SWANSEA UNIVERSITY

DISSERTATION PROJECT

Connecting Stochastic Simulations and Deterministic Models of Chemical Reactions

---

Author:
James CHALK

Supervisor:
Dr. Noemi PICCO

A project submitted in fulfillment of the requirements for the degree of Bachelor of Science

in the

Mathematical Medicine Group,
Faculty of Science and Engineering

April 5, 2024


Prifysgol
Abertawe
Swansea
University

SWANSEA UNIVERSITY

Abstract

Faculty of Science and Engineering

Bachelor of Science

Connecting Stochastic Simulations and Deterministic Models of Chemical Reactions

by James CHALK

In this project on mathematical modelling of chemical reactions, I investigate the link between Stochastic Simulation and Deterministic Models. I create MATLAB models of stochastic simulations and compare them by plotting them next to a deterministic Master ODE, which I have derived. The code that I have written has been included in this text where necessary to explain and support my arguments. However, for more basic examples or variations of models I have included all of my code in a bulk file which is linked to in the appendix (A) should it interest the reader. I have introduced some new theories and terminology which I have formulated throughout this project. Still, this project does not assume that the reader has any prior knowledge of stochastic analysis or probability theory and so when this is used I have explained it in full.

I have taken an example-based approach to this project, beginning with a basic model within the context of biological processes and slowly adding depth and conditions to my models. This changes the characteristics of my model and how it compares to the Master ODE. By using MATLAB tools such as ODE45 I was able to create more complex systems of equations that interacted with my stochastic models in different ways. This project has made an effort to justify and explore that relationship. After developing the core understanding of a biological system and its stochastic models, we compared the results to its stochastic mean or its ODE solution with a given initial condition. In basic examples, these are often equivalent which I have given proof of. In the later chapters I then go on to discuss how the connection between the two methods of modelling breaks down, and the characteristics of the modelling types begin to become apparent. Adding noise into the system or using systems with multiple steady states leads to the connection becoming weaker, this project looks into the causes of that and the subsequent effect on the model.

By the end of this paper, the reader should be able to understand the mathematical link between Stochastic Simulations and Deterministic Modelling and apply that logic to biological processes such as the decay of a molecule and more complex chemical reactions. This allows the reader to understand the differences between the two modelling types as well as their similarities, hopefully aiding the reader to choose the correct one going forward in their own research. It also explores the vital role of technology within research on a deeper level with the use of programs like MATLAB which allow us to calculate these systems with ease.

Contents

Chapter 1

Motivation & Introductory Examples

Stochastic simulations are a type of mathematical modelling which involve the use of random numbers to study the behaviour of a system. In the context of biological processes, stochastic simulations can be used to study the behaviour of molecules, cells, or even entire organisms. For example, a stochastic simulation might be used to study the movement of molecules in a cell, or the growth and division of cells in a tissue.

One of the key advantages of stochastic simulations is that they can capture the inherent randomness and uncertainty that is present in many biological systems. For example, the movement of a cell follows a probabilistic distribution and a stochastic simulation can capture this probabilistic behaviour more accurately than a deterministic model due to its high degree of volatility.

Deterministic modelling, on the other hand, involves using mathematical equations to describe the behaviour of a system. In the context of biological processes, deterministic models are often used to study the behaviour of large numbers of molecules or cells. For example, a deterministic model might be used to study the population dynamics of a species of animals or the spread of a disease through a population.

Deterministic modelling does allow for the analysis of complex systems in a mathematically rigorous way. By using equations to describe the behaviour of a system, it is possible to make precise predictions about how the system will behave under different conditions. This, in turn, leads to increasingly intense and demanding simulations that are expensive to run in terms of data and the power required. Stochastic simulations can take these same variables and account for them much more efficiently by treating them as random variables.

To introduce the topic to the reader we will look at some motivational examples of the connection between stochastic simulations and deterministic modelling and why one may be favoured over another in a given scenario. The outcome of this first chapter is to lay out the ground-works for this topic in such a way that it introduces and explains the areas that we will go on to talk about in more detail. The key difference in this chapter is that we are doing so without the 'distraction' of biological processes (the context this project is working within) to separate the concept from the maths to aid initial understanding. Thus making room for the consideration of many other applications of stochastic modelling and differential equations and their relationship.

1.1 Introductory Examples

Let us start by looking at an example to bring some motivation. These two examples will give an insight into the logic that is used throughout the paper. First, we must understand that a stochastic model is one in which we cannot know the exact outcome, whereas a deterministic one can be precisely calculated.
What we will see from these examples is that the line between these two seemingly separate fields of modelling is more blurred than one might initially assume. Some deterministic models can appear stochastic, and vice versa, some stochastic simulations can give what looks like calculated results.

1.1.1 Galton Board

The Galton Board is a famous mathematical experiment in which a ball is dropped onto a pyramid of pins and falls into corridors depending on the direction of the ball. The ball has a 50% chance of falling left or right at each pin. This was first used as a way to visualise binomial distribution, as seen below (Fig 1.1). When we run the model $10,000$ times we always obtain the above distribution of results. We are doing something random, but it always produces the same distribution. i.e, a stochastic model can give deterministic-looking results.

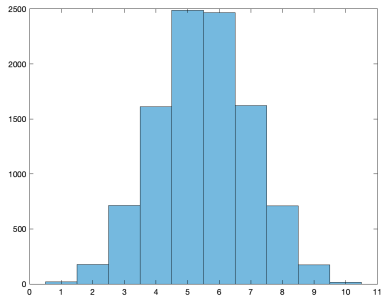


FIGURE 1.1: Distribution of 10,000 marbles on a Galton Board

1.1.2 The Monte-Carlo Method

Another example of a stochastic simulation that gives deterministic-looking results would be the Monte-Carlo method. Say we wanted to measure the price of a company's stock value over time. What we could do is look at the history of the stock price and use that to find a mean value and extrapolate that into the future. However, this is not a particularly accurate method for stocks due to the versatility of the market. Instead what investors can do is use a random sampling of data from previous information and use it to predict trends. An analyst will add many factors into a distribution, such as inflation, asset returns, tax rates, and even lifespans. The result is a distribution of portfolio sizes with the probabilities of it supporting their client's desired needs.

The analyst will then use the Monte-Carlo simulation to determine the expected value and probability that an investment portfolio is acceptable for the client. This method allows one to view many different outcomes given different paths the investment portfolio may take, such as stock volatility or asset allocation at different time periods. The analyst will look at varying degrees of risk, savings and investment strategies to arrive at a distribution of acceptable portfolios along with the probability of arriving at them by the date required desired timeframe. The image below (Fig 1.2) is an example of such a simulation of a portfolio. It is worth noting that if we were to create many more realisations of this stochastic model, we can obtain a stochastic mean of these outcomes which will correspond with a deterministic model that we can plot on the same graph. We will discuss this further in chapter 2. (Technologies (2023))

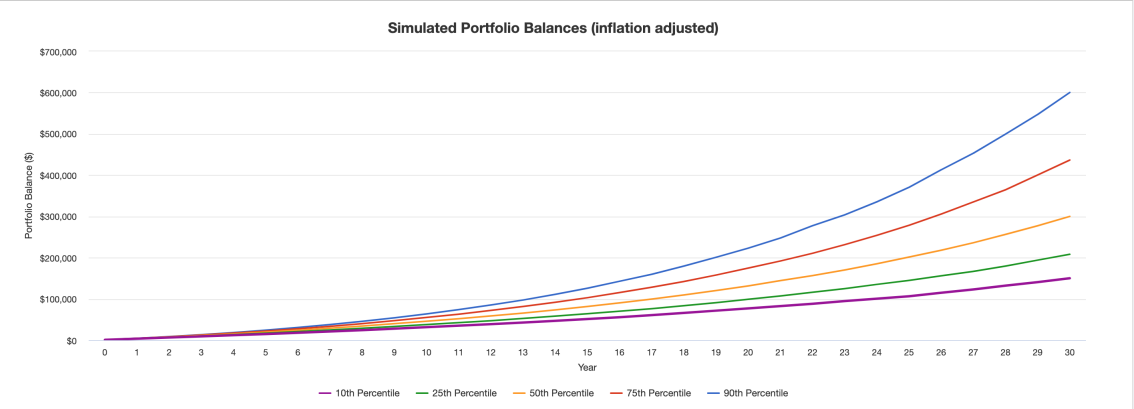


FIGURE 1.2: Five possible outcomes at varying probabilities of occurrence for an example portfolio over 30 years

1.2 Motivation and Wider Scope

This connection is present in many fields of science and technology and is used to accurately model and predict a variety of scenarios.

- Finance - Financial analysts use stochastic simulations to model the behaviour of stock markets, currencies, and other financial instruments. By using random variables to represent factors such as changes in interest rates, market volatility, and consumer behaviour, analysts can study how different scenarios are likely to impact the performance of a portfolio. They will still however model these simulations against a deterministic mean that they can compute by using past data and identify trends that they can project forward into the future. The combination of these two methods will greatly increase their accuracy.
- Engineering - analysts might use a deterministic model to study the behaviour of a bridge. The model might include fixed values and assumptions about the materials used to build the bridge and the loads that it is designed to support. This can be compared against stochastic simulations of weather and population density over time. By using these models to simulate the behaviour of the bridge under different conditions, the analysts can identify potential weaknesses and make recommendations for improvements.

Chapter 2

A Basic Example in the Context of Biological Processes

This chapter is based on the source material (Erban and Chapman (2020)) which has taught me the content necessary to create these examples and discuss these ideas. The initial rudimentary example of this project can be found in this book however the code used in all stochastic models in this project is my own.

2.1 The Deterministic Approach

To introduce the biological context of the project, and see a real working example of stochastic vs. deterministic modelling we can begin with the most simple example of a system. We can monitor molecules leaving the system through decay only, an independent chemical reaction:

$$A \xrightarrow{k} \emptyset. \tag{2.1}$$

Where k is equal to the rate that the molecule leaves the system and $\emptyset$ is equal to any chemical system we are not modelling. In this model, we are choosing to track decay. However, any interference that could cause a molecule to change its state and leave our observable system would be displayed in the same way.

In this model, we will track the number of molecules in our system, A, at a time, t by $a(t)$, and k will be the rate of change of A. This can be represented as an ordinary differential equation. Let us first comprise the deterministic model.

By using differentiation by definition on $a(t)$ we get the equation:

$$\begin{aligned} a(t+dt) &= a(t) - ka(t)dt \\ \frac{a(t+dt) - a(t)}{dt} &= -ka(t) \end{aligned}$$

as $dt \longrightarrow 0$ we get the Ordinary Differential Equation (ODE):

$$\frac{da}{dt} = -ka. \tag{2.2}$$

We can solve this analytically by integrating both sides:

$$\frac{d}{dt}a(t) = -ka(t). \tag{2.3}$$

Which has the form $y' + P(x)y = 0$, ($P(x) = k$) so it can be solved as a linear inhomogeneous

1st Order Differential Equation by means of substitution of variables.

$$\frac{dy}{y} = -P(x)dx \tag{2.4}$$

$$\int \frac{1}{y}dy = -\int P(x)dx \tag{2.5}$$

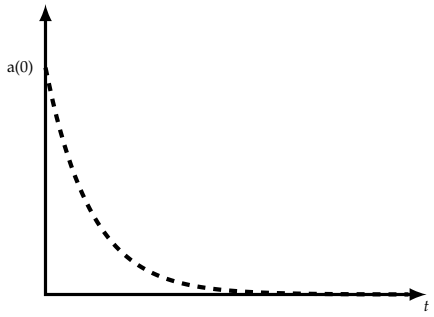
$$\log(|y|) = -\int P(x)dx \tag{2.6}$$

$$|y| = e^{-\int P(x)dx}. \tag{2.7}$$

So substituting back in, we have the solution to the ordinary differential equation:

$$a(t) = a(0)e^{-kt}. \tag{2.8}$$

Which can be plotted to find our expected mean decay rate over time, which you can see in the figure below:



2.2 The Stochastic Approach

We can then predict statistically correct answers for this model by using what is known as a Stochastic Simulations Algorithm (SSA). This algorithm uses probability distribution formulae to create increasingly accurate predictions of a model. One of three things can happen during our decay process:

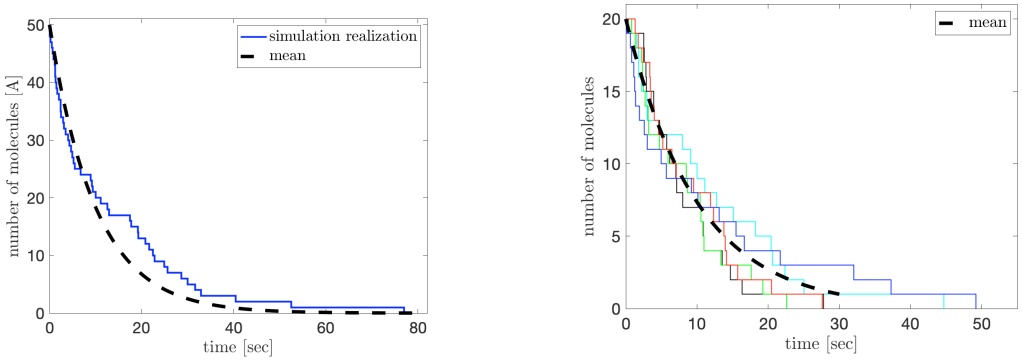
1. Nothing Happens
2. A Single Decay Occurs
3. Multiple Molecules Decay Simultaneously.

For sufficiently small Δt (the smaller, the more accurate our predictions) the probability that more than one reaction happens within a single time step gets so small that it can be safely neglected and so our probability becomes a slightly simpler system. We can do this by defining $k \cdot \Delta t$ to be equal to the probability that a molecule decays.
So our formulae are:

- Probability that decay occurs: $A(t)kdt$
- Probability that decay does not occur: $1 - A(t)kdt$.

In order to execute our algorithm we would need to do thousands of calculations that are relatively complex and time-consuming. Because of this, we will be making great use of the program MATLAB to do these computations for us, and plot them nicely to obtain a visualisation of the SSA against the ODE mean. You will be able to find all of the code that has led to these models and figures in the glossary.

We will use our probability formula above (2.8) but assign the generalised variables some values for the purpose of this example. Below are some sections of the MATLAB code created to produce the diagrams (Fig 2.1).



(A) (B)

FIGURE 2.1: *An 'expensive', un-optimised SSA running 1 & 5 realisations of the simulation against the mean*

LISTING 2.1: Example 1

```
1  rand('state', 10)
2  k1=0.1;
3  Ainitial=50;
4  deltat=0.001;
5  noruns=1;
6     for i=noruns
7          A=Ainitial;
8          k=1;
9          time=0;
10         Aplot(1,i)=A;
11         timeplot(1,i)=0;
12            while (k<Ainitial+1)
13                 random=rand(1,1);
14                 time = time+deltat;
15
16                 if (random<k1*A*deltat)
17                      A=A-1;
18                      k=k+1;
19                      Aplot(k,i)=A;
20                      timeplot(k,i)=time;
21                 end
22            end
23     Aplot(k+1,i)=A;
24     timeplot(k+1,1)=time;
25     end
```

We begin by defining several variables that are used in the simulation, such as the rate of decay (k1), the initial number of molecules (Ainitial), and the time step (deltat).

Next, we set up a for loop that will run the simulation 'noruns' times. Inside the for loop, we use a while loop to simulate the decay of the molecules over time. At each step of the simulation, we generate a random number and compare it to the probability of decay. If the random number is less than the probability, we remove a molecule from the system and update the results of the simulation.

After the while loop has finished, we save the simulation results in the arrays (Aplot), (timeplot) which we then use to plot the images (Fig 2.1) above.

2.3 Refining Our SSA

We will now go on to investigate something specific to this example which can introduce insight into developing SSAs in a general sense, something which is useful to keep in mind as SSA's become more complex later on in the project. In the figures above, a stochastic simulation looks like a rough estimate of the more direct deterministic method but this isn't always the case. We can calculate 'stochastic fluctuations' about the mean. This will show us how much each realisation of the SSA can vary from the mean line you can see in (Fig 2.1 (b)). We can try to create a more accurate model to reduce this fluctuation from the mean. One way that we can increase the accuracy would be to decrease our time-step Δt in order to more accurately locate the time in which a molecule has decayed. However, in doing this we vastly increase our computational demand and the simulation becomes more and more expensive and quickly inefficient.

What we can notice is that for $A(0) = 50$ in our example we have checked around $80,000$ time-steps, meaning usually when checking a timestep we have run the calculation only to find that nothing has happened, this is wasting precious computational power and is quite a slow method. It seems we will require a new approach. We can shift from focusing on progressing through time and instead progress in decay. We can do this by calculating the exact time, $t + \tau$ at which the next decay will occur, instead of jumping forward in time and then looking back to see if one has happened. We can then move the simulation forward τ seconds to arrive at the next change in our system.

As we can find on [pg.4] of the source material (Erban and Chapman (2020)), $f(A(t), s)$ ds is the probability that a given molecule decays during the time interval $[t + s, t + s +$ ds) where ds is an infinitesimal time-step. In order for this to be true then it must be the case that there was no decay during the interval before that, $[t, t + s)$. We can then denote $g(A(t), s)$ to be the probability that a reaction does not occur during $[t, t + s]$, and then we can see that the probability $f(A(t), s)$ ds can be computed as the product $g(A(t), s) \cdot A(t + s)k$ ds. Further to this, since we have stated in this short proof that no reaction has occurred during the time interval $[t, t + s]$ then $A(t + s) = A(t)$ so:

$$f(A(t), s) \text{ ds} = g(A(t), s) \cdot A(t)k \text{ ds}. \tag{2.9}$$

We can then create a separate time interval, $[t, t + \sigma + d\sigma]$, $\sigma > 0$ to glove $g(A(t), s)$. Using very similar logic as above we can find that the probability that no reaction occurs during this time period is the product of the probability that no reaction occurs in the interval $[t, t + \sigma]$ and the probability that no reaction occurs in the interval $[t + \sigma, t + \sigma + d\sigma]$. And so

$$g(A(t), \sigma + d\sigma) = g(A(t, \sigma)[1 - A(t + \sigma)kd\sigma.$$

And again, for all of this to happen in the first place we must have had no decay occurring during $[t, t + \sigma]$ and so $A(t + \sigma) = A(t)$ which allows us to simplify the above similarly to (2.9)

$$g(A(t), \sigma + d\sigma) = g(A(t, \sigma)[1 - A(t)kd\sigma]$$

$$\frac{g(A(t), \sigma + d\sigma) - g(A(t), \sigma)}{d\sigma} = -A(t)k \cdot g(A(t, \sigma). \tag{2.10}$$

This (2.10) is a finite difference approximation, also known as differentiation from first principles and because of this, as $d\sigma$ approaches 0 the difference term $((g(A(t), \sigma + d\sigma) - g(A(t), \sigma)$ become infinitely small and as such we are left with a partial differential equation that we can solve, using a function,g, a constant, k and a variable, σ using the separation of variables method.

$$\frac{dg}{d\sigma} = -A(t)k \cdot g(A(t), \sigma).$$

The function $g(A(t), \sigma)$ can be written as a product of two functions:

$$g(A(t), \sigma) = f(A(t)) \cdot h(\sigma)$$

.

Substituting this into the PDE and dividing both sides by $f(A(t))$:

$$\frac{dh}{h(\sigma)} = -A(t)k \cdot d\sigma$$

.

Integrating both sides with respect to σ, we obtain:

$$\ln|h(\sigma)| = -A(t)k \cdot \sigma + C$$

.

Where C is the constant of integration. Solving for $h(\sigma)$, we get:

$$h(\sigma) = Ce^{-A(t)k\cdot\sigma}$$

.

Using the initial condition $g(A(t), 0) = 1$, we get:

$$h(0) = Ce^{-A(t)k\cdot 0} = 1$$

.

Therefore, $C = 1$, and $h(\sigma) = e^{-A(t)k\cdot\sigma}$.
Substituting this back into the expression for $g(A(t), \sigma)$, we obtain:

$$g(A(t), \sigma) = f(A(t)) \cdot e^{-A(t)k\cdot\sigma}$$

Where $f(A(t)) = 1$ from the initial condition $g(A(t), 0) = 1$,
Thus, the solution is:

$$g(A(t), \sigma) = e^{-A(t)k\cdot\sigma}$$

.

We can then rewrite (2.9) as

$$f(A(t), s)\ \text{ds} = A(t)k \cdot e^{-A(t)ks}\ \text{ds}. \tag{2.11}$$

This function is in the form of $\lambda e^{-\lambda x}$ which is the Probability Mass Function for the exponential distribution so we now know that the time interval $[t, t+\tau]$ is distributed by the exponential distribution. We can use this in our stochastic simulations to help us generate random numbers distributed by (2.15).

We can create an auxiliary function as found in the source material for this example (Erban and Chapman (2020) (a fit-for-purpose, placeholder function):

$$F(\tau) = e^{-A(t)k\tau} = \int_{\tau}^{\infty} f(A(t), s)\ \text{ds} \tag{2.12}$$

Where $f(\tau)$ represents the probability that the time to the next reaction is greater than τ, and is monotone decreasing for $A(t) > 0$. With the use of (2.11) and (2.12) we obtain

$$\begin{aligned}\int_{F^{-1}(b)}^{F^{-1}(a)} f(A(t), s)ds &= \int_{F^{-1}(b)}^{F^{-1}(a)} A(t)k \cdot e^{-A(t)ks}ds \\ &= -\int_{F^{-1}(b)}^{F^{-1}(a)} \frac{dF}{ds}ds = -F[F^{-1}(a)] + F[F^{-1}(b)] = b - a.\end{aligned}$$

Given an algorithm that generates a random number r uniformly over $(0, 1)$, we can generate the time of the next reaction by setting

$$r = F(\tau) = e^{-A(t)k\tau}.$$

Solve for τ to obtain

$$\tau = \frac{1}{A(t)k} ln \left[\frac{1}{r}\right]. \tag{2.13}$$

This is an important realisation in the refinement of our models and while plotting another model using this new SSA will not necessarily show an obviously "more accurate" set of realisations against our deterministic mean line, the usefulness of finding fresh perspectives to create more accurate findings should not be overlooked. To close this example, let us work backwards from this discovery, τ, and use it to find the mean of $A(t)$ over infinitely many realisations, which also will lead us to the ODE in our example, $\frac{da}{dt} = -ka$.

If $p_n(t)$ is the probability that there are n molecules of A at time t, then $A(t) = n$. For an infinitesimal time interval $[t, t + dt]$ there are exactly two ways for $A(t + dt) = n$ to happen. Either, $A(t) = n$ so no reaction has occurred, or $A(t) = n + 1$ and one molecule was degraded. This can be written as

$$p_n(t + dt) = p_n(t) \cdot (1 - kndt) + p_{n+1}(t) \cdot k(n + 1)dt$$

Which we can rearrange to create the differential equation:

$$\frac{p_n(t + dt) - p_n(t)}{dt} = k(n + 1)p_{n+1}(t) - = knp_n(t)$$

Passing to the limit $dt \longrightarrow 0$, we find the 'chemical master equation' for our system

$$\frac{dp_n}{dt} = k(n + 1)p_{n+1} - knp_n \tag{2.14}$$

This, at first glance, is an infinite system of ODEs for $p_n, n = 0, 1, 2, 3, ...)$, but with $A(0) = n_0$ we know that there are never more than n_0 molecules in the system. This means that $p_n = 0$ for $n > n_0$ and so the system shrinks down to $n_0 + 1$ ODEs for $p_0, p_1, ..., p_{n_0}$ and at $n = n_0$ we have:

$$\frac{dp_{n_0}}{dt} = -kn_0 p_{n_0}$$

Solving this with the initial condition $p_{n_0}(0) = 1$, since there are n_0 molecules of A at time $t = 0$, we obtain:

$$p_{n_0}(t) = e^{-kn_0 t}$$

Which we can substitute into (2.14) for $p_{n_0-1}(t)$ to obtain:

$$\frac{d}{dt} p_{n_0-1}(t) = kn_0 \cdot e^{-kn_0 t} - k(n_0 - 1)p_{n_0-1}(t)$$

Which we can solve with initial condition $p_{n_0-1}(0) = 0$:

$$p_{n_0-1}(t) = e^{-k(n_0-1)t} \cdot n_0(1 - e^{-kt})$$

This means that by induction, the general case is equal to:

$$p_n(t) = e^{-knt} \cdot \frac{n_0!}{n!(n_0 - n)!} \cdot (1 - e^{-kt})^{n_0-n} \tag{2.15}$$

This equation (2.15) is equal to the probability mass function of the binomial distribution with parameter e^{-kt} which allows us to speak more accurately about the system (2.1) with initial condition $A(0) = n_0$. For example, we can now say $A(t) = n_0$ with probability $p_n(t)$ rather than just $A(t) = n_0$ which is a more complete look at the model what we see is that $n_0 \cdot e^{-kt}$ is equal to the solution of (2.2). We can use this new information to derive the mean value of $A(t)$ over infinitely many realisations.

We can say that the mean over time is equal to the sum of the number of molecules multiplied by the probability of them occurring from $n = 0$ to $n = n_0$, which can be written

as:

$$M(t) = \sum_{n=0}^{n_0} np_n(t)$$

Using (2.19),

$$M(t) = \sum_{n=0}^{n_0} np_n(t) \tag{2.16}$$

$$= \sum_{n=0}^{n_0} n \cdot e^{-knt} \cdot \frac{n_0!}{n!(n_0-n)!} \cdot (1-e^{-kt})^{n_0-n} \tag{2.17}$$

$$= n_0 \cdot e^{-kt} \cdot \sum_{n=0}^{n_0} \cdot \frac{(n_0-1)!}{(n-1)!((n_0-1)-(n-1))!} \cdot (1-e^{-kt})^{(n_0-1)-(n-1)} \cdot (e^{-kt})^{n-1} \tag{2.18}$$

$$= n_0 \cdot e^{-kt} \tag{2.19}$$

The cancellation between (2.18) and (2.19) is due to the binomial theorem being applied to the identity, ie, $(1-e^{-kt}) + (e^{-kt})^{n_0-1} = 1$.

As we can conclude from above, what we have in (2.19) is equal to the solution (2.8) with initial condition $n_0 = A(t)$. In other words, the stochastic mean could in this example have been solved by solving the ODE but this is not true for most systems of reactions. With this stochastic mean (2.18) we can plot the line against stochastic realisations to calculate and analyse the fluctuations around it, however, in this example, this will look identical to Fig(2.1) due to the mean being equal to the deterministic ODE we were comparing to earlier in the example.

Chapter 3

Adding Complexity

3.1 A Model With Two Reactions

In this section we will begin to look at how adding complexity to a model in different ways can affect how stochastic and deterministic models change their behaviour. We often need to shift perspective in order to deal with the new layers of complexity or find new ways of tackling a problem.

Propose we have a model with the chemical reactions below, occurring in a container of volume v:

$$A \xrightarrow{k_1} \emptyset \qquad \emptyset \xrightarrow{k_2} A. \tag{3.1}$$

Where the first reaction describes (2.1), and the second reaction describes the production of A with a rate of k_2 from outside of our observable space, denoted by $\emptyset$. This means that we now need to consider that a molecule, A now has a chance to be created at a probability of $k_2 v dt$ where v is the volume of the system. This is done to take into account the size of the system, and that if we were to half the physical system by dividing the container the system was taking place in by 2, the production rate would also halve. This is important because the total space available is a real-world factor to consider in our models.

In order to create the ODE from our chemical reactions (3.1) we can make use of a new theorem.

Definition 3.1.1 (The Law of Mass Action)**.** The Law of Mass Action states that the production rate of a reaction is directly proportional to the product of the input population sizes. Specifically, if

$$ay \xrightarrow{r} by. \tag{3.2}$$

is the reaction of interest then the production rate is proportional to

$$ry^a \tag{3.3}$$

And the accompanying ODE is

$$\frac{dy}{dt} = (b-a)ry^a. \tag{3.4}$$

The above definition has been taken from Dr. Noemi Picco's lecture notes (Picco (2023)) [Definition 5, pg 18] on MA-182 Biomathematics. From Def(3.1.1) we can now formulate the ODE for these chemical reactions to be

$$\frac{da}{dt} = -k_1 a + k_2. \tag{3.5}$$

Which we can solve using the same method at (2.3)

$$\frac{da}{dt} = -k_1 a + k_2.$$

Separating the variables a and t, we get:

$$\frac{1}{-k_1 a + k_2} da = dt.$$

Integrating both sides, we get:

$$-\frac{1}{k_1} \ln(-k_1 a + k_2) = t + C.$$

where C is an arbitrary constant of integration. Solving for a, we get:

$$-k_1 a + k_2 = e^{-k_1(t+C)}.$$

Simplifying, we get:

$$a(t) = \frac{k_2}{k_1} - Ce^{-k_1 t}.$$

Where $C = (\frac{k_2}{k_1} - a(0))$ is the constant of integration determined by the initial condition $a(0) = a_0$. Then, substituting the constant of integration,

$$a(t) = \frac{k_2}{k_1} - (\frac{k_2}{k_1} - a_0)e^{-k_1 t}.$$

Which can be plotted to show the values of this ODE over time as seen below

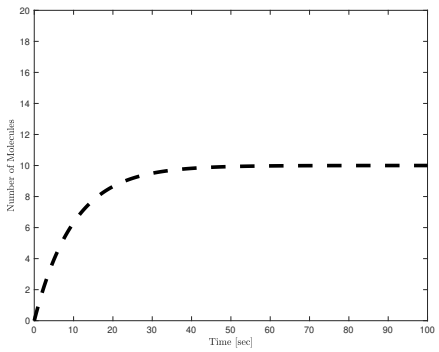


FIGURE 3.1: *A plot of our ODE (3.5) over time*

Now that we have the deterministic half of our comparison, let us build the stochastic system to complement it. Once again we will be creating an algorithm in MATLAB, this time we want it to do the following things:

1. Compute two random numbers r_1, r_2 which are normally distributed in $(0, 1)$
2. Compute the probability $\alpha_0 = A(t)k_1 + k_2 v$
3. Compute the exact time $(t + \tau)$ in which a future reaction occurs as in (2.13)
4. Find $A(t + \tau)$

The code for this can be found in the appendix, as this follows a very similar structure to the previous examples, with some slight changes to the probability function α_0 but can be seen below (Fig 3.2).

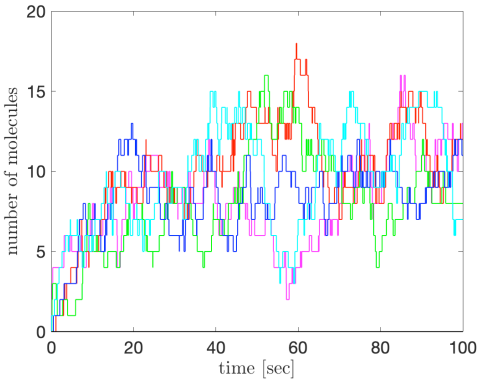


FIGURE 3.2: *A model showing 5 realisations of (3.1) with* $A(0) = 0$, $k_1 = 0.1$, $k_2 = 1$,

This image is quite useful, as it shows the drastic differences that can occur between realisations and also visualises the consequences of our more complicated chemical reaction. $A(t)$ is fluctuating between 0 & 18 however, for the best idea of what is happening, let us overlay these plots as we have done before below (Fig 3.3)

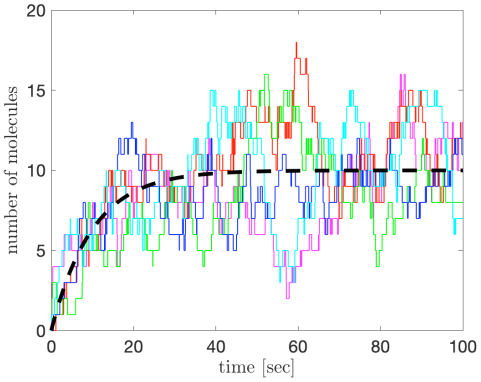


FIGURE 3.3: *(Fig 3.1) & (Fig 3.2) Overlayed*

We can clearly see in (Fig 3.3) that the stochastic simulation follows the ODE line fairly closely with some spiking exceptions. What we could say about this is that the ODE describes the general behaviour of the model over time, whereas the SSA simulates the stochastic fluctuations around the average behaviour.

3.2 Analysing The System

Let us have a closer look at how we can further analyse the behaviour of this model, and what this relationship can tell us about both types of modelling. There are lots of ways to look at a differential equation analytically so let us look at a few of them to gain more insight into what is happening.

- Comparison Of Results

We can see from the plots that the ODE gives us a smooth curve, whereas the stochastic model gives us fluctuations. These fluctuations get more dramatic the later into the simulation we get. This is because as the number of molecules decreases, the probability that a molecule decays also decreases. Another affecting factor is that the deterministic solution assumes that the number of molecules of A changes continuously over time, whereas the stochastic trajectories show discrete changes due to the random occurrence of individual reactions. Our model only had a small number of molecules in the system at one time which gives

inherent randomness the chance to cause large spikes in the model, if there was a massive number of molecules $A(t) > 10000$ then these fluctuations wouldn't stand out so much and the behaviour would more closely follow the predicted behaviour of the ODE 'mean line.'

- Sensitivity Test

A sensitivity analysis can be used to study how changing the parameters of a model can affect its behaviour. In our case, we can change k_1 and k_2 to see what will happen given different constant rates of change.

What you will see is if you fix k_2 and alter k_1, you reduce the value of the equilibrium point that the ODE finds very quickly. You only need to increase k_1 by 0.1 to half the steady state. This is due to us dividing by k_1 in the model and because $k_2 = 1$, $\frac{1}{0.1} = 10$, $\frac{1}{0.1} = 5$. Conversely, the rate at which k_2 changes the model is approximately half, but as it increases it has the opposite effect in which it raises the equilibrium point of the model as it's the numerator of our fraction. This is discussing only non-negative values for k_1, k_2 as they are rates of change, which are fundamentally non-negative, a positive rate of decay is an oxymoron.

These two forms of analysis have helped explain why the model behaves as it does, especially in the context of decay and production. A lot more can be learnt by using these analysis styles from more complex chemical reactions but it seems there is a limit to using the Law of Mass action to create ODEs and solve our models in this way. Gillespie himself (Gillespie (1977)) said that the use of the Law Of Mass Action in this context of biology is fundamentally flawed. If you have chemical species, A & B that react to create a molecule, C and a concentration for these species, a, b & c respectively then as they travel through a time interval $[t, t + dt]$ they will trace a collision volume. The number of reactions that take place (no. of C created) will be proportional to the number of molecules A & B that are within this volume at some constant rate k_1. the change in c can be therefore written as $k_1 \cdot a \cdot b\delta t$. Letting δt tend towards zero, we can form a differential equation as we have in each of our examples thus far. However as Gillespie pointed out as δt goes towards zero, the number of A & B molecules in the (small) collision volume, being either zero or one, is no longer proportional to concentration a, b. This is only ever really accurate when the number of molecules involved is sufficiently large that "averaging" the number of molecules A, B in the collision volume over many molecules i.e. when $A(t)$, $B(t)$ are huge samples.

3.3 Gillespie Algorithm

Nonetheless, it is important to zoom out slightly and understand that the Stochastic Simulation Algorithms we have been using (2.3) so far fall under the blanket term of a Gillespie Algorithm. Because of its importance in this field, we can briefly have a look at the formulation of this algorithm (thankfully, a lot of the work has been done in previous sections of this project.)

This section has been adapted from the source material (Erban and Chapman (2020)) & D. Gillespie's work (Gillespie (1977)).

Consider a system of n chemical reactions. Let $\alpha_i(t)$ be the function denoting the likelihood (propensity function) of the ith reaction occurring, $i = 1, 2..., n$, at time t, so $a_i(t)$ dt is the probability that the ith reaction occurs during the time interval $[t, t + dt)$. In such a system, a Gillespie SSA can be used which consists of the following steps:

1. Generate two random numbers r_1, r_2 which are normally distributed between $(0, 1)$
2. Compute the propensity function $\alpha_0 = \sum_{i=1}^{n} \alpha_i(t)$
3. Compute τ, defined in (2.13) to be $\tau = \frac{1}{\alpha_0} \cdot ln[\frac{1}{r_1}]$
4. Compute which reaction occurs at time $[t + \tau]$.
   Find j such that $r_2 \geq \frac{1}{\alpha_0} \sum_{i=1}^{j-1} \alpha_i(t)$ and $r_2 < \frac{1}{\alpha_0} \sum_{i=1}^{j} \alpha_i(t)$

This means the jth reaction has taken place, so we can change the number of reactants and products for the jth reaction and then loop back to step 1 for time $t + \tau$. The limits of this can be when n number of reactions has been met, or the number of molecules hits zero. It could also be some threshold that Gillespie shows examples of in (Gillespie (1977)).

Chapter 4

Exploring The Differences In Stochastic and Deterministic Modelling

As we start to explore chemical reactions that aren't quite as straightforward, we start to see discrepancies in the information given by the stochastic and deterministic methods. In this final chapter, we are going to look at a few examples where some key characteristics of the model are missed entirely by the ODE due to the noise present in the SSA. These systems are not uncommon and are not isolated cases; noise, in the form of oscillations over time or stochastic resonance, is often overlooked by the accompanying ODE derived from the same reaction system.

Noise is defined as the random fluctuations or variations in a stochastic process that are not predictable or controllable. These fluctuations can be modelled using random variables or random functions that are added to the deterministic components of the process. In some cases, the noise may be modelled as a normal distribution with a mean of zero and a specified variance or standard deviation as we have done in our previous examples. In other cases, more complex probability distributions or stochastic processes may be used to model the noise.

Noise can be disruptive to a process, however, it is a much more realistic way of modelling a reaction-diffusion process. In fact, it can be so disruptive it can cause a stochastic process to be pulled from its deterministically predicted steady state and find a new one that was not previously expected.

4.1 Systems with Multiple Steady States

When a system has more than one steady state we find that the behaviours of the two types of modelling shift over a long enough period of time. Let us consider the example in (Erban and Chapman (2020)) with the following four chemical reactions:

$$3A \underset{k_2}{\overset{k_1}{\rightleftarrows}} 2A \qquad A \underset{k_4}{\overset{k_3}{\rightleftarrows}} \emptyset. \tag{4.1}$$

Which has the deterministic ODE model which can be derived from The Law of Mass Action (Def 3.1.1)

$$\frac{da}{dt} = -k_1 a^3 + k_2 a^2 - k_3 a + k_4. \tag{4.2}$$

We can rewrite this in terms of $a(t)v$ commonly denoted $\overline{A}(t)$ which is the average number of molecules of A in the volume v with concentration $a(t)$:

$$\frac{d\overline{A}}{dt} = -\frac{k_1}{v^2}\overline{A}^3 + \frac{k_2}{v}\overline{A}^2 - k_3\overline{A} + k_4 v. \tag{4.3}$$

Given the constants $k_1 = 2.5 \times 10^{-4}, k_2 = 0.18, k_3 = 37.5, k_4 = 2200$ As picked by Erban & Chapman to create this illustrative example (Erban and Chapman (2020)). You can solve

this equation at $\frac{d\overline{A}}{dt} = 0$ to find the steady states of the system, of which we find three.

$$\overline{A}_{(s,1)} = 100 \qquad \overline{A}_{(s,2)} = 220 \qquad \overline{A}_{(s,3)} = 400. \tag{4.4}$$

where $\overline{A}_{(s,1)}$ and $\overline{A}_{(s,3)}$ are stable and $\overline{A}_{(s,2)}$ is unstable. To show this, we find the derivative of the function at each fixed point and evaluate the outcome. If positive, we know the fixed point is unstable and if negative it must be stable which we know from teachings throughout our undergraduate degree.

$$f(x)' = -0.00075x^2 + 0.36x - 37.5$$

$$f(100)' = -9, \quad f(220)' = \frac{27}{5}, \quad f(400)' = -\frac{27}{2}.$$

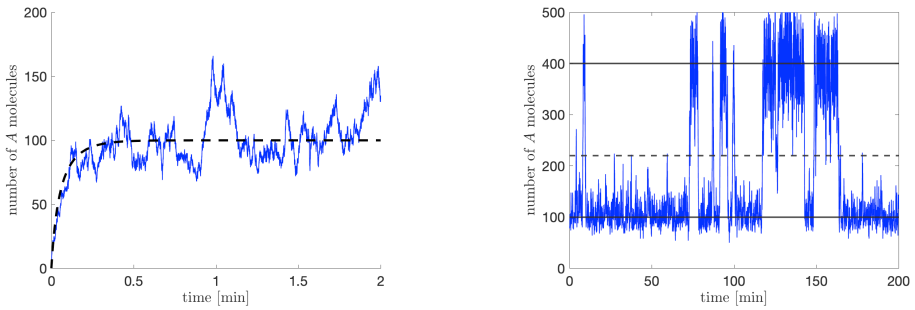
Since $f(100)'$ and $f(400)'$ are both negative we know that both $\overline{A}_{(s,1)}$ and $\overline{A}_{(s,3)}$ are stable fixed points of this function. The solution to ODE (4.2) converges to one of the steady states dependent on the initial condition inputted.

$$\text{If, } \ \overline{A}(0) \in [0, \overline{A}_{(s,2)}) \ \text{ (4.2) Converges to } \ \overline{A}_{(s,1)}.$$

$$\text{If, } \ \overline{A}(0) > \overline{A}_{(s,2)} \ \text{ (4.2) Converges to } \ \overline{A}_{(s,3)}.$$

If we continue this experiment as we have with Chapter 2, our next step would be to set $A(0) = 0$ and begin with no molecules in the system. However, as we have found above, this will inevitably lead to the convergence of $\overline{A}_{(s,1)}$, where plotting the accompanying rendition of the Gillespie algorithm will look very similar to our examples thus far (Fig 4.1a).

The differences emerge once we allow the simulation to run over a much longer period. This gives more opportunity for noise to affect the simulation. We can see in the diagrams (Fig 4.1b) that given sufficient time, the random fluctuations can pull the system towards other steady states. The stability of the steady states is important here too, notice how the system spends most of its time at its ODE-preferred steady state, and then after that, it spends time at its secondary stable steady state. The unstable steady state $\overline{A}_{(s,2)}$ is visited for the least amount of time and always results in the model trending back toward one of the other, stable, steady states. When we compare this to the ODE plot (Fig 4.1a) we see that no such fluctuations are picked up. The function finds a steady state, trends towards it and stays there. This is the true behaviour of our two methods of modelling shining through as the deterministic ODE is showing the average behaviour of the system over a large time period. Given an infinite timescale, the ODE would show a very accurate depiction of the model and its behaviour but on smaller scales, the stochastic modelling is able to pick up these fluctuations.



(A) (B)

FIGURE 4.1: *A short, 2minute simulation vs. a 200minute simulation showing the system fluctuating to different steady states due to noise*

4.2 Stochastic Focusing

(Paulsson et al. (2000)) discuss something known as Stochastic Focusing, a phenomenon which visualises the fluctuations of a system between steady states and shows how fluctuations in one chemical species can drive the entire system to a different steady state than the one predicted by the deterministic model. Consider the system:

$$\emptyset \xrightarrow{k_1} A \xrightarrow{k_2} B \xrightarrow{k_3} \emptyset. \tag{4.5}$$

$$A + C \xrightarrow{k_4} C.$$

$$\emptyset \underset{k_6}{\overset{k_5}{\rightleftarrows}} C. \tag{4.6}$$

Where a product, B is created from A when in the presence of a catalyst, C. What's happening above is A is created, from a source we are not observing in the model, with a rate k_1 and B is produced from A with a rate k_2 and degrades to outside of our system at a rate k_3. This reaction then has the added complexity of the presence of species C which is a reaction that causes the degradation of A with rate k_4 which is in turn affected by the production rate of C, k_5 and its decay rate k_6. By using the Law of Mass Action (Def 3.1.1) we can find the system of ODEs as follows:

$$\frac{dA}{dt} = k_1 - k_2 A - k_4 CA \tag{4.7}$$

$$\frac{dB}{dt} = k_2 A - k_3 B \tag{4.8}$$

$$\frac{dC}{dt} = k_5 - k_6 C. \tag{4.9}$$

We can find this system's equilibrium points as before by letting derivatives in (4.7)-(4.9) equal zero and solve for A, B and C.

The number of molecules,A, B & C, are produced depending on rates k_1 to k_6 as follows:

$$A = \frac{k_1 k_6}{k_2 k_6 + k_4 k_5} \qquad B = k_6 \frac{k_1 k_2}{k_3 (k_2 k_6 + k_4 k_5)} \qquad C = \frac{k_5}{k_6}.$$

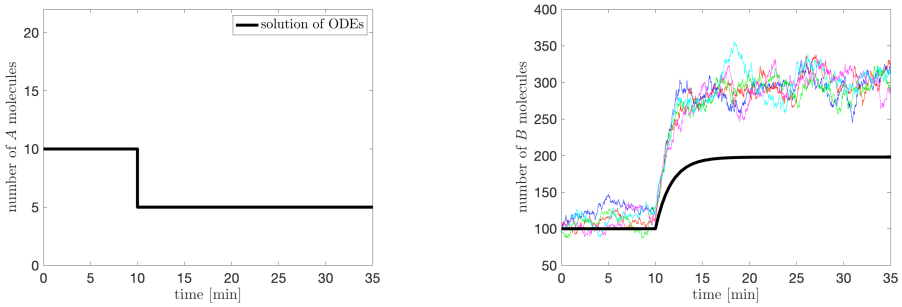
If the value of the rate constants were changed, we would expect the number of B molecules to find a steady state according to the above equation. Let us suppose that the decay of molecule C (k_5) halves at some point throughout the simulation. Now when $k_4 k_5 \gg k_2 k_6$ we would have

$$B \approx k_6 \frac{k_1 k_2}{k_3 k_4 k_5}.$$

i.e the amount of B molecules in the system will, at most, double. This can be seen (Fig 4.2b) by the black line, a solution to the ODE with initial conditions $(10, 100, 0)$ for $A(0)$, $B(0)$ & $C(0)$ respectively, with rate constants $k_1 = 100s^{-1}, k_2 = 1000s^{-1}, k_3 = 1s^{-1}$, & $k_4 = 9900s^{-1}, k_6 = 100$

Where k_5 is a varying rate constant which halves at time 10 minutes.

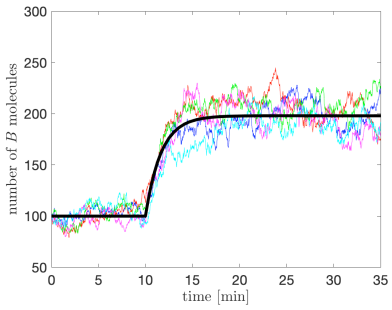
$$k_5 = \begin{cases} 10^3 sec^{-1}, & \text{for} 5 < 10min, \\ 5 \times 10^2 sec^{-1}, & \text{for } t \geq 10min. \end{cases}$$



(A) (B)

FIGURE 4.2: *(A) The rate of production halving from $k_5 = 10000$ to $k_5 = 5000$. (B) The number of molecules produced by the ODE increases significantly less than simulated by five realisations of Gillespie's Algorithm.*

What we see from the simulation of this system over time is that when simulating the six reactions the number of A molecules fluctuates around the initial value given. Halving k_5 we see that the value of A also halves as expected. However, the number of B molecules produced more than doubles, which breaches the expected limit set by the deterministic system. In order to see that this increase in B is in fact due to stochastic fluctuations of A we can hold A at specific values. When we simulate only the first four reactions in our original system we see that the simulation in fact matches the deterministic model's behaviour. (Fig 4.3)



(A)

FIGURE 4.3: *five realisations of a stochastic simulation using only constants k_1 to k_4*

What we are able to determine from this is that when A does not fluctuate, the system is essentially a linear system of two ODEs B & C for a constant, A. This has now become a system similar to our previous examples, and so the ODE becomes the mean over a long period of time, known at the time evolution, (Erban et al. (2007)) of the stochastic simulation.

For the rates given in this example, we can calculate with the code used to create the plots for this example the average number of B molecules created. After 15s an average of 264 (standard deviation 17.3) has been created, whereas a solution of the ODEs (4.7)-(4.9) shows that there should be 198 molecules of B. If we only use the non-fluctuating ODEs for our SSA (4.7) & (4.8), with 100 realisations, then we see that in fact 198 is the new mean number of B molecules, lining up with the deterministic process. This allows us to conclude that given a chemical species' innate random fluctuation about a mean value, a system can be altered to a new steady state. The stochastic method then is much better at modelling a more sensitive, real-world system with random kinetic properties like that of cells moving around in a petri dish (over a short period of time.)

4.3 Developing This Idea

At the conclusion of this project, we can discuss how one might expand on this idea of introducing noise into a system to analyse the differences between stochastic simulations and deterministic modelling. One avenue that could be explored is a higher-dimensional experiment. Even though all examples shown previously are 2D for the sake of simplicity and understanding, we have already mentioned the skill that stochastic simulations have at accounting for nuance that would only be present in real-world systems. Therefore it would be a logical progression to consider what would happen in the 3D world.

One example of this is the Belousov-Zhabotinsky Reaction (Zhabotinsky (1991)) in which they found that in an experiment of mixing chemicals, you can visualise fluctuations of concentrations of the reactants and catalysts which then lead to spontaneous changes in the behaviour of the system. They found oscillations between steady states in the form of patterns in the reaction that are not predicted by the deterministic model.

Another idea to develop this project further could be to look at stochastic resonance. This is the idea that even a small amount of noise added to a system can shift it to states that are different to its deterministic counterpart. Random fluctuations of molecules known as the signal chemical push a system into what's known as a 'limit cycle' which only occurs with the noise in the system present. Without this, it reaches an equilibrium that lines up perfectly with its system of ODEs (Erban and Chapman (2020) & Muratov et al. (2005)).

Chapter 5

Conclusion & Final Reflection

5.1 Conclusion

At the conclusion of this project, we have effectively connected Stochastic Simulations and Deterministic Modelling. Through the use of MATLAB, we have laid deterministic ODEs over several realisations of stochastic simulation to compare their outputs over time. We've seen how we can derive ODEs from chemical reaction equations. For simple examples, these ODEs act as time-evolution models for stochastic simulations. We've also seen, however, that the law of mass action is flawed and when we introduce more complex examples the ODE begins to incorrectly project this average. Another avenue would need to be explored when there are variable rates of k. One area we could explore to resolve this is stochastic differential equations. This is an area of maths which can help to solve many of the difficulties faced. It combines both areas of mathematics; the deterministic average over time and the stochastic fluctuations visible on small data sets over a short period.

5.2 Final Reflection

The creation of this project has been the largest piece of writing in my life to date. Throughout this process, I have learnt a lot both about my own skill set as a mathematician and as a writer. Citing and referencing scientific journals was something new to me and I enjoyed looking for relevant sources outside of what my supervisor Dr Noemi Picco gave me. I also taught myself some areas of maths that I did not know before starting this project, such as The Law of Mass Action and MATLAB, so that I could explain myself confidently. To do this I read papers on the topics as well as used lecture notes from modules that I didn't take to teach myself the necessary content, especially MA-182 Biomathematics. I developed my confidence in solving differential equations throughout this paper using a range of methods, both old and new, which was a great refresher of my second year. Mathematics wasn't the only thing being challenged and refreshed: the LaTeXrequired in this project was another thing that I needed to learn more about. I developed my knowledge on this while writing, and often found myself going back to improve sections of my dissertation, for example knowing how I could have made a diagram clearer or more presentable.

However, the project was not without its struggles. I have countless corrupted or unsuitable programs in my MATLAB folder, which was the largest obstacle I faced in this piece of work. I had never created a stochastic simulation in MATLAB before this project, so being able to write one independently from scratch is something I'm proud of. The first month or so of researching and writing was purely trial and error in attempts at coding. I would constantly be researching code, testing code I had found, or trying to decipher code supplied by my reference material- it was a real challenge. Eventually, I started to understand what my project was going to be (it changed a lot throughout my writing) and that helped me to know what simulations I would need, their purpose and what they were trying to show the reader.

Another area I struggled with was understanding which direction to take the project. I wrote the first two chapters early on in the year and then took a break to focus on other studies and to research more. When I returned, the direction and style I wanted for my dissertation changed slightly. I found it difficult to shift the tone of the paper without it ruining the flow, so I ended up rewriting a lot of sections to make it more readable. This ate

into my schedule, and I drifted from my project plan and the time goals I had set for each chapter/section.

Overall, however, I am proud to say I completed the project punctually and gave myself two weeks to proofread my dissertation and make it the best it could be. I made a conscious effort to uphold good grammar and a professional, academic writing style throughout while still making it engaging to the reader.

To summarize, I have been able to outline a valuable discussion on Stochastic Simulations, Deterministic Modelling and their connections and differences. This taught a new reader on the topic the skills to manipulate and analyse both types of modelling, understand why a modelling process is giving you the information, and why it's useful, as well as what connects them.

Appendix A

Code

All code used in this project can be found at https://github.com/JamesChalk22/Dissertation for your consideration.

Bibliography

Sillicon Cloud Technologies. Portfolio visualiser. `https://www.portfoliovisualizer.com/monte-carlo-simulation`, 2023. Accessed: 22-03-2023.

Radek Erban and S Jonathan Chapman. *Stochastic modelling of reaction–diffusion processes*, volume 60. Cambridge University Press, 1 2020.

Dr. Noemi Picco. Ma-182 biomathematics. 2023.

Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81:2340–2361, 1977.

Johan Paulsson, Otto G Berg, and Måns Ehrenberg. Stochastic focusing: fluctuation-enhanced sensitivity of intracellular regulation. *Proceedings of the National Academy of Sciences*, 97(13):7148–7153, 2000.

Radek Erban, Jonathan Chapman, and Philip Maini. A practical guide to stochastic simulations of reaction-diffusion processes. *arXiv preprint arXiv:0704.1908*, 2007.

Anatol M Zhabotinsky. A history of chemical oscillations and waves. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 1(4):379–386, 1991.

Cyrill B Muratov, Eric Vanden-Eijnden, and E Weinan. Self-induced stochastic resonance in excitable systems. *Physica D: Nonlinear Phenomena*, 210(3-4):227–240, 2005.