

Session 1: Modelling foundations and frameworks

Mathematics and mathematical modelling are the perfect companion to experimental results and observations. In this first session, we will start by introducing some foundations of modelling which will allow us to formulate our own models. You may know some of the ideas discussed today, but we want to get everyone to the same level!

1.1: Why modelling?

So, why would you want to write down and analyse a mathematical model anyway? You have done your experiments and have some results that seem to make sense biologically, so what is the point? Mechanistic mathematical modelling allows us to incorporate a list of biological observations and/or assumptions, convert them to mathematical ones, and conduct mathematical experiments to determine their effects. It's not simply about replicating an experiment's results and confirming them with maths, but to go beyond. This might be to extrapolate beyond the constraints of an experiment, to test the relative contributions of each observation or to give understanding to why we see an observation.

As an example, in recent work, I have been looking at the effects that a defensive symbiont which protects its host by making transmission by a harmful parasite more difficult, can have on the host population, both ecologically and evolutionarily. It has been observed in nature that this protection mechanism can be selected for, and can cause hosts to benefit from the mutualistic relationship between it and the symbiont. However our modelling suggests that for this to occur, the cost to the symbiont must be sufficiently small, otherwise the symbiont can drive itself extinct.

This result and interpretation comes with some caveats - the biggest being that we have assumed that the symbiont suffers a cost to its transmission as it invests more into protecting its host, and so our results must be interpreted according to that. Mathematical modelling is not a silver bullet - it can't solve every problem - but it can go some way to increase our understanding.

1.2: Mathematical foundations

In this section we will introduce mathematical foundations. This will mostly be definitions for various mathematical concepts, but we will also introduce some mathematical ideas that are commonly used, and we will employ throughout these sessions.

1.2.1: Definitions

Definition: Function

A function can be thought of as a machine which is provided with something and then outputs something different. We often use the letters f , g or h to denote a function, with the input in brackets alongside (for example $f(y)$ says apply the function f to the input y). Some examples might be the identity function $f(y) = x$ (the right-hand side of the equality is the output), the quadratic function $f(y) = y^2$ or the exponential function $f(y) = \exp\{y\} = e^y$.

Definition: Derivative

The derivative of a function with respect to its input is the *gradient* of the function. We would write, for example, the derivative of the function f with respect to its input y as

$$\frac{df}{dy} = f'(y).$$

This derivative can also be thought of as a flux or flow. For example, if the input of the function is time, t , then the derivative $f'(t)$ can be the change per unit time of the function. If you are interested, the exact definition of the derivative is

$$\frac{df}{dy} = \lim_{h \rightarrow 0} \left(\frac{f(y+h) - f(y)}{h} \right),$$

although this is not important for these sessions.

Definition: Equilibria/steady state

An equilibria or a steady state (both mean the same thing, and I will use interchangeably) is a state where it no longer changes in time, and is usually considered as the temporal "endpoint" of a model. A second way of thinking of this is that if I start with a population of size N , and that N is an equilibrium, then the population will remain at size N for all time. This doesn't mean that hosts become immortal and can never die, but what it does mean is that if a host dies, they will be replaced by a new one, meaning that we always remain at N individuals.

1.2.2: Common modelling ideas

Now that we have some groundwork, we are able to introduce some concepts used when we model mechanistically. These may be principles which guide our approach, or a common type of assumption which deserve a little more attention and will be used commonly

Occam's razor

Occam's razor is what should guide any form of modelling, or science in general, and says that if you have two competing explanations for the same phenomena, then the one that requires the fewest assumptions is likely to be the correct one. In a practical sense, this means that we should always start with the fewest possible number of mathematical assumptions to explain something and verify if those results make biological or mathematical sense. If they don't, then we can add further assumptions until we are happy.

The law of mass action

The law of mass action is a common assumption that is used when two or more agents/species/individuals are to interact. It states that the number of possible interactions between the different agents is proportional to the product of the number of the different agents. For example, consider an infection event where the disease is passed on through direct contact between a susceptible and infected individual. Then the number of those interactions between S susceptible and I infected is proportional to SI .

Others

There are many other interaction types that exist, which we will not look at in too much detail here, but will be mentioned if you wanted to know more. For enzymatic interactions, you may choose a "Michealis-Menten" type interaction, for effects where low densities cause a drop in population size you can use "allee effects". These might occur when considering the pollination of plants. If there are a small number of plants, the chances of significant pollination, and hence reproduction, will be low, and so the population may decrease. Once you increase above a critical mass, pollination becomes sufficient for population growth and the population increases.

1.3: Mathematical frameworks

We now begin to look at different frameworks that we can use to mathematically model. Technically, statistical modelling will fall into this category as a framework, however as we are interested in mechanistic modelling, we will restrict our view to continuous and discrete time modelling.

1.3.1: Continuous vs discrete time

We differentiate these two modelling frameworks by how they treat and progress time. We will begin with discrete time as it is intuitively easier to understand. We will then move on to continuous time systems.

When we model using a discrete-time framework, we are making the implicit assumption that our dynamics (or time evolution) of our system happens in generations, and that the number of entities at a current generation depends on the previous generations. Mathematically, suppose we are interested in the number of something labelled as X at a generation n , which we label as X_n . Then we write a *difference equation*, or DE for short, for the evolution of X :

$$X_{n+1} = f(X_n, X_{n-1}, \dots, X_1, X_0),$$

with an initial condition given by $X_0 = x$. Typically, the value of X_{n+1} will depend on the most recent few generations, so from now on, we will assume that

$$X_{n+1} = f(X_n); \quad X_0 = x.$$

We will see some examples of this in section 1.3.2.

For a continuous time system, we will consider the time evolution of a quantity X at a time t , denoted as $X(t)$. To do this, in the same way as we have for the discrete time system, we need an idea of where the solution is going. For the discrete time system, this was through X_{n+1} , but for a continuous time system, this uses the derivative:

$$\frac{dX}{dt} = f(X),$$

which has an initial condition $X(0) = x$. Again, we will see some examples of this in section 1.3.3 and section 1.4.

1.3.2: Common DEs

Fibonacci numbers

One of the first things we learn in algebra are the Fibonacci numbers, named after Leonardo Fibonacci and was used to model a population of rabbits. Consider a population of rabbits which always come in pairs. We assume that a newly born breeding pair are placed in a field and each breeding pair mates at the age of one month. At the end of the second month (and every subsequent month), they produce exactly one pair of rabbits. This continues indefinitely, with no death. If we denote by F_n the number of pairs of rabbits after n months, then:

$F_0 = 1$	(First pair)
$F_1 = 1$	(Original pair mate, but won't produce for another month)
$F_2 = 2$	(Original pair produce, creating new pair)
$F_3 = 3$	(Original pair produce, new pair mate)
$F_4 = 5$	(Original and new pair produce)
...	
$F_n = F_{n-1} + F_{n-2}$	

This is a DE, but because it depends on the previous two states, we need to specify two initial condition, usually $F_0 = F_1 = 1$. This has a solution called the Binet formula:

$$F_n = \frac{\varphi^n - (-\varphi)^{-n}}{\sqrt{5}},$$

where φ is the golden ratio and satisfies $\varphi^2 - \varphi - 1 = 0$.

Beverton-Holt model

The Beverton-Holt model is a classic difference equation which models the growth of a population. It is given by the following:

$$X_{n+1} = \frac{rX_n}{1 + X_n/M}. \quad (1.1)$$

Suppose that $r = 2$ and $M = 100$ with an initial condition $X_0 = 1$. Then calculating the first few steps we obtain:

$$\begin{aligned} X_1 &= \frac{2X_0}{1 + X_0/100} = \frac{2}{1.01} \approx 1.98 \\ X_2 &= \frac{2X_1}{1 + X_1/100} = \frac{4/1.01}{103/101} = \frac{400}{103} \approx 3.88 \\ X_3 &\approx 7.48 \\ X_4 &\approx 13.91 \\ &\dots \\ X_n &\rightarrow 100 \quad \text{as } n \rightarrow \infty. \end{aligned}$$

In general, this has a solution

$$X_n = \frac{M(r-1)X_0}{X_0 + (M(r-1) - X_0)r^{-n}},$$

and the long term population is $M(r-1)$, known as an environmental capacity of the population.

1.3.3: Exponential growth

Assume that a population, denoted by $N(t)$ at time t , reproduces with a rate a , that is the average time to reproduce for any given member of the population is $1/a$. Then the rate of change at time t is given by the rate of growth a , multiplied by the number of individuals at time t , $X(t)$:

$$\frac{dX}{dt} = f(X) = aX; \quad X(0) = x.$$

Think of it this way, each individual in the population gives birth with rate a , so the total number of births per unit time is aX . We won't go through in detail how to solve a differential equation, however if you are familiar with calculus, it involves integrating the both sides with respect to time (this ODE is known as separable). However I will tell you the solution which is

$$X(t) = x \exp\{at\},$$

and is otherwise known as exponential growth.

1.4: A one-dimensional example

We will now create a one-dimensional population model which will incorporate both a simple birth and death process. We do this to firstly give a brief look at assumptions which we will continue in the second session, and then to demonstrate some of the things we have learnt today.

1.4.1: An introduction to assumptions

We will make a few assumptions from a biological point of view and then see how these convert to mathematical assumptions. Biologically, suppose we want to investigate how the size of a population is changing over time, and that we only want to look at the birth and death processes. We will make the following biological assumptions:

1. We have a single population who are all the same as each other (i.e. we have no heterogeneity in the population).
2. The only processes we care about are births and deaths. We will assume that births are asexual and occur with a constant rate.
3. All individuals have the same average life-span.

These biological assumptions then convert mathematically in the following way, using Occam's razor and the simplest possible assumptions we can make:

1. We will have a single population, which we will label as N and changes in continuous time t .
2. We suppose that the birth rate b per unit time. The rate of change on the number of individuals caused by births is $bN(t)$.
3. We will write the average life-span of the hosts to be $1/d$ time units. This means that the rate of death that each individual suffers is d (If the lifespan is 2 days, then the rate at which each host dies is 0.5 per day). The rate of change multiplied by the number of individuals is removed per unit time.

These three assumptions yield the following ODE:

$$\frac{dN}{dt} = bN - dN = (b - d)N. \quad (1.2)$$

Note that this is the same as the exponential growth ODE from before, where we replace the a with $b - d$. This means that we know what the solution to this is:

$$N(t) = \exp\{(b - d)t\}.$$

We can look at the long term behaviour of this, which means we investigate what happens as $t \rightarrow \infty$. If we have birth rate larger than death rate, so $b > d$, then $b - d > 0$ and so $N(t) \rightarrow \infty$. The fact that the birth rate being bigger than the death rate yields a population which gets bigger over time makes biological sense, however the fact it grows without bound is problematic. We will revisit this shortly. If the birth rate is smaller than the death rate, $b - d < 0$ and $N(t) \rightarrow 0$, and the population goes to extinction which also makes biological sense.

1.4.2: Modifying a model

We saw previously, that if the birth rate is larger than the death rate, our population grows, which is a property we want, but can grow infinitely large, which is biologically infeasible. This is one of the first lessons we can learn about mechanistic modelling - it won't always work, and if it doesn't, we can add further assumptions. However, it is always best to start simple and add things than it is to start with a very comprehensive model and have to remove them. This is because complex models with lots of terms and parameters are difficult to fit to data, and can have identifiability issues.

Let's ammend our mathematical assumption 2. Instead of allowing births to be generated with a rate b , we will add in a density dependent term, which will slow the generation of new births as we hit a specified "limit". We will also want the population to decrease to that limit if it ever starts above it. Mathematically, this means that the birth term, say $B(N)$ needs to have the following properties for some "limit" K . You can think of this limit as being a cap on resources, which could be space or food.

- $B(N) > 0$ for population sizes between 0 and K ,
- $B(N) < 0$ for population sizes bigger than K , and
- $B(K) = 0$. This ensures that the no new births can occur at this resource cap.

It is worth noting that this limit K is a theoretical upper bound for the population size. Deaths can cause the population to drop below this. The simplest choice for this would be a linear function (i.e. a straight line), which passes through $B(K) = 0$. As an example:

$$B(N) = b \left(1 - \frac{N}{K}\right)$$

. Notice that this satisfies all of the properties above. This means our updated model, which replaces bN in [[REF]] to $B(N)N$, yields:

$$\frac{dN}{dt} = B(N)N - dN = bN \left(1 - \frac{N}{K}\right) - dN,$$

We can tidy this up by defining some new parameters: $a = b - d$ and $\kappa = K(b - d)/b$. Then we obtain the following differential equation:

$$\frac{dN}{dt} = aN \left(1 - \frac{N}{\kappa}\right) \tag{1}$$

Now the question we ask is what can we learn from this? Firstly, when N is small, the population will grow with a rate close to a , which is exponential growth. However, when N is close to κ , the derivative becomes 0, which means we are no longer changing in time.