MMath Group Project - Mathematical Virology

Harrison O'Neill, Matthew Knowles, Sabrina Duxbury, Yuanbo Liang

Department of Mathematics,
University of York,
United Kingdom

Contents

Chapter 1

On Differential Equations in Molecular Virology

1.1 Introduction

1.1.1 Biological Background of Molecular Virology

Before diving into the maths of pathology, we will briefly look at some of the biological processes involved, as a way of giving context to the maths that we explore in this chapter. Viral Pathogenesis is a very complex process with lots of varying factors. It can be somewhat boiled down to a balancing act between the host and the virus. The goal of any virus that wants to survive is to hijack the cells in a host and replicate as much as possible. This is done by a process known as *replication.* [2]. It can take a while for this balancing act to reach equilibrium, and often the case is that viruses naturally become less deadly the longer they are around, in order to preserve a population of potential hosts. Those that kill off their hosts too quickly, inadvertantly kill off theirselves.

There are three stages involved in replication, **Initiation of Infection**, **Replication and Expression of the viral genome** and **Release of mature virions**. The modelling we will undertake primarily focuses on steps 1 and 3.

Viruses are unique in that they bind to specific genes in a host, for example, we will discuss Hepititis B at the end of this chapter, which only binds to liver cells, whereas a virus like SARS or MERS are respiritory illness and so bind to receptors in a host's lungs and trachea. This fact is useful when it comes to actually computing the models we describe here, because it means we can reduce the size of the parameters required- a virus that infects the large intestine will require bigger parameters than one that infects the kidneys.

Virions are the individual particles that bind to certain points on the genome of the cells in the target tissue. Virions may not always get this far if the target cells are aware of the incoming cell as a virus and can prevent binding. This is what happens if someone has immunity by either already having had the virus or being vaccinated. Once bound to the genome they can activate certain genes. This is the 'hijacking' part of the process, as it will cause the infected cells to produce more home-grown infected cells instead of the healthy ones it should be producing. A lot of statisitcal pattern recognition is going on to help develop treatments for viruses and cancers by identifying genes that viruses target, allowing treatments to fight an infection from a genomic level- but we won't discuss that in this chapter. [1]

---

[1]See the work of Dr. Andrew Holding, Department of Biology, University of York - www.andrewholding.com

Now that the biological context for this work has been established, we can begin to look at the mathematical models being employed to describe these systems.

1.1.2   Overview Differential Equations in Molecular Virology

Differential Equations are very relevant in applied mathematics, Virology is no exception. There are two main types of differential equations we will look at: ordinary (ODE) and stochastic (SDE). We assume the reader is familiar with ODEs, but SDE's are slightly more specialised. [4].

Definition 1 *A **Stochastic Differential Equation (SDE)** is a differential equation of the form*

$$dX_t = b(X_t, t)dt + \sigma(X_t, t)dW_t$$

To dissect this definition a bit more, it can be broken into the two constituents:

- $b(X_t, t)$ is the **drift coefficient** of the equation. It describes the deterministic part of the equation.

- $\sigma(X_t, t)dW_t$ is the **diffusion coefficient**. It describes random motion by being a collection of random variables, which is what seperates SDEs from ODEs. If this term was not here, we would simply have an ODE.

We will return to stochastic processes later, but first we take a look at a deterministic model.

1.2   An ODE model for in-host viral dynamics

1.2.1   Deriving the model

Consider a human being. We can split the molecular constituents of this person into 3 categories: **Uninfected cells**, **Cells already infected cells** and **Free virions**. We denote the amount of these present at a time t as $T(t)$, $I(t)$ and $V(t)$ respecitvely. The body of the person in question will produce fresh uninfected cells at a rate λ, however these cells will be removed from this category if they die or become infected. We use k to denote the rate at which uninfected cells become infected per virion. Secondly, these cells die naturally at a rate of μ_T. In the same way, μ_I is the rate at which infected cells are removed, the three options for this are natural cell death, immune response and *lysis*, which is where fresh virions are ejected from the cell, breaking the membrane and killing the cell. Similarly, μ_V is the rate at whcic virions are removed from the host. Finally, N gives the average number of new virions produced during the life of an infected cell.

With all of this in mind, we can derive differential equations for T, I, and V. Starting with T, we have a constant λ feeding new cells, but then we are removing them at a rate k per virion, and a proportion μ_T die of natural cell death. So we have:

$$\frac{dT}{dt} = \lambda - \mu_T T - kTV$$

Then for I, the kTV term from the above becomes positive, because these cells are going from uninfected to infected. These cells also die off at a rate of μ_T which encompasses natural death and the immune system. Putting this together gives:

$$\frac{dI}{dt} = kTV - \mu_I I$$

Finally, for V, we have to include N and μ_I into one term, because N virions can only be ejected from a cell if it dies. THen again we have the term μ_V for the removal of these cells from the host.

$$\frac{dV}{dt} = N\mu_I I - \mu_V V$$

All that is left to do is combine these three into a sytem of ODEs which we call the **Standard Model for In-Host Viral Dynamics** [6].

$$\begin{cases} \frac{dT}{dt} = \lambda - \mu_T T - kTV \\ \frac{dI}{dt} = kTV - \mu_I I \\ \frac{dV}{dt} = N\mu_I I - \mu_V V \end{cases} \tag{1.1}$$

1.2.2 Model of an Unifected Person

Before someone is infected, they have no free virions. With $V = N = I = 0$, we now model the person with two equations.

$$\frac{dT}{dt} = \lambda - \mu_T T \tag{1.2}$$

The solution for this system is given as

$$T(t) = Ae^{-\mu_T t} + \frac{\lambda}{\mu_T} \tag{1.3}$$

Since there are no infected cells, and no free virions, we just have the number of target cells increasing exponentially. Which makes sense because one's body creates more cells than it kills off- or at least, ideally that's what happens.

We can plot a sketch of what this looks like with values $\mu_T = \frac{1}{10} * \lambda$. This can be seen in figure 1. Clearly the values are restircted by my computing power, but the shape of this curve will be the same on an axis ranging across more accurate values.
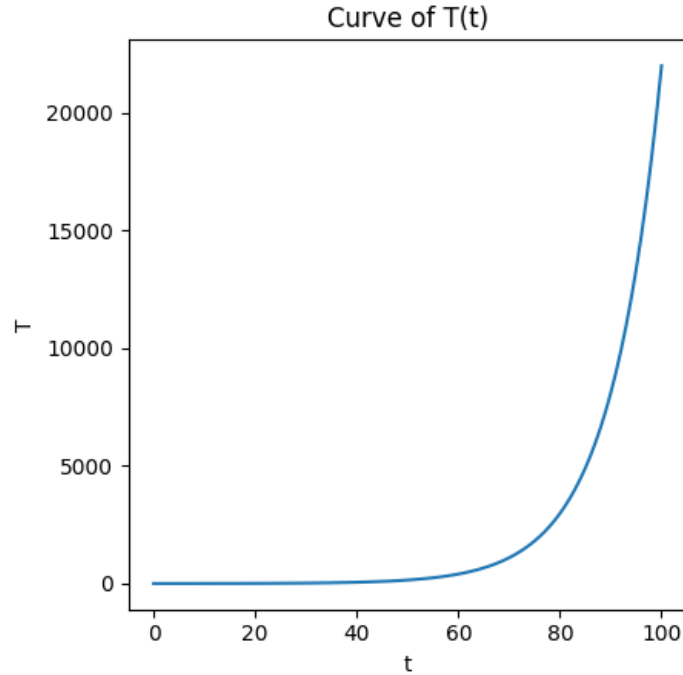
Figure 1.1: Sketch of T(t)

So we see that the exponential shape as mentioned. This is clearly slightly simplified, to make a more realistic model, we wouldn't have $\frac{d\lambda}{dt} = 0$. As a person gets older, their body will produce less cells, which would cause the graph to curve off and eventually stop.

1.2.3   Model for an Infected Person

We now turn to someone who has caught a virus. There are now V free virions present. The important factor is N. To see just how much this factor varies the results, a plot of this system was created in the Python programming language.[2]

We first look at positive N. When N is 1, each time an infected cell dies, only 1 new virion is released, analagous for N=10 and 50. The plotted graphs for these models can be seen below:

We can see that these equations result in a oscilating graph in each case. These coupled differential equations show how the amount of infected vs target cells changes over a given time period. When N is lower, the initial oscillation is greater. From a biological standpoint, this makes perfect sense as the host's body won't be as agressive in attacking the pathogen. In the scenario where N is lower, we can see from the above models that the infection is sustained for longer, whereas when N is larger, the host will be more aggressive in fighting the pathogen, and so the infection won't last as long. Through genetic variations, this is where the virus aims to find a balancing act in performance.

Further, note that here we have not considered medications and treatments, this will be discussed in a later chapter.

---

[2]Python was selected due to it's efficiancy and helpful modules which makes solving these equations much easier.
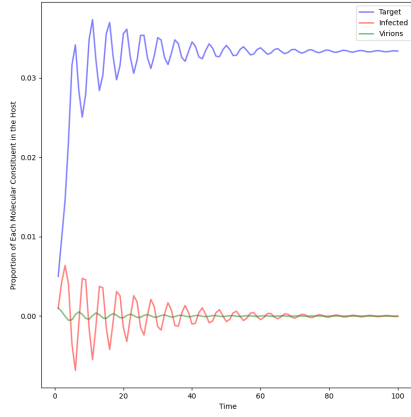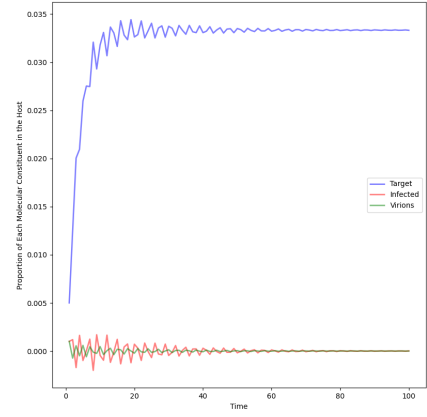
Figure 1.2: ODE Model with N=1
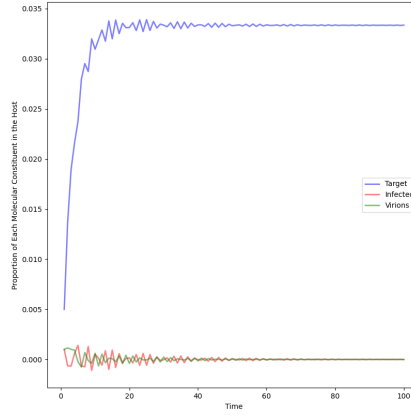

Figure 1.3: ODE Model with N=10

Figure 1.4: ODE Model with N=50

1.2.4 Advancing the ODE model - Parameter Estimation

Differential equations are a great tool for this kind of scientific enquiry, and the model we have seen above has demonstrated it's power in acurately describing reality. But how can we improve it? Firstly, it depends on the virus being modelled. There are many viral infections, HIV, Dengue, SARS-CoV2 to name a few. These individual viruses behave differently internally and externally, the way we model this difference is through th parameters given to the equations. This can be seen intuitively by comparing something like HIV and SARS-CoV2. HIV is an immunodeficiency virus, this means that it destroys the hosts immune system, so our parameters like λ and μ_I will actually decrease over time as the virus takes hold. Comparitively, SARS-based viruses are respiritory illnesses, so they dont taget the hosts defense systems in the same way, and so these parameters will remain roughly the same throughout.

These parameters are derived from analysing large amounts of data. But this can be improved as time goes on by using Bayesian Statistics. This is discussed extensively in a 1989 paper by Zeger et. al [7], in which they focus on the AIDS epidemic in the USA in the 1980s. By using log-linear models, they are able to estimate trends and numbers in the epidemic whilst correcting for new incoming data. It is this correcting aspect which is most relevant to our dis-

cussion. By correcting the models we can get a set of more accurate parameters which can then be used to fine tune our differential equations. This is less to do with the models themselves but it is useful to see how this works. Before doing this, we recall an incredibly powerful theorem:

Theorem 2 *Bayes: Let $\{H_1, \ldots, H_k\}$ be a partition of an event-space $\mathbb{H}$ then for some other event E, Bayes' Theorem gives us:*

$$Pr(H_j|E) = \frac{Pr(E|H_j)Pr(H_j)}{\sum_k Pr(E|H_k)Pr(H_k)} \tag{1.4}$$

Example 3 *Parameter estimation for μ_I*
Suppose there is some disease circulating thorughout a population. We observe some dataset y of infection data. Our parameter of interest is μ_I then Theorem 2 gives us:

$$p(\mu_I|y) \propto p(y|\mu_I)p(\mu_I).$$

Since infection data for epidemics follows a normal distribution, the probability density functions can be subsituted in and the necessary calculations made from this, although this is beyond the scope of this chapter.

1.3 The Stochastic Differential Equation Model

We met the definition of a Stochastic Differetnial Equation (SDE) in the overview, definition 1. We will be focussing now on the $\sigma(X_t, t)dW_t$ term. Firstly, we ask "why use SDEs when we have seen that ODES give a good model?". Intuitively, there is an element of randomness to pathology, just because you have a disease, it doesn't mean you have a 100% chance of passing it on if you come into contact with an infected person. Analagously, a pathogen might not necessarily make it all the way through the pathogenesis process of infecting a cell. So introducing the randomness term can refine our models nicely.

1.3.1 Outline of a Stochastic Process

The σ term is a *Stochastic Process*. We introduce the following definition for a stochastic process as given in [1]:

Definition 4 *A stochastic process is a family of random variables $\sigma(t)$ parametrised by $t \in T \subset \mathbb{R}$. When T is a continuous inteval in $\mathbb{R}$ (As in our case) then we call σ a stochastic process in continuous time.*

In the above definition, we saw how $\sigma(t)$ is a family of random variables, but in our case, which ones?. To get the answer to this, we return to the processes involved in modelling molecular virology.

1.3.2 Deriving the Stochastic Process for Molecular Virology

This subsection is considerably more statiscs-based than the rest of this chapter, although this is very necessary due to the structure of an SDE- you can't have one without the other. Before proceeding, we need the following definition.

Definition 5 *A random variable Y has Poisson Distribution with parameter θ if it has PDF:*

$$p(Y = y|\theta) = \frac{\theta^y}{y!}exp(-\theta).$$

Poisson random variables are very useful for the scenario we are describing- this is because by their nature they do a good job of modelling counts of independant events which occur within the same timeframe and within regions of the same dimension. With this in mind, we can now imporve the model we described in section 1.2.1.

It is clear to see that a lot of the variables we looked at would follow a probability distribution based on other factors. For example, the number of free virions being released from a cell which is infected is N, follows a poisson distribution with parameter which I denote as θ_N.

In addition, we previously described k as a rate at which unifected cells become infected. This rate will likely not remains constant throughout an infection. In a case without medication, the ideal scenario is that $\lim_{t\to\infty} k = 0$ naturally. This means k follows normal distribution with parameters $mu_k, \frac{2}{k}$. At the start of the infection, k will be quite low as the virus begins to take hold, before peaking at the height of the infection and then dropping back down again when the host begins to fight back. Of course, with the introduction of a vaccine or treatment the curve of this distribution will be much flatter, and a lot more compact.

Next up is the rate μ_i, which also follows a normal distribution in the same way k does, but it will 'lag behind' k so to speak. It does this because it may take a while for the body to actually realise it is infected and launch an immune response. So it will take a time difference of δt before μ_I begins to go up. Ideally however, if the curves were fitted onto the same axis, they would cross at some point, at that point, it is at that point the host will start to feel better and the body begins to win the battle against the infection.

Finally we have μ_T which doesn't follow a normal distribution as the others do, but it actually follows an exponential distribution.

Definition 6 *A radnom variable Y has Exponential Distribution if the PDF is:*

$$P(Y = y|\theta) = \theta exp(-\theta y).$$

It's easy (but also a bit morbid) to see why μ_T follows this distribution. The below figure is an exaple plotted in R over a sequence of 1 to 500 and $\theta = 0.2$:

So naturally the actual values on this are exemplified but μ_T follows the same shape- by the time someone reaches the age of 80 they are producing cells at a much slower rate than someone at the age of 8. One may raise a point along the lines of 'Well, you are modelling a lifetime-scaled value in the case of micro system, surely μ_T won't have actually changed that much', and this is a valid point, but because this distribution is continuous, the value of μ_T will have decreased by the time someone is cured of an illness, so it is still important- in the persuit of accuracy- to model this rate with a distribution, as we have done in the others.

So now we actually have all the ingredients for our model. All the rates at how the infection evolves can be encompassed in the σ term of an SDE, and we have a slightly more accurate model for virology.

1.4 Stochastic vs Differential, Which is better? What's the point?

One may say, 'well if the stochastic model is more accurate why bother with the deterministic one?' and this is another reasonable point, but by reffering to the definition of an SDE we can
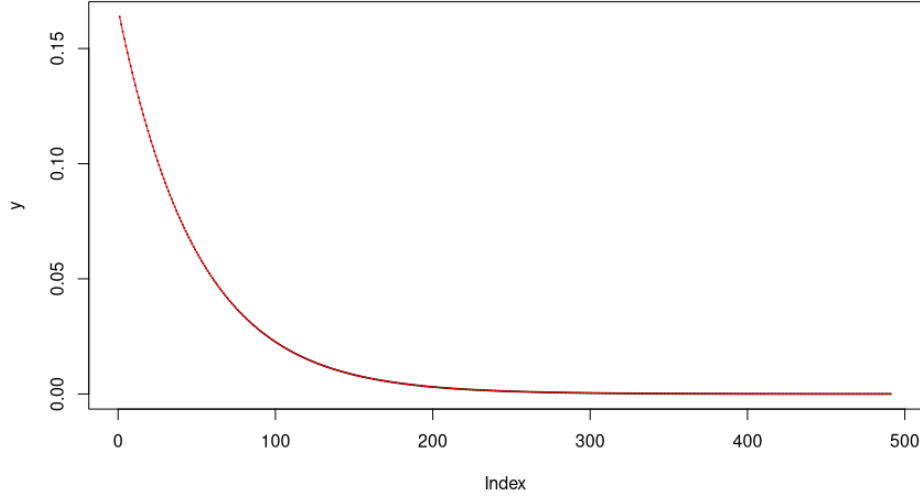
Figure 1.5: Example of an exponential distribution with $\theta = 0.2$

answer this question very inuitively. An SDE is simply an extension of an ODE, so it means if the σ term relating all the distributions of the parameters is small, (which we refer to as there not being much noise) there is no point going through the extra effort to calculate it. This is due to things like computational complexity and the need for the extra 'meta' parameters such as θ_N.

Infact, after considerable numerical simmulation, it was found in a 2017 paper by Mahrouf et. al, [5] that in the case where there is not much 'noise', it is just as well to use the deterministic model as outlined in the first section. I do not have the technical expertise to reproduce these graphs but I do reccomend them to the reader, in particular figure 3, as it gives a good demonstartion of this result.

On the contrary, in the event that there is a considerable amount of noise, which could be caused by a virus being particularly nasty, or a host being partularly ill, the deterministic equation is simply not enough to model the scenario.

It goes without saying modelling in epidemiology has been very relevant, and in the fight against pandemics the macro-scaling models of infection are being used to fight infections, so where does the work we have outlined here come into the fight? The main way this will be used is in the development and assessment of treatments/vaccines, knowing how different parameters such as age affect how a hosts' body will respond is very beneficial in this fight, and so the work being done in this area is abundantely important, not just with the current pandemic, but fighting viruses like HIV and Dengue, as we mentioned earlier.

A lot of work in this field is going to finding out when direct solutions exist to the models we have described, and further, what conditions there has to be on the parameters for solutions to exist. This is particularly crucial in helping with the work described above, because it means we can cater treatments to individual people's needs based on how far along the infection cycle.

9

1.5 Real World Application: Deterministic Model for Hepitis B

Throught this chapter, I have name-dropped different virusses as examples, but we have just talked about 'a virus', so to round this chapter up I want to look at a paper from the Journal of Mathematical Biology, in which they discuss Hepatitis B, and its dynamics. [3]

Hews et al [3] begin by defining the original model used for modelling Hepitis B infection, which is in the form of that in (1.1):

$$\begin{cases} \frac{dx}{dt} = r - dx(t) = \beta v(t)x(t) \\ \frac{dy}{dt} = \beta v(t)x(t) - ay(t) \\ \frac{dv}{dt} = \gamma y(t) - \mu v(t) \end{cases} \tag{1.5}$$

Here, x is the healthy cells, more specifically hepatocytes, which are cells making up liver tissue. y are cells infected with Hepitits B, and again v is the number of free virions. r gives the rate at which new hepatocytes are produced, and removed at rate d. α, β give the rate at which infected cells die, and the *mass-action constant* respecitvely. μ is N in our general model, and γ is the rate new virions are made. So it's clear how the model we outlined has been put into practice.

In figure 1 of the paper, they show a case of chronic Hepitis B, which gives a graph looking very similar to the ones in Figures 2,3,4 of this chapter. Some of the other diagrams given in this paper we don't have analogies for but they tell us a lot about infection dynamics.

To advance their, model, they introduce a new parameter- $R*$ the *The Cellular Vitality Index*, which is a way of describing the behaviour of healthy cells more succinctly. They conclude that the difference between R_0 and $R*$ will tell us a lot about how an infection will play out. Recall R_0 is the basic reproduction number of an infection. As a perfect example of what I mentioned earlier about these models being used to improve treatments- the authors of this paper suggest that novel treatments should look to improve and maintain the value of $R*$ in a host, in order to allow the liver tissue to regenerate and fight an infection.

I believe this model goes a long way in showing the power of mathematical biology and it's applications to medicine. We have seen how we can use the models at our disposal to reccomend treatments, just by evaluating the stability of a differential equation. The more accurate we make these models, the more people we can help.

---

[3]Hepatitis B is a nasty liver infection which in the worst cases causes liver faliure and cancer.

Bibliography

[1] Zdzislaw Brzezniak and Tomasz Zastawniak. *Basic stochastic processes: a course through exercises*. Springer Science & Business Media, 2000.

[2] Alan Cann. *Principles of Moclecular Virology*. Elsevier Academic Press, 2015.

[3] Sarah Hews, Steffen Eikenberry, John D Nagy, and Yang Kuang. Rich dynamics of a hepatitis b viral infection model with logistic heptocyte growth. *Journal of Mathematical Biology*, 60(4):573–590, 2010.

[4] Miranda Holmes-Cerfon. Lecture notes on stochastic differential equations.

[5] Maroane Mahrouf, El Mehdi Lotfi, Mehdi Mazaine, Khalid Hattaf, and Noura Yousif. Stability analysis for stochastic differential equations in virology. *Journal of Advances in Mathematics and Computer Science*, pages 1–12, 2017.

[6] Stephen Pankavich and Christian Parkinson. Mathematical analysis of an in-hot model of viral dynamics with spatial hetrogeneity. *arXiv preprint arXiv:1508.07617*, 2015.

[7] Scott L Zeger, Lai-Chu See, and Peter J Diggle. Statistical methods for monitoring the aids epidemic. *Statistics in Medicine*, 8(1):3–21, 1989.