# Mathematical approach to Biology

**A very short introduction**

Oreste Affatato*

March 4, 2024

An Awesome Publisher

* A not so awesome author

**Disclaimer**
This book contains the material for the lab of mathematical tools applied to Biology, for master students in Biomedicine at Uppsala University. The text is purely intended as an introduction to more sophisticated and rigorous methods and should not be considered a specialized textbook on the subject.



**Colophon**
This document was typeset with the help of KOMA-Script and LaTeX using the kaobook class.
The source code of this book is available at:
https://github.com/fmarotta/kaobook
(You are welcome to contribute!)

**Publisher**
First printed some time ago by An Awesome Publisher

My methods are really methods of working and thinking; this is why they have crept in everywhere anonymously.

– Amalie Emmy Noether

# Preface

This is the handout for the computer lab on Mathematics applied to Biology. It is a very simple introduction to the field of Biomathematics with some R programming. The overall aim is to offer you valuable tools that for sure might become useful in your future career. In doing that, I also hope to convey the beauty of these methods.
I hope that there are not so many mistakes. I am only a researcher at beginner level and a constant amateur in life. My efforts here are to be consider a gentle push for you to read better texts. Anyway, I hope to convey the beauty of the scientific method, as one of the most valuable tools humankind has to uncover what we call truth.
P.S. I will be grateful to our students that will indicate to me mistakes in the text. I will also greatly appreciate any suggestion to improve the lab.

*Oreste*

# Contents

# 1 Introduction



In the next sections you can find some material on very basic concepts that will be very important for the lab. Many of you are surely very familiar with this ideas, but for someone it may be useful to read again a bit.
In this introduction we are also going to start drawing the main distinction that is going to guide our short journey into this kind of methods: the difference between the continuum and the discrete.

## 1.1 Continuous Mathematics

Let's start recalling the concept of *set*. Actually, this is a very problematic concept in mathematics[1], but we'll just use the naive definition.

1: I suggest to the curious reader to check the Russell's paradox.

**Definition 1.1.1** *A set is a collection of objects, called elements of the set. A set is called finite if it has a finite number of objects, while it is infinite if it has infinite objects.*

In the case of a finite set you can just write down its elements, while in the case of a infinite set you should write down the common characteristic of the elements.

$$A = \{a, b, c\}$$

A is a finite set, with three elements. B is the set of all the positive natural numbers bigger than 2

$$B = \{x \in \mathbb{N} \mid x > 2\}$$

Where $\mathbb{N}$ is the set of all the natural numbers. When we want to say that a a certain element belongs to a set, for example that the element $a$ belongs to the set $A$, we write $a \in A$.
Now we will introduce a slightly more complex idea. In biology, we are interested in quantities like the concentration of a reagent, the number of individuals in a certain population, the signal sent by a firing neuron... All of these quantities are related to others. The size of a population may change over time, concentrations of chemical reagents may vary due to other chemical element, the prevalence of a disease may vary in space in a given city or region. We describe these relations through the concept of *function*. Very basically speaking, a function is a rule to associate different elements of a set to element of another set. For example, in a case in epidemiology we may connect via a function the number of individual with a certain disorder with the region in which they live. So in one set we have all the individuals with that disorder and in the other the various places of a region. A function is an association of people in the

first set with places in the other set.
Let's now give a more rigorous definition.

**Definition 1.1.2** *A function is a relation that associates **each** element of a given set X to **one** element of another one Y. X is called domain of the function, while Y is the codomain.*

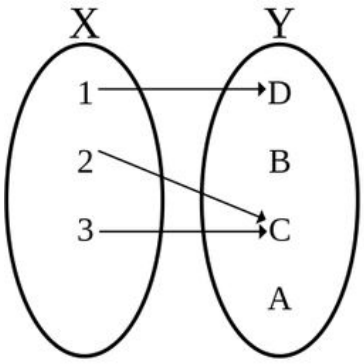



**Figure 1.1:** Example of a function. Source: Wikipedia.

In this case we will say that this function is defined in $X$ and that has values in $Y$. What the function $f$ does is simply to transform each element in $X$ to an element in $Y$. So in the case of the example of Figure Figure 1.1 the value 1 is transformed into D by the function $f$. We may write that $f(1) = D$, so $f$ applied to 1 gives D as an output.
Speaking in general terms, the fact that a function $f$ has $X$ as domain and $Y$ as codomain, may be written as $f : X \to Y$. To highlights better the fact that $f$ takes from $X$ the inputs and that transforms those inputs into elements contained in $Y$. Speaking in general case, given $x \in X$ as an input and $y \in Y$ the corresponding output we may write more synthetically that $y = f(x)$, which means that $f$ transforms a given $x$ into $y$*.
In our lab we are going to focus our attention to a specific class of mathematical function. Our main variables of interest will depend on time, concentration of other variables etc. Such variables can get any possible value in the real line, at least in principle. They can assume continuous values. Since the core variables are continuous as well as their outcomes, we will focus our attention of functions defined in $\mathbb{R}$ with values in $\mathbb{R}$. Thus the functions we are interested in are the functions $f : \mathbb{R} \to \mathbb{R}$.

### Some important functions

We will now briefly recall some functions that we'll be using in our lab. All of the functions below will be $f : \mathbb{R} \to \mathbb{R}$.

- the linear function

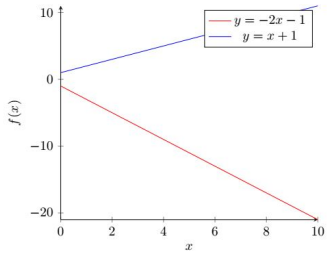



**Figure 1.2:** Graph of two linear functions.

The linear function is a typical real function.
The general form of the linear function is the following

$$f(x) = mx + q$$

and in the usual real plane it has of course the shape of a straight line (see Figure 1.2). The factor $m$ is the slope of the line while the $q$ is the intercept to the y-axis.

- the quadratic function

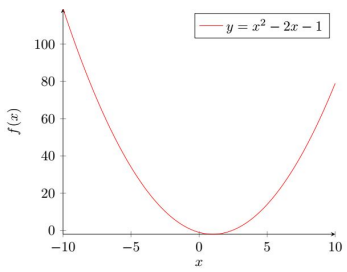



**Figure 1.3:** Graph of a parabola.

The general form of the quadratic function is the following

$$f(x) = ax^2 + bx + c$$

and in the real plane it has the shape of a parabola (see Figure 1.3).

* Remember that $x$ is an element of $X$ while $y$ is an element of the set $Y$. In mathematical terms, $x \in X$ and $y \in Y$.

- the exponential function

The general form of the exponential function is the following

$$f(x) = a^x$$

Where $a$ is any positive number. Remember that $a$ is fixed, while $x$ is the variable. The shape of this function is in Figure 1.4

Notice from the graph that if $a > 1$, than the curve grows for increasing values of $x$. For values in which $0 < a < 1$ the situation is reversed. For increasing values of $x$ the exponential tends to zero. Notice also of course that if $a = 1$ or $a = 0$ we don't have any exponential at all.
From the graph you may also notice that $f(0) = a^0 = 1$.
$a$ is called the *basis* of the exponential and we'll mostly used $e$, the *Napier's number* as basis for our exponentials. $e$ is an irrational number, $e = 2.718....$

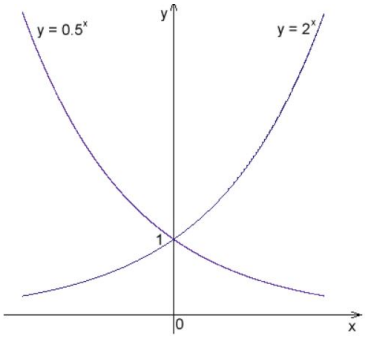



**Figure 1.4:** Graph of an exponential. Source: Wikipedia.

- the logarithmic function

The general form of the exponential function is the following

$$f(x) = \log_a x$$

Where $a$ is the *basis* of the logarithm, while $x$, the variable, is the *argument* of the logarithm. While the exponential has a very straightforward interpretation, as it is very similar to the power of a number, the logarithm is a bit tricky.
The idea is the following, the logarithm is the inverse of the exponential. So the result of a logarithm, its output, is which exponent you should put to the basis in order to get the argument.
For example, what is the logarithm in basis 3 of 9, i.e. $\log_3 9$? Which exponent should I put on 3 to get 9? 2. Thus, $\log_3 9 = 2$, because $3^2 = 9$. This was a simple case, $\log_3 5$ is more complex, for example. But it is ok, the computer is going to do the calculation, we just need to understand the concepts.
The graph of the logarithm is in Figure 1.5

A very important logarithm is the one in basis $e$. This is called *natural logarithm*, $\log_e x = \ln x$.

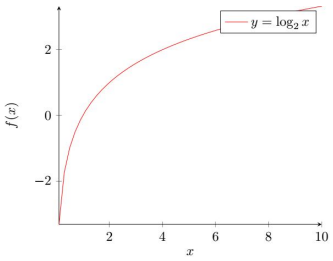



**Figure 1.5:** Graph of a logarithm.

### Derivation

Now we are going to recall briefly the concept of derivative. This idea is strictly related to the problem to find the rate of variation over time of a certain quantity[†].
Imagine you have a function that describes the trajectory of a material body. So this functions relates the positions of this body to the various instants of time. Once we know this function, we indeed have a good amount of information and knowledge about its movements, but this is not the end of the story. For instance, we don't know a key feature of interest, such as its velocity. So now we want to know the rate of

[†] This and related concepts were introduced by Isaac Newton and Gottfried Leibniz independently in the 17th century.

variation of this function, how fast this particle changes position over time. Basically we want to calculate this velocity, knowing its trajectory. Such a problem led Newton to set the basics of *mathematical analysis*, or *calculus*, a knowledge that at the beginning was applied to the movements of stars and planets, but later it was applied to virtually any scientific field.
What he noticed is that the straight line, the linear function we already briefly recalled, has a very clear slope, the coefficient $m$. The bigger the $m$, the higher the slope, i.e. faster the function changes over $x$ (imagine over time). A general function doesn't have this feature, the slope may change continuously. He noticed that he could apply this property of the lines to general curves. The idea is that a general function, a general curve, is going to vary, but *locally*, if we zoom in a lot in the graph it seems like a straight line. Think about the Earth. Of course is not flat, but we in general don't notice the curvature, we perceive that the Earth is flat. Because locally it is flat, but globally is a spherical solid.
The same applies to our function. The idea is approximate our general function locally with a straight line. With this process we have a line with the same slope with our function (locally). Repeating this process for all the points in the graph of our function we obtain global information about the rate of variation of our function (see Figure 1.6).

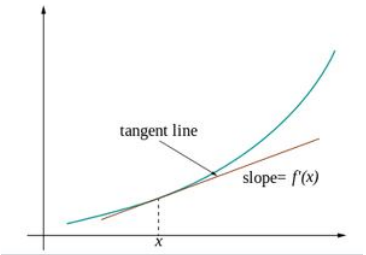



**Figure 1.6:** Tangent line approximating locally a function. Source: Wikipedia.

The slope we obtain with this calculation is called the *derivative* of the function $f$ at a point $x$. Given the function $f$, its derivative in $x$ is written as $f'(x)$.
Now we are going to give to this process a more sound mathematical basis.

We want to calculate the slope of the line tangent to the function in Figure 1.7 at a fixed point $x$. We start drawing a secant, passing through $(x, f(x))$ and another point $(x + h, f(x + h))$, where of course $h$ is the distance between $x$ and $x + h$. We notice that when $h$ tends to zero, i.e. the distance between these two points tends to zero, of course $x + h$ tends to $x$, but also the secant line tends to the tangent line of the curve in $x$ (see Figure 1.8).

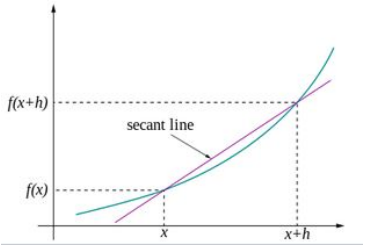



**Figure 1.7:** Approximating the tangent with secants. Source: Wikipedia.

We are going to use this idea, and the definition of slope of a straight line, in order to get the definition of derivative.
From Figure 1.7 we calculate the slope of the secant line. By definition

$$m = \frac{f(x+h) - f(x)}{x + h - x} = \frac{f(x+h) - f(x)}{h}$$

As we said, when $h$ tends to zero, this $m$ tends to be the slope of the tangent of the curve in $x$. If this limit exists and it is finite it is the *derivative* of the function in $x$.

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

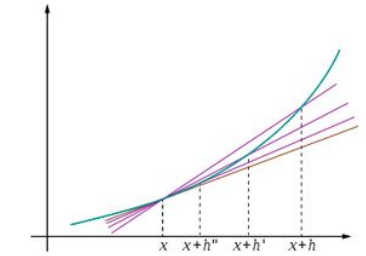



**Figure 1.8:** Secants converging into tangent. Source: Wikipedia.

Actually, the process of calculating derivatives is algorithmical and we may use softwares to calculate them. We'll just recall here some important derivatives and some properties.

- the derivative of a sum of functions is equal to the sum of the derivatives

$$[f + g]'(x) = f'(x) + g'(x)$$

- derivatives do not affect multiplication by a constant. Let's consider a constant $k$

$$[kf]'(x) = kf'(x)$$

- rule of derivation of multiplication of functions

$$[fg]'(x) = f'(x)g(x) + f(x)g'(x)$$

- rule of derivation of division of functions

$$\left[\frac{f}{g}\right]'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{g^2(x)}$$

Let's see now some common derivatives.
The constant function never varies, so its derivative is always zero. $f(x) = k$ implies $f'(x) = 0$.

- given a power $f(x) = x^n$, the derivative is $f'(x) = nx^{n-1}$;
- given an exponential $f(x) = e^x$, the derivative is $f'(x) = e^x$;
- given a logarithm $f(x) = \ln(x)$, the derivative is $f'(x) = \frac{1}{x}$;

**Differential equations**

Most of you may remember the usual algebraic equations. Here's an example.

$$x + 1 = 6$$

The solution of such an equation is a number that, once put into the equation, will produce and identity (i.e. something always true). In this case is 5.

$$5 + 1 = 6$$

$$6 = 6$$

In this case we got just one solution, but there can be more than one. Consider the following equation

$$x = x$$

This is true for every number, since every number is equal to itself. In this case we have infinite solutions. Or consider this other case

$$x = x + 1$$

Here we don't have any solution, as no number is equal to itself plus one. This brief review of such elementary concepts is essential to extend the basic idea of algebraic equation to differential equation. In its core, the idea is the same, we are looking for a solution, but in the case of differential equations, the solution is not a single number, but an entire function. As in the previous cases, we may have one solution, more than one, or no solutions at all. Differential equations are fundamental because in many cases we know the flow, the evolution in time of a certain quantity, but we don't know its explicit form. Solving a differential equation we go from what we know about the rate of change of this phenomenon to its actual explicit form.
Let's consider functions depending on just one variable (e.g. time) $f : \mathbb{R} \to \mathbb{R}$. A *differential equation* is a relation between an unknown function, its derivatives, and also the independent variable (time in our example). The following is a general way to state this

$$F(t, f, f') = 0$$

This is a first order differential equation since it appears just the first derivative. In general, by *order* of a differential equation we mean the maximum order of derivative that appears in the equation. Let's consider an easy example.

$$\frac{df}{dt} = 0$$

This is a first order differential equation. The solution is a function whose derivative is always zero. We know that each constant has this property. So for example $f(t) = 6$ a function that is always equal to 6 (for any time) should indeed solve the equation. Let's see

$$\frac{df}{dt} = \frac{d}{dt}(6) = 0$$

So this is a solution. Are there others? Yes, every constant is a solution. Every function of the form $f(t) = k$, with $k$ constant value, is of course a solution. So we have infinite solutions. This may be a problem since in many application we want a single solution for our models, not infinite. In order to avoid that problem we add the so called *initial condition*. Basically, we want that the solution satisfies the differential equation, but also that it respects a prescribed initial conditions, i.e. at a certain time it has a specific value. This will fix one solution, among the infinite possible. A differential equation plus an initial condition is called *Cauchy problem*. Let's consider the following

$$\begin{cases} \frac{df}{dt} = 0 \\ f(0) = 4 \end{cases}$$

In this case we are looking for a function whose derivative is always zero and that at time $t = 0$ the solution is equal to 4. The previous solution that we proposed, $f(t) = 6$, is not a good one anymore, since in zero is still equal to 6, not to 4. The only possible solution is $f(t) = 4$. As you see, there is only one possible solution for this Cauchy problem.

The Cauchy problem, in the end, is our model. We gather our knowledge about a phenomenon and we produce a differential equation. From that we can compute the solution. Using experimental data we may use the initial condition and we may analyse how the theoretical curve is fitting the experimental curve. Our hope is that our theoretical explanation may produce a model that has a good fit with the data. This can get us closer to the real nature of the phenomenon we are investigating.

## 1.2 Discrete Mathematics

Here we recall very briefly some ideas in discrete mathematics. Everything will be developed better in due course.
Why do we need discrete mathematics? Why can't we just use the mathematics of the continuum for our models? The answer is simple: because everything in nature is discrete.
The flow of water, as you know, is made by many tiny molecules moving, even if it may appear as a continuum. We use the continuum approach because it is a good approximation of reality, since at many levels of experiments and observations we don't really deal with the true nature of water. Thus, treating it as a continuum is a good basis for any model that does not dwell too deep into this phenomenon. The positive aspect of this approximation is that it holds true in many applications.
In some contexts, by the way, we are investigating the phenomenon at a scale in which we just cannot ignore the discrete structure of reality. If you are doing, for example, experiments on water molecules at atomic level, at this stage you cannot treat the system as a continuum.
The same holds true in our application to biology. Imagine we are studying the dynamic of a population of a certain species of animals in a certain environment. We are interested in the rate of growth as well as in the rates of deaths. What is most likely to happen is that they will have a certain period during the year in which they mate and then reproduce. Then, the cubs will be born after a defined period of time. So, the total population will increase on average of a certain amount, but just after that specific time. The new generation doesn't follow in a continuous flow in many contexts, and the discrete nature of the phenomenon must be taken into account. In this brief introduction (and better later on in these notes) we will develop the mathematical tools to create models that can take into account the discrete nature of such phenomena.
Recall that $\mathbb{N}$ is the set of all the natural numbers, so $\mathbb{N} = \{0, 1, 2, 3, ...\}$. In the previous section, the functions were depending on time (or in general a continuous variable), but in the discrete case they will depend on separated time intervals, each labelled via a natural number. So there will be the zeroth interval (or instant), the first, the second, and so on. We will allow our functions to have real values, just the time will be discrete[‡]. So we are interested in functions defined in the set of natural numbers, but with values in the set of the reals.

**Definition 1.2.1** *A sequence is a function defined in $\mathbb{N}$ with values in $\mathbb{R}$,*

[‡] As in the case of the population dynamic, the period of pregnancy after the mating is fixed, while the average number of individual in the population after that period may be real, as it is an average.

*$f : \mathbb{N} \to \mathbb{R}$. Generally, a sequence is written as $\{x_n\}$, where $x_n$ is the general term of the sequence.*

Let's consider an example: $x_n = 2n$. I defined the sequence by defining the general term. This is the sequence of the even numbers, i.e. $2, 4, 6, \ldots$. As you see, you just need to define the general term to know the behaviour of the sequence.
Very much like the case of the differential equations, also here we have functional equations in which the solution is a sequence. These are called *difference equations*. We'll be dealing with a special case of these difference equations, the so called *recurrence relations*, in which basically the next step in the sequence is defined by the previous one[§]. Consider the following example

$$x_{n+1} = 3x_n$$

This is a very simple recurrence equation. The solution is a sequence that, once substituted in the recurrence equation, will always give an identity. In order for a sequence to be a solution you of course have to explicitly identify it. In this case, a solution could be $3^n$. Let's see if it works

$$x_{n+1} = 3x_n$$

$$3^{n+1} = 33^n = 3^{n+1}$$

$$1 = 1$$

Summing up, a general recurrence equation would have the following aspect

$$x_{n+1} = f(n, x_n)$$

A solution of such an equation is a sequence that inserted into this equation will always give an identity.

[§] Of course it could be more general than that.

# Why Mathematics? 2



Why to learn mathematics in a biomedical master program? To be more precise, we are going to learn mathematical modelling, but still why bother with that? Life sciences are becoming more and more quantitative and therefore there is an increasing need for learning quantitative and analytical thinking.
There are also many successful cases of applications of mathematical tools in biological research. A good example is the work from Hodgkin and Huxley on the action potential. Their mathematical model created the basis of the modern neuroscience and for this they received the Nobel Prize in physiology. There are many more important cases, but we decided to justify our approach using a concrete example, to see the power of mathematics in action.
Much of this notes are inspired by a very good introductory book on mathematical biology [1] (unluckily not available in English). More can be found in a fundamental, but more advanced text [2].

## 2.1 Ho's model on HIV

The *Human Immunodeficieny Viruses* (HIV) over time cause the acquired immunodeficiency syndrome (AIDS). The rate of reproduction of this kind of virus is quite high *in vivo*, but the immune system may slow down the progression of the HIV for a certain period of time.
HIV attacks mainly CD4 T cells, a class of helper T cells. When the concentration of CD4 T cells, which is normally around 1000 cells/$\mu L$, decreases to 200 cells/$\mu L$ the patient is considered to have AIDS[1].
At the very beginning, researchers thought that the very long period between the HIV infection and the manifestation of AIDS symptoms it was a period of latency of the virus. The main idea was that the mechanisms involved was in general quite slow. Now, we know that many complex mechanisms are involved and at very different time scales (see Figure 2.1). To understand these mechanisms is essential to improve the treatment available to cure this disease.

1: Of course this is not the only criterium used, we are simplifying the discussion for pedagogical reasons.

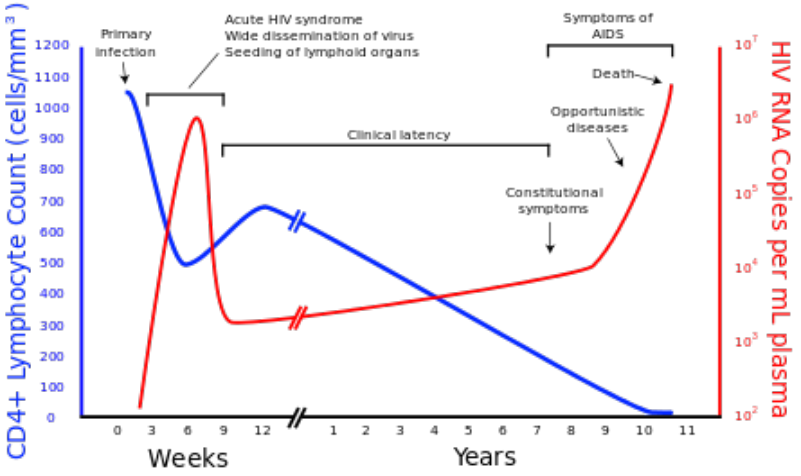



**Figure 2.1:** Progression of HIV infection. Source: Wikipedia.

An essential step forward in understanding the dynamic of this infection was made by David Ho, in the mid nineties. We are now going to have a look to Ho's paper [3].
As we already said, people thought that the HIV was in latency period, at the beginning of the infection. The concentration of HIV and antibodies was almost constant over time, while CD4 T cells were slowly diminishing. To understand weather HIV was active or not, Ho's group perturbed its activity via administration of inhibitor of HIV-1 protease (ABT-538) on 20 patients. However, to understand the results we need a model first.
Let's call $V(t)$ the concentration of the virus at a certain time $t$. This function describes how changes the virus population over time. This is the quantity we want to know in order to see how fast the virus is proliferating. To get this knowledge we need some other information and to translate this information into a quantitative language. Since this $V(t)$ is the concentration of the virus at a certain time, what can change it? This concentration may grow because of the reproduction process of the HIV; but this concentration may also decrease because these cells may die or because of the action of the immune system.
Taking all of this into account let's call $P$ the concentration of viral cells produced in the unit of time and $E$ the rate of elimination of viral cells[2] (due to the mechanisms described). We are here assuming that in this period of the infection, those values ($P$ and $E$) are basically constant over time [3].
Now that we have all these ingredients, we may start building up our model. As we already said, the variation of the concentration of HIV depends on two mechanisms: creation of new cells, death of other cells.

2: By this we mean the percentage of viral cells eliminated in the unit of time.

3: Remember our hypothesis? The concentration of the viral cells is overall constant during the latency period.

variation of V(t) = creation of new cells - death of some cells in the population

In mathematical terms, the variation of a function is defined by its derivative. So

$$\frac{dV}{dt} = P - EV$$

As people thought at that time, in this latency phase the concentration of the virus should be constant over time. This means that the variation over time must be equal to zero.

$$\frac{dV}{dt} = 0$$

And this means

$$P - EV = 0$$

Solving for the concentration we finally get that $V = P/E$. Let's call this specific concentration the concentration at equilibrium $V_e$. So $V_e = P/E$. Now, in Ho's experiment the protease is inhibited. Let's assumed that it is perfectly inhibited, to simplify a bit our calculations. So, if we administer this inhibitor at a certain time, let's say, $t = 0$, for all the future instants, $t > 0$, we have that there will be no more viral cells, so $P = 0$[4]. Thus, our equation becomes

$$\frac{dV}{dt} = -EV$$

This is a quite easy equation to solve. How to get the solution is not important at the moment, let's keep focusing on the meaning of all this process. The solution is an exponential function. Since no more cells are produced, the only mechanisms left are the ones that kills viral cells. So the concentration of viral cells is must decrease to zero over time (and our model tells us that this is going to happen in an exponential way, due to our assumptions)[5]. The solution is

$$V(t) = V_0 e^{-Et}$$

So this is the *explicit* form of your function, how the concentration of viral cells $V$ behaves over time. $V_0$ is your starting point, the population you had at the first instant, what in mathematics is called the *initial condition*. So it is basically the value that your function has at the instant $t = 0$. So $V(0) = V_0$. Let's have a quick check on our solution[6].

$$V(0) = V_0 e^{-E0} = V_0 e^0 = V_0$$

But let's remember that at the beginning of the treatment we were in the latency period of the virus and in this regime we know not only that the population is constant over time, but also that in our model has a very specific value, which is $V_e = P/E$. Thus our initial condition, our initial population, is the concentration at equilibrium. Thus, in the end the solution is

$$V(t) = V_e e^{-Et} = \frac{P}{E} e^{-Et}$$

See Figure 2.2 for a simple plot of this function.

4: The authors are assuming the the action of the inhibitor is perfect.

5: This may be obvious for you, biologically speaking you already know that. But the important thing is that in this way you can make actual predictions. You may know *exactly* how this process is going to behave. Thus you can have stronger basis to check your hypotheses.

6: We are just evaluating the function in the value zero, $t = 0$. So wherever you find a $t$ in your equation, just replace it with zero and then do the calculations.

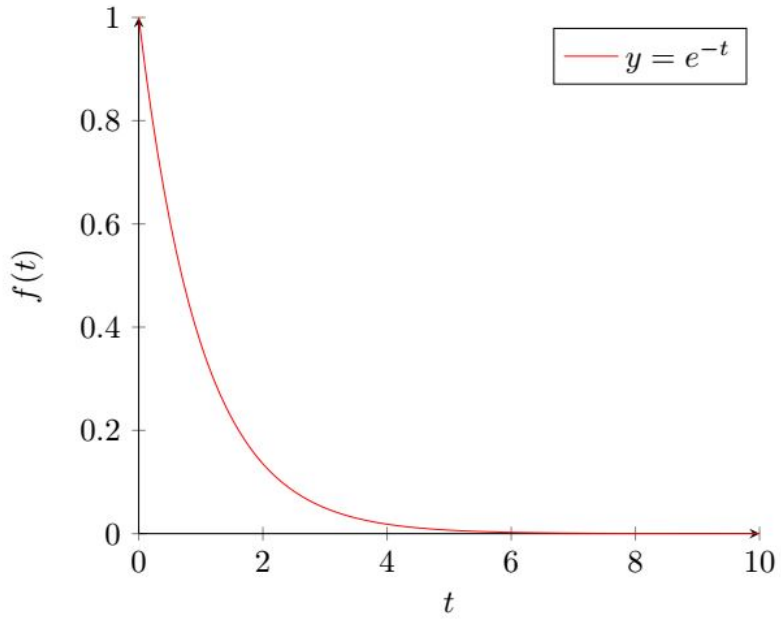



**Figure 2.2:** Plot of exponential function. In this case, $E = 1$ and $P/E = 1$.

Just bear in mind that $P/E$ is our estimation for the initial concentration and that $E$ is the rate of elimination of viral cells (the bigger the E the faster this curve goes to zero).
Fitting this theoretical curve with the experimental data we can obtain the value of $E$ (see Figure 2.3).

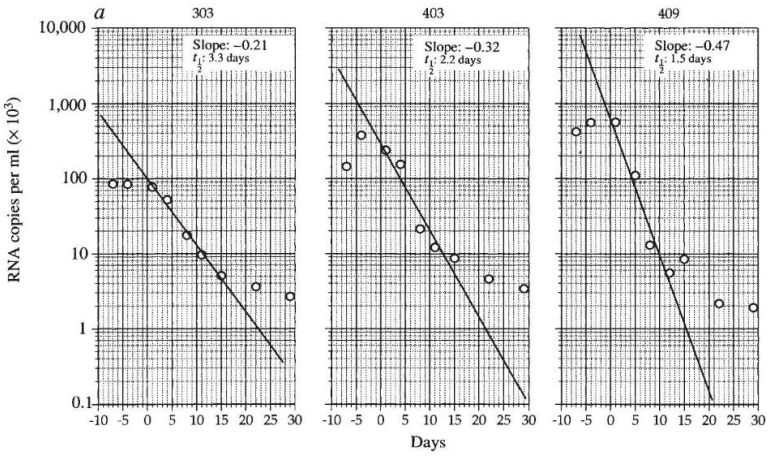



**Figure 2.3:** Data interpolation from Ho's experiment. Source: [3].

In this case they used the logarithmic scale in which the exponential has the shape of a straight line. We get this by just applying the logarithm to both sides of our equation.

$$\ln\left(V(t)\right) = \ln\left(V_e e^{-Et}\right)$$

$$\ln\left(V(t)\right) = \ln(V_e) - Et$$

7: See the next chapter for more details.

Which is the equation of a line, given the fact that $V_e$ is constant[7]. From the slope of the fitted data we can get a value for $E$. What they measured was

$$E = 0.34 \pm 0.06 \quad d^{-1}$$

Where $d$ stands for days, being the unit of measure cells over time.
From the data interpolation we can get an estimation also for $V_e$, which in this particular experiment was around $10^6$, $10^7$.
Using these data we may answer our research question, what is the dynamic of the infection, how fast is the virus reproducing in this latency period. The answer for this question is the parameter we called $P$, the reproduction rate. We know from our theoretical model that

$$V_e = P/E$$

Inverting this formula we get $P$

$$P = V_e E$$

From the experiment we measured $V_e$ and $E$. Thus, substituting

$$P \simeq 0.33 \quad 10^6 \quad cells/(mLd)$$

So we discovered that the virus is not quiescent, but it is reproducing at an high rate. This was a fundamental discovery that led to an important change in our understanding of HIV.

## 2.2 Conclusive remarks

You might have noticed that to build our model we simplified the phenomenon under study. We assumed many feature of the actual mechanism that are very unlikely. We assumed that the inhibitor was perfectly accomplishing its task, we assumed that in the latency phase (that now we may call pseudo-latency) the population of cells was constant over time, we didn't consider the interaction of the viral cells with the environment or with other cells of the immune system.
All of this was done on purpose. Considering the nature of a phenomena in all its complexity would certainly lead to a very complex model that would be impossible to solve. But that is not the point, and indeed that is never the point.
What we want to know it is the essence of a phenomenon, what are its basic characteristics, its fundamental structure. We can isolate all of these from the rest, since we want to isolate the signal from what its just noise. This noise is indeed part of the phenomenon, as nature is intrinsically complex and leads to such a variety, but it is not the most important part. This is the nature of modelling. We create models of the phenomena because we want to understand their nature at the deepest level, dividing signals from noise. And we use mathematics to translate our biological intuition into quantitative variables that we can measure via an experiment. The mathematical framework, in this sense, has two main advantages (among others). It enables us to test if our model is correct, if it matches actual data. So we know that (given due simplifications) our understanding of a phenomenon is essentially correct. But not only this.

Once the model is established it enables us to make prediction. In the same conditions, but in a new environment, we know from our model that the phenomenon will manifest as it did in the past experiments. This is a tremendous achievement and has vast applications in the life sciences (ecology, pharmacology...).

# Discrete mappings 3



## 3.1 Malthusian growth

Ecology is the science that studies the evolution, the growth, of living organisms in interaction with the environment in which they live. The word ecology derives from the Ancient Greek $o\iota\kappa o\varsigma$ which means "home" and $\lambda o\gamma o\varsigma$ which means "reason, word", but also "study, discipline". So ecology would literally mean the study of home. I think it is super cute.
So, imagine you are a researcher in ecology. Off you go for your new adventure in some wild environment. Your task is to study the dynamic of a population of a previously unknown animal in that environment. To do that, you register the number of animal in several time points, to have an idea of the change over time of the size of the population.
Unfortunately, your evil supervisor asks you to use the few data points you collected so diligently to make prediction on the evolution of the population over time. This task looks already more difficult. What can influence the dynamic of this population? Predators most surely have a strong impact. Also, the amount of available resources. Moreover, in that very environment there are thousands of more organisms that during their life cycles interact with each other, so all may have a more or less indirect effect on the growth of your target population. How do we solve this?
Let's start from a simple problem. Solving this might reveal the key to solve the more complicated original one. Amongst the many mechanisms that can change the dynamic of this population over time we will consider for the moment only the most basic ones: reproduction and death. After a certain and fixed interval of time some of those animals will have completed their cycle of reproduction, while some of them will be dead. Let's call $p_n$ the population at the $n^{th}$ interval of time. So $p_0$ would be the population at the beginning of our observations, $p_1$ the population after one reproductive cycle, and so on.
During a typical cycle a percentage $\alpha$ of individuals will give birth to $k$ new individuals. In the same time a percentage $\beta$ of individuals will die*. Given the fact that we assumed that just birth and death may influence the variation of the population, our model is determined by only these three parameters: $\alpha, \beta, k$.
The population at the next step will be the sum of several contributions: the previous population, plus the new individuals born, minus the individuals that died.

$$p_{n+1} = p_n + k\alpha p_n - \beta p_n$$

We see that we can rewrite this expression.

$$p_{n+1} = (1 + k\alpha - \beta)p_n = \mu p_n$$

* Clearly, in a real experiment those parameters will be empirically measured.

**Figure 3.1:** Thomas Malthus (The Rookery, 13 February 1766 – Bath, 29 December 1834). Clearly happy because he found a new equation. Source: Wikipedia.

Where $\mu = 1 + k\alpha - \beta$. In the end, the dynamic of this population is regulated just by the parameter $\mu$. This is known as the *Malthusian model*, from the reverend Thomas Malthus that first studied such a dynamic. Malthus is also sadly known for the political suggestions made in matter of births in a nation in virtue of the consequences of such an equation. Let's examine such consequences.
As we said, the entire dynamic is controlled by the parameter $\mu$. Knowing $\mu$ will enable us to predict qualitatively the entire dynamic of the population.
First of all, we notice that $\mu$ cannot be negative. This would imply that $p_{n+1}$ is negative as well, which does not make sense, since it represents the number of individuals of a population. If $0 < \mu < 1$ it means that the population in the next step $p_{n+1}$ is a fraction of the population in the preceding step $p_n$. Since $\mu$ is fixed in our model, this means that in any future step the population will keep being a fraction of the population of the step before, thus continuously decreasing until reaching the point of extinction. As a matter of fact, having $0 < \mu < 1$ implies that the death rate is higher than the birth rate. If $\mu > 1$ this means that the birth rate is higher than the death rate and the population keeps growing.
This is from an abstract point of view [1], but now that we developed the intuition, we would like to practically see how this population evolves in time[2]. We'll use R to plot the dynamic of this population.
At first, we have to define our parameters, in this case just $\mu$.

1: And it could be proven mathematically

2: Malthus' point was to control the births in order to prevent the humanity falling to extinction.

```
#Parameters of the model
mi = 0.7
```

Note that I chose $0 < \mu < 1$, so we already know what to expect after enough time has passed.
Our equation, as explained in the *Introduction* part, is a recurrence equations. So we have to decide how many times we want to apply the equation, how many birth cycles we want to observe.

```
#Number of cycles
cycles = 20
```

Now we have to create our main variable, the population.

```
#Declaration of varibales
p = 0
```

I just arbitrarily decided to set the variable to zero, just to create it. This is a vector that will contain the population in the different intervals of time, i.e. different cycles. Now, a very important point. We have to declare the initial population, the number of individuals present when we began our observations, our experiments.

```
p[1] = 20
```

Now we have to set the variable time, that will go from the initial observation until the last cycle we want to investigate.

```
#Declaration of the time variable
time = 1:cycles
```

This means that this vectors goes from 1 (first observation, first cycle) until the value of the variable *cycles*, i.e. the number of cycles we want to observe.
Now we have to implement our recurrence equations. We have to iterate

our function for the chosen amount of times (20 in this case). In order to do this, we'll use a *for* cycle.

```
#Iteration of the Malthus' equations
for (i in 2:cycles) {
  p[i] = mi*p[i - 1]
  }
```

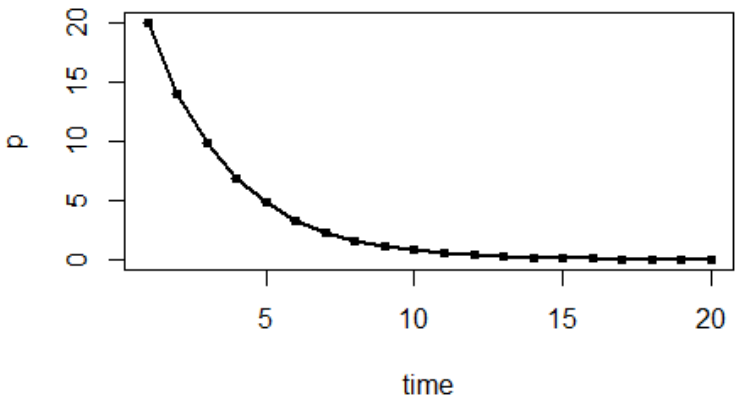
After all the iteration we want to plot our results, population in function of time, in order to have a visual representation of this dynamic. Our code should be this one, in the end.

**Listing 3.1**: Final script.

```
#Parameters of the model
mi = 0.7
#Number of cycles
cycles = 20
#Declaration of varibales
p = 0
p[1] = 20
#Declaration of the time variable
time = 1:cycles
#Iteration of the Malthus' equations
for (i in 2:cycles) {
  p[i] = mi*p[i - 1]
  }
plot(time, p, type="o", ylab="p", xlab="time", pch=20, lwd=2)
```

The following should be our results[3]. As you can see, the population reaches extinction.

3: Technical note here. "o" means that we want to overplot dots and lines, as you can see. "pch" is the type of dot we want in the plot, I chose the number 20 which are the squares. "lwd" is to specify the line width. You can search on Google for further options.




### 3.1.1 Exercises

1. Try with different values of $\mu$ and iterations and see what happens.
2. Try to run the model changing all the parameters. Describe all the possible main outcome scenarios.
3. Have a look at all the previous simulations, check the R output. How big is the size of the population at the end?

## 3.2 Logistic growth

You are very happy with your job, the model seems to work pretty well. However, there is something weird about it. You notice that the Malthusian model is not realistic in the long run. Consider for example the case of $\mu > 1$. If we wait long enough, the population is going to be infinitely big. This of course cannot happen. The Malthusian model can work well just in the short run.
As we said, in the long run the population is growing bigger and bigger and the environment in which this evolution is taking place sets a lot of constrains on the growth. Resources are limited and cannot sustain and infinite growth. Thus we need to take all of this into account in our model.
Let's go back to our equation.

$$p_{n+1} = p_n + k\alpha p_n - \beta p_n$$

We can take into consideration the environmental pressure letting our birth and death rate depending on the actual population: $\alpha = \alpha(p)$, $\beta = \beta(p)$. The new birth rate can be considered as constituted by an actual birth rate term $\alpha_0$ minus the environmental pressure depending on the population $ap_n$, as the bigger is the population the higher is the competition for the resources. In the end, $\alpha = \alpha_0 - ap_n$. A similar reasoning applies for the death rate. It is made by an actual death rate term $\beta_0$ plus the environmental pressure $bp_n$. In the end we have $\beta = \beta_0 + bp_n$. Substituting in our equation we have

$$p_{n+1} = p_n + k(\alpha_0 - ap_n)p_n - (\beta_0 + bp_n)p_n$$

Doing some calculations we get

$$p_{n+1} = (1 + k\alpha_0 - \beta_0)p_n - (ka + b)p_n^2$$


**Figure 3.2:** Pierre François Verhulst (Bruxelles, 28 October 1804 – Bruxelles, 15 February 1849). He's obviously looking away from Malthus as his model is just ridiculous. Source: Wikipedia.

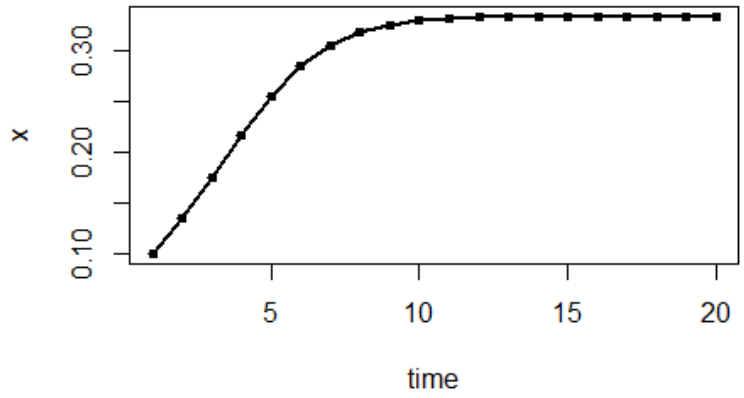
The model is ready, but we can further simplify it. Before proceeding with that, as we did, we need to see what can we predict. Taking into consideration the environmental pressure we easily see that the population cannot grow indefinitely. Limited resources, intra/inter-species competition and other mechanisms will allow the growth until a possible limit. The maximum population sustainable is a given environment is called *carrying capacity*. If the starting population is way below this value, then there is enough resources to allow for a growth until the maximum possible. If the starting population is above that limit, the environmental pressure will impose a decreasing population. These are the main features of the *logistic model*, first proposed by the Belgian mathematician Verhulst in the $19^{th}$ century.
We can simplify the equation, by just rescaling the population. We make the following substitution

$$p_n = \frac{1 + k\alpha_0 - \beta_0}{ka + b} x_n$$

into the main equation and we get

$$x_{n+1} = \mu x_n(1 - x_n)$$

where $\mu = 1 + k\alpha_0 - \beta_0$. The new variable may take values between zero and one, $0 < x_n < 1$, and represent a rescaled version of the real population. In order for also $x_{n+1}$ be in the interval $[0, 1]$, the parameter $\mu$ must be $0 < \mu < 4$.




### 3.2.1 Exercises

1. Write the program to plot a solution logistic equation;
2. Try different values of $\mu$ and see what happens;
3. What happens if $\mu < 0$ or $\mu > 4$? Do these solutions make sense?

### 3.2.2 The problem of the stability and the emergence of chaos

You have probably now noticed that the dynamic of the system changes dramatically with the parameter $\mu$. What does this mean?
First of all, we noticed that we have two main final outcome for the logistic growth. If the survival conditions are too bad, the population is doomed to extinction. Otherwise, the population may thrive and reach the maximum sustainable value, in normal conditions. So, two main solutions: reaching extinction or the carrying capacity.
Another thing we may ask is: once we reach one of these two conditions, for how long the system is going to stay in that state? In more formal words, we are asking ourselves if these two final populations are *stable* for this system.
In the case of the extinction, is easy to solve this issue. There is no one left, therefore no more cycles of reproduction. This means that the extinction is stable, is going to be like this forever[4].
For the other solution, the situation is quite different. We see that for values of $\mu$ inside the interval $(1, 3)$ the population seems to reach the equilibrium and then stay there for very long time. So for $\mu \in (1, 3)$ this

4: This is true of course, under the assumption of our model. It is clear that the immigration of other animals into that very environment may change this situation. This and other mechanisms may prevent extinction or restore the population there after the extinction.

equilibrium seems stable. However, if $\mu \in (3, 4)$ the final populations starts oscillating.

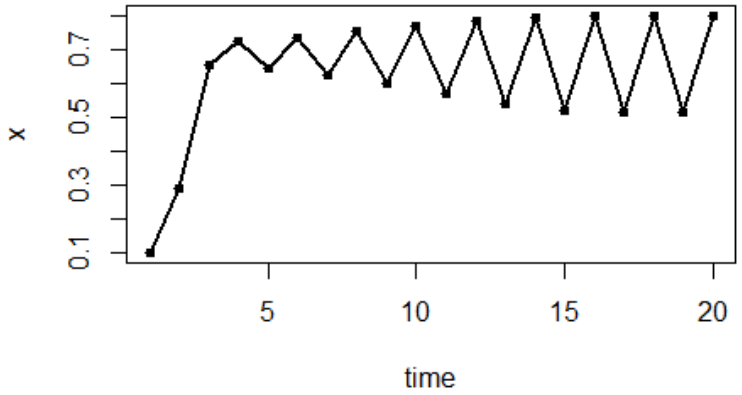



**Figure 3.3:** Robert McCredie May, Baron May of Oxford (Sydney, 8 January 1936 – Oxford, 28 April 2020). A shocked Robert May the moment when he discovered chaos. Source: Wikipedia.

It seems that for these values of $\mu$ the populations is orbiting around a certain value. So it doesn't reach a fixed value forever, but it keeps oscillating around it. You can see that the period of the oscillation is fixed. Now the situation is a bit more complex, but it seems to behave at least regularly. It is not fixed any more, but oscillating within certain boundaries.

If we keep increasing $\mu$ we will decrease the period of oscillation. And indeed the researchers a long ago thought that that was it. Increasing the value of $\mu$ is going to keep our population alive, just oscillating differently according to that parameter.

The problem is that this is just wrong. It has been discovered by the biomathematician Robert May in the Seventies, that after a certain value, the dynamic of the system is completely *chaotic*. We cannot foresee the behaviour of this population. It won't oscillate regularly, it will just reach all the possible values in a unpredictable manner. Eventually will reach extinction.

This was indeed a great discovery as it showed as that the chaos may emerge even in a very simple dynamical system. From a theoretical perspective this was a great progress. From a more practical perspective, we now know that in a real situation not only adverse conditions may lead to extinction. Good conditions may indeed increase the value of $\mu$ leading our system into the chaotic region. Once there the system is out of control and will behave in unpredictable ways (even going to extinction).

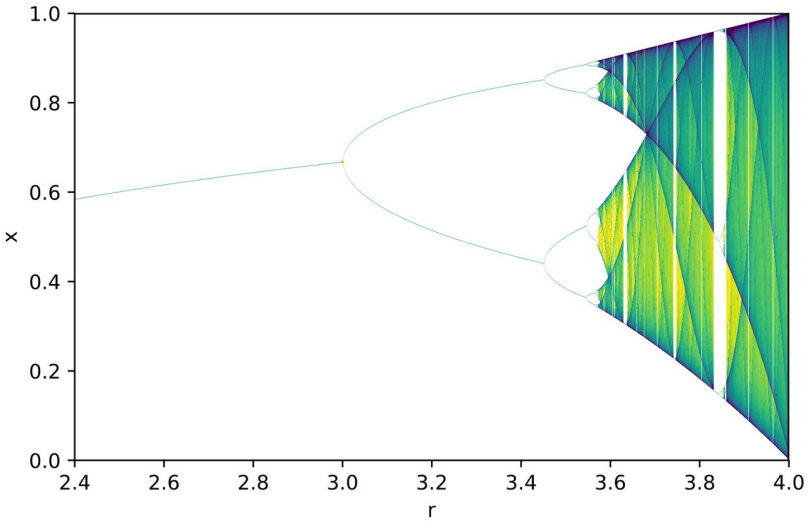



**Figure 3.4:** Logistic map plot as a function of the parameter $r$ ($\mu$ in our example). We can see here clearly the chaotic behaviour. Source: Wikipedia.

## 3.3 Conclusion and final remarks

In this part we studied two very important models. They are the basic models used in ecology and can be generalized easily to describe more complex situations.
Keeping in mind the assumptions of each models, one can use this mathematical tools to predict the behaviour of a certain system. This in the field of ecology is very important as it is a strong method we had to measure the impact of any given factor on a certain environment. We may be able to actually assess the impact of the human action on a certain population of animals in a certain region and see (even if approximately) the consequences of our actions. Do not underestimate the power of such models because even with such a simple mathematics we could have avoided many natural catastrophes.
The power of this model does not rely only on the prediction of future outcomes, but also on the fact that may describe many different phenomena that apparently have nothing in common with ecology. That is the beauty of mathematics. It enables us to see the hidden pattern beneath complex phenomena. And what happens is that very different phenomena are intrinsically very similar. Our task is to unveil the hidden nature of reality, and mathematics is one of our best ally.

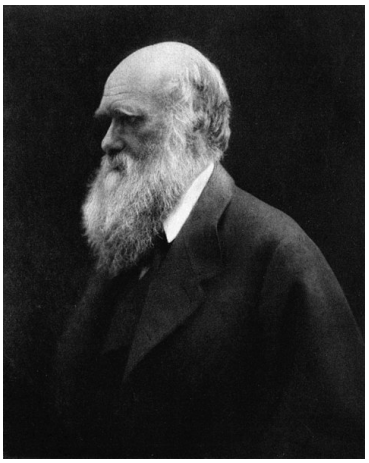


**Figure 3.5:** Charles Robert Darwin (12 February 1809 – 19 April 1882). An example of a famous ecologist. Source: Wikipedia.

## 3.4 Final exercise: growth and evolution

To become a famous ecologist, you have to somehow include the action of the evolution in your study. In this sense, you might want to go back to the simple Malthus model and add the evolution factor. Once again, let's start simple, that is the basis to build up more complicated things.
After rigorous assessments, you discovered that within the population you are studying there are two different sub-groups with slightly different phenotypes. After more accurate studies, you notice that actually these two different sub-groups have slightly different *fitness*. By fitness we mean the expected number of offspring for each individual on average.

Let's call our two sub-groups $P$ and $Q$. In the first case, we would have

$$P_{n+1} = f_P P_n$$

In the second it would be

$$Q_{n+1} = f_Q Q_n$$

Where $f_P$, $f_Q$ are the fitnesses of the sub-groups. The total population would be

$$N_n = P_n + Q_n$$

This information might be useful, because it might be just easier to talk in relative terms. So, we can consider the *frequencies* of the two types in the overall population.

$$p_n = \frac{P_n}{N_n}$$

And the other one...

$$q_n = \frac{Q_n}{N_n}$$

Write a script in R that shows the evolutions of the frequencies $p_{n+1}$, $q_{n+1}$ over time. Plot the solution on the same graph (you might want to use *matplot* to have both curves on the same graph). Discuss the results in terms of the theory of evolution.

# Continuous mappings 4



## 4.1 Epidemiology of infectious diseases

After the your field study away from home you decide to come back and change career. Ecology is not really your thing, so you decided to come back and think a bit about your future plans. You get yourself a beer, sit on the couch and watch the news. Soon you discover that a contagious disease is spreading in your town. You decide that you want to help. So you join the epidemiological team working on the case.
Let's assume that the diffusion of this infectious disease happens in a closed population of $N$ individuals. We will assume to be in the case in which the population is varying so that the continuous approximation is indeed a valid one.
We will make some assumptions in order to simplify our model:

- all the individuals are equally exposed to the infection;
- all the individuals are equally contagious;
- the probability of transmission of the disease $\tau$ is fixed;
- after the infection, the individual is immediately contagious;
- the individuals are contagious during our observation time;
- the infection doesn't change the behaviours of the sick ones.

This is indeed a very simple model, but for a limited amount of time this may represent good enough many actual infectious diseases.
Let's call $I$ the number of infected and $H$ the number of healthy people. We notice that given our assumptions $I+H = N$. Of course these numbers are going to change over time, so $I = I(t)$ and $H = H(t)$ are functions of time, but we will be still in the case of a closed system, so in any time it is true that $I(t) + H(t) = N$.
What we want to know is of course how is going to change over time the number of infected individuals. This of course will depend on the number of individuals already infected, by the number of people that could potentially be infected and by the probability to get the infection. Summing up what we already know, the variation of the number of infected individuals over time should be directly proportional to the probability of contagion, the number of people already infected and the number of healthy ones.

$$\frac{dI}{dt} \propto \tau IH$$

We set as $k$ the constant of proportionality and this will measure, together with $\tau$, the spreading of the infections.

$$\frac{dI}{dt} = k\tau IH$$

We can simplify further this equation, grouping together the two constants ($\lambda = k\tau$) and remembering that $H = N - I$. With this last substitution everything will depend on $I$ in the end.

$$\frac{dI}{dt} = \lambda I(N - I)$$

This equation already looks familiar. We can further simplify it. Let's make a change in variables: instead of speaking of the absolute numbers of infected individuals, we can just study the frequency of infected ones, i.e. $i = I/N$. This is a number in the interval $[0, 1]$, as $N$ is the overall population, the maximum number of people. In the end, we obtain

$$\frac{di}{dt} = N\lambda i(1 - i)$$

Putting together the constant, $\mu = N\lambda$, we obtain the usual equation of the logistic growth.

$$\frac{di}{dt} = \mu i(1 - i)$$

We already solved such an equation, in the discrete case. We should not assume that the behaviour of the solution in the continuous case would be exactly the same. For instance, the fact the we are dealing with the continuous approximation cancels out the emergence of chaos.
We should ask ourselves how the situation is going to change due to the variation of the parameter $\mu$. In this case we are dealing with the variation of the number of infected people, the derivative of the infected over time, while in the other we were dealing with the number of people in the next cycle of time. So in this present case, $\mu$ could be a negative or a positive number, as a negative derivative means just that our function is decreasing. So if $\mu < 0$ we are expecting that the number of infected people is decreasing. If $\mu > 0$, the derivative is positive and the number of infected ones is of course increasing. It is not going to increase forever, as here as well we have an environmental pressure: we have a maximum amount of people available, i.e. $N$. This means that maximum $N$ people may be infected, as our system is supposed to be closed.
In our model $\mu$ is used to take into account the rate of infections, so natural values would be the ones for which $\mu > 0$. Of course, if in another model we would take into account the action of drugs, vaccines or other measures to contain the spread, we would allow for $\mu$ to have negative values as well.
Let's try to solve this equation with R. First of all we have to install the package DeSolve.

```
install.packages("deSolve")
```

Now we can use the functions of that package to solve our equation.

```
library(deSolve)
```

At first we have to create a function that takes into account the general feature of our solution: the independent variable (the time $t$), the initial condition and the parameters.

```
logistic = function(t, state, parameters) {
  with(as.list(c(state, parameters)), {
    di =  mi*i*(1 - i)
    list(c(di))})}
```

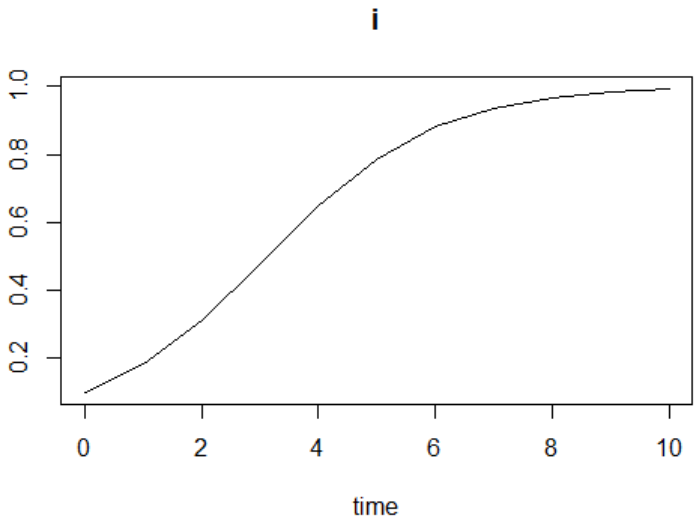
Now we should give the value of the parameter $\mu$ ($\mu = 0.7$, for example), the initial conditions, i.e. the initial fraction of infected people ($i = 0.1$, let's say) and then the time span of our observation as well as the spacing of the intervals of time (like until 10 months, where each step is of one month).

```
parameters = c(mi = 0.7)
state      = c(i = 0.1)
times      = seq(0, 10, by = 1)
```

In the end we call the function *ode* to solve the equation with the values we chose, and then print it. The whole code should be something like this

```
library(deSolve)
logistic = function(t, state, parameters) {
  with(as.list(c(state, parameters)), {
    di =  mi*i*(1 - i)
    list(c(di))})}
parameters = c(mi = 0.7)
state      = c(i = 0.1)
times      = seq(0, 10, by = 1)
out = ode(y = state, times = times, func = logistic, parms =
    parameters)
plot(out)
```

This is the result.



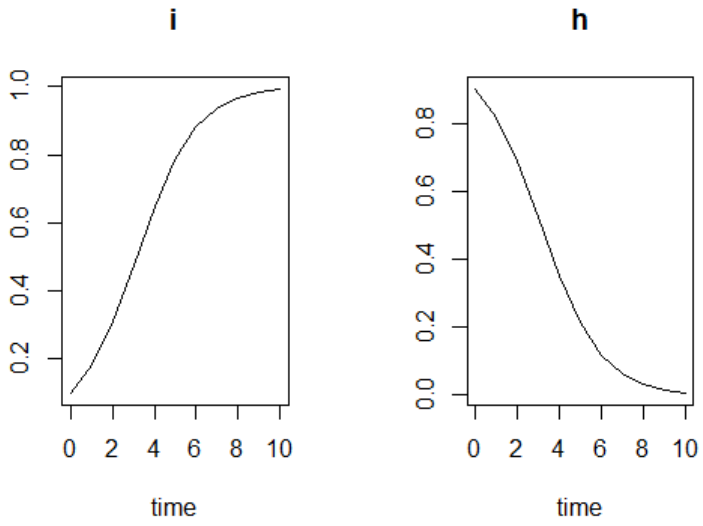

Since we know that the healthy people are related to the infected ones via the equation $i + h = 1$, we may write an equation for the healthy people as well. Doing the substitution $i = 1 - h$ into the logistic equation for the infected individuals we get an equation for the healthy as well

$$\frac{dh}{dt} = -\mu h(1 - h)$$

We can include this equation in our code as well and see the evolution of this part of the population as well. What we expect is, of course, that

when the number of infected is rising, the number of the healthy people is decreasing at the same pace.

```
library(deSolve)
logistic = function(t, state, parameters) {
  with(as.list(c(state, parameters)), {
    di =  mi*i*(1 - i)
    dh =  -mi*h*(1 - h)
    list(c(di, dh))})}
parameters = c(mi = 0.7)
state      = c(i = 0.1, h = 0.9)
times      = seq(0, 10, by = 1)
out = ode(y = state, times = times, func = logistic, parms =
    parameters)
plot(out)
```




### 4.1.1 Exercises

1. Try to change the parameter, initial condition and timing and see what happens.
2. Again, as in the discrete examples, the dynamic of the system seems to be bound to certain trajectories. Can you guess which ones?

## 4.2 Kermack-McKendrick approach: SIR Model

1: The SIR model can be further generalized in many different other models.

Our first model for the infectious diseases is quite simple. It could be improved in the SIR model, a very important model in epidemiology[1]. In this framework the total population is divided into three subgroups: the susceptible $S$, the infectious $I$ and the recovered $R$. We will assume that the following is the dynamic of the infection

$$S \to I \to R$$

This means that people from the susceptible may get infected and after that they may heal and gain permanent immunity. Before proceeding any further we should make some assumptions

- there is no time of incubation, each infected is immediately contagious;
- the probability of meeting with any member of any group is the same;
- probabilities of contagion and healing are constant.

In order to have a complete picture of this phenomenon we have to understand the evolution in time of all the three subgroups. Starting from the susceptible ones, we know that the rate of variation of the people in this group will be directly proportional to both the actual number of susceptibles and the number of people already infected. This means the bigger the $S$ group is, the easier is going to spread the infection, as well as because of the higher is the number of infected $I$. We will introduce the (positive) constant of proportionality $\alpha$ that will measure how fast is spreading the infection. Summing up we have

$$\frac{dS}{dt} = -\alpha SI$$

The minus sign is due to the fact that because of the spreading infection the number of susceptibles is decreasing, so the rate of variation must be negative.
At the same time, all the people going away from the $S$ group are going to join the $I$ group, i.e. the susceptibles get infected. So $I$ is gaining the same amount we lose on average in $S$ ($\alpha SI$). In the meantime, some infectious individuals are getting cured, so basically they are going from the group $I$ to the group $R$. Let's call $\beta$ the average rate of healing. Putting everything together we get

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

Now, we have to consider the last movement. All the people that get cured, leave the group of the infected ones and go to the group of the recovered.

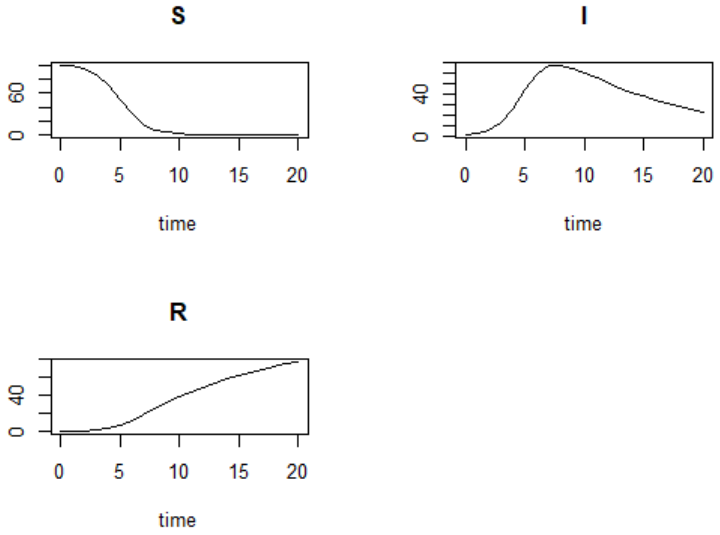
$$\frac{dR}{dt} = \beta I$$

We can see easily that all these equations are strongly entangled, as any movement from one group to another depends on what is happening on the others. Thus we have to solve this system

$$\begin{cases} \frac{dS}{dt} = -\alpha SI \\ \frac{dI}{dt} = \alpha SI - \beta I \\ \frac{dR}{dt} = \beta I \end{cases}$$

Now that we have the system we should try to predict what may happen in this conditions. The number of susceptibles is going to decrease steadily, because the infection is going to spread faster and faster as the number

of infected ones is increasing. The number of infected ones is going to increase, for the reason already sated, but cannot increase forever. First of all the amount of people is limited, as we are in a closed system ($N = S + I + R$). Second, at a certain pace (but continuously), the infected ones get permanent immunity once recovered. So the infected individuals are going to reach a maximum level and then we expect this number to decrease as the number of people with permanent immunity is increasing as well. For the last group the situation is pretty straightforward now. The recovered are going to increase with time and eventually the population in its entirety is going to be permanently immune.
This is more or less the result




### 4.2.1 Exercises

1. Try to write down the code for this system and change the values of the parameters $\alpha$ and $\beta$, the initial conditions ($S_0 = 100$, $I_0 = 1$, $R_0 = 0$, for example) and the time of observation and see what happens.

## 4.3 Final exercise: VEGF receptor inhibitor

2: The argument is adapted from [4].

3: This argument can be found at the beginning of Chapter 14 of [4].

In this exercise we try to create a model for cancer therapy[2].
There are many types of drugs used in cancer therapy, in this section we will focus on a particular subgroup of drugs that aim at making hostile the environment for tumor growth. The story goes as follows[3]: among other things, the tumor needs oxygen in order to grow. To accomplish this purpose the tumor produces vascular endothelial growth factor (known also as VEGF) which in turn attracts endothelial cells to form new blood vessels. These new blood vessels will supply more oxygen and various nutrients to the tumor.
So far the main actors in this drama are tumor cells and endothelial cells. Therefore we indicate with $c$ the concentration of cancer cells and with $e$ the concentration of endothelial cells. With $v$ we then indicate the

concentration of VEGF and with $w$ the concentration of oxygen[4]. Let's start then to build up our model.
For the rate of variation of cancer cells we might consider two most important factors, as usual: generating and deleting factors. The equation should look like the following

$$\frac{dc}{dt} = \lambda_1 ec\left(1 - \frac{c}{K}\right) - \delta_1 c$$

Let's begin with the easy part. $\delta_1 c$ is the elimination part where the $\delta_1$ represents the death rate of cancer cells. The generating part is the usual logistic growth. In this form, $K$ represents the carrying capacity. Instead of the usual constant growth factor $\mu$, here we have a changing growth factor $\lambda_1 e$. This means that the general pace of growth is $\lambda_1$, but this term is modulate by the concentration of endothelial cells $e$. Higher the concentration of these cells, faster the growth. The equation for the VEGF is similar, but the production is only due to cancer cell proliferation

$$\frac{dv}{dt} = \lambda_2 c - \delta_2 v$$

Oxygen tends to be consumed by both tissues (with a rate $\delta_3$) and by cancer cells (with a rate $\delta_4$)

$$\frac{dw}{dt} = -\delta_3 w - \delta_4 cw$$

The growth of the endothelial cells is controlled by the following equation

$$\frac{de}{dt} = \lambda_3 v - \delta_5 ce - \delta_6 e$$

This is the model for the normal proliferation. Now we consider the drug as well. Avastin, in particular, is a drug that inhibits VEGF receptor. A way to model this would be to replace $\lambda_2$ with $\lambda_2/(1 + A)$ where $A$ is proportional to the amount of drug delivered to the patient. Therefore, the final equation would be

$$\frac{dv}{dt} = \frac{\lambda_2}{1 + A} c - \delta_2 v$$

Write down the script to provide a numerical solution for this problem. Modify the model to have a more realistic picture of reality.
Suggested parameters: $\lambda_1 = 0.7$, $\lambda_2 = 0.3$, $\lambda_3 = 0.1$, $\delta_1 = 0.2$, $\delta_2 = 0.1$, $\delta_3 = 0.13$, $\delta_4 = 0.12$, $\delta_5 = 0.01$, $\delta_6 = 0.02$, $K = 1.7$, $A = 100$.

4: I didn't choose the letter $o$ as it is a bit unfortunate in mathematics.

# Conclusions 5

In this brief lab we discussed some examples of application of mathematics in biology. All the models we discussed are very simple, but can allow us to have a deeper understanding of the phenomenon we are studying. But of course there are more sophisticated methods that one can use to investigate reality. One could use probability theory to incorporate the role of uncertainty. One could use the tools of game theory to study human or animal behaviour. It is possible to study human physiology taking into account more than one variable and solving partial differential equations. The possibilities are endless. Mathematics can offer a gigantic variety of analytical tools to address our scientific questions, to formalize the phenomenon we want to study, to get a deeper understanding. With the power of mathematics you could really study anything, from the tiniest of the cells replicating in a Petri dish up to the complex features of evolution along the millennia. From simple to complicated, from deterministic to random, mathematics can offer you the grammar to translate your biological ideas into quantitative models. These models will give you unique power of understanding the phenomena you are studying.

However, mathematics is much more than that. Much more than numbers, formulas and equations. It is a way to formalize in a rigorous way your thinking. It is a general approach, pure thinking. Mathematics is most of all a method, a method in the original sense. It is enquiry, it is a path. Methods derives from the Greek words $\mu\epsilon\tau\alpha$, which conveys the idea of proceeding, and $o\delta o\varsigma$, which means road. It has the same meaning of the Chinese character Dao, the Path. It is the path to reach higher understanding, in the realm of nature. As the quote from Amalie Noether at the beginning suggests, it is so general that you can go anywhere.


**Figure 5.1:** Dao. Source: Wikipedia.

**Figure 5.2:** Amalie Emmy Noether (Erlangen, 23 March 1882 – Bryn Mawr, 14 April 1935). In my beginning is my end. Source: Wikipedia.

# Bibliography

Here are the references in citation order.


[1] Giuseppe Gaeta. *Modelli Matematici in Biologia*. Springer, 2007 (cited on page 9).

[2] James Dickson Murray. *Mathematical Biology, Vol. I and II*. Third. Springer, 2002 (cited on page 9).

[3] David D. Ho et al. 'Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection.' In: *Nature* 373 (1995), pp. 123–126. DOI: https://doi.org/10.1038/373123a0 (cited on pages 10, 12).

[4] Ching Shan Chou and Avner Friedman. *Introduction to Mathematical Biology - Modeling, Analysis, and Simulations*. Springer, 2016 (cited on page 28).