Introducing Mathematical Biology

Introducing Mathematical Biology

AN OPEN EDUCATION RESOURCE

ALEX BEST

Contents

Introduction

All models are wrong but some are in this textbook.
(Based on a tweet by @alackles).

WHAT IS THIS?

In this *online, interactive and free* textbook, you will be guided through how to develop and analyse mathematical models to ask questions about a variety of problems from biology and medicine.

WHAT MATHEMATICS WILL WE USE?

In the general field, mathematical models for biological and medical systems can take a range of forms, including difference equations, ordinary or partial differential equations, stochastic models, individual-based computational models, not to mention more data-driven approaches and much more besides. The mathematical models we will cover here are all in the form of *ordinary differential equations* – some linear and some non-linear. For a few of the models you can also play with some *interactive simulation models* in Python (for which no prior experience of coding is needed).

WHAT APPLICATIONS WILL WE SEE?

In terms of applications, mathematical modelling plays an increasingly important role in almost any area of life sciences that you'd care to think of. Here we will focus on a few key areas:

- *Population ecology* – single and interacting species including competition and predator-prey systems.
- *Infectious diseases* – classic epidemic models, host-parasite systems and evolution.
- *Immunology and cell dynamics* – within-host disease interactions and simple cancer dynamics.
- *Gene networks* – regulatory feedback loops in both one and two gene systems.
- *Pharmacokinetics* – single and repeated doses of intravenous and orally adminstered drugs.

WHO IS THIS AIMED AT?

I hope that anyone who is interested in learning about how to model biological systems will be able to get something out of this resource. While some degree of mathematical knowledge is assumed, optional background review material is provided for those who need a bit more detail. I'd see the main audience for this book as:

- Undergraduate and postgraduate mathematics students who have not studied mathematical biology before.
- Undergraduate and postgraduate life-sciences students with an interest in modelling.
- Researchers or analysts in fields such as ecology, public health, immunology or pharmacology who are interested in modelling approaches.
- A-level/post-16 students studying mathematics who are keen to see some university-level material.

HOW DO I USE THIS BOOK?

I have written the textbook imagining that you would work through it in order from start to finish. Of course, if you are short on time and have a particular interest in a certain application you are very welcome to focus on just one (or two or more) sections of the material.

In terms of mathematical (and for that matter, biological) background I have aimed the core material assuming a reader has taken the first year or two of an undergraduate mathematics degree – and is therefore familiar with concepts such as what an ordinary differential equation (ODE) represents, how to solve first-order linear ODEs, as well as more general material such as geometric series, properties of exponentials, etc. If you are not familiar with these concepts you may need to look at some other materials to give you some background (for which I suggest browsing the Pressbooks Directory), but I hope that you can at least follow the thinking behind the methods. The additional background review chapters provided here are for those who have perhaps studied the first year of a mathematics degree but have not yet come across non-linear ordinary differential equations.

I have tried to make the textbook fairly interactive. In pretty much every chapter you will see blue boxes like below where you are strongly encouraged to have a go at some of the working yourselves.

Exercises

Boxes like this are used for problems and bits of working for you to have a go at on your own. When you've done that, you should click the button below.

Click for solution

After you've made your attempt, you can now reveal a worked solution.

Sometimes these are stand-alone problems, but they can also involve doing bits of working that are important for developing the material. I therefore strongly encourage you to have a go at the problem and check over the solution before continuing.

CODING

In addition, some boxes provide you with Python code that you can use to explore the models in more detail. No previous coding experience is required to run these – simply copy and paste the provided code into a Python program and run it. For those who are new to coding, it is an increasingly large part of mathematical modelling and, while it is not an essential part of this course, I hope you will take the chance to learn some further skills. There are many ways to install Python on your computer, but you can *download a full distribution for free* by visiting https://docs.anaconda.com/anaconda/install/ and selecting the appropriate download. This comes with two nice pieces of software for running Python code in – Spyder and Jupyter Notebooks, of which Spyder is perhaps marginally more straightforward for a new user.

REFERENCES

The content of a textbook like this is built up over many years of study and teaching. I have tried my best to reference the resources that went into each chapter's material in the *Chapter references* at the end of each page, with a full reference list at the end. These might also provide a starting point for further study of any areas that you have a particular interest in.

ACCESSIBILITY

There is a particular challenge to the mathematics community in producing mathematical notes that meet even minimum accessibility standards. Thanks to recent developments, content produced as HTML webpages using MathJax has solved many problems. This content has been tested for screenreaders using NVDA and MathPlayer in Firefox and the mathematical content could be read. All images have long description alt-text provided. If you discover any issues with accessibility please do contact me using the anonymous feedback form and I will see if I can fix it.

WHAT IS AN OPEN EDUCATION RESOURCE?

Traditional textbooks, both printed and e-books, can be prohibitively expensive for both individuals and even large organisations, continuing bias in who has access to teaching and learning materials. Open Education Resources are created and licensed for users to own, use and even modify. As such, **this online resource is free for anyone to retain, reuse, revise, remix and redistribute** (under the condition that any published use of it is cited).

Being online, the hope is also that you can use the book in a way that suits you, for example making it easier to jump between sections, allowing some extra background material and providing Python code.

FEEDBACK

To help me understand how this textbook is being used and perhaps inform future edits, please consider filling in this anonymous feedback form to provide some details.

PART I

POPULATION ECOLOGY

CHAPTER 1

Single population models

A FIRST POPULATION MODEL

Our aim in this textbook is to model the dynamics of populations over time. By 'population' we simply mean some collection of individuals that are subject to the same underlying mechanisms. For example, we may consider the human population of Sheffield, or the tapir population of South America, or the maize crop population of a field in Nigeria. In later chapters we will move to smaller biological scales, considering perhaps less intuitive definitions of populations, such as of cells in the body, or proteins within cells. However, a key message to take from this textbook is that we can consider pretty much any biological populations in the same way from a modelling perspective, and subject to the same fundamental biological mechanisms.

We will model these dynamics using *ordinary differential equations*, and our focus will be on how the size of a population varies over time. We will not consider where individuals are in space – this would likely require extending our methods to *partial differential equation* models, and plenty of excellent courses and textbooks can be found to explore such systems. When we talk about the 'size' of a population, we might think we mean the number of individuals. However, as we will be using ordinary differential equations we need our variables to be continuous, not discrete. We will therefore instead keep track of a population's *density* – the number of individuals within some unit area.

Introducing our mathematical notation, we might call $N(t)$ the density of individuals within our population at time t. We now wish to write down an ordinary differential equation that describes its dynamics – that is what causes the population to increase or decrease. Our first modelling challenge, then, is to decide what mechanisms we should include. What mechanisms *would* lead a population to change in size? Using some biological intuition for the simplest possible population we can think of, we might expect our model to look something like,

$$\frac{dN}{dt} = \text{births} - \text{deaths}.$$

Now to make any mathematical progress we will need to decide on functional forms for each of these mechanisms. This is another important modelling decision, and will depend on the specific biology. There are three main forms we might use:

- *constant rate*, $\text{births} = b$. This would be suitable when new individuals appear in the environment at some constant rate, for example due to migration, or production of cells by the body.
- *per-capita rate*, $\text{births} = bN$. This would be suitable when each individual produces offspring at a certain rate, such that there are more offspring produced when the population is larger.
- *density-dependent rate*, $\text{births} = b(N)N$. This would be suitable if the per-capita rate is not fixed but instead varies with the population size, for example if high density means more competition for resources and thus lower growth. The form of this function will again depend on the biology, though there are common functions we tend to use.

For our basic model, per-capita rates for birth and death seem the most obvious choice. Our model is thus,

$$\begin{aligned}\frac{dN}{dt} &= bN - dN \\ &= rN,\end{aligned}$$

where the parameter $r = b - d$ is the growth rate of our population. We therefore have a first-order, linear ordinary differential equation (ODE). Such ODEs are readily solved using either separation of variables or integrating factors. In this textbook I have made the assumption that readers are comfortable with such methods. If you are not, you might like to look at a textbook or online material for solving linear ordinary differential equations, but in the first few exercises I will also provide fairly detailed worked solutions. Have a go at the exercise below to solve this first population model.

Exercises

Either by *separation of variables* or *integrating factors* show that the solution to our model is,

$$N(t) = N(0)e^{rt},$$

where $N(0)$ is the density at $t = 0$.

Click for solution

I will use separation of variables to solve this equation. First we gather all the terms in our ODE with N to the left-hand side and everything else to the right-hand side, giving,

$$\frac{dN}{N} = rdt.$$

Then we integrate both sides to find,

$$\ln(N) = rt + C,$$

where C is a constant of integration. If we then take the exponential of both sides we get,

$$N(t) = e^{rt+C} = C_1 e^{rt},$$

where $C_1 = e^C$. To replace this constant we use the initial condition that at time $t = 0$ we have a density of $N(0)$. Substituting these values in we find that $C_1 = N(0)$, leaving us with the final solution,

$$N(t) = N(0)e^{rt},$$

as required.

This predicts exponential growth of the population for $r > 0$ and exponential decline for $r < 0$.

THE LOGISTIC EQUATION

We would usually assume that births are greater than deaths, meaning $r > 0$ and our population will continue growing to infinitely large densities. This is, we can hopefully agree, a tad unrealistic. We should not lose heart at this point, though. A model that produces unrealistic results is still helpful to us, because it will point the way towards how we can develop an improved version (remember the adage, *all models are wrong but some are useful*).

What might we have failed to take into account in our model? One answer is that as the population grows the environment will become limiting in some way. For example, individuals need space and nutrients to live. These resources will be consumed by the population more and more as the population grows, likely leading to reduced birth or increased death rates. So while our assumptions may hold at low population densities, they will break down at higher densities. Our simple model has shown us that we need to take this into consideration if we are to understand the population dynamics.

Given this, choosing a density-dependent growth rate starts to look like a better assumption. A relatively simple and intuitive way to set a limit on population growth is to assume that there is a fixed *carrying capacity* for the population, for example due to the available space or nutrients. We assume that as the population approaches this carrying capacity, the overall growth rate approaches zero, and once the population goes above this value the growth rate becomes negative. The simplest implementation of this is to assume that r decreases linearly with N:

$$r(N) = r_0 \left(1 - \frac{N}{K}\right),$$

where r_0 (the *basic reproductive rate*) and K (the carrying capacity) are positive constants. Our ordinary differential equation for the dynamics of our population is then

$$\frac{dN}{dt} = r_0 N \left(1 - \frac{N}{K}\right).$$

This equation is known as the *logistic growth equation* and is an example of *intra-specific competition*.

A USEFUL STEP: NON-DIMENSIONALISATION

We can continue our analysis with the model in this form again using methods for solving ordinary differential equations. However, before continuing with our analysis we shall introduce the useful technique of *non-dimensionalising* our model. This is often – though by no means always – done to models to simplify some of the later working. It can both reduce the number of parameters in the model and give us insight in to the scales at which biological processes operate.

Non-dimensionalisation involves combining variables and/or parameters from the model in biologically and mathematically helpful ways. For example, a natural scale for the population level is K, so we can define the non-dimensional population variable $n = N/K$ (n is non-dimensional because both N and K have units 1/[density], so n is a dimensionless number). The equation then becomes

$$\frac{dn}{dt} = \frac{d(N/K)}{dt} = \frac{1}{K}\frac{dN}{dt} = r_0 n(1-n).$$

We can also note that a natural time scale for the dynamics is $1/r_0$, such that we can take a new non-dimensional time variable, $\tau = r_0 t$, and the final non-dimensional model equation is

$$\frac{dn}{d\tau} = \frac{dn}{d(r_0 t)} = \frac{1}{r_0}\frac{dn}{dt} = n(1-n).$$

As you can see, by using two substitutions we have arrived at a very simple equation.

MODEL ANALYSIS

Even though this equation is non-linear (we have an n^2 term), we can still find an explicit solution for it. By separation of variables, we can write this as,

$$\int \frac{dn}{n(1-n)} = \int r_0 \, d\tau,$$

We cannot immediately integrate the left-hand side of this. However, we can use *partial fractions* to rewrite

the fraction on the left-hand side as the sum of two terms (it turns out to be $1/n + 1/(1-n)$). We can now integrate these to reach an initial solution of,

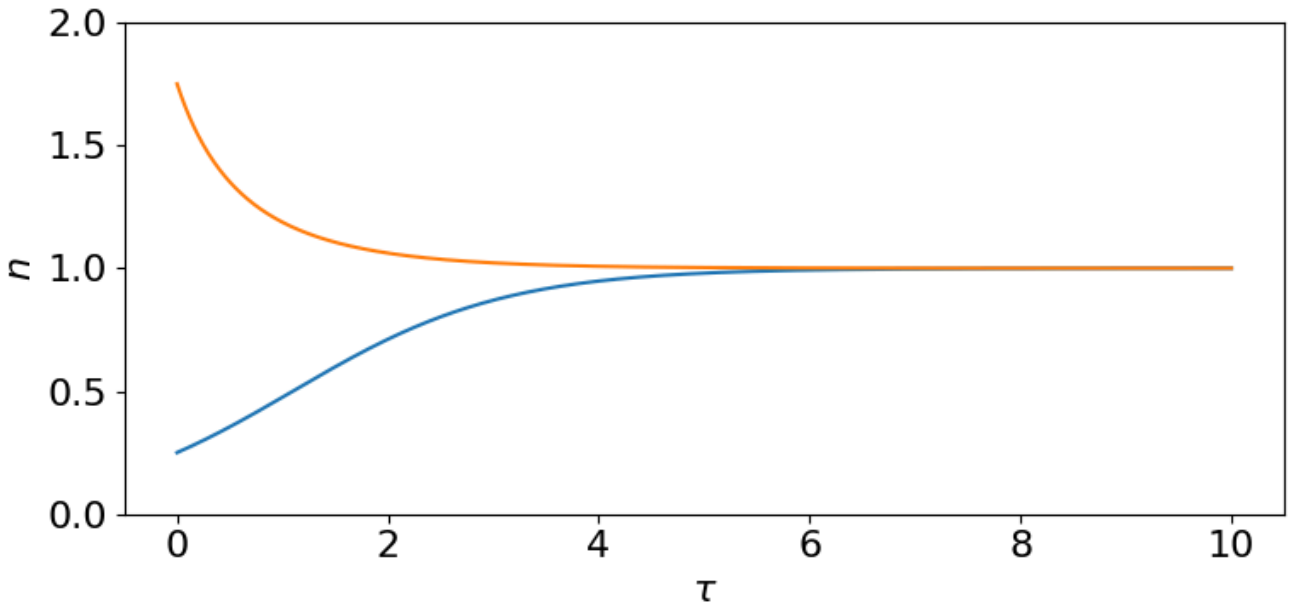
$$\ln(n) - \ln(1-n) = r_0\tau + C,$$
$$\implies \frac{n}{1-n} = e^C e^{r_0\tau},$$

where C is a constant of integration. Letting the initial condition be $n(0) = n_0$ and rearranging, we arrive at the final solution,

$$n(\tau) = \frac{n_0}{n_0 + (1-n_0)e^{-\tau}}.$$

(Alternatively a suitable substitution can be chosen to evaluate the integrals. It is often true in mathematical modelling that there are multiple ways to arrive at a solution, and embracing this diversity of approach can be very useful).

Examples of the (non-dimensional) solution for two different initial densities are shown in the figure.



Two time-courses of the non-dimensional logistic model for different initial densities.

While this has been useful to get an explicit solution, we can in fact gain a good deal of *qualitative* information about this model directly from the ODE without having to spend the few minutes involved in deriving that explicit solution. Specifically, we can see:

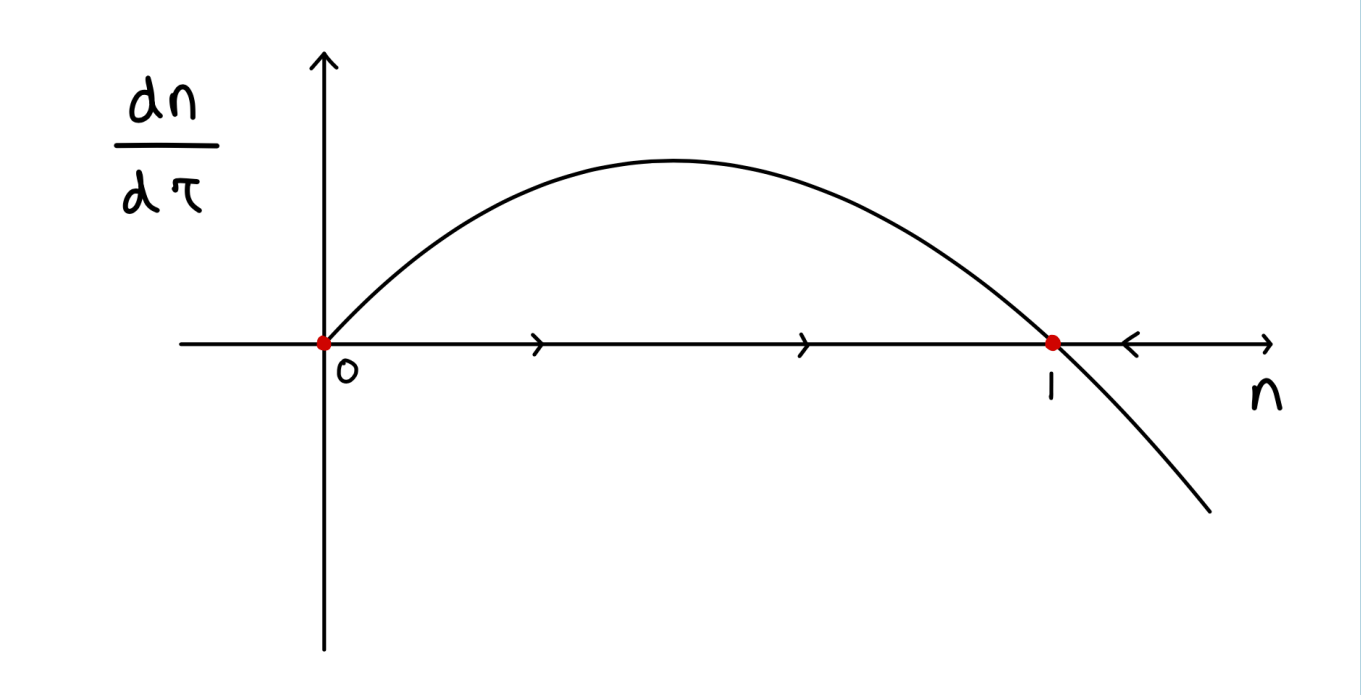
- for $0 < n < 1$, $dn/d\tau > 0$,
- for $n > 1$, $dn/d\tau < 0$,

(note that it makes no biological sense to worry about $n < 0$). There are two values of n for which $dn/d\tau = 0$, at $n = 0$ and $n = 1$, which are the *equilibria* of the system. Therefore for any starting value of $n > 0$ the population should move towards a final state of $n = 1$, suggesting that the equilibrium at $n = 0$ is *unstable* and the equilibrium at $n = 1$ is *stable*. We can think about this more by sketching the curve $dn/d\tau$ vs n. If you are unfamiliar with *linear stability analysis* you should have a read of the relevant background review chapter.

Exercises

Sketch the curve $dn/d\tau$ vs n and use it to identify the key aspects of the system.

Click for solution



A phase diagram of the logistic model, showing an unstable equilibrium at $n = 0$ and a stable equilibrium at $N = 1$.

On this sketch we can mark our equilibria at $n = 0$ and $n = 1$ and then consider the wider behaviour by considering the curve $dn/d\tau$ as a function of n. The highest power here is n^2 so the curve will look like a quadratic, and since the n^2 term is negative it will initially increase, then peak, and then decrease.

For convenience let us call $dn/d\tau = f(n)$, and then look at what happens nearby the equilibria by *linearisation*.

- At $n = 0$, the gradient of the curve df/dn is positive, meaning $dn/d\tau > 0$ for $n > 0$ and $dn/d\tau < 0$ for $n < 0$ (we don't really mind about the actual values in this latter case since $n < 0$ but it helps for picturing the gradient). As such if we start with a population density a small distance away from $n = 0$ we will always end up moving further away from it – it is unstable.
- At $n = 1$, we have df/dn negative, meaning $dn/d\tau > 0$ for $n < 1$ and $dn/d\tau < 0$ for $n > 1$, As such if we start with a population density a small distance away from $n = 1$ we will always move in towards this equilibrium – it is stable.

We can also mark on the qualitative direction of trajectories on the x-axis, giving a *phase line*. Notice, then, that we have fully described the possible behaviours of this system without having to explicitly solve the equation.

CASE STUDY: SPRUCE BUDWORM

Spruce budworm is an insect that lives in spruce and fir forests in the USA and Canada (see chapter references). While they mostly exist at low numbers, there are periodic explosions of budworm populations that devastate the forests. Let us assume that on their own, the budworm populations follow the logistic growth model we have just studied. An additional feature of budworm dynamics is that they are predated by birds. We might then describe the dynamics of budworm by the following equation,

$$\frac{dN}{dt} = r_0 N \left(1 - \frac{N}{K}\right) - P(N),$$

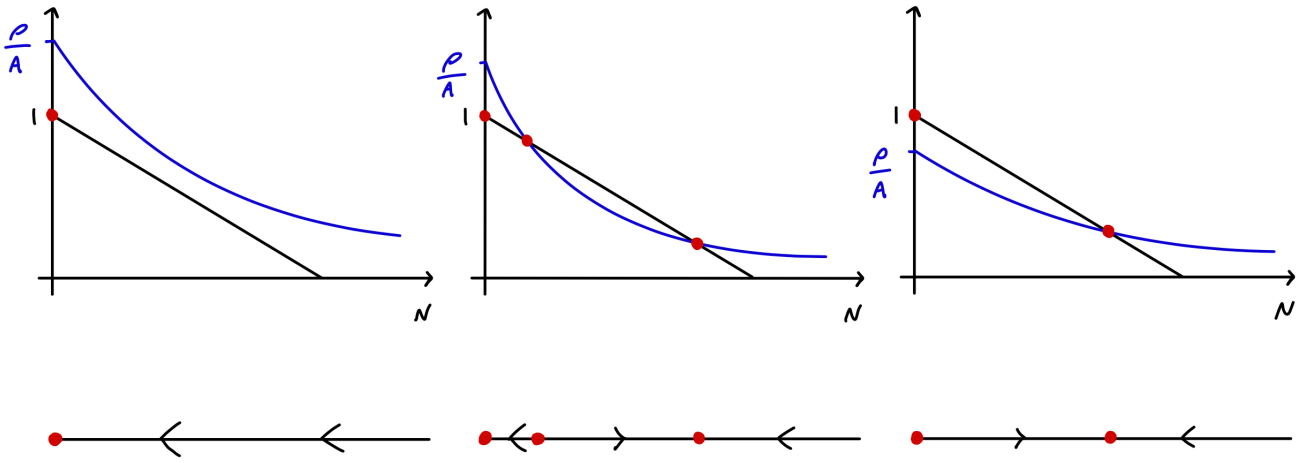
where $P(N)$ describes the rate of predation (note that we do not explicitly consider the bird population dynamics). What should $P(N)$ look like? Should it be constant, a fixed per-capita rate, or some other density-dependent function? First, we might reasonably assume that as more food becomes available, the more the birds will eat, so predation should increase with budworm density. However, birds cannot eat an infinite amount, and so the rate of predation should therefore saturate at high budworm densities. We therefore need a density-dependent function. We might then give an explicit model as,

$$\frac{dN}{dt} = r_0 N \left(1 - \frac{N}{K}\right) - \frac{\rho N}{N + A}$$

This form for the predation function is a very commonly used one in mathematical biology models (in this context it is often known as a *Holling type II functional response*). What are the possible outcomes of this model? We might start by finding equilibria as before, but we will quickly find ourselves bogged down. Instead, let's take a qualitative approach and use curve-sketching to explore the possible outcomes. The above model can be written as,

$$\frac{dN}{dt} = N\left[r(N) - p(N)\right],$$

where $r(N) = r_0(1 - N/K)$ and $p(N) = \rho/(N + A)$. We know that an equilibrium occurs where $r(N) - p(N) = 0$, which will be when the two curves cross. Below we sketch these two curves, $r(N)$ and $p(N)$, for the three possible orientations, separated by how high the predation parameter ρ is.



Three phase diagrams for the spruce budworm model, plotting $p(N)$ (blue) and $r(N)$ (black).

It is worth stressing that $N = 0$ is always an equilibrium even though the two curves may not cross here – if you look back at the ODE we factored N out of the equation. What happens in each case?

1. Predation by the birds is very high, the two curves never cross, and $dN/dt < 0$ for all $N > 0$.

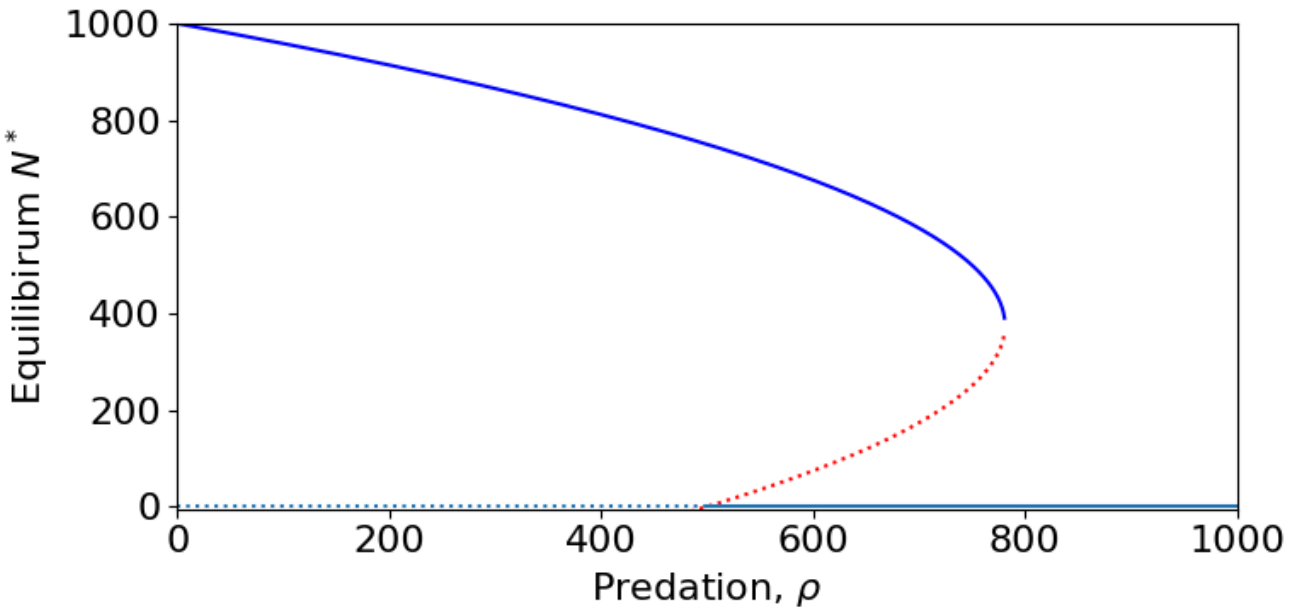
Therefore the insect population is always reducing until it reaches extinction, with $N = 0$ a stable equilibrium.

2. Predation has weakened somewhat and two new equilibria appear, one stable and one unstable. We now have bistability – where two different equilibria are stable, but which one we go to depends on the initial conditions.
3. Predation is very weak and we now only have one equilibrium for $N > 0$, at high budworm densities. We also see now that $N = 0$ is no longer stable, and the budworm population will always reach this high equilibrium density.

What do these results tell us? That pest outbreaks such as spruce budworm are likely to be driven by the pressure from predation. We can also see that, due to the bistability, sudden outbreaks can occur if there are relatively small changes to the environment. We have achieved this insight without having to do any detailed mathematical derivations, but by applying mathematical reasoning and biological interpretation – these are two key skills we will develop over this course.

BIFURCATION DIAGRAM

This model reveals that as we alter a parameter value, the behaviour of the model qualitatively shifts, either through switching the stability of two already existing equilibria or through the creation of completely new equilibria. Such transitions are called *bifurcations*, which you can read further on in the relevant background chapter if needed. A useful approach is to draw a bifurcation diagram, which effectively puts all of the information on our diagrams above into one single plot, allowing us to see at a glance what the possible behaviours and transitions are. The bifurcation diagram for this model is shown below.



Bifurcation diagram for spruce budworm model. Solid lines denote stable equilibria and dashed lines unstable equilibria. Parameter values: $r_0 = 2, K = 1000, A = 250$.

This shows that if ρ is small, we have two equilibria – one at $N^* = 0$ that is unstable and one for high N^* that is stable. Therefore throughout this region, the population will be attracted to the equilibrium with a large population (though its value shrinks as predation increases). At $\rho = 100$ a *transcritical bifurcation* occurs where a further equilibrium (that was at negative values and so hadn't been drawn) collides with $N^* = 0$ – causing itself to be unstable and the zero equilibrium to become stable – and then increases. In this second

region we therefore have bistability between extinction and the higher (but again, decreasing) equilibrium. At around $\rho = 300$ the two non-zero equilibria collide and disappear in a *saddle-node bifurcation*. This leaves the final region where only the extinction equilibrium exists and is stable.

Explore the model

Click the button below to reveal Python code that you can use to explore the behaviour of this model in more detail. Paste this into your chosen Python software and run the model. This should produce two plots – a time course for two initial N values and a plot of how the rates vary with N – that match case (1) above.

On line 12 of the code you should see "`rho=20`". Using the workings above as a guide, change the value of ρ to further visualise what the dynamics look like in the three different cases. Can you choose a value of ρ that clearly shows bistability in the time-courses?

If you feel confident, feel free to adjust other parameter values and initial conditions to explore the behaviour.

Click for code

```
 # Import the necessary libraries
import numpy as np
import matplotlib.pyplot as plt
from scipy.integrate import solve_ivp
 # Function for population dynamics called 'budworm'
def budworm(N,t):
    dN = r0*N*(1-N/K)-rho*N/(N+A)
    return dN
 #Parameter values
rho=20
r0=5
K=10
A=1
 # Initial conditions
N01=[8]
N02=[2]
 # Time points to use
 tc = np.linspace(0, 10, 1000)
 # Run the model using 'solve_ivp' with the 2 different ICs
Nc1 = solve_ivp(budworm, [tc[0],tc[-1]], N01, t_eval=tc)
Nc2 = solve_ivp(budworm, [tc[0],tc[-1]], N02, t_eval=tc)
 # Nullclines
nn=np.linspace(0,12,100)
```

```
rnull=r0*(1-nn/K)
pnull=rho/(nn+A)
 # Plotting commands
plt.rcParams['figure.figsize'] = [12, 4]
plt.rcParams.update({'font.size': 16})
fig, (ax1, ax2) = plt.subplots(1, 2)
ax1.plot(tc, Nc1.y[0], "r")
ax1.plot(tc, Nc2.y[0], "r:")
ax1.set(xlabel='Time', ylabel='Densities')
ax1.axis([0,10,0,10])
ax2.plot(nn,rnull,'r',label='r(N)')
ax2.plot(nn,pnull,'k',label='p(N)')
ax2.set(xlabel='$N$', ylabel='Rates')
ax2.legend()
ax2.axis([0, 12, 0, 10])
```

Key Takeaways

- Parameters in a model represent rates of change, and their functional form implies how that process works.
- The logistic equation goes to a long-term equilibrium, and we do not need to explicitly solve it to understand its key behaviours.
- Sudden outbreaks of spruce budworm can occur if there is a small change in the level of predation.

Chapter references

- The spruce budworm model is based on an example from the textbook, *Mathematical Biology 1* by Murray.
- The content in the *Population ecology* section was influenced by the textbook, *Mathematical Biology 1* by Murray.

CHAPTER 2

Interacting populations 1: competition

CASE STUDY: RED AND GREY SQUIRRELS

Populations rarely (if ever) exist in isolation. In reality, the growth rate of a given population depends not only on itself, but also on other populations that it interacts with either directly or indirectly. Such interactions lead to a range of ecological relationships, including competition for resources, predation, mutualism, parasitism and more besides. In the next chapters we will study models that represent a few of these different interactions.

Our first example will be of two species that compete for a common resource. This is sometimes known as the *Lotka-Volterra model* (though I personally try to avoid using labels named after people – they can be confusing and ingrain bias) or *interspecific competition*. To guide our thinking we will have a case study in mind, namely red and grey squirrels – well-known competitors in the UK – but we could equally consider microbes competing for substrates, plants competing for water or nutrients or a whole host of species competing with one another for territory or food. As with our previous single population models, we begin by considering what factors might contribute to the birth and death mechanisms that control the rates of changes of the two populations. We will assume:

- There is only a limited supply of food for squirrels to eat. This means that even in the absence of interspecific competition, the environment would have a limited carrying capacity (i.e. a maximum number of squirrels that it could support). The population dynamics for each species on its own can be modelled using a logistic growth equation (that does include *intraspecific competition* between individuals in the same population).
- The effect of one species on the other is proportional to its population size (since the more there are, the less food, territory and other resources will be available). Therefore there will be a term in each equation causing a reduction in growth that is given by the product of the two species' densities.

We also assume that the carrying capacity for the two species of squirrel can be different (perhaps because one species can eat a wider variety of food than the other), and that the two species can have different susceptibilities to competition (perhaps because one species is more timid than the other). Without interspecific competition, we could model the two populations using two logistic growth equations:

$$\frac{dR}{dt} = R(a - R)$$
$$\frac{dG}{dt} = G(c - G),$$

where R and G represent the respective population densities of red and grey squirrels. Note that we have already done a little non-dimensionalising here, and that a and c are thus the respective carrying capacities (since if, for example, $R > a$, we have $dR/dt < 0$). To include competition between the species, we add an additional term to each equation:

$$\frac{dR}{dt} = R(a - R - bG)$$
$$\frac{dG}{dt} = G(c - dR - G),$$

where b and d are the strengths of interspecific competition. While this model is only a little more complicated than the logistic model, this system cannot be solved explicitly since it is *non-linear*. We will therefore need to take a qualitative approach to understanding what can possibly happen in this model. However, we can still gain considerable insight in to the system by applying the tools of *dynamical systems*. See the appropriate background review chapters on phase portraits and linear stability analysis if needed.

ANALYSIS

PHASE PORTRAITS

A good way to visualise what is happening here is through a *phase portrait*. As discussed in more detail in the background review chapter, our phase portrait can be built up using the following procedure:

1. Draw axes of the two key variables.
2. Determine how much of the phase plane is biologically feasible.
3. Calculate nullclines and draw them on your plot.
4. Mark on equilibria where two (different) nullclines cross.
5. Work out the qualitative directions of travel in each of the regions separated by the nullclines and draw arrows on your phase portrait to show these direction fields.
6. Optionally use linear stability analysis to get a clearer picture of behaviour around the equilibria.
7. Sketch on some sample trajectories.

In this example our two axes will be our population densities, R and G, and we know both densities must be non-negative. Now let us find the nullclines:

$$\frac{dR}{dt} = 0 \implies R = 0 \quad \text{or} \quad G = \frac{a - R}{b}$$
$$\frac{dG}{dt} = 0 \implies G = 0 \quad \text{or} \quad G = c - dR$$

We therefore have a nullcline along each axis, and two nullclines which are decreasing straight lines on our plot. Qualitatively, there are four different ways we can configure these two nullclines:

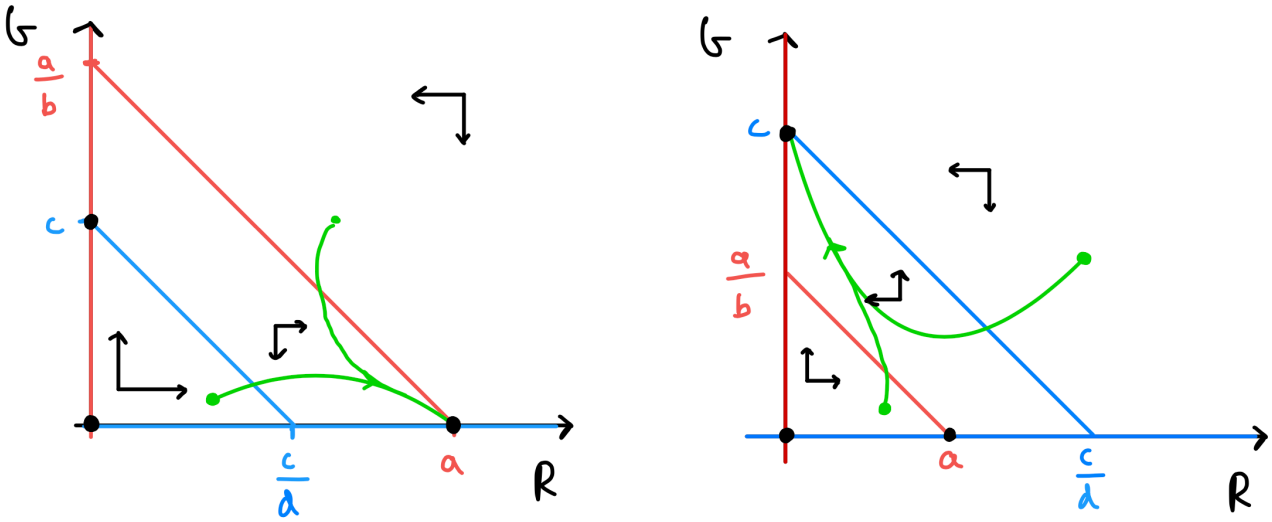
1. Red starts above Grey and they do not cross (in the biologically feasible region).
2. Grey starts above Red and they do not cross (in the biologically feasible region).
3. Grey starts above Red and they do cross.
4. Red starts above Grey and they do cross.

When drawing the direction fields we can say that,

- If R and G are both small (we are below both nullclines), $dR/dt > 0$ and $dG/dt > 0$.
- If R and G are both large (we are above both nullclines), $dR/dt < 0$ and $dG/dt < 0$.
- If we are above the R-nullcline but below the G-nullcline, $dR/dt < 0$ but $dG/dt > 0$.

- If we are below the R-nullcline but above the G-nullcline, $dR/dt > 0$ but $dG/dt < 0$.

This gives us enough information to sketch out the four phase portraits and assess the qualitative behaviour. The first two cases that we listed above are sketched out below. Work your way through the procedure above and see how these have been built up.



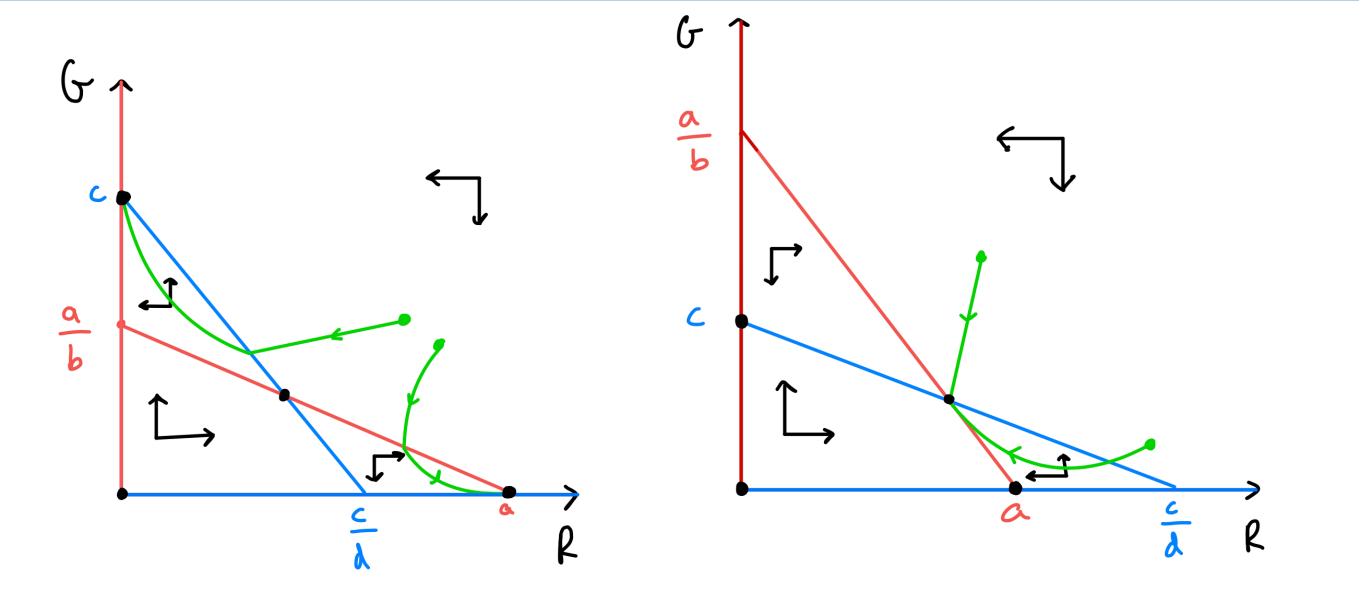
Two phase portraits for the competition model. Red lines represent nullclines for red squirrels, blue lines nullclines for grey squirrels, and orange curves are example trajectories. In the first case trajectories all move to the red-only equilibrium, and in the second case they move to the grey-only equilibrium.

Exercises

Sketch out the two remaining qualitatively different phase portraits.

Click for solution

The remaining two phase portraits should look like this:



The two remaining phase portraits for the competition model. Red lines represent nullclines for red squirrels, blue lines nullclines for grey squirrels, and orange curves are example trajectories. In the third case trajectories can go to either the red-only or grey-only equilibrium depending on the initial condition, and in the fourth case trajectories all move to the coexistence equilibrium.

Let's consider what these phase portraits show. Firstly, what are the possible equilibria? We see that equilibria at $(0,0)$, $(a,0)$ and $(0,c)$ are always present (an R-nullcine and G-nullcline always cross at these values). These respectively mean no squirrels present, only red squirrels present and only grey squirrels present. Then we *sometimes* have another equilibrium – if $a > bc$ and $c > ad$ or $a < bc$ and $c < ad$ – where both species are present.

Which of these equilibria would we end up going to in each case? Looking through our four phase portraits in turn we find,

1. R "wins" and settles to a steady population level, while G goes extinct.
2. G "wins" and settles to a steady population level, while R goes extinct.
3. The outcome depends on the starting condition, with either R or G eventually winning out and sending the other population to extinction.
4. R and G co-exist at steady non-zero population levels.

We have therefore sketched out all the possible outcomes in a competition model without finding explicit solutions for the two populations (remember, such solutions do not even exist). We can see that in theory two species can coexist, but it requires a particular balance of competition and carrying capacities. If this isn't met we will see *competitive exclusion* where one of the species is ultimately driven to extinciton by the other. It is this latter case that appears to be occurring with the two squirrel species in the UK – with Red squirrels rapidly declining while Grey squirrels spread throughout the UK (as an aside, it is increasingly appreciated that an infectious disease – squirrelpox virus – is playing a significant role in this replacement).

LINEAR STABILITY ANALYSIS

Beyond this graphical method it might be useful if we can formally characterise these equilibria, and in particular assess our conclusions from the phase portraits about when each equilibrium will be an attracting end-point of the dynamics (*stable*) and when not (*unstable*). Much like we did with the logistic model, we can assess the local stability of steady states of a system by linearising around the equilibrium, but now we are working in two dimensions we need to do this by finding the eigenvalues of the Jacobian matrix, as discussed in the background review chapter.

To briefly summarise the methods, let us suppose that $dR/dt = f(R, G)$ and $dG/dt = g(R, G)$. Then the Jacobian matrix is made up of the partial derivatives of f and g,

$$J = \begin{pmatrix} \frac{\partial f}{\partial R} & \frac{\partial f}{\partial G} \\ \frac{\partial g}{\partial R} & \frac{\partial g}{\partial G} \end{pmatrix},$$

taken at the equilibrium values of R^* and G^* (I often use a * to denote an equilibrium – it is not essential but I find it helps remind me what I am doing). We can then use this Jacobian to find the eigenvalues at an equilibrium, which will tell us whether that point is stable or unstable. In particular, if,

- $\lambda_1 < 0$ and $\lambda_2 < 0$, the equilibrium is stable;
- $\lambda_1 > 0$ and $\lambda_2 > 0$, the equilibrium is unstable;
- $\lambda_1 < 0$ and $\lambda_2 > 0$, the equilibrium is a saddle.

Alternatively, in this 2-dimensional system, we can use the *trace* and *determinant* of our Jacobian to determine stability:

- $tr(J) = \frac{\partial f}{\partial R} + \frac{\partial g}{\partial G}$,
- $\det(J) = \left[\frac{\partial f}{\partial R}\right] . \left[\frac{\partial g}{\partial G}\right] - \left[\frac{\partial f}{\partial G}\right] . \left[\frac{\partial g}{\partial R}\right]$.

The signs of these two quantities combine to tell us about the stability of the equilibrium, which is best seen through the following diagram:

diagram of stability condition based on trace and determinant

Figure: Stability of an equilibrium using the trace and determinant of the Jacobian.

It is also useful to note that $tr(J) = \lambda_1 + \lambda_2$ and $\det(J) = \lambda_1 \lambda_2$. Also, note that for a 2×2 Jacobian, if one of the off-diagonal entries is zero, the eigenvalues are simply the two entries in the main-diagonal.

Let's write down the generic Jacobian for our system,

$$J = \begin{pmatrix} a - 2R^* - bG^* & -bR^* \\ -dG^* & c - dR^* - 2G^* \end{pmatrix},$$

We can use the information from the Jacobians to derive the eigenvalues at each equilibria explicitly. In particular we can use the signs of the trace and determinant to establish the stability of each equilibrium.

The extinction equilibrium

At $(0, 0)$, the Jacobian is just

$$J = \begin{pmatrix} a & 0 \\ 0 & c \end{pmatrix},$$

which has $tr = a + c > 0$, $\det = ac > 0$ and $tr^2 - 4\det = (a-c)^2 > 0$, where tr and $\det$ are the trace and determinant of the matrix. Thus $(0,0)$ is always an unstable node.

The two single-species equilibria

At $(a,0)$, we have

$$J = \begin{pmatrix} -a & -ab \\ 0 & c - ad \end{pmatrix},$$

which has $tr = -a + (c - ad)$, $\det = -a(c - ad)$ and $tr^2 - 4\det = [a + (c - ad)]^2 > 0$. In fact, with the 0 in the off-diagonal we can just read the eigenvalues off as $\lambda_1 = -a$ and $\lambda_2 = c - ad$. Looking at our phase portraits, in our 1st and 3rd cases we see $c - ad < 0$ (look at the horizontal axis), meaning both eigenvalues are negative (and real) so we have a stable node. In our 2nd and 4th cases our phase portraits reveal $c - ad > 0$, meaning we have one positive and one negative eigenvalue, so we have a saddle .

Using a similar argument, we can show that $(0,c)$ is a stable node in our 2nd and 3rd cases, and a saddle point in the 1st and 4th case.

The coexistence equilibrium

In the cases so far we have substituted the equilibrium values for R^* and G^* into the Jacobian. However, we have not yet actually written down what these are for the coexistence equilibrium. Finding these expressions leads to some rather annoying algebra, but we can actually assess its stability without having to find these expressions. Instead, we can note that this equilibrium lies at the intersection of the two 'off-axis' nullclines. That means we know that at this equilibrium we have $a - R^* - bG^* = 0$ and $c - dR^* - G^* = 0$. Substituting these into $\mathbf{J}$, we get

$$J = \begin{pmatrix} -R^* & -bR^* \\ -dG^* & -G^* \end{pmatrix},$$

which has $tr = -(R^* + G^*) < 0$, $\det = (1 - bd)R^*G^*$ and $tr^2 - 4\det = (R^* - G^*)^2 + 4bdR^*G^* > 0$. Everything hinges on the sign of $\det = (1 - bd)R^*G^*$. Remember it is only our 3rd and 4th cases we need to look at here.

In our 3rd case, if we look at our phase portraits we see we have $a - bc < 0$ and $c - ad < 0$. Putting these together we find we have $a < bad$ and so $1 < bd$. Therefore $\det < 0$ and the coexistence equilibrium is a saddle.

In contrast, in our fourth case we have $a > bc$ and $c > ad$, meaning $a > bad$ and so $1 > bd$. Therefore $\det > 0$ and the coexistence equilibrium is a stable node.

Notice that these conditions again agree with what we found from our phase portraits. We can also go a step further and say that coexistence requires the product of the two competition terms to be small.

Explore the model

Click the button below to reveal Python code that you can use to explore the behaviour of the competition model in more detail. This will produce two plots – a time course for two sets of initial conditions and a phase portrait.

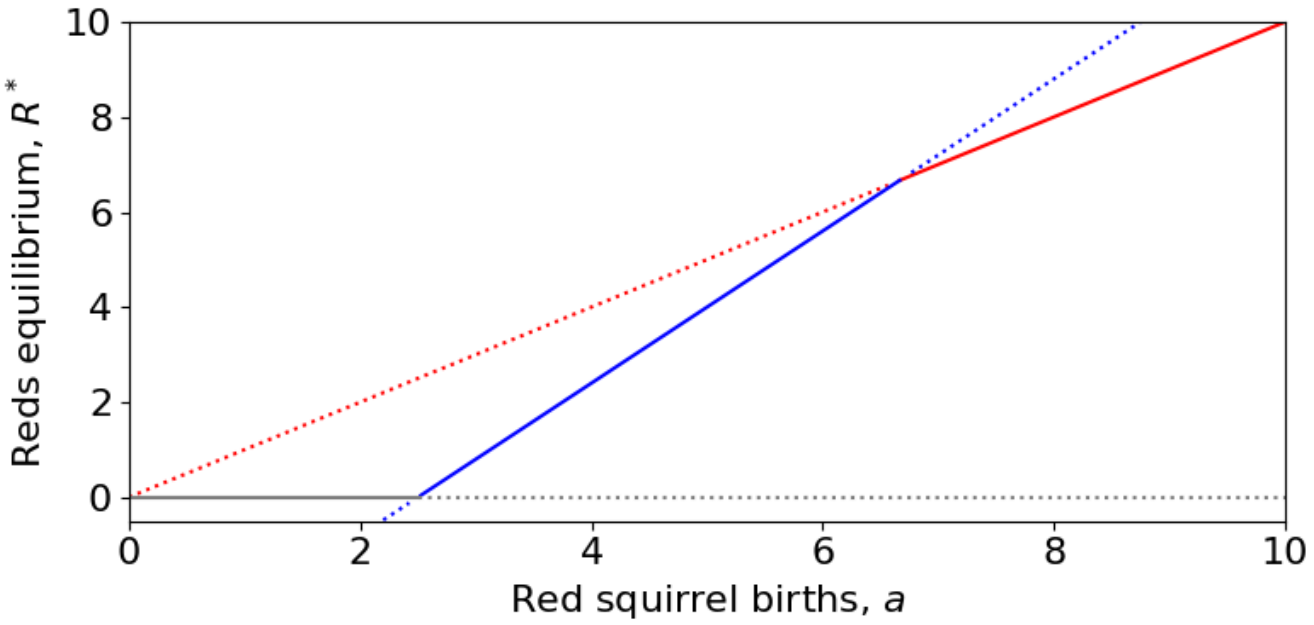
Try varying some of the parameter values to find the four different qualitative outcomes.

Click for code

```
# Import the necessary libraries
import numpy as np
import matplotlib.pyplot as plt
from scipy.integrate import solve_ivp
# Options to make the plots the right size
plt.rcParams['figure.figsize'] = [12, 4]
plt.rcParams.update({'font.size': 16})
# Function for the dynamics called 'competition'
def competition(t,N):
    # Rename the variables for ease
    R=N[0]
    G=N[1]
    # The ODEs
    dR = R*(a - R - b*G)
    dG = G*(c - G - d*R)
    return [dR,dG]
# Parameter values
a=3
b=0.5
c=3
d=0.75
# Initial conditions
R0_1=3
G0_1=3
R0_2=0.5
G0_2=0.5
N0=[R0_1,G0_1]
N1=[R0_2,G0_2]
# Time points to use
tc = np.linspace(0, 10, 1000)
# Run the model using 'solve_ivp'
Nc = solve_ivp(competition, [tc[0],tc[-1]], N0, t_eval=tc)
Nd = solve_ivp(competition, [tc[0],tc[-1]], N1, t_eval=tc)
```

```
# Plotting commands
fig, (ax1, ax2) = plt.subplots(1, 2)
ax1.plot(tc, Nc.y[0], "r", label="Reds")
ax1.plot(tc, Nc.y[1], "k", label="Greys")
ax1.plot(tc, Nd.y[0], "r:")
ax1.plot(tc, Nd.y[1], "k:")
ax1.set(xlabel='Time', ylabel='Densities')
ax1.legend()
ax1.axis([0,10,0,5])
rr=np.linspace(0,10,5)
rnull=(a-rr)/b
gnull=c-d*rr
ax2.plot(Nc.y[0],Nc.y[1],'b')
ax2.plot(Nd.y[0],Nd.y[1],'b')
ax2.plot(rr,rnull,'r')
ax2.plot(rr,gnull,'k')
ax2.axis([0, 5, 0, 5])
ax2.set(xlabel='Red Squirrels', ylabel='Grey Squirrels')
```

A BIFURCATION DIAGRAM



Bifrucation diagram for the competition model. The grey line marks the Grey-only equilibrium, the blue line the coexistence equilibrium and the red line the Red-only equilibrium. Solid lines denote a stable equilibrium and dashed lines an unstable equilibrium. Parameter values, $b = 0.5, c = 5, d = 0.75$.

We can also plot a bifurcation diagram of the system. This is slightly complicated since we now have two variables and multiple parameters, but we can just focus on how one parameter impacts one variable. Here we see three different transcritical bifurcations at different points as we vary the red squirrel carrying capacity, a:

- When a is low, the Grey-only equilibrium (grey line with $R = 0$) is stable, the Red-only equilibrium (red line) is unstable and the coexistence equilibrium (blue) is negative (and unstable).
- When a is intermediate, the coexistence equilibrium becomes positive and stable and the Grey-only equilibrium becomes unstable through a transcritical bifurcation. The Red-only equilibrium remains unstable.
- When a is high, the coexistence equilibrium becomes unstable (and actually negative in G, though this plot does not show that) and the Red-only equilibrium becomes stable through a transcritical bifurcation. The Grey-only equilibrium remains unstable.

(There is a third transcritical bifurcation at 0 where the Red-only and Grey-only equilibria collide, but this makes no real difference to the dynamics since we assume $a > 0$).

Key Takeaways

- Linear stability analysis allows us to classify the behaviour of a system near an equilibrium.
- Competing species can coexist with each other in the long term, provided the effects of interspecific competition are quite weak.
- Very often there will be *competitive exclusion*, where only one species can survive in the long term, whenever the effects of interspecific competition is quite strong.

Chapter references

- The content in the *Population ecology* was influenced by the textbook, *Mathematical Biology 1* by Murray.

CHAPTER 3

Interacting populations 2: predator-prey

A SIMPLE PREDATOR-PREY MODEL

In the previous chapter we studied a first example of interacting populations with two species who compete with one another largely indirectly. Another common interaction that is much more direct is predation. Many species rely on others as a primary food source: foxes and rabbits; lions and zebras; ladybirds and aphids to name just a few. Such predator-prey interactions will have a similar structure to the competition models we considered last time, except that now one species (the predator) benefits from the interaction.

We will not start with the logistic model this time, but instead think of the simplest possible way we can model the dynamics of a prey with density N and predator with density P,

$$\frac{dN}{dt} = N(a - bP)$$
$$\frac{dP}{dt} = P(-c + dN).$$

The prey has a birth rate a, no intraspecific competition and its death rate is only due to predation. This death from predation depends on the predator density (with parameter b) with more predators leading to higher prey death and no saturation of predation at higher densities. In contrast, the predator's birth rate depends on its ability to predate, and so depends on the prey density (with parameter d). It has a simple death rate however at per-capita rate, c. We might expect the parameters b and d to be linked, since one controls the amount of predation and the other the ability to reproduce which itself depends on predation. In some models you will see the predator birth term written as something like θbPN, where θ is the 'conversion' of energy from feeding into new births.

We have already identified a few simplifications we have made here even compared to our first few models – no intraspecific competition and continually increasing predation. A good approach to modelling is always to start with the simplest reasonable model and then build increasing complexity as needed. This is also the 'classic' predator-prey model you will see elsewhere, so it seems sensible to cover it here.

Let's begin our analysis by finding the equilibria of the model. There is an extinction equilibrium at $(N, P) = (0, 0)$, and a coexistence equilibrium at $(N, P) = (c/d, a/b)$. The stability of these steady states is calculated, as before, by finding the Jacobian of the system.

Exercises

Write out the Jacobian matrix for the predator-prey system.

Click for solution

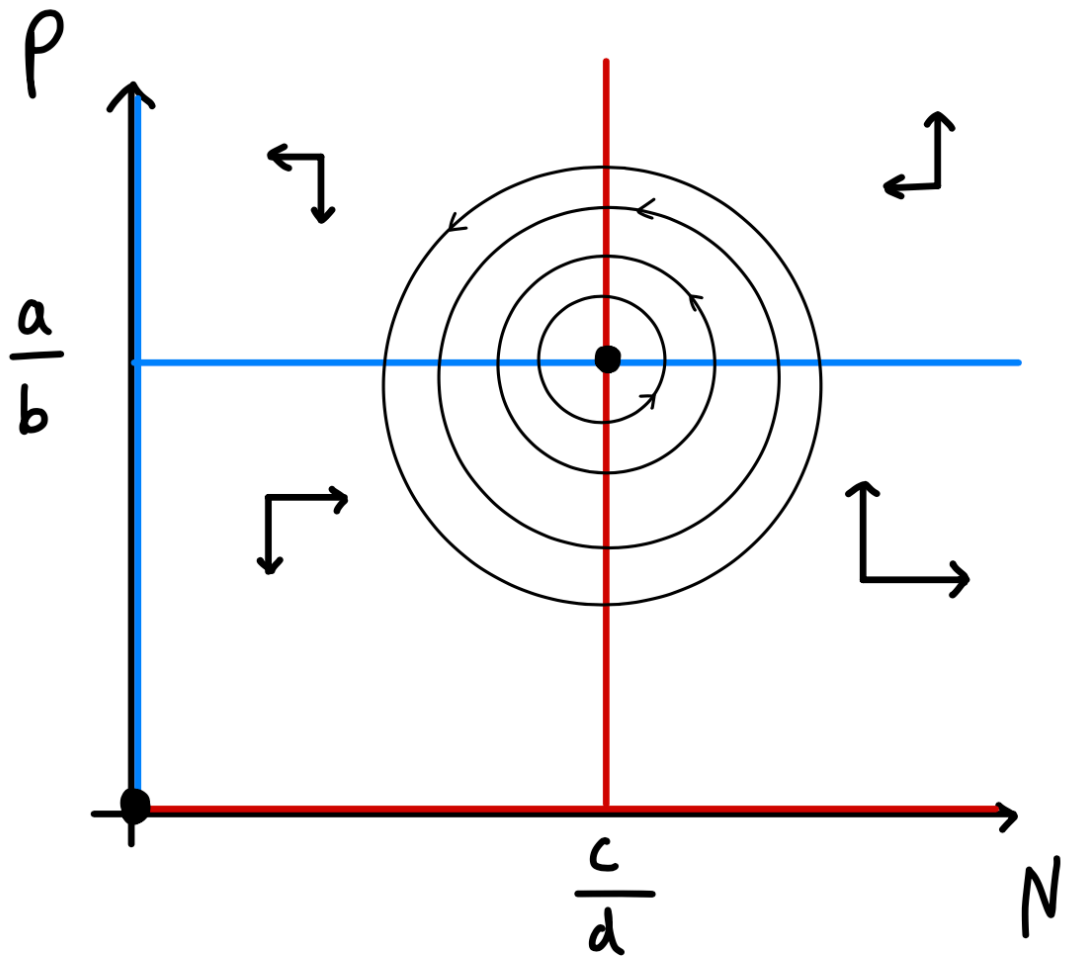
Recall that to write down the Jacobian we go through the ODEs in turn and form a matrix of the partial derivatives. For this model that gives us,

$$J = \begin{pmatrix} a - bP^* & -bN^* \\ dP^* & -c + dN^* \end{pmatrix},$$

We can quickly see that the extinction equilibrium, $(N, P) = (0, 0)$, is always unstable (since $a > 0$). What about the coexistence equilibrium? This evaluates to:

$$J = \begin{pmatrix} 0 & -bc/d \\ ad/b & 0 \end{pmatrix},$$

Here, $tr = 0$ and $\det = ac > 0$. These conditions mean that the eigenvalues at this equilibrium have zero Real part and are purely Imaginary. The behaviour near the steady state is therefore 'centres', a family of neutrally-stable closed orbits around the equilibrium. We therefore predict that these two populations will constantly cycle. With low predation the prey density increases; this produces more food for predators, and so the predator numbers begin to rise; this in turn starts to push the prey numbers back down; finally, with their food supplies falling, the predator numbers also drop down. A phase portrait of these dynamics would look like this,



Phase portrait for the simple predator-prey model. The blue lines give the prey nullclines and the red lines the predator nullclines, and the black curves are sample trajectories. These reveal centres.

Such a result initially seems appealing- this fluctuating behaviour makes intuitive sense and seems in line with some data from real systems. However, being centres these cycles are not structurally stable. The fact that if we make a change to the initial values we end up on a slightly different cycle is definitely mathematically unsatisfying and also questionable biologically. Let us see if we can update our model and get a stronger result.

A MORE REALISTIC MODEL

From our previous models we might already think of two additions to make the model more realistic. One is that we ignored intraspecific competition (i.e. no carrying capacity), meaning prey in particular could grow to infinite levels. Another is that we assumed predation was linearly dependent on prey density, whereas in the spruce budworm model we argued it should be saturating. Including these assumptions modifies the model to,

$$\frac{dN}{dt} = N\left(a - \frac{bP}{N+A} - \alpha N\right)$$
$$\frac{dP}{dt} = P\left(-c + \frac{dN}{N+A}\right).$$

Again we can start by looking for equilibria. We can see we still have the equilibrium at $(0,0)$, and some coexistence equilibrium (N^*, P^*). There is now also a new 'prey-only' equilibrium at $(N,P) = (a/\alpha, 0)$. To assess stability we will need to also update our Jacobian, which is now,

$$J = \begin{pmatrix} a - \frac{bP}{N+A} - 2\alpha N + \frac{bPN}{(N+A)^2} & \frac{-bN}{N+A} \\ \frac{dAP}{(N+A)^2} & -c + \frac{dN}{N+A} \end{pmatrix},$$

The extinction equilibrium will remain a saddle always, which can be seen by using the Jacobian or simply considering the dynamics of the two species near that point. Let us look in more detail at the two other equilibria.

Prey-only equilibrium

In this case the Jacobian reduces to,

$$J = \begin{pmatrix} -\alpha N & \frac{-bN}{N+A} \\ 0 & -c + \frac{dN}{N+A} \end{pmatrix},$$

Noting that we have a 0 in the bottom-left, we can read off our eigenvalues as $\lambda_1 = -\alpha N^* < 0$ and $\lambda_2 = -c + dN^*/(N^* + A)$. Therefore stability depends on λ_2, which is actually the growth rate of predators near this equilibrium (i.e. with near-zero predator densities). If the growth rate is negative the prey-only equilibrium is stable, but if it is positive the prey-only equilibrium is a saddle, which seems like what we would expect.

Coexistence equilibrium

As in the example with competition, we have not yet written out the densities for the prey and predators at the coexistence equilibrium, but we can still assess its stability by using the fact we are at the equilibrium. This means we know that $a - bP/(N + A) - \alpha N = 0$ and that $-c + dN/(N + A) = 0$. Substituting these into the Jacobian leads us to,

$$J = \begin{pmatrix} -\alpha N + \frac{bPN}{(N+A)^2} & \frac{-bN}{N+A} \\ \frac{dAP}{(N+A)^2} & 0 \end{pmatrix},$$

In this case we need to use the trace and determinant to assess stability. We can find that the determinant is given by, $\det = bdAN^*P^*/(N^* + A)^3 > 0$, and the trace by, $tr = -\alpha N^* + bN^*P^*/(N + A)^2$. Stability therefore depends on whether the trace is positive or negative, which we can see depends on the balance of intraspecific competition in the prey and predation: if competition is high relative to predation, we will have a negative trace and the equilibrium is stable; if predation is high relative to competition, we will have a positive trace and the equilibrium is unstable.

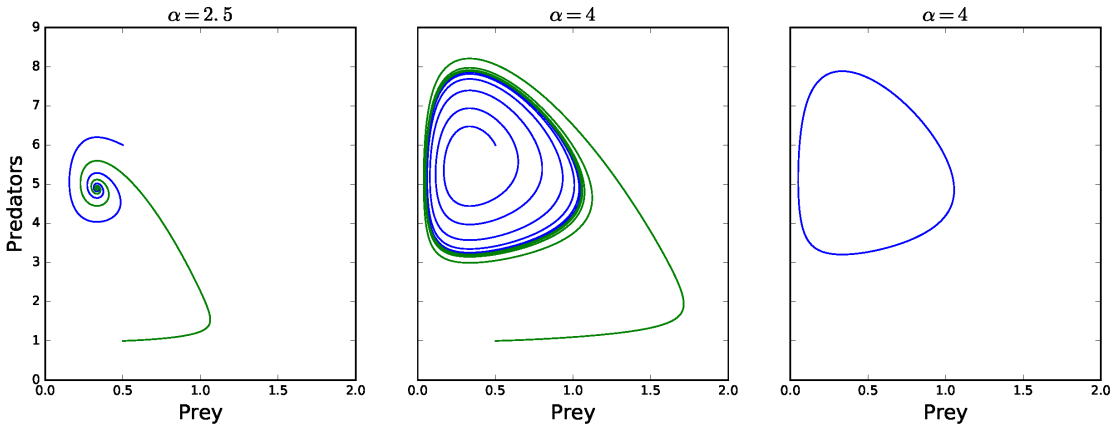
Notice that if we vary one parameter, α, say, we could move the system from being a stable spiral to being an unstable spiral, since we move from a negative to a positive trace. This is a change in stability, and therefore a bifurcation, but one that we have not yet seen in practice and which cannot occur in one-dimensional systems.

HOPF BIFURCATION

This transition is called a Hopf bifurcation. In this case changing a parameter turns a stable spiral in to an

unstable spiral (or vice versa). This is a particularly important behaviour, because as we initially move into the unstable spiral region, a unique, stable closed orbit emerges from the equilibrium (see the background review chapter). This results in the population cycling/oscillating, and the orbit is called a *limit cycle*. It is not very easy to draw a bifurcation diagram in this case, but we can see the process by looking at the phase portraits as we vary a parameter.

The example in the figure below comes from our predator-prey model we have just examined. Here we see we initially have a stable spiral into the coexistence equilibrium, but that as we increase α the equilibrium loses stability and starts to spiral away. However, this outward trajectory does not continue to spiral out but tends towards a closed orbit (which can be seen clearly in the right-hand figure where I only plotted later time-points). An initial condition from further away starts to spiral in but also approaches this closed orbit. This is a much stronger result for population cycles than the centres we saw in the simple model. Hopf bifurcations have important consequences as populations which are fluctuating may be much harder to measure or control.



Example of a Hopf bifurcation and emergence of a limit cycle in the predator-prey model with a type II functional response. Parameter values: $= 5, b = 1, c = 0.5, d = 2, A = 1$.

Explore the model

Use the Python code below to explore the emergence of the limit cycle in this model. The code produces plots of time-courses and a phase portrait for two initial conditions. The given parameter values lead to a stable spiral. Try gradually reducing the value of α to see the limit cycle emerge and grow.

Click for code

```
# Import the necessary libraries
import numpy as np
import matplotlib.pyplot as plt
```

```
from scipy.integrate import solve_ivp
# Options to make the plots the right size
plt.rcParams['figure.figsize'] = [12, 4]
plt.rcParams.update({'font.size': 16})
#Function for the dynamics called 'predprey'
def predprey(t,x):
    # Rename the variables for ease
    N=x[0]
    P=x[1]
    # The ODEs
    dN = N*(a-b*P/(N+A)-alpha*N)
    dP = P*(-c+d*N/(N+A))
    return [dN,dP]
# Parameter values
a=4
b=1
c=0.5
d=2
alpha=4
A=1
# Initial conditions
N0_1=1.5
P0_1=6
N0_2=0.5
P0_2=4
N0=[N0_1,P0_1]
N1=[N0_2,P0_2]
# Time points to use
tc = np.linspace(0, 50, 1000)
# Run the model using 'solve_ivp' for two different ICs
Nc = solve_ivp(predprey, [tc[0],tc[-1]], N0, t_eval=tc)
Nd = solve_ivp(predprey, [tc[0],tc[-1]], N1, t_eval=tc)
# Plotting commands
fig, (ax1, ax2) = plt.subplots(1, 2)
ax1.plot(tc, Nc.y[0], "r", label="Prey")
ax1.plot(tc, Nc.y[1], "k", label="Predators")
ax1.plot(tc, Nd.y[0], "r:")
```

```
ax1.plot(tc, Nd.y[1], "k:")
ax1.set(xlabel='Time', ylabel='Densities')
ax1.legend()
ax1.axis([0,50,0,10])
nn=np.linspace(0,10,100)
nnull=(a-alpha*nn)*(nn+A)/b
pnull=A*c/(d-c)
ax2.plot(Nc.y[0],Nc.y[1],'b')
ax2.plot(Nd.y[0],Nd.y[1],'g')
ax2.plot(nn,nnull,'r')
ax2.axvline(x=pnull)
ax2.axis([0, 5, 0, 5])
ax2.set(xlabel='Prey', ylabel='Predators')
ax2.axis([0,2,0,10])
```

Key Takeaways

- The simplest predator-prey model produces *centres* – structurally unstable cycles.
- A more realistic predator-prey model can produce stable *limit cycles*.
- We call the transition from an equilibrium to a limit cycle a *Hopf bifurcation*.

Chapter references

- The content in the *Population ecology* section was influenced by the textbook, *Mathematical Biology 1* by Murray.

PART II

INFECTIOUS DISEASE

CHAPTER 4

Epidemics in human populations

THE SUSCEPTIBLE-INFECTED-RECOVERED MODEL

At the time of writing, it seems reasonable to say that mathematical modelling of disease spread has never been as topical or well-known amongst the general public. Many modellers and scientists have been contributing to our understanding of the Covid-19 pandemic, from research being conducted and published, to directly advising government policy, to public engagement and information. In the next few chapters we will cover the fundamentals of the mathematical models that are the basis of most of this modelling work. The models that have been used to advise governments are of course rather more detailed than those you will see here (and we'll discuss some extensions as we go along), but there is a common engine behind them all.

Our first few models focussed on the dynamics of ecological populations. As part of this we have assumed that every individual within each population is identical. Such a generalisation has the benefit of reducing our model to a low number of variables. But often certain individuals within a population are fundamentally distinct from one another. In this case, we might want to divide our population up in to different *compartments*. The choice of how to divide our population is often dictated by the biological question we are seeking to address. For example we may wish to have an age-structured model, with adults and juveniles, or perhaps individuals with different roles in a population such as foragers or workers.

For the next few chapters we will focus on *epidemic models*, where we will need to use such a compartment model. Broadly, we will think about disease dynamics within populations of humans, animals or plants that act as *hosts* for diseases such as viruses or bacteria, often termed *parasites* or *pathogens*. We will initially assume we are interested in infectious diseases of human populations, and this will guide our model design. When thinking about the spread of disease through a population, while we might have some interest in the total population at any time, it is more likely that we want to know about individuals' infection status – are they currently infected, for example. This will be the basis of our compartment model. Let's draw a schematic diagram to picture this process:

Schematic of the SIR model.

Before a disease emerges, all individuals in the population are *susceptible*. Once exposed to the disease,

individuals may then become *infected* (and also *infectious*). Finally, individuals will eventually fight off the disease to become *recovered* and *immune*.

This is the 'SIR' model (i.e. **S**usceptible, **I**nfected, **R**ecovered), and has been central to our understanding of disease dynamics in human populations since it was first proposed in the 1920s. Disease is transmitted by 'direct contact' between infected and susceptible hosts and is a 'mass action' process, meaning we do not take any account of contact networks and the risk of infection is proportional to the total density of infection in the population. When hosts have recovered, they gain immunity and cannot be infected again (this is a key aspect of vertebrate – and especially human – immune systems, but is not necessarily true of invertebrates, plants, etc. We will consider models more appropriate for such populations later). β controls the rate at which contacts between S and I hosts cause infection, and γ is the rate of recovery. A small personal pet peeve – β itself is *not* a rate; it has the wrong units. In this initial example we have not included any birth or death processes, as we assume that infection happens at a much faster timescale than demography. By considering our schematic diagram above we can write down the dynamics of our system as a set of ODEs:

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

with the total population, $N = S + I + R$. There are some useful definitions it is worth making at this point:

- βI is often called the force of infection (this really is the rate of infection);
- I/N gives the prevalence of infection;
- $1/\gamma$ is the infectious period.

One thing to spot is that $dN/dt = dS/dt + dI/dt + dR/dt = 0$. This means that the total population stays at a constant size (if we stop and think for a moment you might realise that this was built into our model all along since we have no births, deaths or migration). We can argue that this is a reasonable assumption for a short epidemic outbreak in long-lived human populations. An advantage of this assumption is that we can eliminate one variable, with $R = N - (S + I)$ and ignore the dR/dt equation. This is because since the total population is always N, once we've worked out the densities of S and I we can immediately calculate the remaining density who must be R. We therefore can write down the simplified model as,

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

ANALYSIS

Once again we have what looks to be a fairly simple set of ODEs (two variables and two parameters), but there exists no explicit solution due to the non-linear term βSI. However, we can again explore the qualitative behaviour of the model by identifying equilibria of the system, assessing their stability and sketching phase portraits.

First let's look for equilibria. The only way both ODEs can be 0 simultaneously is if $I = 0$ (and S may take any value less than N). This means we have a line of non-isolated equilibria where there are no infected

individuals in the population. This is another special case for stability as we would find that one of the eigenvalues is always zero. We will consider quite what is going on here when we sketch the phase portraits.

Before we do that, let us think a bit more about the general behaviour of the model. It is often the case with emerging diseases that they will eventually burn out, as we seem to be predicting here. But it is still important for us to know whether there can be an *epidemic*, when an initially small amount of infection causes a large outbreak of disease in the population (even if it eventually tends to zero). Taking this definition, for an epidemic to occur we need to have $dI/dt > 0$ initially. Before an outbreak, the initial densities are $S(0) \approx N, I(0) \approx 0$. This gives,

$$\frac{dI}{dt} \approx (\beta N - \gamma) I$$

This gives us a condition for an epidemic:

- $\beta N - \gamma < 0 \implies$ disease dies out.
- $\beta N - \gamma > 0 \implies$ epidemic.

So even though the ultimate dynamic is for the disease to die out, we can have an epidemic if the initial amount of infection outstrips the rate of recovery.

Let's now draw phase portraits of our SIR model for the two scenarios we have found above. To find the nullclines we take,

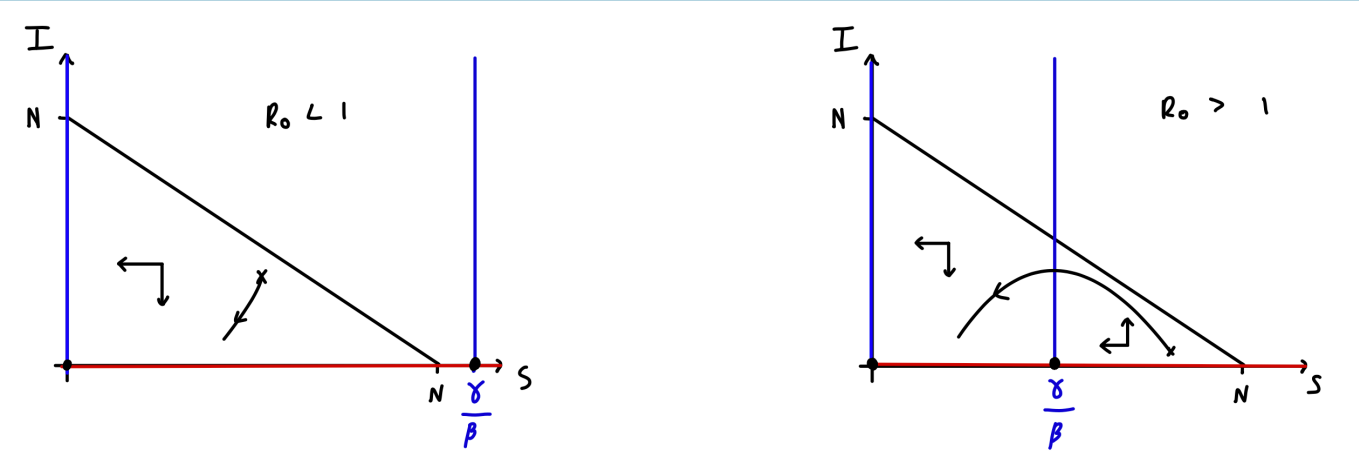
$$\frac{dS}{dt} = 0 \implies S = 0, I = 0$$
$$\frac{dI}{dt} = 0 \implies S = \frac{\gamma}{\beta}, I = 0$$

Exercises

Draw the two qualitatively different phase portraits for this system, one where $\gamma/\beta > N$ and one where $\gamma/\beta < N$.

Click for solution

The two plots are sketched below. The first thing to notice is that in both plots only a triangular region bounded by the axes and $I = N - S$ is feasible. That is because if we took other positive values of S and I we would have $S + I > N$ which would not be possible given our fixed population size. In both diagrams we can quickly see that S should be decreasing everywhere (since $dS/dt < 0$). In the first diagram we only have one region to worry about the qualitative direction of flow of I, since the vertical nullcline at $S = \gamma/\beta$ is outside the feasible region. In the second diagram we have two regions, towards the right I increases and towards the left it decreases.



Phase portraits of the SIR model. The blue lines give the infected nullclines and the red line the susceptible nullclines. The black line marks the region of biological feasibility, and the curves example trajectories.

Remember where two (different) nullclines intersect is necessarily an equilibrium. Since both ODEs give a nullcline at $I = 0$ we reinforce the fact that we have a continuum of equilibria along the line $I = 0$. Using our equations we can then find out in what regions S and I will be increasing or decreasing. Again, note it doesn't matter that we can't say precisely the direction of a trajectory anywhere on our plot; knowing the qualitative direction of travel gives us all the information that we need.

In the first case, the vertical nullcline at $S = \gamma/\beta$ does not appear in the feasible region, so actually has no impact on the dynamics. In fact we will always just drop down towards $I = 0$ and the disease dying out. In the second case, the phase portrait is divided into two regions. If we start in a region to the right of the nullcline, the infected density initially increases – we have an epidemic – before crossing the vertical nullcline and approaching $I = 0$ again. Therefore the equilibria where $S < \gamma/\beta$ appear to be locally stable (which it turns out is equivalent to the second eigenvalue being negative).

The divide between these two cases is whether $N > \gamma/\beta$ or not. We can rearrange this to the condition $\beta N/\gamma > 1$. This ration is known as R_0, so we have the if $R_0 > 1$ we get an epidemic but if $R_0 < 1$ the disease dies out.

THE BASIC REPRODUCTIVE RATIO, R_0

There is a very strong chance you have heard the terms R_t or R_0 in discussions about the speed of spread of Covid-19. We define $R_0 = \beta N/\gamma$ as the basic reproductive ratio of the disease: 'The average number of secondary infections caused by one infected host in an otherwise disease-free population.' Note that βN gives the total infections caused in a disease-free population, and $1/\gamma$ is the infectious period. So R_0 gives the number of infections caused by an individual in the time that it's infected near the start of an epidemic. The more general term, $R_t = \beta S/\gamma$ gives the number of new infections per case later in the epidemic when many individuals have already been infected and/or recovered. Because R_0 has this rather intuititve definition, it is a value that can often be estimated or measured. A number of estimates for R_0 of different

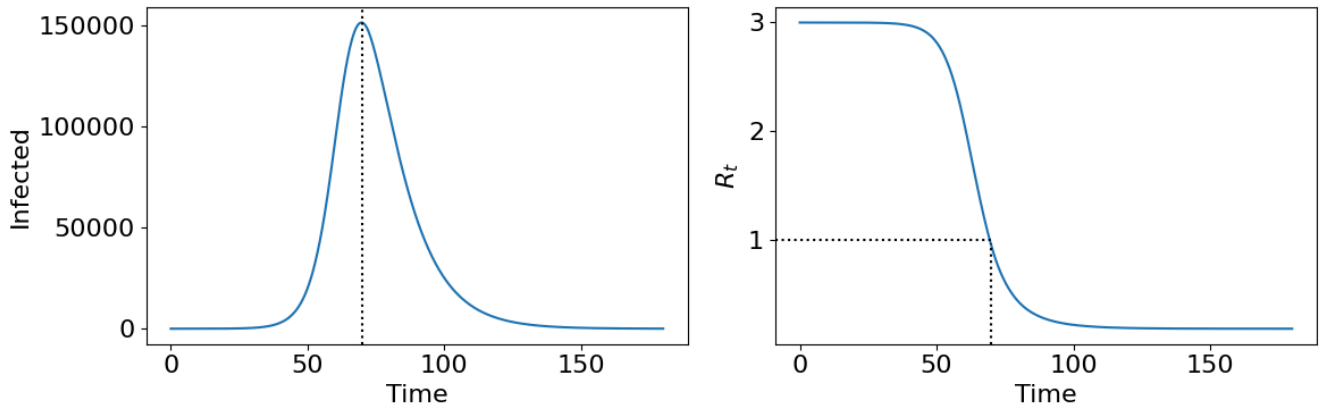
diseases are given below. Remember, this figure tells you (roughly) how many new people an infected person will infect.

Disease	R_0
Flu	1-3
Covid-19	2-5
SARS	2-5
HIV	2-5
Smallpox	5-7
TB	8-10
Measles	12-18

Estimates of R_0 for some important human diseases.

THE EPIDEMIC CURVE

The phase portraits demonstrate again that an epidemic should always burn out and the population return to being disease-free. Interestingly the phase portraits also suggest that not all of the population will get infected during an epidemic, since we usually do not end with $S = 0$ and $I = 0$ (which would mean $R = N$). As we have already said, due to the non-linearity of the system we are unable to find an exact solution and therefore cannot say precisely what the dynamics should look like. However, it is useful to understand what the epidemic curve – the number of people infected over time – looks like. With some approximations it is possible to express this as a mathematical formula, but we shall just look at some numerical output as shown in the left-hand side of the figure. This gives a characteristic bell-shaped curve. This is in fact the sort of curve we see from data from many real-world epidemics, suggesting that even our simple model is capturing something fundamental about disease spread. In the right-hand side we show how the reproductive ratio R_t changes during the epidemic. Notice that the peak of the epidemic is precisely at $R_t = 1$. This will be important later when we discuss the idea of *herd immunity.*



An epidemic curve based on a populaion of N=500,000 and $R_0 = 3$ and how the value of R_t changes during the epidemic. The dashed line marks the peak of the epidemic, which is exactly when $R_t = 1$.

Explore the model

Use the Python code below to explore the SIR model. In particular, this code will take an alternative approach by presenting the output from a *stochastic simulation algorithm*.

When using differential equations – as we are focussing on in this course – the models are *deterministic* meaning for the same parameter values and initial conditions you get the exact same dynamcis every time. In a *stochastic* model we account for variation in the parameters – for example in the ODE model a recovery rate of $\gamma = 1/7$ is effectively interpreted as saying everyone recovers after exactly 7 days, whereas in a stochastic model we instead assume the recovery rate of each individual is drawn from a distribution where 1/7 is the mean. This means we get slightly different time-courses every time we run the model.

The code below will run 20 stochastic simulations as well as the deterministic equivalent. How good an approximation is the deterministic model to the stochastic one? Can you think of ways you might communicate the variation in the time-courses to policy makers or the public? Try changing the value of R_0 to see how the behaviour and differences change.

Click for code

```
# Import the necessary libraries
import numpy as np
import matplotlib.pyplot as plt
from scipy.integrate import solve_ivp
# Options to make the plot the right size
plt.rcParams['figure.figsize'] = [8, 4]
plt.rcParams.update({'font.size': 16})
#Parameter values
N=1000
GAMMA=1/14
R0=2.5
BETA=R0*GAMMA/N
REPS=20 # Number of times to run the stochastic model
I0=2 # Starting number of infected individuals
# STOHASTIC MODEL CODE
# Function to count numbers in each compartment
def count_type(type):
    current_type=0;
    for i in range(0,N):
        for j in range(0,N):
            if grid[i,j]==type:
                current_type=current_type+1
    return current_type
# Function to check the current scale
```

```
 def findscale():
     S=susceptibles[-1]
     I=infecteds[-1]
     #Set relative parameter values
     scale=GAMMA*I+BETA*S*I
     return scale
 # Main function
 for reps in range(0,REPS):
     # Set initial conditions
     tsteps=[0]
     infecteds=[I0]
     susceptibles=[N-I0]
     current_t=0
     # Main run of stochastic model
     while current_t<180:
         # Find time step
         scale=findscale()
         dt = -np.log(np.random.rand()) / scale
         current_t=tsteps[-1]
         tsteps.append(dt+current_t)
         #Find which event happens
         if np.random.rand()<GAMMA*infecteds[-1]/scale: #Event is recovery
             infecteds.append(infecteds[-1]-1)
             susceptibles.append(susceptibles[-1])
         else: #Event is transmission
             infecteds.append(infecteds[-1]+1)
             susceptibles.append(susceptibles[-1]-1)
         if infecteds[-1]==0:
             break
     # Plot latest run
     if reps==0:
         plt.plot(tsteps,infecteds,':',color='blue',alpha=0.5,label='Stochastic mode
l')
     else:
         plt.plot(tsteps,infecteds,':',color='blue',alpha=0.5)
 # ODE MODEL
 # Function to run model called 'disease'
```

```
def disease(t,x):
    sdot=-BETA*x[0]*x[1]
    idot=BETA*x[0]*x[1]-GAMMA*x[1]
    return sdot, idot
# Time points to use in ODE model
ts=np.linspace(0,180,2000)
# Run ODE model using 'solve_ivp'
xx1=solve_ivp(disease,[ts[0],ts[-1]],[N-I0,I0],t_eval=ts)
# Plotting commands
plt.plot(ts,(xx1.y[1]),'k',label='ODE model')
plt.legend()
plt.xlabel('Time')
plt.ylabel ('No. Infected')
plt.xlim(0,180)
```

Key Takeaways

- We can model diseases using a compartment framework called the SIR model.
- An epidemic will occur when dI/dt is initially positive, but in the simplest model will always reach an equilibrium where $I = 0$.
- We can measure how quickly a disease initially spreads using R_0, and this tells us with there will be an epidemic or not.

Chapter references

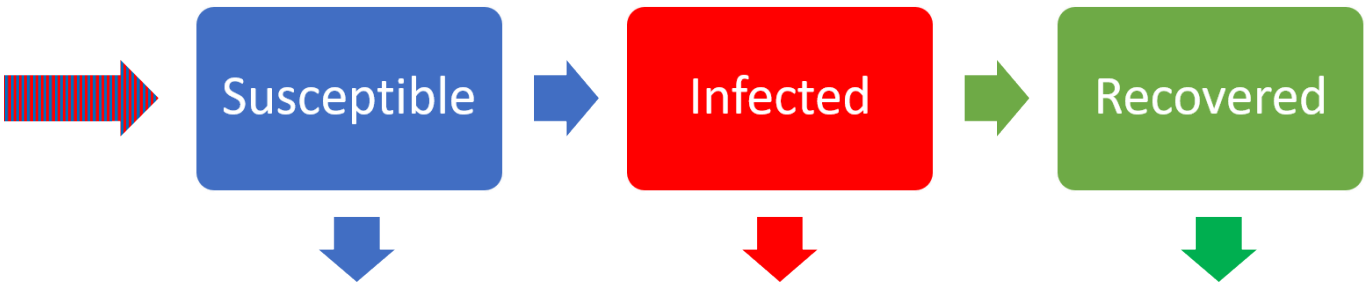
- The content in the *Infectious diseases* section was influenced by the textbooks, *Mathematical Biology 1* by Murray and *Modelling Infectious Diseases in Humans and Animals* by Keeling & Rohani.

CHAPTER 5

The SIR model with demographics

INCLUDING BIRTHS AND DEATHS

In our initial epidemic model in the previous chapter we had only two mechanisms involved – infection and recovery. Let us now add some increased detail into the model by including birth and death processes. Given the timescales of epidemic processes, this likely means looking at dynamics over a much longer time period, at least in human populations. Let's begin by drawing a schematic of the system again:



Schematic of the SIR model with births and deaths now included.

We will make the simplifying assumption that the birth rate and death rate are equal, so that the population would be at equilibrium in the absence of disease (this is a questionable assumption for many ecological populations, but perhaps not too unreasonable for modern human populations). If we assume all individuals produce offspring at rate μ, that everyone is born susceptible, and that every individual has the same death rate, also μ, this leads us to the equations,

$$\frac{dS}{dt} = \mu N - \beta SI - \mu S$$
$$\frac{dI}{dt} = \beta SI - (\gamma + \mu)I$$
$$\frac{dR}{dt} = \gamma I - \mu R$$

Once more, $dN/dt = 0$ so we can eliminate $R = N - (S + I)$ (again, this is somewhat by design – if the rates of birth and death were not equal this simplification could not be made). One thing to note here is that the infectious period has changed to be $1/(\gamma + \mu)$. When we define R_0, therefore, in our updated model we will have $R_0 = \beta N/(\gamma + \mu)$.

ENDEMIC DISEASE

At this point we could non-dimensionalise our system as we have seen previously. This would allow us to reduce the number of parameters in our model to make life easier, as well as revealing potentially useful information about the scales involved. You will see some studies do this and others don't. For now, let us continue with our analysis without doing so, since it means we retain the clear biological meanings of all of our parameters and variables.

First we should find the equilibria of our system, where $dS/dt = 0$ and $dI/dt = 0$ simultaneously. Recall in the previous model this only happened when $I = 0$, with this condition satisfying both ODEs and leaving us with a line of equilibria. Now, we find more 'standard' unique equilibria. There are two cases where $dI/dt = 0$:

- $I^* = 0$, giving $dS/dt = 0 \implies S^* = N$;

This means the population is disease-free.

- $S^* = \dfrac{\gamma + \mu}{\beta}$, giving $dS/dt = 0 \implies I^* = \dfrac{\mu(N - S^*)}{\beta S^*} = \dfrac{\mu}{\beta}\left(\dfrac{N}{S^*} - 1\right)$;

This means the disease is *endemic*.

We can do a bit more manipulation of that last equilibrium,

$$I^* = \frac{\mu}{\beta}\left(\frac{N}{S^*} - 1\right) = \frac{\mu}{\beta}\left(\frac{\beta N}{\gamma + \mu} - 1\right) = \frac{\mu}{\beta}(R_0 - 1)\,.$$

While both expressions are equally correct, the latter one is nice in that R_0 is included in the expression. We now have two important terms from epidemiology. An *epidemic* occurs when a disease initially spreads through a population. A disease is *endemic* when it remains at a steady level within the population.

Having found these equilibria we now want to classify their stability. Let's look at the Jacobian for our system,

$$J = \begin{pmatrix} -\beta I^* - \mu & -\beta S^* \\ \beta I^* & \beta S^* - (\gamma + \mu) \end{pmatrix}$$

We shall now deal with each of our equilibria in turn.

Disease-free equilibrium

$$J = \begin{pmatrix} -\mu & -\beta N \\ 0 & \beta N - (\gamma + \mu) \end{pmatrix}$$

Having a zero in an off-diagonal (for a 2×2 matrix) makes our life much easier, as we can just read off the eigenvalues as the two diagonal entries. In this case we therefore have $\lambda_1 = -\mu$ and $\lambda_2 = \beta N - (\gamma + \mu) = (\gamma + \mu)(R_0 - 1)$. Therefore, if $R_0 < 1$, the disease-free equilibrium is unstable.

Endemic equilibrium

$$J = \begin{pmatrix} -\mu R_0 & -(\gamma + \mu) \\ \mu(R_0 - 1) & 0 \end{pmatrix}$$

In this case we cannot read off our eigenvalues, so we shall instead look at the signs of the trace and determinant:

- $tr(J) = -\mu R_0 < 0$
- $\det(J) = \mu(\gamma + \mu)(R_0 - 1)$

Since the trace is always negative, from the determinant we see that if $R_0 > 1$, the endemic equilibrium is stable.

The stable equilibrium can be a node or a spiral depending on the precise parameter values. Generally, high γ, intermediate R_0 and low μ make a spiral more likely.

Putting these results together we can see that when $R_0 > 1$ the disease will become endemic.

PHASE PORTRAITS

We can explore these two scenarios further with phase portraits. First we remember that again we must have $S + I \leq N$, so we only have a triangular region in our phase portraits to really worry about. To find the nullclines:

$$dS/dt = 0 \implies I = \frac{\mu}{\beta}\left(\frac{N}{S} - 1\right)$$

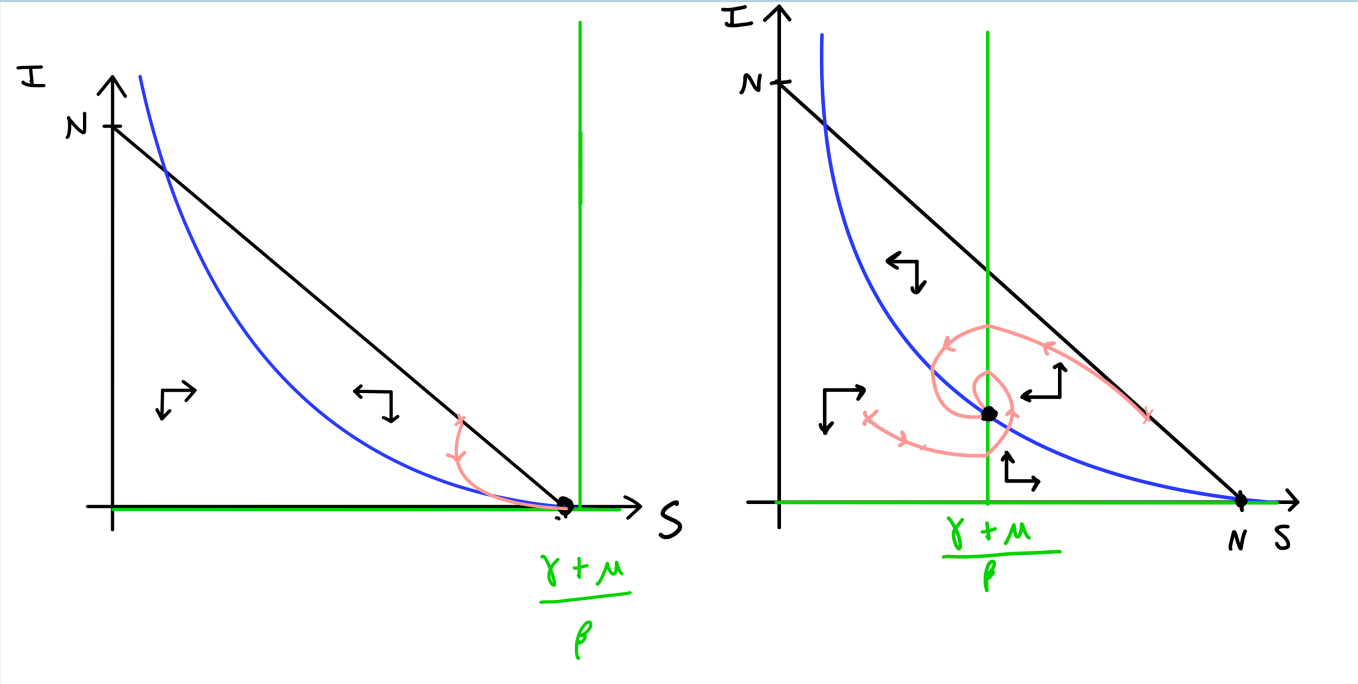
$$dI/dt = 0 \implies I = 0, S = \frac{\gamma + \mu}{\beta}.$$

Exercises

Draw the two qualitatively different phase portraits for this system.

Click for solution

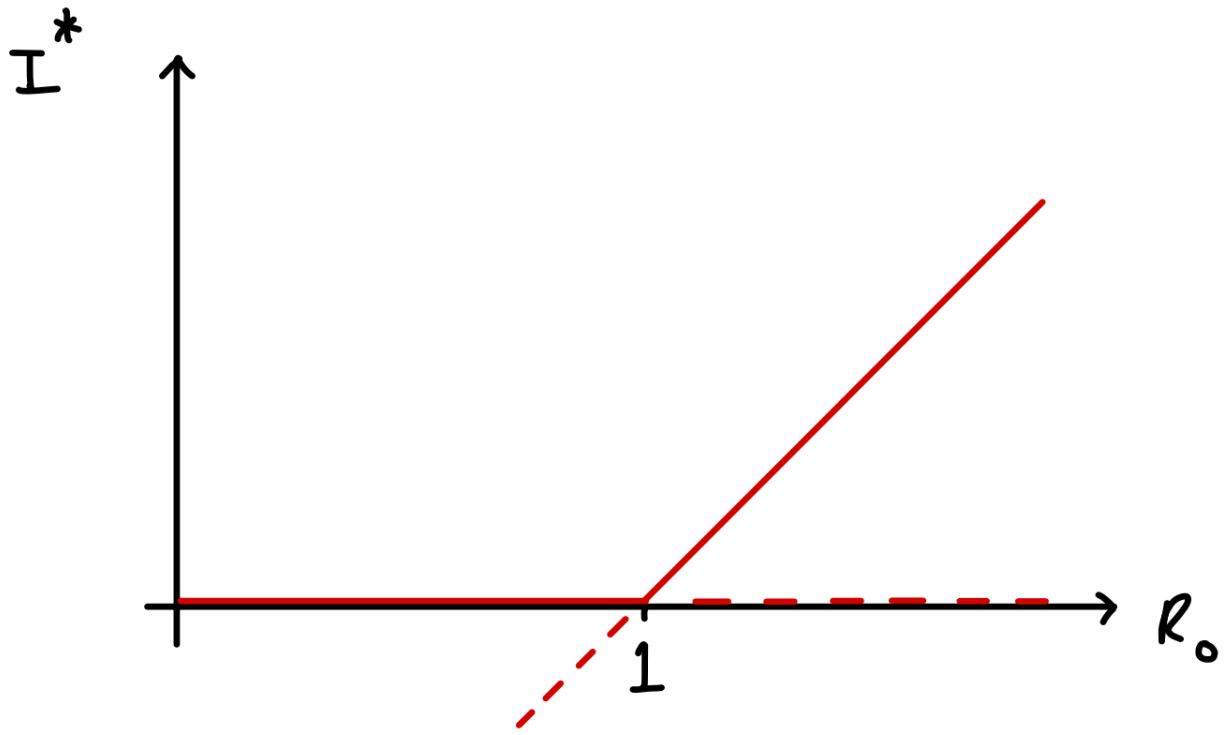
The two phase portraits are sketched below. If you start sketching a phase portrait you should find that the only way you can produce two qualitatively different phase portraits is again about the placement of the vertical nullcline, now given by $S = (\gamma + \mu)/\beta$, which in turn relates to whether or not $R_0 > 1$ or not. In the first diagram, all trajectories will eventually approach the bottom-right corner at the disease-free equilibrium. In the second diagram, you should see the new endemic equilibrium in the middle where the two nullclines cross. Trajectories approach this equilibrium, moving around it anti-clockwise (as I once read, if you can't remember clockwise/anti-clockwise directions, clockwise is the way you dial a phone).



Two phase portraits for the SIR model. Blue lines give the susceptible nullclines, green lines the infected nullclines, and the black line the boundary of biological feasibility, with trajectories in pink.

We see nice demonstrations of the behaviour we predicted from our stability analysis here. In particular, when $R_0 > 1$ the disease will persist in the long-term.

R_0 is a quantity that is often estimated for diseases when they emerge, and gives a good indication of how contagious they are and (as we shall see below) how easy they will be to eradicate. We can also use R_0 as a useful bifurcation parameter to draw a bifurcation diagram of this system below. Note again the key value of $R_0 = 1$ where the transcritical bifurcation occurs.



Bifurcation diagram of the SIR model, plotting the equilibria of infected densities in terms of R_0.

VACCINATION

Suppose we want to prevent a disease spreading through a vaccination programme. What proportion, p, of the population do we need to vaccinate to succeed? If our vaccine works perfectly then the pre-infection density of susceptibles reduces from N to $N(1-p)$. Recall that for the disease to die out we want $R_0 = \beta N/(\gamma+\mu) < 1$. After vaccination this becomes,

$$\frac{\beta N(1-p)}{\gamma+\mu} \implies p > 1 - \frac{1}{R_0}$$

Note that not all of the population needs to be vaccinated – just enough to prevent the disease from spreading freely. This is important as there may be certain groups, those undergoing medical treatment, for example, who cannot be vaccinated. This population-wide protection created by significant vaccination is called *herd immunity*. The higher the R_0 of the disease, the greater proportion of the population that needs to be vaccinated.

Exercises

A novel infectious disease has been detected in a small town with a population of 10,000 people. The yearly birth and death rates of the population are estimated to be 0.02. From previous outbreaks in other towns the estimated transmission coefficient is $\beta = 7.5 \times 10^{-4}$ and the (yearly) recovery rate is 4.48. A vaccine exists that will perfectly prevent infections from occurring. What proportion of the population must be vaccinated to achieve herd immunity?

Click for solution

Using the values given we can find the value for R_0 as,

$$R_0 = \frac{\beta N}{\gamma + \mu} = \frac{7.5}{4.5} = \frac{5}{3}.$$

Since the vaccine prevents all infections we can use our herd immunity threshold as above to give,

$$\begin{aligned} p &> 1 - \frac{1}{R_0} \\ &> 1 - \frac{3}{5}. \end{aligned}$$

We therefore need to vaccinate at least 40% of the population to achieve herd immunity.

A POINT ABOUT HERD IMMUNITY

This method to calculate the herd immunity threshold relies on using R_0 as our key parameter, which is the relative speed of spread *if the initial population is disease-free*. During the Covid-19 pandemic, proposals were made for a "herd immunity strategy" that would essentially rely on the disease dying away once enough of the population – the herd immunity threshold – had immunity, but immunity that was acquired after being infected, not through vaccination. There are two big drawbacks to this approach (see chapter references).

Firstly, as we have seen here, if the disease persists for any length of time – such that birth and death processes start to introduce population turnover – we'd be unwise to assume the disease would naturally burn out. There are a whole host of other factors that would also allow for a disease to persist in the long-term, including waning immunity and lack of immunity to mutations to name just two.

Secondly, and related to R_0, if we use the herd immunity threshold as calculated above, the idea is that we have protected (almost) the whole population until a vaccine was developed, and then we can move enough individuals out of the susceptible compartment to stop the disease from ever spreading. If instead we allow the disease to spread the key parameter is no longer R_0 but R_t, as discussed in the last chapter. We know R_t equals 1 exactly at the peak of an epidemic, and at this point the proportion that has been infected will be the same as the herd immunity threshold. However, in this approach we now have a potentially large proportion of the population who are still infected, as well as a large proportion who remain susceptible. So while the number of infections will start to fall (because we are in the region of $R_t < 1$) we will still have a considerable number of onward infections. Say that we calculate that a disease has $R_0 = 3$, meaning the herd immunity threshold is 66.7%. As we have just said, that is now the proportion that has been infected *at the peak of the infection*. By the time the infection has actually died out, the proportion infected will be over 90%.

The lesson here is to be careful about how you interpret models!

Key Takeaways

- We can extend the SIR model to be more realistic for long-term predictions by including births and deaths.
- We now get an endemic equilibrium, where the disease will remain in the population over the long term.

- R_0 remains a key quantity, and is helpful in determining the threshold for *herd immunity*.

Chapter references

- The discussion on herd immunity is based on the paper by Best & Ashby (2022).
- The content in the *Infectious diseases* section was influenced by the textbooks, *Mathematical Biology 1* by Murray and *Modelling Infectious Diseases in Humans and Animals* by Keeling & Rohani.

CHAPTER 6

Diseases of ecological populations

INTRODUCTION

In the models of infectious diseases we have covered so far we have been focussing on human populations. However, infectious diseases are also extremely important to the dynamics of many animal, plant and microbial populations (for example, Foot and Mouth, Bovine TB, Wheat Rust, etc.). We should therefore extend our models to consider these cases. There are quite a few changes to our models that we may wish to make so that our models are more reflective of wider ecological populations. Three particular ones we will consider here are:

- We shall now assume that there is no immunity. Hosts can still recover from infection, but they will simply become susceptible once again. (The degree to which even simple organisms possess adaptive immunity is in fact a fascinating question, but for now we will assume it is absent).
- We can no longer reasonably assume that births and deaths are equal.
- We should include the fact that disease causes significant damage to hosts (something we have so far ignored). In particular, we shall assume that infected hosts suffer an additional mortality, or *virulence* at rate α.

Putting these assumptions together with our previous model, we have a new SIS (Susceptible-Infected-Susceptible) model given by,

$$\frac{dS}{dt} = bN - \beta SI - dS + \gamma I$$
$$\frac{dI}{dt} = \beta SI - (d + \alpha + \gamma)I.$$

A CASE STUDY: CONTROLLING RABBITS IN AUSTRALIA

Prior to European settlement, there were no rabbits in Australia (see chapter references). Initially bred for food, their numbers stayed low until a small number were released for hunting purposes on an estate. Within 10 years rabbit numbers were well into the millions.

How might we model this initial population growth? During this early period, we might actually suppose that our very first population growth model provided a good approximation of rabbit dynamics,

$$\frac{dN}{dt} = bN - dN = rN$$

predicting exponential growth for $b > d$ (births greater than deaths). We previously criticised this model for not having a carrying capacity, but for populations of rabbits in the hundreds or thousands with the whole continent of Australia to exploit, we might argue that in fact there wasn't much limiting their growth.

A number of control strategies were attempted through the late 19th- and early 20th-centuries targeting rabbits. In 1950, the myxoma virus was deliberately released in to the rabbit population. From a modelling perspective, the effect of this is to transform the system to the SIS model we introduced above, that is,

$$\frac{dS}{dt} = bN - \beta SI - dS + \gamma I$$
$$\frac{dI}{dt} = \beta SI - (d + \alpha + \gamma)I.$$

We want to use this model to answer a simple question: will the introduction of the disease successfully control rabbit numbers? Initially we had exponential growth. What we then need to know is whether introducing the disease can limit this growth, perhaps to an equilibrium or even complete eradication.

We can answer this question by analysing our model through our usual methods. First, what are the equilibria of this model? There is an extinction equilibrium at $(S, I) = (0, 0)$, but this is quickly shown to be always unstable. Also, if $I = 0$ then we return to exponential growth of the susceptible rabbit population when $b > d$. But there is also an endemic equilibrium, at,

$$S^* = \frac{d + \alpha + \gamma}{\beta},$$
$$I^* = \frac{(b - d)S^*}{\beta S^* - b - \gamma} = \frac{(b - d)(d + \alpha + \gamma)}{\beta(\alpha - (b - d))}.$$

To determine whether this equilibrium is stable, and so whether the disease can indeed control the rabbit population to this equilibrium level, we need to look at the Jacobian.

Exercises

Write out the general Jacobian for this system.

Click for solution

Taking the partial derivatives of the ODEs and putting them in the Jacobian matrix we get,

$$J = \begin{pmatrix} b - d - \beta I^* & b - \beta S^* + \gamma \\ \beta I^* & \beta S^* - (\alpha + d + \gamma) \end{pmatrix}.$$

Substituting in our equilibrium values for S^* and I^* we find that,

$$J = \begin{pmatrix} b - d - \frac{(b-d)(d+\alpha+\gamma)}{(\alpha-(b-d))} & b - d - \alpha \\ \frac{(b-d)(d+\alpha+\gamma)}{(\alpha-(b-d))} & 0 \end{pmatrix}$$

From here we can find that $\det = (b - d)(d + \alpha + \gamma)$, which is positive provided $b > d$. Stability will therefore depend on the trace. With some re-arranging, we have $tr = (b - d)(b + \gamma)/[(b - d) - \alpha]$. Since we are assuming $b > d$ then the numerator is positive. For the fixed point to be stable we require $tr < 0$ which therefore requires $\alpha > (b - d)$ in the denominator. So, if the rate of virulence (infection-

induced mortality) is greater than the rate of intrinsic population growth, the endemic equilibrium is stable. Otherwise it is unstable and we would expect exponential growth.

So, what can we conclude here? That for an exponentially growing population of rabbits, provided the disease is sufficiently virulent, introducing the myxoma virus will control the rabbit population to an equilibrium level. We have therefore shown that an infectious disease can be used as an effective control mechanism against biological invasions.

A FREE-LIVING PARASITE MODEL

There are even more different additional assumptions we may look to include in a model representing infectious disease spread in wildlife populations. We'll now look at a case where we assume that infection is not simply through direct contact between susceptible and infected hosts, but where the parasite is *free-living* in the environment, and infection occurs when susceptible hosts pick up these free-living stages. Letting P be the density of free-living parasites, one way to model this would be as follows,

$$\frac{dS}{dt} = bS - dS - \beta SP$$
$$\frac{dI}{dt} = \beta SP - (d+\alpha)I$$
$$\frac{dP}{dt} = \phi I - \mu P.$$

The key differences to our previous models are:

- The transmission term βSP now requires contact between susceptible hosts and free-living parasite stages for infection to occur.
- Free-living parasites are produced at a constant rate by infected hosts.
- Free-living parasites die at rate μ, and we assume the loss of parasites due to infection can be ignored.

As ever with modelling, many of these processes could be formed differently or have different underlying assumptions, but this gives us a model we can work with.

What equilibria does this model give us? We will have a trivial equilibrium of nothing present, and since we again do not include density-dependence in the susceptible host's dynamics we will not have a disease-free equilibrium (if the parasite is not present, meaning $I = P = 0$, then the host population increases without bound). This just leaves an endemic equilibrium.

Exercises

Find the endemic equilibrium densities of S^*, I^* and P^* for this system.

Click for solution

Using the first ODE we can take out X as a factor, which tells us,

$$a - b - \beta P = 0$$
$$\implies P^* = \frac{a-b}{\beta}.$$

Then from the third ODE we have,

$$\phi I = \mu P$$
$$\implies I^* = \frac{\mu P^*}{\phi}$$
$$\implies I^* = \frac{\mu(a-b)}{\phi\beta}$$

Finally we use the second ODE to find an expression for S^*,

$$\beta SP - (\alpha + b)I = 0$$
$$\implies S^* = \frac{(\alpha+b)I^*}{\beta P^*}$$
$$\implies S^* = \frac{(\alpha+b)\mu}{\phi\beta}.$$

What can we say about stability here? We still need to use the Jacobian to assess stability, but it is now a 3×3 matrix,

$$J = \begin{pmatrix} 0 & 0 & -\beta S^* \\ \beta P^* & -\alpha - b & \beta S^* \\ 0 & \phi & -\mu \end{pmatrix}.$$

We have not covered how to assess stability in a three-dimensional system in any case so far but you can read about it in the background review chapter on linear stability analysis. As a brief summary here, the general rule for stability is that if all the eigenvalues from the Jacobian matrix are negative an equilibrium is stable, and if any eigenvalues are positive the equilibrium is unstable (or a saddle). We have seen that for two-dimensional systems we can sometimes find these eigenavlues directly, and in more complicated cases we use the trace-determinant condition. In fact, this latter condition is a special case of stability criteria called *Routh-Hurwitz conditions*. This is what we need to use here. Often this can be very messy and hard to get meaningful results out of, but sometimes we are lucky and things fall out easily. This case is ... in-between.

Firstly we write out the *characteristic equation* given by $\det(J - \lambda I_3) = 0$ where I_3 is the 3×3 identitiy matrix. This gives,

$$\det(J - \lambda I_3) = \lambda\left[(\alpha + b + \lambda)(\mu + \lambda) - \beta S^* \phi\right] + \beta S^* \beta P^* \phi = 0.$$

To use the Routh-Hurwitz criteria we need to write this out in the form $\lambda^3 + a_2\lambda^2 + a_1\lambda + a_0$. For our equation this gives,

$$\begin{aligned}\det(J-\lambda I_3) =&\lambda^3+\\ &\lambda^2[\alpha+b+\mu]+\\ &\lambda[(\alpha+b)\mu-\beta S^*\phi]+\\ &\beta S^*\beta P^*\phi=0.\end{aligned}$$

The Routh-Hurwitz criteria tell us that for a 3×3 system, stability requires, $a_2, a_1, a_0 > 0$ and $a_2 a_1 > a_0$. Looking at our equation, we can see that a_2 and a_0 are definitely positive. However, if we substitute in the value for S^* we find that $a_1 = 0$. This same result also means that we never satsify $a_2 a_1 > a_0$, meaning that this equilibrium is never stable.

So if the endemic equilibrium is never stable, what does happen in this system? It would be surprising if the disease never persisted and we only ever had exponential growth of hosts. It is difficult to prove mathematically, but the answer is that the system exhibits limit cycles (like we saw in the predator-prey model). An example of these cycles are shown in the plot below, showing really quite sharp and extreme cycles, with rapid outbreaks of infection followed by collapse to near-extinction. This change occurs because the free-living parasite has introduced a delay to the infection process. The rate of infection experienced by a susceptible no longer depends on the current infected density but on what that density was a few time-steps ago when it was producing the free-living stages. In general, introducing delays into models has the tendency to cause cycles.

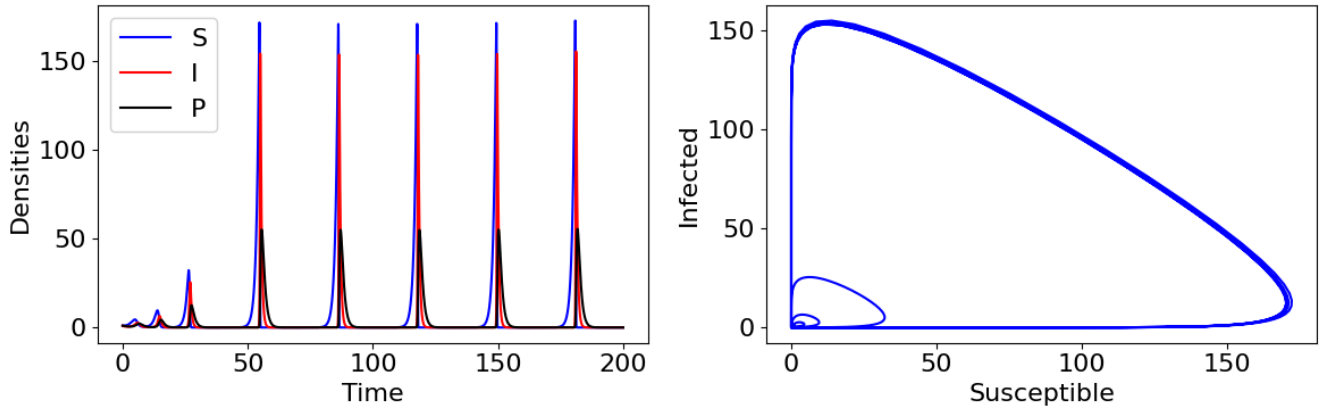


Figure: Time-course and phase portrait from the free-living parasite model. Parameter values: $a=2, b=1, \alpha=1, \beta=1, \phi=1, \mu=1$ *with initial conditions* $S(0)=1, I(0)=1, P(0)=1$.

Explore the model

Use the Python code below to explore the free-living parasite model. This will produce plots of the time-courses and a phase-portrait for S and I, showing very large cycles. Using the parameter values, calculate what the equilibrium values would be – how much larger do the densities grow on the cycles than the equilibria? Try changing a few parameters to see how the nature of the cycles change.

You may well find that if you change parameter values the model begins to run slowly or even stops entirely and sends an error message. This is an example of a *stiff* ODE system. Roughly, this is a system where the dynamics have two very different time-scales – for example in this case the dynamics spend a good deal of time with densities close to 0 and not changing, followed by very rapid increases – and standard numerical ODE solvers often struggle to run such models. Approaches to overcome this include using an alternative solver (you can add the optional argument, `method='Radau'`

to the end of the `solve_ivp` function to do this here) or *log-transforming* the model to take new variables $\bar{S} = \ln S$, etc, and solving for the transformed model of $d\bar{S}/dt$, etc.

Click for code

```
# Import the necessary libraries
import numpy as np
import matplotlib.pyplot as plt
from scipy.integrate import solve_ivp
# Options to make the plots the right size
plt.rcParams['figure.figsize'] = [12, 4]
plt.rcParams.update({'font.size': 16})
# Function for dynamics called 'freeliving'
def freeliving(t,x):
    # Rename variables for ease
    S=x[0]
    I=x[1]
    P=x[2]
    # The ODEs
    dS = a*S-b*S-beta*S*P
    dI = beta*S*P-(alpha+b)*I
    dP = phi*I-mu*P
    return [dS,dI,dP]
# Parameter values
a=2
b=1
alpha=1
mu=1
beta=1
phi=1
# Initial conditions
S0=1
I0=1
P0=1
N0=[S0,I0,P0]
# Time points to use
tc = np.linspace(0, 200, 10000)
# Run model using 'solve_ivp'
Nc = solve_ivp(freeliving, [tc[0],tc[-1]],N0, t_eval=tc)
```

```
# Plotting code
fig, (ax1, ax2) = plt.subplots(1, 2)
ax1.plot(tc, Nc.y[0], "r", label="S")
ax1.plot(tc, Nc.y[1], "k", label="I")
ax1.plot(tc, Nc.y[2], "b", label="P")
ax1.set(xlabel='Time', ylabel='Concentrations')
ax1.legend()
ax2.plot(Nc.y[0],Nc.y[1],'b')
ax2.set(xlabel='S', ylabel='I')
```

Key Takeaways

- If we model disease in ecological populations we need to make some changes to our model, for example including disease-induced deaths.
- If a disease is sufficiently deadly, it can be used as a biological control of a pest species.
- If disease is caused by a free-living paraiste stage, this can lead to cycles of infection.

Chapter references

- The rabbit-myxoma model is based on an example in a paper by Anderson and May (1982).
- The free-living parasite model is based on an example in the paper by Anderson and May (1981).
- The content in the *Infectious diseases* section was influenced by the textbooks, *Mathematical Biology* by Murray and *Modelling Infectious Diseases in Humans and Animals* by Keeling & Rohani.

CHAPTER 7

Evolution and adaptive dynamics

INTRODUCING EVOLUTION

This chapter could equally well belong in the population ecology section, or indeed a section of its own, but as I will focus on applications of evolution to disease systems we will examine it here.

In the models we have looked at so far, all individuals within a population (or compartment of a population) are identical – they have the same birth rates, death rates, risk of infection and so on. However, we often see mutations arise within populations, leading to offspring – and therefore a subpopulation – with slightly different rates. These two *strains* of the organism – the *resident* and the *invader* – will then compete with one another, with potentially one strain dominating (such a scenario could be applied to any sort of invading species, but our focus will be on mutants generated through reproduction). This gives rise to the potential for *evolution*.

The basic idea of natural selection is that strains of an organism that are the most successful – that is they make the best use of their resources, live the longest and, ultimately, produce the most offspring – are likely to be better represented in future generations. Evolution is the repeated process by which new mutants are generated and these most successful strains come to dominate. We often think of evolution as an immensely slow process – for example the evolution of humans from primates – but in large populations with short generation times, microbial populations for example, it can be witnessed in experiments lasting just a few days.

Like most biological processes, evolution is fundamentally complicated. An important element of mathematical modelling is deciding what assumptions we need to include, and what we will neglect. Broadly, evolutionary models take one of two approaches. One set of models focus on the genetics behind evolution. Mutation is ultimately a genetic process, and it is the genes that determine the behaviour of organisms. However, these models can have little to say about what it is that drives evolution. The other set of models focus on evolutionary ecology; that is, how the ecology (population dynamics) drives evolution (adaptive dynamics). This is because it is the ecological environment that creates selection that favours different strains. In these models we often assume very simple genetics, but are able to gain insight in to when we might expect organisms to evolve in a certain way. It is this second approach that we shall be introducing here, and in particular an approach developed through the 1990s called *adaptive dynamics* or more generally *evolutionary invasion analysis*.

RESIDENT DYNAMICS

To give us some focus, we will think about evolution in the context of infectious disease (this is my own area of research, so I had to really). In particular we will consider how a parasite might evolve to best exploit its host. We start by writing down the population dynamics of the resident strain in the absence of the mutant. We will take a variation of our wildlife disease model:

$$\frac{dS}{dt} = (b - qN)S - dS - \beta SI$$
$$\frac{dI}{dt} = \beta SI - (d + \alpha)I,$$

where $N = S + I$ is the total population size. Notice that here reproduction of hosts is reduced due to competition (q), creating an emergent carrying capacity for the population in the absence of disease. We will not go through the full analysis of the resident dynamics here as it is very similar to the wildlife disease model, but you are encouraged to carry this out yourself.

The key points are:

- There is a trivial equilibrium at $S = I = 0$. This equilibrium is never stable and biologically fairly irrelevant.
- There is a disease-free equilbirium at $S_{df} = (b - d)/q, I_{df} = 0$. This is stable for $R_0 = \frac{\beta S_{df}}{d + \alpha} < 1$.
- There is an endemic equilibrium at $S_e = (d + \alpha)/\beta, I_e = (b - qS_e - d)/(\beta + q)$. This is stable for $R_0 = \frac{\beta S_{df}}{d + \alpha} > 1$, which is true whenever $I_e > 0$.

FITNESS

We will assume that an organism, here a parasite, is able to evolve one or more *traits*. Pragmatically, this means that one or more parameters of the model will evolve. The key quantity for analysing evolution is *fitness*. This can be formally defined in slightly different ways, but it is ultimately a measure of how successful a given strain is in its environment. For our model, the fitness of a particular strain will be defined as its exponential growth rate.

Let us assume that initially all individuals in the population are the resident strain. We will then try and determine the fitness of a mutant strain, which has a small difference in one or more traits (we will usually assume two traits vary as we expect there to be a *trade-off*: if a strain is 'stronger' in one area we expect it to pay for this by being 'weaker' in another. However, these two traits will be linked by some defined trade-off function, meaning we can ultimately reduce it down to thinking about what happens to just one trait). We will make three key assumptions:

1. Mutants arise very rarely, meaning we can assume that the resident has reached an equilibrium of its population dynamics. This is called a *separation of timescales* assumption; the population dynamics are 'fast' and the adaptive dynamics are 'slow'.
2. The mutant is initially rare, meaning it has a very small density (initially) relative to the resident.
3. The mutant is very similar to the resident, with just a small difference in the value of the parameter(s) that evolve.

The mutant's growth rate depends both on its own strategy and on the 'environment', by which we mean the ecological conditions created by the current resident. Let us say that the current resident has some strategy x. We can write the fitness of this resident – that is its growth rate – as $r(x, E_x)$, where E_x represents the environment. In this case we must have $r = 0$, because we have assumed the resident is at equilibrium, so it should be neither growing nor shrinking.

Now consider a mutant type with strategy x_m invading a resident environment. The fitness of this mutant is

$r(x_m, E_x)$. In all likelihood $r \neq 0$. If $r > 0$ then the mutant population will grow, meaning the mutant can invade the current resident. In that case it may coexist with the old resident, or more commonly will completely replace it. If $r < 0$ it will die out.

Remember we assumed mutants are very similar to the resident (so the mutation size is small). This means we can take a *Taylor expansion* of the mutant's fitness up to the linear term as,

$$r(x_m, E_x) = r(x, E_x) + \left.\frac{\partial r}{\partial x_m}\right|_{x_m=x} (x_m - x) + \dots$$

Since $r(x, E_x) = 0$, the direction of selection – whether it is positive or negative – is entirely dictated by the fitness gradient, $[\partial r/\partial x_m]_{x_m=x}$. If the mutant has a trait value $x_m > x$ then for that mutant to invade there must be a positive fitness gradient.

We can picture there being a sequence of these invasion events by mutants. We start with resident type A, and assume a mutant of type B arises. Where that mutant is successful it will usually go on to replace the type A population and now the resident population is type B. Then a new mutant of type C arises and attempts to invade, and so on. Given small mutations, we can now think of the dynamics of the evolving trait as being a dynamical system in its own right; the *adaptive dynamics*. If we define a (rather vague) evolutionary timescale, $T >> t$ (with t the ecological timescale), then the change in the trait value x will be given by,

$$\frac{dx}{dT} = \mu \left.\frac{\partial r}{\partial x_m}\right|_{x_m=x}.$$

The parameter $\mu > 0$ contains information about how quickly new mutants appear (it is beyond the scope of this course to consider this in detail, so we shall just assume it is a constant). Since the fitness gradient will depend on the resident equilibrium, you might begin to see how the separation of timescales assumption works: we first have to solve the population dynamics to its equilibrium, then solve the adaptive dynamics. Note that μ only scales the 'speed' of evolution. Therefore if only one organism is evolving (so relative speeds do not matter) we will generally only need to focus on the fitness gradient to determine the long-term evolutionary behaviour.

Given this definition, we might then expect to look for 'equilibria' of these adaptive dynamics, like we have in all of our population models so far. That is, we would expect evolution of the trait x to continue until a point is reached where the fitness gradient is zero. This is indeed what we do, but things are a little more complicated and there is some new terminology associated with it. Firstly, an 'equilibrium' of adaptive dynamics is known as a *singular strategy*,

$$\frac{dx}{dT} \propto \left.\frac{\partial r}{\partial x_m}\right|_{x_m=x} = 0.$$

What can happen at such a singular strategy? Perhaps confusingly, there are now two different types of stability we need to consider.

Evolutionary stability

Perhaps even more confusingly, our first term, evolutionary stability, is not really 'stability' in a traditional mathematical sense at all. Instead it relates to a long-standing ecological idea of the *Evolutionarily Stable Strategy (ESS)*. In fact this term asks, if the strategy is adopted by the current resident, can it be invaded by any (nearby) mutants? Mathematically, this is given by,

$$\left.\frac{\partial^2 r}{\partial x_m^2}\right|_{x_m=x} < 0$$

You might spot that the 'peak' or 'trough' will always be at $r = 0$, so all nearby mutants either have positive or negative fitness. In other words, it asks whether the strategy is a local fitness maximum or not.

Convergence stability

However, just because a strategy is uninvadable that does not mean that it will be reached through the evolutionary process of mutations and invasions. Instead we need to assess whether the strategy is locally attracting. This term is called convergence stability and is a more direct counterpart of classic stability.

$$\frac{d}{dx}\left(\frac{\partial r}{\partial x_m}\Big|_{x_m=x}\right) = \frac{\partial^2 r}{\partial x_m^2}\Big|_{x_m=x} + \frac{\partial^2 r}{\partial x_m \partial x}\Big|_{x_m=x} < 0$$

Note that this is the sum of the evolutionary stability condition and an additional second derivative.

Evolutionary stability and convergence stability are independent properties. This therefore provides four potential outcomes at an evolutionary singular point:

Evol. Stable	Conv. Stable	Outcome
✓	✓	Continuously Stable Strategy (CSS)
×	×	Repeller
✓	×	Garden of Eden
×	✓	Evolutionary Branching Point

- A *Continuously Stable Strategy (CSS)* is both locally attracting and uninvadable. It is therefore a long-term end-point of evolution.
- A *repeller* is neither attracting nor uninvadable. Therefore evolution will always take the system away from this strategy.
- A *Garden of Eden* is uninvadable and so would be a local fitness maximum were it ever reached. However, since it is not attracting, evolution will always take the system away from this strategy. In that sense it usually behaves roughly the same as a repeller.
- An *evolutionary branching point* is locally attracting but invadable, that is a fitness minimum. In this case evolution will drive the population to this strategy, but once there all nearby mutants can invade. What happens then is that the population divides in to two coexisting strategies either side of the singular strategy (subject to a couple of extra assumptions that we'll ignore here), and evolution continues. This is an important outcome since it creates diversity directly from the evolutionary process.

THE EVOLUTION OF PARASITE VIRULENCE

Let us take these theoretical ideas and apply them to our infectious disease model to see what it all really means. We will assume here that it is the parasite that is able to evolve, and in particular that it can evolve to alter the amount of transmission, β. We might well expect a disease to 'want' to infect and spread as quickly as possible (though we should do our best not to be quite so anthropomorphic). A classic assumption in the theoretical literature is that there is a transmission-virulence (mortality) trade-off for parasites. The argument is that for a parasite to increase its transmission rate it must grow more quickly, and this faster growth is likely to cause greater damage and hence greater virulence. As such we might define a trade-off function $\alpha = \alpha(\beta)$ which is an increasing function.

We shall assume that we have parameter values such that $R_0 > 1$ and the endemic equilibrium is stable. How do we define the fitness for the parasite? In this model we do not track explicitly the number of parasites, so we don't have a growth rate for them as such. However, we do track the density of host individuals infected

with parasites, and we can argue it is reasonable to say that this number of infected hosts is a good measure of the growth and success of the parasite, and therefore we will take the growth rate of infected hosts to be the fitness. We then need to write out the dynamics of the mutant parasite, I_m, interacting with a population of resident hosts and parasites. These are given by,

$$\frac{dI_m}{dt} = \beta_m S_e I_m - (\alpha(\beta_m) + d) I_m.$$

If we had any terms with mutant densities higher than first order we could set these to be 0, since the mutant is assumed to be rare. As we can see, that is not the case here (but it often will be). As such to find the fitness for the mutant parasite, i.e. its growth rate, we can write $dI_m/dt = rI_m$, where r is the growth rate given by,

$$r = \beta_m S_e - (\alpha(\beta_m) + d)$$

The next step is to find the fitness gradient. We take the derivative of r with respect to β_m and then substitute in $\beta_m = \beta$ (this substitution is about having small mutation steps). This fitness gradient gives,

$$\frac{\partial r}{\partial \beta_m}\Big|_{\beta_m=\beta} = S_e - \alpha'(\beta)$$

If this fitness gradient is positive the parasite will increase the transmission rate, and hence also increase its virulence; if the fitness gradient is negative the parasite will reduce the transmission rate, and hence also decrease the virulence. Which occurs depends on the steepness of the transmission-virulence trade-off $\alpha(\beta)$. Where the two terms are equal we will have a singular strategy.

Let us assume we are at a singular point, so that must mean $S_e = \alpha'(\beta)$. What are the stability conditions at this singular point? Firstly we have evolutionary stability,

$$\frac{\partial^2 r}{\partial \beta_m^2}\Big|_{\beta_m=\beta} = -\alpha''(\beta)$$

So we have that the singular strategy will be evolutionarily stable whenever the curvature (second derivative) of the trade-off at the singular strategy is positive – this means that increased transmission is 'increasingly costly' in terms of virulence. For convergence stability we have the sum of the last term and,

$$\frac{\partial^2 r}{\partial \beta_m \beta}\Big|_{\beta_m=\beta} = \frac{dS_e}{d\beta}.$$

What does this term evaluate to? Using the expression we found for S_e, we have,

$$\begin{aligned}\frac{dS_e}{d\beta} &= \frac{\alpha'(\beta)}{\beta} - \frac{\alpha(\beta)+d}{\beta^2} \\ &= \frac{1}{\beta}\left(\alpha'(\beta) - \frac{\alpha(\beta)+d}{\beta}\right) \\ &= \frac{1}{\beta}(\alpha'(\beta) - S_e)\end{aligned}$$

Now recall that we are evaluating this second derivative at the singular strategy, and we found above that at the singular strategy we have that $S_e - \alpha'(\beta) = 0$. Therefore the term above also evaluates to zero. The final convergence stability term, then, is identical to the evolutionary stability term,

$$\frac{\partial^2 r}{\partial \beta_m^2}\Big|_{\beta_m=\beta} + \frac{\partial^2 r}{\partial \beta_m \partial \beta}\Big|_{\beta_m=\beta} = -\alpha''(\beta)$$

So what does this tell us? That whenever the curvature of the trade-off is positive, then the singular strategy is both evolutionarily stable and convergence stable, and therefore a CSS. In contrast, whenever the curvature is negative, then the singular strategy is a repeller. Therefore we can never get evolutionary branching in this model (but will happen in many other models).

HOW DOES OPTIMAL PARASITE VIRULENCE VARY?

Now let's go to a bit more detail. Let us assume we have chosen parameters and a trade-off such that we predict a singular strategy that is a Continuously Stable Strategy (CSS). Now let us ask what would happen if we changed the system by increasing the host's birth rate, b. What would we expect to happen? Should the parasite increase or decrease its transmission (and therefore virulence)?

The answer, in fact, is neither. Changing the host's birth rate will have no effect on parasite evolution. This may initially seem puzzling. After all, we have changed the environment (we would presume by increasing the size of the host population), and we know that the host birth rate appears in the population dynamics. Let's look at the parasite's fitness gradient, which is the driver of evolution,

$$\frac{\partial r}{\partial \beta_m}\Big|_{\beta_m=\beta} = S_e - \alpha'(\beta) = \frac{\alpha(\beta)+d}{\beta} - \alpha'(\beta).$$

As we can see, the host birth rate does not appear anywhere in the fitness gradient, and this is why it does not impact parasite evolution. Parasite evolution is simply determined by how infective the mutant is compared to the resident and how much more quickly it dies. We might have expected increasing the birth rate to lead to more infections since there are more hosts. However, it is only the equilibrium density of susceptible hosts that matters, and this value is actually not impacted by host births at all.

Now let us ask how parasite evolution might depend on the (natural) death rate, d. As we can see, this parameter does appear in the parasite's fitness since it partly controls the susceptible density. Let's make our lives easier by choosing a functional form for the trade-off as,

$$\alpha(\beta) = \alpha_{\min} + \beta^2$$

(This is actually not a very good choice of trade-off form in general – think what happens to its gradient when $\beta = 0$ – but it is good enough for now). Since $\alpha'(\beta) > 0$ we know from above that this will produce a CSS (both evolutionarily stable and convergence stable). The fitness now becomes,

$$\frac{\partial r}{\partial \beta_m}\Big|_{\beta_m=\beta} = S_e - \alpha'(\beta) = \frac{\alpha_{\min}+\beta^2+d}{\beta} - 2\beta.$$

We can solve this to find the singular strategy,

$$\frac{\partial r}{\partial \beta_m}\Big|_{\beta_m=\beta} = 0$$

$$\implies \frac{\alpha_{\min}+\beta^2+d}{\beta} - 2\beta = 0$$

$$\implies \beta = \sqrt{\alpha_{\min}+d}$$

Therefore the parasite's evolutionarily stable transmission rate is an increasing function of the host's death rate. In other words, we expect the parasite to evolve higher transmission – and therefore higher virulence – against hosts that die more quickly.

Why might this be? In general, what might we expect the parasite's optimum strategy to be? Recall we defined the term R_0 to be a measure of how infective the parasite is. We would expect the parasite to evolve to be as infectious as possible, and therefore to maximise R_0. A knock-on effect of this is that we should therefore expect the parasite to minimise the susceptible population, as this would mean it is exploiting the host to the best of its ability. Remember that $S_e = S_{df}/R_0$, clearly highlighting this link.

Looking at our equilibrium expression, we see that increasing the death rate, d, leads to a bigger susceptible density, S_e (and so a lower R_0). Since the parasite's optimum strategy should be to minimise S_e (maximise R_0), we would therefore expect evolution to cause the parasite to exploit the host more. That means evolving higher transmission, and by the trade-off, higher virulence. Note that this means than when hosts die more quickly due to natural causes, the parasite is evolving to increase the disease-induced mortality as well.

Key Takeaways

- We can model evolution by thinking about repeated invasions of rare mutants into resident populations.
- At singular points there are two types of stability to check: evolutionary stability and convergence stability.
- Parasites will evolve higher transmission and virulence when the host death rate increases.

Chapter references

- The theoretical underpinnings are developed from the paper by Geritz et al. (1998).

PART III

IMMUNE AND CELL DYNAMICS

CHAPTER 8

A within-host Covid-19 model

A MODEL FOR COVID-19

In the last few chapters we were thinking about the dynamics of disease at the scale of human or ecological populations. Often, though, we will be more concerned with the health of individual patients, particularly from a medical perspective. In these cases, we want to know how a virus or bacteria grows and develops inside the body, and how our immune system interacts with it. To study these dynamics we therefore require a *within-host model*. As we move down in biological scales, for the biology novice some of the terminology and ideas will now start to feel less familiar, but it is important to remember that we are still dealing with populations. But whereas we previously thought of populations as groups of humans or animals, we are now looking at populations of cells or virus particles.

In our first example we will look at a relatively simple model of virus-cell dynamics that loosely describes the interaction between Covid-19 virus particles and a type of human immune cell called the T-cell (see chapter references). We assume that the virus grows in the body according to classic logistic growth that we met back in the first chapter, that it decays at some background rate, but that it is also killed by T-cells. T-cells themselves are produced by the body at some constant, background rate, but also *proliferate* when they sense virus is present. Finally the T-cells also decay. We can write down our model as follows,

$$\frac{dV}{dt} = pV(1 - V/K) - \beta VT - \mu V$$
$$\frac{dT}{dt} = s + rT\left(\frac{V^2}{V^2 + c^2}\right) - \delta T$$

As usual, all the parameters are positive and we will also assume that $r > \delta$. There are a few things to notice about this model:

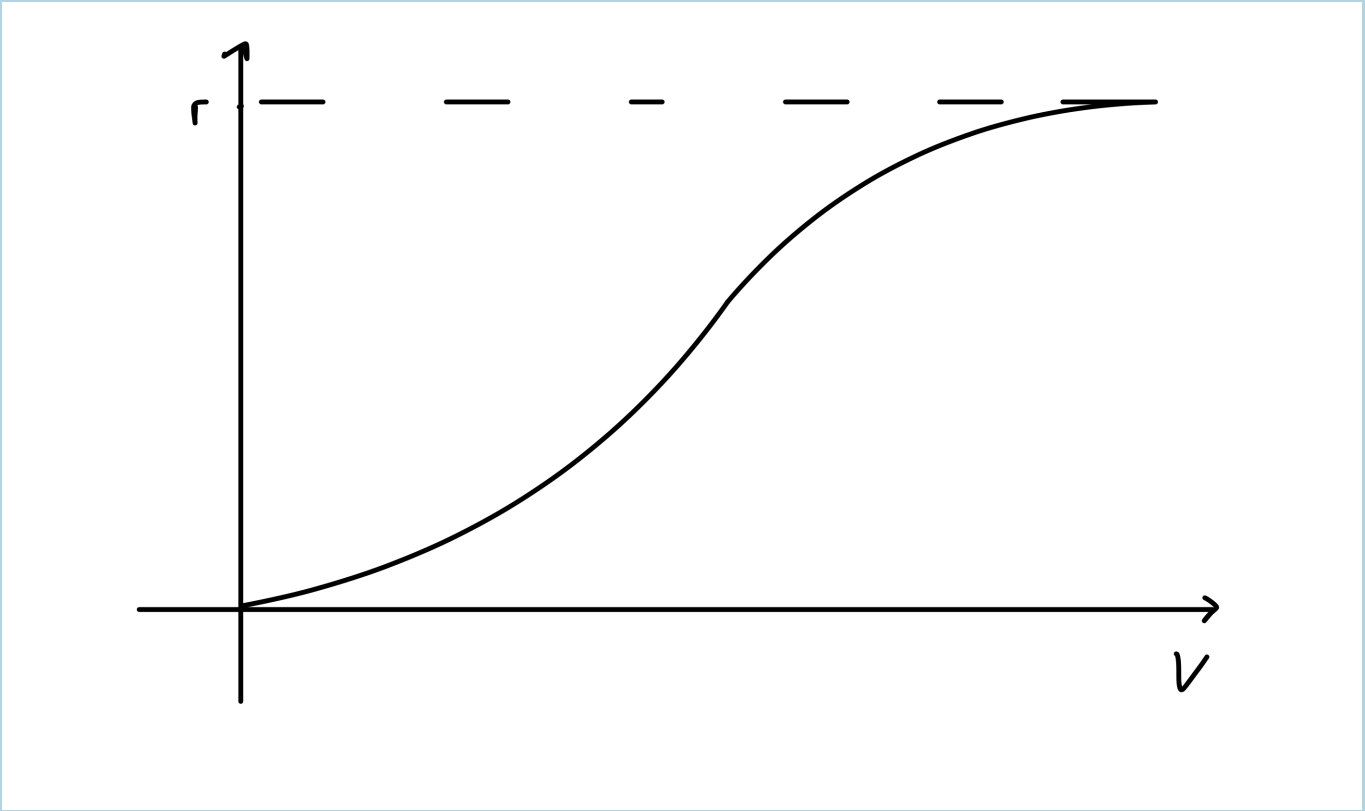
- When a virus is killed by a T-cell, there is no corresponding killing of the cell itself.
- T-cell production is constant, not a per-capita rate.
- The proliferation function is sigmoidal – while virus density is small little proliferation occurs, but past a threshold value it rapidly increases. Quite a lot of the analysis is going to depend on knowledge of what this proliferation function looks like.

Exercises

Sketch the curve $r\left(\frac{V^2}{V^2 + c^2}\right)$ and note down its key features.

Click for solution
The curve is sketched below. The key points to note are that:

- It starts at 0.
- It is strictly increasing.
- It tends towards r.
- It has a sigmoidal shape.



As you will be getting used to, we will first consider the equilibria of this system. If we factorise the first equation we find two possible ways to make it equal zero:

- $V = 0$
- $p(1 - V/K) - \beta T - \mu = 0 \implies T = (p(1 - V/K) - \mu)/\beta$

Substituting $V = 0$ into the second equation, we can find that this requires $T = s/\delta$ for an equilibrium. This will be the virus-free equilibrium, with no virus and T-cells present at their background density.

For the second equilibrium we can see that the result of making the substitution of T into our second ODE is likely to leave something quite complicated. Indeed, there may even be more than one equilibrium produced. To save ourselves a lot of tedious algebra let us instead examine the phase portrait.

A GRAPHICAL ANALYSIS

We have a two-dimensional system and both variables must be non-negative to make biological sense. We can use the workings we used when finding the equilibria to help us find the nullclines. The nullclines from the first equation give us:

- $V = 0$ – a straight vertical line.
- $T = p(1 - V/K) - \mu)/\beta$ – a decreasing straight line starting at $(p - \mu)/\beta$ and crossing 0 at $V = K(1 - \mu/\beta)$.

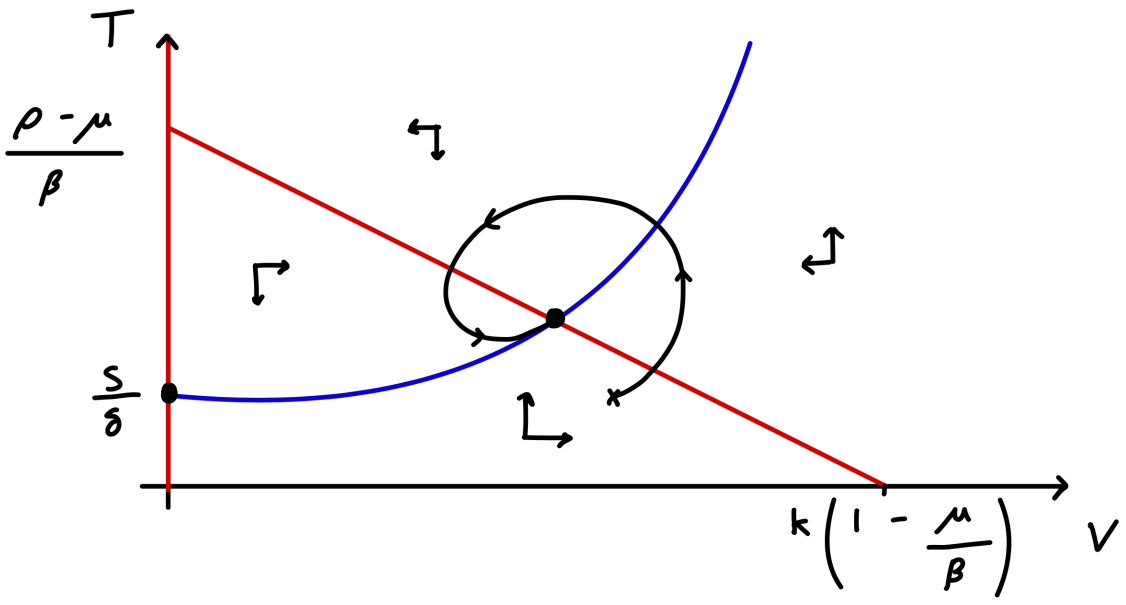
After a couple of lines of working we can find the single nullcline from the second equation to be,

- $$T = \frac{s}{\delta - r\left(\frac{V^2}{V^2+c^2}\right)}.$$

What does this nullcline look like? It starts at $T = s/\delta$ when $V = 0$ and as V gets very large we find $T \to s/(\delta - r)$. In between these extremes we have this sigmoidal proliferation function to worry about. Note that $V^2/(V^2 + c^2)$ varies from 0 to 1 and is strictly increasing and that $r > \delta$. This tells us two things:

1. The nullcline is strictly increasing.
2. At some point the denominator will pass through zero (some further work finds that techncially this happens twice, but one of the values occurs for $V < 0$). After this happens we will have $T < 0$.

One additional thing to note is that the term, $V^2/(V^2 + c^2)$ remains small up to reasonably high values of V. Therefore this term will initially not effect the nullcline much as V increases, meaning our nullcline is initially quite flat. We can put all of this together to form our phase portrait.



Phase portrait of the within-host Covid model. Blue lines represent the virus nullclines, and red lines the cell nullclines, with the black curve giving an example trajectory.

What we see, then, is that there is only one equilibrium when the virus is present, and for the assumptions we have made about relative parameter values, it appears to be a stable spiral. We can additionally note that the initial dynamics are for a rapid rise in virus concentration but little change in T-cell density, before the virus becomes more prevalent and T-cell proliferation begins. We can also note that the virus density at the equilibrium is considerably lower than it would be without T-cells doing their thing. Examining the phase portrait, we might also conclude that the virus-free equilibrium will only become stable if we can move the T nullcline far enough upwards that the two non-zero nullclines no longer intersect. This would require $s/\delta > (p-\mu)/\beta$. Let us see if we can confirm that conclusion with linear stability analysis.

LINEAR STABILITY ANALYSIS

The general Jacobian for the system is given by,

$$J = \begin{pmatrix} p - \frac{2pV^*}{K} - \beta T^* - \mu & -\beta V^* \\ rT^* \frac{2V^* c^2}{(V^{*2}+c^2)^2} & r\frac{V^{*2}}{V^{*2}+c^2} - \delta \end{pmatrix}$$

Virus-free equilibrium

Since $V^* = 0$ both the top-right and bottom-left entries are zero, meaning we can read off the eigenvalues as the two terms on the main diagonal:

- $\lambda_1 = p - \beta T^* - \mu$
- $\lambda_2 = -\delta$

The second eigenvalue is negative, so everything depends on the first. Re-arranging we require $T^* > (p-\mu)/\beta$. Recall that for the virus free equilibrium we have $T^* = s/\delta$. Therefore this equilibrium is only stable when $s/\delta > (p-\mu)/\beta$, exactly as we surmised was the case from our phase portrait.

Virus-present equilibrium

You will recall that we never wrote down an expression for this equilibrium. You might remember from previous chapters, however, that this does not usually stop us making conclusions about its stability. We know from the phase portrait that there is only one equilibrium to worry about. We also know from the equilibrium conditions that:

- $p(1 - V^*/K) - \beta T - \mu = 0$
- $rV^{*2}/(V^{*2} + c^2) - \delta = -s/T^*$.

Substituting these values in to the Jacobian we now find,

$$J = \begin{pmatrix} -\frac{pV^*}{K} & -\beta V^* \\ rT^* \frac{2V^* c^2}{(V^{*2}+c^2)^2} & -s/T^* \end{pmatrix}$$

Now we must assess stability based on the trade and determinant conditions. These are:

$$tr(J) = -\frac{pV^*}{K} - \frac{s}{T^*} < 0$$

$$\det(J) = \frac{psV^*}{KT^*} + r\beta V^* T^* \left(\frac{2V^* c^2}{V*^2 + c^2} \right) > 0$$

This means that whenever the equilibrium at $V^*, T^* > 0$ exists, it must be stable. And we know from the phase portrait that the condition for it to exist is the exact opposite of the condition for stability of the virus-free equilibrium.

Key Takeaways

- We can build models of cell-virus interactions in much the same way as we did for interactions between whole organisms.
- The virus will reach a non-zero equilibrium provided it grows quickly and there is limited killing by T cells.
- The sigmoidal proliferation rate makes T cells initially slow to respond to an infection.

Chapter references

- This chapter is based on a model from Hernandez-Vargas & Velasco-Hernandez (2020).

CHAPTER 9

A within-host HIV-I model

CASE STUDY: HIV-I

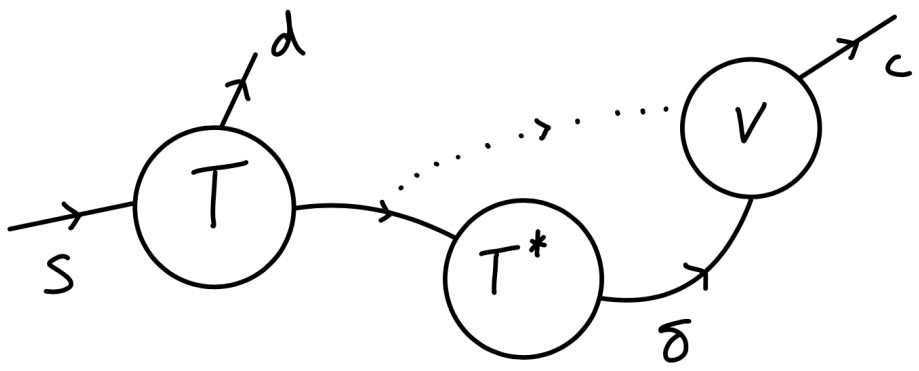
In this chapter we will look at another example of a within-host disease model, and in fact one of the first such models to be developed. Human immunodeficiency virus (HIV) infects immune cells called 'CD4+ T-cells', which form an important part of the human immune system. HIV enters these T-cells, replicates within the cell and then releases new virus particles in to the bloodstream. Immediately after first infection, the virus grows rapidly and produces common infection symptoms in the patient. After a few months, these symptoms disappear and the virus concentration reduces to a lower, but steady, level. This 'asymptomatic period' can last for years, with the virus density staying roughly constant, and the concentration of T-cells very slowly dropping. Eventually, the T-cell density becomes so low that the patient's immune system is no longer effective, a condition called aquired immunodeficiency syndrome (AIDS). The patient is then at risk from life-threatening opportunistic infections. From the 1990s, much work has been done to explore the dynamics of HIV and to produce potential treatments, including the development of mathematical models (see chapter references).

A (PRE-TREATMENT) MODEL

Let's start by establishing our variables and drawing a schematic of what is happening in this system. The variables that we want to keep track of (before treatment, at least), are:

- Healthy T-cells (T)
- Infected T-cells (T^*)
- Virus particles (V)

We will only keep track of the concentration of virus particles in the bloodstream, not the 'intracellular' concentrations. Our schematic should look like this:



Schematic of the HIV-I model.

Healthy T-cells are produced at some constant rate s by the body (not a per-capita rate), but also decay at some rate d. These healthy cells become infected through contact with virus, assuming a mass-action process, with coefficient k. These infected T-cells then decay at some rate δ. New virus particles are produced when infected T-cells die, with N giving the average number of virus particles produced by each cell. The virus then also decays at some rate c. The system can therefore be expressed with the following set of ODEs:

$$\frac{dT}{dt} = s - dT - kVT$$
$$\frac{dT^*}{dt} = kVT - \delta T^*$$
$$\frac{dV}{dt} = N\delta T^* - cV$$

Looking at the equations, we can see that there is an HIV-free equilibrium for $(T, T^*, V) = (s/d, 0, 0)$. We have another example here of a 3-dimensional system. As we saw in the free-living parasite model, assessing stability here can be a bit more complicated, often reying on using the *Routh-Hurwitz criteria*. As it turns out, for the HIV-free equilibrium things simplify fairly nicely to the point that we can actually directly calculate one of the eigenvalues, and then just use the trace-determinant conditions (themselves the Routh-Hurwitz criteria for a 2-dimensional system) to determine the sign of the other two.

The general Jacobian is,

$$J = \begin{pmatrix} -d - kV & 0 & -kT \\ kV & -\delta & kT \\ 0 & N\delta & -c \end{pmatrix}.$$

Subsituting in the equilibrium values this becomes,

$$J = \begin{pmatrix} -d & 0 & -ks/d \\ 0 & -\delta & ks/d \\ 0 & N\delta & -c \end{pmatrix}.$$

To find the eigenvalues we want to write out the *characteristic equation* – the determinant of the matrix of $(J - \lambda I_3)$, where I_3 is the *identity matrix*. The characteristic equation is,

$$(-d - \lambda)\left[(-\delta - \lambda)(-c - \lambda) - \frac{ksN\delta}{d}\right] = 0$$

This is already partly factorised, making our life much easier. The first bracket tells us that one eigenvalue is

$\lambda = -d < 0$. To find the remaining two eigenvalues we look at what is inside the square bracket, wich can be re-written as,

$$\lambda^2 + (c+\delta)\lambda + \delta c - \frac{ksN\delta}{d} = 0$$

We could use the quadratic formula to solve for the eigenvalues explicitly, but we can also note that this is just the characteristic equation for a two dimensional system, with $-(c+\delta)$ being the trace and $\delta c - \frac{ksN\delta}{d}$ being the determinant. We can then see that we definitely have a negative trace, and that the determinant is positive – meaning the equilibrium is stable – if $c > ksN/d$. In other words, the decay rate of the virus must be high, and the infection rate and production rate must be low.

Of course, when a patient is to undergo treatment, they will not be virus-free. We are therefore more interested in the pre-treatment infection steady-state, and we will denote these with a subscript 0. Setting the equations for dT^*/dt and dV/dt to zero, yields,

$$N\delta T_0^* = cV_0 \text{ and } kV_0T_0 = \delta T_0^* \implies T_0 = \frac{c}{Nk}$$

Solving $dT/dt = 0$, then gives,

$$V_0 = \frac{sN}{c} - \frac{d}{k}$$

and, by substituting back in,

$$T_0^* = \frac{kV_0T_0}{\delta}$$

We now have expressions for each of these equilibrium values. We won't prove the stability here as it gets pretty complicated, but it will be stable whenever the infection-free equilibrium is unstable.

TREATMENT

Assume a patient arrives with their virus and T-cell concentrations at the pre-treatment steady state. We then wish to put them on a course of treatment. There are many different treatments for HIV that act in different ways. We will focus on an early treatment strategy of using 'protease inhibitors'. These drugs did not prevent infection, but meant that any new virus produced inside infected T-cells were non-infectious. The system therefore becomes:

$$\frac{dT}{dt} = s - dT - kV_IT$$
$$\frac{dT^*}{dt} = kV_IT - \delta T^*$$
$$\frac{dV_I}{dt} = -cV_I$$
$$\frac{dV_{NI}}{dt} = N\delta T^* - cV_{NI}$$

Initially this looks like we have made our lives even harder since we now have *four* equations to deal with! However, it turns out we can simplify things quite a bit. In particular, in the early stages of treatment it is reasonable for us to assume that the number of healthy T-cells stays roughly constant at its pre-treatment level $T(t) = T_0$. This means dT/dt can be ignored and that T_0 can be considered a constant. We also know that before treatment, all virus particles were of the infective type, such that $V_I(0) = V_0$ and $V_{NI}(0) = 0$

. If we look at the equation for dV_I/dt we can see it is in fact linear, and since we have just stated what its initial value is, we can solve it fairly quickly using separation of variables.

$$V_I(t) = V_0 e^{-ct}$$

Given that $T(t) = T_0$, we now find that dT^*/dt is also now linear,

$$\frac{dT^*}{dt} + \delta T^* = kV_I T_0$$

It may not initally be obvious that this is linear, since the right-hand side had $V_I \times T_0$. However, T_0 is a constant and we already have an expression for $V_I = V_0 e^{-ct}$. This can therefore be solved using an integrating factor. Let's go through the working for this.

Taking an integrating factor of $e^{\delta t}$ we can write this as,

$$\frac{d}{dt}\left[T^* e^{\delta t}\right] = kV_0 T_0 e^{-ct} e^{\delta t}$$

Integrating both sides with respect to t we then get,

$$T^* e^{\delta t} = \frac{kV_0 T_0}{\delta - c} e^{-ct} e^{\delta t} + C$$

where C is a constant of integration. After rearranging we reach our initial result,

$$T^* = \frac{kV_0 T_0}{\delta - c} e^{-ct} + Ce^{-\delta t}.$$

Earlier we found the pre-treatment equilibrium for $T_0^* = kV_0 T_0/\delta$, so we can say that when $t = 0$ we have,

$$\frac{kV_0 T_0}{\delta} = \frac{kV_0 T_0}{\delta - c} + C.$$

We then rearrange to find the constant,

$$C = \frac{-ckV_0 T_0}{\delta(\delta - c)}$$

and then substitute back in to get our final solution,

$$T^*(t) = \frac{kT_0 V_0}{\delta(\delta - c)}\left(\delta e^{-ct} - ce^{-\delta t}\right).$$

Phew! Now it's your turn....

Exercises

Using similar methods show that,

$$V_{NI}(t) = \frac{cV_0}{\delta - c}\left(\frac{c}{\delta - c}\left(e^{-\delta t} - e^{-ct}\right) + \delta t e^{-ct}\right)$$

Click for solution

Firstly we can write this equation out in the correct form for using an integrating factor,

$$\frac{dV_{NI}}{dt} + cV_{NI} = N\delta T^* = \frac{NkT_0V_0}{\delta - c}\left(\delta e^{-ct} - ce^{-\delta t}\right).$$

The integrating factor is therefore e^{ct}, which leads us to,

$$\int \frac{d}{dt}\left(V_{NI}e^{ct}\right)dt = \int \frac{NkT_0V_0}{\delta - c}\left(\delta - ce^{-\delta t + ct}\right)dt.$$

Computing these two integrals we find,

$$V_{NI}e^{ct} = \frac{NkT_0V_0}{\delta - c}\left(\delta t - \frac{c}{c-\delta}e^{-\delta t + ct}\right) + A$$

$$\implies V_{NI} = \frac{NkT_0V_0}{\delta - c}\left(\delta t e^{-ct} - \frac{c}{c-\delta}e^{-\delta t}\right) + Ae^{-ct}$$

where A is the constant of integration.

At the start of treatment we'd have $V_{NI}(0) = 0$. Substituting this in tells us,

$$A = \frac{NkT_0V_0}{\delta - c}\left(\frac{c}{c-\delta}\right),$$

and so we have (after a little rearranging),

$$V_{NI} = \frac{NkT_0V_0}{\delta - c}\left(\delta t e^{-ct} + \frac{c}{\delta - c}(e^{-\delta t} - e^{-ct})\right).$$

Finally we use our assumption that $T_0 = c/(Nk)$ to reach the final solution.

This can be added to the solution for $V_I(t)$ to find the overall density of virus particles in the bloodstream at any time point after treatment has begun. After estimating the values of each of the model parameters, researchers plotted the model's predictions against data of individual patient's virus concentrations, and found that that the model provides a remarkably good fit. So even though we have a very simplified model of HIV and immune cell interactions and the effects of treatment, we can get useful and applicable insights.

Key Takeaways

- A patient will only be disease-free if the virus decay rate is high, and its infection rate and burst size is low.
- Under a simplifying assumption the treatment model becomes linear, so we can solve the model explicitly.
- The treatment model predicts exponential decay of the virus, and matches well to real data.

Chapter references

- This chapter is based on models by Perelson and Nelson (1999).

CHAPTER 10

Introducing models of cancer dynamics

INTRODUCING CANCER MODELS

More than one in three people will develop some form of cancer during their lifetime. Given this, it is no surprise that cancer research is a highly active field, and, relevant to this course, that there are a great many researchers interested in modelling the dynamics of cancer growth and its treatments. It can also mean it is a difficult subject for some to think about – I have personally lost much-loved family to cancer, and it is only recently I have felt able to start teaching content in this area – but the more research we can do, the sooner better treatments and even cures will come.

In this chapter we will examine a number of different model forms for thinking about how cancer tumours – which we essentially treat as localised populations of cells – grow or *proliferate* over time (see chapter references). For the most part we will not do any in-depth analysis of the models here, but we will look at one specific model in more detail in the next chapter.

SINGLE VARIABLE MODELS

From a mathematical viewpoint we can essentially think about the growth of a population of cancer cells in much the same way as we would the growth of any population. We have seen some of these model structures already in this textbook, but will cover them again here for completeness.

LINEAR GROWTH (WITH OR WITHOUT MORTALITY)

The most basic model for growth of a population of cancer cells, c – even more basic model than we have examined before – would be for linear growth, given by,

$$\frac{dc}{dt} = r.$$

While you may spot a few flaws in such a model (for example we have positive proliferation even with zero cancer cells), it has been used to describe the dynamics of certain cancers. There is also no mortality of cells here, or *shrinkage* of the tumour. Such a term is readily added to give,

$$\frac{dc}{dt} = r - kc.$$

We can solve this model either by separation of variables or an integrating factor to give,

$$c(t) = \frac{r}{k} + \left(c(0) - \frac{r}{k}\right) e^{-kt}.$$

This suggests the tumour will tend towards an intermediate size of $c = r/k$ (which we can also see by just looking for the equilibrium from the ODE).

EXPONENTIAL GROWTH (WITH OR WITHOUT MORTALITY)

The more classic example of population growth we saw at the very start of this resource was that of exponential growth, where the population growth depends on the current density, that is,

$$\frac{dc}{dt} = rc.$$

There might then be a question of whether r is purely the proliferation rate or, as we assumed earlier, the difference between proliferation and shrinkage. Let's say that we do include a separate shrinkage term, the equation can still be solved using separation of variables,

$$\begin{aligned} \frac{dc}{dt} &= rc - kc, \\ \implies c(t) &= c(0)e^{(r-k)t}, \end{aligned}$$

giving either exponential growth or decay of the tumour depending on whether proliferation or shrinkage is greater.

LOGISTIC GROWTH

Again, as we saw back in chapter 1, there is a realism problem with exponential growth in that it predicts growth to infinite numbers of cells. We introduced one approach to deal with this which is to assume a linear decline in the growth rate, r, as the population density increases, leading to a carrying capacity at size K after which the population declines. This is given by the ODE,

$$\frac{dc}{dt} = r_0 c\left(1 - \frac{c}{K}\right).$$

We solved the non-dimensionalised version of this earlier, with the full solution here being given by,

$$c(t) = \frac{Kc(0)}{c(0) + (K - c(0))e^{-r_0 t}}.$$

GOMPERTZ GROWTH

The logistic equation assumes the growth rate decreases linearly with the population density, but experimental studies have indicated that the decrease in cell proliferation is often closer to exponential. We can readily make our density-dependent growth rate take a different functional form to represent different scenarios. One classic example that has been used to good effect in cancer modelling is the Gompertz model, with the ODE given as,

$$\frac{dc}{dt} = r_0 c \ln\left(\frac{K}{c}\right).$$

Now the individual-level growth rate (I'd like to say *per-capita*, but it has been pointed out to me that cells do not have heads) is $r_0 \ln(K/c)$. This equals 0 when $c = K$, retaining the meaning of K as a carrying capacity. Note we do have an issue that we cannot let $c = 0$. What is the solution to this?

Exercises

Using the substitution $u = \ln\left(\frac{K}{c}\right)$, show that the solution to the Gompertz model is,

$$c(t) = K\left(\frac{c(0)}{K}\right)^{e^{-rt}}.$$

Click for solution

If we're given a substitution we can assume we want to replace all instances of c with some function of u. If $u = \ln(K/c)$, we can find that $du/dc = -1/c$. Putting this all together we can rewrite the ODE as follows,

$$\begin{aligned} \frac{1}{c\ln(K/c)}\frac{dc}{dt} &= r \\ \implies -\frac{1}{u}\frac{du}{dc}\frac{dc}{dt} &= r \\ \implies \frac{1}{u}\frac{du}{dt} &= -r. \end{aligned}$$

At this point we can now use separation of variables to find the solution for u,

$$\begin{aligned} \ln(u) &= -rt + A \\ \implies u &= A_1 e^{-rt} \\ \implies \ln\left(\frac{K}{c}\right) &= A_1 e^{-rt} \end{aligned}$$

where A is the constant of integration and $A_1 = e^A$. If we have density $c(0)$ at $t = 0$ we can find that $A_1 = \ln(K/c(0))$. Then if we remember that $\ln(A/B) = \ln(A) - \ln(B)$, this means that $-\ln(A/B) = \ln(B/A)$. Using this fact we can then say,

$$\begin{aligned} \ln\left(\frac{K}{c}\right) &= \ln\left(\frac{K}{c(0)}\right)e^{-rt} \\ \implies \ln\left(\frac{c}{K}\right) &= \ln\left(\frac{c(0)}{K}\right)e^{-rt} \\ \implies \frac{c}{K} &= e^{\left[\ln\left(\frac{c(0)}{K}\right)e^{-rt}\right]} \\ \implies \frac{c}{K} &= \left(\frac{c(0)}{K}\right)^{e^{-rt}}. \end{aligned}$$

We then just multiply the K to the other side to reach the required solution.

This results in a rather more complex solution that we saw for logistic growth, but at the benefit of a curve that often fits real data much better.

VOLUME-BASED GROWTH

Let's return to the exponential growth model for a moment. A more general form for this is given by,

$$\frac{dc}{dt} = rc^b.$$

It has been suggested that the best choice for the power is not $b = 1$ (as we implicitly assumed in our approach to the exponential model) but $b = 2/3$. Why might that be? Unlike an ecological population, a cancer tumour forms as a roughly spherical object. If it has volume V, then its surface area scales with $V^{2/3}$. If we assume that all resources that the tumour needs must enter through the outer edge of the tumour, then its growth rate will be dependent on its surface area rather than its volume. Note that this need not be limited to the exponential growth model, and may equally well form the growth rate in more complex models.

TWO VARIABLE MODELS

All of the model forms we have looked at so far assume that there is only one variable of interest – the density (or maybe volume) of a cancer tumour. However, there are many reasons why we may wish to explore models with two or more variables. We will explore some of these here.

PROLIFERATING AND QUIESCENT CELLS

Not all cancer cells in a tumour are growing all the time, and this may have impacts on the overall dynamics. We might therefore choose to separate out the cells into those that are proliferating and those that are not, which we call *quiescent*. A simple model structure for this case would be,

$$\frac{dP}{dt} = f(P) - m_1 P + m_2 Q$$
$$\frac{dQ}{dt} = m_1 P - m_2 Q.$$

The function $f(P)$ describes the growth dynamics of the tumour, likely using one of the model forms we saw earlier. There is then a simple linear transfer of cells between the proliferative and quiescent states. The ability to solve this system will depend on the nature of $f(P)$. If it is linear we will be able to solve the system explicitly. Otherwise we would use our qualitative approaches to find the long-term behaviour. Let's look at a quick example here.

Exercises

Find the possible equilibria for P and Q for the system,

$$\frac{dP}{dt} = f(P) - m_1 P + m_2 Q$$
$$\frac{dQ}{dt} = m_1 P - m_2 Q.$$

when there is logistic growth of proliferating cells with basic growth rate r_0 and the carrying capacity *for all cells* is K.

Click for solution

If we have logistic growth with the carrying capacity determined for all cells we can write our system as,

$$\frac{dP}{dt} = r_0 P\left(1 - \frac{P+Q}{K}\right) - m_1 P + m_2 Q$$
$$\frac{dQ}{dt} = m_1 P - m_2 Q.$$

Since this is non-linear it looks like we will not be able to solve it explicitly. Instead, let us determine the equilibria and their stability. If we set the second ODE to 0 we find $Q=Pm_1/m_2$. If we substitute this into the first ODE and set it to 0 we obtain,

$$r_0 P\left(1 - \frac{P(1+m_1/m_2)}{K}\right) = 0.$$

The long-term equilibria are therefore either,

- $P^* = 0$, meaning $Q^* = 0$;
- $P^* = \dfrac{Km_2}{m_1+m_2}$, meaning $Q^* = \dfrac{Km_1}{m_1+m_2}$.

In the first case the tumour is absent, while at the second the tumour is present, and its total size is at its carrying capacity, but only a proportion of those cells are proliferating.

We can check the stability of these equilibria by writing out the Jacobian,

$$J = \begin{pmatrix} r_0 - 2\frac{r_0 P^*}{K} - \frac{r_0 Q^*}{K} - m_1 & -\frac{r_0 P^*}{K} + m_2 \\ m_1 & -m_2 \end{pmatrix}.$$

At the no tumour equilibrium this reduces to,

$$J = \begin{pmatrix} r_0 - m_1 & m_2 \\ m_1 & -m_2 \end{pmatrix},$$

which gives $tr = r_0 - m_1 - m_2$ and $\det = -r_0 m_2$. Since the determinant is negative, this equilibrium is always a saddle. For the equilibrium where the tumour is present, if we substitute in the equilibria and then do some cancelling we have,

$$J = \begin{pmatrix} -\frac{r_0 m_2}{m_1+m_2} - m_1 & -\frac{r_0 m_2}{m_1+m_2} + m_2 \\ m_1 & -m_2 \end{pmatrix}.$$

Here we have $tr = -\dfrac{r_0 m_2}{m_1+m_2} - m_1 - m_2 < 0$ and $\det = r_0 m_2 > 0$, so it is definitely stable. Substituting in some values, it is quite easy to find examples where $tr^2 - 4\det < 0$, meaning we can have a stable spiral into the equilibrium. Note that this is different to if we just had proliferating cells with logistic growth, where no such damped oscillations would be possible.

DRUG-RESISTANT AND DRUG-SENSITIVE CELLS

An important question when thinking about treatment strategies for tumours is whether they can develop drug resistance. If so, we might divide our population into two compartments: sensitive and resistant. We would assume drug-sensitive cells have their proliferation rate reduced through treatment but drug-resistant cells do not. Different assumptions might then be made over whether resistance is a pre-existing trait or if it can be acquired. Similarly, we might explore whether cells can switch back and forth between being sensitive and resistant. Such a model might be represented as,

$$\frac{dS}{dt} = f_S(S) - m_1 S + m_2 R$$
$$\frac{dR}{dt} = f_R(R) + m_1 S - m_2 R.$$

This looks similar to our previous model for proliferating and quiescent cells, except here both types of cell can proliferate, just at different rates. We might also think about whether the transition rates, m_1 and m_2, depend on the level of treatment.

SUMMARY

We have seen just a few examples of model structures here, and there are many more we haven't covered, for example directly including the effects of treatment or the immune system. In the next chapter we will see another specific example where we think about how a tumour both grows and impacts its own carrying capacity through *angiogenesis*.

It is worth stressing that these different model forms are not necessarily limited to cancer modelling, and you may find that some of these different scenarios are more or less suited to the system you are interested in.

Key Takeaways

- We can model growth of tumours in many different ways.
- Single variable models can often be solved explicitly, and include examples such as logistic growth and Gompertz growth.
- Two variable models allow us to explore cell dynamics in more biological detail.

Chapter references

- The content in this chapter is based on works by Gerlee (2013) and Yin et al. (2018).

CHAPTER 11

A model of cancer volume dynamics

A MODEL FOR ANGIOGENESIS

In the last chapter we looked at a number of different ways we might model cancer dynamics, but did minimal analysis of these. In this chapter we will take a more in-depth look at one particular model where we consider the phsyical growth of a cancerous tumour (see chapter references).

We can model the size of a tumour by the number of cancer cells that make it up. These dynamics may actually be well represented by the logistic growth model we studied in the very first lecture (other forms, such as the Gompertz model, are often used instead, but the logistic model will do just fine). That is because tumours tend to slowly increase in size at first, then rapidly grow and finally saturate to a finite size due to resource limitations such as physical space and blood supply. Therefore, the density of cancer cells, c, in a tumour may be expected to obey the dynamics,

$$\frac{dc}{dt} = r_0 c \left(1 - \frac{c}{K}\right)$$

where r_0 is the basic growth rate and K the carrying capacity.

An important feature of tumour growth, however, is that they can change their environment as they grow through *angiogenic* factors, such that they can both stimulate and inhibit their own growth. For example, as the tumour grows it can physically create more space to grow in to, as well as secrete chemicals that encourage blood vessels to grow. On the other hand as the tumour grows it may cause damage to the existing blood supply. By affecting their environment in this way, cancer cells are changing their own carrying capacity. Therefore, while we previously assumed a fixed carrying capacity, K, it now makes sense to treat this is a dynamic variable that may grow or shrink over time depending on these angiogenic processes. The model that is proposed for these dynamics is,

$$\frac{dK}{dt} = \phi c - \theta K c^{2/3} = c\left[\phi - \theta K c^{-1/3}\right].$$

Cells stimulate the growth of blood vessels, and hence the carrying capacity, at individual-level rate ϕ. The inhibition rate, with parameter θ, is rather more complicated and stems from the argument we made in the previous chapter about the volume of tumour, but applied to how tumour growth will damage local blood vessel networks.

ANAYLSIS

As usual, we will proceed by finding possible equilibria, classifying their stability and drawing phase portraits.

We already know from our previous work on the logistic model that,

$$\frac{dc}{dt} = 0 \implies c = 0 \text{ or } c = K.$$

Substituting these values in to our second equation, we find that,

- if $c = 0$, dK/dt=0 for any K, and therefore there are a continuum of equilibria with no tumour but a positive carrying capacity.
- if $c = K$, $dK/dt = 0 \implies c[\phi - \theta K^{2/3}] = 0$. There are two different cases here. Firstly we could have $c = K = 0$ or we could have $c = K = (\phi/\theta)^{3/2}$.

There are therefore two qualitatively different long-term solutions: either the tumour is absent (with the carrying capacity possibly at zero) or it is maintained.

We should now look at the stability of these equilibria. The Jacobian of the system can be found to be,

$$J = \begin{pmatrix} r\left(1 - 2\frac{c}{K}\right) & \frac{rc^2}{K^2} \\ \phi - \frac{2}{3}\theta K c^{-1/3} & \theta c^{2/3} \end{pmatrix}$$

The general case of $c = 0, K > 0$ is problematic because of the lower-left entry. Of course this was already a special case, being a line of equilibria. We will look at this again when we draw the phase portrait. Let us first focus on the special case of $c = K = 0$. In fact, we know that we have $c = K$ in all of these cases, which leads to a number of simplifications.

Exercises

Write out the general Jacobian for this system with $c = K$.

Click for solution

Substituting in $c = K$ we find,

$$J = \begin{pmatrix} -r & r \\ \phi - \frac{2}{3}\theta K c^{2/3} & \theta c^{2/3} \end{pmatrix}$$

If we now look specifically at $c = K = 0$ we find,

$$J = \begin{pmatrix} -r & r \\ \phi & 0 \end{pmatrix}$$

Here we have $tr = -r < 0$ and $\det = -r\phi < 0$, meaning this equilibrium is always unstable.

For the full equilibrium, since we know that $K = (\phi/\theta)^{3/2}$, this simplifies to,

$$J = \begin{pmatrix} -r & r \\ \frac{1}{3}\phi & -\phi \end{pmatrix}$$

We therefore have, $tr = -r - \phi < 0$ and $\det = 2r\phi/3 > 0$, meaning the equilibrium is always stable. Together, then, these results suggest that the tumour will always grow to a fixed size and, even with very low growth, low stimulation and high inhibition, the tumour will persist.

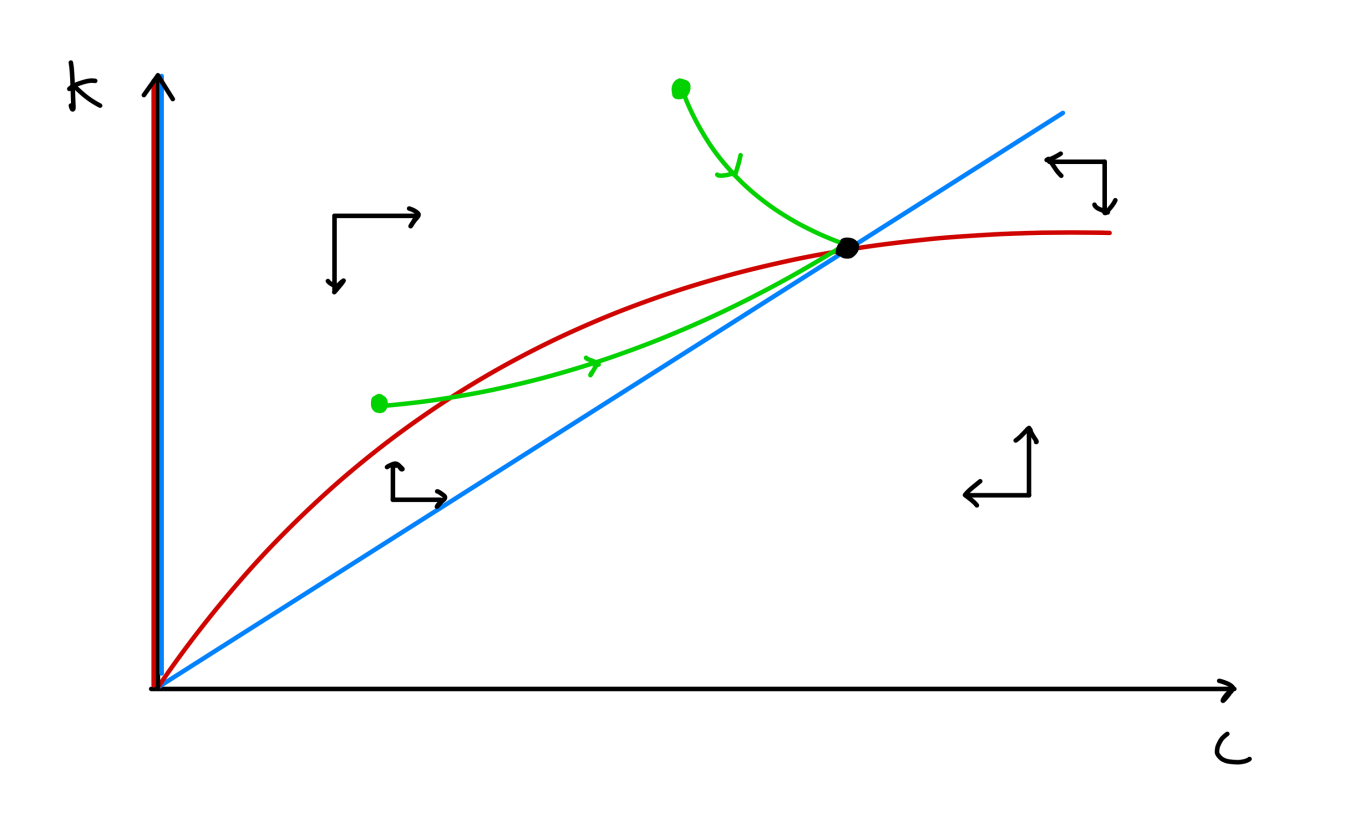
We can explore the behaviour further through plotting the phase portrait. As ever, we need to find the nullclines, determine the qualitative direction of flow, and sketch some trajectories.

Exercises

Sketch the phase portrait for this system.

Click for solution

The phase portrait is sketched below. This shows trajectories must always approach the equilibrium where the size of the tumour exactly equals its carrying capacity.



TREATMENT

There are many ways to treat cancer, including through forms of chemotherapy and radiotherapy and an array of medication. Suppose we administered an anti-angiogenic drug, that reduces the carrying capacity of the tumour by limiting the blood supply. We will take a simple assumption that the drug causes a constant per-capita reduction in the carrying capacity (in reality it would be rather more complicated than this). We can model this by updating the second equation to include an additional mortality term, μ,

$$\frac{dK}{dt} = \phi c - \theta K c^{2/3} - \mu K.$$

After a few lines of working we can find that this changes the equilibrium values to be,

$$c^* = K^* = \left(\frac{\phi - \mu}{\theta}\right)^{3/2},$$

or $c^* = K^* = 0$.

Similarly the Jacobian changes to,

$$J = \begin{pmatrix} -r & r \\ \phi - \frac{2}{3}\theta K^{2/3} & -\theta K^{2/3} - \mu \end{pmatrix}$$

and after substituting in the tumour equilibrium we find,

$$J = \begin{pmatrix} -r & r \\ \frac{1}{3}\phi + \frac{2}{3}\mu & -\phi \end{pmatrix}$$

As before we have $tr = -r - \phi < 0$, but we now have $\det = r\frac{2}{3}[\phi - \mu]$. Therefore when μ is high enough the equilibrium becomes unstable (and in fact it is quickly seen that this would be at the point when the equilibrium becomes negative, suggesting a transcritical bifurcation). We could also demonstrate this by re-plotting the phase portrait, which is left as an optional exercise.

Key Takeaways

- This model of cancer growth looks a lot like the logistic model, but with the carrying capacity also dynamic.
- Without treatment a tumour will always grow to fill its space.
- With treatment, the tumour can be eradicated provided the effect is strong enough.

Chapter references

- The model in this chapter is a much simplified version of that proposed by Hanhfeldt et al. (1999).

PART IV

GENE NETWORKS

CHAPTER 12

Introducing gene networks

A QUICK GUIDE TO CELLS AND GENETIC NETWORKS

In the next few chapters, we will look at models for the regulation of gene expression in cells, and how simple feedback loops underlie some of the basic dynamics that cells exhibit: regulation, switching, and oscillation.

From experience, undergraduate mathematics students are often familiar enough with ideas in ecology or epidemiology to follow the biological aspects of our models well enough without too much extra detail. However, many have little knowledge of cell dynamics and genetics. So, before we start introducing the models themselves we will have a short primer of what it is we are trying to model (and I hope any biologists will forgive me my simplifications and mistakes!)

All living things are made of cells – self-contained structures bounded by a cell membrane (and sometimes also a cell wall). Many organisms exist predominantly as single cells (e.g. bacteria, amoebae), while others exist as patterned collections of cells (e.g. animals, most plants). While different cells possess and maintain a well-defined identity, they are far from static structures, and depend on balanced dynamical processes. Mathematical modelling plays an important role in understanding these dynamical processes and how they are regulated in cells.

Unicellular organisms encounter variable environments, and therefore need to be able to adapt their behaviour. For example, a bacterium living in your gut will have to adapt its metabolism to whatever food you present it with, an example where a cell must switch between different behaviours. Multicellular organisms often develop from a single cell through sequential rounds of cell division. However, the cells in multicellular organisms do not remain identical to each other, but take on stable and well-defined characteristics, so we can talk about skin cells, liver cells, muscle cells, blood cells, etc. in a meaningful way. This adoption of distinct well-defined characters is called *differentiation*, and again requires cells to be able to switch their behaviour.

How do cells manage to change or switch their state in a coherent way? They use a combination of information from their environment and internal mechanisms. Central to the internal mechanisms are what we call gene regulatory networks. Almost all cells contain large structures centred on DNA (deoxyribonucleic acid), basically a set of long, linear sequences of letters (A, C, G and T). This sequence is highly stable, is accurately copied during cell division, and is identical in all cells of a multicellular organism. The full DNA sequence in a cell is referred to as the genome.

How is the genome involved in the regulation of the state of a cell, either stably maintaining it or switching it? A key concept is that of the gene, and the basic unit of dynamics that we will concern ourselves with here is the *expression* of a gene. For our purposes, a gene is a defined subset of the DNA sequence that can code for (can be copied, or *transcribed* to) an mRNA molecule. In turn the mRNA produced from the gene codes for (can be *translated* into) a particular protein molecule. We will refer to the regulated production and degradation of the mRNA and protein corresponding to a gene as the expression of that gene. The amount of the gene itself does not change, and so the dependent variables of our models will be the concentrations of mRNA and protein in a cell.

Initially, we will consider the regulated expression of a single gene. Later, we will consider how genes can interact by regulating each others' expression. We will again focus exclusively on differential equation models,

in which the concentrations/amounts of mRNA and protein are represented by continuous variables. These are essentially population models, just like the ones in the textbook so far.

A FIRST MODEL

In these models, we will consider the dynamics of gene expression in two scenarios:

1. The expression of the gene is regulated by some external transcription factor (whose concentration may or may not be a function of time).
2. The expression of the gene is regulated by its own protein product (which is therefore a transcription factor).

The models have two variables: $M(t)$, the concentration of mRNA, and $P(t)$, the concentration of protein. The dynamics of $M(t)$ and $P(t)$ are regulated by production and degradation. These are similar to what we would formerly have called births and deaths, but those terms don't really make biological sense to use here. In this context, the molecular processes underlying the production of mRNA and protein are called transcription and translation, respectively.

We will make the following assumptions:

1. The rates of mRNA and protein degradation are proportional to their concentration (i.e. degradation is linear).
2. The rate of translation is proportional to the mRNA concentration (i.e. translation is linear).

Then the general form of the models, written as a pair of ordinary differential equations, is

$$\frac{dM}{dt} = R(t) - \mu M$$
$$\frac{dP}{dt} = kM - \nu P.$$

where μ is the degradation rate of mRNA, ν is the degradation rate of protein, k is the translation rate, and $R(t)$ is the transcription rate. Note that for the model to make sense biologically, μ, ν and k must be positive constants, and $R(t) \geq 0$. Note also that if $R(t)$ is not constant, then the model will not go to an equilibrium.

CONSTANT TRANSCRIPTION

Let's get as far as we can analysing this model and then think about this tricky translation term $R(t)$. Starting with the dynamics of mRNA expression, we see that it is linear in M and independent of P, so we can solve it in isolation by using an integrating factor. We have

$$\frac{dM}{dt} + \mu M = R(t).$$

Exercises

Using an integrating factor, show that the initial solution to this ODE is,

$$M(t) = e^{-\mu t} \int e^{\mu t} R(t) dt.$$

Click for solution

The integrating factor will be $e^{\mu t}$. We now multiply both sides by this integrating factor to find,

$$e^{\mu t} \frac{dM}{dt} + e^{\mu t} \mu M = e^{\mu t} R(t)$$
$$\implies \frac{d}{dt} e^{\mu t} M = e^{\mu t} R(t),$$

since the left-hand side is the inverse of the product rule for differentiation. The form of the left-hand side is one that we can readily integrate (that is the whole purpose of using an integrating factor after all). However, because $R(t)$ is some (as yet) unknown function of time we cannot yet compute that integral. This means we can write down the solution as,

$$e^{\mu t} M = \int e^{\mu t} R(t) dt$$
$$\implies M(t) = e^{-\mu t} \int e^{\mu t} R(t) dt.$$

Note that we didn't account for the constant of integration when we integrated the left-hand side; we will deal with this if/when we integrate the term on the right-hand side.

Note we don't have a fully explicit solution here; it is clear that $R(t)$ is going to be very important for how the time course plays out. Let us initially assume that $R(t) = r$, that is, we have constant transcription. As such we can simplify the equation to,

$$M(t) = re^{-\mu t} \int e^{\mu t} dt.$$

This gives us a function we can integrate, so our solution becomes,

$$M(t) = \frac{r}{\mu} e^{-\mu t} \left[e^{\mu t} + C \right],$$
$$= \frac{r}{\mu} \left[1 + Ce^{-\mu t} \right],$$

where C is a constant of integration. Now let us assume we know the initial concentration of mRNA at time $t = 0$ is M_0. We can use this to find our constant of integration as,

$$C = \frac{M_0 \mu}{r} - 1.$$

Substituting back in allows us to reach the final expression of,

$$M(t) = \frac{r}{\mu} + \left[M_0 - \frac{r}{\mu} \right] e^{-\mu t}.$$

From this we can see that as $t \to \infty$ the concentration will tend towards an equilibrium value of $M = r/\mu$. Also, as the expression shows we have exponential decay or growth, we can get an idea of the

speed at which the concentration approaches the equilibrium by finding its *half-life*. In particular, assume that $M_0 = 0$ (i.e. there is no mRNA until time 0 when we "switch the gene on"). This means,

$$M(t) = \frac{r}{\mu}(1 - e^{-\mu t}).$$

Half the equilibrium value will be $r/(2\mu)$. If we call the time at which this value is reached τ, then we have,

$$\frac{r}{2\mu} = \frac{r}{\mu}(1 - e^{-\mu \tau}).$$

Some inspection finds that this requires $e^{-\mu\tau} = 1/2$, meaning $\tau = \ln(2)/\mu$. Therefore the faster the degradation rate of the mRNA, the more slowly it reaches its equilibrium.

What about the time-course of $P(t)$? Now that we have an expression for $M(t)$ we can also turn the equation for $dP(t)/dt$ into a linear ODE, and can again use an integrating factor.

$$\begin{aligned}
&\frac{dP}{dt} + \nu P = kM(t) \\
\implies &\frac{d}{dt}[Pe^{\nu t}] = kM(t)e^{\nu t} \\
\implies &P(t) = ke^{-\nu t}\int M(t)e^{\nu t}dt.
\end{aligned}$$

Making progress relies on us having a nice function for $M(t)$ that we can integrate. Let us again assume $M(0) = M_0 = 0$ and also that $P(0) = P_0 = 0$. We just found that in this case $M(t) = r/\mu(1 - e^{-\mu t})$. Therefore,

$$\begin{aligned}
P(t) &= \frac{kr}{\mu}e^{-\nu t}\int (1 - e^{-\mu t})e^{\nu t}dt, \\
&= \frac{kr}{\mu}e^{-\nu t}\int (e^{\nu t} - e^{(\nu-\mu)t})dt.
\end{aligned}$$

We need to consider the cases where $\nu = \mu$ and $\nu \neq \mu$ separately.

$\nu = \mu$

$$\begin{aligned}
P(t) &= \frac{kr}{\mu}e^{-\mu t}\int (e^{\mu t} - 1)dt, \\
&= \frac{kr}{\mu}e^{-\mu t}\left(\frac{1}{\mu}(e^{\mu t} - 1) - t + C\right)
\end{aligned}$$

With the initial condition $P(0) = 0$ we can find that $C = 0$, and we can re-arrange slightly to give,

$$P(t) = \frac{kr}{\mu^2}\left(1 - (1 + \mu t)e^{-\mu t}\right)$$

$\nu \neq \mu$

$$\begin{aligned}
P(t) &= \frac{kr}{\mu}e^{-\nu t}\int (e^{\nu t} - e^{(\nu-\mu)t})dt, \\
&= \frac{kr}{\mu}e^{-\nu t}\left(\frac{e^{\nu t}}{\nu} - \frac{e^{(\nu-\mu)t}}{\nu - \mu} + C\right).
\end{aligned}$$

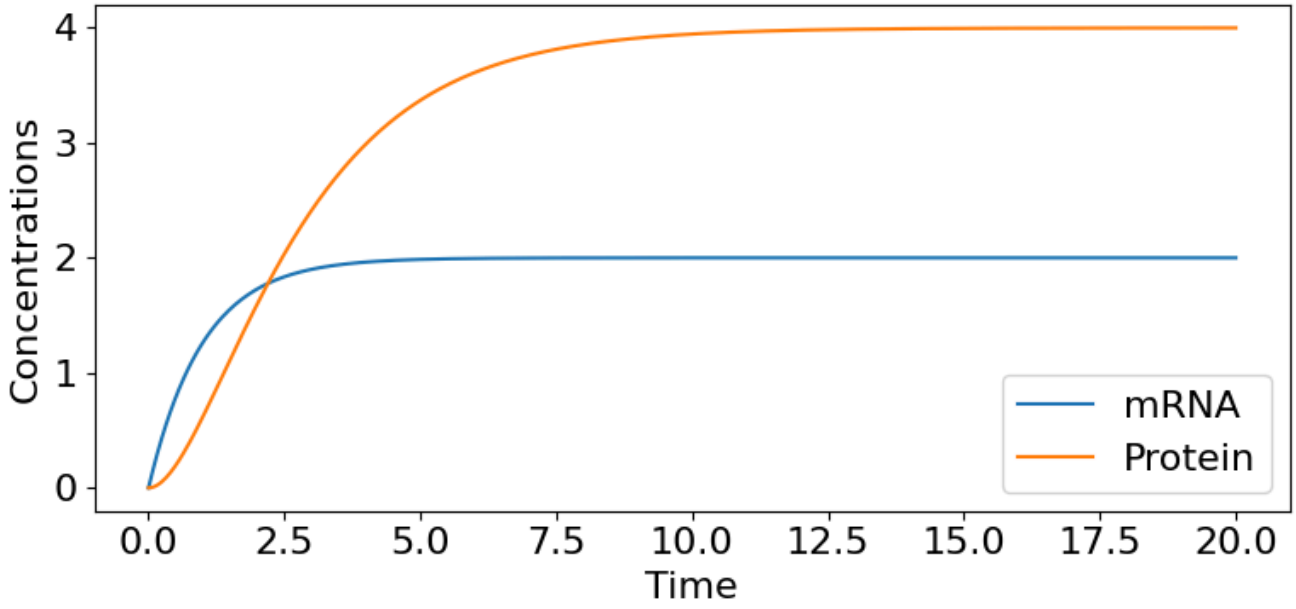
Hopefully you can see from this expression why we needed to take the case $\nu = \mu$ separately. Now we just need to find C. Setting $t = 0$ and $P = 0$ in the initial solution we find,

$$0 = \frac{kr}{\mu}\left(\frac{1}{\nu} - \frac{1}{\nu - \mu} + C\right).$$

Rearranging this gives $C = \mu/(\nu(\nu - \mu))$. Substituting this in and putting everything over a common denominator gives,

$$\begin{aligned} P(t) &= \frac{kr}{\mu}e^{-\nu t}\left(\frac{(\nu - \mu)e^{\nu t}}{\nu(\nu - \mu)} - \frac{\nu e^{(\nu-\mu)t}}{\nu(\nu - \mu)} + \frac{\mu}{\nu(\nu - \mu)}\right) \\ &= \frac{kr}{\mu\nu}\left(1 + \frac{\mu}{\nu - \mu}e^{-\nu t} - \frac{\nu}{\nu - \mu}e^{-\mu t}\right). \end{aligned}$$

In either case we see that as $t \to \infty$, $P \to kr/(\mu\nu)$. The figure below shows the time-course of the system for some example parameter values.



Time-course of single gene model, with $r = 2, \mu = 1, \nu = 0.5, k = 1$ *and initial concentrations* $M(0) = P(0) = 0$.

So, in this model, even when we started with 0 concentrations of both mRNA and protein, both eventually tend towards equilibrium values, with the speed of that approach largely dictated by the mRNA degradation rate.

OSCILLATING TRANSCRIPTION

We allowed that the transcription rate, $R(t)$, might depend on time but then just assumed it was constant. What if we instead assume we have an oscillating transcription rate? We will often find that we end up with expressions that are too complicated to integrate explicitly (and in that case we would probably just numerically integrate it on a computer), but if we take a sinusoidal function we can make some progress.

Let's take the transcription function $R(t) = r_0(1 + \delta\sin(\omega t))$. This has average input r_0, but varies between 0 and $2r_0$, with the period depending on ω. Returning to our earlier expression, we now need to solve,

$$M(t) = e^{-\mu t} \int e^{\mu t} r_0 \left[1 + \delta \sin(\omega t)\right] dt$$
$$= r_0 e^{-\mu t} \left[\int e^{\mu t} dt + \delta \int e^{\mu t} \sin(\omega t) dt\right].$$

The first integral is identical to what we had before, so we can just copy that down again. The second may initially look a bit scary, but is actually quite a classic example of applying the method of *integration by parts*. This is sometimes called a "product rule for integration", with the trick being that when you have an exponential multiplied by a sine or cosine, after a couple of steps you get the same integral back again (if you do not know how to do this, you could try asking a friendly mathematician to talk you through it, or look it up online, but we will also go through it in some detail in the exercise solution).

Exercises

Show that the initial solution for $M(t)$ is given by,

$$r_0 e^{-\mu t} \left[\frac{1}{\mu} e^{\mu t} + \frac{\delta e^{\mu t}}{\mu^2 + \omega^2} (\mu \sin(\omega t) - \omega \cos(\omega t)) + C\right],$$

where C is a constant of integration.

Click for solution

We will focus first on just finding the solution to the integral,

$$I = \int e^{\mu t} \sin(\omega t) dt,$$

which we will do through integration by parts. Let $u = e^{\mu t}$ and $dv = \sin(\omega t)$. Then we use these to find $du = \mu e^{\mu t}$ and $v = -\cos(\omega t)/\omega$. The integration by parts formula is that,

$$\int u dv = uv - \int v du,$$

so we have,

$$I = -\frac{1}{\omega} e^{\mu t} \cos(\omega t) + \frac{\mu}{\omega} \int e^{\mu t} \cos(\omega t) dt,$$

(we should also have a constant of integration but we shall leave that out until the very end). We now have one nice term, but another tricky looking integration. Strange as it may seem, if we do integration by parts again we will make a lot of progress. Let's take some new variables $u = e^{\mu t}$ and $dv = \cos(\omega t)$, meaning $du = \mu e^{\mu t}$ and $v = \sin(\omega t)/\omega$. This gives is an updated solution of,

$$I = -\frac{1}{\omega} e^{\mu t} \cos(\omega t) + \frac{\mu}{\omega} \left[\frac{1}{\omega} e^{\mu t} sin(\omega t) - \frac{\mu}{\omega} \int e^{\mu t} \sin(\omega t) dt\right],$$
$$= \frac{1}{\omega^2} e^{\mu t} \left[-\omega \cos(\omega t) + \mu \sin(\omega t)\right] - \mu^2/\omega^2 \int e^{\mu t} \sin(\omega t) dt.$$

This may look like we just keep making things worse! However, if you look at the remaining integral you might spot that

this is actually the integral we set out to find to start with. Now that might initially seem like a problem, but it is in fact a big help. We can now write this as,

$$I = \frac{e^{\mu t}}{\omega^2}[-\omega\cos(\omega t) + \mu\sin(\omega t)] - \mu^2/\omega^2 I,$$

$$\implies \frac{\mu^2+\omega^2}{\omega^2} I = \frac{e^{\mu t}}{\omega^2}[\mu\sin(\omega t) - \omega\cos(\omega t)]$$

$$\implies I = \frac{e^{\mu t}}{\mu^2+\omega^2}[\mu\sin(\omega t) - \omega\cos(\omega t)]\,.$$

Now we can put this together with the solution to the more straightforward integral (and remembering our constant of integration) to find,

$$r_0 e^{-\mu t}\left[\frac{1}{\mu}e^{\mu t} + \frac{\delta e^{\mu t}}{\mu^2+\omega^2}(\mu\sin(\omega t) - \omega\cos(\omega t)) + C\right].$$

We now need to complete this solution by adding in the initial condition. If we again take the initial condition that $M(0) = 0$, you should be able to show that,

$$C = \frac{\delta\omega}{\mu^2+\omega^2} - \frac{1}{\mu},$$

which leads us to the final – somewhat long – solution,

$$M(t) = r_0\left[\frac{1}{\mu} + \frac{\delta}{\mu^2+\omega^2}(\mu\sin(\omega t) - \omega\cos(\omega t)) + \left(\frac{\delta\omega}{\mu^2+\omega^2} - \frac{1}{\mu}\right)e^{-\mu t}\right].$$

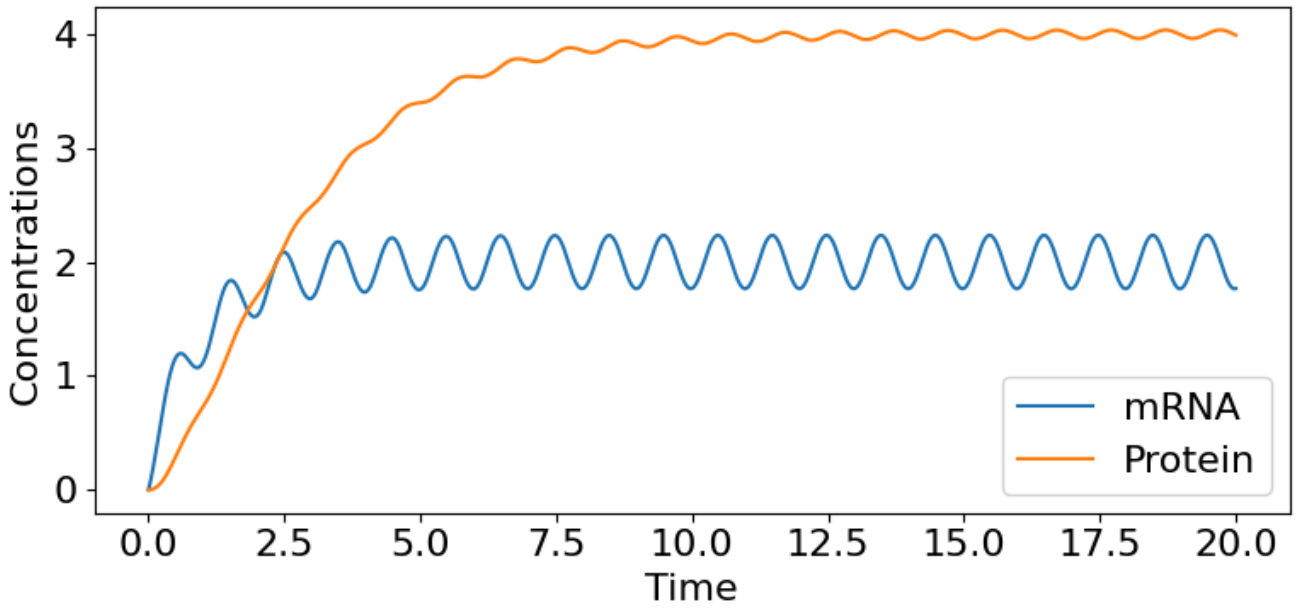
As long as this is, it can readily be broken into three parts – a constant term, an oscillating term and an exponentially decaying term. After a long time we can assume the exponentially decaying term stops having much effect, meaning the solution reduces to,

$$M_{long-time}(t) = r_0\left[\frac{1}{\mu} + \frac{\delta}{\mu^2+\omega^2}(\mu\sin(\omega t) - \omega\cos(\omega t))\right].$$

Using a few more results about trigonometric functions we can reduce this a bit further to,

$$M_{long-time}(t) = r_0\left[\frac{1}{\mu} + \frac{\delta}{\sqrt{\mu^2+\omega^2}}\cos(\omega t - \theta)\right],$$

where $\theta = tan^{-1}(\omega/\mu)$. This then tells us that, in the long-term, the mRNA concentration oscillates around the value of r_0/μ (the same value as in the constant case), with the same period as the inputted transcription rate (because the 'time' term in the cosine is ωt)), but delayed behind the input by amount θ. We can follow a similar approach to get an expression for $P(t)$ as well if we wish but we won't cover that here. A time-course of these dynamics is shown below, showing the regular fluctuations in expression.



Time-course of single gene model with oscillating transcription, with $r_0 = 2, \mu = 1, \nu = 0.5, k = 1, \delta = 0.75, \omega = 2\pi$ *and initial concentrations* $M(0) = P(0) = 0$.

Key Takeaways

- We can model the dynamics of cells, genes and proteins in much the same way as we modelled larger populations.
- In a simple model of gene expression, the equations are linear and can be solved with integrating factors.
- If *transcription* is constant we expect an equilibrium of gene expression, but if it oscillates we expect oscillating expression.

Chapter references

- The content in the *Gene networks* section is based on the unpublished *Mathematical Biology* lecture notes developed by Nick Monk.

CHAPTER 13

Autoregulation 1: auto-repression

REGULATORY FEEDBACK LOOPS

In our first example, transcription of mRNA was due to some external signal or resource. However, many genes encode transcription factors that directly regulate the rate of transcription of the coding gene. This produces a small feedback circuit. Representing the concentrations of mRNA and protein by $M(t)$ and $P(t)$ again, we can now write,

$$\frac{dM}{dt} = f(P) - \mu M$$
$$\frac{dP}{dt} = kM - \nu P.$$

Now the concentration of protein directly influences the transcription of mRNA, producing this *autoregulation* feedback loop. The function $f(P)$ must be bounded above, since transcription must have some maximal rate. Furthermore $f(P) \geq 0$ (the production rate cannot be negative). Therefore $f(P)$ must be non-linear, which makes the system defined by the equations above difficult, if not impossible, to solve explicitly.

To gain insight into the possible dynamics of the system, we will use qualitative analysis based around the construction of the phase portraits and linear stability analysis.

CONSTRUCTING THE PHASE PORTRAIT

NULLCLINES

The nullclines of the system are given by,

$$\frac{dM}{dt} = 0 \implies M = f(P)/\mu$$
$$\frac{dP}{dt} = 0 \implies M = \nu P/k.$$

These are both single lines/curves in the phase plane.

EQUILIBRIA

Putting the two nullcline equations equal to one another, we find a single equation for the possible equilibrium values of P:

$$\frac{\mu\nu}{k}P^* = f(P^*).$$

In that case we can then find $M^* = \nu P^*/k$ to be the corresponding equilibrium for mRNA. We can therefore focus our attention on determining the nature of any solutions for P^* for different forms of the function $f(P)$.

LINEARISATION

As in previous examples, we assess stability around an equilibrium by linearisation, through looking at the Jacobian. For this model our Jacobian is,

$$J = \begin{pmatrix} -\mu & f'(P^*) \\ k & -\nu \end{pmatrix}.$$

We then check the trace and determinant to assess stability:

- $tr(J) = -\mu - \nu$
- $\det(J) = \mu\nu - kf'(P^*)$.

The trace is always negative, limiting stability to a stable spiral or node, or an unstable saddle. Which outcome we get depends on the determinant, which in turn depends on the gradient of our transcription function at the equilibrium, $f'(P^*)$. We will now consider a few examples of what this function might look like.

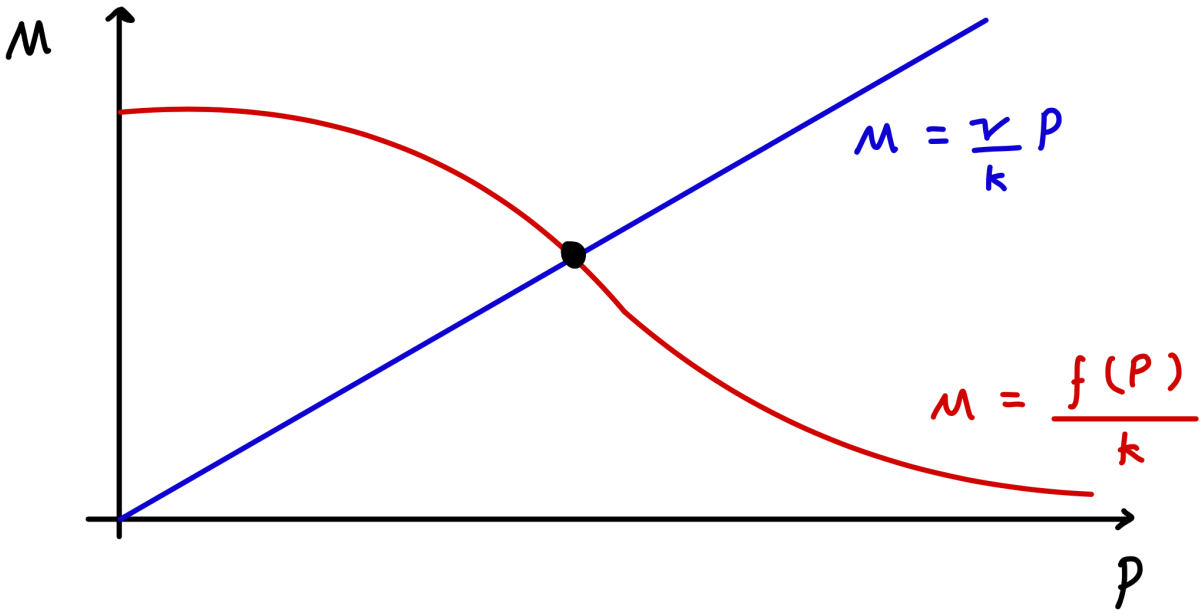
AUTO-REPRESSION: A NEGATIVE FEEDBACK

If the protein product of a gene acts to reduce the rate of transcription, we describe it as a *transcriptional repressor*. The feedback circuit is described as *auto-repression*. In this case, $f'(P) \leq 0$, i.e. $f(P)$ is a decreasing function. For the most part we will not write down explicit forms for $f(P)$ but just rely on our qualitative knowledge about it.

Recall the equation for steady state values of P^* is

$$\frac{\mu\nu}{k}P^* = f(P^*).$$

Since $f(P)$ is strictly decreasing, this has a unique solution for $P^* > 0$, as shown graphically below.



Nullclines of the auto-repression model. The blue line is the protein nullcline and the red curve is the mRNA nullcline.

We therefore have one unique equilibrium. Let us call the value of $f'(P^*) = \phi < 0$ at the equilibrium. As such we can say that $\det(J) = \mu\nu - k\phi$ is positive, meaning the equilibrium is stable. Whether the equilibrium is a node or a spiral then depends on the value of,

$$\begin{aligned} tr^2 - 4\det &= (\mu+\nu)^2 - 4\nu\mu + 4k\phi \\ &= (\mu-\nu)^2 + 4k\phi. \end{aligned}$$

Firstly, we can see in the special case that $\mu = \nu$ (the two degradation rates are equal), we just have $4k\phi < 0$, meaning we will have a stable spiral. More generally, there is a critical value,

$$\phi_c = -\frac{(\mu-\nu)^2}{4k},$$

such that,

- if $\phi > \phi_c$ then we have a stable node,
- if $\phi < \phi_c$ then we have a stable spiral.

Recalling that $\phi < 0$, we see that $f(P)$ must be sufficiently steep at the equilibrium for it to be a spiral. The steepness of $f(P)$ (i.e. the modulus of the derivative of $f(P)$) can be thought of as the *sensitivity* of the rate of transcription to changes in the protein concentration; high sensitivity means that a small change in the concentration of P results in a large change in the rate of transcription. So sensitive regulation is more likely to lead to a stable spiral.

SKETCHING THE PHASE PORTRAIT

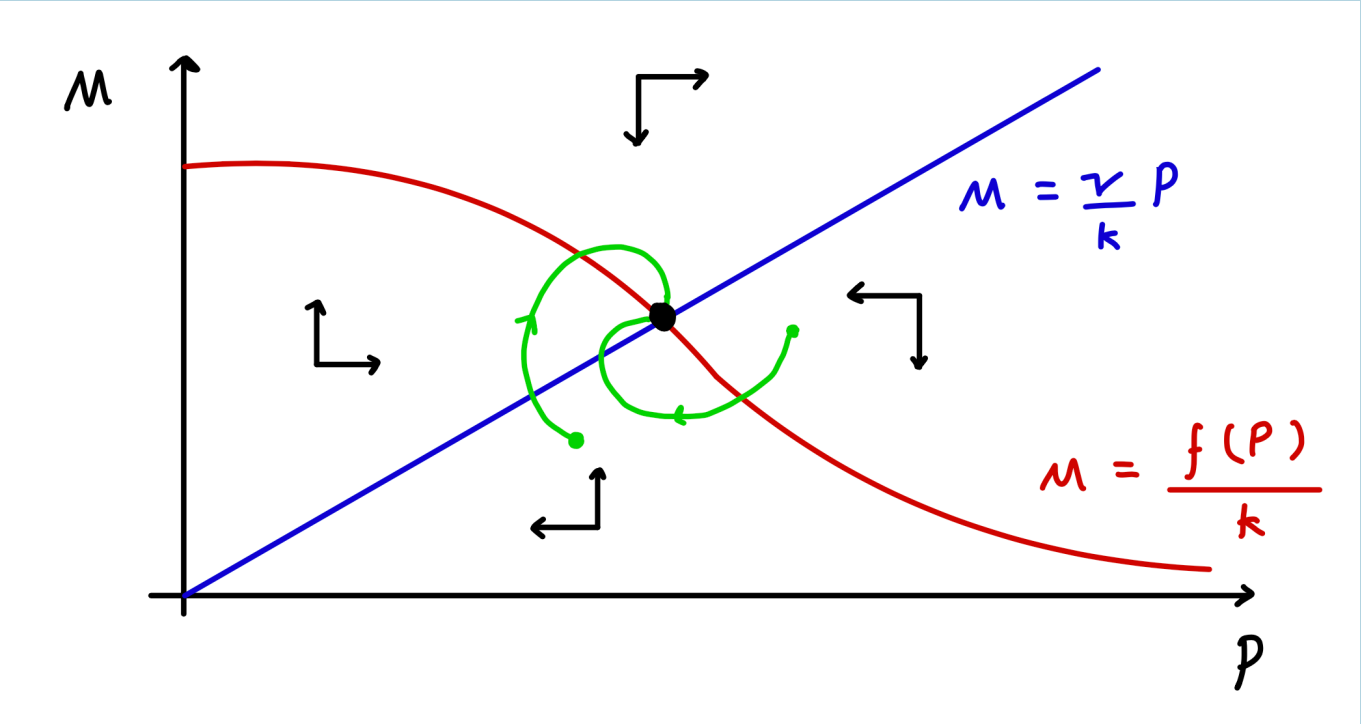
We can now sketch the phase portrait for the case of an auto-repressive gene. Note that it is easiest to draw the phase portrait with the variable P on the x-axis, and M on the y-axis.

Exercises

Using the nullclines above as a starting point, sketch the phase portrait showing how trajectories behave in this system.

Click for solution

The phase portrait is sketched below.



Note that if the equilibrium is a spiral, then trajectories spiral in a clockwise direction. The time courses $M(t)$ and $P(t)$ will be damped oscillations about their equilibrium values, with a peak in P following a peak in M. This makes sense biologically, since mRNA acts as a template for production of protein (by translation).

Key Takeaways

- An autoregulatory gene network means transcription of mRNA is controlled by the concentration of protein.
- An auto-repressive gene can be modelled by the transcription being a decreasing function of protein concentration.
- An auto-repressive gene has a single equilibrium that can either be a stable node or spiral, depending on the model parameters.

Chapter references

- The content in the *Gene networks* section is based on the unpublished *Mathematical Biology* lecture notes developed by Nick Monk.

CHAPTER 14

Autoregulation 2: auto-activation

AUTO-ACTIVATION: A POSITIVE FEEDBACK

In the last chapter we introduced the general model for an auto-regulatory circuit for mRNA and protein dynamics,

$$\frac{dM}{dt} = f(P) - \mu M$$
$$\frac{dP}{dt} = kM - \nu P$$

and looked at the specific case of auto-repression where $f'(P) < 0$. We will now look at the opposite form of autoregulatory genetic networks. If the protein product of a gene acts to increase the rate of transcription, we describe it as a *transcriptional activator*. The feedback circuit is described as *auto-activation*. In this case, $f'(P) \geq 0$, i.e. $f(P)$ is an increasing function.

The equation for steady state values of P is

$$\frac{\mu\nu}{k}P = f(P)$$

We know that $f(P)$ is increasing and bounded above, so what can we conclude about the number of equilibria? Depending on how weird and wonderful we make $f(P)$, there must be at least one, and there can be any odd number of equilibria. Realistically we'd expect one or three, since we probably wouldn't choose a form for $f(P)$ that is too wild.

As before, the Jacobian is given by,

$$J = \begin{pmatrix} -\mu & f'(P^*) \\ k & -\nu \end{pmatrix},$$

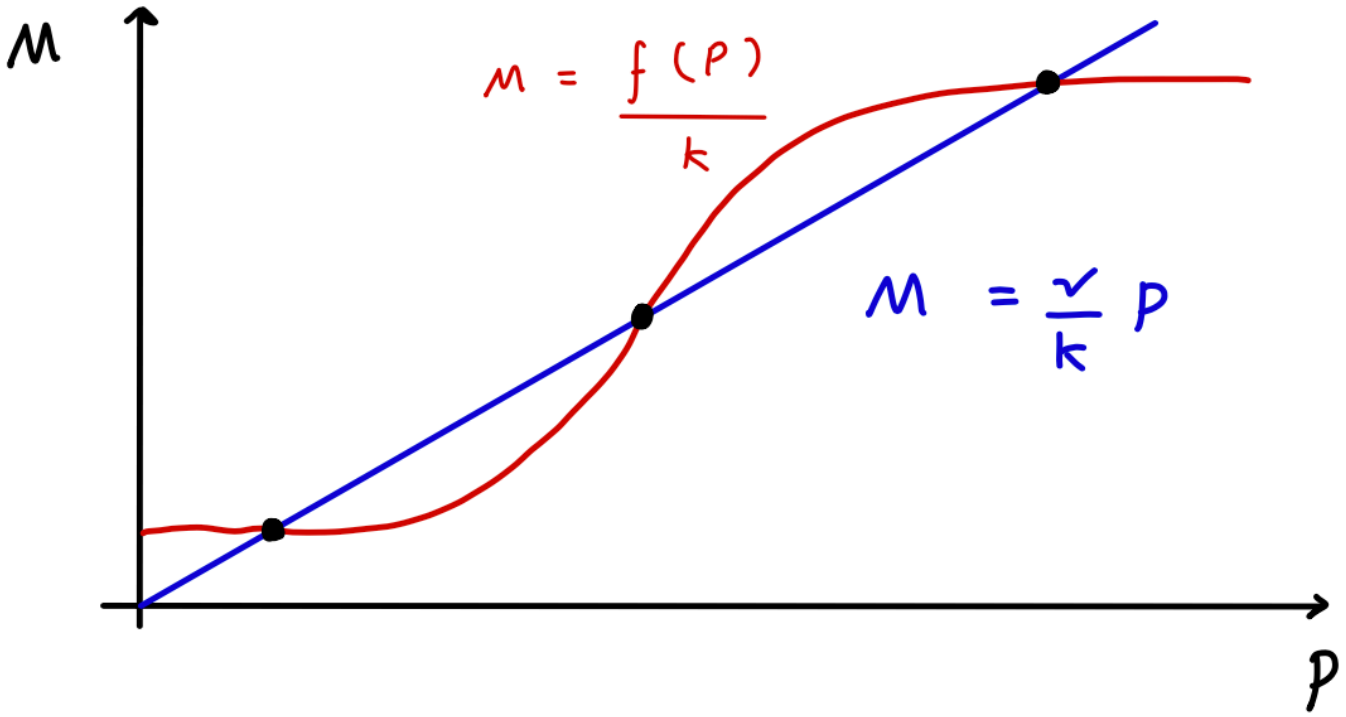
but now $f'(P^*) > 0$. The trace remains $tr(J) = -\mu - \nu < 0$. The determinant is again given by $\det(J) = \mu\nu - k\phi$, where $\phi = f'(P^*)$, but because $\phi < 0$ it can be either positive or negative. We therefore have that,

- if $\phi < \mu\nu/k$, the equilibrium is stable,
- if $\phi > \mu\nu/k$, the equilibrium is a saddle.

We can go slightly further in the first case, by recalling that $tr^2 - 4\det = (\mu - \nu)^2 + 4k\phi > 0$. Therefore if the equilibrium is stable it will definitely be a node and not a spiral.

USING PHASE PORTRAITS

As it turns out, using this information, when we look at a phase portrait we can immediately spot whether an equilibrium is a stable node or a saddle. Consider the sketched nullclines in the figure below. This identifies three equilibria where the two nullclines cross in this case. The gradient of the blue line is ν/k and the gradient of the red line is $f'(P)/\mu = \phi/\mu$. As we have just found, which of these two terms is greater – i.e. which is the steeper – controls the stability of each equilibrium. Therefore in the two cases where the blue line is steeper the equilibria are stable nodes, and in the case where the red line is steeper we have a saddle.



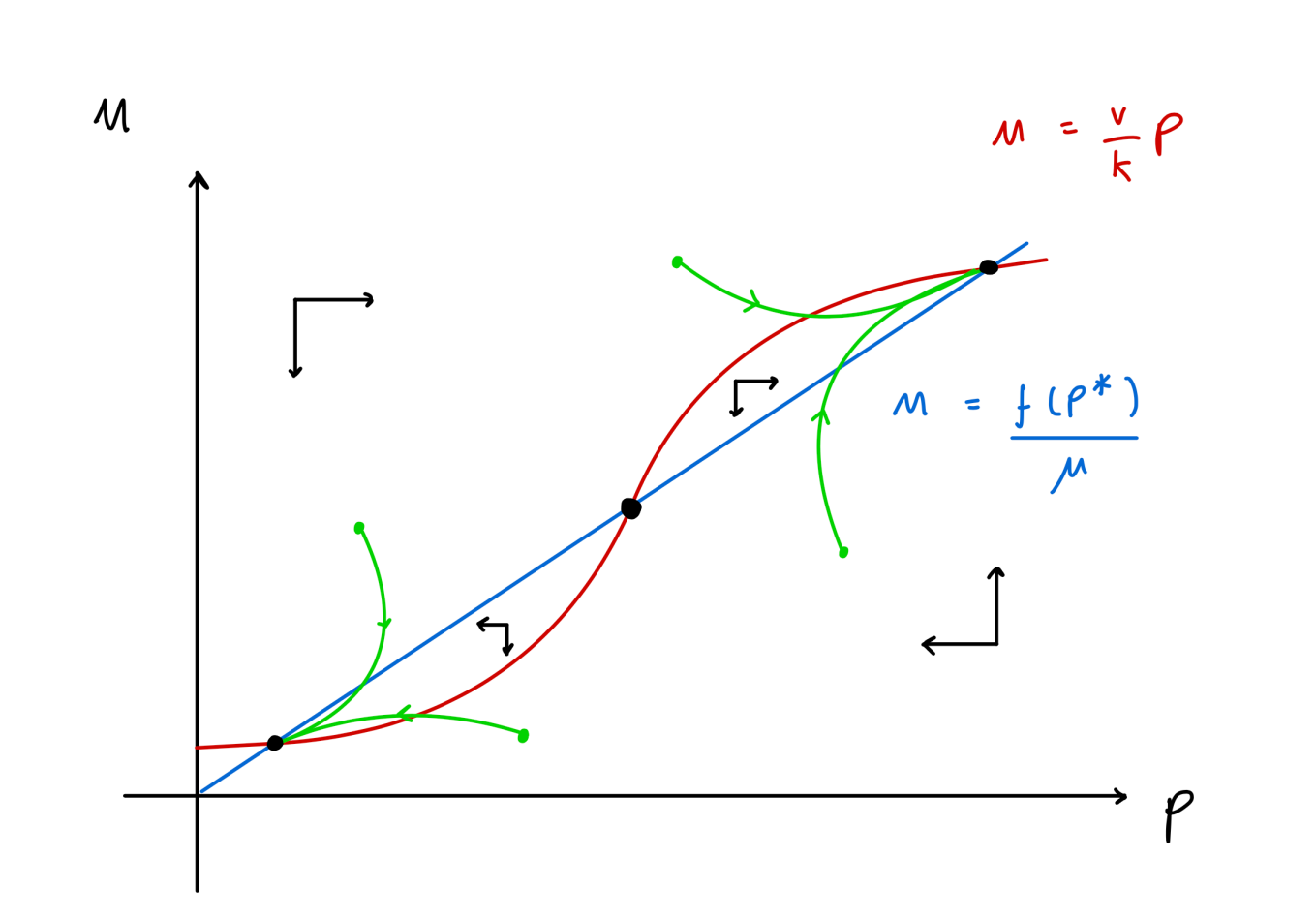
Nullclines for the auto-activation model. The blue line is the protein nullcline and the red curve is the mRNA nullcline.

Nullclines of auto-regression model
Figure: example of nullclines for auto-repression model.

Exercises

Fill in the rest of the phase portrait, showing how trajectories behave in this system.

Click for solution
The full phase portrait is sketched below.



This system demonstrates *bistability* – there are two possible stable endpoints separated by an unstable saddle (we saw a similar case back when we looked at the competition model). This particular case is often called a *bistable switch* because the system can switch between high and low expression of the gene with small changes to the initial condiitons.

It is worth noting that we could equally have drawn the phase portrait with just one equilibrium. These must result in a single stable node (since the red nullcline must start from above and then cross below the blue nullcline), but can be at high or low gene expression depending on quite how we draw the two nullclines, as we will see below.

AN EXAMPLE TRANSCRIPTION FUNCTION

You will have noticed we sketched our $f(P)$ nullclines in a sigmoidal shape. We saw something similar with the spruce budworm, predator-prey and within-host Covid-19 models. In this biological context they are often called *Hill functions*. The general formula for such a curve is,

$$f(P) = \alpha + \beta \frac{P^n}{\theta^n + P^n},$$

where α, β, θ and n are all positive constants. The curve has the following key features:

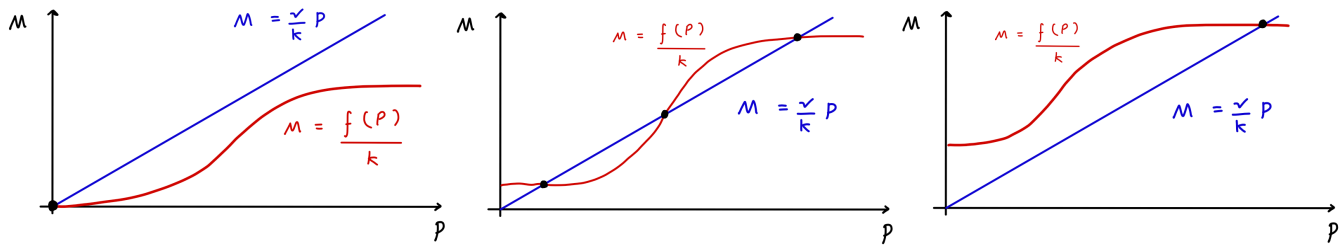
- The curve starts at $f(0) = \alpha$,
- The gradient $f'(P) > 0$,
- θ is the *half-saturation constant*, with $f(\theta) = \alpha + \beta/2$,
- n controls how strong the sigmoidal shape is, and we usually assume $n \geq 2$,
- As $P \to \infty$, $f(P) \to \alpha + \beta$.

Thinking more about the contribution of n, as this becomes larger, the transition from low to high transcription rates becomes more sudden and steep – we would describe this as increasing the *sensitivity* of the rate of transcription.

To see how we can use the parameters of the Hill function to switch gene expression from a low to high state, consider the effect of increasing the value of α. If we start with $\alpha = 0$ and ν/k large enough, then the only steady state is a stable node at $P = 0$ as in the first figure below. We say the system is *monostable*. In terms of the biology of the system, the gene is turned off.

If we now increase the value of α, then provided the steepest part of the function has a gradient that is greater than ν/k , there is a range of values of α for which the system has three steady states. For this range of values of α, the system is bistable, and has two possible stable steady state expression levels as shown in the second diagram.

If we increase α even further, then the system again has only one stable node steady state (monostability), but now it corresponds to a high level of expression of P, as in the final diagram. By increasing the value of α , we have therefore switched the gene from a stable "off" state ($P = 0$) to a stable "on" state (P high).



Three nullcline diagrams for the auto-activation model as we increase α. Labellings are as in the earlier figures.

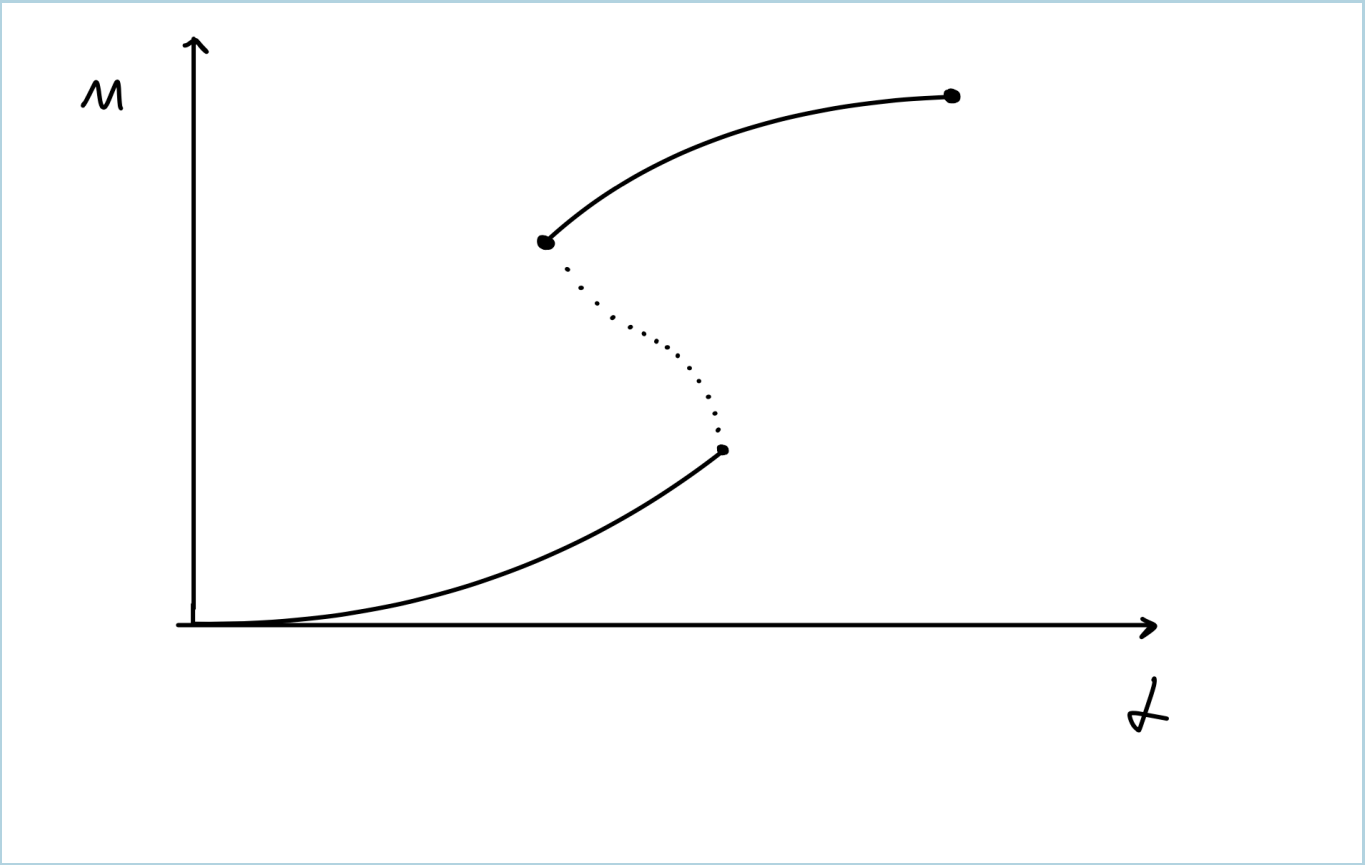
Exercises

Using the three diagrams to guide you, sketch a bifurcation diagram of the system, showing the stable and unstable equilibria as α is varied.

Click for solution

We can use the three phase portraits to guide us. When $\alpha = 0$ then the only equilibrium is at $M = 0$, so we can mark on this starting point. As we increase α, the nullcline will move upwards, causing a similar increase to the

equilibrium, so the equilibrium will start to increase. We know this equilibrium is stable so we draw it with a solid line. Once we reach a critical value of α, the two new equilibria emerge at a higher value of M, so we initially have a single point that diverges into two lines – the higher one is stable so is drawn with a solid line, but the central one is a saddle so is drawn with a dashed line. As α increases the outer two equilibria continue to higher values of M, but the central one moves down. Eventually it collides with the lower equilibrium and the two disappear. The full diagram is shown below.



α can be considered as an external input to the system – a transcription rate that is independent of P. We sometimes refer to α as a *basal transcription rate*. If the system is initially in the bistable region (with three steady states), but is in the low expression state (which is a stable node), then a transient increase in α can push the system into the monostable region, where only a high stable steady state exists. In other words, expression of the gene is turned on. If α is now returned to its original value, the system will remain in the steady state corresponding to a high level of expression. Thus, a transient externally-imposed impulse can stably switch the bistable system from a low to a high level of expression of an auto-activating gene. This is an example of *hysteresis* – the response of the system depends not only on its current input, but also on its history of past inputs. Whether or not such switching occurs depends on the speed and magnitude of the transient increase in α.

Key Takeaways

- We can model auto-activation of a gene by having a transcription rate that is a positive function of P.
- This usually produces one stable equilibrium or three equilibria: an unstable saddle surrounded by two stable nodes.
- Using a sigmoidal function to model transcription, we can see how a gene can be switched on and off by potentially small changes to the input.

Chapter references

- The content in the *Gene networks* section is based on the unpublished *Mathematical Biology* lecture notes developed by Nick Monk.

CHAPTER 15

Longer negative feedback networks

A SIMPLIFIED 3-VARIABLE MODEL

We saw in chapter 13 that a two-component negative feedback circuit (the auto-repressive gene system) possesses a unique steady state, which is always stable. This underlies *homeostasis*; the tendency of a negative feedback system to return to its equilibrium when perturbed (for small perturbations, at least).

However, negative feedback circuits can also generate sustained oscillations. One way this can occur is if we extend our two-variable model to a model including at least three variables. This was first proposed in the context of regulated gene expression in the 1960s, and the prototype three-component model is often referred to as the *Goodwin oscillator* (see chapter references). There are a number of ways of thinking about what might underlie a model of auto-regulatory gene expression that would contain more than two variables. One example would be that in fact the protein product of the gene is translated/produced in one place – the cytoplasm – but regulates transcription in another – the nucleus. We could therefore consider three variables: mRNA, cytoplasmic protein and nuclear protein.

Here, we will look at a slightly simplified model to Goodwin's original, but the principles remain the same. Rather than focusing on a specific example of gene expression, we will consider a simple model of three interacting components, X, Y, and Z, which form a negative feedback loop by regulating each other. In particular:

- Increasing X leads to increased production of Y,
- Increasing Y leads to increased production of Z,
- Increasing Z leads to *decreased* production of X.

Put together there is therefore a negative feedback in the system (since increasing X ultimately leads to decreased production of X). For simplicity we will assume that the degradation rates of all three components are the same, so our model can be written as,

$$\frac{dX}{dt} = f_1(Z) - \mu X$$
$$\frac{dY}{dt} = f_2(X) - \mu Y$$
$$\frac{dZ}{dt} = f_3(Y) - \mu Z.$$

Each of the functions f_i are non-negative and bounded as we assumed before. Given our assumptions in the bullet points we also have that $df_1/dZ < 0$ and $df_2/dx, df_3/dY > 0$.

Equilibria of this system occur where the following three conditions are simultaneously satisfied:

$$\mu X = f_1(Z)$$
$$\mu Y = f_2(X)$$
$$\mu Z = f_3(Y).$$

How many equilibria can we expect? If we let $F_i = f_i/\mu$, then we have $X = F_1(F_3(F_2(X))) = G(X)$. We know that $G(X)$ must be a decreasing function of X (intuitively it must be because we know we have a negative feedback loop, but we can also show it by the chain rule). By the same argument as in the two-component auto-repression model, then, we can say that there must be a single equilibrium (because X is an increasing straight line and $G(X)$ is some decreasing function, they can only cross once).

LINEAR STABILITY ANALYSIS

The Jacobian for this system may be 3 dimensional but doesn't look too horrid,

$$J = \begin{pmatrix} -\mu & 0 & \phi_1 \\ \phi_2 & -\mu & 0 \\ 0 & \phi_3 & -\mu \end{pmatrix},$$

where ϕ_i are the derivatives of f_i at the equilibrium point. We have seen before that 3-dimensional Jacobians cannot be solved as easily as other models we have met. However, because we kept this system (relatively) simple we can actually check its stability directly from the *characteristic equation*, a bit like when we looked at the HIV model. To find this we take,

$$\det(J - \lambda I) = (-\mu - \lambda)^3 + \phi_1\phi_2\phi_3.$$

We then look for solutions for λ to where this equation equals 0. That is,

$$(\mu + \lambda)^3 = \phi_1\phi_2\phi_3 = -\chi^3 < 0.$$

Here we have made the product of our three gradients equal to a new parameter $-\chi^3$. The minus sign is to remind us that this product is negative (so in fact $\chi > 0$) and the cubed is to remind us that this is really the composite of three parameters. The solution to this is then,

$$\lambda = (-1)^{1/3}\chi - \mu.$$

Somewhat susprisingly then, we find ourselves needing to find the cubed root of -1 in order to understand a real-world biological problem. Recalling that we can write $e^{-i\pi} = -1$ (I bet you didn't expect to see Euler's relation in this textbook!), we have that the three possible solutions are,

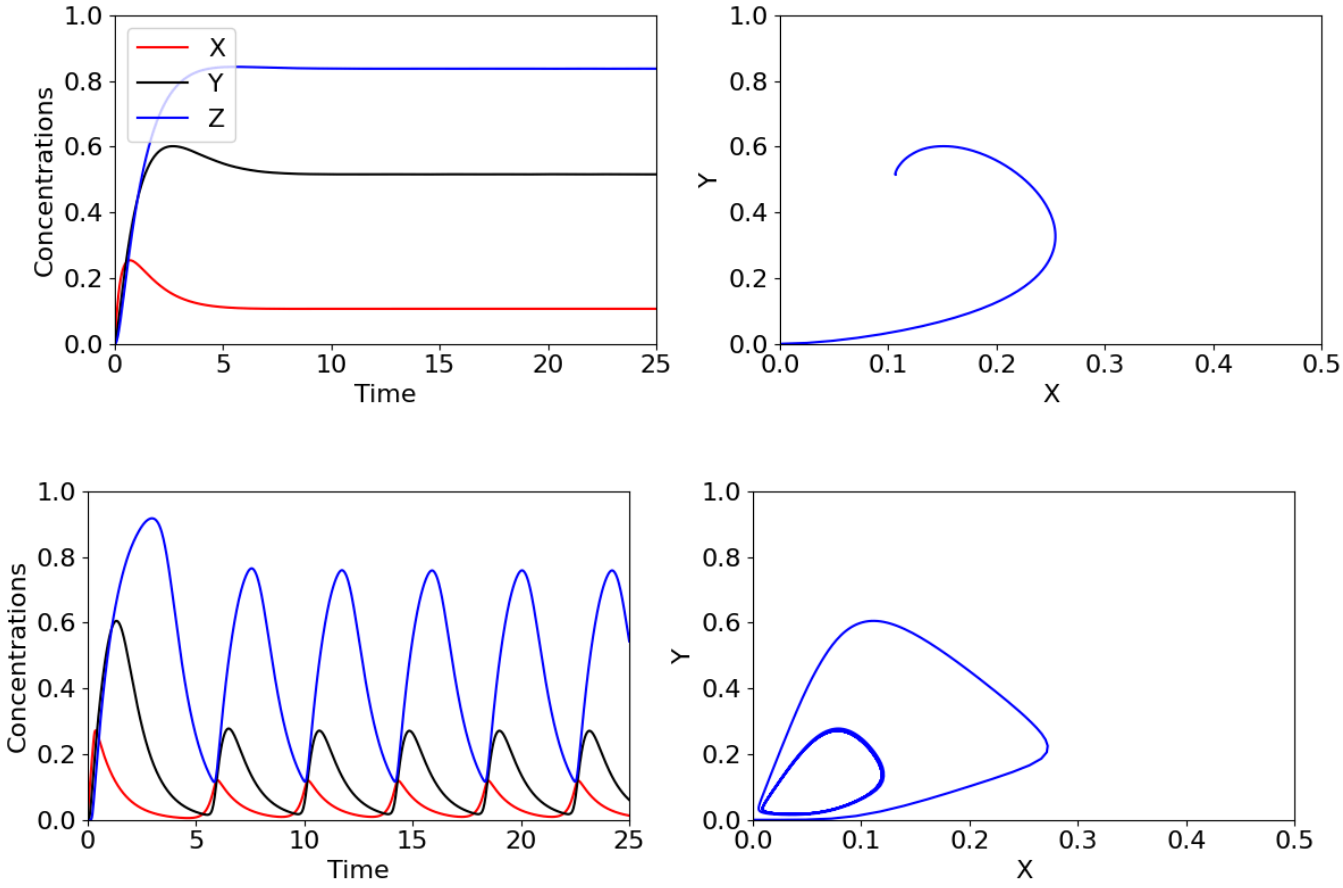
- $\lambda_1 = -\chi - \mu$,
- $\lambda_2 = e^{i\pi/3}\chi - \mu$,
- $\lambda_3 = e^{-i\pi/3}\chi - \mu$.

We can see that λ_1 is definitley negative. However the other two eigenvalues are complex, and we can find that they have real part $Re(\lambda_{2,3}) = -\mu + \chi/2$. Putting this all together we find that,

- if $\chi < 2\mu$, all the eigenvalues have negative real part and the equilibrium is a stable spiral (we know it is a spiral because two of the eigenvalues are complex),
- if $\chi > 2\mu$, two of the eigenvalues have positive real part and the equilibrium is an unstable spiral.

We have seen a case before where stability moves from a stable spiral to an unstable spiral, when we looked at the predator-prey model. You may recall this produced a *Hopf bifurcation*, where a *limit cycle* emerges. The

same phenomenon occurs here, with sustained oscillations in the expression of the gene arising. Time-courses and phase portraits of the dynamics before and after the Hopf bifurcation are shown in the figure below, with large oscillations in concentrations shown in the lower plots.



Time-courses and phase portraits of the three-variable model using the equations shown in the 'Explore the model' box below. In both cases, $\mu =, \theta = 0.1, X(0) = 0, Y(0) = 0, Z(0) = 0$. *Top:* $n = 1$*, giving a stable spiral into an equilibrium. Bottom:* $n = 4$*, giving a limit cycle.*

The three-component feedback loop we have considered here demonstrates that negative feedbacks can do more than homeostasis (stability of a system around a fixed point); it can also result in sustained oscillations. Such oscillations are a central feature of many biological systems. A prominent example is the roughly 24-hour period circadian oscillatory gene expression that occurs in both plants and animals. These oscillations, which are critical for coordinating metabolic activity, etc. with the rotation of the Earth, are generated by a slight elaboration of the mechanism we have studied here. Indeed, the Goodwin oscillator provided an early model for studying the properties of circadian oscillations. Other oscillations, with a wide range of oscillatory periods, are also generated by negative feedbacks.

Explore the model

Using the Python code below, explore how sustained oscillations can emerge in the three variable negative feedback model. The dynamics are governed by the following ordinary differential equations:

$$\frac{dX}{dt} = -\mu + \frac{\theta^n}{\theta^n + Z^n}$$
$$\frac{dY}{dt} = -\mu + \frac{X^n}{\theta^n + X^n}$$
$$\frac{dZ}{dt} = -\mu + \frac{Y^n}{\theta^n + Y^n}$$

Initially we have set $n = 1$, and the resulting dynamics are for a slight spiral into the stable equilibrium. Gradually increase the value of n in steps of 0.5 to see the limit cycle emerge and grow.

Click for code

```
# Import the necessary libraries
import numpy as np
import matplotlib.pyplot as plt
from scipy.integrate import solve_ivp
# Options to make the plots the right size
plt.rcParams['figure.figsize'] = [12, 4]
plt.rcParams.update({'font.size': 16})
# Function for three variable model called 'goodwin'
def goodwin(t,x):
    # Rename the variables for ease
    X=x[0]
    Y=x[1]
    Z=x[2]
    # The ODEs
    dX = -mu*X + theta**n/(theta**n+Z**n)
    dY = -mu*Y + X**n/(theta**n+X**n)
    dZ = -mu*Z + Y**n/(theta**n+Y**n)
    return [dX,dY,dZ]
# Parameter values
mu=1
theta=0.1
n=1
# Initial conditions
X0=0
Y0=0
Z0=0
N0=[X0,Y0,Z0]
```

```
# Time-points to use
tc = np.linspace(0, 50, 1000)
# Use 'solve_ivp' to run the model
Nc = solve_ivp(goodwin, [tc[0],tc[-1]],N0, t_eval=tc)
# Plotting commands
fig, (ax1, ax2) = plt.subplots(1, 2)
ax1.plot(tc, Nc.y[0], "r", label="X")
ax1.plot(tc, Nc.y[1], "k", label="Y")
ax1.plot(tc, Nc.y[2], "b", label="Z")
ax1.set(xlabel='Time', ylabel='Concentrations')
ax1.legend()
ax1.axis([0,50,0,1])
ax2.plot(Nc.y[0],Nc.y[1],'b')
ax2.axis([0, 2, 0, 2])
ax2.set(xlabel='X', ylabel='Y')
ax2.axis([0,1,0,1])
```

Key Takeaways

- We can create a 3 variable negative feedback circuit with two positive regulatory functions and one negative.
- In a simplified system we can calculate the eigenvalues directly.
- We find limit cycles can emerge for steep enough regulation funcions, such that gene expression regularly oscillates.

Chapter references

- This model is a simplified version of that proposed by Goodwin (1965).
- The content in the *Gene networks* section is based on the unpublished *Mathematical Biology* lecture notes developed by Nick Monk.

CHAPTER 16

A two-gene toggle switch

A GENERAL MODEL FORM

In the previous genetic models we have assumed that one gene exists in isolation and its expression is controlled by its 'own' mRNA and protein concentrations. Now we will extend our model to look at two interacting genes. The central property of the circuit is that of cross-repression: the product of gene 1 represses the transcription of gene 2, and vice versa.

We will take a fairly general model set-up of two interatcing genes. To keep things a bit more simple we will assume that the two mRNAs have the same degradation rate as each other, that the two proteins have the same degradation rate as each other, and that the functions describing the two transcription and translation terms are identical, then we can represent the cross-repression circuit by the following four ODEs:

$$\frac{dm_1}{dt} = f(p_2) - \mu m_1$$
$$\frac{dp_1}{dt} = km_1 - \nu p_1$$
$$\frac{dm_2}{dt} = f(p_1) - \mu m_2$$
$$\frac{dp_2}{dt} = km_2 - \nu p_2.$$

where μ and ν are the degradation rates of the mRNA and protein, respectively, k is the per capita translation rate, and $f(p)$ is a monotonic decreasing function representing regulated transcription. Note that we have assumed such a high degree of symmetry in the equations only for notational and algebraic simplicity; the behaviour that we will study applies also to non-symmetrical models. Even with these simplifications, we are still faced with a 4-dimensional system to analyse! As we will see below, however, if we think about the problem in the right ways we can still gain considerable insight.

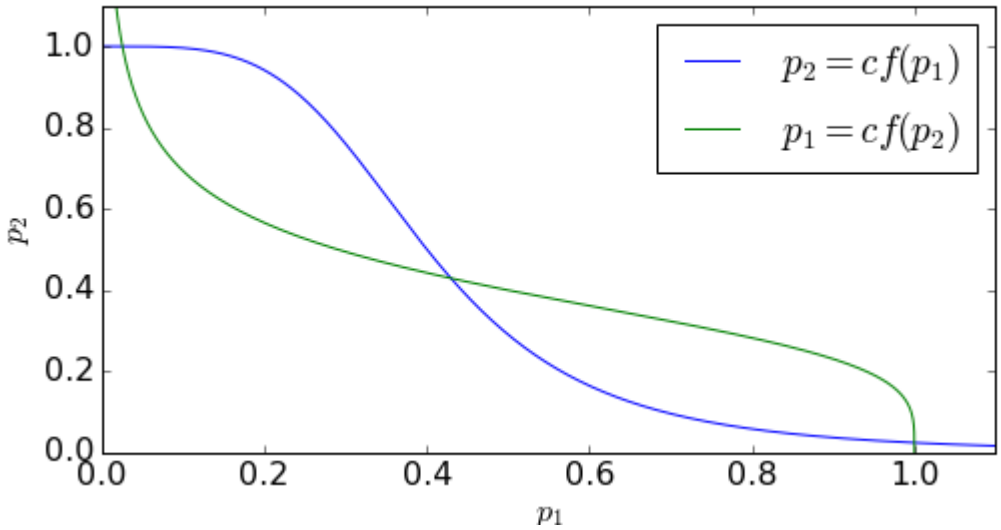
LINEAR STABILITY ANALYSIS

One particular issue with having a four-dimensional system is that it is impossible to sketch out a phase portrait. Instead we will proceed for now by linearising around the equilibria to assess stability. If we set each of the four ODEs to 0, we reach the following expresisons for equilibria,

$$\frac{\mu\nu}{k} p_1 = f(p_2)$$
$$\frac{\mu\nu}{k} p_2 = f(p_1).$$

Since μ, ν and k are positive constants, we can think of this as roughly giving us two expressions,

$p_1 = cf(p_2)$ and $p_2 = cf(p_1)$. Since $f(p_i)$ is positive but decreasing, if we take our standard sigmoidal shape of regulation function (i.e. a Hill function), we have two general scenarios that can occur here. Below we plot an example of the curves these produce with p_2 as a function of p_1.



Equilibria curves from two-gene toggle switch model.

Each curve is a continuum of points along which two of the ODEs are zero. Therefore equilibria of our system occur where these two curves intersect. As such we either have just one equilibrium (where $p_1 = p_2$ – and $m_1 = m_2$ – because of the symmetry we've assumed) or, as we have in the example here, three equilibria (one at $p_1 = p_2$ and another two either side of this). As ever, we assess the stability of this system by considering the Jacobian matrix.

Exercises

Write out the 4×4 Jacobian matrix for this system, letting $\phi_i = df(p_i)/dp_i$ at the equilibrium.

Click for solution

As before with Jacobians, we each line to represent each ODE, and each column to be the partial derivatives with respect to the relevant variable. So for this 4×4 case we have,

$$J = \begin{pmatrix} -\mu & 0 & 0 & \phi_2 \\ k & -\nu & 0 & 0 \\ 0 & \phi_1 & -\mu & 0 \\ 0 & 0 & k & -\nu \end{pmatrix},$$

We ave seen that with a 3×3 Jacobian, if it is not too complicated, we can still assess stability by writing out the characteristic equation. The same holds here with our 4×4 Jacobian. Thankfully we have a large number of 0s, and we can find that the characteristic equation reduces to,

$$\det(J - \lambda I) = (\nu + \lambda)^2(\mu + \lambda)^2 - k^2\phi_1\phi_2.$$

Similarly to last time let us take $\chi = \sqrt{\phi_1\phi_2}$. Then we are looking for solutions of,

$$(\nu + \lambda)(\mu + \lambda) = \pm k\chi,$$
$$\implies \lambda^2 + (\mu + \nu)\lambda + \mu\nu \pm k\chi = 0$$

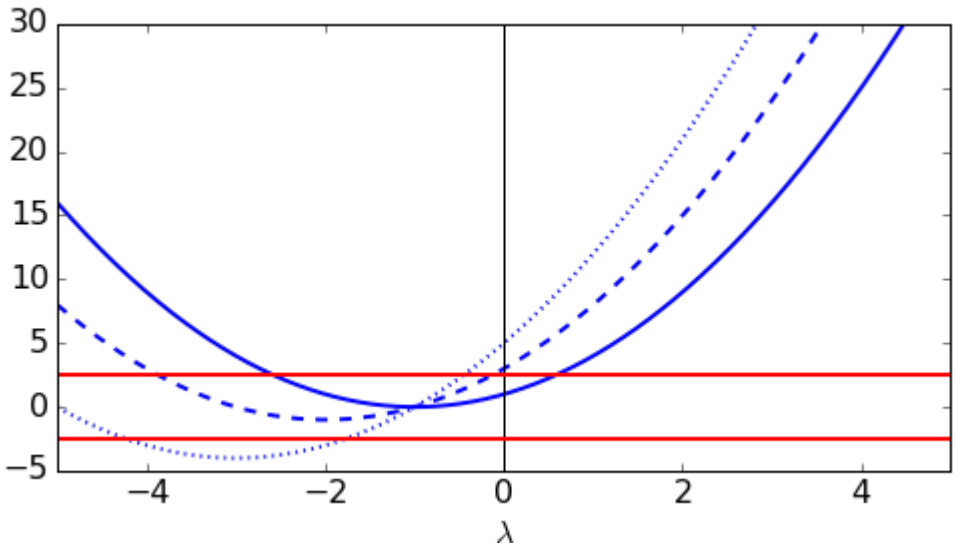
We are therefore reduced to a quadratic equation, which we can solve in the standard way, finding,

$$\lambda = \frac{-(\mu + \nu) \pm \sqrt{(\mu + \nu)^2 - 4\mu\nu \pm 4k\chi}}{2}$$

These eigenvalues may or may not be complex. The important aspect for stability is what happens to the real parts of the eigenvalues. Provided $k\chi < \mu\nu$, we must have negative real parts for all four eigenvalues (because the square root is definitely no bigger than $\mu + \nu$). However, if $k\chi > \mu\nu$ and we take the '+' term outside the square root and the '+' term inside the square root, the eigenvalue will be positive (and entirely real).

We can try and understand what is going on here a bit more by sketching out the equation $(\nu + \lambda)(\nu + \lambda) = \pm k\chi$, as below. In each case we plot $(\mu + \lambda)(\nu + \lambda)$ in blue as a function of λ and look for where it intersects the red lines for $\pm k\chi$, which will be solutions to our characteristic equation. Notice that $(\mu + \lambda)(\nu + \lambda)$ intersects the vertical axis at $\mu\nu$. We have three cases:

1. The two curves intersect twice (solid blue curve), once for $\lambda < 0$ and once for $\lambda > 0$. We can see from the plot that in this case $\mu\nu < k\chi$. The other two eigenvalues must be complex (but we know they have negative real part from above).
2. The two curves again intersect twice (dashed blue curve), but now $\lambda < 0$ at both crossing points. We can see that in this case $\mu\nu > k\chi$. Again, the other two eigenvalues must be complex but we know they have negative real part.
3. The two curves intersect four times (dotted blue curve), and in each case $\lambda < 0$. We therefore have 4 purley real eigenvalues. Some further work can reveal that the transition between cases 2 and 3 occurs at $k\chi = (\mu - \nu)^2/4$.



Solutions to the characteristic equation for the two-gene toggle switch. Blue curves give different functions of $(\mu + \lambda)(\nu + \lambda)$ *and the red lines the value of* $\pm k\chi$.

Putting this all together, then, we can either have 4 eigenvalues with negative real parts and so a stable

equilibrium (when $\mu\nu > k\chi$), or 3 eigenvalues with negative real parts but one with positive real part and so an unstable equilibrium (when $\mu\nu < k\chi$).

LINKING TO THE EQUILIBRIA

The challenge now is to work out how each of these cases for stability links to the equilibria we identified earlier. Consider the curves we sketched out in the first figure. Let us think about the respective slopes of the red and blue curves,

$$\frac{dp_2}{dp_1} = \frac{\mu\nu}{k\phi_2}$$
$$\frac{dp_2}{dp_1} = \frac{k\phi_1}{\mu\nu}.$$

In our first case (one equilibrium), we see the red curve is steeper (more negative) at the crossing point, meaning,

$$\frac{\mu\nu}{k\phi}(k\phi)^2 \implies \mu\nu > k\chi.$$

(Notice we know that $\phi_1 = \phi_2 = \phi$ at this equilibrium because of the symmetry. We also change the sign in the inequality because $\phi < 0$.) We know from above that $\mu\nu > k\chi$ means that the equilibrium is stable.

We can see that in the second case (three equilibria), at the two non-central equilibria the red line is again steeper. By the same arguments we can conclude that these two cases will both be stable (although $\phi_1 \neq \phi_2$, our definition of $\chi = \sqrt{\phi_1\phi_2}$ is enough). At the central equilibrium we now have that the blue line is steeper. This means our inequalities all reverse and we find that now $\mu\nu < k\chi$. As we found above, this means that this central equilibrium is unstable.

BIOLOGICAL SIGNIFICANCE

We often find mutually repressing genes in cells. Our modelling shows that if the cross-regulation is weak, or the gene products are stable, then the first case holds, and the two genes weakly hold each other in check, resulting in steady, intermediate expression of both genes.

However, if for some reason either (a) the cross-regulation becomes stronger (ϕ increases), or (b) the gene products become destabilised (μ and/or ν decrease), then the stalemate state can become unstable, and the second case holds. Now the stalemate state behaves locally like a saddle point, and the state of the system is driven to one of the two new equilibria. In these states, one or the other of the two genes "wins", and becomes expressed strongly, while expression of the other gene reduces to a low level. The cell has thus switched its gene expression state to either (A ON, B OFF) or (A OFF, B ON). Once one of these states is reached, the system is stable to perturbations.

Such switching of gene expression underlies the basic "fate" decisions that cells have to make during processes such as embryonic development. Since the two protein products of the genes are transcription factors, they can regulate other genes. In this way, a single toggle switch can result in the switching on or off of large collections of genes in a cell, giving coherent choices between different cell fates.

Key Takeaways

- We can model a two-gene system with a four-dimensional model.

- We can still find equilibria and determine their stability by finding the characteristic equation and using some graphical methods and reasoning.
- The two gene system can produce a *switch* whereby the system reaches one of two possible fates.

Chapter references

- The content in the *Gene networks* section is based on the unpublished *Mathematical Biology* lecture notes developed by Nick Monk.

PART V

PHARMACOKINETICS

CHAPTER 17

Single intravenous bolus dose

WHAT IS PHARMACOKINETICS?

Pharmaco-kinetics is concerned with how drugs move around the body (see chapter references). In reality this is a highly complex process, as drugs are absorbed, distributed, metabolised and eliminated through various parts of the body. We will make various assumptions that mean we can simplify these processes. In particular we will assume the body is made up of just one or two compartments through which the drugs move. We will also simplify the processes by which drugs are administered. Our focus will be on how the concentration of a particular drug in the body changes over time. We are therefore continuing with ordinary differential equations as our go-to tool, but it is worth stressing that we have moved slightly away from what we might classically call *biology* (since nothing is actually living in these models) and more into *medicine*.

A SINGLE DOSE MODEL

Let's start with the simplest (yet realistic) model we can think of, and then we will gradually build more complexity in to it. Let us assume that:

- The body can be considered as a single compartment (effectively the bloodstream),
- There is rapid infusion of the drug in to the body as a single dose, called an *intravenous bolus*,
- Only one dose of the drug is administered,
- The drug is eliminated from the body at a rate proportional to the drug concentration.

Let $C(t)$ be the concentration of the drug in the bloodstream at time t. If the rate of elimination of the drug is k, we can write,

$$\frac{dC}{dt} = -kC,$$

with some initial concentration $C(0) = C_0 > 0$. Note that we only have a negative term on the right-hand side, indicating the concentration of the drug is always decreasing. This fits with our assumptions above intuitively, since we assume no additional drug is added after $t = 0$, and so all that can happen to the drug is it is eliminated by the body.

This is a linear ordinary differential equation that is almost identical to the very first model that we met in this textbook (just with a minus sign). Using the same method of separation of variables we can find that,

$$C(t) = C_0 e^{-kt}.$$

We thus predict exponential decay of the drug concentration.

CALCULATING k

Suppose we wish to find out what the elimination rate is for a particular drug. All we need are two data-points for the concentration at known times. One of these could be the starting concentration (C_0 at $t = 0$), and suppose we take a measurement of the concentration, C_1 after t_1 hours. If we take logs of both sides of our solution we find,

$$\begin{aligned} &\ln(C_1) = \ln(C_0) - kt_1 \\ \implies &k = \frac{\ln(C_0) - \ln(C_1)}{t_1}. \end{aligned}$$

In fact, if we were to have multiple data points we simply need to know the gradient of the line of $\ln(C(t))$ vs t. Given that k is a rate it will have units t^{-1}. Hourly elimination rates for real drugs tend to be in the range $k \in [0.02, 0.4]$.

A NOTE ON C_0

We assumed above that we would know the initial concentration, C_0. This seems obvious since we know what the dose was. However, we should note that *dose* and *concentration* are not the same thing. The dose is the actual amount of the drug administered, while the concentration is the relative amount of drug in the body as a proportion of the volume of the body. We can convert between the two by noting that $C_0 = dose/volume$. For slightly complex biological reasons, the effective volume is not necessarily as simple as just calculating a patient's actual body (or blood) volume and varies from drug to drug. For simplicity, however, we will assume it is a fixed value for each drug in every patient.

DRUG HALF-LIFE

Given that we expect exponential decay of drug concentration over time, we cannot give a precise time when the concentration would be exactly zero. However, we can calculate a *half-life* for the drug based on its elimination rate. Assume we know the concentration is C_a at some time point. Then to find the half-life we need to know the time until the concentration is $C_a/2$. Using the logged solution from above, we have,

$$\begin{aligned} &\ln(C_a/2) = \ln(C_a) - kt_{1/2} \\ \implies &t_{1/2} = \frac{\ln(C_a) - \ln(C_a/2)}{k} \\ \implies &t_{1/2} = \frac{\ln(2)}{k}. \end{aligned}$$

Exercises

Suppose the half-life of a drug is known to be 3 hours, and the effective volume for the drug in a patient is 30 litres. What should the initial dose be such that after 4 hours the concentration of the drug in the patient is 2.5 mg/L?

Click for solution

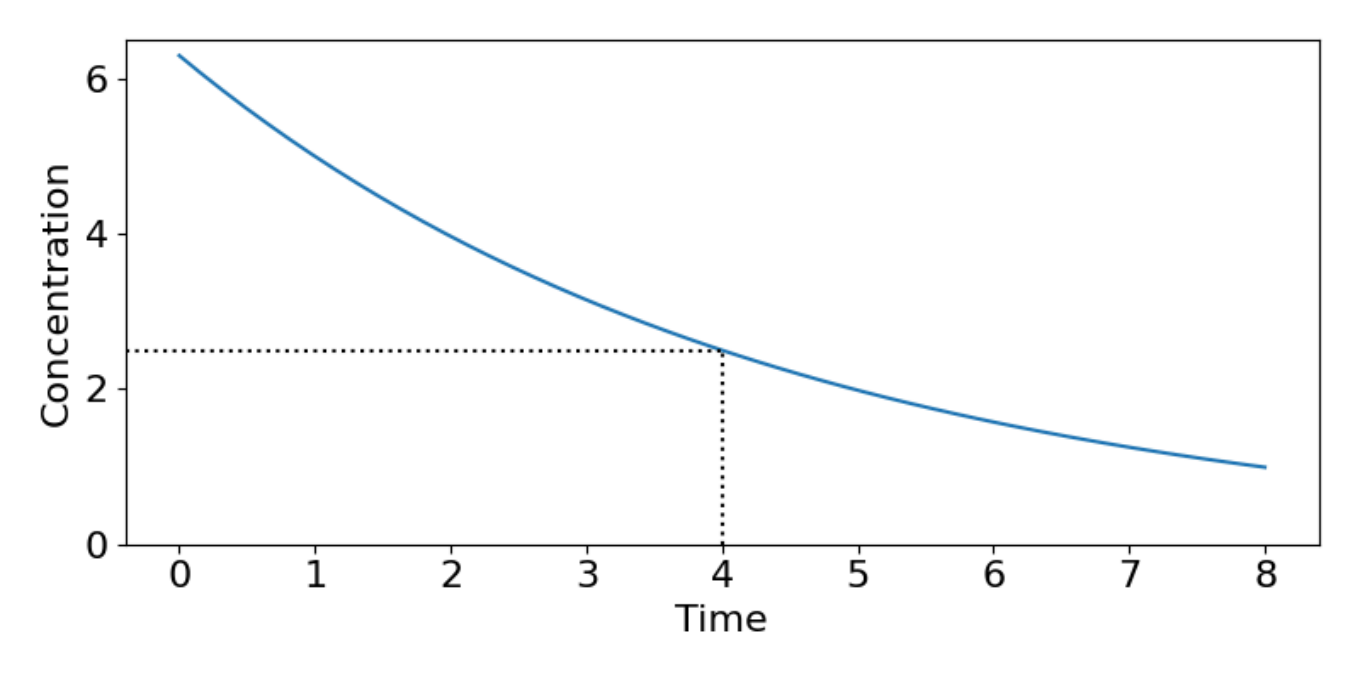
First, we can find the elimination rate, k, directly from the half-life, using,

$$t_{1/2} = \frac{\ln(2)}{k}$$
$$\implies k = \frac{\ln(2)}{t_{1/2}}$$
$$\implies k = 0.231.$$

Next we will find out what the initial concentration must have been. We know that $k = 0.231$ and that $C(t = 4) = 2.5$. We can therefore use,

$$C(t) = C_0 e^{-kt}$$
$$\implies C_0 = C(t)e^{kt}$$
$$\implies C_0 = 2.5e^{0.231 \times 4} = 6.297 \text{mg/L}.$$

The final step is to convert this to the actual dose by multiplying through by the volume to get dose $= 6.297 \times 30 \approx 190$mg. A plot of the resulting dynamics is shown in the figure below.



Drug concentration for a single intravenous bolus dose. $C_0 = 6.297$ *and* $k = 0.231$.

Key Takeaways

- Pharmacokinetics means modelling drug concentrations in the bloodstream.

- A simple model of drug decay can be derived using a linear ordinary differential equation.
- We can use the half-life of the drug concentration to parameterise the model.

Chapter references

- The content in the *Pharmacokinetics* section is based on the ebook, *Basic Pharmacokinetics* by Bourne.

CHAPTER 18

Repeated intravenous bolus doses

WHY REPEATED DOSES MIGHT BE A PROBLEM

In our first pharmacokinetic model we assumed that only one dose of the drug was given to the patient, and then thought about how its concentration decayed over time. In certain circumstances this may well be all that happens – for example you may often have taken one dose of paracetamol for a headache. However, for ongoing, chronic conditions, a patient will take repeated doses of a drug. This leads to an important question: what is the optimal dosage and time interval between doses for a particular drug? We would certainly like to ensure the patient always has a minimal amount of the drug in their system so that it is always being effective. However, we also need to ensure the patient never has too much drug in their system, avoiding overdoses.

There is a potential problem, though. In the previous model we could assume that we started with no drug in the bloodstream. After repeated doses, however, the drug will accumulate. Even if we wait a very long time for the 2nd dose, there will still be some small amount of the 1st dose left. That means that when we add on the new dose we will end up with a greater concentration than we had initially. If this keeps on adding up, there is a danger we will eventually reach dangerous levels of the drug in the bloodstream. Let's try and figure out if this really will be a problem by considering an example.

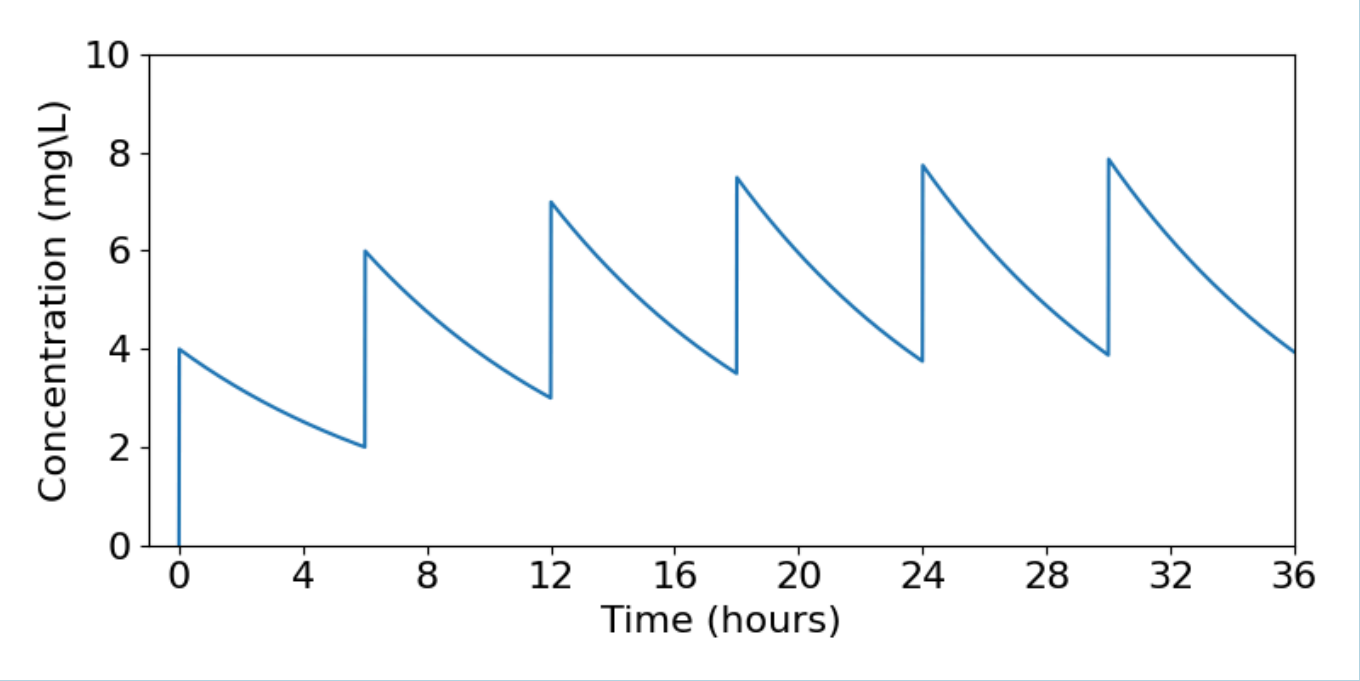
Exercises

A drug with a half-life of 6 hours is given to a patient in doses of 100mg every 6 hours, with an effective volume of 25 litres. Write out the concentrations (i) immediately *after* each dose is given and (ii) immediately *before* the next dose is given for the first 6 doses. Do these concentartions seem to be increasing without stopping or do they seem to level off?

Click for solution

Dose no.	Concentration after dose	Concentration before dose
1	4	2
2	6	3
3	7	3.5
4	7.5	3.75

The concentrations for the first four doses are shown in the table above. A full time course of the concentrations is shown in the plot below. Eventually the concentrations of the drug simply fluctuate between 8mg/L and 4mg/L. There is thus a limit to accumulation, because eventually the constant amount being added becomes equal to the density-dependent amount being removed. It therefore looks like we could design our drug dosage regime to safely oscillate between some finite maximum and minimum concentrations.



Time-course of concentrations from repeated intravenous bolus doses, with half-life of 6 hours, dose of 100mg/L and volume of 25L.

MATHEMATICAL DERIVATION

As a good scientist, you should always be sceptical of being presented with one case. Will repeated doses *always* balance out in the end like this? Or have I been sneaky in presenting a unique example to you? To check this, the key is to spot that in fact this process is given by a geometric series. Recall from our first model that if the dynamics of the drug are given by,

$$\frac{dC}{dt} = -kC,$$

then,

$$C(t) = C_0 e^{-kt}.$$

Suppose that we give a dose of the drug every τ hours. We will also label the concentrations with a subscript to denote which number dose we are considering. Thus τ hours after the first dose, but immediately before the second dose, we have,

$$C_1(\tau) = C_0 e^{-k\tau}.$$

The second dose is then administered. Immediately after this second dose (re-starting 'time since last dose' as $t = 0$) we have,

$$C_2(0) = C_1(\tau) + C_0 = C_0(1 + e^{-k\tau}).$$

This quantity is therefore the new initial condition, and the dynamics for the next τ hours will now be given by,

$$C_2(t) = C_0(1 + e^{-k\tau})e^{-kt},$$

and exactly τ hours after this second dose, but just before the third,

$$C_2(\tau) = C_0(1 + e^{-k\tau})e^{-k\tau}.$$

This pattern continues. For example immediately after and τ hours after the third dose we have, respectively,

$$C_3(0) = C_0(1 + e^{-k\tau})e^{-k\tau} + C_0 = C_0(1 + e^{-k\tau} + e^{-2k\tau}),$$
$$C_3(\tau) = C_0(1 + e^{-k\tau} + e^{-2k\tau})e^{-k\tau}.$$

This is beginning to look a little messy. To help matters, let $R = e^{-k\tau}$. Then we can re-write these two expressions as,

$$C_3(0) = C_0(1 + R + R^2),$$
$$C_3(\tau) = C_0(R + R^2 + R^3).$$

From here we can extrapolate to the concentrations after the n-th dose,

$$C_n(0) = C_0(1 + R + R^2 + \ldots + R^{n-1}) = \sum_{i=1}^{n} C_0 R^{n-1},$$

$$C_n(\tau) = C_0(R + R^2 + R^3 + \ldots R^n) = \sum_{i=1}^{n} C_0 R^n.$$

These are geometric series! A nice property of such series is that we can express this as a simple fraction as follows,

$$\begin{aligned} C_n(0) &= C_0(1 + R + R^2 + \ldots + R^{n-1}) \\ RC_n(0) &= C_0(R + R^2 + \ldots + R^n) \\ C_n(0) - RC_n(0) &= C_0(1 - R^n) \\ C_n(0) &= C_0 \frac{1 - R^n}{1 - R} = C_0 \frac{1 - e^{-nk\tau}}{1 - e^{-k\tau}}. \end{aligned}$$

By similar reasoning we can find that,

$$C_n(\tau) = C_0 R \frac{1 - R^n}{1 - R} = C_0 e^{-k\tau} \frac{1 - e^{-nk\tau}}{1 - e^{-k\tau}}.$$

This also leads to an equation for the concentration at any time, t, after the n-th dose,

$$C_n(t) = C_0 e^{-kt} \frac{1 - e^{-nk\tau}}{1 - e^{-k\tau}}.$$

We can then look at what happens in the long-term by letting $n \to \infty$ to find the minimum and maximum concentrations of the drug after many doses. Because we know that $e^{-k\tau} < 1$, we can say that $e^{-nk\tau} \to 0$ as $n \to \infty$, meaning we have,

$$C_\infty(0) = C_+ = \frac{C_0}{1 - e^{-k\tau}}$$

for the maximum concentration, and,

$$C_\infty(\tau) = C_- = \frac{C_0 e^{-k\tau}}{1-e^{-k\tau}}$$

for the minimum.

Notice that this approach works so long as $e^{-k\tau} < 1$, but this is always true for the biologically realistic assumptions that $k, \tau > 0$. So we will always reach finite maxima and minima that the concentration fluctuates between.

Exercises

Suppose that a drug has a half-life of 4.5 hours, and is given in 200mg doses every 8 hours with V=25. Clinical guidelines suggest that, in the long-term, the maximum safe concentration of the drug is 20mg/L, and that a minimum concentration of 7.5mg/L is needed for the drug to be effective. Does this drug dosage regime fit within these guidelines?

Click for solution

Using this information we can find that $C_0 = 8$mg/L, $k = 0.154$ and therefore $e^{-k\tau} = 0.292$. We can then substitute these numbers in to our expression to find,

$$C_+ = \frac{8}{1-0.292} = 11.299\text{mg/L}$$

for the maximum concentration, and,

$$C_- = \frac{8 \times 0.292}{1-0.292} = 3.299\text{mg/L}$$

for the minimum. Therefore this dosage regime is always safe (since the maximum is less than 20mg/L) but for large periods of time not effective (since considerable time will be spent with concentrations lower than 7.5mg/L).

DESIGNING A DRUG DOSAGE REGIME

In that last exercise you were told a dose amount and time interval, calculated the (long-term) maximum and minimum concentrations, and then checked back to see if they fitted some given guidelines. Now let's think about the problem a different way. Suppose we wish to design our dosage regime so that it fits the clinical guidelines exactly – that is a maximum concentration of 20mg/L and a minimum of 7.5mg/L. What *should* the dosage amount and time interval be? We already know that $k = 0.154$. First, notice that,

$$\frac{C_\infty(\tau)}{C_\infty(0)} = R = e^{-k\tau}$$

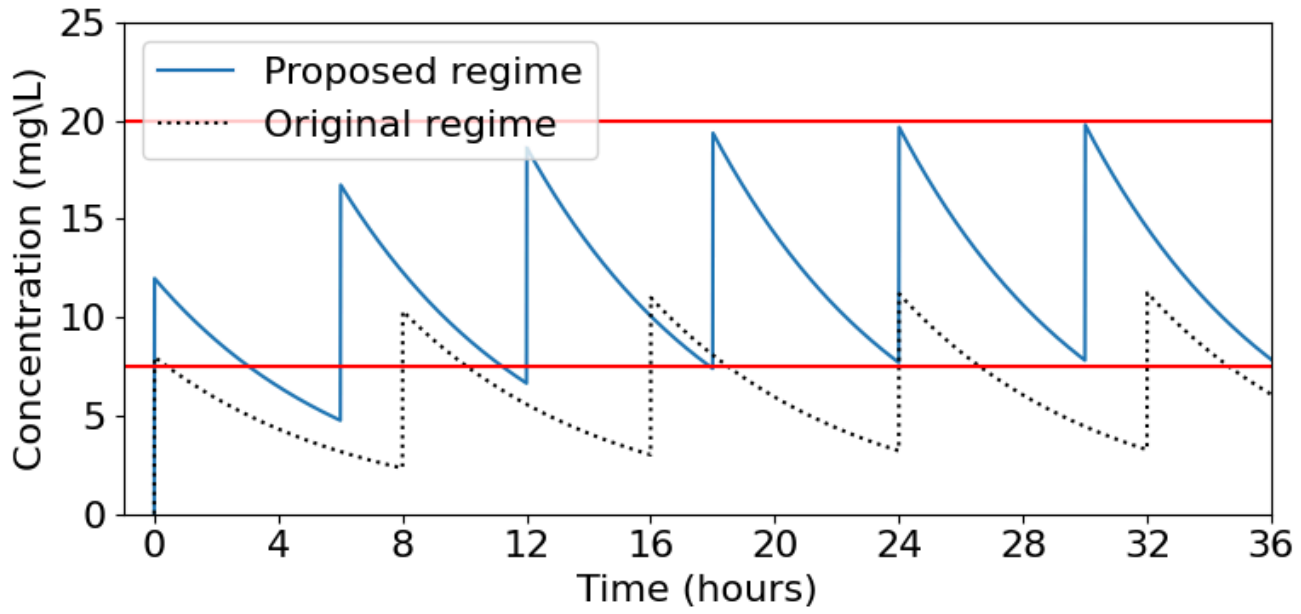
$$\implies \frac{7.5}{20} = e^{-0.154\tau}.$$

By taking logs and re-arranging this we get $\tau = -\ln(0.375)/0.154 = 6.369$h. Let's round this to the more realistic time frame of 6 hours.

Then by considering the maximum we have,

$$
\begin{aligned}
& C_0 = C_\infty(0)(1 - e^{-k\tau}) \\
\implies & C_0 = 20(1 - e^{-0.154\times 6}) \\
\implies & C_0 = 12.061\text{mg/L}.
\end{aligned}
$$

Given that $V = 25$, this means a dose of 301.525mg, which we might round down (because we definitely wish to avoid overdosing) to the more realistic 300mg. Thus, based on our knowledge of the drug, we would recommend a dose of 300mg every 6 hours for maximum, yet safe, efficacy. (A quick check shows that our rounding means this actually gives a maximum of 19.90mg/L and a minimum of 7.90mg/L.) A time course of this proposed regime is shown in the figure below, along with the original regime which we can see was much less efficient.



Time-courses of repeated intravenous bolus doses with the maximum and minimum suggested doses marked with red lines. For the original regime we had a dose of 200mg/L and a dose interval of 8 hours. For the proposed regime we have a dose of 300mg/L and a dose interval of 6 hours. In both cases the half-life is 4.5 hours and the blood volume is 25L.

LOADING DOSE AND MAINTENANCE DOSE

A downside of the approach we have just described is that it might take a long time for the drug concentration to reach these long-term maximum and minimum values. What we might consider doing, then, is to initiate treatment with a larger *loading dose* that takes the concentration to (or near) the maximum concentration, with following dosages given as a *maintenance dose*. The loading dose should come as close as possible to immediately reaching the maximum concentration. For our example above, then, we might take a loading dose of $20 \times 25 = 500$mg. The maintenance dose and timing would then be identical to previously (300mg every 6 hours) since this regime is already balanced to fluctuate between the maximum and minimum concentrations.

We can still write a general solution by considering the dynamics over subsequent doses. Calling the loading dose C_L and the maintenance dose C_M, some time t after the fist dose we have,

$$C_1(t) = C_L e^{-kt}.$$

Immediately after the second dose at time τ, re-starting time as $t = 0$, we have,

$$C_2(0) = C_1(\tau) + C_M = C_M + C_L e^{-k\tau},$$

and the dynamics for the next τ hours will now be given by,

$$C_2(t) = [C_M + C_L e^{-k\tau}]e^{-kt}.$$

Then, immediately after the third dose we have,

$$C_3(0) = C_M + [C_M e^{-k\tau} + C_L e^{-2k\tau}].$$

Note that we can then re-write this as,

$$C_3(0) = C_M[1 + e^{-k\tau} + e^{-2k\tau}] + (C_L - C_M)e^{-2k\tau}.$$

and for any time t after this we would have,

$$C_3(t) = C_M[1 + e^{-k\tau} + e^{-2k\tau}]e^{-kt} + (C_L - C_M)e^{-2k\tau}e^{-kt}.$$

We can see the first term is exactly the same as we had previously and will again be given by a geometric sum. There is then the addition of a second term to do with the extra loading dose. The final expression for the dynamics in this case are thus given by,

$$C_n(t) = C_M e^{-kt}\frac{1 - e^{-nk\tau}}{1 - e^{-k\tau}} + (C_L - C_M)e^{-k[(n-1)\tau + t]}.$$

Explore the model

Use the Python code below to explore the model of repeated intravenous bolus doses with a loading dose and maintenance dose. The initial code shows a maximum and minimum concentration as horizontal bars and the dynamics of the drug concentration for some loading dose, maintenance dose and time interval. Using the approach above, determine a dosing regime that will fit the given maximum and minimum concentrations as perfectly as you can. Try testing out different values of the parameters to see how different drug regimes will look.

Click for code

```
# Import the necessary libraries
import numpy as np
import matplotlib.pyplot as plt
# Options to make the plots the right size
plt.rcParams['figure.figsize'] = [6, 4]
plt.rcParams.update({'font.size': 16})
# Parameter values
max=25 # The maximum safe concentration
min=10 # The minimum effective concentration
cl=300 # Loading dose
cm=300 # Maintenance dose
v=25 # Bloodstream volume
tau=6 # Time between doses
k=0.2 # Decay rate
```

```
C_l=cl/v #Loading dose concentration
C_m=cm/v #Maintenance dose concentration
maxtime=tau #Time we want to run the dynamics for each dose
maxdose=6 #How many doses we will follow
steps=100 #How many time steps per hour to plot
#Create arrays to hold the concentration and time data of a suitable size.
# Assume we will take 100 time steps per hour (then add 1 to include t=0).
C=np.zeros(maxtime*maxdose*steps+1)
t=np.zeros(maxtime*maxdose*steps+1)
#loop over n=maxdose doses
for n in range(1,maxdose+1):
    #loop over time steps for each dose
    for i in range(1,maxtime*steps+1):
        t[(n-1)*maxtime*steps+i]=i/steps+(n-1)*maxtime
        Cload=(C_l-C_m)*np.exp(-k*((n-1)*maxtime+t[i]))
        Cmain=C_m*np.exp(-k*t[i])*(1-np.exp(-(n*k*maxtime)))/(1-np.exp(-k*maxtime))
        C[(n-1)*maxtime*steps+i]=+(Cload+Cmain)
#Create plot
plt.plot(t,C)
#Add the horizontal lines for the max and min
plt.axhline(y=min, color='r', linestyle='-')
plt.axhline(y=max, color='r', linestyle='-')
#Control various plot properties
plt.ylim(0,30)
plt.xlim(-1,maxtime*maxdose)
plt.xticks([0,4,8,12,16,20,24,28,32,36])
plt.xlabel('Time (hours)')
plt.ylabel('Concentration (mg\L)')
```

Key Takeaways

- Repeated doses of a drug can be modelled as a geometric series.
- Eventually there will be a balance between the maximum and minimum concentrations, and we can use this to design dosage

regimes to fit clinical guidelines.

- We can also use a loading dose to avoid low concentrations at early time points.

Chapter references

- The content in the *Pharmacokinetics* section is based on the ebook, *Basic Pharmacokinetics* by Bourne.

CHAPTER 19

Single and repeated oral doses

DEVELOPING A 2-COMPARTMENT MODEL

So far we have assumed that as soon as the drug is administered it is immediately present in the bloodstream since it is administered intravenously. However, many drugs are administered in other ways, such as orally through tablets. In this case the drug will first reach the gastrointestinal (GI) tract where it dissolves and is gradually absorbed into the bloodstream. In this case we need to include two compartments in our model – one for the amount of the drug in the GI tract (which we might denote by X_G), and one for the amount of the drug in the bloodstream (X_B). Notice that I have used the word *amounts* here, as the concentration would have little meaning in the GI tract.

Let's initially assume that there is a one-off dose of the drug. The amount of the drug in the GI tract cannot be increased, and simply reduces as it is absorbed in to the bloodstream. Hence, we can describe the dynamics of this first compartment as,

$$\frac{dX_G}{dt} = -aX_G.$$

For the second compartment we can assume all the drug that leaves the GI tract arrives in the bloodstream, and this then decays as before as the body uses up the drug. Hence, the second equation will be,

$$\frac{dX_B}{dt} = V\frac{dC}{dt} = aX_G - kX_B = aX_G - kVC.$$

Even before deriving the solutions to these equations, we can get a good qualitative picture of what might happen. The amount of drug in the GI tract will decay exponentially down towards zero. Initially, the drug concentration in the bloodstream will be very low, suggesting little being lost due to decay, but the amount being absorbed in to the bloodstream would be relatively high, meaning the concentration will initially increase. As time goes on, the amount of drug left in the GI tract will decrease until a point is reached that the absorption of new drug is less than the decay of existing drug, and the bloodstream concentration will reduce. Eventually we would expect the bloodstream concentration to also approach zero.

Let's show this formally. We can solve this pair of equations in turn. The first is in a fairly simple form and just yields,

$$X_G(t) = X_G(0)e^{-at}.$$

We can then substitute this in to the second equation, to give,

$$\frac{dX_B}{dt} = V\frac{dC}{dt} = aX_G(0)e^{-at} - kVC.$$

This can be re-arranged in to the form,

$$\frac{dC}{dt} + kC = \frac{aX_G(0)}{V}e^{-at}.$$

Written in this form, we can see it is possible to solve this equation by use of an integrating factor.

Exercises

Show that, given the initial bloodstream concentration is 0, the explicit solution for the concentration is,

$$C(t) = \frac{aX_G(0)}{V(k-a)}(e^{-at} - e^{-kt}).$$

Click for solution

The integrating factor here will be e^{kt}. We multiply through every term by this to get,

$$e^{kt}\frac{dC}{dt} + e^{kt}kC = \frac{aX_G(0)}{V}e^{-at}e^{kt},$$
$$\frac{d}{dt}e^{kt}C = \frac{aX_G(0)}{V}e^{(k-a)t}.$$

Integrating both sides then gives,

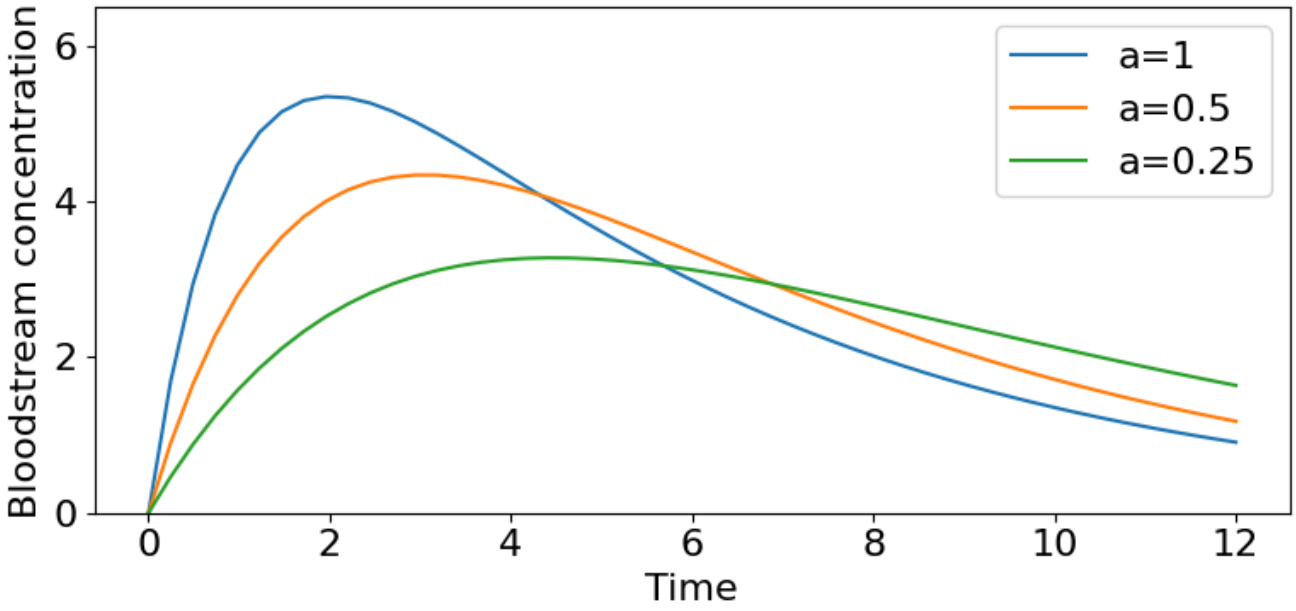
$$e^{kt}C = \frac{aX_G(0)}{V(k-a)}e^{-at}e^{kt} + B$$
$$\implies C(t) = \frac{aX_G(0)}{V(k-a)}e^{-at} + Be^{-kt}.$$

where B is the constant of integration. Then we can use that the initial concentration in the bloodstream should be 0, meaning $C(0) = 0$, to find $B = -aX_G(0)/(V(k-a))$. This can then be substituted back in to find,

$$C(t) = \frac{aX_G(0)}{V(k-a)}(e^{-at} - e^{-kt}),$$

as required.

The resulting concentration over time for some chosen parameter values is shown in the figure below. (Note, there is one special case we need to consider separately when $a = k$, which is left as an optional exercise).



Timecourses of the bloodstream concentrations of an orally-administered drug for 3 different rates of absorption. In all cases $k = 0.2, V = 25, X_G(0) = 200$.

Consider our general solution, as shown in the figure. Notice that in theory the maximum possible concentration in the bloodstream would be $X_G(0)/V$ (which here would be 8mg/L) but due to the balance of absorption and decay this value is never reached (by some distance). What is the maximum concentration then? This peak is where the curve of the concentration becomes flat, which is by definition when $dC/dt = 0$. Hence we can find when it occurs by finding,

$$\begin{aligned}
&\frac{aX_G(0)}{V}e^{-at} - kC = 0 \\
&\implies \frac{aX_G(0)}{V}\left[e^{-at} - \frac{k}{k-a}(e^{-at} - e^{-kt})\right] = 0 \\
&\implies ke^{kt} - ae^{-at} = 0 \\
&\implies \ln\left(\frac{k}{a}\right) = (k-a)t \\
&\implies t = \ln\left(\frac{k}{a}\right)\frac{1}{k-a}.
\end{aligned}$$

The faster the drug is absorbed into the bloodstream, the faster the concentration increases and the higher levels it reaches. However, this is at a cost of a faster loss of the drug as well. Plugging the values of a in to the formula above, we find the peak concentration moves from 5.35mg/L at ~2 hours at the highest absorption rate, to 3.28mg/L at ~4.5 hours for the slowest rate of absorption.

REPEATED ORAL DOSES

In this initial model we assumed a single oral dose. But, much like the intravenous bolus case previously, we will often give repeated doses of a drug. How will these concentrations change over time? We would probably expect to see curves rather like the previous case, with accumulation of the drug until some balance of dose and decay is achieved. In fact, we can see that for the amount of drug in the GI tract, $X_G(t)$, we will get exactly

the same equations as for the intravenous bolus case. For the concentration of the drug in the bloodstream it is a bit more complicated, so we will think about things a little differently.

Let us assume that a dose is given every τ hours and let t be the time since the last dose was given (meaning we always have $0 < t < \tau$). Sometime after the first dose, but before the second dose, the concentration of the drug, from above, is,

$$C(t) = \frac{aX_G(0)}{V(k-a)}(e^{-at} - e^{-kt}).$$

Now, sometime after the second dose the concentration will be made up of some contribution from the first dose plus a contribution from the second dose, such that,

$$C(t) = C_1(t) + C_2(t) = \frac{aX_G(0)}{V(k-a)}(e^{-a(t+\tau)} - e^{-k(t+\tau)}) + \frac{aX_G(0)}{V(k-a)}(e^{-at} - e^{-kt}).$$

Factoring out some terms and noting that $e^{-a(t+\tau)} = e^{-at}e^{-a\tau}$ gives,

$$C(t) = \frac{aX_G(0)}{V(k-a)}(e^{-at}(1 + e^{-a\tau}) - e^{-kt}(1 + e^{-k\tau})).$$

Extrapolating we can see that after n doses we will have,

$$C(t) = \frac{aX_G(0)}{V(k-a)}(e^{-at}(1 + e^{-a\tau} + \ldots + e^{-(n-1)a\tau}) - e^{-kt}(1 + e^{-k\tau} + \ldots + e^{-(n-1)k\tau})).$$

We can see that there are two geometric series here, which will mean we can re-write this as,

$$C(t) = \frac{aX_G(0)}{V(k-a)}\left(e^{-at}\left(\frac{1-e^{-na\tau}}{1-e^{-a\tau}}\right) - e^{-kt}\left(\frac{1-e^{-nk\tau}}{1-e^{-k\tau}}\right)\right).$$

As before, we can find the maximum and minimum values by considering what happens as $n \to \infty$ and noting that the minimum concentration will always be as the new dose is taken, i.e. $t = 0$. Then,

$$C_\infty(0) = \frac{aX_G(0)}{V(k-a)}\left(\left(\frac{1}{1-e^{-a\tau}}\right) - \left(\frac{1}{1-e^{-k\tau}}\right)\right).$$

It may be that we can we assume that by this point there is negligible new drug being absorbed in to the body (in the case that the previous dose has virtually run out), and in that case we can take $e^{-a\tau} \approx 0$ and therefore,

$$C_\infty(0) \approx \frac{aX_G(0)}{V(k-a)}\left(-\frac{e^{-k\tau}}{1-e^{-k\tau}}\right).$$

Calculating the maximum concentration is rather more tricky in this case as it is not simply at the start or end of the dosing interval. What we can do is to find the average concentration of the drug during one dose period (τ), defined as the area under the curve for one dose period at steady state,

$$\bar{C} = \frac{1}{\tau}\int_0^\tau C_{ss}(t)dt.$$

In fact we can make things easier by noting that this is equal to the whole area under the curve of the first dose on its own,

$$\bar{C} = \frac{1}{\tau}\int_0^\tau C_{ss}(t)dt = \frac{1}{\tau}\int_0^\infty C_1(t)dt.$$

Computing this integral we find,

$$\begin{aligned}\bar{C} &= \left(\frac{1}{\tau}\right)\left(\frac{aX_G(0)}{V(k-a)}\right)\int_0^\infty (e^{-at}-e^{-kt})dt \\ &= \left(\frac{1}{\tau}\right)\left(\frac{aX_G(0)}{V(k-a)}\right)\left[\frac{-e^{-at}}{a}+\frac{e^{-kt}}{k}\right]_0^\infty \\ &= \left(\frac{1}{\tau}\right)\left(\frac{aX_G(0)}{V(k-a)}\right)\left(\frac{-1}{a}+\frac{1}{k}\right) \\ &= \frac{X_G(0)}{Vk\tau}.\end{aligned}$$

This average concentration can be useful in its own right. For instance, notice this means that a low dose at short intervals will give an equivalent average concentration as a high dose at long intervals. It can also be used to construct a crude approximation of the maximum concentration using $C_+ = \bar{C} + (\bar{C} - C_-)$.

EXAMPLE PROBLEMS

Suppose an oral drug is prescribed for a patient with $k = 0.25, a = 0.4, V = 25$ and a dose of 400mg is taken every 6 hours. Suppose that guidelines state the maximum safe concentration of the drug is 10mg/L and the minimum effective concentration is 5mg/L. Is this regime safe and effective?

Let us start by finding the minimum and maximum concentrations reached after repeated doses. If we do not make the assumption that absorption is negligible when the new dose is taken, we have,

$$C_- = \frac{aX_G(0)}{V(k-a)}\left(\left(\frac{1}{1-e^{-a\tau}}\right)-\left(\frac{1}{1-e^{-k\tau}}\right)\right).$$

Substituting in the values we have this gives, $C_- = 8.0$mg/L. We can also find the average concentration at steady-state using,

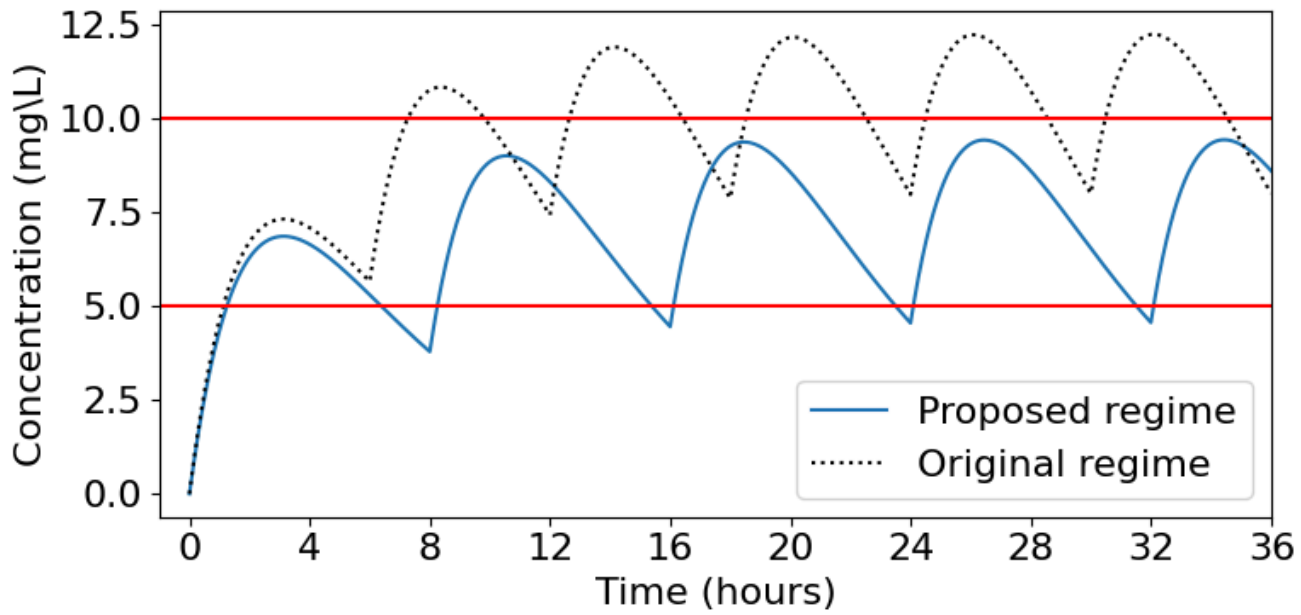
$$\bar{C} = \frac{X_G(0)}{Vk\tau},$$

which gives $\bar{C} = 10.67$mg/L and we can use this to *estimate* the maximum concentration as $C_+ = 13.3$ mg/L (the plot below reveals this is in fact an over-estimate). Comparing these values to the guidelines, it is clear that this regime would reach doses which are much too high. Let us therefore devise a new regime.

Firstly, we can find that $\bar{C} = (C_+ - C_-)/2 = 7.5$mg/L. Substituting in to the formula for $\bar{C}$ we can say,

$$X_G(0) = 46.9\tau.$$

We can substitute this back in to the equation for the minimum concentration, however there is no easy way to solve this analytically. What we can do is to solve it numerically. To do this we could plot the right-hand side for varying τ and look for where it equals 5. Doing this we find that the optimum timing is $\tau = 7.2$ hours, which means $X_G(0) = 332$mg. We might well look to take a more realistic value of $\tau = 8$ hours, giving $X_G(0) = 375$mg. The resulting dynamics for both the original and proposed regimes are shown below, revealing that our proposed regime indeed does a much better job of fitting the guidelines.



Time-courses for the original and proposed dosage regimes of a repeated oral dose drug, showing the maximum and minimum suggested concentrations as red horizontal lines. In both cases $k = 0.25, a = 0.4, V = 25$*. In the original regime the dose was 400mg every 6 hours. In the proposed regime the dose is 375mg every 8 hours.*

Key Takeaways

- For an orally-administered drug we need two compartments – the GI tract and the bloodstream.
- With one dose, the bloodstream concentration initially increases, peaks and then heads towards zero.
- With multiple doses, we again reach a balance in the long-term between maximum and minimum concentrations in both compartments.

Chapter references

- The content in the *Pharmacokinetics* section is based on the ebook, *Basic Pharmacokinetics* by Bourne.

CHAPTER 20

A two compartment bolus model

A BLOODSTREAM-TISSUE MODEL

For our final chapter (well done for making it!) we will return to the scenario of intravenous bolus doses. In our earlier models we assumed the body acted as a single compartment for the drug, basically considering the drug concentration in the bloodstream. When we looked at the orally administered drug models we assumed there was now a second compartment where the drug had to enter the body (the GI tract) before it could move on to the bloodstream. In that case things simplified because the dynamics of the first compartment were very simple – exponentially decreasing to zero – and only concentrate on what happened in the bloodstream.

Now we will consider a more complicated case where there are two compartments, but the drug can now move in and out of both of those compartments. A biological example might be the bloodstream as our main compartment, and tissue as the second compartment. We will assume the drug is again administered by an intravenous bolus and so appears immediately in the bloodstream. The concentration of the drug in the bloodstream will again be given by C, and the concentration in the tissues will be given by D. While in the bloodstream the drug is used up and is eliminated. The drug can also pass in to tissues at some rate a and can pass back from tissues to the bloodstream at rate b. We will assume the drug is not eliminated while in the tissue. Our mathematical model will thus look like,

$$\frac{dC}{dt} = -kC - aC + bD$$
$$\frac{dD}{dt} = aC - bD.$$

SOLVING A SYSTEM OF LINEAR EQUATIONS

Since both equations explicitly depend on both variables, we cannot handle this case in quite the same way as the previous pharmacokinetic models. We also cannot solve the system one equation at a time – i.e. apply an integrating factor to solve one, then substitue that solution into the second. However, we have a well established toolbox for dealing with a system of linear ordinary differential equations. I have not provided any explicit background review material for these methods, but will try to take the explanations quite slowly.

First note that we can write this out in matrix form, that is,

$$\frac{d}{dt}\begin{pmatrix} C \\ D \end{pmatrix} = \begin{pmatrix} -k-a & b \\ a & -b \end{pmatrix}\begin{pmatrix} C \\ D \end{pmatrix}.$$

For a system like this we expect to see solutions of the form,

$$\begin{pmatrix} C(t) \\ D(t) \end{pmatrix} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} e^{\lambda_1 t} + \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} e^{\lambda_2 t}.$$

This has become an *eigenvalue/eigenvector problem* since, substituting these desired solutions into our matrix equation we have,

$$\lambda_1 \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} e^{\lambda_1 t} = \begin{pmatrix} -k-a & b \\ a & -b \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} e^{\lambda_1 t},$$

and similarly for λ_2.

To solve such a problem we first find the eigenvalues. Calling the matrix of parameters M, we set $\det(M - \lambda I_2) = 0$, where I_2 is the two-dimensional identity matrix, which leads to the characteristic equation,

$$\lambda^2 + (k + a + b)\lambda + kb = 0.$$

This does not factorise particularly nicely. Let us simplify things a little by choosing some new constants α and β such that,

$$\alpha + \beta = k + a + b$$
$$\alpha\beta = kb.$$

This helps as it means the characteristic equation is now,

$$\lambda^2 + (\alpha + \beta)\lambda + \alpha\beta = (\lambda + \alpha)(\lambda + \beta) = 0,$$

for which the solutions are just $\lambda = -\alpha$ and $\lambda = -\beta$. If and when the time comes to find out what values α and β take, we would need to substitute them in to the quadratic formula to find,

$$\alpha, \beta = \frac{(\alpha + \beta) \pm \sqrt{(\alpha + \beta)^2 - 4\alpha\beta}}{2},$$

and we have defined both $\alpha + \beta$ and $\alpha\beta$ in terms of the parameters a, b and k that we do know.

We have therefore narrowed down our general solution to be,

$$\begin{pmatrix} C(t) \\ D(t) \end{pmatrix} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} e^{-\alpha t} + \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} e^{-\beta t}.$$

To find the eigenvectors we substitute each eigenvalue back in and solve,

$$\begin{pmatrix} -k-a-\lambda & b \\ a & -b-\lambda \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Let us take $\lambda = -\alpha$. Using the bottom line this gives the equation,

$$au_1 + (\alpha - b)u_2 = 0 \implies au_1 = (b - \alpha)u_2.$$

Thus for some constant, κ_u, taking $u_1 = \kappa_u(b - \alpha)$ and $u_2 = \kappa_u a$ will always satisfy this. (An aside: you might wonder why we used the bottom line here and not the top. When dealing with these eigenvector problems we usually expect to get basically the same expression whichever line we use, but here it is not obvious that is true. However, by the time you get to substituting in the initial conditions it turns out that you do indeed get the exact same final solutions – feel free to check!). The solution for $\lambda = -\beta$ will look almost identical. We can then update our final solution for the concentration of drug in the bloodstream as,

$$\begin{pmatrix} C(t) \\ D(t) \end{pmatrix} = \kappa_u \begin{pmatrix} b - \alpha \\ a \end{pmatrix} e^{-\alpha t} + \kappa_v \begin{pmatrix} b - \beta \\ a \end{pmatrix} e^{-\beta t}.$$

We can determine the values of the two remaining constants, κ_u and κ_v by substituting in initial conditions. We know that at $t = 0$ the patient has just received the dose, so the concentration of the drug in the bloodstream should be at its maximum (i.e. the dose adjusted by the effective volume) and the concentration in the tissue should be zero. These two conditions tell us,

$$C(0) = C_0 = \kappa_u(b - \alpha) + \kappa_v(b - \beta)$$
$$D(0) = 0 = \kappa_u a + \kappa_v a.$$

The second expression tells us that we have $\kappa_u = -\kappa_v$. Substituting this into the first expression gives,

$$\begin{aligned} C_0 &= \kappa_u(b-\alpha) - \kappa_u(b-\beta) \\ &= \kappa_u(-\alpha+\beta) \\ \implies \kappa_u &= \frac{C_0}{\beta-\alpha}. \end{aligned}$$

We can therefore arrive at our final solution for the concentration of the drug in the bloodstream as,

$$C(t) = C_0\left[\left(\frac{\alpha-b}{\alpha-\beta}\right)e^{-\alpha t} - \left(\frac{\beta-b}{\alpha-\beta}\right)e^{-\beta t}\right].$$

This is the sum of two negative exponentials, suggesting eventually the concentration will tend to zero, as we should expect. The final solution for the drug concentration in the tissue can be found similarly.

Exercises

Suppose a dose of 200mg is given to a patient whose bloodstream has $V = 10$L. The drug is eliminated at rate $k = 0.15$, and the transition rates are found to be $a = 0.2, b = 0.15$. What is the concentration of the drug in the bloodstream 4 hours after it has been administered?

Click for solution

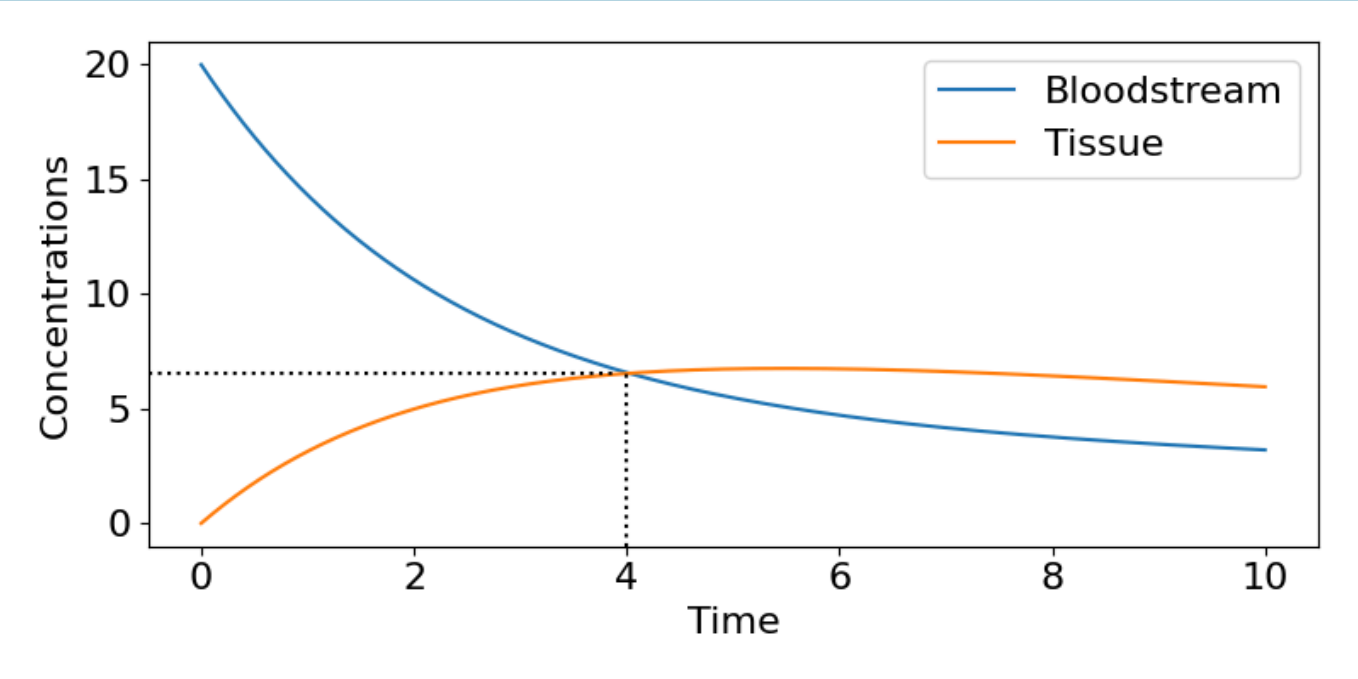
To substitute these values in to our general solution, there are a few values we need to calculate first. These are,

$$\begin{aligned} C_0 &= dose/V = 200/10 = 20 \\ \alpha+\beta &= k+a+b = 0.5 \\ \alpha\beta &= kb = 0.0225 \\ \alpha &= \frac{1}{2}\left((\alpha+\beta) + \sqrt{(\alpha+\beta)^2 - 4\alpha\beta}\right) = 0.45 \\ \beta &= \frac{1}{2}\left((\alpha+\beta) - \sqrt{(\alpha+\beta)^2 - 4\alpha\beta}\right) = 0.05. \end{aligned}$$

We can therefore write down our solution as,

$$C(t) = 20\left[0.75e^{-0.45t} + 0.25e^{-0.05t}\right].$$

All that remains is to substitute in $t = 4$ to find that after 4 hours the concentration of the drug will be 6.57mg/L. The plot below shows the full dynamics. It also shows that, purely by chance, 4 hours is roughly the time when the concentrations of the drug in the bloodstream and tissue are equal.



Plot of the concentrations in both the bloodstream and tissue for the 2 compartment model, with $k = 0.15, a = 0.2, b = 0.15, V = 10L$ *and a dose of 200mg.*

Key Takeaways

- We can extend a model to include more compartments, such as tissues, by having more variables.
- Systems of linear ODEs can be solved to find explicit solutions for each variable.
- With a single dose, the concentrations in all compartments will eventually decay towards zero.

Chapter references

- The content in the *Pharmacokinetics* section is based on the ebook, *Basic Pharmacokinetics* by Bourne.

PART VI

BACKGROUND REVIEWS

CHAPTER 21

Phase portraits

A GRAPHICAL ANALYSIS

In analysing a model we would often like to visualise the dynamics of a system. Most commonly we might do this by plotting the time-courses of our variables, showing how the density of our populations change over time. If we have two populations – as we regularly do in this course – we'd therefore be plotting two curves together, which can make such plots a bit messy, certainly if we want to look at different initial conditions. An alternative is to sketch a *phase portrait*. In this plot we leave time implicit, and instead plot how the two densities change together.

Suppose we have two populations with densities X and Y for which we have ordinary differential equations describing their dynamics. We can imagine a plot that takes X and Y as the two axes. We could then mark on the two densities at their initial values, $X(0)$ and $Y(0)$ as a single point on the plot. We might then go forward in time a little and mark on a second point. We can then continue moving forward in time and adding on points, then 'join the dots' to form a *trajectory*, showing how the two densities change over time.

This approach assumes we know the actual densities at various time-points, but as we have seen, for most non-linear models we do not have that luxury without using numerical solvers in a programming package. However, we can get an idea of the *qualitative* behaviour by sketching certain details onto a phase portrait.

ALGORITHM

The basic algorithm for construcing a phase portrait is as follows:

1. *Draw axes of the two variables.*
   We can only really sketch the phase portrait for two-dimensional systems. Usually we'd have the variable for our first ODE on the horizontal axis and the second on the vertical axis, but sometimes it makes sense to go the other way around (particularly if the nullclines are awkward for one of the two ODEs – see step 3).
2. *Determine how much of the phase plane is biologically feasible.*
   An advantage of mathematical biology models is that we rarely need to worry about negative densities, so we can usually just draw the upper-right quadrant of the plane. Sometimes we can reduce it even more – see the SIR models in chapters 4 and 5 for examples
3. *Calculate nullclines and draw them on your plot.*
   Nullclines are curves on your plot along which *one* of the ODEs is equal to zero. In turn, set each ODE to zero and (assuming you placed X on your horizontal axis) re-arrange it into the form $Y = f(X)$. This will hopefully give you a curve you can sketch, though may occasionally look unpleasant. Because the ODE is zero along this line, trajectories must cross these nullclines either vertically or horizontally depending on which ODE it came from – again assuming X is on the horizontal axis, the

nullcline that gives $dX/dt = 0$ must be crossed vertically, since the densities should not be changing in the X direction.

4. *Mark on equilibria where two (different) nullclines cross.*
   Remember that each nullcline is where *one* of the ODEs is zero. Therefore, where they intersect it must be that *both* of the ODEs are zero, and that is the definition of an equilibrium. Mark these points clearly.
5. *Work out the qualitative directions of travel in each of the regions separated by the nullclines and draw arrows on your phase portrait to show these direction fields.*
   This is often the most challenging part. The nullclines divide the plot up into regions where each ODE is either positive or negative. We don't especially mind about their actual values, just what the sign is. In each region, look at the ODEs and try and infer whether each ODE is positive or negative. You could do this by choosing helpful values to plug in ("what if X and Y are both really small?") or by thinking about the nullcline equations. Then draw on horizontal/vertical arrows in each region corresponding to this sign. For example, if X is on the horizontal axis, then $dX/dt > 0$ means X is increasing, so trajectories move to the right.
6. *Optionally use linear stability analysis to get a clearer picture of behaviour around the equilibria.*
   If we want more detail, we can apply linear stability analysis to work out whether each equilibrium is stable or unstable, or a node or a cycle, etc. I generally do not worry too much about this step unless I am really unsure how to draw on the trajectories.
7. *Sketch on some sample trajectories.*
   Using all of the rules you have just found, sketch on a couple of trajectories. Choose some starting point, then look at what direction you must be travelling in from your direction fields. Draw your curve in that direction until you come to a nullcline. Make sure you cross the nullcline in the correct horizontal or vertical direction, then look at what the direction is in your new region of the phase portrait and continue your line. Most of the time you will end up moving towards an equilibrium point, where if you have used step 6 you will have more detail of how to finish the trajectory, though often it will be obvious from the way the nullclines and direction fields are arranged.

If you follow these steps, you should be able to construct any of the phase portraits needed in this course and beyond.

CHAPTER 22

Linear stability analysis

INTRODUCTION

We know that if an ordinary differential equation is equal to zero for some density, that is an *equilibrium* of the system, meaning that if we reach that exact density, the model would say we stay at that density forever. Two important and related questions are (1) "are we are ever likely to reach that equilibrium?" and (2) "if we move a small distance away from that equilibirum will we be pulled back towards it?" These questions are about the *stability* of an equilibrium.

ONE-DIMENSIONAL SYSTEMS

Suppose we have some population with density X whose dynamics are governed by some ODE dX/dt. We can imagine sketching dX/dt as a function of X. Where the curve passes through zero we have an equilibrium. How can we tell if that equilibrium is stable?

If we look at the curve close to the equilibrium there are generically two forms it could take:

1. The curve starts above zero and crosses to below zero.
2. The curve starts below zero and crosses to above zero.

(In theory we could have a curve that just 'touches' zero and then goes back the same way, but this would be a very special type of point).

In case 1, for values of X less than the equilibrium we have $dX/dt > 0$, so the value of X will be increasing towards the equilibrium. Similarly, for values of X greater than the equilibrium we have $dX/dt < 0$, so the value of X will be decreasing towards the equilibrium. In both directions, we are pulled in towards the equilibrium so it is *stable*.

In case 2, for values of X less than the equilibrium we have $dX/dt < 0$, so the value of X will be decreasing away from the equilibrium. Similarly, for values of X greater than the equilibrium we have $dX/dt > 0$, so the value of X will be increasing away from the equilibrium. In both directions, we are pulled away from the equilibrium so it is *unstable*.

This hopefully makes sense in terms of sketching the curve, but we can calculate this stability directly from the ODE. If we think about those two cases, what we want to know is whether the curve defined by dX/dt is decreasing (case 1) or increasing (case 2) at the equilibrium. In other words, if we were to approximate the dynamics as a straight line, would its gradient be positive or negative? We can find this by *linearising* the system. For ease of writing, let us set $dX/dt = f(X)$. We can then take a *Taylor expansion* of our system near the equilibrium, which we'll call X^*, as,

$$f(X) = f(X^*) + \frac{df(X^*)}{dX}(X - X^*) + \frac{1}{2}\frac{d^2f(X^*)}{dX^2}(X - X^*)^2 + \dots$$

Now, since we are near the equilibrium we know that $f(X^*)$, that is the value of dX/dt at the equilibrium X^* is zero. Again, assuming we are near the equilibrium we can also say terms like $(X - X^*)^2$ are extremely small. We can therefore approximate the dynamics near the equilibrium by the linear system $dX/dt = (df(X^*)/dX)(X - X^*)$. Notice that this involves precisely the gradient we have talked about, $df(X^*)/dX$.

In short, then, to determine whether an equilibrium is stable in a one-dimensional system, we take the derivative of the ODE with respect to the variable, substitute in the value of the variable at the equilibrium, and check whether it is positive or negative.

TWO-DIMENSIONAL SYSTEMS

We made the case for stability depending on the linearised system in a one-dimensional system from a graphical argument about the gradient of the ODE itself near an equilibrium. For a two-dimensional system this is slightly harder to picture, but in fact the basic argument is the same. If we are near an equilibrium we can take a Taylor expansion of our system and approximate the dynamics by the linear terms. Now assume we have a generic system given by,

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

and that there is an equilibrium at (X^*, Y^*). The Taylor expansion of the first ODE near to the equilibrium gives,

$$f(X,Y) = f(X^*,Y^*) + \frac{df(X^*,Y^*)}{dX}(X - X^*) + \frac{df(X^*,Y^*)}{dY}(Y - Y^*) + \frac{1}{2}\frac{d^2f(X^*,Y^*)}{dX^2}(X - X^*)^2 + \frac{1}{2}\frac{d^2f(X^*,Y^*)}{dY^2}(Y - Y^*)^2 + \frac{d^2f(X^*,Y^*)}{dXdY}(X - X^*)(Y - Y^*) + \dots$$

and similarly for $g(X, Y)$. Making the same assumptions about being near the equilibrium we arrive at the linearised system,

$$\frac{dX}{dt} = \frac{df(X^*,Y^*)}{dX}(X - X^*) + \frac{df(X^*,Y^*)}{dY}(Y - Y^*)$$
$$\frac{dY}{dt} = \frac{dg(X^*,Y^*)}{dX}(X - X^*) + \frac{dg(X^*,Y^*)}{dY}(Y - Y^*).$$

This leaves us with a two-dimensional linear system. We won't go into deep detail here as to how to analyse such systems – if you need more background any textbook or lecture notes on *planar linear systems of ODEs* will give you the detail you need. We will jump ahead and say that we know that, if we set $x = X - X^*$ and $y = Y - Y^*$, then the solution to such a system can be written down as,

$$x(t) = x_1 e^{\lambda_1 t} + x_2 e^{\lambda_2 t}$$
$$y(t) = y_1 e^{\lambda_1 t} + y_2 e^{\lambda_2 t}.$$

In this case our equilibrium is at $(x, y) = (0, 0)$. We can therefore say that we would move towards the equilibrium – meaning it is stable – if $\lambda_1, /lambda_2 < 0$, otherwise we will move further away from it – meaning it is unstable. These two values are the *eigenvalues* of the system. These can be found by writing out the *Jacobian matrix* of the system, which is given by,

$$J = \begin{pmatrix} \frac{df(X^*,Y^*)}{dX} & \frac{df(X^*,Y^*)}{dY} \\ \frac{dg(X^*,Y^*)}{dX} & \frac{dg(X^*,Y^*)}{dY} \end{pmatrix}.$$

Formally we find the eigenvalues by writing out the *characteristic equation*, given by $\det(J - \lambda I) = 0$ where I is the identity matrix. This will give us a quadratic equation in λ that we can try and solve. Often however we will just look at the signs of *trace* and *determinant* of J, since these dictate the signs of the eigenvalues. In particular,

- if $\det(J) > 0$ and $tr(J) < 0$, the equilibrium is stable,
- if $\det(J) > 0$ and $tr(J), 0$, the equilibrium is unstable,
- if $\det(J) < 0$ the equilibrium is a saddle (one eigenvalue is positive, the other negative).

We can add further detail by noting that if the eigenvalues are complex we will see *spirals* (and this occurs when $tr^2 - 4\det < 0$), but if they are purely real we will see *nodes*.

HIGHER-DIMENSIONAL SYSTEMS

The approach outlined above for two-dimensional systems can be extended to three-dimensional or higher systems, but is generally much harder work. Ultimately we look to linearise the system and find the eigenvalues. However, since the characteristic equation will now be a cubic or even higher, there is no simple formula to solve it. Sometimes we will get lucky and it will nicely factorise – a couple of examples like this are seen in chapters 9 and 15 – but usually it is too complicated. We can also use an extension of the trace-determinant approach. In fact, this is the special case for two-dimensions of a method for testing stability called the *Routh-Hurwitz criteria*. This approach involves looking at the coefficients of the characteristic equation (the first and last of which are always the trace and determinant of your Jacobian) and checking their sign. Again, sometimes we can use this approach without too much trouble (like in chapter 6) but it is often challenging.

CHAPTER 23

Bifurcations

INTRODUCING BIFURCATIONS

In most of the models in this textbook we find that we can get a number of qualitatively different outcomes depending on the parameters. Very often we find that equilibria and their stability vary as we change different parameters, and specific values where equilibria collide, appear and change stability. These are important transitions known as *bifurcations* and are one of the most important aspects of a dynamical system. Bifurcations tell you when and how you can expect discrete shifts in behaviour as you change one or more parameters.

There are only a few different types of bifurcation that are possible, and we will cover the fundamental ones here. In all cases we will see how changing a particular parameter leads to a change in the stability, or existence of, the equilibria in the model. We will largely do this by looking at *bifurcation diagrams*. These plot the location and stability (solid lines denote stable equilibria, dashed lines unstable) of equilibria as parameters vary.

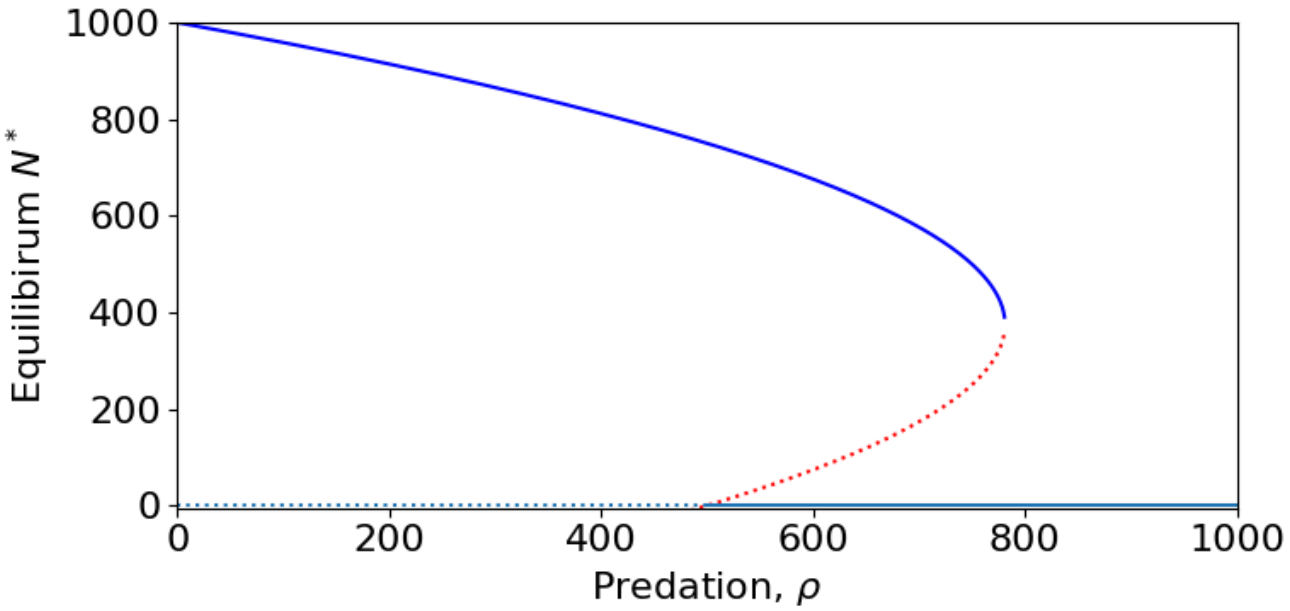
The first three examples can all be seen in one-dimensional (i.e. one variable) models as well as in higher dimensions. (There is some formal mathematical work and terminology behind bifurcations that we will not concern ourselves with here. If you want to know more, the textbook *Nonlinear Dynamics and Chaos*, by Strogatz is a very good place to start.)

STANDARD BIFURCATIONS

TRANSCRITICAL BIFRUCATION

The transcritical bifurcation is perhaps the most common form of bifurcation in a mathematical biology model. It occurs when two equilibria intersect and swap stability. It is so common because we will usually have a trivial equilibrium where nothing exists, which will lose stability when it intersects with an equilibrium where *something* exists.

An example is shown in the diagram below (which is actually from the spruce budworm model in chapter 1). For $\rho < 100$ the $N = 0$ equilibria is unstable. Technically a second equilbiria exists at $N < 0$, which would be locally stable, but is not biologically feasible. For $\rho > 100$ we see that now the $N = 0$ equilibrium is stable, with the second equilibrium now present at $N > 0$ being unstable. The point where these two equilibria meet and swap stability, $\rho = 100$ is the transcriticial bifurcation.



Example of a bifurcation diagram showing a transcritical and saddle-node bifurcation (taken from the spruce budworm model in chapter 1). Solid lines denote stable equilibria and sashed lines unstable equilibria.

SADDLE-NODE BIFURCATION

The transcritical bifurcation occurs when (at least) one equilibria exists over a large range of the parameter (i.e. the trivial equilibrium). However, there are also cases where equilibria can be created or destroyed as a parameter is varied. These are called saddle-node or blue-sky bifurcations (the former because what gets created at the bifurcation is a stable-unstable pair of equilibria; the latter because the equilibria appear out of the clear blue sky)

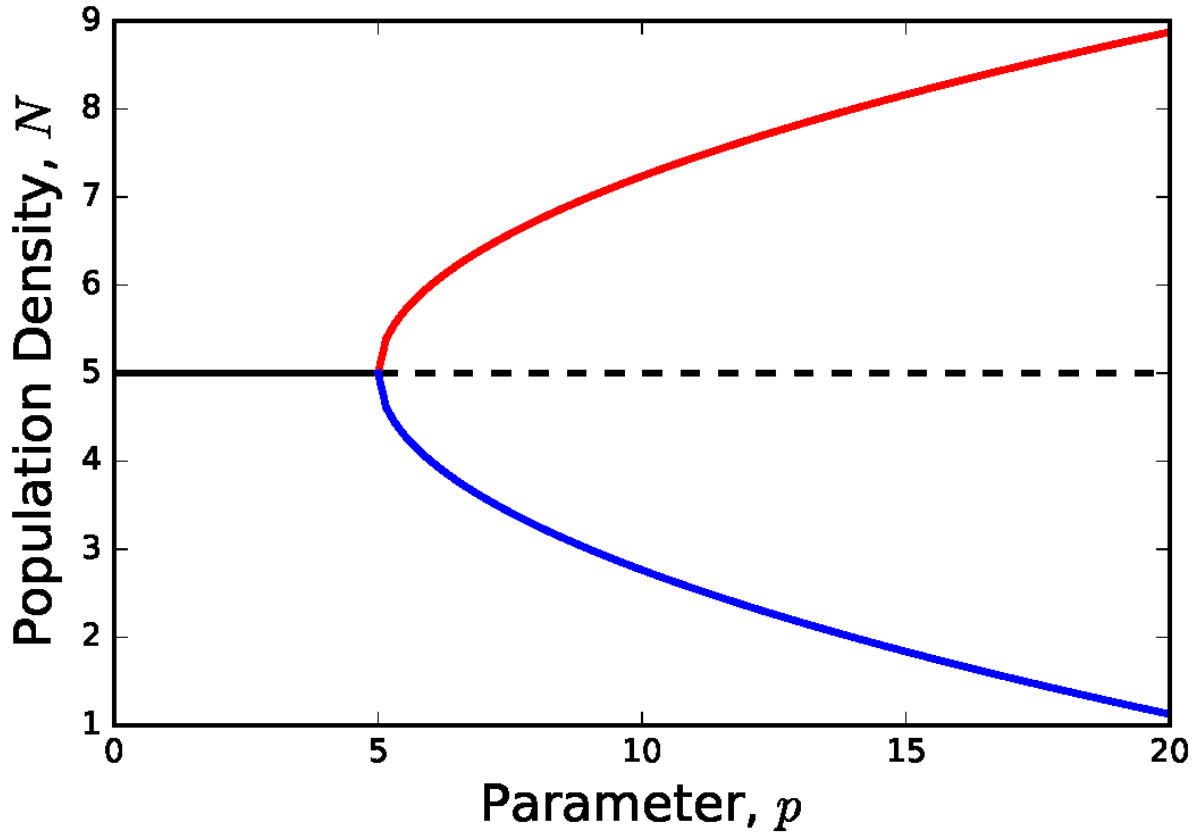
An example of a saddle-node bifurcation can again be seen from the spruce budworm model. For $latex\ rho\gt300$ we see that only the $N = 0$ equilibrium exists. As ρ is reduced the two non-zero equilibria appear, one stable and the other a saddle, and diverge.

Saddle-node bifurcations thus have very important consequences for any biological system. It means that there may not be a gradual and steady decrease down to extinction (as in the transcritical case), but a very sudden crash. This is often termed 'a catastrophe' and is a focus of much research to understand the risk of sudden extinctions in real populations. For example, models of fisheries management show that if the rate of fishing becomes too high we may see sudden declines in fish stocks. Moreover, the decline is very hard to reverse as a small decrease in the harvesting rate does not easily return the system to the stable equilibrium as the saddle is still pushing the population densities lower. This is an example of *hysteresis*.

PITCHFORK BIFURCATION

The third type of standard bifurcation – the pitchfork – does not occur in any of the models covered in this textbook (it is usually a result of strong symmetry in a system). However, for completeness we shall briefly mention it here.

The reason for the naming of this bifurcation is obvious by looking at the bifurcation diagram below. At low value of the bifurcation parameter there is a single stable equilibria. At the bifurcation point this existing point remains but loses stability, and two new stable equilibria emerge.



A made up example of a pitchfork bifurcation.

HOPF BIFURCATION

The final type of bifurcation we will cover here is only seen in systems of two or more dimensions. A Hopf bifurcation occurs when varying a parameter alters an equilibrium that was a stable spiral into an unstable spiral (or vice versa). This is a particularly important behaviour, because as we initially move into the unstable spiral a unique, stable closed orbit emerges from the equilibrium, resulting in the population cycling/oscillating. It is not very easy to draw a bifurcation diagram in this case, but we can usually see the process by looking at the phase portraits as we vary a parameter and seeing how the limit cycle emerges. Examples of this are seen in chapters 3 and 15.

Final thoughts and acknowledgements

SUMMARY

You have reached the end of this open education resource on introducing mathematical biology – well done! In every chapter we have given 3 key takeaways to remember. What are 3 key takeaways from this whole book?

Key Takeaways

- We can build mathematical models for biology or medicine by considering the mechanisms that will increase or decrease the whole or parts of the population.
- We can apply a range of tools from dynamical systems – from explicit solutions, to qualitative analysis, to computational simulations – to infer how populations will change over time and under different parameter regimes.
- Even simple models can provide us with very useful insight into biological and medical systems.

FURTHER STUDY

I hope that this open education resource has proved useful to someone out there. If anyone ever finds themselves having read to this point and thinking they have found it useful, please do feel free to let me know using the anonymous feedback form, as this helps me understand how the textbook is being used.

I also hope this textbook might inspire you to further study, whether within a specific area of mathematical biology or medicine, or more generally in mathematical modelling. This has been a bit of a whistle-stop tour of investigating questions from across biology and medicine using ordinary differential equations. As stated at the very start of the book, this is just one method for approach modelling questions, and the more mathematically or computationally adventurous amongst you will find many different technical approaches out there. Similarly, the biological questions are by no means limited to the topics covered here. As a first step to further study, you might find it useful to browse some of the references for some more in-depth study of particular subjects. You could also search for mathematical biology research groups at a nearby university or research body (a quick plug: click here to go to the webpages for our own research group) – on the whole we are a friendly bunch and will generally welcome enquiries from those interested in pursuing further study.

ACKNOWLEDGEMENTS

It is a cliche, but this work would truly not have been possible without the support from a number of wonderful people. Firstly a huge thanks to **Oscar Devine-Wright**, who worked as an assistant in preparing the book and undertook many helpful activities, including producing the wonderful hand-drawn figures.

I also picked the brains of my colleagues in the mathematical biology research group at the University of

Sheffield many times during the process. An extra special mention goes to **Nick Monk**, with whom I taught the *Mathematical Biology* module at Sheffield for many years, the content of which formed the backbone of this book. Further thanks go to **Alex Fletcher** and **Jonathan Potts** for their helpful thoughts and feedback. I'd also like to thank my erstwhile PhD students who read parts of the content and gave useful feedback, **Prerna Singh**, **Megan Oliver**, **Charly Pain** and **Ryan Grossman**.

I must thank the **Open Education Resources** team at the Unviersity of Sheffield Library. Importantly, the team provided me with funding that allowed me to focus on developing these materials in early 2023. I'd particularly like to thank **Helen Moore** and **Maria Mawson** for their support in every stage of the development. I must also thank the **UK-US Fulbright Commission** for awarding me a visiting scholarship in 2021, during which the idea behind this textbook started to form.

I'd like to thank members of the **Mathematics and Statistics Digital Accessibility Working Group**, especially **Jenny Hughes**, as it was conversations as part of this group which first led me to exploring the world of Open Education Resources and online materials. I'd also like to mention the team behind the **Jupyter Book** initiative, as this was the first platform I used to create free, online content for my lectures and they unwittingly built my knowledge and confidence of how to develop online materials.

ABOUT THE AUTHOR

I am a lecturer in the School of Mathematics and Statistics at the University of Sheffield, specialising in mathematical biology. I obtained a BSc Mathematics and Philosophy from the University of Durham in 2005 and an MRes Mathematics in the Living Environment from the University of York in 2006. My PhD was titled *Modelling the evolution and coevolution of host defence* under the supervision of Prof Mike Boots in the Animal and Plant Sciences department also at the University of Sheffield. I have been teaching undergraduate and postgraduate maths courses since 2013 and I was a Fulbright Scholar in 2021.

My teaching focusses on mathematical modelling, guiding students through how to build and analyse models for real-world systems. I place a strong emphasis on embedding equity, diversity and inclusion into all I do and am passionate about encouraging and supporting students from minoritised groups to succeed in mathematics.

I live on the outskirts of Rotherham with my wife and children.

References

The material that formed this book has been developed over a number of years as part of my teaching at the University of Sheffield, particularly in modules of *Mathematical Biology*, *Mathematical Modelling of Natural Systems* and *Mathematics and Statistics in Action*. This is very much a 'standing on the shoulders of giants' production, and I have tried my best to acknowledge the various sources that have fed into this material, either directly or through some general osmosis. For each chapter I have tried to list the works that were directly used in developing the content, and their detailed citation information is listed below. I particularly want to acknowledge the following resources that have played an indirect but important part in choosing and explaining the content:

- **Mathematical Biology** by Murray – most people in the field will know this as the definitive textbook for mathematical biology teaching, and I suspect more ideas than I realise have come from this classic.
  Murray, J. (2002). *Mathematical Biology I: An Introduction*. New York: Springer.
- **Modelling Infectious Diseases in Humans and Animals** by Keeling & Rohani – again, a bit of a classic in the more specific field of epidemiological modelling.
  Keeling, M. and Rohani, P. (2007). *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press.
- The **Basic Pharmacokinetics** e-book and webpages by Bourne – a fantastic resource for learning about pharmacokinetic models.
  Bourne, D. (2022). *Basic Pharmacokinetics v1.5.5.* Apple iTunes bookstore, https://itunes.apple.com/us/book/basic-pharmacokinetics/id505553540?mt=11
- **Nonlinear Dynamics and Chaos** by Strogatz – there are a great many books devoted to the general field of dynamical systems, but I would argue this is by far the most accessible for an undergraduate mathematician.
  Strogatz, S. (2000). *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry and Engineering*. Westview Press.

Further reference list

- Anderson, R. and May, R. (1981). The population dynamics of microparasites and their invertebrate hosts. *Phil. Trans. Roy. Soc,. B*, 291:451–524.
- Anderson R. and May, R. (1982). Coevolution of hosts and parasites. *Parasitology,* 85:411-26
- Best, A. and Ashby, B. (2022). Herd immunity. *Curr. Biol.,* 31:R174-R177.
- Geritz, S. Kisdi, E., Meszena, G. and Metz, J. (1998). Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. *Evol. Ecol.* 12:35-57.
- Gerlee, P. (2013). The model muddle: in search of tumor growth laws. *Cancer Research*, 73:2407-2411.
- Goodwin, B. (1965). Oscillatory behavior in enzymatic control processes. *Adv. Enzyme Regul.,* 3:425–428.

- Hahnfeldt, P., Panigrahy, D., Folkman, J., Hlatky, L. (1999). Tumor development under angiogenic signaling: a dynamical theory of tumor growth, treatment response, and postvascular dormancy. *Cancer Research*, 59:4770-4775.
- Hernandez-Vargas, E. and Velasco-Hernandez, J. (2020). In-host mathematical modelling of COVID-19 in humans. *Annual Reviews in Control*, 50:448-456.
- Perelson, A. and Nelson, P. (1999). Mathematical analysis of HIV-1 dynamics in vivo. *SIAM Review*, 41:3-44.
- Yin, A., Moes, D., van Hasselt, J., Swen, J. and Guchelaar H-J. (2019). A review of mathematical models for tumor dynamics and treatment resistance evolution of solid tumors. *CPT Pharmacometrics Syst Pharmacol.*, 8:720–737.