Calculus Revisited

Chris Wymant

Department of Infectious Disease Epidemiology,
Imperial College London

February 2015

Acknowledgement: I wrote this while funded by ERC Advanced Grant PBDR-339251.

Recap

Definition of a derivative: $$\frac{dy}{dx} := \lim_{\delta x \to 0} \frac{y(x + \delta x) - y(x)}{\delta x}$$

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: $y = x^2$

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

(Remember that $\delta x \neq \delta \times x$: it is a single quantity written using two symbols, representing "a small change in x".)

Differentiating x^n

The importance of this result will become apparent at the end.

$$\begin{aligned}
\frac{d(x^n)}{dx} &= \lim_{\delta x \to 0} \frac{(x+\delta x)^n - x^n}{\delta x} \\
&= \lim_{\delta x \to 0} \frac{\left(\sum_{i=0}^{n} {}^nC_i \, x^i \delta x^{n-i}\right) - x^n}{\delta x} \\
&= \lim_{\delta x \to 0} \frac{x^n + nx^{n-1}\delta x + \mathcal{O}(\delta x^2) - x^n}{\delta x} \\
&= nx^{n-1}
\end{aligned}$$

where ${}^nC_i$ is the number of ways one can choose i things from n things (the *Binomial Coefficient*) and $\mathcal{O}(\delta x^m)$ means any collection of terms that all contain a factor of δx raised to a power at least as large as m.

(Discussion of what the hell just happened.)

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. (In general, however, the order in which you perform operations changes the result – consider something as simple as 'add 1' and 'multiply by 2'.) This is what we mean when say that differentiation is a *linear* operation.

By induction you can see that this holds when adding more than two functions – let $g(x)$ be the sum of two other functions, and repeat.

Linearity II

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

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

Which is intuitive, hopefully: if you multiply a function by a constant A, the new function's gradient/slope is A times larger. (Actually, this follows immediately from the previous slide, since we could have considered differentiating the same function added together A times.) Combining this rule with the previous slide:

$$\frac{d}{dx}\big(Af(x) + Bg(x)\big) = A\frac{df}{dx} + B\frac{dg}{dx},$$

a general statement of the property of linearity.

Integration is the Reverse of Differentiation

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

(Discussion.)

Recap: 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.

Not Just Geometry

Why are we so interested in finding the slopes of curves and the areas under them?

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.

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.

Recap: The Chain Rule

If we can write y as a function of g only, and g depends on x only (so that y really only depends on x), then

$$\frac{dy}{dx} = \frac{dy}{dg}\frac{dg}{dx}, \tag{1}$$

the *chain rule*. Note that this sort of looks obvious, because "it looks like dg should cancel". In fact it is a non-trivial result (which I won't prove here) because $\frac{dy}{dg}$ does not mean dy divided by dg – it is defined as a limit, see the very first thing written on the first slide. The fact that it looks like dg should cancel can help you to remember the equation, however.

e.g. $y = (\sin x)^3$: write $y = g^3$, with $g = \sin x$. Using (1):

$$\begin{aligned}\frac{dy}{dx} &= \frac{d(g^3)}{dg}\frac{d(\sin x)}{dx} \\ &= 3g^2 \cos x \\ &= 3(\sin x)^2 \cos x\end{aligned}$$

The concept of an operator I

Recap: a *function* is something to which we give a number, from which we get a number.
Some functions in action:
$f(x) = 2x + 1: \quad f(0) = 1,\ f(1) = 3,\ f(10) = 21$
$sin(x): \quad sin(0) = 0,\ \sin(1) = 0.84147(5d.p.),\ \sin(\pi/2) = 1$

An *operator* is something to which we give a function, from which we get a function.
Any function can also be an operator. For example consider $\sin(x)$ as an operator: if we give it the function $x + 2$, we get back the function $\sin(x + 2)$. If we give it e^x, we get back $\sin(e^x)$.
However we can have other kinds of operators: notably $\frac{d}{dx}$.

The concept of an operator II

We give the operator $\frac{d}{dx}$ any function and get back another function, equal to first one's derivative or gradient.
Just as you can't say what number $\sin(x)$ is without specifying what the number x is, it does not make sense to ask what "$\frac{d}{dx}$" is equal to: it needs a function to operate on.

Now, each time you've written something like $y = x^2 \implies \frac{dy}{dx} = 2x$, or even $y = mx + c \implies \frac{dy}{dx} = m$, what you've really done is the same thing you're always doing in maths: *doing the same thing to both sides of the equation*. In this case, applying the operator $\frac{d}{dx}$ to both sides.

The Chain Rule Revisited I

When we have $y = y(g)$, $g = g(x)$,

$$\frac{dy}{dx} = \frac{dy}{dg}\frac{dg}{dx},$$

$$\text{or} \quad \frac{d}{dx}y = \frac{dg}{dx}\frac{d}{dg}y,$$

$$\text{so} \quad \frac{d}{dx} = \frac{dg}{dx}\frac{d}{dg}$$

This last form is an equality of operators. ($3 = 2 + 1$ is an equality of numbers; $(x+1)^2 = x^2 + 2x + 1$ is an equality of functions.)

Similarly, thinking only about a relationship between y and x, forgetting about g, we have

$$\frac{d}{dx} = \frac{dy}{dx}\frac{d}{dy}$$

This means that if you start with some function of y equals some other function of x, you don't have to rearrange it to the form $y = f(x)$ before you can differentiate it...

The Chain Rule Revisited II

... for example:

$$
\begin{aligned}
e^y &= x \\
\frac{dy}{dx}\frac{d}{dy}e^y &= \frac{d}{dx}x \\
\frac{dy}{dx}e^y &= 1 \\
\frac{dy}{dx} &= \frac{1}{x}
\end{aligned}
$$

In this way you would have learned what the derivative of $\ln x$ was, if you didn't know already (because $e^y = x \implies y = \ln x$).

Exponential Growth In Theory

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

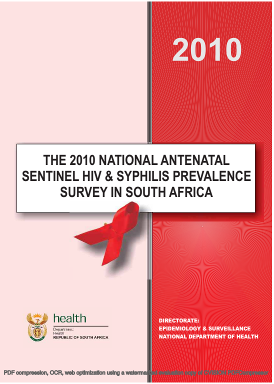
1. population size, neglecting competition for resources / density dependent mortality;
2. number of actively dividing cells (e.g. cancerous or bacterial) or viruses, with the same caveat as above;
3. number of people infected with a disease, if the probability of contact between two infecteds or an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

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

$$\frac{dN}{dt} \propto N \quad \text{or} \quad \frac{dN}{dt} = rN, \quad \text{for some } r$$
$$\implies N \propto e^{rt} \quad \text{or} \quad N = Ae^{rt} \quad \text{or} \quad \ln(N) = rt + \ln(A),$$

for some A.

Exponential Growth In Practise I



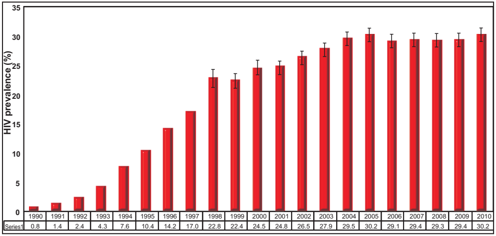
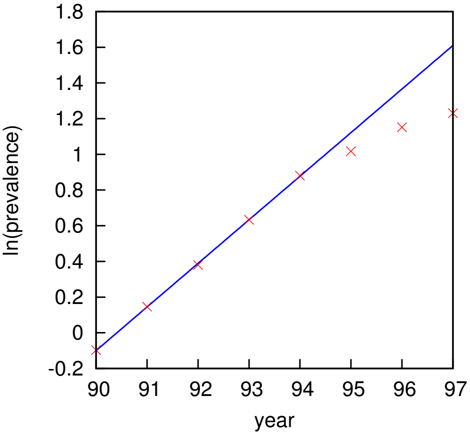


Figure 3: HIV prevalence trends among antenatal women, South Africa 1990 to 2010. The estimates from 2006 are based on a different sample to the previous years.

Exponential Growth In Practise II



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

$$\underbrace{\qquad\qquad}_{R^2 = 0.9998}$$

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

So for $1990 \leq t \leq 1994$,

prevalence $\propto e^{0.244\ \text{year}^{-1}\ t}$

Exponential Decay of Survival Probabilities 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,

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

Exponential Decay of Survival Probabilities 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. If in reality we have a different survival distribution, you need[1] a *non-Markovian* description, for which what happens next depends on the past as well as the present.

---

[1] A possible alternative is replacing one state by several, with different hazards for leaving.

Equilibrium

(Also called the *steady state.*)

Wiktionary: "the condition of a system in which competing influences are balanced, resulting in no net change".

Example 1: constant population size $\iff$ rate of birth + immigration = rate of death + emigration.
Example 2: a population is split into multiple groups, and people are moving from one group to another, but in such a way that the numbers in each group remain constant.

Equilibrium $\implies \frac{d}{dt}$(any system variable) $= 0$
Imposing this condition for all of the system variables will typically give a set of simultaneous equations.

Aside: Simultaneous Equations

N equations are generally required to solve for N variables simultaneously. e.g. $N = 2$:

$$3x + 2y = 7,$$
$$x + 5y = 4$$

Rearranging the first equation gives $x = \frac{1}{3}(7 - 2y)$; substituting this into the second gives

$$\frac{1}{3}(7 - 2y) + 5y = 4,$$
$$\therefore\ y = \frac{5}{13}$$

Substituting this into one of the two original equations gives $x = \frac{27}{13}$.

For $N \geq 3$ you could rearrange your first equation to get one variable as a function of all the others, then substitute into the second; then rearrange this to get a different variable as a function of all the others, and substitute into the third; ... Or use a computer.

Equilibrium Example I

The simplest SIR model, with no demography:

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

Equilibrium $\implies$ all these equations vanish. This happens if and only if $I = 0$.

Equilibrium Example II

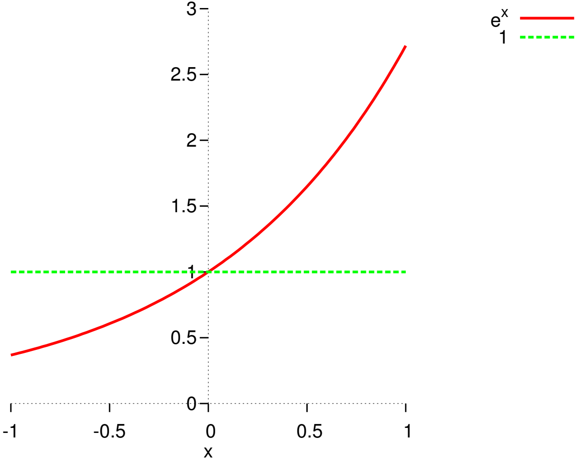
The simplest SIS model, with no demography:

$$\frac{dS}{dt} = -\frac{\beta SI}{N} + \gamma I$$
$$\frac{dI}{dt} = +\frac{\beta SI}{N} - \gamma I$$

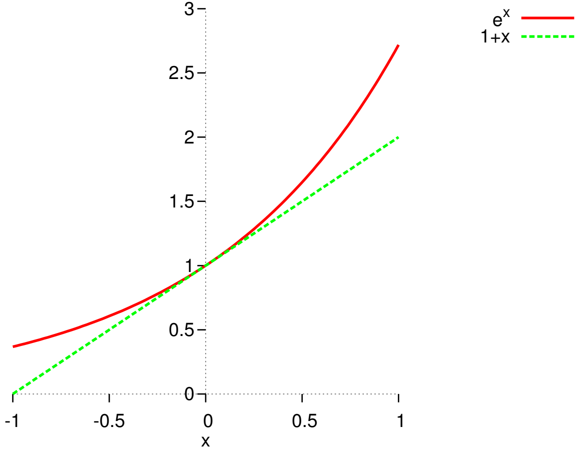
Equilibrium $\implies$ both of these equations vanish $\implies$ either $S = N\gamma/\beta$, and $I = N - S = N(\beta - \gamma)/\beta$; or $I = 0$.

Why is the size of γ relative to β important?

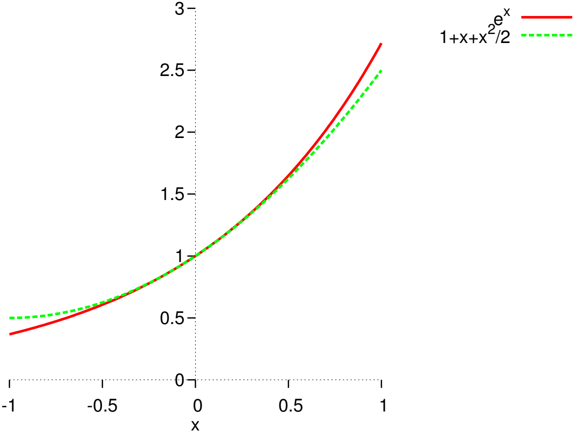
Recap: Functions Can Be Approximated By Polynomials I



Recap: Functions Can Be Approximated By Polynomials II



Recap: Functions Can Be Approximated By Polynomials III



Functions as Infinite Polynomials

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots \tag{2}$$

where $n! = n \times (n-1) \times (n-2) \dots \times 2 \times 1$, and $0! = 1$.
An exact equality, not an approximation. NB the right-hand side contains an infinite number of things added together!
This is a *Power* or *Maclaurin* Series. Using Eq. (2) and linearity:

$$\begin{aligned}
\frac{de^x}{dx} &= \sum_{n=0}^{\infty} \frac{1}{n!} \frac{dx^n}{dx} \\
&= \sum_{n=1}^{\infty} \frac{1}{n!} n x^{n-1} \\
&= \sum_{n=1}^{\infty} \frac{1}{(n-1)!} x^{n-1} \\
&= \sum_{n=0}^{\infty} \frac{x^n}{n!} = e^x
\end{aligned}$$

Functions as Infinite Polynomials: Homework

... so, reassuringly, differentiating this power series we get back the same thing.

Homework for the keen: in exactly the same way, differentiate

- $\sin x = \sum_{n=0}^{\infty}(-1)^n \frac{x^{2n+1}}{(2n+1)!} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots$
- e^{2x} (hint, in Eq. (2) on the last slide, carefully replace x by $2x$)

and integrate
$\cos x = \sum_{n=0}^{\infty}(-1)^n \frac{x^{2n}}{(2n)!} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots$.

For the curious: using the simple rules for differentiating exponentials, cos and sin, differentiate e^{ix}, and $\cos x + i \sin x$, where i is the constant $\sqrt{1}$ (so $i^2 = -1$). Compare what you have before and after in each case. What do you notice? Try writing down their power series.