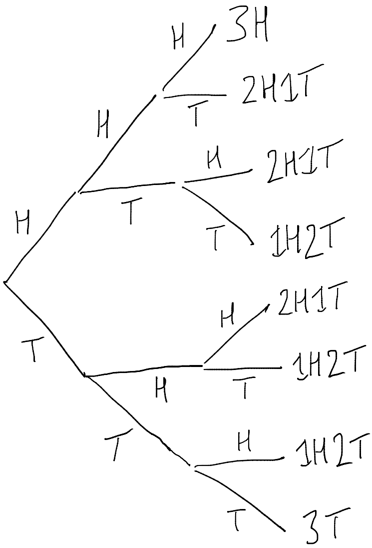
Illustrating why $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
($\cup$ means 'or'; $\cap$ means 'and')

Independency, Mutual Exclusivity II

Exercise: think about which of the following are independent, which are mutually exclusive, and which are neither.

- Person A has flu. Person B has flu. (They live in the same house.)
- Person A has flu. Person B has flu. (They live in different countries.)
- Lab A gives a false positive for the blood test. Lab B gives a false positive for the blood test.
- The patient's age is zero to eighteen. The patient's age is eighteen or older.
- The patient is male. The patient is female.
- Study A concludes something. Study B concludes the same thing. (They used the same patients.)

Illustrating All Possibilities: Trees



Toss a fair coin three times. Recalling that the probabilities of independent outcomes (subsequent tosses) are multiplied, and those of mutually exclusive outcomes (i.e. distinct possibilities) are added, the probability of getting

- 3 heads is $\frac{1}{8}$,
- 2 heads is $\frac{3}{8}$,
- 1 head is $\frac{3}{8}$,
- 0 heads is $\frac{1}{8}$.

Problem Session 7

Three patients are infected with a disease, each having a chance of dying of 1 in 6. What is the probability that at least one of them dies?
(Advanced: if there are N patients, what is the probability that at least one of them dies?)

The Mean

You probably know how to calculate the *mean* of a list of numbers: sum them, then divide by how many there are.
Equivalently: assign each distinct value a probability equal to the number of times it occurs in the list divided by how many values there are in the list, then add all distinct values each multiplied by (or 'weighted by') its probability.
In general, for a discrete random variable X, the mean or expected value is

$$E[X] = \langle X \rangle = \sum_x x\, P(X = x)$$

We multiply each possible value X can take by its probability, then add them all together.

The Variance

More generally, for the expected value of a function of X, we weight the function of x (instead of just x itself) by the probability of having that value of x:

$$E[f(X)] = \langle f(X) \rangle = \begin{cases} \sum_x f(x)\, P(X = x), & \text{if } X \text{ is discrete} \\ \int_x f(x) p(x) dx, & \text{if } X \text{ is continuous} \end{cases}$$

A particularly important example is the *variance*: the expected value of the difference from the mean, squared. This measures how disperse the distribution is. Letting $E[X] = \mu$ for clarity:

$$E[(X - \mu)^2] = \sum_x (x - \mu)^2 \, P(X = x)$$

Combining Calculus and Probability: Survival I

Imagine a state for which the hazard of leaving is a constant, λ. This means that at any time t, *provided you haven't left already*, the probability that you leave in the next small time interval δt is approximately $\lambda \delta t$. The approximation becomes exact as $\delta t \to 0$. (Conversely it gets worse as δt gets larger; for example when $\delta t > \frac{1}{\lambda}$ it tells you the probability is more than 1!). Let $P(t)$ be the probability that you are still in the state at time t. Then the probability that you leave between t and $t + \delta t$ is

$$P(t) - P(t + \delta t) \approx P(t) \times \lambda \delta t$$

The factor of $P(t)$ on the right-hand side accounts for the 'provided' caveat above – it is the probability that you didn't leave before t, which we need to include if we're considering departure between t and $t + \delta t$. Rearranging,

$$\frac{P(t + \delta t) - P(t)}{\delta t} \approx -P(t) \times \lambda$$

Combining Calculus and Probability: Survival II

Letting $\delta t \to 0$, the approximation becomes exact and the left-hand side becomes a derivative:

$$\frac{dP(t)}{dt} = -\lambda P(t) \quad \text{so} \quad P(t) = Ae^{-\lambda t} \quad \text{for some } A$$

Defining t more precisely as the time since you were last known to be in the state, $P(0) = 1$ (an *initial condition*) and so $A = 1$.

NB probabilities for an individual = fractions of a population, assuming individuals are independent.

It was the (*memoryless* or *Markovian*) assumption that the hazard of leaving the state is independent of the time spent in the state that forced the survival distribution to be an exponential decay.

Matrix Basics

What's a Matrix?

As mentioned before, we can collect together a set of N numbers into a single object called a vector, $\mathbf{v}$ say, the ith component of which we call v_i: $\mathbf{v} = (v_1, v_2, \ldots v_N)$.
We can also collect together a set of $P \times Q$ numbers into a rectangular array called a matrix, $\mathbf{M}$ say; the element (number) in row i and column j is called M_{ij}:

$$\mathbf{M} = \begin{pmatrix} M_{11} & M_{12} & \ldots M_{1P} \\ M_{21} & M_{22} & \ldots M_{2P} \\ & \vdots & \\ M_{Q1} & M_{Q2} & \ldots M_{QP} \end{pmatrix}$$

To Multiply a Vector by a Matrix

We can multiply a vector **v** by a matrix **M**, provided the width of the matrix equals the number of components of the vector. The result is a vector $\mathbf{M} \cdot \mathbf{v}$ of the same height as the matrix.

How do we do this?

In words: multiply each component of the original vector by each element of the *first* row of the matrix, pairwise, then sum them: this gives the *first* component of the resulting vector. Then multiply each component of the original vector by each element of the *second* row of the matrix, pairwise, then sum them: this gives the *second* component of the resulting vector, etc.

As an equation: the ith component of the resulting vector is

$$(\mathbf{M} \cdot \mathbf{v})_i = \sum_j M_{ij} v_j$$

An example – a 3×3 matrix:

$$\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} ax_1 + bx_2 + cx_3 \\ dx_1 + ex_2 + fx_3 \\ gx_1 + hx_2 + ix_3 \end{pmatrix}$$

The Next-Generation Matrix $\mathbf{K}$, part I

Split the population up into different types. Define K_{ij} as the total number of i-type people (or people in group i, if you like) infected by one j-type person throughout the course of their infection.

i and j run over the same set of possible types (type '1', type '2' etc.) so $\mathbf{K}$ is square. The diagonal entries of $\mathbf{K}$ are the numbers of infections people transmit to the same type as themselves; off-diagonal entries enumerate transmissions between different types. A *diagonal matrix* (all off-diagonal elements are zero) describes no mixing: people only transmit to their own group.

$\sum_i K_{ij}$ (i.e. the sum of the jth column) is the total number of people infected by one j-type person throughout the course of their infection.

e.g. $\quad \mathbf{K} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \implies$ a person in group 1 infects a people in group 1 and c people in group 2; a person in group 2 infects b people in group 1 and d people in group 2.

An Aside: Understanding a *Generation* of Infecteds

Consider the exact analogy between infection as a process of 'epidemiological birth' (producing a new infected individual) and the process of actual birth, i.e. population growth. Date of infection/transmission corresponds to date of birth. Just as the individuals in one generation of a family do not all have the same age, one generation of individuals has a variety of dates of infection. Individuals in one generation of a family may even have non-overlapping lives: a child may die before his/her sibling is born, giving non-overlap within generation one; death of a grandchild before his/her cousin is born gives non-overlap in generation two; etc. Thus, infecteds in the same generation can have infections that do not overlap in time.

As the Nth generation of a family can be connected to the 'root' of the family through N processes of birth but no fewer, the Nth generation of infecteds in an epidemic can be connected to the founder through N transmission contacts but no fewer.

The Next-Generation Matrix $\mathbf{K}$, part II

If we write the numbers of people of each type, in a given generation of infecteds, in a vector $\mathbf{N} = (N_1, N_2, \ldots)$, then the numbers in the next generation are given by the vector $\mathbf{K} \cdot \mathbf{N}$. The numbers in the generation after that are given by the vector $\mathbf{K} \cdot (\mathbf{K} \cdot \mathbf{N})$. Multiplication by $\mathbf{K}$ gives the next generation, hence the name.

Problem Session 8

A population has two kinds of people, A and B. A-type people, when infected, will go on to infect 2 A-type people and 0.5 B-type people, on average. B-type people infect 1 B-type person and 1 A-type person. An epidemic starts with one of each kind of person infected (the 'zeroth' generation). How many people of each type are infected in the second generation?