Advanced Mathematical Biology IV (MATH4411)

Dr Denis Patterson, Durham University

2025-06-01

Table of contents

Michaelmas Term Overview

In Michaelmas term of this course, we will study a range of stochastic models that allow us to analyse populations of biological agents moving and interacting in both space and time. We will consider general individual-based models in the form of stochastic simulation algorithms accounting for reactions among chemical species, or interactions between individuals in a population. We show how these models can, in various limits, give rise to the same continuum-level descriptions obtained via conservation laws at the macroscale. However, we will also demonstrate novel features of stochastic individual-based models, such as stochastic resonance and noise-induced attractor switching, that are not present in the deterministic macroscale models.

A key focus will be on designing and understanding stochastic simulation algorithms to simulate various stochastic processes on a computer. Simultaneously, we will develop a toolkit of analytic approaches to studying these processes that will validate and complement direct numerical simulations. We begin by studying how to simulate spatially homogeneous reaction systems, before moving on to more complex spatially extended systems later in the course. Applications will include models from ecology, neuroscience, chemistry, genetics and cell biology.

Content

The content of Term 1 is divided into 4 Chapters, some of which will be longer than others as we devote more time earlier in the course to foundational concepts that will be needed throughout.

- **Chapter 1: Stochastic Simulation of Chemical Reactions**

  - Stochastic models of 1st, 2nd, and higher order chemical reaction systems
  - "Naive" versus Gillespie stochastic simulation algorithms
  - The chemical master equation
  - Stationary distributions and probability generating functions

- **Chapter 2: Deterministic vs Stochastic Models**

  - Disagreement between ODEs and stochastic systems
  - Stochastic resonance
  - Finite-size effects and stochastic forcing

- **Chapter 3: Stochastic Differential Equations (SDEs)**

  - Computational definition of SDEs
  - Examples with drift, diffusion and bistability
  - Introduction to the Fokker-Planck equation (and Chemical F-P equation)
  - Mean switching times and bistability

- **Chapter 4: Stochastic Reaction-diffusion Models**

  - Modeling diffusion with SDEs
  - Discrete approach to diffusion
  - Spatially discrete approach to reaction-diffusion
  - SDE approach to reaction-diffusion
  - Pattern formation in stochastic models

Lectures, Problem Classes & Homeworks

Lectures will primarily be used to present new material, but will also feature computer demonstrations of the algorithms and models presented. As such, students are encouraged to bring along their own laptops to both lectures and problem classes so that they can also run the examples themselves. The MATLAB code for all of the examples will be available on the course Github page:

https://github.com/patterd2/MATH4411_Adv_Math_Bio

You don't need any prior coding experience in MATLAB to run this code; mostly you will just want to tweak parameter values, analyse the output of the code, and compare the algorithms to the pseudocode in the lecture notes.

Problem classes will not contain any new material, but will focus on the presentation of additional examples, discussion of lecture material, and solving problems from the problem sheets. There will be 4 short homework assignments that will be graded for the purposes of formative assessment (i.e. letter grade of A-D with additional comments and feedback). Homeworks will be due every two weeks with the first homework due in week 4 (precise submission instructions are available on Ultra and on the homework question sheets themselves).

	Activities	Content
Week 1	Introductory lecture, 1 lecture	Chapter 1
Week 2	2 lectures, 1 problem class	Chapter 1
Week 3	2 lectures, HW1 due	Chapter 1
Week 4	2 lectures, 1 problem class	Chapter 2
Week 5	2 lectures, HW2 due	Chapter 2
Week 6	2 lectures, 1 problem class	Chapter 3
Week 7	2 lectures, HW3 due	Chapter 3

	Activities	Content
Week 8	2 lectures, 1 problem class	Ch. 3/Ch. 4
Week 9	2 lectures, HW4 due	Chapter 4
Week 10	1 lecture, 1 rev. class + extra prob. sheet	Chapter 4

Additional Reading

The lecture notes are designed to be sufficient for the course and hence students do not need to purchase a textbook to successfully complete the course. References for additional reading will also be given at the end of each chapter. The main reference for the course content in Michaelmas term is:

- **Erban, R. & Chapman, S.J.** *Stochastic Modelling of Reaction-Diffusion Processes*, Vol. 60, Cambridge University Press, 2020.

Supplementary material and further reading:

- **Durrett, R. & Levin, S.** *The importance of being discrete (and spatial)*, Theoretical Population Biology, 46(3), 363-394, 1994.
- **Keener, J.P.** *Biology in Time and Space: A Partial Differential Equation Modeling Approach*, Vol. 50, American Mathematical Society, 2021.
- **Murray, J.D.** *Mathematical Biology: II: Spatial Models and Biomedical Applications*, Vol. 3, Springer, 2003.

Contact Information

For questions or clarifications on any of the above, please come speak to me in lectures, office hours or drop me an email at denis.d.patterson@durham.ac.uk.

1 Stochastic Simulation of Chemical Reactions

The goal of this chapter is to introduce stochastic chemical reaction processes and some algorithms to simulate them numerically. We will also develop some analytic tools to study these processes and begin to discuss similarities and differences between deterministic and stochastic models of the same phenomena.

1.1 Stochastic Simulation of Degradation

1.1.1 Introducing Stochastic Simulation Algorithms (SSAs)

We begin with the simplest possible chemical reaction process. Consider an experiment starting with n_0 molecules of a chemical species denoted by A; A reacts at rate k in the absence of any stimulus and upon reacting, a molecule of A transitions into some other species that is not of interest to us.

We can represent this reaction process more succinctly as:

$$A \xrightarrow{k} \emptyset.$$
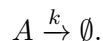
The symbol $\emptyset$ here simply denotes a species not of interest in the experiment, rather than meaning that a molecule of A disappears after reacting. Note also that we are not considering the spatial extent of this experiment; we are assuming that the system is spatially homogeneous and that each molecule of A reacts exactly as described above regardless of its spatial position (an assumption we will relax later in the course). We also choose to model this system as evolving in continuous time.

Let $A(t)$ denote the number of molecules of species A in the experiment at time t. Suppose $dt > 0$ is a small quantity (think $\epsilon > 0$ that will later be sent to zero) and consider the dynamics of the process in the time interval $[t, t + dt)$. There are 3 possibilities for what can happen in $[t, t + dt)$ and we assign them the following probabilities:

- **No reactions occur:**
  $\mathbb{P}[A(t + dt) = A(t)] \approx 1 - A(t)kdt + \mathcal{O}(dt^2)$
- **Exactly one reaction occurs:**
  $\mathbb{P}[A(t + dt) = A(t) - 1] \approx A(t)kdt + \mathcal{O}(dt^2)$

- **More than one reaction occurs:**
  $\mathbb{P}[A(t + dt) < A(t) - 1] \approx \mathcal{O}(dt^2)$

If dt is sufficiently small, we can neglect the $\mathcal{O}(dt^2)$ terms and thus ignore the possibility of multiple reactions occurring "simultaneously".

> **i** Note
>
> From its definition, the degradation process has the **Markov property**, i.e.
>
> $$\mathbb{P}[A(t + dt) = n \mid A(s), s \leq t] = \mathbb{P}[A(t + dt) = n \mid A(t)].$$
>
> Can you give an intuitive explanation of why this is the case? Since it is posed in continuous time, A is formally called a **Markov jump process**.

Deferring the mathematical analysis of this process for now, how could we simulate a sample path of the process to find the number of molecules of A remaining at time t given $A(0) = n_0$? We could try to find a chemical with reaction rate k and conduct the actual experiment, or we can simulate the process on a computer! To this end, we may introduce the following "naive" stochastic simulation algorithm (SSA) for the process:

At time $t = 0$, set $A(0) = n_0$, then:

1. Generate a random number $r \sim U([0, 1])$.
2. If $r < A(t)k\Delta t$, then

   - set $A(t+\Delta t) = A(t)-1$, set $t = t+\Delta t$, and go back to step 1, else set $A(t+\Delta t) = A(t)$, set $t = t + \Delta t$, and go back to step 1.

We can stop this procedure whenever there are no molecules of A left to react or when t reaches some maximal stopping time T that we choose in advance. The quantity $\Delta t > 0$ is some fixed discretisation parameter that we have chosen a priori.

But does this algorithm really simulate the process described above?

To understand the motivation for the steps, recall that a uniform random variable $r \sim U([0, 1])$ has CDF given by

$$F(x) = \mathbb{P}[r < x] = \begin{cases} 0, & x < 0, \\ x, & x \in [0, 1], \\ 1, & x > 1. \end{cases}$$

Hence the probability of one reaction occurring in the interval $[t, t + \Delta t)$ is $\mathbb{P}[r < A(t)k\Delta t] = A(t)k\Delta t$, assuming that $A(t)k\Delta t < 1$. This emphasises the need to choose Δt sufficiently small for the algorithm to have the appropriate reaction probabilities. In fact, we want to choose $\Delta t \ll 1/A(t)k$ because this ensures that the probability of multiple reactions taking place in $[t, t + \Delta t)$ is very small, which is what we want according to the definition of the process above. However, there is a trade-off between accuracy and speed here: a smaller Δt means lower probability of multiple reactions and numerical error, but a smaller Δt also means more compute time (and more intervals in which no reactions take place).

Virtually all modern programming languages have algorithms for generating (pseudo-) random numbers and we can use such routines to generate the uniformly distributed random numbers called for in step 1 to execute the **naive SSA**. The output of some simulations of the degradation process using this algorithm are shown in Figure 1.1. The left panel shows two sample paths compared to what we called the **deterministic mean**. The motivation for this comparison is that a principle called **the law of mass action** can be invoked to yield a simple deterministic model of the reaction process.

> **i** Note
>
> The law of mass action states that "the rate of a chemical reaction is directly proportional to the product of the activities or concentrations of the reactants."

If we assume our chemical is at unit volume (for simplicity), then A reacts at a constant rate and so the deterministic model is given by the linear ODE

$$\frac{d}{dt}A(t) = -kA(t), \quad A(0) = n_0.$$

This equation approximately describes the evolution of the mean of the stochastic model and is readily solved to show that

$$A(t) = n_0 e^{-kt}, \quad t \geq 0.$$

The deterministic model of the mean behaviour is strictly decreasing in time and thus does not capture the qualitative behaviour of individual trajectories very well. However, the right panel of Figure 1.1 shows that it captures the behaviour of the average over a large number of simulations quite well; the stochastic mean at time t is calculated as the average value at that time over 20 simulations of the process.

It is worth noting that the naive SSA is very inefficient in this example as we chose our parameters to ensure a very small probability of multiple reactions occurring in $[t, t + \Delta t)$. In fact, we had $\mathbb{P}[\text{one reaction in } [t, t + \Delta t)] \approx 0.01$. But this means that step 1 is mostly wasted as we generate lots of random numbers that we don't use! We can be more efficient than that...
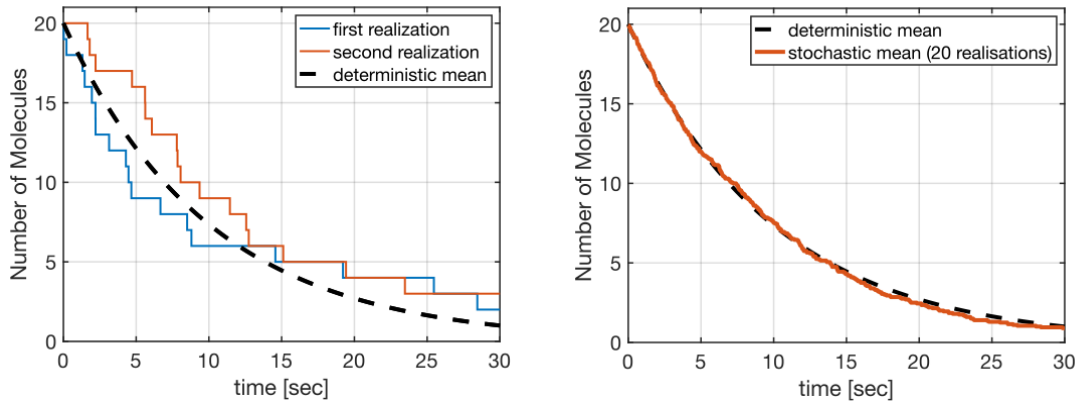
Figure 1.1: Left: Some paths of the degradation process compared to the corresponding deterministic model of the same process. Right: Stochastic mean compared to the mean predicted by the deterministic model. Parameters: $n_0 = 20$, $k = 0.1$, $\Delta t = 0.005$.

Exercise 1.1

Open the course Github page and try running the MATLAB script:
`CH1_naive_SSA_degradation.m`

1. How does the difference between the deterministic and stochastic means change as you vary the number of realisations?

2. What happens if you set $\Delta t = 1$ and how can we explain this behaviour?

From an efficiency standpoint, we can do much better than the naive SSA, but this advancement requires a change of perspective. In developing the naive SSA we focused on matching the transition probabilities of the underlying process on short time intervals, stepping forward a short time step and then repeating the probability matching step. We might instead ask: **When is the next reaction going to take place?** If we know when the next reaction will take place, then we can just skip ahead to that time, let the reaction occur, and repeat this process.

The problem we have now is that the time of the next reaction is random. But since it is just some (continuous) random variable related to the reaction rate, we can compute its distribution. To this end, let τ denote the next reaction time and take $t, s > 0$. Define the function $f(A(t), s)ds$ to be the probability that no reaction occurs in the interval $[t, t+s]$ and that exactly one reaction occurs in $[t+s, t+s+ds]$. $f(A(t), s)$ is the PDF of τ, the random variable giving the next reaction time, and so by computing f, we can identify the distribution of τ.

Suppose there are n molecules of A at time t, then

$$
\begin{aligned}
f(A(t), s)ds &= \mathbb{P}\left[A(t + s + ds) = n - 1,\ A(t + s) = n \mid A(t) = n\right] \\
&= \underbrace{\mathbb{P}\left[A(t + s + ds) = n - 1 \mid A(t + s) = n\right]}_{\text{one reaction in } [t+s,\, t+s+ds]} \\
&\quad \times \underbrace{\mathbb{P}\left[A(t + s) = n \mid A(t) = n\right]}_{\text{no reactions in } [t+s]}.
\end{aligned}
$$

Letting $g(A(t), s)$ denote the probability that no reaction occurs in the interval $[t, t + s]$, we can write the formula above more succinctly as

$$
\begin{aligned}
f(A(t), s)ds &= g(A(t), s)A(t + s)kds \\
&= g(A(t), s)A(t)kds,
\end{aligned}
$$

where the second equality holds because we are conditioning on no reactions occurring in $[t, t + s]$.

For any $\sigma > 0$, using the definition of $g(A(t), s)$ yields that

$$
\begin{aligned}
g(A(t), \sigma + d\sigma) &= g(A(t), \sigma)\left[1 - A(t + \sigma)kd\sigma\right] \\
&= g(A(t), \sigma)\left[1 - A(t)kd\sigma\right]
\end{aligned}
$$

where the last equality follows because $A(t) = A(t + \sigma)$ if there is no reaction in $[t, t + \sigma]$. Rearranging the equality above and letting $d\sigma \downarrow 0$ then gives that g obeys the ODE

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

Note that the derivative is with respect to σ here, not t! Solving this ODE gives

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

since $g(A(t), 0) = \mathbb{P}[\text{no reaction at time } t] = 1$. Plugging this back into the previous formula and cancelling ds on both sides shows that

$$
f(A(t), s) = A(t)ke^{-A(t)ks}.
$$

Recall that the PDF of an exponential random variable with rate parameter λ is given by

$$
f(x) = \lambda e^{-\lambda x}, \quad x \geq 0,
$$

and hence we can conclude that $\tau \sim \text{Exp}[kA(t)]$.

> **i Note**
>
> In general, the waiting times (time intervals between reactions/jumps) of a Markov jump process are exponentially distributed.

We want to simulate the degradation process and so we now need to simulate exponentially distributed random numbers in order to find the next reaction time at each step. Previously, we assumed that our computer could generate random numbers that are uniformly distributed on $[0, 1]$, but fortunately, there is a simple way to generate an exponential random variable from a uniform one.

Suppose $U \sim U([0, 1])$ and F denotes the CDF of a continuous random variable. If F is invertible, then $V = F^{-1}(U)$ is distributed according to F. To see this, just calculate the CDF of V as follows:

$$\mathbb{P}[V \leq x] = \mathbb{P}[F^{-1}(U) \leq x] = \mathbb{P}[U \leq F(x)] = F(x).$$

Can you justify all of the equalities in the calculation above? This procedure is called the **inverse transform method** and allows us to sample from any continuous distribution with an invertible CDF given uniform random numbers.

The CDF of the exponential distribution with rate $\lambda > 0$ is given by

$$F(x) = \begin{cases} 0, & x < 0, \\ 1 - e^{-\lambda x}, & x \geq 0, \end{cases}$$

and hence

$$F^{-1}(x) = \begin{cases} -\frac{1}{\lambda}\log(1 - x) = \frac{1}{\lambda}\log\left(\frac{1}{1-x}\right), & x \in (0, 1), \\ 0, & \text{else.} \end{cases}$$

Thus, if $U \sim U([0, 1])$, then $\log(1/(1 - U))/\lambda \sim \text{Exp}(\lambda)$. Furthermore, $1 - U \sim U([0, 1])$ and so $\log(1/U)/\lambda \sim \text{Exp}(\lambda)$ as well; this is the formula that we will use to generate exponentially distributed random numbers in the algorithms that follow.

> **Exercise 1.2**
>
> 1. Calculate the inverse CDF for the exponential distribution and verify the formula above!

2. Apply the inverse transform method to the Pareto distribution, which has PDF given by

$$f(x) = \begin{cases} \frac{\alpha \, \beta^{\alpha}}{x^{\alpha+1}}, & x \geq \beta 0, \\ 0, & \text{else.} \end{cases}$$

We're now ready to introduce our **more efficient SSA**:

At time $t = 0$, set $A(0) = n_0$, then:

1. Generate a random number $r \sim U([0, 1])$.
2. Compute the next reaction time by calculating

$$\tau = \frac{1}{A(t)k} \log(1/r).$$

3. Set $t = t + \tau$, set $A(t + \tau) = A(t) - 1$ and go back to step 1.

Figure 1.2 below shows the results of some simulations of the degradation process using the more efficient SSA. There is a notable speed-up compared to the naive SSA when plotting large numbers of realisations. We take advantage of this to illustrate further differences between the deterministic model and stochastic one in the right panel of Figure 1.2; this panel shows a comparison between the estimated PMF of the process at $t = 10$ and the deterministic mean at the same time. We used 200 paths of the process for this plot but more are probably required to fully capture the PMF of even this simple process, further highlighting the need for efficient SSAs!

Exercise 1.3

Open the course Github page and try running the MATLAB script:
`CH1_more_efficient_SSA_degradation.m`
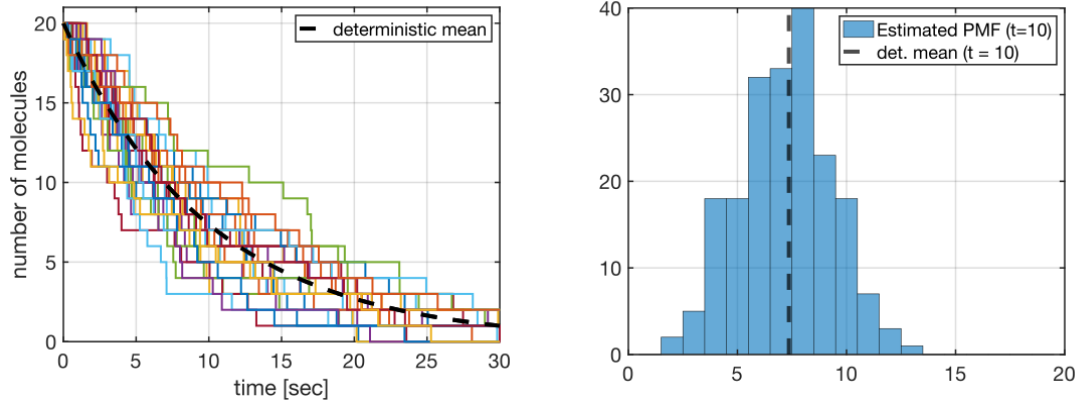
Figure 1.2: Left: Some paths of the degradation process simulated using the more efficient SSA. Right: Estimated PMF of the degradation process based on 200 realisations at time $t = 10$. Parameters: $n_0 = 20$, $k = 0.1$.

1.1.2 The Chemical Master Equation

We now have a couple of ways to simulate the degradation process numerically, but we also need some mathematical tools to analyse the process analytically. In modelling, there is very often an interplay between theory and numerics; we typically want to prove as much as possible about a model rigorously and use numerics to generate conjectures and address questions that are not possible to answer with theoretical tools.

To this end, we introduce the **chemical master equation** for the degradation process by letting

$$P_n(t) = \mathbb{P}[A(t) = n], \quad t \geq 0, \quad n \in \mathbb{N}.$$

For a fixed value of t, $P_n(t)$ is simply the probability mass function of the random variable $A(t)$, so we have a collection of PMFs indexed by (continuous) time. Our goal is to deduce an evolution equation for how $P_n(t)$ changes over time. As always, we assume there is an infinitesimally small quantity dt, so that $\mathcal{O}(dt^2)$ terms are negligible, and begin by considering the event $\{A(t + dt) = n\}$. There are two ways we can arrive at this event, either

- $A(t) = n$ and there were no reactions in $[t, t + dt)$, **or**
- $A(t) = n + 1$ and there was one reaction in $[t, t + dt)$.

Hence

$$P_n(t + dt) = \underbrace{P_n(t)(1 - kndt)}_{\text{no reactions in } [t,t+dt)} + \underbrace{P_{n+1}(t)(k(n + 1)dt)}_{\text{one reaction in } [t,t+dt)}.$$

14

Rearrangement yields

$$\frac{P_n(t + dt) - P_n(t)}{dt} = k(n + 1)P_{n+1}(t) - knP_n(t),$$

and letting $dt \downarrow 0$ thus gives the *system* of ODEs

$$\frac{d}{dt}P_n(t) = k(n + 1)P_{n+1}(t) - knP_n(t), \quad t \geq 0, \quad n \geq 0.$$

> **i** Note
>
> In the theory of Markov jump processes, this equation is called the **Kolmogorov Forward Equation** (sometimes just the Kolmogorov equation) for the process. You might see this terminology used in other texts or papers.

The function $P_n(t)$ contains all of the information we could ever want to know about the stochastic process, but to obtain this information we need to solve the system of equations above! In general, this is far from a trivial task for complex processes with many chemical species and reactions (and that is before we even consider adding spatial extent to the system!). However, we can solve this system directly for the degradation process.

We can simplify the system immediately by thinking about the initial conditions of the process. Since $A(0) = n_0$, we have

$$P_{n_0}(0) = 1, \quad P_n(0) = 0 \quad \forall n \neq n_0.$$

Moreover, the process $A(t)$ is non-increasing because we can never produce any new molecules of A. Hence

$$P_n(t) = 0, \quad \forall n > n_0.$$

These two facts already provide a significant simplification because we now have a finite system of ODEs (after starting with a countable number of ODEs to solve) and we have the initial condition at $t = 0$ for each value of n from above. Our strategy from here is to start with the equation for P_{n_0} and try to solve the system iteratively (until we hopefully see a pattern). The equation for P_{n_0} reads:

$$\frac{d}{dt}P_{n_0}(t) = k(n_0 + 1)P_{n_0+1}(t) - kn_0 P_{n_0}(t), \quad P_{n_0}(0) = 1.$$

But, as we just noted, $P_{n_0+1}(t) = 0$ so we can immediately solve to find that

$$P_{n_0}(t) = P_{n_0}(0)e^{-kn_0 t} = e^{-kn_0 t}.$$

Next tackle the equation for $P_{n_0-1}(t)$, which now reads:

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

Thankfully, this equation is still linear in P_{n_0-1}, but it is inhomogeneous so we need to use the variation of constants formula to solve it. We thus obtain

$$P_{n_0-1}(t) = P_{n_0-1}(0)e^{-k(n_0-1)t} + e^{-k(n_0-1)t}\int_0^t e^{k(n_0-1)s}kn_0 e^{-kn_0 s}\,ds$$

$$= n_0 e^{-k(n_0-1)t}\left(1 - e^{-kt}\right).$$

Now we begin to see the pattern: We can take the solution for P_{n_0-1}, plug it into the equation for P_{n_0-2} and apply the variation of constants formula to then solve that linear inhomogeneous ODE, and carry on this procedure until we reach P_0. In fact, we can show by induction that the general formula for P_n is given by

$$P_n(t) = \begin{cases} 0, & n > n_0, \quad t \geq 0, \\ \binom{n_0}{n}e^{-knt}(1 - e^{-kt})^{n_0-n}, & 0 \leq n \leq n_0, \quad t \geq 0. \end{cases}$$

We could have guessed the case $n > n_0$: there is no chance that we can have more than n_0 molecules of A! The second part of the formula is less obvious but should look familiar: at each fixed time t, this is the PMF of a Binomial random variable with parameters n_0 (number of trials) and e^{-kt} (probability of success per trial). Succinctly, we may write

$$P_n(t) \sim \text{Binomial}(n_0,\, e^{-kt}).$$

Given the solution to the **chemical master equation**, we can answer virtually any question regarding the process. In practice, we may want to know about its average behaviour, its fluctuations (which are often characterised by the variance), or the probability of it hitting a certain value by a given time. We can compute the mean of the process, $M(t)$, directly from the formula above as follows:

$$M(t) := \mathbb{E}[A(t)] = \sum_{n=-\infty}^{\infty} nP_n(t) = \sum_{n=0}^{n_0} nP_n(t)$$

$$= \sum_{n=0}^{n_0} n \binom{n_0}{n} e^{-knt}(1 - e^{-kt})^{n_0-n}$$

$$= n_0 e^{-kt} \sum_{m=0}^{n_0-1} \binom{n_0-1}{m} \left(e^{-kt}\right)^m (1 - e^{-kt})^{(n_0-1)-m}$$

$$= n_0 e^{-kt}.$$

This exactly matches the deterministic model for the mean behaviour that we obtained by solving the ODE obtained via the law of mass action. However, this is not true in general and it is also worth noting that the deterministic ODE can't tell us more detailed information about the paths, like their fluctuations or hitting times. For more complicated examples (coming in later chapters), we will see more dramatic disagreement between the mean behaviour predicted by the law of mass action and the dynamics of the underlying stochastic process.

> **Exercise 1.4**
>
> 1. Complete the inductive step from $n_0 - k$ to $n_0 - (k+1)$ to obtain the formula above.
>
> 2. Use the more efficient SSA code to show that with enough realisations we obtain an approximate Binomial distribution, in agreement with the formula above.

1.2 Stochastic Simulation of Production & Degradation

1.2.1 The production/degradation process

We next consider a variation on the degradation process where we now add production of species A to the set of possible reactions. We are still assuming that there is some container of volume ν that holds the molecules of chemical A and that we can neglect the spatial extent of the experiment. The new reaction process can be written as:

$$A \xrightarrow{k_1} \emptyset, \quad \emptyset \xrightarrow{k_2} A.$$
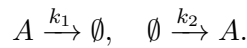
Once more, the $\emptyset$ simply denotes some chemical species not of interest, rather than denoting that A is being produced out of thin air! Given that the volume of the container is ν, we will assume that production of A is proportional to $k_2\nu$ so that the input scales appropriately with the system size. It is worth remarking that the units of k_1 will be $\sec^{-1}$ and those of k_2 are $\sec^{-1}m^{-3}$ because production was chosen to scale with volume. This all leads to the following characterization of the transition rates in a (small) interval $[t, t + dt)$:

- No reactions occur:
  $\mathbb{P}[A(t + dt) = A(t)] \approx 1 - A(t)k_1 dt - k_2 \nu dt + \mathcal{O}(dt^2)$
- One molecule of A is produced:
  $\mathbb{P}[A(t + dt) = A(t) + 1] \approx k_2 \nu dt + \mathcal{O}(dt^2)$
- One molecule of A is degraded:
  $\mathbb{P}[A(t + dt) = A(t) - 1] \approx A(t)k_1 dt + \mathcal{O}(dt^2)$
- More than one reaction occurs:
  $\mathbb{P}[\text{multiple reactions in } [t, t + dt]] \approx \mathcal{O}(dt^2).$

Based on the transition rates outlined above, we can define the **propensity function** of the process

$$\alpha(t) := A(t)k_1 + k_2 \nu.$$

This definition is motivated by the fact that the probability of a reaction occurring in $[t, t + dt)$ is given by $\alpha(t)dt$ $(+\mathcal{O}(dt^2)$ terms). We might also call the propensity function the **total reaction rate** of the process.

Next we outline a version of our **efficient SSA** that we claim simulates the production/degradation process described above:

At time $t = 0$, set $A(0) = n_0$, then:

1. Generate random numbers r_1, $r_2 \sim U([0, 1])$.

2. Compute the propensity function for the process, i.e.

$$\alpha(t) = k_1 A(t) + k_2 \nu.$$

3. Compute the next reaction time by calculating

$$\tau = \frac{1}{\alpha(t)} \log(1/r_1).$$

4. Compute the number of A molecules at time $t = t + \tau$:

$$A(t + \tau) = \begin{cases} A(t) + 1, & \text{if } r_2 < k_2 \nu/\alpha(t), \\ A(t) - 1, & \text{if } r_2 \geq k_2 \nu/\alpha(t), \end{cases}$$

and set $t = t + \tau$.

Go back to step 1.

Does this scheme correctly simulate the production/degradation process?

Step 3 in the scheme generates an exponentially distributed random number with rate parameter $\alpha(t)$ as the next reaction time. We could consider a single reaction, as in the last section, by letting **production or degradation** be one reaction, the reaction rate of this simplified process would be:

$$
\begin{aligned}
\mathbb{P}[\text{one reaction in } [t, t+dt]] &= \mathbb{P}[\text{production in } [t, t+dt) \text{ or degradation in } [t, t+dt)] \\
&= \mathbb{P}[\text{production, no degradation}] + \mathbb{P}[\text{no production, degradation}] \\
&= (k_1 A(t) dt)(1 - k_2 \nu dt) + (k_2 \nu dt)(1 - k_1 A(t) dt) \\
&= (k_1 A(t) + k_2 \nu) dt = \alpha(t) \, dt.
\end{aligned}
$$

In other words, the single reaction or reaction/no-reaction process has reaction rate $\alpha(t)$ and we showed in the previous section that the waiting time before its next reaction time is exponential distributed with rate parameter $\alpha(t)$, thus justifying step 3.

Step 4 tells us, conditional on a reaction having taken place in $[t, t+dt)$, whether a production or degradation reaction took place. Since $r_2 \sim U([0,1])$, we can calculate the (conditional) probabilities of production or degradation:

$$
\begin{aligned}
\mathbb{P}[\text{production}] &= \mathbb{P}\left[r_2 < \frac{k_2 \nu}{k_1 A(t) + k_2 \nu}\right] = \frac{k_2 \nu}{k_1 A(t) + k_2 \nu}, \\
\mathbb{P}[\text{degradation}] &= \mathbb{P}\left[r_2 \geq \frac{k_2 \nu}{k_1 A(t) + k_2 \nu}\right] \\
&= 1 - \frac{k_2 \nu}{k_1 A(t) + k_2 \nu} = \frac{k_1 A(t)}{k_1 A(t) + k_2 \nu}.
\end{aligned}
$$

These calculations were facilitated by the fact that, given positive rate constants, volume and a non-zero number of molecules, we have

$$
\frac{k_2 \nu}{k_1 A(t) + k_2 \nu} \in (0, 1), \quad \frac{k_1 A(t)}{k_1 A(t) + k_2 \nu} \in (0, 1).
$$

We conclude from these computations that the relative probabilities (and they are true probabilities being in $(0, 1)$) of production and degradation are proportional to their rates, which at least seems a sensible state of affairs!

A more rigorous justification of steps 3 and 4 is to note that **conditional on the current value of the process**, $A(t)$, the production and degradation reactions are independent single-reaction processes (until the next reaction occurs). Therefore the waiting time until the next

19

degradation reaction is exponentially distributed with rate parameter $k_1 A(t)$ and the waiting time until the next production reaction is exponentially distributed with rate parameter $k_2\nu$. Hence the waiting time until the next reaction time for the production/degradation system is simply the minimum of these two waiting times. Once the next reaction takes place, $A(t)$ changes, these two (independent) clocks are reset and we wait again to see which takes place first.

Suppose $E_D \sim \text{Exp}(k_1 A(t))$ and $E_P \sim \text{Exp}(k_2\nu)$ so that $\tau = \min(E_D, E_P)$ is the waiting time until the next reaction in the production/degradation process. Crucially, E_D and E_P are independent here because $A(t)$ is fixed. What is the distribution of τ? Compute the CDF of τ as follows:

$$
\begin{aligned}
\mathbb{P}[\tau > x] = \mathbb{P}[\min(E_D, E_P) > x] &= \mathbb{P}[E_P > x, E_D > x] \\
&= \mathbb{P}[E_D > x]\,\mathbb{P}[E_P > x] \\
&= e^{-k_1 A(t)x}e^{-k_2\nu x} = e^{-(k_1 A(t)+k_2\nu)x} \\
&= e^{-\alpha(t)x}, \quad x > 0.
\end{aligned}
$$

Thus $\tau \sim \text{Exp}(\alpha(t))$, as claimed before. Moreover, we can directly compute the probability that a production reaction occurs next by computing $\mathbb{P}[\text{production}] = \mathbb{P}[E_P < E_D]$ since this is just the probability that one exponential random variable is less than another (and they are both independent!). Carrying out this calculation shows that

$$
\mathbb{P}[\text{production}] = \mathbb{P}[E_P < E_D] = \frac{k_2\nu}{k_1 A(t) + k_2\nu},
$$

in agreement with the formulae above.

> **Exercise 1.5**
>
> Directly compute $\mathbb{P}[E_P < E_D]$ to show that the formula above holds.

Figure 1.3 shows some sample paths of the production/degradation process generated using the more efficient SSA. All sample paths begin with $A(0) = 0$ but, for this parameter choice, production outpaces degradation initially before the process levels off somewhat around 10 molecules (on average). The deterministic mean in this plot is generated by using the law of mass action to derive a deterministic model for the average behaviour of the system.

> **Exercise 1.6**
>
> 1. Open the course Github page and try running the MATLAB script: `CH1_production_degradation.m`
> 2. Use the law of mass action to derive a deterministic model for the produc-
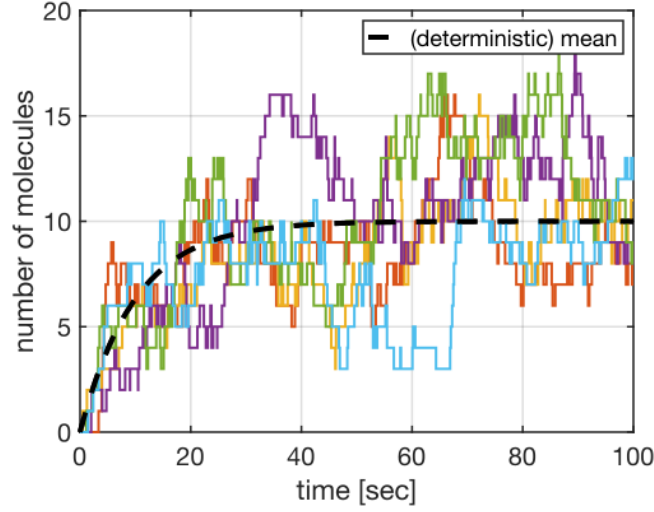
Figure 1.3: Sample paths of the production/degradation process generating using the more efficient SSA and compared to the deterministic mean; the deterministic mean was obtained by solving the deterministic model generated by the law of mass action. Parameters: $k_1 = 0.1$, $k_2 = 1$, $\nu = 1$, $A(0) = n_0 = 0$.

tion/degradation process and verify the formula for the deterministic mean in the code above.

1.2.2 Deriving the chemical master equation

To derive the chemical master equation for the production/degradation process we again consider the question: How can the number of molecules change between time t and time $t + dt$? Suppose we have n molecules at time $t + dt$, i.e. $A(t) = n$, then we either had $n + 1$ molecules at time t and a degradation reaction took place in the interval $[t, t + dt)$, **or** we had $n - 1$ molecules and a production took place in $[t, t+dt)$, **or** we had n molecules at time t and nothing happened in $[t, t + dt)$! Letting $P_n(t) = \mathbb{P}[A(t) = n]$, we can express this probabilistically as follows:

$$P_n(t + dt) = P_n(t)\left[1 - k_1 n dt - k_2 \nu dt\right] + P_{n+1}(t)\left[k_1(n + 1)dt\right] + P_{n-1}(t)\left[k_2 \nu dt\right].$$

Rearranging and letting $dt \downarrow 0$ thus yields

$$\frac{d}{dt}P_n(t) = k_1(n + 1)P_{n+1}(t) - k_1 n P_n(t) + k_2 \nu P_{n-1}(t) - k_2 \nu P_n(t), \quad \text{for all } n \geq 0,$$

21

with the convention that $P_n(t) \equiv 0$ for all $n < 0$.

Unlike the pure degradation process, there is no upper bound on the number of molecules in the production/degradation process and hence we can't reduce this countably-infinite system of ODEs to a finite system like we could previously. In practice, we would hope to show (or somehow know by other means!) that $P_n(t)$ tends to zero as $n \to \infty$; this would justify truncating the system at some large but finite value of n in order to solve it numerically.

Even for this relatively simple process, the chemical master equations are challenging to solve analytically. Instead, we will compute the mean and variance processes, i.e.

$$M(t) = \sum_{n=0}^{\infty} n P_n(t), \quad V(t) = \sum_{n=0}^{\infty} (n - M(t))^2 P_n(t),$$

as these are typically more tractable and still offer considerable insight into the dynamics of the process.

To derive an evolution equation for the mean, multiply the master equation by n and sum over n:

$$\frac{d}{dt} \sum_{n=0}^{\infty} n P_n(t) = k_1 \sum_{n=0}^{\infty} n(n+1) P_{n+1}(t) - k_1 \sum_{n=0}^{\infty} n^2 P_n(t)$$

$$+ k_2 \nu \sum_{n=0}^{\infty} n P_{n-1}(t) - k_2 \nu \sum_{n=0}^{\infty} n P_n(t)$$

$$= k_1 \sum_{m=0}^{\infty} m(m-1) P_m(t) - k_1 \sum_{n=0}^{\infty} n^2 P_n(t)$$

$$+ k_2 \nu \sum_{m=0}^{\infty} (m+1) P_m(t) - k_2 \nu \sum_{n=0}^{\infty} n P_n(t)$$

$$= -k_1 \sum_{n=0}^{\infty} n P_n(t) + k_2 \nu \sum_{n=0}^{\infty} P_n(t)$$

$$\iff \frac{d}{dt} M(t) = -k_1 M(t) + k_2 \nu,$$

where the final equality used the fact that $\sum_{n=0}^{\infty} P_n(t) = 1$.

> **i Note**
>
> In the preceding calculation we freely exchanged the limiting operations of differentiation

and infinite summation, i.e. we implicitly claimed that

$$\frac{d}{dt}\sum_{n=0}^{\infty} nP_n(t) = \sum_{n=0}^{\infty} n\frac{d}{dt}P_n(t).$$

Changing the order of limiting operations can, and often does, change the result of a calculation. For example, consider the function $f(n,m) = n/(n+m)$ and let both n and m tend to infinity, order matters here! However, we can interchange limits without worrying about changing the result when the conditions of **uniform convergence** are satisfied. You can assume we have uniform convergence throughout this course, unless explicitly noted otherwise.

It is straightforward to then solve this linear inhomogeneous ODE for $M(t)$ to show that

$$M(t) = M(0)e^{-k_1 t} + \frac{k_2 \nu}{k_1}\left(1 - e^{-k_1 t}\right), \quad t \geq 0.$$

Similarly, but with slightly more pain along the way, we can deduce an evolution equation for $V(t)$. This begins by simplifying the definition of $V(t)$ first:

$$V(t) = \sum_{n=0}^{\infty} (n - M(t))^2 P_n(t) = \sum_{n=0}^{\infty} (n^2 - 2nM(t) + M(t)^2)P_n(t)$$

$$= \sum_{n=0}^{\infty} n^2 P_n(t) - 2M(t)\sum_{n=0}^{\infty} nP_n(t) + M(t)^2 \sum_{n=0}^{\infty} P_n(t)$$

$$= \sum_{n=0}^{\infty} n^2 P_n(t) - 2M(t)^2 + M(t)^2$$

$$= -M(t)^2 + \sum_{n=0}^{\infty} n^2 P_n(t).$$

We need to write the sum involving n^2 in terms of $M(t)$ and $V(t)$ to obtain a closed system so go back to the master equation, multiply across by n^2 and sum over n:

$$\frac{d}{dt}\sum_{n=0}^{\infty} n^2 P_n(t) = k_1 \sum_{n=0}^{\infty} n^2(n+1)P_{n+1}(t) - k_1 \sum_{n=0}^{\infty} n^3 P_n(t)$$

$$+ k_2 \nu \sum_{n=0}^{\infty} n^2 P_{n-1}(t) - k_2 \nu \sum_{n=0}^{\infty} n^2 P_n(t)$$

$$= k_1 \sum_{n=0}^{\infty} (-2n^2 + n)P_n(t) + k_2 \nu \sum_{n=0}^{\infty} (2n+1)P_n(t)$$

Thus,

$$\frac{d}{dt}V(t) + 2M(t)\frac{d}{dt}M(t) = -2k_1(V(t) + M(t)^2) + k_1M(t) + 2k_2\nu M(t) + k_2\nu,$$

where the last implication requires us to insert the simplified $V(t)$ to replace the sums with n^2 terms. Tidying this up leaves us with the following evolution equation for $V(t)$:

$$\frac{d}{dt}V(t) = -2k_1V(t) + k_1M(t) + k_2\nu, \quad t \geq 0.$$

Equation above is another linear inhomogeneous ODE and can be solved given that we know $M(t)$. Instead, we focus on an approach that tends to be tractable for more complex systems, namely, to focus on the large time or asymptotic behaviour of the process. To this end, define the quantities M_s and V_s by

$$M_s := \lim_{t\to\infty} M(t), \quad V_s := \lim_{t\to\infty} V(t).$$

Assuming that these limiting quantities are well-defined for this system (which is far from a given in general!), we can compute them by solving the steady-state versions of the evolution equations (setting the time derivatives to zero):

$$0 = -k_1M_s + k_2\nu, \quad 0 = -2k_1V_s + k_1M_s + k_2\nu.$$

Thus

$$M_s = V_s = \frac{k_2\nu}{k_1}.$$

This already gives us good information on the average behaviour of the system, and fluctuations around that average, for large times but we can actually take our asymptotic analysis a step further. To do so, define the **stationary distribution** ϕ of the production/degradation process by:

$$\phi(n) := \lim_{t\to\infty} P_n(t), \quad n \geq 0.$$

If the limit above is well-defined for each $n \geq 0$, then we can compute the stationary distribution by solving the steady-state version of the chemical master equation, i.e.

$$0 = k_1(n+1)\phi(n+1) - k_1n\phi(n) + k_2\nu\phi(n-1) - k_2\nu\phi(n), \quad n \geq 0,$$

24

where $\phi(n) \equiv 0$ for all $n < 0$. We can solve this system recursively by starting at $n = 0$ and trying to then guess the general form of the solution. For $n = 0$, we have

$$0 = k_1 \phi(1) - k_2 \nu \phi(0) \implies \phi(1) = \frac{k_2 \nu}{k_1} \phi(0).$$

Since ϕ is a PMF, we have the additional normalisation constraint that $\sum_{n=0}^{\infty} \phi(n) = 1$ and this will allow us to determine $\phi(0)$ later if we can find the general formula for ϕ up to a multiplicative constant. It can be shown by induction that

$$\phi(n) = \frac{1}{n!} \left(\frac{k_2 \nu}{k_1} \right)^n e^{-k_2 \nu / k_1}, \quad n \geq 0,$$

and hence, for large times, the production/degradation process is approximately Poisson distributed with parameter $k_2 \nu / k_1$.

Figure 1.4 shows a comparison of the stationary distribution with the estimated PMF of the process at different times points. In all simulations, $A(0) = 0$ and we can see how the bias of the initial condition begins to disappear gradually as we take larger and larger time intervals. By $t = 100$, the estimated PMF is already in very close agreement with the asymptotic behaviour predicted by the stationary distribution.
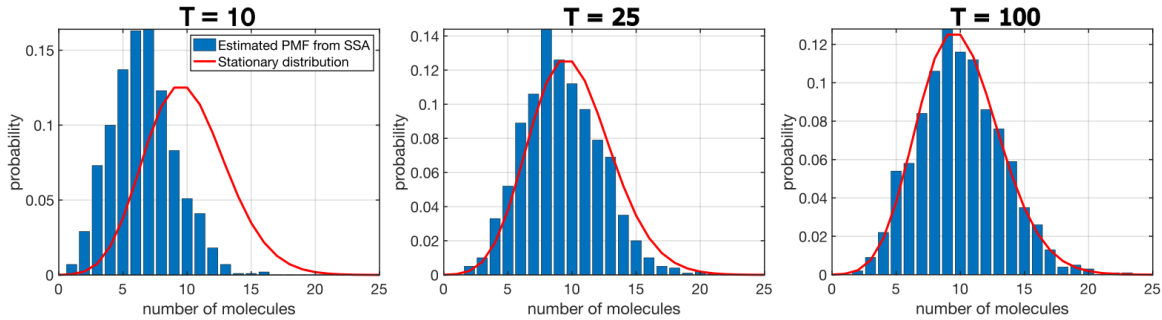
Figure 1.4: Simulations of the production/degradation process starting with $A(0) = 0$ with the PMF estimated at different times and compared to the stationary distribution. Parameters: $k_1 = 0.1$, $k_2 = 1$, $\nu = 1$, $A(0) = n_0 = 0$.

Exercise 1.7

Open the course Github page and try playing with the MATLAB script:
`CH1_production_degradation_CME.m`
Try varying the initial conditions and the time interval parameter T to observe how the mean and skewness of the estimated PMF vary.

1.3 Higher Order Chemical Reactions

Thus far we have only dealt with chemical reaction processes in which reactions depend linearly on the number of molecules of a given species. In reality, most chemical reactions of interest involve two or more molecules interacting. For example, a simple **second order** chemical reaction would be the reaction process:

$$A + B \xrightarrow{k} C,$$
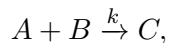
where a molecule of A and a molecule of B react at rate k to produce a molecule of C. We once more neglect spatial extent here but clearly the A and B molecules would need to come together in space to cause such a reaction. We are assuming the molecules are in some container with volume ν and hence the reaction rates should naturally scale with volume. For example, if the volume is increased, then A and B molecules will take longer on average to bump into each other (assuming there is no attraction/repulsion between them and that both species are distributed uniformly randomly in the container). Thus we will assume

$$\mathbb{P}\left[\text{one reaction between an } A \text{ and a } B \text{ molecule in } [t, t+dt]\right] \approx \frac{k}{\nu}dt + \mathcal{O}(dt^2),$$

so that the reaction rates scale appropriately with volume. The total number of possible A-B pairs that could react together at time t is $A(t)B(t)$ and hence

$$\mathbb{P}\left[\text{one reaction in } [t, t+dt]\right] \approx \frac{k}{\nu}A(t)B(t)dt + \mathcal{O}(dt^2).$$

A slightly different type of second order chemical reaction is the process

$$A + A \xrightarrow{k} C,$$
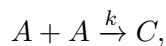
where two molecules of A join together by ions or bonds to form a new chemical species C, a process called **dimerisation**. In this case, the reaction rate of the system will depend on the number of possible A-A pairs, which is given by

$$\binom{A(t)}{2} = \frac{A(t)(A(t)-1)}{2}.$$

Thus the propensity function for the dimerisation process above is given by

$$\alpha(t) = \begin{cases} \frac{kA(t)(A(t)-1)}{2\nu}, & A(t) \geq 2, \\ 0, & A(t) < 2. \end{cases}$$

Table 1.1 below lists some simple reaction processes of different orders, along with their propensity functions ($\alpha(t)$) and the units of their reaction rates (k).

Table 1.1: Examples of some simple low order chemical reactions and their corresponding propensity functions.

Reaction	order	$\alpha(t)$	units of k
$\emptyset \xrightarrow{k} A$	zero	$k\nu$	$m^{-3} \sec^{-1}$
$A \xrightarrow{k} \emptyset$	first	$kA(t)$	$\sec^{-1}$
$A + B \xrightarrow{k} \emptyset$	second	$kA(t)B(t)/\nu$	$m^{-3} \sec^{-1}$
$A + B + C \xrightarrow{k} \emptyset$	third	$kA(t)B(t)C(t)/\nu^2$	$m^{-6} \sec^{-1}$

Implicit in all of these processes is the assumption that the molecules are **well-mixed** spatially, meaning that a molecule of any species is equally likely to encounter a molecule of any other species. Later in the course we will consider biological processes where this well-mixedness assumption is very strongly violated, such as in chemotaxis in cell biology (in which cells are attracted to/repulsed by other cells) or prey that move to avoid predators in ecological models.

Exercise 1.8

Write down the propensity functions for the following reaction processes:
1. $2A + B \xrightarrow{k} \emptyset$
2. $A + 2B + C \xrightarrow{k} \emptyset$
3. $2A + 3B \xrightarrow{k} \emptyset$

1.4 Stochastic Simulation of Dimerisation

The chemical master equations for complex higher-order reaction processes are typically very hard to solve analytically. In this section we will use an example of a dimerisation process to introduce the probability generating function approach to solving the chemical master equations. Dimerisation in chemistry is the process where two identical or similar molecules, known as monomers, join together to form a single, larger molecule called a dimer – we will use it as the simplest example of a second order chemical reaction process.

1.4.1 Probability Generating Functions

Before introducing the probability generating function (PGF) of a stochastic process, we may define the PGF of a random variable as follows:

For a discrete random variable X taking values in the set $\{x_0, x_1, \dots\}$, its **probability generating function** G is the function

$$G : [-1, 1] \mapsto \mathbb{R}, \quad G(x) = \sum_{n=0}^{\infty} x^n \mathbb{P}[X = x_n].$$

The PGF of a random variable contains all of the information about the distribution of the random variable. In particular, it is straightforward to show from the formula above that

$$\mathbb{P}[X = x_n] = \frac{1}{n!} G^{(n)}(x) \Big|_{x=0},$$

where $G^{(n)}(x)$ denotes the nth derivative of G with respect to x. This gives us a very direct way to recover the PMF of a random variable once we know its PGF. Similarly, the PGF and the moments of a random variable are related by the formula:

$$\mathbb{E}\left[\frac{X!}{(X-k)!}\right] = G^{(k)}(x) \Big|_{x=1}, \quad k \geq 0.$$

The formula above refers to the kth factorial moment of the random variable but from this we can deduce formulae for the mean and variance, which are given by

$$\mathbb{E}[X] = \frac{d}{dx} G(x) \Big|_{x=1}, \quad \mathrm{Var}[X] = \left(\frac{d^2}{dx^2} G(x) + \frac{d}{dx} G(x) - \left(\frac{d}{dx} G(x)\right)^2\right) \Big|_{x=1}.$$

For example, if we consider a Poisson distributed random variable Y with parameter $\lambda > 0$, its PGF is given by

$$G_Y(x) = \sum_{n=0}^{\infty} x^n \frac{\lambda^n e^{-\lambda}}{n!} = e^{-\lambda} \sum_{n=0}^{\infty} \frac{(x\lambda)^n}{n!} = e^{\lambda(x-1)}.$$

We can thereby calculate the mean of Y as suggested above:

$$\mathbb{E}[Y] = \frac{d}{dx} G_Y(x) \Big|_{x=1} = \lambda e^{\lambda(x-1)} \Big|_{x=1} = \lambda.$$

1.4.2 Dimerisation Process Analysis

Consider the second-order stochastic reaction process given by

$$A + A \xrightarrow{k_1} \emptyset, \quad \emptyset \xrightarrow{k_2} A.$$
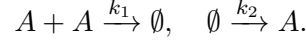
The propensity function of the process is thus given by

$$\alpha(t) = \frac{k_1}{\nu} A(t)(A(t) - 1) + k_2 \nu,$$

where ν denotes the system volume and we have absorbed the constant $1/2$ owing to the number of possible A pairs into the rate constant k_1. The more efficient SSA applied to the dimerisation process takes the following form:

_____

At time $t = 0$, set $A(0) = n_0$, then:

1. Generate random numbers r_1, $r_2 \sim U([0,1])$.
2. Compute the propensity function $\alpha(t)$.
3. Compute the next reaction time by calculating

$$\tau = \frac{1}{\alpha(t)} \log(1/r_1),$$

   and set $t = t + \tau$.
4. Compute the number of A molecules at time $t = t + \tau$:

$$A(t + \tau) = \begin{cases} A(t) + 1, & \text{if } r_2 < k_2 \nu / \alpha(t), \\ A(t) - 2, & \text{if } r_2 \geq k_2 \nu / \alpha(t), \end{cases}$$

   and go back to step 1.

_____

We defer simulating the process for now and instead proceed directly to its analysis via the chemical master equations. Adopting our usual approach, we see that

$$P_n(t + dt) = P_n(t) \left[1 - \frac{k_1}{\nu} n(n-1) dt - k_2 \nu dt \right]$$
$$+ \frac{k_1}{\nu}(n+1)(n+2) P_{n+2}(t) dt + k_2 \nu P_{n-1}(t) dt.$$

Rearranging and letting $dt \downarrow 0$, we obtain the system of ODEs:

$$\frac{d}{dt}P_n(t) = \frac{k_1}{\nu}(n+1)(n+2)P_{n+2}(t) - \frac{k_1}{\nu}n(n-1)P_n(t) + k_2\nu P_{n-1}(t) - k_2\nu P_n(t),$$

with the convention that $P_n \equiv 0$ for $n < 0$.

Clearly, it will not be particularly fun or easy to solve this system! Worse still, it is not possible to write a closed system of equations for the mean and variance of this process via manipulating the chemical master equations (as we did before). Instead we proceed via the probability generating function of the process:

$$G : [-1,1] \times (0,\infty) \mapsto \mathbb{R} : \quad G(x,t) = \sum_{n=0}^{\infty} x^n P_n(t).$$

As we can see, the PGF of a stochastic process is defined analogously to the PGF of a random variable, except now we have an extra argument to account for the time at which we are evaluating the process $A(t)$. Formulae for the PMF, mean and variance of $A(t)$ are exactly the same as those shown above. Therefore, if we can obtain the PGF of the process, it contains all of the same information as the PMF and is thus functionally equivalent to solving the chemical master equation!

To derive an evolution equation for the PGF of the process, multiply the master equation across by x^n and sum over n to show that:

$$\frac{\partial}{\partial t}\sum_{n=0}^{\infty} x^n P_n(t) = \frac{k_1}{\nu}\sum_{n=0}^{\infty} x^n(n+1)(n+2)P_{n+2}(t) - \frac{k_1}{\nu}\sum_{n=2}^{\infty} x^n n(n-1)P_n(t)$$

$$+ k_2\nu\sum_{n=0}^{\infty} x^n P_{n-1}(t) - k_2\nu\sum_{n=0}^{\infty} x^n P_n(t).$$

Next note the identity

$$\frac{\partial^2}{\partial x^2}G(x,t) = \sum_{n=2}^{\infty} n(n-1)x^{n-2}P_n(t),$$

which we will use to simplify the sums above. Changing the indices in the 1st and 3rd sums yields

$$\frac{\partial}{\partial t}G(x,t) = \frac{k_1}{\nu}\sum_{n=2}^{\infty}n(n-1)x^{n-2}P_n(t) - \frac{k_1}{\nu}x^2\sum_{n=2}^{\infty}n(n-1)x^{n-2}P_n(t)$$

$$+ k_2\nu x\sum_{n=0}^{\infty}x^n P_n(t) - k_2\nu\sum_{n=0}^{\infty}x^n P_n(t)$$

$$\implies \frac{\partial}{\partial t}G(x,t) = \frac{k_1}{\nu}(1-x^2)\frac{\partial^2}{\partial x^2}G(x,t) + k_2\nu(x-1)G(x,t).$$

Thus this is a second-order PDE for the PGF of the dimerisation process and if we can solve it, we will recover all of the information contained in the PMF of the process. To solve the PDE, we need to supply some initial and boundary conditions. Firstly,

$$G(x,0) = \sum_{n=0}^{\infty}x^n P_n(0), \quad x \in [-1,1]$$

so we can compute the initial condition for each x given the initial conditions of the process. We also need boundary conditions at $x = 1$ and $x = -1$. At $x = 1$, we have

$$G(1,t) = \sum_{n=0}^{\infty}P_n(t) = 1$$

and evaluating the PDE at $x = -1$ gives us

$$\frac{\partial}{\partial t}G(-1,t) = -2k_2\nu G(-1,t)$$

so that $G(-1,t) = G(-1,0)e^{-2k_2\nu t}$. These conditions plus the PDE thus constitute a well-posed initial-value problem that could be solved analytically or numerically, although this is not straightforward either!

Instead of trying to solve the PDE directly, we will analyse the asymptotic behaviour of the process via the **stationary probability generating function** G_s, which is given by:

$$G_s : [-1,1] \mapsto \mathbb{R}, \quad G_s(x) = \lim_{t\to\infty}G(x,t) = \sum_{n=0}^{\infty}x^n\phi(n),$$

where ϕ is the stationary distribution of the process. We can then obtain the asymptotic mean, $M_s := \lim_{t\to\infty}M(t)$, the asymptotic variance, $V_s := \lim_{t\to\infty}V(t)$, and the stationary distribution, ϕ, via the stationary PGF. Since the stationary PGF, G_s, does not depend on t, the PDE becomes

$$0 = \frac{k_1}{\nu}(1 - x^2)\frac{d^2}{dx^2}G_s(x) + k_2\nu(x - 1)G_s(x),$$

which simplifies to the second order ODE

$$G_s''(x) = \frac{k_2\nu^2}{k_1}\frac{1}{1 + x}G_s(x), \quad x \in (-1, 1).$$

The general solution of this ODE can be written as

$$G_s(x) = C_1\sqrt{1 + x}I_1\left(2\sqrt{\frac{k_2\nu^2(1 + x)}{k_1}}\right) + C_2\sqrt{1 + x}K_1\left(2\sqrt{\frac{k_2\nu^2(1 + x)}{k_1}}\right),$$

where C_1, C_2 are constants and the modified Bessel functions I_1 and K_1 are two independent solutions to the equation

$$z^2I_n''(z) + zI_n'(z) - (z^2 + n^2)I_n(z) = 0.$$

Since we are now studying the stationary PGF, we use our old boundary conditions as $t \to \infty$, i.e.

$$\lim_{t\to\infty} G(1, t) = 1, \quad \lim_{t\to\infty} G(-1, t) = \lim_{t\to\infty} G(-1, 0)e^{-2k_2\nu t} = 0.$$

The functions I_1 and K_1 obey

$$I_1(z) \sim \frac{z}{2} \text{ as } z \downarrow 0, \quad K_1(z) \sim \frac{1}{z} \text{ as } z \downarrow 0$$

and combining this with the boundary conditions allows us to deduce that $C_2 = 0$ and

$$C_1 = \left[\sqrt{2}I_1\left(2\sqrt{\frac{2k_2\nu^2}{k_1}}\right)\right]^{-1}.$$

We thus have an explicit formula for $G_s(x)$ in terms of the modified Bessel function I_1:

$$G_s(x) = \sqrt{1 + x}\,I_1\left(2\sqrt{\frac{k_2\nu^2(1 + x)}{k_1}}\right)\left[\sqrt{2}I_1\left(2\sqrt{\frac{2k_2\nu^2}{k_1}}\right)\right]^{-1}, \quad x \in [-1, 1].$$

This formula can be evaluated numerically as the function I_1 is implemented in most mathematical software (such as in MATLAB or appropriate Python packages).

1.4.3 Simulation of the Dimerisation Process

In this section, we summarise and conclude our analysis of the dimerisation process. We have a number of tools and approaches we can use to understand the dynamics of this process, including:

- **direct simulation** (using the more efficient SSA),
- **analytic approaches**: chemical master equation, PGF, stationary PGF, etc.
- **deterministic modelling** via the law of mass action.

In order to compare all of these tools in a unified analysis, we need to apply the law of mass action to develop a deterministic model of the dimerisation process. This yields the **nonlinear ODE**:

$$\frac{d}{dt}a(t) = -2k_1 a(t)^2 + k_2,$$

where $a(t) = A(t)/\nu$ is the concentration of A molecules. We can solve this ODE numerically or solve the steady-state version to see that it predicts the long-run mean of $A(t)$ will be $\nu\sqrt{k_2/2k_1}$.

Figure 1.5 presents a synthesis of the three forms of analysis of the dimerisation process outlined above. In the left panel, we show sample paths of the process, together with the long-run mean obtained via analytic calculations (red dashed line) and the solution to the ODE model given by law of mass action (dashed black line). The law of mass action predicts a slightly lower mean value than the long-run mean, M_s, and these quantities differ in general for this model (which we will show in problem classes). We also plot $M_s \pm 2\sqrt{V_s}$ in the left-hand panel to show that these bounds give a good idea of the fluctuations of the process around its long-run mean. This works because Chebyshev's inequality tells us that any distribution with a finite mean and variance has approximately 75% of its mass within two standard deviations of either side of the mean (where the standard deviation here is $\sqrt{V_s}$).

The right-hand panel of Figure 1.5 shows the estimated PMF of the dimerisation process at $t = 100$ using $1,000$ sample paths of the process obtained via the more efficient SSA. The solid red line denoted by "analysis" in the legend is the stationary distribution of the process calculated from an analytic formula based on the calculations of the preceding section (see problem classes for more details!).

Exercise 1.9

Open the course Github page and try playing with the MATLAB script:
`CH1_dimerisation_process.m`
Try varying the parameters (time interval, initial conditions, reaction rates) to observe how the analysis presented above changes.
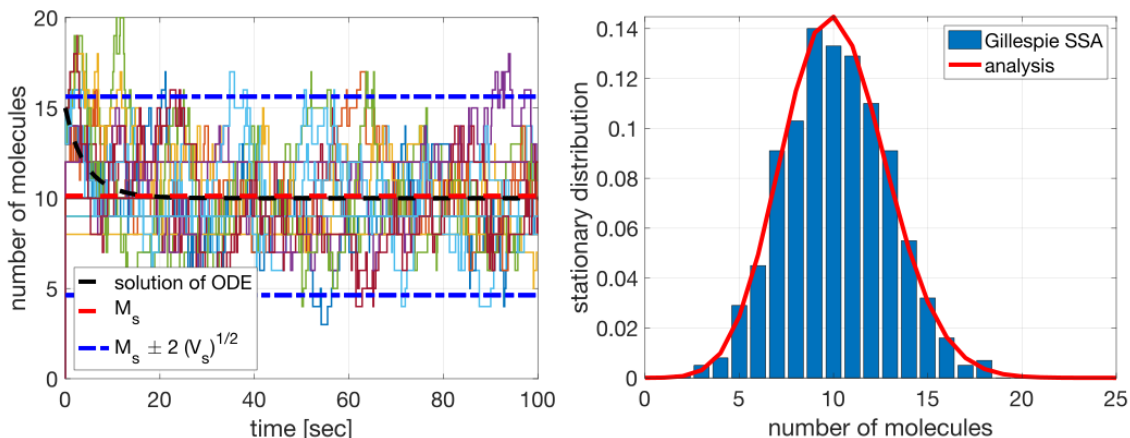
Figure 1.5: Left: Sample paths compared to the solution of the mass action ODE model and to the long run mean and variance. Right: Comparison between the stationary distributed obtained analytically and the estimated PMF from simulations. Parameters: $A(0) = 15$, $k_1 = 0.005$, $k_2 = 1$.

1.5 The Gillespie Stochastic Simulation Algorithm

1.5.1 Gillespie SSA Formulation

The "more efficient SSA" that we introduced in this chapter is a special case of the so-called **Gillespie algorithm** for simulating Markov jump processes. The Gillespie algorithm is a classic example of Stigler's law of eponymy, i.e. the "rule" that scientific discoveries are not named after the person who discovered them. Physicist Dan Gillespie popularised the algorithm that now bears his name in a 1977 paper concerning the simulation of chemical systems [@gillespie1977], but the algorithm was known some 35 years prior. The Gillespie algorithm was first implemented in 1950 by English mathematician and statistician David George Kendall on the Manchester Mark 1 computer.

In this section, we formulate the Gillespie algorithm for a general stochastic reaction process with $q \geq 1$ chemical reactions. Let $\alpha_i(t)$ denote the propensity function of the ith reaction. Note that we don't need to specify the number of chemical species involved; what matters is the number of distinct possible reactions.

At time $t = 0$, set the initial number of molecules in each species, then:

1. Generate random numbers r_1, $r_2 \sim U([0,1])$.

2. Compute the propensity function of the system:

$$\alpha_0(t) = \sum_{i=1}^{q} \alpha_i(t).$$

3. Compute the next reaction time by calculating

$$\tau = \frac{1}{\alpha_0(t)} \log(1/r_1),$$

and set $t = t + \tau$.

4. Figure out which of the q reactions took place by finding the integer j such that:

$$r_2 \geq \frac{1}{\alpha_0(t)} \sum_{i=1}^{j-1} \alpha_i(t), \quad r_2 < \frac{1}{\alpha_0(t)} \sum_{i=1}^{j} \alpha_i(t).$$

Carry out reaction j, i.e. adjust the number of molecules in each species to account for reaction j occurring at time $t = t + \tau$. Go back to step 1.

The justification for this algorithm's correctness is a natural generalisation of the rationale we provided for the two species case. Step 3 is asserting that the next reaction time is exponentially distributed with parameter $\alpha_0(t)$, where $\alpha_0(t)$ is the sum over all of the q individual reaction rates. To see that this is correct, we can once again note that conditional on the current state of the system, the waiting time until the ith reaction occurs is $\tau_i \sim \text{Exp}(\alpha_i(t))$ and each pair of waiting times τ_i, τ_j are (conditionally) independent for $i \neq j$. Hence,

$$\tau = \min\left(\tau_1, \tau_2, \ldots, \tau_q\right),$$

where τ is the waiting time for the entire system. Using the mutual independence of the individual waiting times, we have

$$\mathbb{P}[\tau > x] = \mathbb{P}[\tau_1 > x, \ldots, \tau_q > x] = \prod_{i=1}^{q} \mathbb{P}[\tau_i > x] = e^{-x \sum_{i=1}^{q} \alpha_i(t)} = e^{-\alpha_0(t)x}.$$

Therefore, $\tau \sim \text{Exp}(\alpha_0(t))$, as claimed.

Step 4 ensures that we have

$$\mathbb{P}[\text{reaction } i \text{ occurs} \mid \text{some reaction occurs}] = \frac{\alpha_i(t)}{\alpha_0(t)},$$

which can be verified as the correct probability by computing $\mathbb{P}[\min(\tau_1, \ldots, \tau_q) = \tau_i]$. Intuitively, we can think of the condition in step 4 as breaking up the interval $[0, 1]$ into q subintervals

35

where the length of subinterval j is proportional to $\alpha_j(t)$. We then draw $r_2 \sim U([0,1])$ and see which subinterval r_2 falls in; if r_2 falls in subinterval j, then reaction j takes place.

There are many ways to improve upon the algorithm outlined above and this can be especially important for efficiently simulating processes that have a large number of reactions per unit time. Some improvements exploit special structure in the processes we are simulating, other improvements, such as the tau-leaping method [@cao2006efficient], are more broadly applicable. For example, we only need to update the propensity functions $\alpha_i(t)$ which changed the last time a reaction took place. This often means that most propensity functions do not need to be changed after every reaction and offers significant speed up for systems with large numbers of reactions.

1.5.2 Gillespie SSA Example

Consider the following stochastic reaction process involving dimerisation and production:

$$A + A \xrightarrow{k_1} \emptyset, \quad A + B \xrightarrow{k_2} \emptyset, \quad \emptyset \xrightarrow{k_3} A, \quad \emptyset \xrightarrow{k_4} B.$$
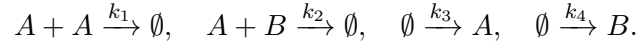
We will use this process to illustrate some of the difficulties that we typically encounter when considering more complex reactions with multiple species and multiple higher-order reactions; both of these features are naturally present in most practical problems of interest.

The propensity functions for the reaction process above are:

$$\alpha_1(t) = \frac{k_1}{\nu} A(t)(A(t) - 1), \qquad \alpha_2(t) = \frac{k_2}{\nu} A(t)B(t),$$
$$\alpha_3(t) = k_3\nu, \qquad\qquad\qquad \alpha_4(t) = k_4\nu.$$

To simulate the process, we need to compute the total propensity function

$$\alpha_0(t) = \alpha_1(t) + \alpha_2(t) + \alpha_3(t) + \alpha_4(t),$$

in order to compute the next reaction time. If τ is the next reaction time, then the step to update the number of A and B molecules according to the Gillespie algorithm will be:

$$A(t + \tau) = \begin{cases} A(t) - 2, & 0 \leq r_2 < \alpha_1/\alpha_0, \\ A(t) - 1, & \alpha_1/\alpha_0 \leq r_2 < (\alpha_1 + \alpha_2)/\alpha_0, \\ A(t) + 1, & (\alpha_1 + \alpha_2)/\alpha_0 \leq r_2 < (\alpha_1 + \alpha_2 + \alpha_3)/\alpha_0, \\ A(t), & \text{else,} \end{cases}$$

and

$$
B(t+\tau) = \begin{cases} B(t), & 0 \leq r_2 < \alpha_1/\alpha_0, \\ B(t) - 1, & \alpha_1/\alpha_0 \leq r_2 < (\alpha_1 + \alpha_2)/\alpha_0, \\ B(t), & (\alpha_1 + \alpha_2)/\alpha_0 \leq r_2 < (\alpha_1 + \alpha_2 + \alpha_3)/\alpha_0, \\ B(t) + 1, & \text{else.} \end{cases}
$$

At this point we could simulate the process directly via our SSA but first we will consider our other typical lines of attack: the chemical master equations and the law of mass action.

As the number of reactions and species increases, so does the complexity of the chemical master equations. In particular, we now need to consider the joint density of the species. To this end, define

$$
P_{n,m}(t) = \mathbb{P}[A(t) = n, \ B(t) = m].
$$

Proceeding in the usual way, we may write

$$
\begin{aligned}
P_{n,m}(t + dt) = & \left[1 - \frac{k_1}{\nu} n(n-1)dt - \frac{k_2}{\nu} nm dt - k_3 \nu dt - k_4 \nu dt \right] P_{n,m} \\
& + \frac{k_1}{\nu}(n+2)(n+1)P_{n+2,m}dt + \frac{k_2}{\nu}(n+1)(m+1)P_{n+1,m+1}dt \\
& + k_3 \nu P_{n-1,m}dt + k_4 \nu P_{n,m-1}dt,
\end{aligned}
$$

where we have suppressed the t arguments on the right-hand side. Letting $dt \downarrow 0$, the chemical master equations are thus given by

$$
\begin{aligned}
\frac{d}{dt}P_{n,m}(t) = & \frac{k_1}{\nu}(n+2)(n+1)P_{n+2,m} + \frac{k_2}{\nu}(n+1)(m+1)P_{n+1,m+1} \\
& + k_3 \nu P_{n-1,m} + k_4 \nu P_{n,m-1} - \frac{k_1}{\nu}n(n-1)P_{n,m} \\
& - \frac{k_2}{\nu}nm P_{n,m} - k_3 \nu P_{n,m} - k_4 \nu P_{n,m},
\end{aligned}
$$

for $n, m \geq 0$, with the standard convention that $P_{n,m} \equiv 0$ if either $n < 0$ or $m < 0$.

Due to the presence of the second-order reactions, we cannot even write a closed system of evolution equations for the mean and variance via the chemical master equations! We could solve these equations, or their steady-state analogue, numerically if we wanted detailed information on the distribution or stationary distribution.

Finally, we can employ the law of mass action to write an approximate deterministic model for the mean behaviour of the process. Letting $a(t) = A(t)/\nu$ and $b(t) = B(t)/\nu$, we obtain the following pair of nonlinear ODEs:

$$\frac{d}{dt}a = -2k_1a^2 - k_2ab + k_3,$$

$$\frac{d}{dt}b = -k_2ab + k_4.$$

Figure 1.6 below shows some sample paths of the process with the corresponding solutions to the deterministic model overlaid for comparison (dashed black lines).
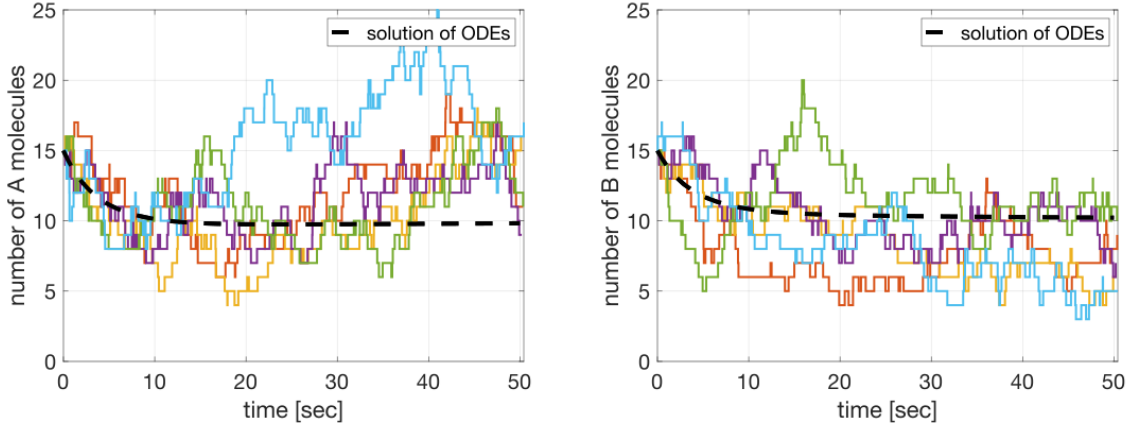
Figure 1.6: Sample paths and the deterministic mean predicted by the law of mass action. Parameters: $k_1 = 0.001$, $k_2 = 0.01$, $k_3 = 1.2$, $k_4 = 1$.

We can gain more information about the dynamics of the process by running more and longer simulations. This allows us to estimate the stationary distribution which characterizes the asymptotic behaviour of the process. Figure 1.7 shows the estimated stationary distribution in the left-hand panel, i.e. the joint PMF of $A(t)$ and $B(t)$ as $t \to \infty$. The right-hand panel of Figure 1.7 shows the marginal stationary distribution of $A(t)$, which can be obtained from the joint stationary distribution via the formula

$$\phi(n) = \sum_{m=0}^{\infty} \phi(n, m),$$

where $\phi(n) := \lim_{t\to\infty} \mathbb{P}[A(t) = n]$ and $\phi(n, m) := \lim_{t\to\infty} \mathbb{P}[A(t) = n,\ B(t) = m]$.

Knowledge checklist

Key topics:

1. Chemical reaction processes (continuous time jump processes)
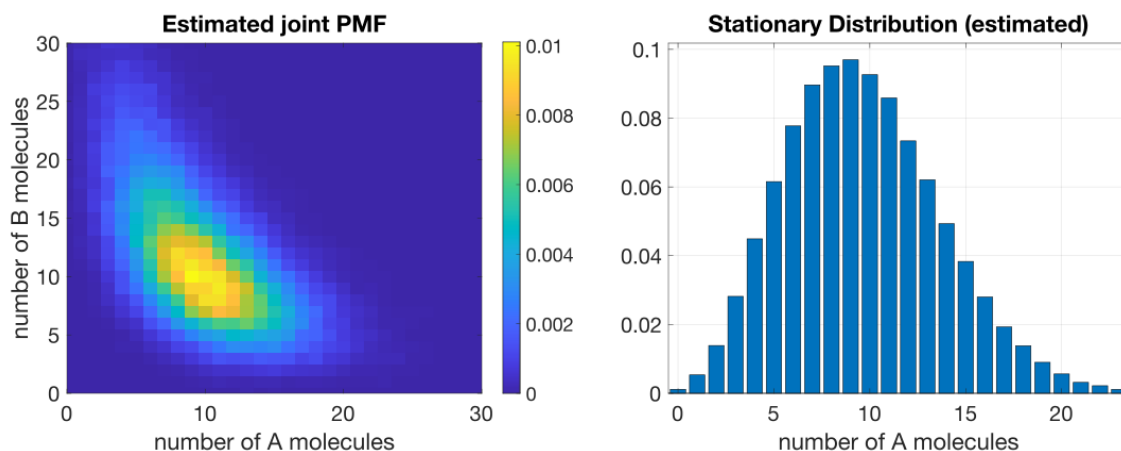
Figure 1.7: Estimated stationary distributions. Parameters: $k_1 = 0.001$, $k_2 = 0.01$, $k_3 = 1.2$, $k_4 = 1$.

2. The law of mass action (approximation)
3. "Naive" & Gillespie SSAs; the inverse transform method
4. Chemical master equations (CMEs)
5. Probability generating functions
6. Stationary distributions

Key skills:

- Define and interpret chemical reaction processes using either:

  – The reaction notation (e.g. $A + B \to C$), or
  – the leading order probabilistic formulations
    (e.g. $\mathbb{P}[A(t + dt) = n] \approx f(A(t))dt + O(dt^2)$).

- Write and analyse mass action approximations of chemical reaction processes
- Write and analyse pseudocode for SSAs (Gillespie and the "naive" SSA)

  – Given a chemical reaction process, write the propensity functions for each reaction (especially for higher order reactions)
  – Comment on the efficiency and justify the correctness of SSAs

- Write and analyse chemical master equations, i.e.

  – computing stationary distributions from CMEs
  – calculating and solving moment equations
  – compute/analyse probability generating functions

- Describe the differences between and advantages/disadvantages of different approaches to studying a given chemical reaction process (i.e. mass action versus SSAs versus analytic approaches).

39

References

2 Deterministic vs Stochastic Models

The goal of this chapter is to build a deeper understanding of the qualitative differences that are possible between deterministic and stochastic models of the same system. We saw in Chapter 1 that the law of mass action can give good predictions of the average behaviour for zero and first order reactions. Even the mean behaviour of higher-order reactions is not perfectly predicted by the analogous deterministic model, although the differences were not pronounced in the examples we have seen thus far. The examples we consider presently show that there can be dramatic differences in the qualitative behaviour of a stochastic reaction process and its deterministic mass action model. Moreover, these novel stochastic dynamics form the basis of current modelling work for a myriad of real-world systems, ranging from biological applications such as cell-fate dynamics and neuroscience, to ecological applications like coral-reef stability and vegetation dynamics [@fung2011alternative;@moris2016transition;@staver2011global].

2.1 Multistable Systems

Consider the following stochastic reaction process involving a single chemical species A:

$$3A \xrightarrow{k_1} 2A, \qquad 2A \xrightarrow{k_2} 3A, \qquad A \xrightarrow{k_3} \emptyset, \qquad \emptyset \xrightarrow{k_4} A. \qquad (2.1)$$
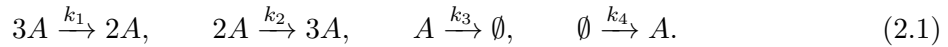
This process involves production, degradation, and several dimerisation-type reactions. If we apply the law of mass action to this process, we can derive a deterministic model for the average behaviour of the concentration of species A, i.e., $a(t) = A(t)/\nu$ where ν is the system volume.

Invoking the law of mass action, multiplying the reaction rates by the concentrations for each reaction, we arrive at the following (approximate) deterministic model for the process:

$$\frac{d}{dt}a(t) = -k_1 a(t)^3 + k_2 a(t)^2 - k_3 a(t) + k_4, \quad t \geq 0. \qquad (2.2)$$

We can solve for the equilibrium solutions of this model by solving the cubic polynomial

$$0 = -k_1 a^3 + k_2 a^2 - k_3 a + k_4 =: f(a),$$

and we can readily evaluate their stability by computing the sign of the "Jacobian", $f'(a)$, of the system at each equilibrium point. The number of steady states and their stability can vary with the system parameters, but, as an example of the possible dynamics, take

$$k_1 = 0.00025, \qquad k_2 = 0.18, \qquad k_3 = 37.5, \qquad k_4 = 2200.$$

In this parameter regime, we find three steady states:

$$\bar{A}_1 = 100, \qquad \bar{A}_2 = 220, \qquad \bar{A}_3 = 400.$$

Linear stability analysis tells us that $\bar{A}_1$ and $\bar{A}_3$ are stable, while $\bar{A}_2$ is unstable. Since the ODE above is one dimensional, $f(0) > 0$, and $f(a) < 0$ for all $a > 0$ sufficiently large, all trajectories must end up at either $\bar{A}_1$ or $\bar{A}_3$ as $t \to \infty$. Trajectories cannot cross the (unstable) solution, $\bar{A}_2$, and thus $\lim_{t \to \infty} A(t) = \bar{A}_1$ if $A(0) < \bar{A}_2$ and $\lim_{t \to \infty} A(t) = \bar{A}_3$ if $A(0) > \bar{A}_2$. Our qualitative analysis of the dynamics is illustrated in the left-hand panel of Figure 2.1 where we plot trajectories against time for different initial conditions with the equilibrium solutions denoted by the dashed black lines.

The right-hand panel of Figure 2.1 is a one-parameter bifurcation diagram of the system above where we have allowed the production rate parameter, k_3, to vary. The values of k_1, k_2, and k_4 are fixed to the same values as above. As k_3 varies, we track the number of solutions and their stability; stable solutions are in red, while unstable solutions are in black. We observe that the system has two stable solutions for k_3 between 36 and 38.9 approximately. In this **bistable regime**, the dynamics are qualitatively as described above (where we had $k_3 = 37.5$). The bistable regime begins and ends with **saddle-node bifurcations** and the system has only one stable solution before/after these bifurcation points.
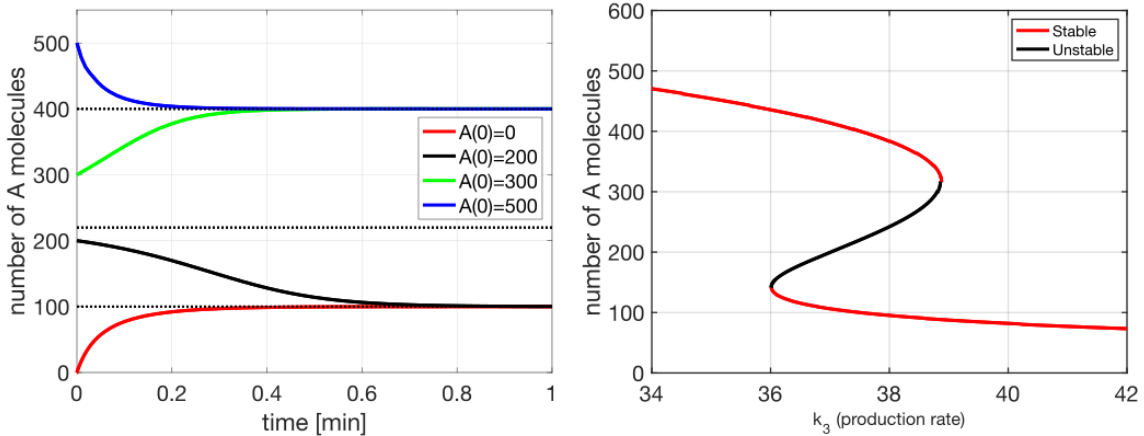
Figure 2.1: **Left:** Solutions of the deterministic model for different initial conditions. **Right:** One parameter bifurcation diagram for the deterministic model varying the production rate parameter k_3. Parameters: $k_1 = 0.00025$, $k_2 = 0.18$, $k_3 = 37.5$ (left), $k_4 = 2200$.

Bifurcation diagrams with the same qualitative structure as in Figure 2.1 are found in many real-world applications and may have very important implications in a given applied scenario.

For example, suppose that the upper stable (red) branch represents a state with high tree cover and the lower stable branch represents a state with low tree cover. We would care enormously if an external process (e.g., climate change) increased the bifurcation parameter past the upper saddle-node bifurcation ($k_3 \approx 38.9$) and caused the system to suddenly drop to a much lower tree cover state! We can imagine similar potential catastrophes if the y-axis represented the abundance of a certain species. **However, if the parameter k_3 does not change value, then the deterministic model predicts either one steady-state solution or the other; there can be no switching between the alternative stable states.**

Exercise 2.1

Write down the propensity function for the process Equation 2.1 and hence sketch the pseudocode for the Gillespie SSA to simulate the process. What is the most efficient way to write the conditions for updating the number of A molecules?

We begin our analysis of the stochastic reaction process by simulating the system using the Gillespie algorithm. Figure 2.2 shows some sample paths of the process (blue) and the corresponding trajectory of the deterministic model Equation 2.2 (red). In the left-hand panel of Figure 2.2, we observe reasonable agreement between the respective solutions as we only consider a very short time interval. However, in the right-hand panel, the stochastic solutions show qualitatively different behaviour once we run the process for long enough. In particular, the blue sample paths cross the unstable solution of the deterministic model ($\bar{A}_2$, middle black dashed line) and switch between spending time near each of the stable solutions of the deterministic model ($\bar{A}_1$ and $\bar{A}_3$). **This switching behaviour is impossible in the deterministic model and is an entirely new phenomenon introduced by the stochasticity of the process** Equation 2.1.
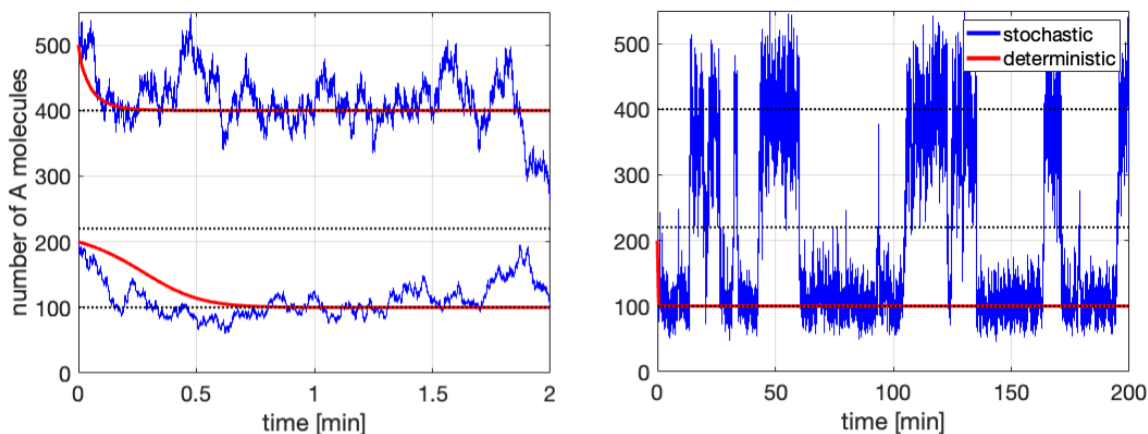

Figure 2.2: Comparison between deterministic model trajectories and simulations of the process Equation 2.1 on short and long time intervals. Parameters: $k_1 = 0.00025$, $k_2 = 0.18$, $k_3 = 37.5$, $k_4 = 2200$.

From an applied perspective, this **noise-induced (attractor) switching** introduces a new possibility when studying systems with alternative stable states. Even if the parameters of the system do not change, the system may spontaneously shift from one stable state to another due to stochastic fluctuations! The **mean switching time** is the average time the system spends in one stable state or the other. In general, it can depend on which state the system starts in as it may be easier to switch in one direction compared to the reverse transition. The mean switching time can give us a measure of how stable the system is to random perturbations and is often used as a measure of a system's stability or "resilience". The mean switching time can be estimated from simulations of a model and there are also analytic approaches for its computation that we will study later in the course.

Exercise 2.2

Open the course Github page and try playing with the MATLAB script

`CH2_bistable_process.m`

to see how the switching frequency varies as you change system parameters.

We can study the chemical master equations of the process Equation 2.1 to gain more insight into the nature of this switching phenomenon and to understand how much time the system is expected to spend in each stable state.

If we consider $dt > 0$ sufficiently small, we can reason in the usual way that the change in the PMF $P_n(t) := \mathbb{P}[A(t) = n]$ over the interval $[t, t + dt)$ is given by

$$
\begin{aligned}
P_n(t + dt) = &\left(1 - \frac{k_1}{\nu^2}n(n-1)(n-2)dt - \frac{k_2}{\nu}n(n-1)dt - k_3 n dt - k_4 \nu dt\right) P_n(t) \\
&+ \frac{k_1}{\nu^2}(n+1)n(n-1)P_{n+1}(t)\, dt + \frac{k_2}{\nu}(n-1)(n-2)P_{n-1}(t)\, dt \\
&+ k_3(n+1)P_{n+1}(t)\, dt + k_4 \nu P_{n-1}(t)\, dt.
\end{aligned}
$$

Rearranging and letting $dt \downarrow 0$ thus yields the chemical master equations:

$$
\begin{aligned}
\frac{d}{dt}P_n = &\frac{k_1}{\nu^2}(n+1)n(n-1)P_{n+1} + \frac{k_2}{\nu}(n-1)(n-2)P_{n-1} + k_3(n+1)P_{n+1} + k_4\nu P_{n-1} \\
&- \frac{k_1}{\nu^2}n(n-1)(n-2)P_n - \frac{k_2}{\nu}n(n-1)P_n - k_3 n P_n - k_4 \nu P_n, \quad n \ge 0,
\end{aligned}
\tag{2.3}
$$

where we have suppressed the t dependence in P_n for brevity. We also adopt our usual convention that $P_n \equiv 0$ for all $n < 0$.

We will now consider the asymptotic behaviour of the process and thus proceed to the steady state version of the CMEs in order to compute the stationary distribution, ϕ. The steady-state version of Equation 2.3 is given by:

$$
0 = \frac{k_1}{\nu^2}(n+1)n(n-1)\phi(n+1) + \frac{k_2}{\nu}(n-1)(n-2)\phi(n-1) + k_3(n+1)\phi(n+1)
$$
$$
+ k_4\nu\phi(n-1) - \frac{k_1}{\nu^2}n(n-1)(n-2)\phi(n) - \frac{k_2}{\nu}n(n-1)\phi(n) - k_3 n\phi(n) - k_4\nu\phi(n),
$$

for $n \geq 0$ and $\phi(n) = 0$ for $n < 0$. Recursively solving these equations, we can write the solution in terms of $\phi(0)$ as

$$
\phi(n) = \phi(0) \prod_{i=0}^{n-1} \frac{(k_2/\nu)i(i-1) + k_4\nu}{(k_1/\nu^2)(i+1)i(i-1) + k_3(i+1)}, \quad n \geq 1. \tag{2.4}
$$

We know that $\sum_{n=0}^{\infty} \phi(n) = 1$, so we can calculate $\phi(n)$ for a large range of values of n and then normalize to find the stationary distribution in practice.

Figure 2.3 shows the results from using 10,000 simulations of the process to estimate the PMF at a large time compared to evaluating the formula Equation 2.4 for the stationary distribution. We see excellent agreement between the analytic and direct simulation approaches with both showing a distinctly bimodal distribution. The peaks of the distribution are centred on the stable states of the deterministic model Equation 2.2, i.e. $\bar{A}_1 = 100$ and $\bar{A}_3 = 400$, with a local minimum between these values around 220, the value of $\bar{A}_2$. Moreover, the peak at $A = 100$ is 8 times higher than the peak at $A = 400$, suggesting that the system will tend to spend much more time close to $A = 100$ in this parameter regime.

We will return to the switching time problem later in the course and develop further tools to estimate the time a stochastic system spends near a stable state before switching to another.

2.2 Self-induced Stochastic Resonance

In this section, we will use a relatively simple chemical reaction process to illustrate another distinct kind of novel noise-induced dynamics known as **stochastic resonance**. Consider the two-species reaction process given by

$$
2A + B \xrightarrow{k_1} 3A, \quad \emptyset \xrightarrow{k_2} A, \quad A \xrightarrow{k_3} \emptyset, \quad \emptyset \xrightarrow{k_4} B. \tag{2.5}
$$
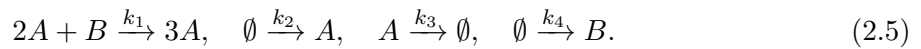
Letting $a(t) = A(t)/\nu$ and $b(t) = B(t)/\nu$, we can use the law of mass action to write down an approximate deterministic description of the average behaviour of the process Equation 2.5. Our deterministic approximation is given by
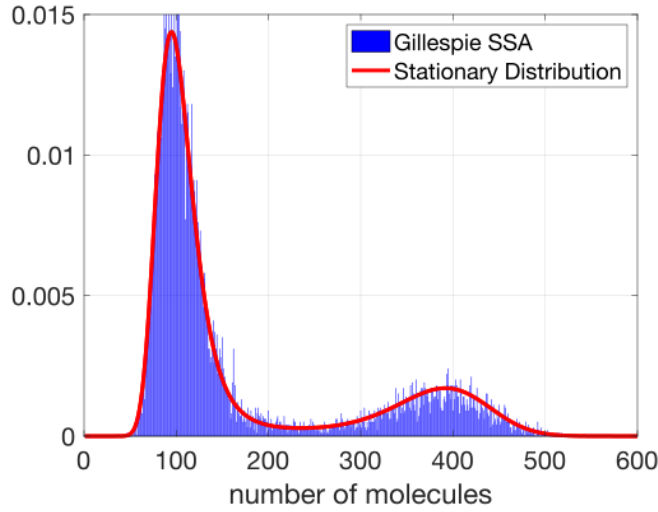
Figure 2.3: Comparison of the stationary distribution and the estimated long term behaviour from 10,000 simulations of the process. Parameters: $k_1 = 0.00025$, $k_2 = 0.18$, $k_3 = 37.5$, $k_4 = 2200$.

$$\frac{d}{dt}a(t) = k_1 a(t)^2 b(t) + k_2 - k_3 a(t),$$
$$\frac{d}{dt}b(t) = -k_1 a(t)^2 b(t) + k_4. \tag{2.6}$$

Exercise 2.3

Write down the pseudocode for the Gillespie SSA to simulate the process Equation 2.5.

We will proceed to simulate the process Equation 2.5 and compare the resulting dynamics with those predicted by the approximate deterministic model Equation 2.6. For the purposes of comparing dynamics, we choose the parameter set

$$k_1 = 0.0004, \qquad k_2 = 50, \qquad k_3 = 10, \qquad k_4 = 25.$$

Figure 2.4 A shows the evolution of the number of A molecules in both the stochastic and deterministic models; we immediately notice that the two models exhibit qualitatively different dynamics. The deterministic model quickly tends to a steady-state, while the stochastic model shows a reasonably regular pattern of oscillations (with some irregularity due to the stochasticity of the process). Figure 2.4 B shows both the A and B molecule evolutions against time and shows regular oscillations in both molecule numbers, with much more abrupt spikes in species

A compared to the more gradual spikes for species *B*. What is the source of this dramatic disagreement between the stochastic and deterministic dynamics?
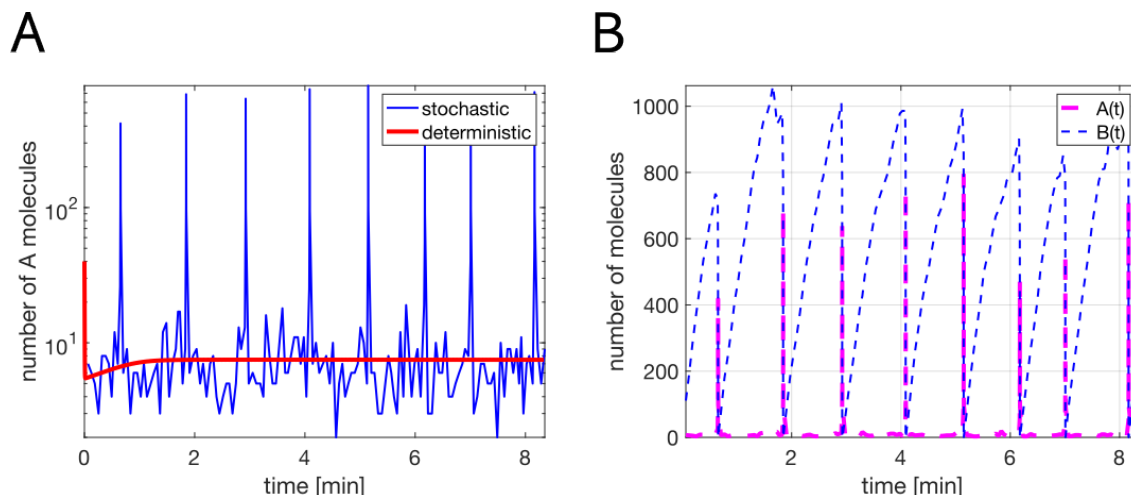
Figure 2.4: A: Comparison of the stochastic dynamics given by Equation 2.5 and deterministic dynamics of the ODEs given by Equation 2.6. B: Time evolutions of the number of *A* and *B* molecules for a realisation of the stochastic process Equation 2.5.

To understand the divergence in qualitative behaviour between the stochastic and deterministic dynamics, we plot the trajectories of both models in the phase space in Figure 2.5. Figure 2.5 A and B show the trajectories of the process Equation 2.5 in blue, with a single path in panel A and multiple paths in panel B. In both panels, the deterministic solution is plotted in red and it approaches the fixed point where the nullclines of the system intersect (green curves). The parabolic-shaped green curve is the *a* nullcline of Equation 2.6 (i.e., $da/dt = 0$), while the almost vertical green line is the *b* nullcline ($db/dt = 0$). In Figure 2.5 C, we plot the direction field of the deterministic system Equation 2.6, along with several solutions of the deterministic system. Blue ticks mark the start of each deterministic solution, with a red tick marking the endpoint of the trajectory. Solutions starting left of the *a* nullcline proceed directly to the fixed point, but trajectories starting to the right of this nullcline undergo a long sojourn to the right, before eventually coming back to the fixed point. **There is only one fixed point of the deterministic system in this parameter regime and it is globally stable**, meaning all trajectories are attracted to it for any initial condition.

The stochastic trajectories follow the red deterministic trajectory up the left-hand part of the *a* nullcline, but when the blue trajectory crosses to the right of the *a* nullcline, it is then sent on a long excursion to the right; this long excursion to the right corresponds to the spike in the number of *A* molecules (and the rapid drop in the number of *B* molecules) observed in Figure 2.4. Figure 2.5 B suggests that the process crosses the *a* nullcline with higher probability the higher the number of *B* molecules, but in principle this crossing could occur anywhere along the nullcline with a large enough stochastic fluctuation. This spontaneous crossing of the
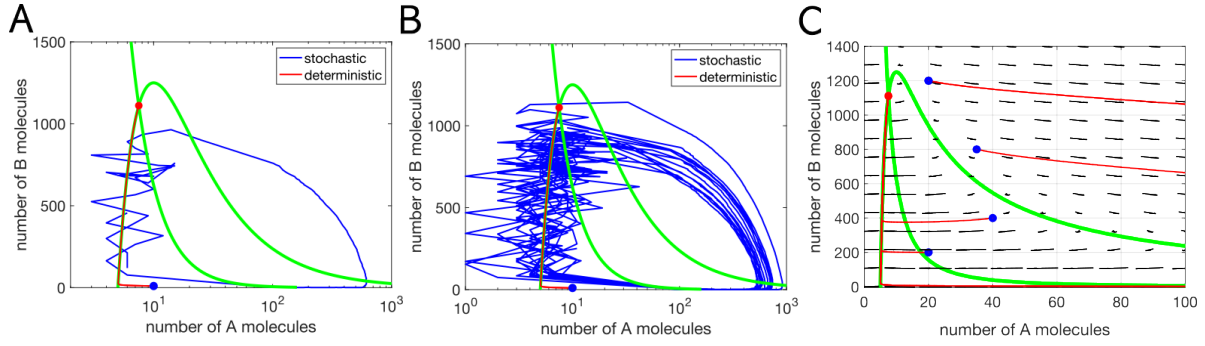
Figure 2.5: A/B: One/multiple trajectories of the process Equation 2.5 plotted in the phase space. C: Direction field and solutions to Equation 2.6 plotted in the phase space. [$k_4 = 25$.]

nullcline is impossible in the deterministic model. We call this phenomenon of noise-induced oscillations **self-induced stochastic resonance**.

If we increase the parameter k_4 from 25 to 100, the dynamics of the deterministic model change and enter an oscillatory regime. Figure 2.6 shows a comparison of the stochastic and deterministic dynamics in this regime. We now observe qualitative agreement between the dynamics, although Figure 2.6 shows that the period of the oscillations is not in exact agreement. The key change here is that increasing k_4 shifts the b nullcline, and hence the fixed point, to the right. Note that the a nullcline does not depend on k_4. Now all deterministic trajectories are sent around the fixed point and take the long excursion to the right, resulting in a stable periodic solution akin to what we observed previously for the stochastic trajectories with $k_4 = 25$.
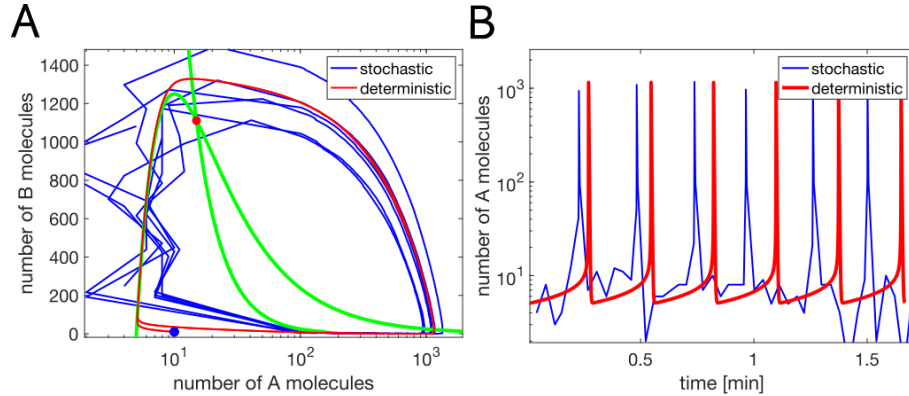
Figure 2.6: A: Comparison of deterministic and stochastic trajectories in the phase space. B: Time evolution of the number of A molecules in the stochastic and deterministic models. [$k_4 = 100$.]

As we increased k_4 from 25 to 100, we passed through a bifurcation point of the ODE

48

Equation 2.6. At this bifurcation, the fixed point where the nullclines intersect became unstable and stable periodic solutions emerged. Thus, another interpretation of the stochastic resonance phenomenon we observed in this model is that the stochastic fluctuations in the process Equation 2.5 brought about the onset of oscillations earlier (i.e., for a lower value of k_4) than predicted by the deterministic model.

The example presented above illustrates that oscillations can be triggered by a noisy process pushing a system over a threshold. In fact, this situation is observed in an array of applications in mathematical biology. For example, this motif is the basis of intense study in neuroscience, where the threshold represents the spiking threshold for a neuron to fire and transmit information to other neurons across the brain. Two of the most well-known models in neuroscience, the Hodgkin-Huxley model and (its simpler phenomenological counterpart) the Fitzhugh-Nagumo model, both exhibit stochastic resonance [@longtin1993stochastic]. Later in the course, we will develop the analytic tools to study stochastic resonance in greater detail and understand how it is impacted by both the level of noise and the specific dynamics (geometry) of the system.

> **Exercise 2.4**
>
> Open the course Github page and try playing with the MATLAB script
>
> `CH2_stochastic_resonance.m`
>
> to see how the stochastic and deterministic dynamics vary as you change the value of the k_4 parameter.
> Can you find the value of k_4 where stable oscillations start in the system Equation 2.6? What do the dynamics near the bifurcation point suggest about the type of bifurcation that occurs?

2.3 Stochastic Focusing

The first two sections of this chapter primarily involved the interplay between nonlinearity and stochasticity; this interplay led to novel noise-induced dynamics, and dramatic disagreement between the stochastic dynamics and the law of mass action predictions. Next, we illustrate a phenomenon that relies on nonlinearity and noise, as well as the discrete nature of the underlying process.

Consider the stochastic reaction process given by

$$\emptyset \xrightarrow{k_1} C \xrightarrow{k_2} B \xrightarrow{k_3} \emptyset, \quad A + C \xrightarrow{k_4} A, \quad \emptyset \xrightarrow{k_5} A \xrightarrow{k_6} \emptyset. \tag{2.7}$$
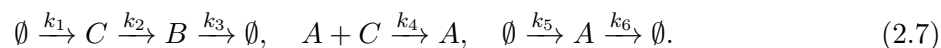
The only second order or higher reaction in the process Equation 2.7 is the middle reaction, in which C molecules are degraded by the presence of A molecules. A influences C, which in turn

regulates the production of B molecules, but A is not itself influenced by C. Hence A is a pure production-degradation process of the type we previously studied. A influences B indirectly via its impact on C.

We begin our analysis of this process by writing down the law of mass action deterministic model for the process. As usual, we write the model for the concentrations (rather than the absolute numbers) of the various chemical species:

$$\frac{d}{dt}a(t) = k_5 - k_6 a(t),$$
$$\frac{d}{dt}b(t) = k_2 c(t) - k_3 b(t), \tag{2.8}$$
$$\frac{d}{dt}c(t) = k_1 - k_2 c(t) - k_4 a(t) c(t),$$

where $a(t) = A(t)/\nu$ and so on.

We can immediately solve for the fixed point of this system,

$$\bar{a} = \frac{k_5}{k_6}, \quad \bar{b} = \frac{k_2 \bar{c}}{k_3}, \quad \bar{c} = \frac{k_1}{k_2 + k_4 \bar{a}},$$

which is stable for all reasonable parameter choices. We will consider a scenario in which almost all of the parameters are fixed, but we will allow the production rate of A, k_5, to switch values as time progresses. Initially, we choose

$$k_1 = 100, \quad k_2 = 1000, \quad k_3 = 0.01, \quad k_4 = 9900, \quad k_6 = 100,$$

and

$$k_5(t) = \begin{cases} 1000, & t < 10, \\ 500, & t \geq 10. \end{cases}$$

Hence if $\bar{A}$ and $\bar{B}$ denote the steady state number of molecules of species A and B respectively, we expect that the system will switch steady states as the value of k_5 switches. In other words, based on the formulae above for $\bar{a}$ and $\bar{b}$, we should have

$$\bar{A} = \begin{cases} 10, & t < 10, \\ 5, & t \geq 10, \end{cases}$$

$$\bar{B} = \begin{cases} 100, & t < 10, \\ 198, & t \geq 10, \end{cases}$$

50
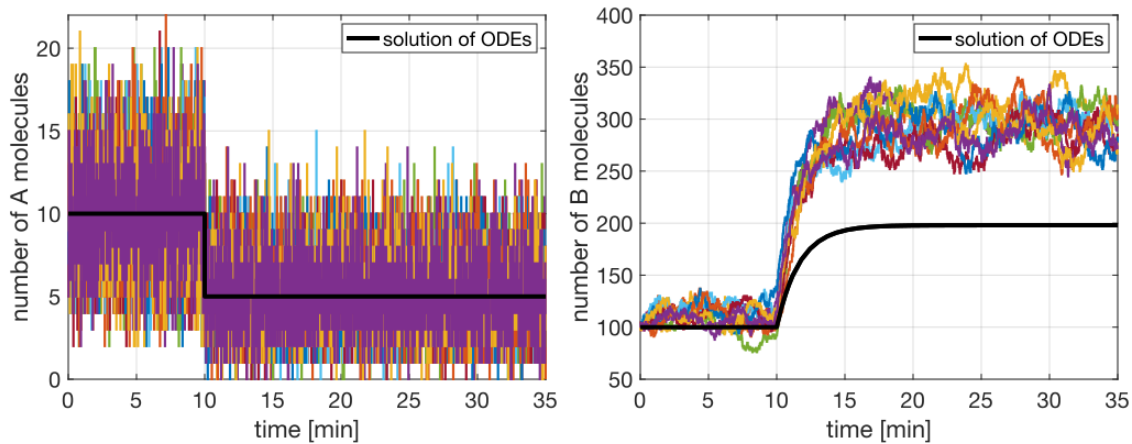
Figure 2.7: Left: Dynamics of the number of A molecules versus time. Right: Dynamics of the number of B molecules versus time.

for this parameter set.

Figure 2.7 shows 10 simulations of the process Equation 2.7 with the solution to the deterministic model Equation 2.8 overlaid in black for comparison. There is excellent agreement between the mean behaviour predicted by Equation 2.8 for the average number of A molecules in the left panel of Figure 2.7. However, we observe a large discrepancy between the number of B molecules observed via the SSA and the mean number predicted by Equation 2.8 after time $t = 10$ in the right-hand panel of Figure 2.7. Species B is more sensitive to the change in k_5 than species A; this increased sensitivity must be somehow due to the way in which this change is transmitted to B via C (whereas A experiences the change in k_5 directly). This phenomenon of enhanced sensitivity to noise is called **stochastic focusing**.

Firstly, we can understand why the agreement is so good between the two models for the number of A molecules. The dynamics of A are a production-degradation process (the number of A molecules is only impacted by the final pair of reactions in Equation 2.7) and thus we know from Chapter 1 that the stochastic mean of A, M_A, obeys the evolution equation

$$\frac{d}{dt}M_A(t) = k_5\nu - k_6 M_A.$$

Hence the long run stochastic mean, $M_{A,s}$, is given by

$$M_{A,s} = \frac{k_5\nu}{k_6},$$

and so for our parameter set, we expect to observe

$$M_{A,s} = \begin{cases} 10, & t < 10, \\ 5, & t \geq 10, \end{cases}$$

which is exactly as shown in the left-hand panel of Figure 2.7. To show that the stochastic fluctuations of A about the mean value $M_{A,s}$ are an essential component of the stochastic focusing phenomenon we observed above, we make $A(t)$ a deterministic process equal to its mean value, i.e., let

$$A(t) = \begin{cases} 10, & t < 10, \\ 5, & t \geq 10. \end{cases} \tag{2.9}$$

Our deterministic model now becomes

$$\begin{aligned} \frac{d}{dt}b(t) &= k_2 c(t) - k_3 b(t), \\ \frac{d}{dt}c(t) &= k_1 - k_2 c(t) - k_4 a(t)c(t), \end{aligned} \tag{2.10}$$

with $a(t) = A(t)/\nu$ where $A(t)$ is given by the formula Equation 2.9. Thus the species B and C are still stochastic processes but take $A(t)$ as deterministic input. The result of simulating this new process is shown in Figure 2.8 with the solution to Equation 2.10 overlaid in black.
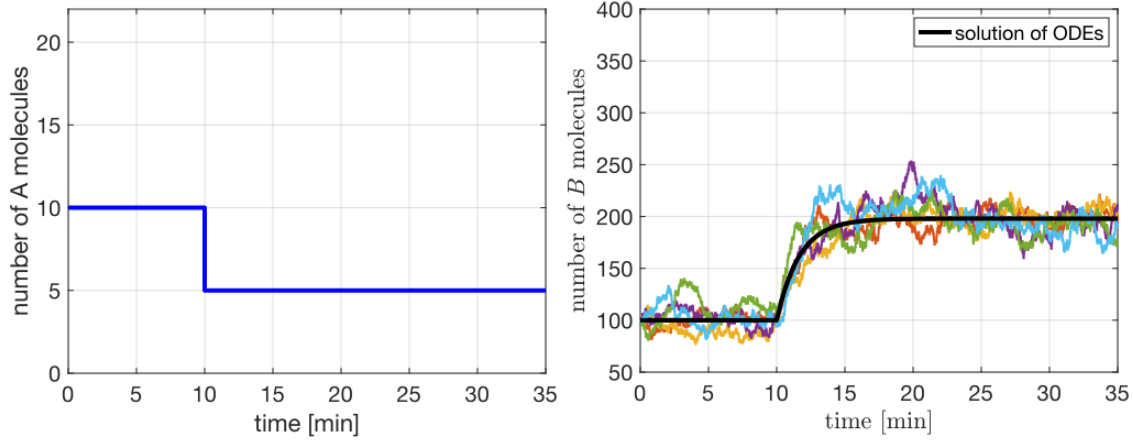
Figure 2.8: Left: The number of A molecules versus time as given by the formula Equation 2.9. Right: Dynamics of the number of B molecules versus time.

As expected, we no longer observe stochastic focusing as the stochastic dynamics now agree very well with the predictions of the deterministic model, confirming that the fluctuations in A are an indispensable aspect of the focusing phenomenon. Moreover, the equations Equation 2.10

are linear and so it can be shown that they are the exact equations for the stochastic means of B and C. Intuitively, we have essentially replaced the 2nd order (nonlinear) reaction in the original process Equation 2.7 with the first order (linear) reaction

$$C \xrightarrow{k_4 A(t)} \emptyset.$$
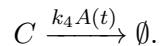
The fluctuations in A and the nonlinear nature of the reaction between A and C are necessary but not sufficient for stochastic focusing. The final ingredient that causes the focusing phenomenon is the discrete nature of the process. In particular, the fact that the number of molecules of C is extremely low in our chosen parameter regime. The average number of C molecules predicted by the law of mass action at steady state is given by

$$\bar{C}_s = \begin{cases} 0.001, & t < 10, \\ 0.00198, & t \geq 10. \end{cases}$$

Thus we can no longer even interpret $\bar{C}_s$ as a number of molecules. In Figure 2.9 we simulate the system with the original parameter set once more and observe the dynamics of the C component. The number of C molecules jumps between zero and one, meaning that sometimes there are no C molecules present to produce B molecules. The presence or absence of C is of course set by the fluctuations in A.
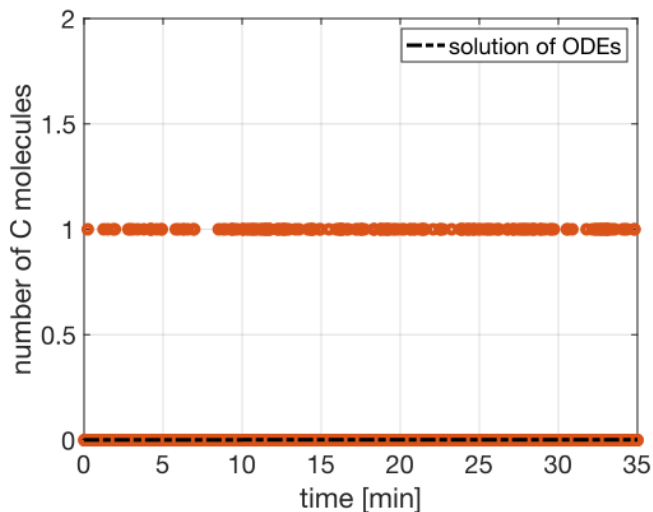
Figure 2.9: Dynamics of the number of C molecules versus time.

To demonstrate that this low number of C molecules is essential for stochastic focusing to occur, we can change the system parameters to ensure a high average level of C molecules. We choose

$$k_1 = 100, \quad k_2 = 0.1, \quad k_3 = 0.01, \quad k_4 = 0.99, \quad k_6 = 100,$$

and this shifts the steady state prediction for $\bar{C}_s$ to

$$\bar{C}_s = \begin{cases} 10, & t < 10, \\ 19.8, & t \geq 10. \end{cases}$$

The stochastic and deterministic dynamics for this new parameter set are shown in Figure 2.10. The average level of C is now much higher and there is now very good agreement between the mass action prediction and the stochastic dynamics of the number of B molecules, i.e., we no longer observe stochastic focusing.
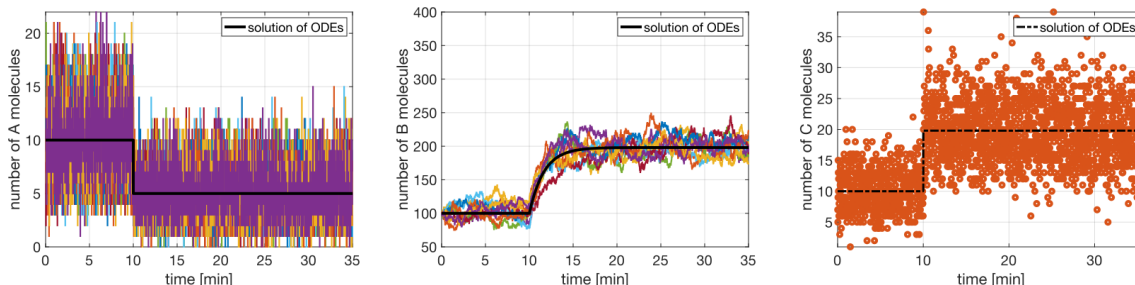
Figure 2.10: Dynamics of the number of A, B and C molecules versus time.

To briefly summarize our analysis of the process Equation 2.7 and the stochastic focusing phenomenon thus far:

- Nonlinearity (higher order reactions) are a necessary component for stochastic focusing (i.e., the reaction $A + C \rightarrow A$ in this example),
- Fluctuations (noise) are an essential component of stochastic focusing,
- Small numbers of molecules in higher order reactions may amplify noise/fluctuations to cause stochastic focusing.

Finally, we will derive a more sophisticated deterministic approximation of the process Equation 2.7 that will allow us to analytically check if a specific system is expected to exhibit stochastic focusing.

Since $A(t)$ is a production-degradation process, we know that its stationary distribution is given by

$$\phi_A(n) = \frac{1}{n!}(M_{A,s})^n e^{-M_{A,s}}, \quad \text{where } M_{A,s} = \frac{k_5}{k_6}$$

i.e., $A(t)$ is approximately Poisson distributed with parameter k_5/k_6 at large times. Next, suppose that $k_4 \gg k_6$ so that C evolves on a much faster time scale than A; this allows us to assume that $A(t)$ is approximately at steady state with respect to C and hence we can assume $A(t)$ is given by the Poisson distribution above.

If $A(t) = n$ at a given time, then C is essentially subject to the reaction dynamics:

$$\emptyset \xrightarrow{k_1} C \xrightarrow{k_2 + k_4 n} \emptyset,$$
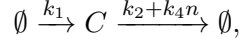
where we write the $\emptyset$ symbol on the right as we choose not to distinguish between C molecules that are degraded and those that become molecules of B; from the perspective of tracking the number of C molecules, these two fates are equivalent. In other words, C is also a production-degradation process and thus we can write down its stationary distribution exactly as we did for A. Therefore

$$\phi_{C,n}(m) = \frac{(\mu_n)^m}{m!} e^{-\mu_n}, \quad \text{where } \mu_n = \frac{k_1 \nu}{k_2 + n k_4 / \nu}.$$

Note that this stationary distribution for C (i.e., the probability that $C(t) = m$ for t large) is dependent on the value of $A(t) = n$ and can thus be thought of as a conditional stationary distribution where we have conditioned on the value of A.

We saw above that stochastic focusing only takes place when the average value of C is much less than 1 and so we want to focus our approximation on the case $\mu_n \ll 1$. In this limit, the Poisson distribution for the behaviour of C tells us that

$$\mathbb{P}[1\, C \text{ molecule present} \mid A(t) = n] = \phi_{C,n}(1) \approx \mu_n,$$

We can then estimate that the average probability of 1 molecule C being present is given by

$$\sum_{n=0}^{\infty} \mu_n \phi_A(n) = \sum_{n=0}^{\infty} \frac{k_1 \nu}{k_2 + n k_4 / \nu} \frac{1}{n!} (M_{A,s})^n e^{-M_{A,s}}, \tag{2.11}$$

where $M_{A,s}$ is the predicted steady-state value of $A(t)$ given above. The formula Equation 2.11 takes into account the small number of C particles as well as the fluctuations in A since we have used the stationary distribution of A in the averaging.

In contrast to the formula Equation 2.11, the law of mass action predicts that the average number of C particles will be

$$\frac{k_1 \nu}{k_2 + M_{A,s} k_4 / \nu}. \tag{2.12}$$

This formula neglects the fluctuations in A by simply using its mean value and also doesn't account for the small number of C molecules in any way. To check for potential stochastic focusing, we can check for substantial disagreement between the predictions for the number of C molecules between the formulae Equation 2.11 and Equation 2.12.

Since B depends linearly on C (i.e., through only a first order reaction), we can estimate the average number of B molecules even for low copy numbers of C using formula Equation 2.11

by simply multiplying by k_2/k_3. If we evaluate these formulae with our original parameter set (which exhibited stochastic focusing), we obtain the estimates

$$M_{B,s} = \begin{cases} 113.1, & t < 10, \\ 316.7, & t \geq 10, \end{cases}$$

which predict exactly the B molecule dynamics observed in Figure 2.7.

> **i** Note
>
> Stochastic focusing is of particular relevance in gene regulatory networks where noisy signals are filtered through complex interaction networks, often with very low copy numbers of specific genes present. Stochastic focusing is a mechanism that allows a regulatory signal to amplify a noisy signal in order to respond hypersensitively for precise control [@eldar2010functional]. If you want to read more about how this phenomenon can be studied mathematically, and applied to understand the dynamics of gene networks, the seminal paper [@paulsson2000stochastic] is the place to start.

Knowledge checklist

Key topics:

1. Multistable systems (noise-induced switching)
2. Stochastic resonance (noise-induced oscillations)
3. Stochastic focusing (nonlinearity and discreteness can amplify noisy signals)

Key skills:

- Identify and explain dramatic differences between stochastic and deterministic models of the same underlying system (especially those caused by stochastic switching, resonance, or focusing).
- Explain the mathematical origin of various noise-induced dynamical phenomena, including stochastic switching, resonance, or focusing.
- Analyse systems to determine if they may exhibit stochastic switching, resonance, or focusing (e.g., using SSAs, mass action approximation, and probabilistic tools).
- Provide examples of specific systems/processes and real-world applications where stochastic switching, resonance, or focusing occur.

References

3 Stochastic Differential Equations

In this chapter we introduce a new class of continuous-time continuous-state stochastic processes that are commonly used for modelling in mathematical biology. Besides being of independent interest, this class of processes will offer more analytic tractability than the chemical reaction processes of the first two chapters of the course. As we have seen in examples and problem classes, analytic calculations for chemical reactions processes with many species and reactions, particularly higher-order reactions, quickly become impossible with just pen and paper! One advantage of these processes is that we can use them to approximate chemical reactions processes and thereby study the complex noise-induced phenomena of Chapter 2 in more detail. The new stochastic processes we introduce will have dynamics described by so-called stochastic differential equations (SDEs), which we can think of as generalizations of ordinary differential equations with random noise added. Instead of a fully rigorous treatment of SDEs, which is an entire course in its own right, we will proceed by discretising time and introducing a "computational definition" of SDEs that gives us the appropriate intuition, enables basic calculations and allows us to perform numerical simulations that produce the correct dynamics.

3.1 Introduction to SDEs

Consider the ordinary differential equation

$$\frac{d}{dt}x(t) = f(x(t)), \quad t > 0, \quad x(0) = x_0. \tag{3.1}$$

We can also write Equation 3.1 in the form

$$dx(t) = f(x(t))\, dt,$$

where we interpret dx as an increment in x and dt as a (small) increment in time. In this context, we mean $dx(t) \approx x(t + dt) - x(t)$ and this naturally leads to the discretisation of the ODE Equation 3.1,

$$x(t + \Delta t) = x(t) + f(x(t))\Delta t, \tag{3.2}$$

where we choose $\Delta t > 0$ as the discretisation parameter. Hence, given the initial condition, we can use this discretisation to numerically approximate the solution to Equation 3.1 by stepping forward in time in steps of Δt (the forward Euler scheme). If f is sufficiently smooth and we choose Δt sufficiently small, we obtain a good approximation and if we let $\Delta t \downarrow 0$, we recover back the differential equation Equation 3.1 exactly (formally, the scheme is first order convergent in Δt).

We could think of Equation 3.2 as a discrete time stochastic process. However, as stated, it is totally deterministic as there is no source of randomness... but we could add some! For example, we could write

$$x(t + \Delta t) = x(t) + f(x(t))\Delta t + G(t),$$

where $G(t)$ is some stochastic process. For reasons that will become clear presently (and some others that won't), we need to insist on more structure than this; we instead choose to extend Equation 3.2 to processes of the form:

$$X(t + \Delta t) = X(t) + f(X(t))\Delta t + g(X(t))\sqrt{\Delta t}\,\xi, \tag{3.3}$$

where ξ is a normally distributed random variable with mean 0 and variance 1. Every time we take a step Δt forward in time, we draw a new realisation of ξ, so we really have a discrete collection of standard normal random variables generating the "noise" in the process Equation 3.3. Fortunately, all modern programming languages can generate normal random variables and hence this makes processes such as Equation 3.3 straightforward to simulate on a computer. Note that if $g \equiv 0$, then Equation 3.3 reduces exactly to Equation 3.2. If g is in some sense small, we can also expect that the dynamics of X will often closely approximate those of the solution to Equation 3.2 (although there will be notable exceptions to this intuition!). Finally, when we write Equation 3.3, we typically restrict t to the discrete set $\{\Delta t, 2\Delta t, \dots\}$ and we can interpolate linearly between these points if we want to know the value of $X(t)$ between time points.

We will refer to equation Equation 3.3 as our **computational definition** of stochastic differential equations (SDEs).

The SDE Equation 3.3 is more often expressed in the form:

$$dX(t) = f(X(t))\,dt + g(X(t))\,dW(t), \tag{3.4}$$

where $W(t)$ is a process whose increments, i.e., $dW(t)$, are normally distributed. In fact, both Equation 3.3 and Equation 3.4 are formal representations of the process $X(t)$. Rigorously, the equation describing the dynamics of $X(t)$ is actually the integral equation

$$X(t) = X(0) + \int_0^t f(X(s))\,ds + \int_0^t g(X(s))\,dW(s), \tag{3.5}$$

which can be formally obtained by integrating Equation 3.4. We can think of Equation 3.5 as equivalent to writing the ODE Equation 3.1 as $x(t) = x(0) + \int_0^t f(x(s))\,ds$ by simply integrating across with respect to t. The final term on the right-hand side of Equation 3.5 is technically a "stochastic integral" and defining such an object rigorously requires developing a whole new theory of integration (which is well beyond the scope of this course). In numerical analysis terms, equation Equation 3.3 is referred to as the **Euler-Maruyama scheme** for discretising Equation 3.5.

You don't need to worry about the details of Equation 3.5 but it is good to be aware of as you may see the same SDE written in different ways in other contexts!

We can immediately write down an SSA for Equation 3.3:

At time $t = 0$, set $X(0)$, then:

1. Generate $\xi \sim N(0, 1)$.
2. Set
$$X(t + \Delta t) = X(t) + f(X(t))\Delta t + g(X(t))\sqrt{\Delta t}\,\xi,$$
   update $t = t + \Delta t$, and go back to step 1.

Example 3.1

Choose $X(0) = 0$, $f \equiv 0$ and $g \equiv 1$. Then Equation 3.3 reads:

$$X(t + \Delta t) = X(t) + \sqrt{\Delta t}\,\xi. \tag{3.6}$$

Suppose $M(t) = \mathbb{E}[X(t)]$ and $V(t) = \text{Var}[X(t)]$. We can then calculate as follows using our computational definition:

$$M(t + \Delta t) = \mathbb{E}[X(t)] + \mathbb{E}\left[\sqrt{\Delta t}\,\xi\right] = M(t).$$

Since $X(0) = 0$, $M(0) = 0$ and thus $M(t) = 0$ for all $t \geq 0$. Similarly, we have

$$\begin{aligned}
V(t + \Delta t) &= \mathbb{E}\left[X(t + \Delta t)^2\right] - M(t + \Delta t)^2 \\
&= \mathbb{E}\left[X(t) + \sqrt{\Delta t}\,\xi)^2\right] \\
&= \mathbb{E}[X(t)^2] + 2\sqrt{\Delta t}\mathbb{E}[X(t)]\mathbb{E}[\xi] + \Delta t\mathbb{E}[\xi^2] \\
&= \mathbb{E}[X(t)^2] + \Delta t \\
&= V(t) + \Delta t.
\end{aligned}$$

Since $V(0) = 0$, $V(t) = t$ for all $t \geq 0$. In other words, the variance of the process Equation 3.6 scales linearly with time. This is a crucial aspect of our choice in adding the noise term when we wrote down Equation 3.3. Moreover, this choice has ensured that both the mean and variance of the process are independent of the time step, suggesting that this was somehow the "right" scaling to choose for the noise term. In fact, we can show that all moments of the process Equation 3.6, i.e., $\mathbb{E}[X(t)^k]$, are independent of step size Δt. Sample paths of the process computed via the SSA above are shown in Figure 3.1 (left panel). We observe that these paths are quite centred around zero with larger fluctuations away from zero as time increases, as expected given that variance scales linearly with time for the process Equation 3.6.

Example 3.2

We can consider higher dimensional examples of SDEs, such as the process

$$\begin{aligned}
X(t + dt) &= X(t) + \sqrt{2D}\,dW_x, \\
Y(t + dt) &= Y(t) + \sqrt{2D}\,dW_y, \\
Z(t + dt) &= Z(t) + \sqrt{2D}\,dW_z,
\end{aligned} \tag{3.7}$$

where $D > 0$ is a parameter and W_x, W_y and W_z are three independent sources of noise. For simulations, as usual, we replace dW_x by $\sqrt{\Delta t}\,\xi$ where $\xi \sim N(0, 1)$ at each time step. The process Equation 3.7 may be interpreted as a stochastic model for a diffusing particle in 3 spatial dimensions, with $D > 0$ denoting the diffusion coefficient. The right-hand panel of Figure 3.1 below shows some sample paths of Equation 3.7 using the SSA outlined above with $(X(0), Y(0), Z(0)) = (0, 0, 0)$ for all simulations. The end point of each simulation is marked with a black dot and as we follow each path, at least qualitatively, we observe what we expect to from a particle diffusing freely in an aqueous medium.

Exercise 3.1

Open the course Github page and try playing with the MATLAB scripts

```
CH3_example_1.m
```

60
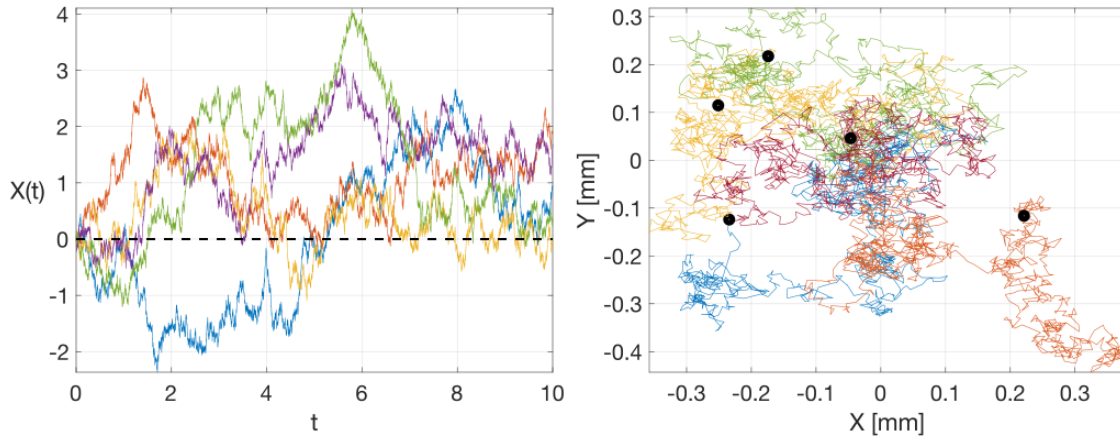
Figure 3.1: Left: Sample paths of the SDE Equation 3.6. Right: Sample paths of the process Equation 3.7 projected onto the X-Y plane. Parameters: $D = 0.0001$.

```
CH3_example_2.m
```

to see how the dynamics vary as you increase the discretisation parameter and the diffusion coefficient.

Example 3.3

Choose $f \equiv 1$ and $g \equiv 1$ so that Equation 3.3 reads

$$X(t + \Delta t) = X(t) + \Delta t + \sqrt{\Delta t}\,\xi = X(t) + dW(t), \quad X(0) = 0. \tag{3.8}$$

In this case, we can compute as in Example 1 to show that

$$M(t + \Delta t) = M(t) + \Delta t \implies M(t) = t,$$

using that $M(0) = 0$ once more.

Figure 3.2 (left) shows some sample paths of the process Equation 3.8. The black dashed line denotes the function $M(t) = t$ and we see that all of the sample paths roughly follow this trend line. This trend line is solely set by the function $f \equiv 1$ which is typically referred to as the "drift" coefficient of the SDE. The function g is called the "diffusion" coefficient and controls the noise level in the process.

Example 3.4

For our final example, choose $f(x) = -k_1 x^3 + k_2 x^2 - k_3 x + k_4$ and $g(x) = k_5$.
Hence we have the SDE

$$X(t + \Delta t) = X(t) + \left(-k_1 X(t)^3 + k_2 X(t)^2 - k_3 X(t) + k_4 \right) \Delta t + k_5 \sqrt{\Delta t}\, \xi. \qquad (3.9)$$

If we set $k_5 = 0$, Equation 3.9 is exactly the ODE we studied in Section 2.1 when we used the law of mass action to derive a deterministic model for the bistable chemical reaction process Equation 2.1. The mass action model for the process Equation 2.1 was given by

$$\frac{d}{dt} a(t) = -k_1 a(t)^3 + k_2 a(t)^2 - k_3 a(t) + k_4, \quad t \geq 0. \qquad (3.10)$$

If we choose reaction rate parameters

$$k_1 = 0.001, \quad k_2 = 0.75, \quad k_3 = 165, \quad k_4 = 10000,$$

then the ODE Equation 3.10 has stable steady states $\bar{A}_1 = 100$ and $\bar{A}_3 = 400$, and an intermediate unstable steady state $\bar{A}_2 = 250$. We saw that the solution to the ODE is not a good model or approximation of the bistable chemical reaction because it cannot switch between the stable fixed points, as the stochastic reaction process can.
In Figure 3.2 we show simulations of the process Equation 3.9 with $k_5 = 200$ and observe that the solution to the SDE can also switch between the stable fixed points of the deterministic model. This illustrates that SDEs offer the potential of better qualitatively matching the dynamics of stochastic reaction processes. Moreover, we can tune the noise level in the SDE to understand how much noise is required to induce switching dynamics and, as we will see presently, we can use the analytic tractability of the SDEs to understand the mean switching time in chemical reaction systems.

Exercise 3.2

Open the course Github page and use the MATLAB script

`CH3_example_4.m`

to investigate how the switching dynamics are impacted by the noise level in Example 3.4 by adjusting the parameter k_5.
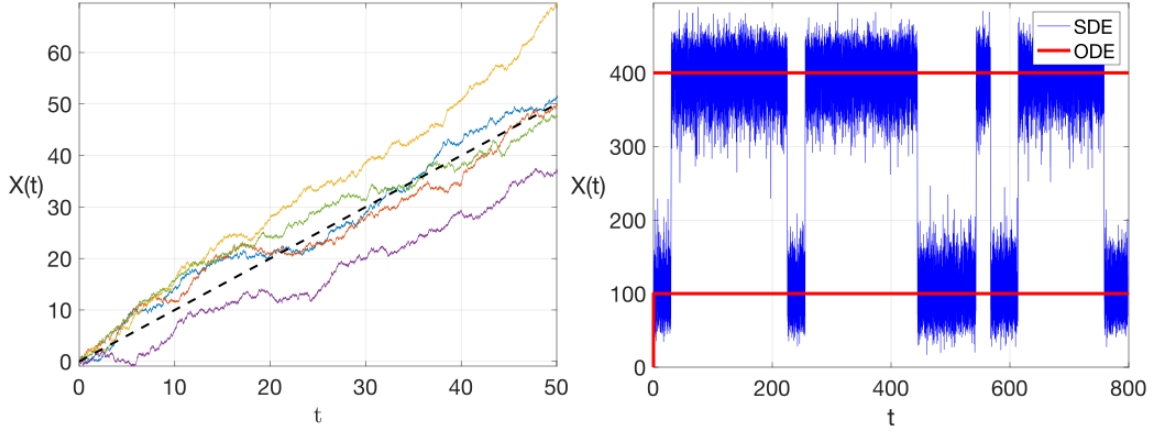
Figure 3.2: Left: Sample paths of the SDE Equation 3.8. Right: Sample path of the process Equation 3.9 and solutions to the corresponding ODE with zero noise.

3.2 The Fokker-Planck Equation

3.2.1 Overview

When we studied chemical reaction processes, we derived the chemical master equations, which can be solved to yield the probability mass function of the process at each point in time. Similarly, if the process $\{X(t) : t \geq 0\}$ solves Equation 3.3 and we denote its probability density function by $p(x, t)$, we can derive an evolution equation for $p(x, t)$.

Since $p(x, t)$ is a PDF, we must have

$$\int_{\mathbb{R}} p(x, t) \, dx = 1,$$

and we have the usual intuitive interpretation of $p(x, t)$ as a proxy for the probability that $X(t)$ is near x at time t, i.e.

$$\mathbb{P}[X(t) \in [x, +dx]] \approx p(x, t) \, dx.$$

We will show that $p(x, t)$ solves the linear second-order PDE:

$$\boxed{\frac{\partial}{\partial t} p(x, t) = \frac{\partial^2}{\partial x^2} \left(\frac{1}{2} g^2(x) p(x, t) \right) - \frac{\partial}{\partial x} \left(f(x) p(x, t) \right), \quad t > 0.} \tag{3.11}$$

63

Equation 3.11 is typically called the Fokker-Planck equation across most fields, although it can also be referred to as the Kolmogorov forward equation or the Smoluchowski equation.

As an example, consider the case when $f \equiv 0$ and $g \equiv \sqrt{2D}$ for some $D > 0$. Then Equation 3.11 reads

$$\frac{\partial}{\partial t} p(x, t) = D \frac{\partial^2}{\partial x^2} p(x, t).$$

In other words, with no drift coefficient and constant diffusion, the Fokker-Planck equation is just the heat equation. Therefore the solution in this case is

$$p(x, t) = \frac{1}{\sqrt{2\pi t D}} e^{-x^2/2Dt}, \quad t \geq 0,$$

if we suppose that the initial condition for the process is $X(0) = 0$; this initial condition corresponds to solving the PDE with the initial distribution being a Dirac delta function (i.e., all mass is at $x = 0$ for $p(x, 0)$). Figure 3.3 (left) shows the solution to the Fokker-Planck equation for this example, along with the estimated PDF from simulations at time $t = 1$. As expected, agreement between the formula and the SSA approach is very close.

When the drift and diffusion coefficients, f and g, are independent of time, we can characterize the long-term behaviour of the SDE Equation 3.3 in terms of the stationary distribution of the process, $p_s(x)$, i.e.

$$p_s(x) = \lim_{t \to \infty} p(x, t).$$

It follows that $p_s(x)$ obeys the steady-state version of the PDE Equation 3.11:

$$0 = \frac{1}{2} \frac{d^2}{dx^2} \left(g^2(x) p_s(x) \right) - \frac{d}{dx} (f(x) p_s(x)). \tag{3.12}$$

In problem classes, we showed that the solution to Equation 3.12 is formally given by

$$\boxed{p_s(x) = \frac{C}{g^2(x)} \exp \left(\int_0^x \frac{2f(y)}{g^2(y)} \, dy \right),} \tag{3.13}$$

where the normalisation constant $C > 0$ is given by

$$C = \left(\int_{\mathbb{R}} \frac{1}{g^2(x)} \exp \left(\int_0^x 2f(y)/g^2(y) \, dy \right) \, dx \right)^{-1}.$$
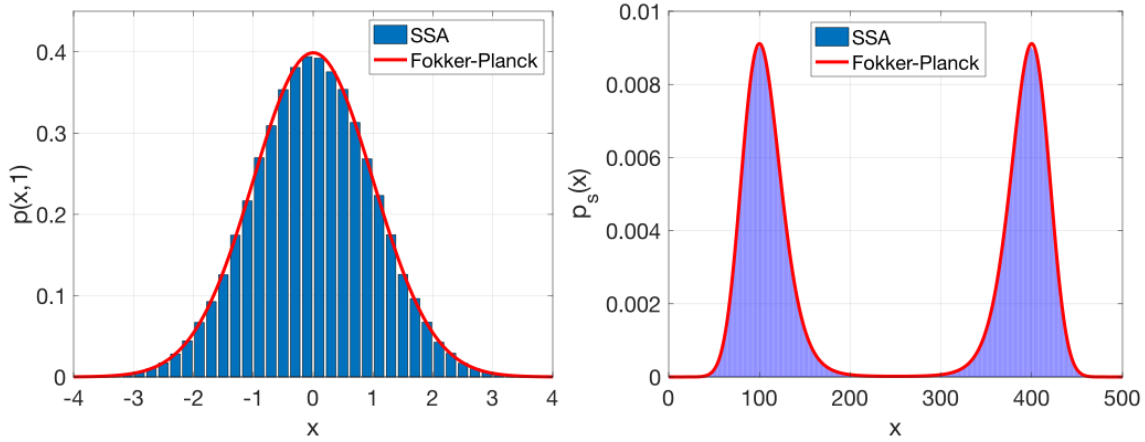
Figure 3.3: Left: Comparison of the Fokker-Planck solution and estimated probability density with $f \equiv 0$ and $g \equiv 1$ at time $t = 1$. Right: Estimated stationary distribution from simulations and stationary distribution obtained via the Fokker-Planck equation for the bistable process Equation 3.9. [Codes: `CH3_FP1.m`, `CH3_FP2.m`]

We saw in our first example above that when drift and diffusion are constant, we obtain a Normal distribution for the process at all times. However, for different choices, both the finite-time and stationary distributions generated can be quite far from Normal. For example, Figure 3.3 (right) shows a comparison of the stationary distribution for the bistable process Equation 3.9 from the last section obtained via both simulations and solving the Fokker-Planck equation using the formula Equation 3.13. Both approaches produce a bimodal distribution in this case, reflecting the stochastic switching between the two stable states at $x = 100$ and $x = 400$ (see Figure 3.2).

3.2.2 Derivation

To derive the Fokker-Planck equation Equation 3.11, we begin by defining the conditional probability density $p(x, t \,|\, y, s)$, which is approximately the probability that $X(t) \in [x, x + dx]$ given that $X(s) = y$ for $s < t$.

Now consider the value of $X(t + \Delta t)$ for some $\Delta t > 0$. We can divide the interval $[s, t + \Delta t]$ into the disjoint intervals $[s, t]$ and $(t, t + \Delta t]$. If we want to arrive at $X(t + \Delta t) = z$, we can think of all possible paths through an intermediate point $X(t) = x$. In other words, we may write

$$p(z, t + \Delta t \,|\, y, s) = \int_{\mathbb{R}} p(z, t + \Delta t \,|\, x, t)\, p(x, t \,|\, y, s)\, dx, \quad s < t. \tag{3.14}$$

Equation 3.14 is called the **Chapman-Kolmogorov equation** and is valid for all $\Delta t > 0$, even though our goal here will be to later let $\Delta t \downarrow 0$ in order to obtain an evolution equation for $p(x, t \,|\, y, s)$.

Next, to alleviate potential regularity issues, we multiply Equation 3.14 across by a smooth test function $\phi(z)$ and integrate over z to obtain:

$$\int_{\mathbb{R}} p(z, t + \Delta t \,|\, y, s)\phi(z)\, dz = \int_{\mathbb{R}} \left[\int_{\mathbb{R}} p(z, t + \Delta t \,|\, x, t)\phi(z)dz \right] p(x, t \,|\, y, s)\, dx,$$

which in the long run will be less confusing to rewrite as

$$\int_{\mathbb{R}} p(x, t + \Delta t \,|\, y, s)\phi(x)\, dx = \int_{\mathbb{R}} \left[\int_{\mathbb{R}} p(z, t + \Delta t \,|\, x, t)\phi(z)dz \right] p(x, t \,|\, y, s)\, dx, \qquad (3.15)$$

where we have simply changed the integration variable on the left-hand side. Now Taylor expand the (smooth test function) ϕ about the point $z = x$ on the right-hand side of Equation 3.15:

$$\int_{\mathbb{R}} \left[\int_{\mathbb{R}} p(z, t + \Delta t \,|\, x, t)\phi(z)dz \right] p(x, t \,|\, y, s)\, dx =$$
$$\int_{\mathbb{R}} \int_{\mathbb{R}} p(z, t + \Delta t \,|\, x, t) \left\{ \phi(x) + \phi'(x)(z - x) + \phi''(x)\tfrac{(z-x)^2}{2} + o((z-x)^2) \right\} p(x, t \,|\, y, s)\, dx.$$
$$(3.16)$$

Now we need to tackle some of the integrals on the right-hand side. Happily, we can rewrite these as (conditional) expectations and compute them using the computational definition of the underlying SDE, i.e., Equation 3.3. Firstly, we have

$$\int_{\mathbb{R}} (z - x)p(z, t + \Delta t \,|\, x, t)\, dz = \mathbb{E}\left[X(t + \Delta t) - x \,|\, X(t) = x \right]$$
$$= \mathbb{E}\left[f(x)\Delta t + g(x)\sqrt{\Delta t}\,\xi \right] = f(x)\Delta t. \qquad (3.17)$$

Similarly, we can compute the second integral as:

$$\int_{\mathbb{R}} (z - x)^2 p(z, t + \Delta t \,|\, x, t)\, dz = \mathbb{E}\left[(X(t + \Delta t) - x)^2 \,|\, X(t) = x \right]$$
$$= \mathbb{E}\left[\left(f(x)\Delta t + g(x)\sqrt{\Delta t}\,\xi \right)^2 \right]$$
$$= \mathbb{E}\left[f^2(x)\Delta t^2 + 2f(x)g(x)(\Delta t)^{3/2}\xi + \xi^2 g^2(x)\Delta t \right]$$
$$= g^2(x)\Delta t + o(\Delta t^2). \qquad (3.18)$$

We have truncated at order Δt since we intend to take $\Delta t \downarrow 0$ later and these higher order terms will vanish; this is also why we truncated our expansion of ϕ at second order. Returning to Equation 3.15 with these estimates and plugging them in yields:

$$\int_{\mathbb{R}} p(x, t+\Delta t \mid y, s)\phi(x)\, dx = \int_{\mathbb{R}} \left\{ \phi(x) + \phi'(x)f(x)\Delta t + \phi''(x)\frac{g(x)^2}{2}\Delta t \right\} p(x, t \mid y, s)\, dx + o(\Delta t^2).$$

With some rearrangement and division by Δt on both sides, we thus have

$$\int_{\mathbb{R}} \frac{p(x, t+\Delta t \mid y, s) - p(x, t \mid y, s)}{\Delta t}\phi(x)\, dx = \int_{\mathbb{R}} \left\{ \phi'(x)f(x) + \phi''(x)\frac{g(x)^2}{2} \right\} p(x, t \mid y, s)\, dx + o(\Delta t).$$
(3.19)

The integrals on the right-hand side of Equation 3.19 can be simplified using integration by parts. Under mild regularity assumptions on f and g, and noting that $\lim_{|x| \to \infty} p(x, t \mid y, s) = 0$, we obtain

$$\int_{\mathbb{R}} \phi'(x)f(x)p(x, t \mid y, s)\, dx = -\int_{\mathbb{R}} \phi(x)\frac{\partial}{\partial x}\left(f(x)p(x, t \mid y, s)\right) dx$$

$$\int_{\mathbb{R}} \phi''(x)\frac{g(x)^2}{2}\, p(x, t \mid y, s)\, dx = \int_{\mathbb{R}} \phi(x)\frac{\partial^2}{\partial x^2}\left(\frac{g^2(x)}{2}p(x, t \mid y, s)\right) dx.$$

Using these identities in Equation 3.19 and letting $\Delta t \downarrow 0$ thus yields

$$\int_{\mathbb{R}} \phi(x)\frac{\partial}{\partial t}\, p(x, t \mid y, s)\, dx = \int_{\mathbb{R}} \phi(x) \left\{ -\frac{\partial}{\partial x}\left(f(x)p(x, t \mid y, s)\right) + \frac{\partial^2}{\partial x^2}\left(\frac{g^2(x)}{2}p(x, t \mid y, s)\right) \right\} dx,$$

or equivalently,

$$\int_{\mathbb{R}} \phi(x) \left\{ \frac{\partial}{\partial t}\, p(x, t \mid y, s) - \frac{\partial}{\partial x}\left(f(x)p(x, t \mid y, s)\right) + \frac{\partial^2}{\partial x^2}\left(\frac{g^2(x)}{2}p(x, t \mid y, s)\right) \right\} dx = 0.$$

The only way this equality can be satisfied for an arbitrary test function ϕ is if the integrand is identically zero, i.e.

$$\frac{\partial}{\partial t}\, p(x, t \mid y, s) = -\frac{\partial}{\partial x}\left(f(x)p(x, t \mid y, s)\right) + \frac{\partial^2}{\partial x^2}\left(\frac{g^2(x)}{2}p(x, t \mid y, s)\right). \qquad (3.20)$$

The PDE Equation 3.20 is valid for any time $s < t$ and any y but for the initial value problem for the SDE Equation 3.3, we simply want the case $s = 0$ and $y = x_0$. Thus we can let

$p(x,t) = p(x,t \,|\, x_0, 0)$ and specialize Equation 3.20 to obtain

$$\frac{\partial}{\partial t} p(x,t) = -\frac{\partial}{\partial x}\left(f(x)p(x,t)\right) + \frac{\partial^2}{\partial x^2}\left(\frac{g^2(x)}{2}p(x,t)\right),$$

which is exactly the claimed Fokker-Planck equation given by Equation 3.11.

3.2.3 Boundary conditions

In the previous section we derived the Fokker-Planck equation under the assumption that the underlying process X took values on all of $\mathbb{R}$. In effect, there is no boundary and hence there is no need for a boundary condition to solve Equation 3.11 in this scenario.

In many practical situations we want to restrict the set of values that the process X can attain and this has natural implications for the Fokker-Planck equations of the process. For example, if we are modelling the positions of particles or cells in a Petri dish, then the domain of the cell positions is probably best modelled as a compact subset of $\mathbb{R}^2$. We must then decide what happens to cells when they come into contact with the boundary of the dish.

Suppose that a particle moves in 1 dimension and its position must remain above the level $x = 0$. We can impose a "reflecting" boundary condition at $x = 0$ to achieve this, i.e., the particle hits the boundary and is pushed back into the admissible domain $(0, \infty)$. If we are using the recurrence relation Equation 3.3 to simulate the process, we can modify our previous SSA for SDEs as follows to account for the addition of a reflecting boundary:

At time $t = 0$, set $X(0)$, then:

1. Generate $\xi \sim N(0,1)$.
2. Compute the possible value of $X(t + \Delta t)$:

$$X(t + \Delta t) = X(t) + f(X(t))\Delta t + g(X(t))\sqrt{\Delta t}\,\xi,$$

3. If $X(t + \Delta t) \geq 0$, then we accept that value. If $X(t + \Delta t) < 0$, then we instead take

$$X(t + \Delta t) = -X(t) - f(X(t))\Delta t - g(X(t))\sqrt{\Delta t}\,\xi.$$

Finally, set $t = t + \Delta t$ and go back to step 1.

In the "reflective step" in step 3 above, we must have $X(t) > 0$. Thus if $X(t + \Delta t) < 0$ in step 2, we must have had

$$f(X(t))\Delta t + g(X(t))\sqrt{\Delta t}\,\xi < 0.$$

In the reflective step, we first reflect the process through zero by replacing $X(t)$ by $-X(t)$. Then we add the positive increment $-f(X(t))\Delta t - g(X(t))\sqrt{\Delta t}\,\xi$ to make the process positive once more.

As an example, consider the SDE Equation 3.3 with $f(x) = -x$ and $g \equiv 1$ with initial condition $X(0) = 0$. We can simulate a path of this process using our SSA with and without the reflecting boundary condition to see the impact of the boundary on the process. To make the comparison even more direct, we will use the same set of normal random variables (ξ's) for the noise terms. In the left-hand panel of Figure 3.4 we see how the paths of the reflected and unreflected processes diverge as soon as they hit zero for the first time. The reflected process (blue) hits zero numerous times but is blocked from crossing into $(-\infty, 0)$, while the red process is unrestricted. We observe that both processes have the same dynamics away from $x = 0$ since they share the same underlying noise process and thus their increments are mostly the same (except for reflection events). The right-hand panel of Figure 3.4 shows the estimated PDF of the reflected and unreflected processes at time $t = 10$. As expected, the PDF of the reflected process is only supported on $[0, \infty)$.
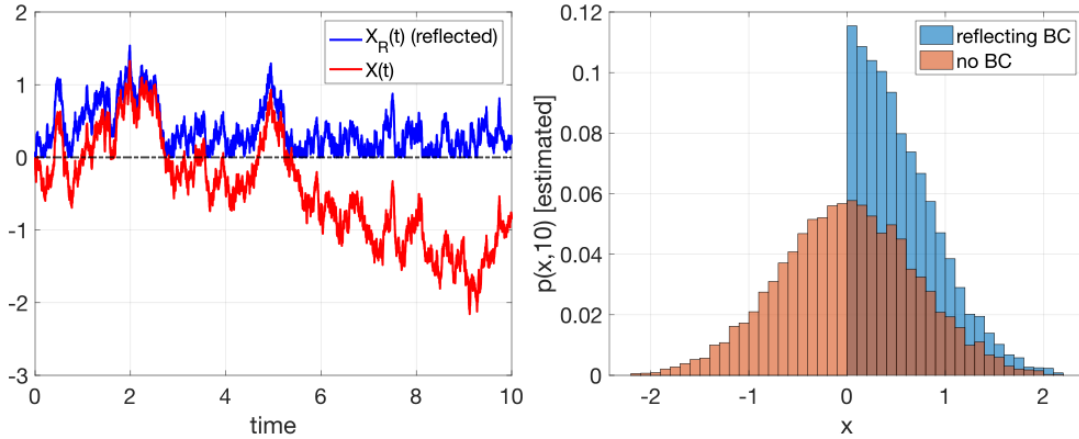
Figure 3.4: Left: Samples path of $X(t)$ and $X_R(t)$ with $f(x) = -x$ and $g \equiv 1$ with identical initial conditions and noise processes (i.e. the same random variables ξ). Right: Estimated PDFs for X and X_R at time $t = 10$. [Code: `CH3_FP3_reflecting.m`]

Figure 3.4 shows how significantly boundary conditions can change the solution to the Fokker-Planck equation, this immediately begs the question: How does the reflecting boundary condition impact the derivation of the corresponding Fokker-Planck equation?

Let $p_R(x, t \mid y, s)$ denote the conditional probability density function of a process X_R solving Equation 3.3 subject to the reflecting boundary condition introduced above. The process X_R can no longer take on negative values with positive probability so $p_R(x, t \mid y, s) = 0$ if $x < 0$ or $y < 0$. The Chapman-Kolmogorov equation Equation 3.14 then becomes

$$p_R(z, t + \Delta t \mid y, s) = \int_{[0,\infty)} p_R(z, t + \Delta t \mid x, t)\, p_R(x, t \mid y, s)\, dx, \quad s < t. \tag{3.21}$$

As before, we can multiply by a smooth test function ϕ and integrate to obtain:

$$\int_{[0,\infty)} p_R(x, t + \Delta t \mid y, s)\phi(x)\, dx = \int_{[0,\infty)} \left[\int_{[0,\infty)} p_R(z, t + \Delta t \mid x, t)\phi(z)dz \right] p_R(x, t \mid y, s)\, dx. \tag{3.22}$$

Next consider the **unreflected** process X, whose conditional density is denoted $p(x, t \mid y, s)$, and whose Fokker-Planck equation we derived in the previous section. We assume that X solves the same SDE as X_R. If $x > 0$, $y > 0$, then a path of X that goes from $X(s) = y$ to $X(t) = -x$, has the same probability as a path of the reflected process that goes from $X_R(s) = y$ to $X_R(t) = x$ due to the symmetry of the reflection step. If our test function ϕ is chosen to be an even function, we then have

$$\begin{aligned}
\int_{[0,\infty)} p_R(z, t + \Delta t \mid x, t)\phi(z)dz &= \int_{[0,\infty)} p(z, t + \Delta t \mid x, t)\phi(z)dz \\
&\quad + \int_{(-\infty,0]} p(z, t + \Delta t \mid x, t)\phi(-z)dz \\
&= \int_{[0,\infty)} p(z, t + \Delta t \mid x, t)\phi(z)dz \\
&\quad + \int_{(-\infty,0]} p(z, t + \Delta t \mid x, t)\phi(z)dz.
\end{aligned}$$

Hence, we have

$$\int_{[0,\infty)} p_R(x, t + \Delta t \mid y, s)\phi(x)\, dx = \int_{[0,\infty)} \left[\int_{\mathbb{R}} p(z, t + \Delta t \mid x, t)\phi(z)dz \right] p_R(x, t \mid y, s)\, dx, \tag{3.23}$$

The inner integral involving p on the right-hand side of Equation 3.23 is exactly the integral we dealt with in the original derivation; we can once more Taylor expand ϕ about $z = x$, simplify and let $\Delta t \downarrow 0$ to obtain

$$\int_{[0,\infty)} \frac{\partial}{\partial t} p_R(x, t + \Delta t \,|\, y, s)\phi(x)\,dx = \int_{[0,\infty)} \left[\phi'(x)f(x) + \phi''(x)\frac{g(x)^2}{2}\right] p_R(x, t \,|\, y, s)\,dx. \quad (3.24)$$

We again apply integration by parts to simplify the integrals on the right-hand side. Doing so yields

$$\int_{[0,\infty)} \phi'(x)f(x)p_R(x, t \,|\, y, s)\,dx = -\int_{[0,\infty)} \phi(x)\frac{\partial}{\partial x}\left(f(x)p_R(x, t \,|\, y, s)\right)dx$$
$$+ \phi(x)f(x)\,p_R(x, t \,|\, y, s)|_{x=0},$$

$$\int_{[0,\infty)} \phi''(x)\frac{g(x)^2}{2}\,p_R(x, t \,|\, y, s)\,dx = \int_{[0,\infty)} \phi(x)\frac{\partial^2}{\partial x^2}\left(\frac{g^2(x)}{2}p_R(x, t \,|\, y, s)\right)dx$$
$$- \phi(x)\left[\frac{\partial}{\partial x}\left(\frac{g^2(x)}{2}p_R(x, t \,|\, y, s)\right)\right]_{x=0},$$

where we used the fact that $\phi'(0) = 0$ because ϕ is an even function. Therefore, for an arbitrary test function ϕ, we have

$$\int_{[0,\infty)} \frac{\partial}{\partial t} p_R(x, t + \Delta t \,|\, y, s)\phi(x)\,dx = -\int_{[0,\infty)} \phi(x)\frac{\partial}{\partial x}\left(f(x)p_R(x, t \,|\, y, s)\right)dx$$
$$+ \int_{[0,\infty)} \phi(x)\frac{\partial^2}{\partial x^2}\left(\frac{g^2(x)}{2}p_R(x, t \,|\, y, s)\right)dx$$
$$+ \left[\phi(x)f(x)\,p_R(x, t \,|\, y, s) - \phi(x)\frac{\partial}{\partial x}\left(\frac{g^2(x)}{2}p_R(x, t \,|\, y, s)\right)\right]_{x=0}$$

The only way this equation can hold for an arbitrary ϕ is if the integrands agree, i.e. p_R obeys the standard Fokker-Planck equation, and, since $\phi(0)$ is arbitrary, we also need the last term on the right-hand side to vanish regardless of the value of $\phi(0)$, i.e.

$$f(x)\,p_R(x, t \,|\, y, s) - \frac{\partial}{\partial x}\left(\frac{g^2(x)}{2}p_R(x, t \,|\, y, s)\right) = 0 \quad \text{at } x = 0. \quad (3.25)$$

Equation 3.25 is thus the condition that p_R must satisfy at $x = 0$ in order to take into account the reflecting boundary condition on the dynamics of the underlying stochastic process X_R. In fact, we can interpret condition Equation 3.25 as a no-flux boundary condition at $x = 0$ as follows: We can write the Fokker-Planck equation Equation 3.11 in the form

71

$$\frac{\partial}{\partial t} p_R(x, t \mid y, s) + \frac{\partial Q}{\partial x} = 0$$

where the probability flux Q is given by

$$Q(x) = f(x) p_R(x, t \mid y, s) - \frac{\partial}{\partial x} \left(\frac{g^2(x)}{2} p_R(x, t \mid y, s) \right).$$

The condition in Equation 3.25 corresponds to taking $Q(0) = 0$, i.e. no flux at the boundary $x = 0$.

A reflecting boundary is not the only possible choice of boundary condition for the Fokker-Planck equation. For example, we could decide to have an "absorbing boundary" in which we terminate trajectories that reach the boundary. This would correspond to ignoring cells which touch the edge of the Petri dish during an experiment. If we set an absorbing boundary at $x = 0$, then the boundary condition for the Fokker-Planck equation is $p(0, t \mid y, s) = 0$, i.e., there is no chance that a particle can reach $x = 0$. This means that the integral $\int_{(0,\infty)} p(x, t)\, dx$ is equal to 1 at time $t = 0$ but its values will decay in time and obey

$$\lim_{t \to \infty} \int_{(0,\infty)} p(x, t)\, dx = 0$$

since all trajectories will eventually hit $x = 0$ and be absorbed if we wait long enough.

Our choice of boundaries is typically motivated by the application we have in mind for the underlying process and other boundary conditions can be implemented using approaches similar to that employed above for the reflecting case.

3.3 The Kolmogorov Backward Equation

3.3.1 Derivation

The Fokker-Planck equation Equation 3.11 tells us how the probability density function of the process changes as time evolves, i.e., how the probability of being in a certain state changes with time. In some situations, we actually want to know how the probability of ending up in a certain state depends on where the process starts, a question that the Fokker-Planck equation is not designed to answer. Instead, for this question, we need the so-called **Kolmogorov backward equation** of the process. If we consider the conditional density function $p(x, t \mid y, s)$, then the Fokker-Planck equation tells us what happens when we vary x, t and the Kolmogorov backward equation gives us information about what happens when y, s are varied. Moreover, we can use the backward equation to derive information about the mean hitting times of states

depending on a given starting position and this will greatly improve our understanding of the noise-induced switching phenomenon we have seen in previous examples.

To derive the Kolmogorov backward equation, we begin by considering the Chapman-Kolmogorov equation with a relabelling of the variables:

$$p(x,t \,|\, y, s - \Delta s) = \int_{\mathbb{R}} p(x,t \,|\, z, s) \, p(z, s \,|\, y, s - \Delta s) \, dz. \tag{3.26}$$

This equation is valid for any $\Delta s > 0$ and our goal will be to take the limit $\Delta s \downarrow 0$ presently. Previously we introduced a smooth test function at this point (for regularity purposes) but this will be a more formal derivation. Thus we will take a rather risky Taylor expansion of $p(x,t \,|\, z, s)$ about the point $z = y$ to obtain

$$p(x,t \,|\, z, s) = p(x,t \,|\, y, s) + (z - y)\frac{\partial}{\partial y}p(x,t \,|\, y, s) + \frac{(z-y)^2}{2}\frac{\partial^2}{\partial y^2}p(x,t \,|\, y, s) + o((z-y)^2).$$

Substituting this expression into the right-hand side of Equation 3.26 yields

$$
\begin{aligned}
p(x,t \,|\, y, s - \Delta s) = {} & p(x,t \,|\, y, s) \int_{\mathbb{R}} p(z, s \,|\, y, s - \Delta s) \, dz \\
& + \frac{\partial}{\partial y}p(x,t \,|\, y, s) \int_{\mathbb{R}} (z - y)p(z, s \,|\, y, s - \Delta s) \, dz \\
& + \frac{\partial^2}{\partial y^2}p(x,t \,|\, y, s) \int_{\mathbb{R}} \frac{(z-y)^2}{2} p(z, s \,|\, y, s - \Delta s) \, dz + o(\Delta s^2).
\end{aligned}
$$

The first integral on the right-hand side of this equation is 1 because it is the integral of a PDF. The second and third integrals are the mean and variance of the process and are dealt with as in the derivation of the Fokker-Planck equation (see equations Equation 3.17 and Equation 3.18). Simplifying these integrals thus yields

$$
\begin{aligned}
p(x,t \,|\, y, s - \Delta s) = {} & p(x,t \,|\, y, s) + f(y)\frac{\partial}{\partial y}p(x,t \,|\, y, s)\Delta s \\
& + \frac{g(y)^2}{2}\frac{\partial^2}{\partial y^2}p(x,t \,|\, y, s)\Delta s + o(\Delta s^2).
\end{aligned}
$$

Rearranging, dividing across by Δs and letting $\Delta s \downarrow 0$ gives us the Kolmogorov backward equation

$$\boxed{-\frac{\partial}{\partial s}p(x,t \,|\, y, s) = f(y)\frac{\partial}{\partial y}p(x,t \,|\, y, s) + \frac{g(y)^2}{2}\frac{\partial^2}{\partial y^2}p(x,t \,|\, y, s).}$$

{#eq-KBE}

If we define the function $d(y) = g^2(y)/2$, then we can write the Fokker-Planck equation more compactly as

$$\frac{\partial}{\partial t}p(x,t) = \frac{\partial^2}{\partial x^2}\left(d(x)p(x,t)\right) - \frac{\partial}{\partial x}\left(f(x)p(x,t)\right)$$

and the Kolmogorov backward equation as

$$-\frac{\partial}{\partial s}p(x,t\,|\,y,s) = f(y)\frac{\partial}{\partial y}p(x,t\,|\,y,s) + d(y)\frac{\partial^2}{\partial y^2}p(x,t\,|\,y,s).$$

The stationary distribution from solving the steady-state version of the Fokker-Planck equation is then given by

$$p_s(x) = \frac{C}{d(x)}\exp\left(\int_0^x \frac{f(y)}{d(y)}\,dy\right), \quad C > 0.$$

3.3.2 Application to mean hitting times

Consider once more the SDE from Example 3.4 earlier in this chapter:

$$X(t+\Delta t) = X(t) + \left(-k_1 X(t)^3 + k_2 X(t)^2 - k_3 X(t) + k_4\right)\Delta t + k_5\sqrt{\Delta t}\,\xi.$$

Using the same parameter set as before, we know that this SDE can exhibit noise-induced switching between the stable steady states $x_{s1} = 100$ and $x_{s2} = 400$ (see Figure 3.2 right). There is also an unstable steady state $x_u = 250$. Our goal will be to estimate how long on average it takes the process to hit the unstable steady state x_u given that it starts below that point. To this end, define the probability $h(y,t)$ by

$$h(y,t) = \mathbb{P}\left[X(t') \in (-\infty, x_u) \; \forall t' < t, \; X(0) = y < x_u\right].$$

In other words, $h(y,t)$ is the probability that the process stays below x_u before time t, given that it started at position y. The quantity $p(x,t\,|\,y,0)dx$ is approximately the probability that a particle is at position y at time zero and in the interval $[x, x+dx]$ at time t; this density obeys the Fokker-Planck (FP) and Kolmogorov backward equations (KBEs). If we supplement the FP and KBEs with the boundary conditions

$$p(x_u, t\,|\,y,s) = p(x,t\,|\,x_u,s) = 0, \quad s < t, \qquad x < x_u,$$

then we ensure that $p(x, t \,|\, y, s) = 0$ if $y \geq x_u$ or $x \geq x_u$. In other words, with these boundary conditions, paths of the particle that cross the level x_u before time t have probability zero. Hence we can write the desired probability as

$$h(y, t) = \int_{-\infty}^{x_u} p(x, t \,|\, y, 0)\, dx,$$

where $p(x, t \,|\, y, 0)\, dx$ is the probability that the process remains in $(-\infty, x_u)$ and lies in $[x, x+dx]$ at time t, given that it started at y initially. Since f, g do not depend explicitly on time, we can then shift time as follows in the expression for h:

$$h(y, t) = \int_{-\infty}^{x_u} p(x, 0 \,|\, y, -t)\, dx.$$

Similarly, shifting time in the Kolmogorov backward equation $(s \mapsto -t)$ yields:

$$\frac{\partial}{\partial t} p(x, 0 \,|\, y, -t) = f(y) \frac{\partial}{\partial y} p(x, 0 \,|\, y, -t) + d(y) \frac{\partial^2}{\partial y^2} p(x, 0 \,|\, y, -t).$$

Integrating over x thus yields

$$\frac{\partial}{\partial t} h(y, t) = f(y) \frac{\partial}{\partial y} h(y, t) + d(y) \frac{\partial^2}{\partial y^2} h(y, t). \tag{3.27}$$

Define the hitting time of x_u starting from y by

$$\tau(y) = \inf\{t > 0 : \ X(t) = x_u, \ X(0) = y\},$$

and thus the average hitting time $\bar{\tau}(y)$ is given by $\bar{\tau}(y) = \mathbb{E}[\tau(y)]$. Formally, we can observe that

$$\mathbb{P}[\tau(y) = t] \approx h(y, t) - h(y, t + dt) \approx -\frac{\partial}{\partial t} h(y, t) dt.$$

Next apply integration by parts to find that

$$\bar{\tau}(y) = -\int_0^\infty t \frac{\partial}{\partial t} h(y, t)\, dt = \int_0^\infty h(y, t)\, dt.$$

Next integrate Equation 3.27 with respect to t and use the initial condition $h(y, 0) = 1$ to show that

$$-1 = f(y)\frac{d}{dy}\bar{\tau}(y) + d(y)\frac{d^2}{dy^2}\bar{\tau}(y), \quad y \in (-\infty, x_u). \tag{3.28}$$

Note that $\lim_{t\to\infty} h(y,t) = 0$ because every trajectory will eventually cross x_u if we wait long enough. We need to impose some boundary conditions in order to solve Equation 3.28. If $y = x_u$, then $p(x, t \mid x_u, s) = 0$, implying that $h(x_u, t) = 0$, which means that

$$\bar{\tau}(x_u) = 0.$$

Next, if we start very far from x_u, then the initial condition should have a negligible impact on how long it takes to hit the level x_u and thus we assume that

$$\frac{d}{dy}\bar{\tau}(y)|_{y=-\infty} = 0.$$

With these boundary conditions in hand, we can solve Equation 3.28 to show that

$$\bar{\tau}(y) = \int_y^{x_u} \frac{1}{d(z)p_s(z)} \int_{-\infty}^z p_s(x)\, dx\, dz.$$

Now we can calculate, for example, how long it takes on average to hit x_u starting from the lower stable state, i.e.

$$\bar{\tau}(x_{s1}) = \int_{x_{s1}}^{x_u} \frac{1}{d(z)p_s(z)} \int_{-\infty}^z p_s(x)\, dx\, dz.$$

Figure 3.5 (right) shows the results of computing the integral formula for $\bar{\tau}$ above for different values of y and compares it to a simulation approach of estimating the hitting time over a large number of paths of the process. The left panel of Figure 3.5 demonstrates one drawback or source of inaccuracy of the direct simulation approach, namely discretisation error. The blue and red paths are essentially the same sample path but the blue path is sampled using a smaller discretisation parameter and hence should be more accurate. We observe that for this particular sample path, the red solution underestimates the hitting time of the level $x_u = 250$. We could take smaller and smaller step sizes for accuracy but of course this will be computationally intensive and demonstrates the advantage of our analytic formula for $\bar{\tau}(y)$.
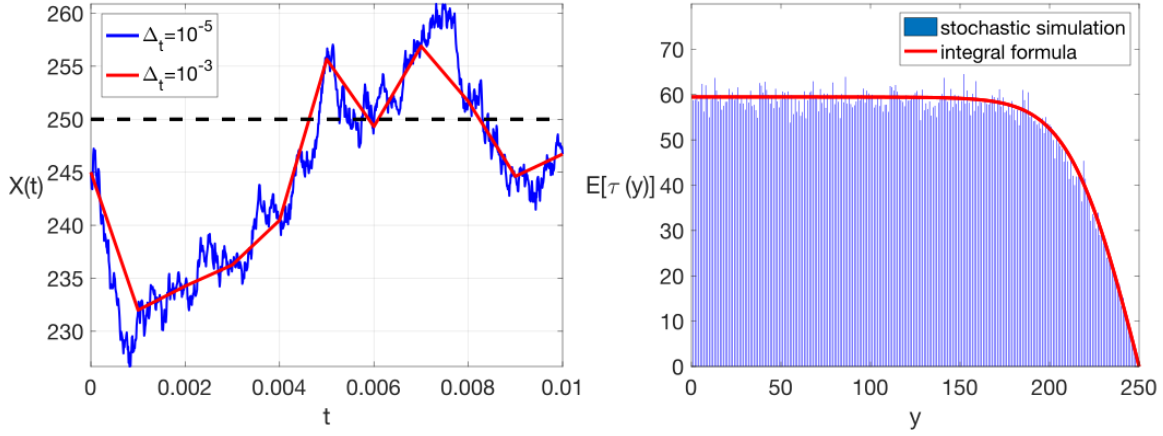
Figure 3.5: Left: Paths of the SDE Equation 3.9 with different values of Δt. Right: Estimates of the mean hitting time of the state $x_u = 250$, $\bar{\tau}(y)$, for different values of the initial condition y. [Code: `CH3_KBE_hitting.m` plus associated .dat files]

3.4 The Chemical Fokker-Planck Equation

Thus far we have seen that SDEs can be useful modelling tools in their own right and that it is more analytically tractable to study certain problems for SDEs than it is for their chemical reaction process counterparts. For example, we developed theory to estimate mean hitting times of states for a general class of SDEs and applied this to the study of systems with noise-induced switching between alternative stable states. We will conclude this chapter by making a connection between SDEs and chemical reaction processes that will allow us to apply our analytic tools for SDEs to reaction processes, thereby enabling us to tackle some problems which were intractable up to now. Our approach will be to approximate the chemical master equations of a reaction process by an appropriate Fokker-Planck equation in the *large molecule number regime*; this approximation is the so-called **Chemical Fokker-Planck equation**.

3.4.1 Derivation for production-degradation processes

Consider our old friend the single-species production degradation process:

$$A \xrightarrow{k_1} \emptyset, \quad \emptyset \xrightarrow{k_2} A. \tag{3.29}$$
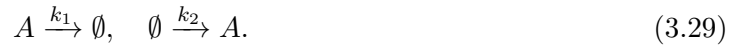
In Chapter 1 we showed that the chemical master equations of the process Equation 3.29 are given by

$$\frac{d}{dt} P_n(t) = k_1(n+1)P_{n+1}(t) - k_1 n P_n(t) + k_2 \nu P_{n-1}(t) - k_2 \nu P_n(t), \quad n \geq 0, \quad t \geq 0.$$

77

Introducing the auxiliary functions

$$h_1(n,t) = k_1 n P_n(t), \quad h_2(n,t) = k_2 \nu P_n(t),$$

we can rewrite the chemical master equations as

$$\frac{d}{dt} P_n(t) = h_1(n+1,t) - h_1(n,t) + h_2(n-1,t) - h_2(n,t), \quad n \geq 0, \quad t \geq 0. \qquad (3.30)$$

The key to our approximation of Equation 3.30 will be to assume that we can concentrate on the case when the number of molecules of A, denoted by n, is large; justifying this assumption would largely depend on the specific dynamics of the process in question. Consider a fixed large number ω such that $n = \eta\omega$ for some continuous variable $\eta > 0$ and set $P_n(t) = P(\eta, t)$. Our auxiliary functions thus become

$$h_1(\eta, t) = k_1 \eta \omega P(\eta, t), \quad h_2(\eta, t) = k_2 \nu P(\eta, t)$$

and Equation 3.30 now reads

$$\frac{\partial}{\partial t} P(\eta, t) = h_1\left(\eta + \tfrac{1}{\omega}, t\right) - h_1(\eta, t) + h_2\left(\eta - \tfrac{1}{\omega}, t\right) - h_2(\eta, t), \quad t \geq 0. \qquad (3.31)$$

Given that ω is a large parameter, we can Taylor expand the right-hand side of Equation 3.31 in powers of ω about the point (η, t). The expansions of h_1 and h_2 will be:

$$h_1\left(\eta + \tfrac{1}{\omega}, t\right) \approx h_1(\eta, t) + \frac{1}{\omega}\frac{\partial}{\partial \eta} h_1(\eta, t) + \frac{1}{2\omega^2}\frac{\partial^2}{\partial \eta^2} h_1(\eta, t) + O(h_1/\omega^3),$$

$$h_2\left(\eta - \tfrac{1}{\omega}, t\right) \approx h_2(\eta, t) - \frac{1}{\omega}\frac{\partial}{\partial \eta} h_2(\eta, t) + \frac{1}{2\omega^2}\frac{\partial^2}{\partial \eta^2} h_2(\eta, t) + O(h_2/\omega^3).$$

Substituting these expansions into Equation 3.31 and truncating at $O(1/\omega^2)$ yields:

$$\frac{\partial}{\partial t} P(\eta, t) = \frac{1}{\omega}\frac{\partial}{\partial \eta} h_1(\eta, t) + \frac{1}{2\omega^2}\frac{\partial^2}{\partial \eta^2} h_1(\eta, t) - \frac{1}{\omega}\frac{\partial}{\partial \eta} h_2(\eta, t) + \frac{1}{2\omega^2}\frac{\partial^2}{\partial \eta^2} h_2(\eta, t).$$

For brevity and notational convenience, we can let $x = \eta\omega$ to obtain the slightly more pleasant expression

$$\frac{\partial}{\partial t} P(x, t) = \frac{\partial}{\partial x} h_1(x, t) + \frac{1}{2}\frac{\partial^2}{\partial x^2} h_1(x, t) - \frac{\partial}{\partial x} h_2(x, t) + \frac{1}{2}\frac{\partial^2}{\partial x^2} h_2(x, t), \qquad (3.32)$$

while bearing in mind that this approximation is now only valid for large values of x by dint of the relationship between n (molecule number) and x. Carefully tracing through our latest change of variables, we have

$$h_1(x,t) = k_1 x P(x,t), \quad h_2(x,t) = k_2 \nu P(x,t),$$

and hence Equation 3.32 can be rewritten as

$$\frac{\partial}{\partial t} P(x,t) = -\frac{\partial}{\partial x}\left(f(x)P(x,t)\right) + \frac{\partial^2}{\partial x^2}\left(d(x)P(x,t)\right), \tag{3.33}$$

where, in our usual SDE/Fokker-Planck notation, we have

$$f(x) = -k_1 x + k_2 \nu, \quad d(x) = \frac{g^2(x)}{2} = \frac{k_1 x + k_2 \nu}{2}. \tag{3.34}$$

Thus Equation 3.33, together with the coefficients from Equation 3.34, are the chemical Fokker-Planck equation for the process Equation 3.29. The solution $P(x,t)$ of Equation 3.33 gives us the approximate probability that we will have roughly x molecules of A present at time t.

At least for large values of x (i.e., large numbers of molecules of A), we can use Equation 3.33 to answer questions about the process Equation 3.29. For example, we can solve the steady-state version of Equation 3.33 to readily obtain an approximate stationary distribution for the production-degradation process. Doing so yields

$$
\begin{aligned}
P_s(x) &= \frac{C}{d(x)} \exp\left(\int_0^x \frac{f(y)}{d(y)}\, dy\right) \\
&= \frac{2C}{k_1 x + k_2 \nu} \exp\left(2\int_0^x \frac{-k_1 y + k_2 \nu}{k_1 y + k_2 \nu}\, dy\right) \\
&= \frac{2C}{k_1 x + k_2 \nu} \exp\left(-2x + 4k_2 \nu \int_0^x \frac{dy}{k_1 y + k_2 \nu}\right) \\
&= 2C \exp\left(-2x + \frac{4k_2 \nu}{k_1} \log(k_1 x + k_2 \nu) - \frac{4k_2 \nu}{k_1} \log(k_2 \nu)\right), \quad x > 0, \quad C > 0.
\end{aligned}
\tag{3.35}
$$

The approximate stationary distribution Equation 3.35 is only valid for $x > 0$ since x represents molecule numbers in this formulation and thus the normalisation constant will be given by

$$C = \frac{1}{2}\left[\int_0^\infty \exp\left(-2x + \frac{4k_2 \nu}{k_1} \log(k_1 x + k_2 \nu) - \frac{4k_2 \nu}{k_1} \log(k_2 \nu)\right) dx\right]^{-1}.$$
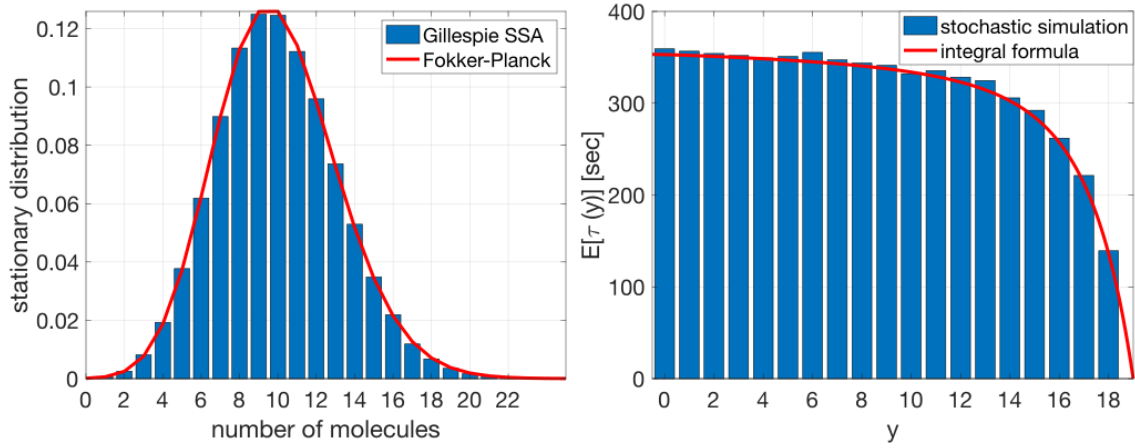
Figure 3.6: Left: Comparison of the stationary distribution Equation 3.35 and estimated stationary distribution from simulations of the process Equation 3.29. Right: Average hitting time of 19 molecules as a function of the initial condition estimated from direct simulations and the integral formula. Parameters: $k_1 = 0.1$, $k_2\nu = 1$. [Code(s): `CH3_CFP_HT1.m` and `CH3_CFP_HT2.m`]

Figure 3.6 (left) shows excellent agreement between the stationary distribution Equation 3.35 from the chemical Fokker-Planck approximation and estimated stationary distribution obtained via direct simulation of the process Equation 3.29.

Armed with an (approximate) analytic expression for the stationary distribution, we can then turn to estimating mean hitting times for the process Equation 3.29. For example, we can define

$$\tau(y) = \inf\{t > 0 : A(t) = 19, \ A(0) = y\}, \quad y \in \{0, 1, \ldots, 18\}.$$

From the theory developed earlier in this chapter, we have

$$\mathbb{E}[\tau(y)] = \int_y^{19} \frac{1}{d(z)P_s(z)} \int_0^z P_s(x)\,dx\,dz,$$

where P_s is given by Equation 3.35 and the function d is given by Equation 3.34. Figure 3.6 (right) shows $\mathbb{E}[\tau(y)]$ as a function of the initial condition (y) compared to the average hitting time obtained via direct simulation of the process. Once more, theory and numerical experiment are in very close agreement with the chemical Fokker-Planck equation offering an accurate approximation without the computational overhead of direct simulations.

3.4.2 General multi-species framework

Before we can write the general chemical Fokker-Planck equation we need to write down the multi-dimensional analogue of the Fokker-Planck equation Equation 3.11. To this end, consider the N-dimensional SDE

$$\mathbf{X}(t + \Delta t) = \mathbf{X}(t) + \mathbf{f}(\mathbf{X}(t))\Delta t + \mathbf{g}(\mathbf{X}(t))\sqrt{\Delta t}\,\boldsymbol{\xi}, \quad t > 0, \tag{3.36}$$

where $\mathbf{X}(t)$ and $\mathbf{f}(\mathbf{X}(t))$ are N-dimensional vectors, $\mathbf{g}(\mathbf{X}(t))$ is an $N \times N$ matrix and $\boldsymbol{\xi}$ is an N-dimensional vector of independent standard Normal random variables. We define the diffusion tensor to be the matrix $\mathbf{D} = \frac{1}{2}\mathbf{g}(x)\mathbf{g}^T(x)$, i.e.,

$$D_{i,j}(\mathbf{x}) = \frac{1}{2}\sum_{k=1}^{N} g_{i,k}(\mathbf{x})g_{j,k}(\mathbf{x}).$$

The Fokker-Planck equation for the N-dimensional process $\mathbf{X}$ solving Equation 3.36 is given by

$$\frac{\partial}{\partial t}P(\mathbf{x}, t) = -\sum_{i=1}^{N}\frac{\partial}{\partial x_i}\left[f_i(\mathbf{x})P(\mathbf{x}, t)\right] + \sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\partial^2}{\partial x_i \partial x_j}\left[D_{i,j}(\mathbf{x})P(\mathbf{x}, t)\right].$$

Now suppose that we have a chemical reaction process in which:

- we have $N \geq 1$ chemical species that can undergo $q \geq 1$ distinct reactions,
- the N-dimensional vector $\mathbf{X}(t)$ will represent the state of the system at time t, i.e., $X_i(t)$ is the number of molecules of species i at time t,
- $\alpha_j(\mathbf{x})$ is the propensity function associated with the jth reaction, where $\mathbf{x} \in \mathbb{R}^N$ is the current state of the system,
- $v_{j,i}$ is the change in the number of molecules in species i, i.e., X_i, that occurs when reaction j takes place.

The chemical master equations for this reaction process can be written in the form:

$$\frac{\partial}{\partial t}P(\mathbf{x}, t) = \sum_{j=1}^{q}\left(\alpha_j(\mathbf{x} - \mathbf{v}_j)P(\mathbf{x} - \mathbf{v}_j, t) - \alpha_j(\mathbf{x})P(\mathbf{x}, t)\right),$$

where $P(\mathbf{x}, t) = \mathbb{P}[\mathbf{X}(t) = \mathbf{x}]$.

Carrying out the same large molecule number approximation as we did for the production-degradation example, we can obtain the multi-dimensional chemical Fokker-Planck equations:

$$\frac{\partial}{\partial t}P(\mathbf{x}, t) = -\sum_{i=1}^{N}\sum_{j=1}^{q}\frac{\partial}{\partial x_i}\left[v_{j,i}\alpha_j(\mathbf{x})P(\mathbf{x}, t)\right]$$

$$+ \frac{1}{2}\sum_{i=1}^{N}\sum_{k=1}^{N}\frac{\partial^2}{\partial x_i \partial x_k}\left[\sum_{j=1}^{q}v_{j,i}v_{j,k}\alpha_j(\mathbf{x})P(\mathbf{x}, t)\right].$$

Knowledge checklist

Key topics:

1. Stochastic differential equations (computational definition and SSAs)
2. The Fokker-Planck equation (stationary distribution, boundary conditions)
3. The Kolmogorov Backward equation (expected hitting times)
4. The Chemical Fokker-Planck equation (approximating chemical reaction processes)

Key skills:

- Use the computational definition of an SDE to compute moments and other related quantities.
- State and derive the Fokker-Planck equation for a given SDE, including applying the appropriate boundary conditions.
- Calculate stationary distributions of SDEs using the steady-state Fokker-Planck equation.
- Compute hitting times for SDEs.
- Explain the impact of boundary conditions on stationary distributions and hitting times.
- Explain how SSAs for SDEs can be modified to account for a range of boundary conditions that occur in biological applications (e.g., reflecting boundaries).
- Write chemical Fokker-Planck approximation for a given chemical reaction process.

4 Reaction-Diffusion Processes

Up to this point, we have considered well-mixed chemical reaction processes and this assumption allowed us to neglect any spatial structure in our models. However, many biological phenomena generate spatial pattern and structure; moreover, this spatial structure is often essential to the function of the system. Hence we need to extend our toolkit to include classes of stochastic processes with explicit spatial extent. To this end, we will introduce several stochastic models of the simplest type of spatial movement: diffusion. We then extend our diffusion processes to reaction-diffusion processes in which particles can not only move, but can also react with one another to exhibit more complex dynamics. We illustrate this theory with some applications to pattern formation in biological and ecological systems.

4.1 Modelling diffusion with SDEs

Our first approach to modelling diffusive motion of a particle (or cell, animal, insect, etc.) is to directly model the dynamics of the position of the particle with a stochastic differential equation (SDE). For example, we can denote the position of the particle at time t in 3-dimensional space by $(X(t), Y(t), Z(t))$. If $D > 0$ is the diffusion coefficient, the dynamics of the particle are given by

$$
\begin{aligned}
X(t + \Delta t) &= X(t) + \sqrt{2D\Delta t}\,\xi_x, \\
Y(t + \Delta t) &= Y(t) + \sqrt{2D\Delta t}\,\xi_y, \\
Z(t + \Delta t) &= Z(t) + \sqrt{2D\Delta t}\,\xi_z,
\end{aligned}
\tag{4.1}
$$

where ξ_x, ξ_y and ξ_z are three independent sequences of standard Normal random variables. The x, y and z coordinates of the particle are thus fully independent of one another and we are really just considering three separate one dimensional processes; we could construct a diffusion process on $\mathbb{R}^d$ in exactly the same way.

The model of diffusion given by Equation 4.1 is continuous in both space and time (although we choose to represent it using our computational definition). We can thus write down the Fokker-Planck equation(s) for the probability density function for the position of the particle directly from Equation 4.1, i.e.

$$\frac{\partial}{\partial t} P_x(\eta_x, t) = D \, \frac{\partial^2}{\partial \eta_x^2} P_x(\eta_x, t),$$

$$\frac{\partial}{\partial t} P_y(\eta_y, t) = D \, \frac{\partial^2}{\partial \eta_y^2} P_y(\eta_y, t), \tag{4.2}$$

$$\frac{\partial}{\partial t} P_z(\eta_z, t) = D \, \frac{\partial^2}{\partial \eta_z^2} P_z(\eta_z, t),$$

where $P_x(\eta_x, t)$ denotes the marginal density of the x coordinate of the process at time t. Since the three components of the process are independent, we can obtain the joint density function of the particle's position, $P(\eta_x, \eta_y, \eta_z, t)$, by multiplication, i.e.

$$P(\eta_x, \eta_y, \eta_z, t) = P_x(\eta_x, t) \, P_y(\eta_y, t) \, P_z(\eta_z, t).$$

Hence we can sum the equations in Equation 4.2 to obtain the Fokker-Planck equation for the joint density of the position:

$$\frac{\partial}{\partial t} P = D \left(\frac{\partial^2}{\partial \eta_x^2} P + \frac{\partial^2}{\partial \eta_y^2} P + \frac{\partial^2}{\partial \eta_z^2} P \right),$$

which is just the heat equation in three-dimensional space. It is straightforward to show that

$$P(\eta_x, \eta_y, \eta_z, t) = \frac{1}{(4D\pi t)^{3/2}} \exp\left(\frac{-\eta_x^2 - \eta_y^2 - \eta_z^2}{4Dt} \right). \tag{4.3}$$

In order to visualize this PDF, we can integrate out the z component to obtain the joint density of the x and y components:

$$P(\eta_x, \eta_y, t) = \int_{\mathbb{R}} \frac{1}{(4D\pi t)^{3/2}} \exp\left(\frac{-\eta_x^2 - \eta_y^2 - \eta_z^2}{4Dt} \right) d\eta_z = \frac{1}{4D\pi t} \exp\left(\frac{-\eta_x^2 - \eta_y^2}{4Dt} \right). \tag{4.4}$$

Figure 4.1 shows the two-dimensional density Equation 4.4 at time $t = 10$ (left). From the previous chapter, we know how to simulate the process Equation 4.1 using the computational definition and Figure 4.1 shows the results of estimating the joint density of $(X(t), Y(t))$ via this approach. To estimate the joint density via simulations, we need to discretize space, count the number of sample paths finishing in each box at $t = 10$ (see Figure 4.1) and then appropriately normalize these counts for the number of paths and area of the boxes.

We can reduce Equation 4.1 to a one-dimensional model of diffusive motion by just considering the first component, i.e.
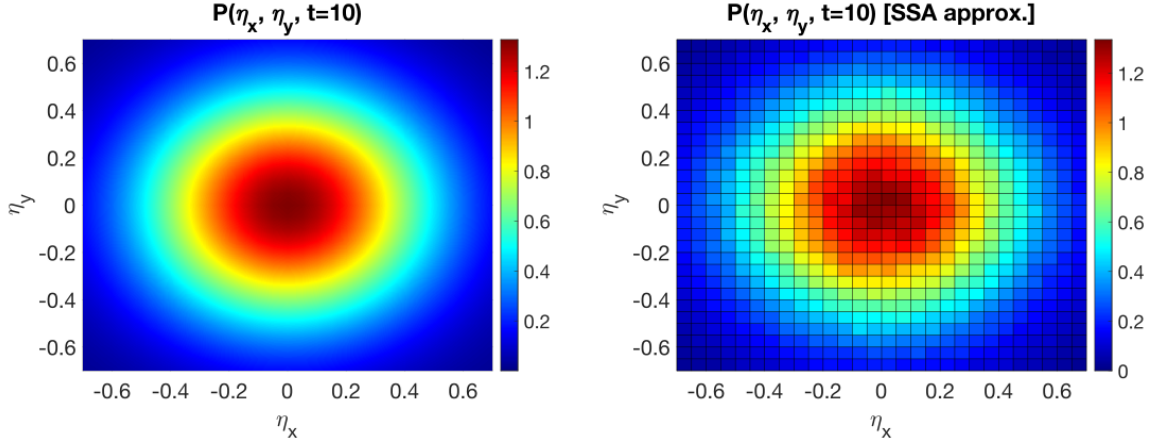
Figure 4.1: Comparison of the 2D density Equation 4.4 at time $t = 10$ (left) with the estimated density computed from direct simulations of the process Equation 4.1 (right). Parameters: $D = 0.0001$. [Code: `CH4_SDE_diffusion.m`]

$$X(t + \Delta t) = X(t) + \sqrt{2D\Delta t}\,\xi. \tag{4.5}$$

The cells or other particles we wish to model often cannot move freely throughout the spatial domain and hence we sometimes need to impose barriers on the domain of $X(t)$ to reflect physical reality. Suppose we want to restrict the domain of $X(t)$ to the interval $[0, L]$ for some $L > 0$ by imposing reflecting boundaries at 0 and L. We can simulate sample paths of Equation 4.5 with these two reflecting boundaries using a variation on the SSA introduced in the previous chapter:

At time $t = 0$, set $X(0)$, then:

1. **Step 1:** Generate $\xi \sim N(0, 1)$.

2. **Step 2:** Compute the possible value of $X(t + \Delta t)$:

$$X(t + \Delta t) = X(t) + \sqrt{2D\,\Delta t}\,\xi,$$

3. **Step 3:** If $X(t + \Delta t) \in [0, L]$, then we accept that value.

   Otherwise, if $X(t + \Delta t) < 0$, then set

$$X(t + \Delta t) = -X(t) - \sqrt{2D\,\Delta t}\,\xi.$$

85

If $X(t + \Delta t) > L$, then set

$$X(t + \Delta t) = X(t) + 2(L - X(t)) - \sqrt{2D\,\Delta t}\,\xi.$$

Finally, set $t = t + \Delta t$ and go back to Step 1.

The reflection steps in the SSA above are highlighted in **blue** with the first instance being a reflection of the process through zero and the second a reflection through the line $\eta_x = L$. Figure 4.2 A shows some sample paths of the process obtained via the SSA above with reflection at $\eta_x = 0$ and $\eta_x = L = 1$.

We can also analyse the 1D diffusion process Equation 4.1 by solving its associated Fokker-Planck equation

$$\frac{\partial}{\partial t} P_x(\eta_x, t) = D\,\frac{\partial^2}{\partial \eta_x^2} P_x(\eta_x, t), \quad \text{for } \eta_x \in (0, L).$$

We additionally need to supply boundary conditions at $\eta_x = 0$ and $\eta_x = L$ to account for the reflection of the process that ensures $X(t) \in [0, L]$ for all $t \geq 0$. From our discussion of reflecting boundary conditions for the Fokker-Planck equation in Chapter 3, we know that we need to specialise condition Equation 3.25; doing so yields the pair of boundary conditions:

$$\left.\frac{\partial}{\partial \eta_x} P_x(\eta_x, t)\right|_{\eta_x=0} = \left.\frac{\partial}{\partial \eta_x} P_x(\eta_x, t)\right|_{\eta_x=L} = 0 \quad \text{for all } t > 0. \tag{4.6}$$

Figure 4.2 A, B and C show how the probability density of $X(t)$, $P(\eta_x, t)$, evolves from an initial distribution that is a Dirac delta at $\eta_x = 0.4$. The influence of the reflecting boundaries at $\eta_x = 0$ and $\eta_x = 1$ become apparent by time $t = 5$ and eventually the distribution of the particle's position approaches a uniform distribution on $[0, 1]$ for large times. The blue histograms in Figure 4.2 are the estimated probability density from the SSA and we observe relatively good agreement between the exact and estimated PDFs at all times ($2, 500$ paths). We could improve the quality of the SSA approximation by decreasing the step size in the simulations, at additional computational cost, or by narrowing the width of the histogram bins.
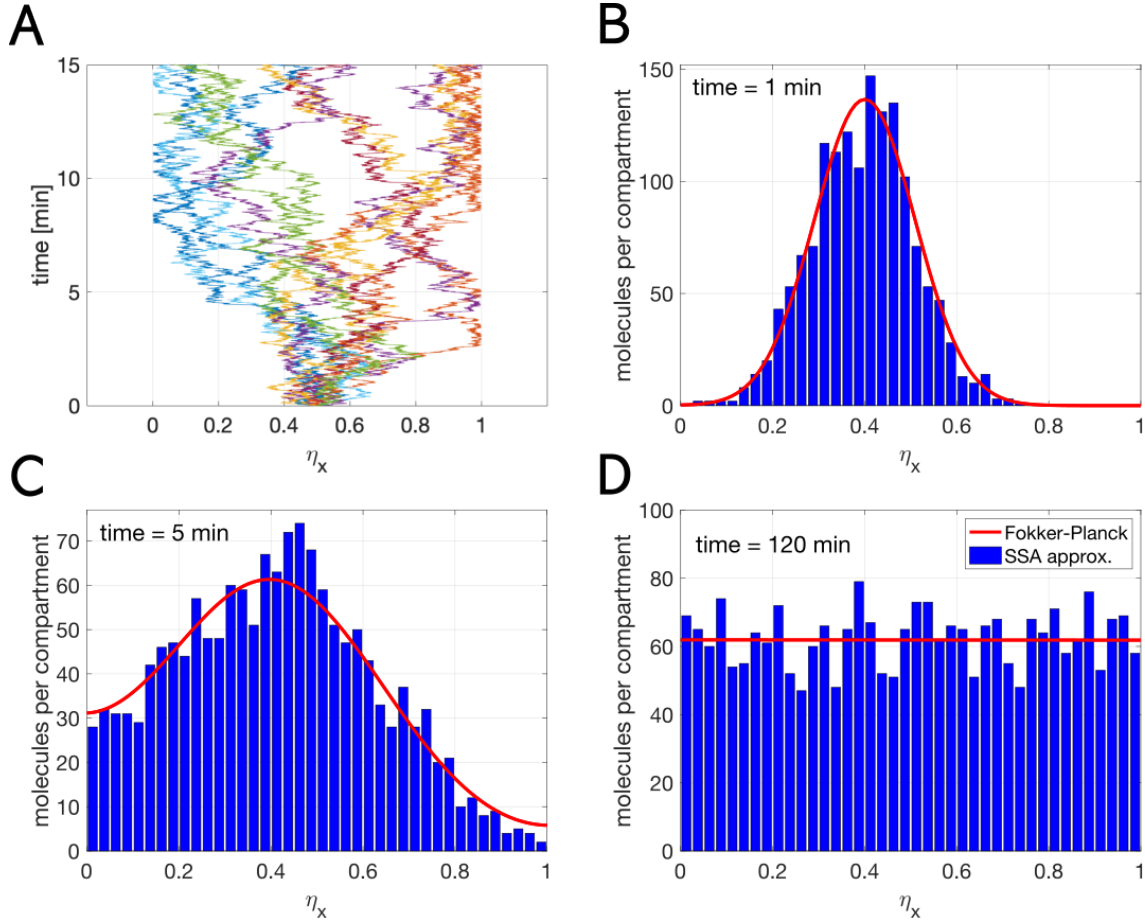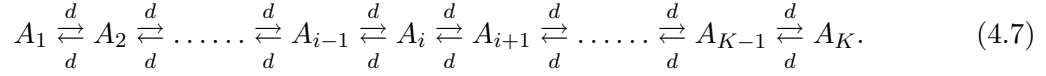
Figure 4.2: A: Sample paths of Equation 4.5 with reflecting boundaries at 0 and $L = 1$. B/C/D: Solution to the 1D Fokker-Planck equation with reflecting (no-flux) boundary conditions as given by Equation 4.6 at times $t = \{1, 5, 120\}$. Parameters: $D = 0.0001$. [Codes: CH4_SDE_reflect.m and CH4_FP_reflect.m]

4.2 A compartmental model of diffusion

In the previous section we modelled diffusive motion by simulating the trajectory of individual particles and then averaging over the trajectories of many particles to understand the properties of the underlying process. We could instead change perspective and consider a collection of particles from the outset and, rather than tracking individual particles, we could record only changes in the number of particles present at each location.

Consider the spatial domain $[0, 1]$ and discretise it into $K = 40$ compartments of equal width, i.e. $h = 1/40$ is our discretisation parameter. Now we can set up a chemical reaction process where each "species" A_i for $i = 1, \ldots, 40$ counts the number of molecules present in compartment i, which is the interval $[(i-1)h, ih]$. Next introduce a parameter $d > 0$ which we call the "hopping rate" that controls the rate at which particles diffuse into neighbouring compartments. We defer for now the discussion of how to choose the discretisation parameter $h > 0$ and the hopping rate d to give a diffusivity rate D corresponding to the diffusion coefficient from the SDE diffusion model Equation 4.1. The dynamics of this process are thus given by:

$$A_1 \underset{d}{\overset{d}{\rightleftarrows}} A_2 \underset{d}{\overset{d}{\rightleftarrows}} \ldots \ldots \underset{d}{\overset{d}{\rightleftarrows}} A_{i-1} \underset{d}{\overset{d}{\rightleftarrows}} A_i \underset{d}{\overset{d}{\rightleftarrows}} A_{i+1} \underset{d}{\overset{d}{\rightleftarrows}} \ldots \ldots \underset{d}{\overset{d}{\rightleftarrows}} A_{K-1} \underset{d}{\overset{d}{\rightleftarrows}} A_K. \tag{4.7}$$

This system has $2(K-1)$ reactions but only K distinct propensities. Let $\alpha_i(t) = dA_i(t)$ and note that this propensity represents two reactions:

$$A_i \overset{d}{\rightarrow} A_{i+1} \text{ (jump right)}, \quad \text{and} \quad A_i \overset{d}{\rightarrow} A_{i-1} \text{ (jump left)}.$$
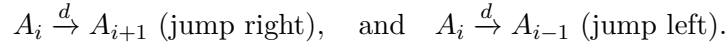
Suppose that we start with N molecules in the system, i.e. $\sum_{i=1}^{K} A_i(0) = N$. Then the total propensity of the system is given by

$$\begin{aligned}
\alpha_0(t) &= \sum_{i=1}^{K-1} \alpha_i(t) + \sum_{i=2}^{K} \alpha_i(t) \\
&= 2 \sum_{i=1}^{K} \alpha_i(t) - \alpha_1(t) - \alpha_K(t) \\
&= 2d \sum_{i=1}^{K} A_i(t) - \alpha_1(t) - \alpha_K(t) \\
&= 2dN - \alpha_1(t) - \alpha_K(t).
\end{aligned} \tag{4.8}$$

We can immediately see how to exploit the additional structure of this process to optimize its simulation via the Gillespie algorithm. In particular, the total number of molecules (N) is always conserved and only two molecule numbers and two propensity functions change at each reaction. Moreover, as Equation 4.8 illustrates, the total propensity function only changes

when there is a reaction impacting compartment 1 or compartment K. Thus we have the following optimized Gillespie algorithm for simulating our compartmental diffusion process:

At time $t = 0$, set the initial molecule numbers in each compartment ($A_i(0)$ for $i = 1, \ldots, K$), then:

1. **Step 1:** Generate $r_1, r_2 \sim U([0, 1])$.
2. **Step 2:** Compute the propensity functions $\alpha_i(t)$ and hence the total propensity function $\alpha_0(t)$ from Equation 4.8.
3. **Step 3:** Compute the time of the next reaction, which takes place at time $t + \tau$, by calculating

$$\tau = \frac{1}{\alpha_0(t)} \log(1/r_1).$$

4. **Step 4:** If $r_2 < \sum_{i=1}^{K-1} \alpha_i / \alpha_0$, then find $j \in \{1, 2, \ldots, K-1\}$ such that

$$r_2 \geq \frac{1}{\alpha_0} \sum_{i=1}^{j-1} \alpha_i \text{ and } r_2 < \frac{1}{\alpha_0} \sum_{i=1}^{j} \alpha_i.$$

   Then adjust the number of molecules as follows:

$$A_j(t + \tau) = A_j(t) - 1, \quad A_{j+1}(t + \tau) = A_{j+1}(t) + 1,$$

   i.e. one molecule in compartment j jumps to the right into compartment $j + 1$.
5. **Step 5:** If $r_2 \geq \sum_{i=1}^{K-1} \alpha_i / \alpha_0$, then find $j \in \{2, \ldots, K\}$ such that

$$r_2 \geq \frac{1}{\alpha_0} \left(\sum_{i=1}^{K-1} \alpha_i + \sum_{i=2}^{j-1} \alpha_i \right) \text{ and } r_2 < \frac{1}{\alpha_0} \left(\sum_{i=1}^{K-1} \alpha_i + \sum_{i=2}^{j} \alpha_i \right).$$

   Then adjust the number of molecules as follows:

$$A_j(t + \tau) = A_j(t) - 1, \quad A_{j-1}(t + \tau) = A_{j-1}(t) + 1,$$

   i.e. one molecule in compartment j jumps to the left into compartment $j - 1$.
6. **Step 6:** Set $t = t + \tau$ and go back to Step 1.

Figure 4.3 below shows some simulations of the compartmental diffusion process in panel A. Note that we have not explicitly implemented a reflecting boundary but particles cannot move left from compartment A_1 or right from compartment A_K and so the process is naturally

confined to the interval $[0,1]$. In panels B, C and D of Figure 4.3 we plot the estimated probability density of the process at various times and compare it to the predictions of the corresponding Fokker-Planck equations for the analogous continuum (SDE) model of diffusion. We see excellent agreement between the compartmental and continuum models because we chose the diffusion coefficient $D = d \times h^2$, where d is the compartmental hopping rate and h is our discretisation parameter. But how do we know that this is the correct relationship between the parameters of the discrete and continuous models of diffusion?

To understand how our discrete and continuous models relate to one another, we need to write down the chemical master equations for the compartmental model. To this end, let $p(\mathbf{n}, t) = \mathbb{P}[\mathbf{A}(t) = \mathbf{n}]$ denote the joint probability mass function of the compartmental process. Next define the discrete shift operators R_i, $L_i : \mathbb{N}^K \mapsto \mathbb{N}^K$ by

$$R_i : [n_1, \ldots, n_i, n_{i+1}, \ldots, n_k] \mapsto [n_1, \ldots, n_i + 1, n_{i+1} - 1, \ldots, n_k], \quad i = 1, 2, \ldots, K-1,$$
$$L_i : [n_1, \ldots, n_{i-1}, n_i, \ldots, n_k] \mapsto [n_1, \ldots, n_{i-1} - 1, n_i + 1, \ldots, n_k], \quad i = 2, 3, \ldots, K.$$

The chemical master equations for the process Equation 4.7 are thus given by:

$$
\begin{aligned}
\frac{d}{dt}p(\mathbf{n}) = & \; d \sum_{j=1}^{K-1} \left((n_j + 1)p(R_j\mathbf{n}) - n_j p(\mathbf{n}) \right) \\
& + d \sum_{j=2}^{K} \left((n_j + 1)p(L_j\mathbf{n}) - n_j p(\mathbf{n}) \right).
\end{aligned}
\tag{4.9}
$$

We can then define the mean of the process Equation 4.7 by $\mathbf{M}(t) = [M_1(t), \ldots, M_K(t)]$ where $M_i(t)$ denotes the mean number of molecules in compartment i and is given by

$$M_i(t) = \sum_{\mathbf{n}} n_i \, p(\mathbf{n}, t) := \sum_{n_1=0}^{N} \sum_{n_2=0}^{N} \cdots \sum_{n_K=0}^{N} n_i \, p(\mathbf{n}, t).$$

With some work, we can show that the evolution equations for the mean are given by

$$
\begin{aligned}
\frac{d}{dt}M_i(t) &= d \left(M_{i+1} + M_{i-1} - 2M_i \right), \quad i = 2, 3, \ldots K-1, \\
\frac{d}{dt}M_1 &= d(M_2 - M_1), \quad \frac{d}{dt}M_K = d(M_{K-1} - M_K).
\end{aligned}
\tag{4.10}
$$

To compare this to the continuum model of diffusion studied in the previous section, we need to convert the mean number of molecules, M_i, to a concentration. If $c(x, t)$ denotes the concentration of molecules at location x, we can make the approximation
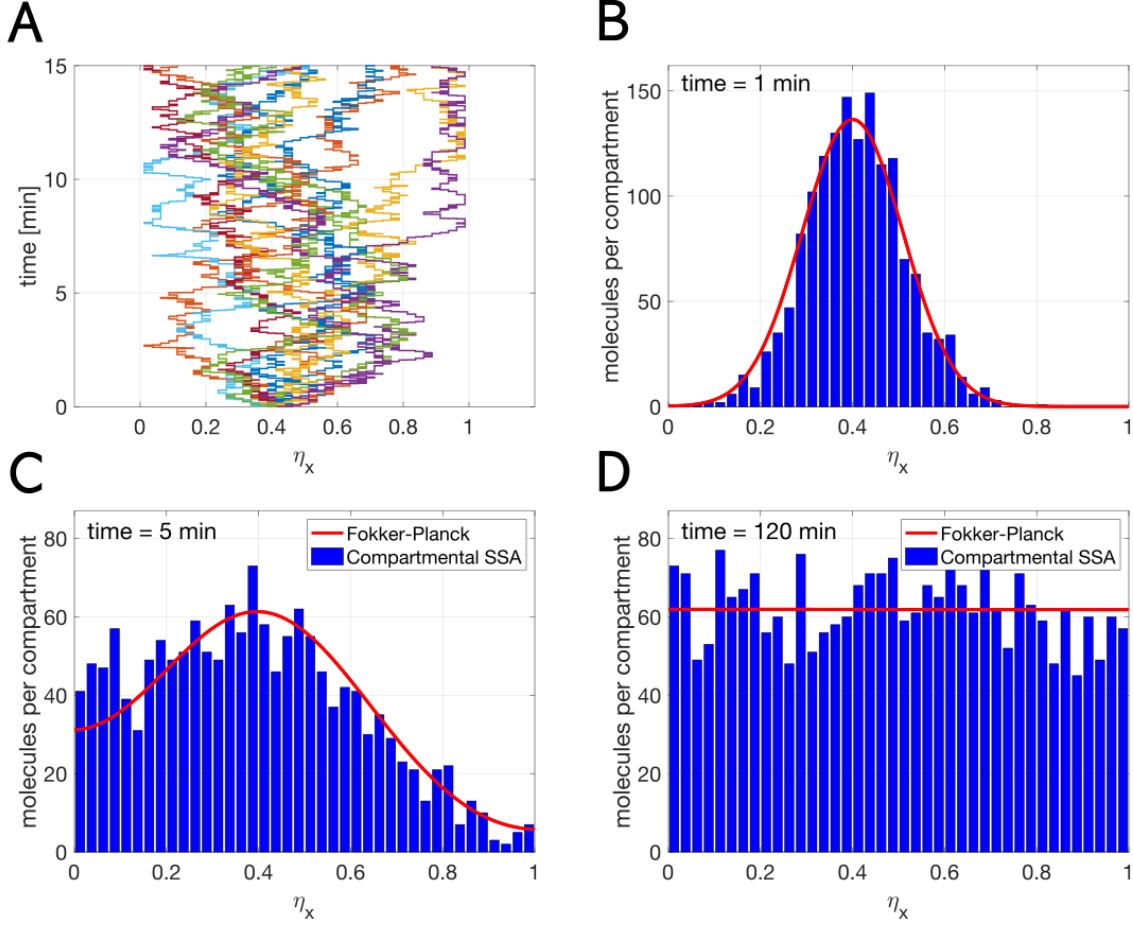
Figure 4.3: A: Sample paths of the compartmental diffusion process with $K = 40$ compartments and $d = D/h^2$. B/C/D: Simulations of the compartmental diffusion process overlaid with solutions to the 1D Fokker-Planck equation with reflecting (no-flux) boundary conditions as given by Equation 4.6 at times $t = \{1, 5, 120\}$. Parameters: $D = 0.0001$, $h = 0.025$. [Codes: `CH4_compartment_diff.m` and `CH4_compartment_FP.m`]

$$c(x_i, t) \approx \frac{M_i(t)}{h}$$

where x_i denotes the center of the ith compartment for $i = 1, \ldots, K$. Using this approximation, divide Equation 4.10 across by h to obtain

$$\frac{\partial}{\partial t}c(x_i, t) = d\left(c(x_i + h, t) + c(x_i - h, t) - 2c(x_i, t)\right).$$

Assuming c is sufficiently smooth, we can Taylor expand the right-hand side to obtain the approximation

$$\frac{\partial}{\partial t}c(x_i, t) \approx dh^2 \frac{\partial^2}{\partial x^2}c(x_i, t).$$

Hence if we choose $d = D/h^2$, where D is the continuum diffusion coefficient, we have

$$\frac{\partial}{\partial t}c(x_i, t) = D\frac{\partial^2}{\partial x^2}c(x_i, t),$$

and the evolution equation for the molecule concentration will match the corresponding Fokker-Planck equation of the SDE model of diffusion. In other words, the compartmental and continuum diffusion models will agree on average.

To fully characterize the distribution of the compartmental diffusion process we additionally need to understand the variance of the process. The variance vector of the process Equation 4.7 is given by $\mathbf{V}(t) = [V_1(t), \ldots, V_K(t)]$, where

$$V_i(t) = \sum_{\mathbf{n}} (n_i - M_i(t))^2 \, p(\mathbf{n}, t) := \sum_{n_1=0}^{N} \sum_{n_2=0}^{N} \cdots \sum_{n_K=0}^{N} (n_i - M_i(t))^2 \, p(\mathbf{n}, t).$$

Deriving a system of evolution equations for the variances requires us to define more generally the covariance matrix $V_{i,j}$ by

$$V_{i,j} = \sum_{\mathbf{n}} n_i n_j p(\mathbf{n}, t) - M_i M_j, \quad i, j = 1, \ldots, K. \tag{4.11}$$

The diagonal entries of the covariance matrix Equation 4.11 are the variances defined above. Multiplying the chemical master equations by n_i^2 and summing over $\mathbf{n}$ yields

$$\frac{d}{dt} \sum_{\mathbf{n}} n_i^2 p(\mathbf{n}, t) = d \sum_{j=1}^{K-1} \left(\sum_{\mathbf{n}} n_i^2 (n_j + 1) p(R_j \mathbf{n}) - \sum_{\mathbf{n}} n_i^2 n_j p(\mathbf{n}, t) \right)$$

$$+ d \sum_{j=2}^{K} \left(\sum_{\mathbf{n}} n_i^2 (n_j + 1) p(L_j \mathbf{n}) - \sum_{\mathbf{n}} n_i^2 n_j p(\mathbf{n}, t) \right). \tag{4.12}$$

Suppose $i \in \{2, \ldots, K-1\}$. Evaluate the first term on the right-hand side of Equation 4.12 for $j = i$ and by changing indices $R_i \mathbf{n} \mapsto \mathbf{n}$:

$$\sum_{\mathbf{n}} n_i^2 (n_i + 1) p(R_j \mathbf{n}) - \sum_{\mathbf{n}} n_i^2 n_i p(\mathbf{n}, t) = \sum_{\mathbf{n}} (n_i - 1)^2 (n_i) p(\mathbf{n}) - \sum_{\mathbf{n}} n_i^2 n_i p(\mathbf{n}, t)$$

$$= \sum_{\mathbf{n}} (-2n_i^2 + n_i) p(\mathbf{n}) = -2V_i - 2M_i^2 + M_i.$$

Now we need to deal with the off-diagonal terms. The term corresponding to $j = i - 1$ in the first sum can be rewritten as

$$\sum_{\mathbf{n}} n_i^2 (n_{i-1} + 1) p(R_{j-1} \mathbf{n}) - \sum_{\mathbf{n}} n_i^2 n_{i-1} p(\mathbf{n}, t) = \sum_{\mathbf{n}} (2n_i n_{i-1} + n_{i-1}) p(\mathbf{n})$$

$$= 2V_{i,i-1} + 2M_i M_{i-1} + M_{i-1}.$$

Every other term in the first sum on the right-hand side of Equation 4.12 with $j \notin \{i, i-1\}$ is zero. The second sum on the right-hand side of Equation 4.12 is handled analogously by addressing first the cases $j = i$ and $j = i + 1$, and showing that all other terms are zero. Carrying out the requisite algebra yields

$$\frac{d}{dt} \sum_{\mathbf{n}} n_i^2 p(\mathbf{n}, t) = d \left(2V_{i,i-1} + 2M_i M_{i-1} + M_{i-1} - 2V_i - 2M_i^2 + M_i \right)$$

$$+ d \left(2V_{i,i+1} + 2M_i M_{i+1} + M_{i+1} - 2V_i - 2M_i^2 + M_i \right).$$

Finally, we arrive at the following set of equations for the variances of the compartments:

$$\frac{d}{dt} V_i = 2d \left(V_{i,i+1} + V_{i,i-1} - 2V_i \right) + d \left(M_{i+1} + M_{i-1} + 2M_i \right), \quad i = 2, \ldots, K-1,$$

$$\frac{d}{dt} V_1 = 2d \left(V_{1,2} - V_1 \right) + d \left(M_2 + M_1 \right), \tag{4.13}$$

$$\frac{d}{dt} V_K = 2d \left(V_{K,K-1} - V_K \right) + d \left(M_{K-1} + M_K \right),$$

However, the system Equation 4.13 is not a fully closed set of equations because it involves covariance terms of the form $V_{i,j}$ for $i \neq j$. Hence we need to also write down evolution equations for the off-diagonal covariance matrix terms; this can be done by repeating variations on the arguments above.

We can also consider the steady state versions of the mean and variance equations to study the long-term behaviour of the compartmental diffusion process. Doing so yields

$$\bar{M}_i = \frac{N}{K} \quad \text{for } i = 1, \ldots, K,$$

and

$$\bar{V}_i = \frac{N}{K} - \frac{N}{K^2}, \quad \bar{V}_{i,j} = \frac{-N}{K^2}, \quad i \neq j.$$

In fact, solving the chemical master equations directly shows that the stationary distribution of the compartmental diffusion process is a multinomial distribution, i.e.

$$p_s(\mathbf{n}) = \frac{C}{n_1! n_2! \cdots n_K!},$$

for an appropriate normalising constant $C > 0$.

4.3 SSAs for reaction-diffusion processes

4.3.1 Compartmental models

We begin with a discrete-space approach that extends our compartmental model of diffusion to incorporate simple reaction dynamics (zero and first order reactions). We begin by defining an essentially one-dimensional spatial domain

$$\Omega = [0, L] \times [0, h] \times [0, h]$$

where we assume $h \ll L$ so that spatial dynamics are only occurring in the x dimension. Formally, we can assume that we have periodic boundary conditions in the y and z directions. Next divide the interval $[0, L]$ into K compartments of width h, as we did before, and let $A_i(t)$ denote the number of molecules present in compartment i at time t. Note that compartment i is the box $[(i-1)h, ih] \times [0, h] \times [0, h]$ and hence the volume of compartment i is h^3. The dynamics of the process will be as follows:

$$A_1 \overset{d}{\underset{d}{\rightleftarrows}} A_2 \overset{d}{\underset{d}{\rightleftarrows}} \ldots \ldots \overset{d}{\underset{d}{\rightleftarrows}} A_{i-1} \overset{d}{\underset{d}{\rightleftarrows}} A_i \overset{d}{\underset{d}{\rightleftarrows}} A_{i+1} \overset{d}{\underset{d}{\rightleftarrows}} \ldots \ldots \overset{d}{\underset{d}{\rightleftarrows}} A_{K-1} \overset{d}{\underset{d}{\rightleftarrows}} A_K,$$

$$A_i \overset{k_1}{\longrightarrow} \emptyset, \qquad i = 1, \ldots, K, \quad \emptyset \overset{k_2}{\longrightarrow} A_i, \quad i = 1, \ldots, K/5.$$

(4.14)
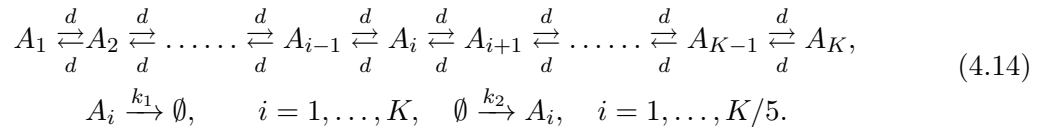
To summarize the dynamics above:

- Molecules diffuse at rate d between all compartments (and we can choose $d = D/h^2$ in order to relate this model to a continuum diffusion approach),
- molecules are produced in the subdomain $[0, L/5]$ at rate k_2 per compartment,
- all molecules in all compartments are subject to degradation at rate k_1.

In order to analyse the dynamics of the process defined by the system of reactions Equation 4.14, we need to write down the reaction-diffusion (RD) master equations for the process. The RD master equations are:

$$
\begin{aligned}
\frac{d}{dt}p(\mathbf{n}) = {} & d\sum_{j=1}^{K-1}\left((n_j+1)p(R_j\mathbf{n}) - n_j p(\mathbf{n})\right) \\
& + d\sum_{j=2}^{K}\left((n_j+1)p(L_j\mathbf{n}) - n_j p(\mathbf{n})\right) \\
& + k_1\sum_{i=1}^{K}\left((n_i+1)p(n_1,\ldots,n_i+1,\ldots,n_k) - n_i p(\mathbf{n})\right) \\
& + k_2 h^3\sum_{i=1}^{K/5}\left(p(n_1,\ldots,n_i-1,\ldots,n_k)p(\mathbf{n})\right).
\end{aligned}
\tag{4.15}
$$

The first two terms on the right-hand side of Equation 4.15 represent diffusive reactions, the third term represents the degradation events and the final term captures the impact of production reactions in the subdomain $[0, L/5]$. We can once more define the mean vector $\mathbf{M}(t) = [M_1(t), \ldots, M_K(t)]$ and write down a set of evolution equations for the components of $\mathbf{M}(t)$. Using the fact that all reactions in Equation 4.14 are either zero or first order, we can directly write down the evolution equations for the stochastic means:

$$
\begin{aligned}
\frac{d}{dt}M_1 &= d(M_2 - M_1) + k_2 h^3 - k_1 M_1, \\
\frac{d}{dt}M_i &= d\left(M_{i+1} + M_{i-1} - 2M_i\right) + k_2 h^3 - k_1 M_i, \quad i = 2, 3, \ldots K/5, \\
\frac{d}{dt}M_i &= d\left(M_{i+1} + M_{i-1} - 2M_i\right) - k_1 M_i, \quad i = K/5+1, \ldots K-1, \\
\frac{d}{dt}M_K &= d(M_{K-1} - M_K) - k_1 M_K.
\end{aligned}
$$

To better compare the stochastic means with a spatially continuous approximation, we let $\tilde{M}_i(t) = M_i(t)/h^3$ to move from molecule numbers to molecule concentrations. This volumetric scaling and choosing $d = D/h^2$ yields:

$$\frac{d}{dt}\tilde{M}_1 = D\frac{\tilde{M}_2 - \tilde{M}_1}{h^2} + k_2 - k_1\tilde{M}_1,$$

$$\frac{d}{dt}\tilde{M}_i = D\frac{\tilde{M}_{i+1} + \tilde{M}_{i-1} - 2\tilde{M}_i}{h^2} + k_2 - k_1\tilde{M}_i, \quad i = 2, 3, \ldots K/5,$$

$$\frac{d}{dt}\tilde{M}_i = D\frac{\tilde{M}_{i+1} + \tilde{M}_{i-1} - 2\tilde{M}_i}{h^2} - k_1\tilde{M}_i, \quad i = K/5 + 1, \ldots K - 1, \tag{4.16}$$

$$\frac{d}{dt}\tilde{M}_K = D\frac{\tilde{M}_{K-1} - \tilde{M}_K}{h^2} - k_1\tilde{M}_K.$$

From the definition of the derivative, we have the following first-order finite-difference approximation:

$$\frac{d}{dx}f(x) = \lim_{h\downarrow 0}\frac{f(x+h) - f(x)}{h} \approx \Delta f(x) := \frac{f(x+h) - f(x)}{h}.$$

Hence we can make the same finite-difference approximation of the second derivative:

$$\frac{d^2}{dx^2}f(x) \approx \Delta^2 f(x) = \frac{f(x+h) - 2f(x) + f(x-h)}{h^2}.$$

Therefore we can interpret the evolution equations Equation 4.16 as a finite-difference approximation of the following (deterministic) reaction-diffusion PDE:

$$\frac{\partial}{\partial t}m(x,t) = D\frac{\partial^2}{\partial x^2}m(x,t) + k_2 \mathbf{1}_{\{x \in [0, L/5]\}} - k_1 m(x,t), \quad x \in (0, L), \tag{4.17}$$

with reflecting boundary conditions:

$$\left.\frac{\partial}{\partial x}m(x,t)\right|_{x=0} = \left.\frac{\partial}{\partial x}m(x,t)\right|_{x=L} = 0,$$

where $\mathbf{1}$ denotes the indicator function.

Figure 4.4 below shows the results of a single simulation of the reaction-diffusion process Equation 4.14 with the estimated PMF plotted at different times in blue. The solution to the approximating reaction-diffusion PDE Equation 4.17 is plotted for comparison in red and we observe quite close agreement between the two models, even for a single realisation of the stochastic process. Of course, the reaction-diffusion approximation is only expected to capture the mean behaviour and will not model fluctuations about the mean of the process.
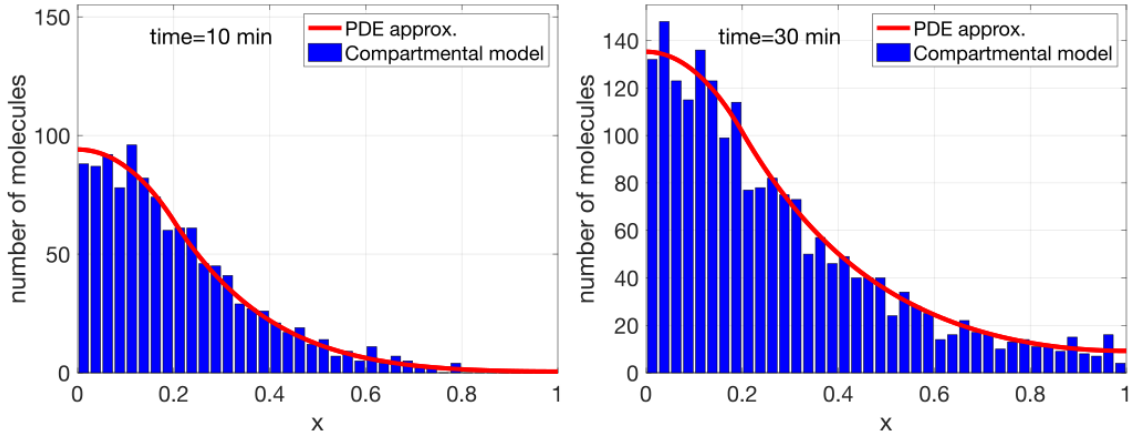
Figure 4.4: Simulations of the compartmental reaction-diffusion process Equation 4.14 with the estimated PMF in blue and the PDE approximation Equation 4.17 shown in red. Parameters: $D = 0.0001$, $h = 0.01$, $k_1 = 0.001$, $k_2 = 0.3125$, $L = 1$. [Code: `CH4_compartment_RD.m`]

4.3.2 SDE-based reaction-diffusion models

As with the modelling of diffusion, we could take the alternative approach of tracking the movements and reactions of individual particles, rather than working with a discretised spatial domain of compartments. To illustrate this alternative approach, we will build a spatially continuous model of the compartmental reaction-diffusion process Equation 4.14. We will once more consider the spatial domain to be $\Omega = [0, L] \times [0, h] \times [0, h]$ with $h \ll L$ so that we only model spatial movement in the x dimension. We now want the continuum model to capture the following attributes:

- Molecules diffuse at rate D on $[0, L]$ with reflecting boundaries at $x = 0$ and $x = L$,
- molecules are produced in the subdomain $[0, L/5]$ at rate k_2,
- all molecules are subject to degradation at rate k_1.

To simulate the stochastic process outlined above, we will discretise time (rather than space) by introducing a finite-size time step $\Delta t > 0$. The spatial dynamics of each particle X_i will be given by the following diffusive SDE:

$$X_i(t + \Delta t) = X_i(t) + \sqrt{2D\Delta t}\, \xi. \tag{4.18}$$

Then at each time t we will evolve the process forward in time by carrying out the following steps:

- **Step 1:** For each molecule, compute its x position at time $t+\Delta t$ according to Equation 4.18 (obeying the reflecting boundary conditions at $x = 0$ and $x = L$).
- **Step 2:** For each molecule, generate $r_1 \sim U([0,1])$. If $r_1 < k_1\Delta t$, remove the molecule (degradation reaction occurs).
- **Step 3:** Generate $r_2 \sim U([0,1])$. If $r_2 < (k_2 h^2 L/5)\Delta t$, then generate $r_3 \sim U([0,1])$ and introduce a new molecule at position $x = r_3 L/5$ (production reaction occurs).

We don't further specify the details of step 1, since we have previously detailed the algorithm to simulate the diffusive SDE Equation 4.18.

If we simulate the reaction-diffusion process using this algorithm and estimate the probability density of the number of molecules we obtain results virtually indistinguishable to those shown in Figure 4.4 (assuming we choose $D = dh^2$). But why should this new algorithm produce statistically accurate realisations of the diffusion-production-degradation process?

The algorithm above essentially couples our computational definition of SDEs to the "naive" SSA for chemical reaction systems that we introduced at the very beginning of the course. If we choose Δt sufficiently small that $\Delta t k_1 \ll 1$, then we correctly reproduce the degradation reaction in step 2, because

$$\mathbb{P}[\text{molecule } i \text{ gets degraded in } [t,\, t + \Delta t]] \approx k_1 \Delta t + O(\Delta t^2).$$

Similarly, new molecules are produced in the region $[0, L/5]$, which has volume $h^2 L/5$, and so

$$\mathbb{P}[\text{new molecule produced in } [t,\, t + \Delta t]] \approx k_2 h^2 L/5 \Delta t + O(\Delta t^2).$$

As with the degradation reactions, we will need to choose the time step sufficiently small to ensure accuracy of the algorithm. In this case, we need $k_2 h^2 L/5\Delta t \ll 1$ and this can also be achieved by choosing h sufficiently small since h is the spatial discretisation parameter. Finally, if a molecule is produced at a given time step, then it was equally likely to have been produced at any location in $[0, L/5]$ and so we simply assign it to a uniformly random location in $[0, L/5]$ in the second part of step 3.

Note that the ordering of the steps in the algorithm here does not matter. If we move molecule i in the SDE updating step, but then that molecule gets degraded at the same time step, it does not impact the dynamics of any other molecule. We also check the degradation reaction for **every** molecule at each time step so it is possible for multiple molecules to be degraded at each time step but this is not a problem since the molecules are not interacting with each other! In other words, we don't have to worry about a case in which two molecules get degraded in one time step but the presence of one of these molecules would have influenced the subsequent dynamics of the other; we will come back to these potential issues in the next section.

One other difference to this continuum space approach in which we track molecule trajectories is that we need to dynamically track the number of particles. From a computational perspective, the size of our system is not fixed and we need to dynamically update the number of SDEs that we are solving and discontinue any SDEs that correspond to molecules that have been degraded. We also need to introduce and solve a new SDE when a new molecule enters the system via a production reaction. There are numerous ways to handle this complication during the implementation of the algorithm but it is worth factoring in when comparing the computational cost of alternative modelling approaches.

4.3.3 Reaction-Diffusion processes with higher-order reactions

In the preceding sections, we introduced a discrete space and a continuous space approach to analysing and simulating stochastic reaction-diffusion processes. In both cases, we restricted our examples to zero and first order reactions, and both approaches were straightforward to implement. Of course, most models of interest in biology will involve higher-order reactions (i.e. interactions between at least two different species) and this forces us to answer some questions which simply did not arise with only zero or first order reactions. In particular, if two molecules are required for a reaction to occur, how close do they need to be in space for the reaction to occur? Should the reaction probability depend on the distance between molecules or always occur if they are "close enough"?

Consider the following illustrative example of a compartmental reaction-diffusion process: There are two types of molecules, type A and type B, present in the pseudo-one-dimensional domain

$$\Omega = [0, L] \times [0, h] \times [0, h],$$

and as usual we only model their movement in the x dimension with the assumption that $h \ll L$. The domain $[0, L]$ is discretised into K compartments of width h. The number of A molecules in compartment i, the interval $[(i-1)h, ih]$, at time t is given by "species" $A_i(t)$ and, similarly, $B_i(t)$ denotes the number of B molecules in compartment i. In terms of reactions, posit that the process has the following dynamics:

- A molecules move to adjacent compartment at hopping rate d_A ($d_A = D_A/h^2$) and B molecules have hopping rate d_B ($d_B = D_B/h^2$),
- **molecules can only react with other molecules in their own compartment**,
- Two A molecules react at rate k_1 to produce a product not of interest,
- An A and a B molecule react at rate k_2 to produce a product not of interest,
- A molecules are degraded at rate k_3 in $[0, L]$,
- B molecules are degraded at rate k_4 in $[0, L]$,
- A molecules are produced at rate k_5 in $[0, L]$,
- B molecules are produced at rate k_6 in $[3L/5, L]$,

The second bullet point above is making explicit the crucial assumption about which molecules can react with each other, i.e. only those within the same spatial compartment. This is essentially assuming that molecules are well-mixed within each compartment, so that the process has two different spatial scales. The macro scale is the interval $[0, L]$ and the molecules are not well-mixed on this scale (we can have spatial pattern), but on the micro scale (within a compartment) molecules are well-mixed spatially and there is no spatial structure within the compartments.

In our usual reaction notation, we can represent this process as follows:

$$A_1 \underset{d_A}{\overset{d_A}{\rightleftarrows}} A_2 \underset{d_A}{\overset{d_A}{\rightleftarrows}} \ldots \ldots \underset{d_A}{\overset{d_A}{\rightleftarrows}} A_{i-1} \underset{d_A}{\overset{d_A}{\rightleftarrows}} A_i \underset{d_A}{\overset{d_A}{\rightleftarrows}} A_{i+1} \underset{d_A}{\overset{d_A}{\rightleftarrows}} \ldots \ldots \underset{d_A}{\overset{d_A}{\rightleftarrows}} A_{K-1} \underset{d_A}{\overset{d_A}{\rightleftarrows}} A_K,$$

$$B_1 \underset{d_B}{\overset{d_B}{\rightleftarrows}} B_2 \underset{d_B}{\overset{d_B}{\rightleftarrows}} \ldots \ldots \underset{d_B}{\overset{d_B}{\rightleftarrows}} B_{i-1} \underset{d_B}{\overset{d_B}{\rightleftarrows}} B_i \underset{d_B}{\overset{d_B}{\rightleftarrows}} B_{i+1} \underset{d_B}{\overset{d_B}{\rightleftarrows}} \ldots \ldots \underset{d_B}{\overset{d_B}{\rightleftarrows}} B_{K-1} \underset{d_B}{\overset{d_B}{\rightleftarrows}} B_K,$$

$$A_i + A_i \xrightarrow{k_1} \emptyset, \quad A_i + B_i \xrightarrow{k_2} \emptyset, \quad i = 1, 2, \ldots, K,$$

$$A_i \xrightarrow{k_3} \emptyset, \quad B_i \xrightarrow{k_4} \emptyset, \quad \emptyset \xrightarrow{k_5} A_i, \quad i = 1, \ldots, K,$$

$$\emptyset \xrightarrow{k_6} B_i, \quad i = 3K/5 + 1, \ldots, K.$$

(4.19)
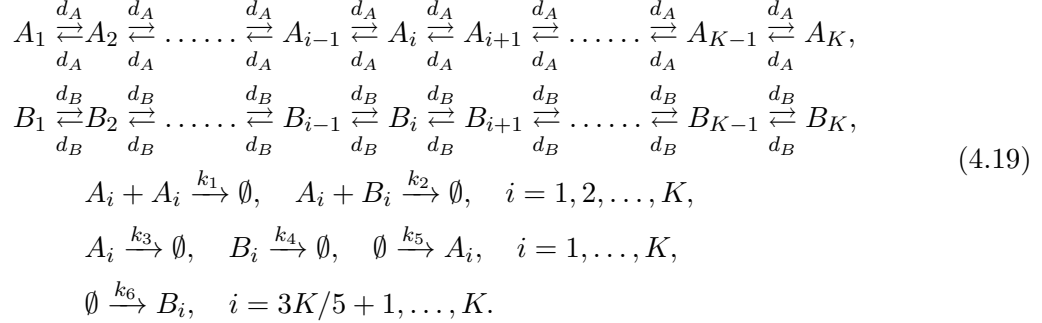
We now have a (potentially very high-dimensional) reaction-diffusion process that can be simulated using the standard Gillespie algorithm. Since we have second-order reactions, we can no longer easily write down a closed system of equations for the stochastic mean of the process. We can however make a law of mass action approximation of the process based on an extension of the mass action principle we used for well-mixed systems. Firstly, let $a(x, t)$ and $b(x, t)$ denote the concentrations of molecules of species A and species B at location $x \in [0, L]$. These concentrations can be approximated from the compartmental model as

$$a(x_i, t) \approx \frac{A_i(t)}{h^3}, \quad b(x_i, t) \approx \frac{B_i(t)}{h^3},$$

where x_i denotes the center of compartment i. If we set $d_A = d_B = 0$, then the standard law of mass action would give us the following approximate model for the average number of particles at each location:

$$\frac{\partial}{\partial t} a(x, t) = -2k_1 a^2 - k_2 ab - k_3 a + k_5,$$

$$\frac{\partial}{\partial t} b(x, t) = -k_2 ab + k_6 \mathbf{1}_{\{x \in [3L/5, L]\}} - k_4 b,$$

(4.20)

for each $x \in [0, L]$. To extend the law of mass action to our spatial model, we simply include diffusion of both species A and B at the appropriate rates, and add the no flux boundary conditions at $x = 0$ and $x = L$, to yield:

$$\frac{\partial}{\partial t}a(x,t) = -2k_1 a^2 - k_2 ab - k_3 a + k_5 + D_A \frac{\partial^2}{\partial x^2}a(x,t),$$

$$\frac{\partial}{\partial t}b(x,t) = -k_2 ab + k_6 \mathbf{1}_{\{x \in [3L/5,L]\}} - k_4 b + D_B \frac{\partial^2}{\partial x^2}b(x,t), \qquad (4.21)$$

$$\left.\frac{\partial}{\partial x}a(x,t)\right|_{x=0} = \left.\frac{\partial}{\partial x}a(x,t)\right|_{x=L} = \left.\frac{\partial}{\partial x}b(x,t)\right|_{x=0} = \left.\frac{\partial}{\partial x}b(x,t)\right|_{x=L} = 0.$$

Figure 4.5 below shows the results of simulating one realisation of the stochastic process and estimating the PMF at time $t = 200$, along with the solution of the associated law of mass action approximation from Equation 4.21. Because we are comparing a model of the approximate mean behaviour with a single realisation of the process, there is some disagreement due to the noisy nature of the underlying process but both solutions qualitatively agree in terms of the respective domains of dominance of the A and B molecules.
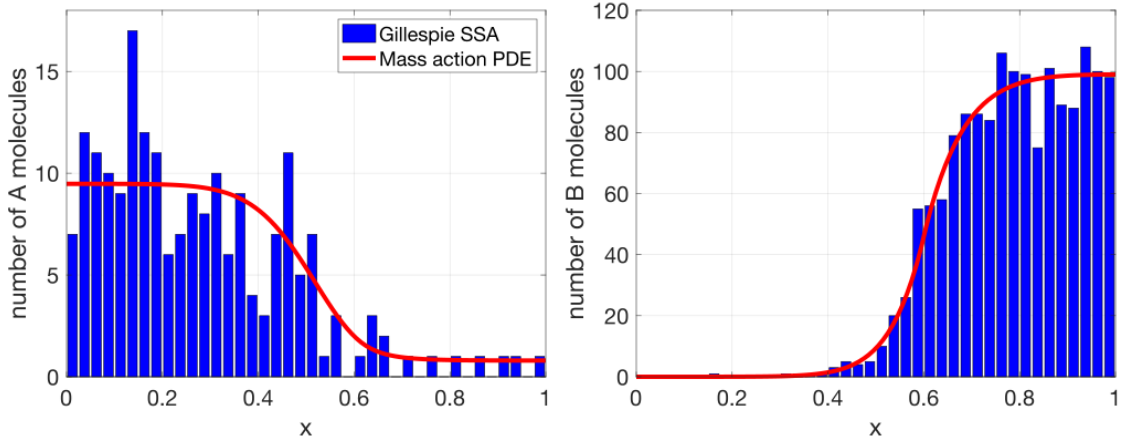

Figure 4.5: Simulations of the compartmental reaction-diffusion process Equation 4.19 with the estimated PMF in blue and the PDE approximation Equation 4.17 shown in red. Parameters: $D_A = D_B = 0.0001$, $h = 0.01$, $k_1 = 0.03$, $k_2 = 0.3$, $k_3 = 0.0001$, $k_4 = 0.0001$, $k_5 = 0.0000001$, $k_6 = 0.000001$, $L = 1$. [Code: `CH4_RD_compartment.m`]

To conclude our discussion of the compartmental approach to modelling reaction-diffusion processes, we need to address a key limitation of the method outlined above. Specifically, we need to discuss the accuracy of the method as it relates to the choice of the compartment size h (also thought of as the spatial discretisation parameter). We saw previously that we could derive reaction-diffusion PDEs in the limit as $h \downarrow 0$ when we have only zero and first order reactions. In fact, in these cases, the reaction-diffusion PDEs are exactly the evolution equations for the stochastic mean of the system as $h \downarrow 0$. However, when we have higher-order (nonlinear) reactions, there is trade-off between accurately modelling the diffusion of molecules and the reactions of molecules as $h \downarrow 0$. In particular, we model diffusion ever more accurately as we decrease h, but we model nonlinear reactions less accurately as we decrease h.

To understand this drawback of compartmental diffusion, we will revisit the stochastic dimeri-
sation process from Chapter 1, i.e. the well-mixed single-species chemical reaction process with
dynamics:

$$A + A \xrightarrow{k_1} \emptyset, \quad \emptyset \xrightarrow{k_2} A. \tag{4.22}$$

In Chapter 1 we saw that we could express the stationary distribution of the process Equa-
tion 4.22 in terms of the modified Bessel function of the first kind and we hence calculated the
stationary mean of the process, M_s, exactly. This process can thus serve as a good benchmark
since we understand its properties fully. We now extend the dimerisation process to the spatial
domain

$$\Omega = [0, L] \times [0, L] \times [0, L],$$

which we discretise into K compartments of width h in each dimension. In other words, we
divide Ω into K^3 cubes each of volume h^3, where compartment (i, j, k) is the compact interval
$[(i-1)h, ih] \times [(j-1)h, jh] \times [(k-1)h, kh]$ for $(i, j, k) \in \{1, \ldots, K\}^3$. We then let $A_{i,j,k}(t)$
denote the number of molecules of species A present in the compartment (i, j, k) at time t. As
in the previous example, only molecules in the same compartment can react with one another
so the reactions in Equation 4.22 become:

$$A_{i,j,k} + A_{i,j,k} \xrightarrow{k_1} \emptyset, \quad \emptyset \xrightarrow{k_2} A_{i,j,k}. \tag{4.23}$$
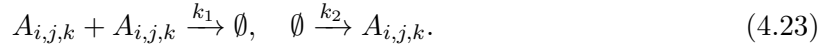
We also assume that molecules can diffuse at rate D_A $(= d_A h^2)$, but since there is production
and degradation across the entire domain we don't expect this to result in any spatial structure
in the solution, other than that resulting from stochastic fluctuations. In three dimensions,
diffusion corresponds to 6 possible movements of a molecule; the cube of volume h^3 that the
molecule resides in has 6 faces (unless on a boundary) and a molecule may diffuse into a
neighbouring cube corresponding to moving through each one of these faces.

The propensity functions for degradation and production in each compartment of our spatial
process are:

$$\alpha_{i,j,k,1}(t) = \frac{k_1 A_{i,j,k}(t)(A_{i,j,k}(t) - 1)}{h^3}, \quad \alpha_{i,j,k,2}(t) = k_2 h^3,$$

and the propensity function for diffusion in compartment (i, j, k) is $D_A A_{i,j,k}(t)/h^2 = d_A A_{i,j,k}(t)$.
We can also compute the total number of A molecules in the system at time t as

$$A(t) = \sum_{i,j,k} A_{i,j,k}(t)$$

and the long-run stochastic mean can then be defined as

$$M(h) := \lim_{t \to \infty} \mathbb{E}[A(t)],$$

where we write the long-run mean as a function of h in anticipation of studying its behaviour as $h \downarrow 0$. If our SSA works as we would hope, we should obtain $\lim_{h \to 0} M(h) = M_s$.
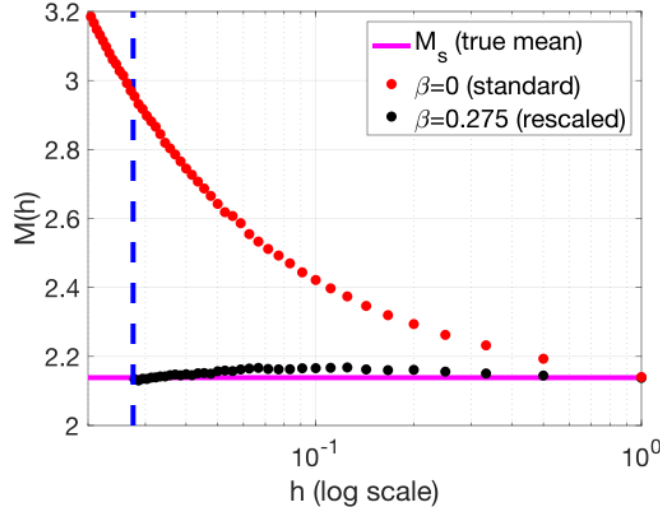

Figure 4.6: Simulations of the compartmental spatial dimerisation process using the standard Gillespie algorithm ($\beta = 0$) and the modified algorithm with rescaled dimerisation propensities ($\beta = 0.275$). Parameters: $D_A = 0.0001$, $L = 1$.

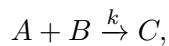We can simulate the spatial dimerisation process via the Gillespie algorithm by choosing $h = L/K$ and increasing the value of K to send $h \to 0$. The results of these simulations are plotted in Figure 4.6 with the red dots indicating the estimated value of $M(h)$ from standard Gillespie simulations and the solid magenta line showing the true value of M_s for the well-mixed version of the dimerisation process. Since we have only a single species of molecules diffusing, we do not expect any spatial structure or influence on the total molecule count and M_s and $M(h)$ should agree, but we actually see $M(h)$ increasing and diverging from M_s as h decreases. It turns out that $\lim_{h \to 0} M(h)$ does not exist and the SSA is not convergent as $h \to 0$. This is because the dimerisation reaction doesn't occur often enough as $h \to 0$ and hence we conclude that the compartmental SSA approach is only valid for h sufficiently large! This is exactly the opposite situation when we use a finite difference scheme to numerically approximate the solution to a PDE; we expect to see convergence of the scheme as we refine the mesh by taking the spatial discretisation parameter h smaller and smaller.

The fact that we cannot take h arbitrarily small is immediately disconcerting since we want h small to ensure that we model diffusion accurately. We generally want $h \ll L$ to be able

to accurately reflect spatial variation in the solution. However, for this dimerisation process we need to take $h \gg k_1/D_A$ in order to ensure that the nonlinear reactions aren't artificially suppressed, leading to a range of acceptable values for h:

$$k_1/D_A \ll h \ll L.$$

Similarly, if we instead had a two species reaction of the form:

$$A + B \xrightarrow{k} C,$$

with A and B diffusing at rates D_A and D_B respectively, then we would need $h \gg k/(D_A + D_B)$ to accurately reflect this reaction. Erban and Chapman (2009) have shown that for the spatial dimerisation process, it is possible to remedy this issue by rescaling the propensity function of the dimerisation reaction to

$$\alpha_{i,j,k,1}(t) = A_{i,j,k}(t)(A_{i,j,k}(t) - 1)\frac{D_A k_1}{D_A h^3 - \beta k_1 h^2}.$$

For $\beta = 0$, this is the standard Gillespie algorithm but for $\beta > 0$ the algorithm can be refined to accurately recover the value of M_s, as shown by the black dots in Figure 4.6. For a given value of K, the optimal value of β can be determined analytically for this process. However, this is a challenging problem in general, and the optimal choice of h for accuracy in both the reaction and diffusive dynamics is an area of active research.

4.4 Applications to biological pattern formation

There are many examples of regular spatial pattern formation in biology and ecology, including zebra stripes, spotted and striped fish, and vegetation patterns. Turing's pattern forming mechanism has long been studied as a simple and parsimonious way for self-organized patterns to form in the absence of external cues or other exogenous spatial structure. Turing patterns are typically studied mathematically in PDE-based models, but, as we have seen earlier in this chapter, these models are often coarse-grained approximations that capture average behaviour and neglect fluctuations and finite-size effects. To conclude our investigations of spatially extended stochastic processes, we will thus present an example of a "Turing pattern" in a stochastic reaction-diffusion system.

Consider two species of molecules, A and B, reacting and diffusing in our standard pseudo one-dimensional domain

$$\Omega = [0, L] \times [0, h] \times [0, h],$$

with the assumption that $h \ll L$. We only model diffusive movement in the x coordinate and discretise the interval $[0, L]$ into K compartments, with compartment i given by $[(i-1)h, ih]$. We choose Schnakenberg reaction kinetics so that the dynamics of the model are given by:

$$
\begin{aligned}
&A_1 \underset{d_A}{\overset{d_A}{\rightleftharpoons}} A_2 \underset{d_A}{\overset{d_A}{\rightleftharpoons}} \ldots \ldots \underset{d_A}{\overset{d_A}{\rightleftharpoons}} A_{i-1} \underset{d_A}{\overset{d_A}{\rightleftharpoons}} A_i \underset{d_A}{\overset{d_A}{\rightleftharpoons}} A_{i+1} \underset{d_A}{\overset{d_A}{\rightleftharpoons}} \ldots \ldots \underset{d_A}{\overset{d_A}{\rightleftharpoons}} A_{K-1} \underset{d_A}{\overset{d_A}{\rightleftharpoons}} A_K, \\
&B_1 \underset{d_B}{\overset{d_B}{\rightleftharpoons}} B_2 \underset{d_B}{\overset{d_B}{\rightleftharpoons}} \ldots \ldots \underset{d_B}{\overset{d_B}{\rightleftharpoons}} B_{i-1} \underset{d_B}{\overset{d_B}{\rightleftharpoons}} B_i \underset{d_B}{\overset{d_B}{\rightleftharpoons}} B_{i+1} \underset{d_B}{\overset{d_B}{\rightleftharpoons}} \ldots \ldots \underset{d_B}{\overset{d_B}{\rightleftharpoons}} B_{K-1} \underset{d_B}{\overset{d_B}{\rightleftharpoons}} B_K, \\
&2A_i + B_i \xrightarrow{k_1} 3A_i, \quad \emptyset \xrightarrow{k_2} A_i, \quad i = 1, 2, \ldots, K, \\
&A_i \xrightarrow{k_3} \emptyset, \quad \emptyset \xrightarrow{k_4} B_i, \quad i = 1, \ldots, K.
\end{aligned}
\tag{4.24}
$$
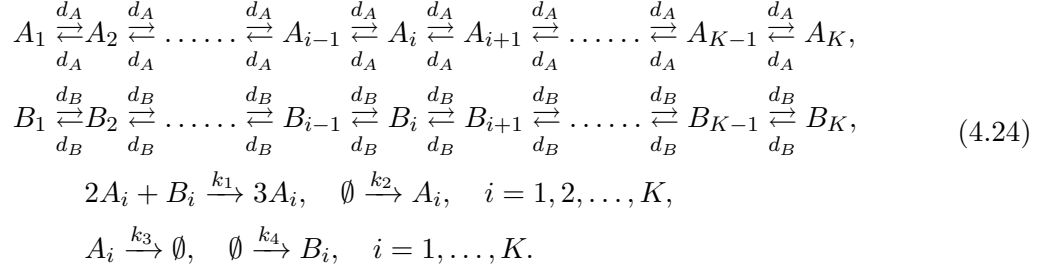
When choosing the hopping rates, d_A and d_B, we employ the standard relationship between the compartmental hopping rates and the continuum diffusion coefficients, i.e.

$$
d_A = \frac{D_A}{h^2}, \quad d_B = \frac{D_B}{h^2}.
$$

In the standard nomenclature of Turing pattern formation, species A should act as the slower diffusing activator and species B should play the role of the more quickly diffusing inhibitor species. Thus we choose the ratio of the diffusion coefficients as

$$
\frac{D_B}{D_A} = 100, \quad \text{with } D_A = 10^{-5} \text{ and } D_B = 10^{-3}.
\tag{4.25}
$$

For this example, the reaction rates are chosen according to:

$$
k_1/h^6 = 10^{-6}, \quad k_2 h^3 = 1, \quad k_3 = 0.02, \quad k_4 h^3 = 3.
$$

Figure 4.7 below shows a single realisation of the process Equation 4.24 with $K = 40$ compartments and the reaction and diffusion rates given above ($L = 1$, $h = 0.025$). Panels A and B show the time-space evolution of the process from a homogeneous initial condition. After around 500 seconds, we begin to see the emergence of spatial structure in the abundances of both the A and B molecules, although there is significant variation over time due to the underlying stochasticity of the process. Panels C and D show the state of the system at time $t = 2000$ seconds and we observe clear spatial patterning, especially in species A, which has very large amplitude patterns. Since this is a single realisation of the process, the roughness of the patterns is to be expected and we could average over many realisations if we wanted to estimate the mean behaviour, which would display much greater regularity.

To gain additional insight into the process Equation 4.24, we can write down a relatively simple mass action approximation of the average behaviour in the usual way. Letting $a(x, t)$ and $b(x, t)$ denote the A and B molecules concentrations, our mass action model is:
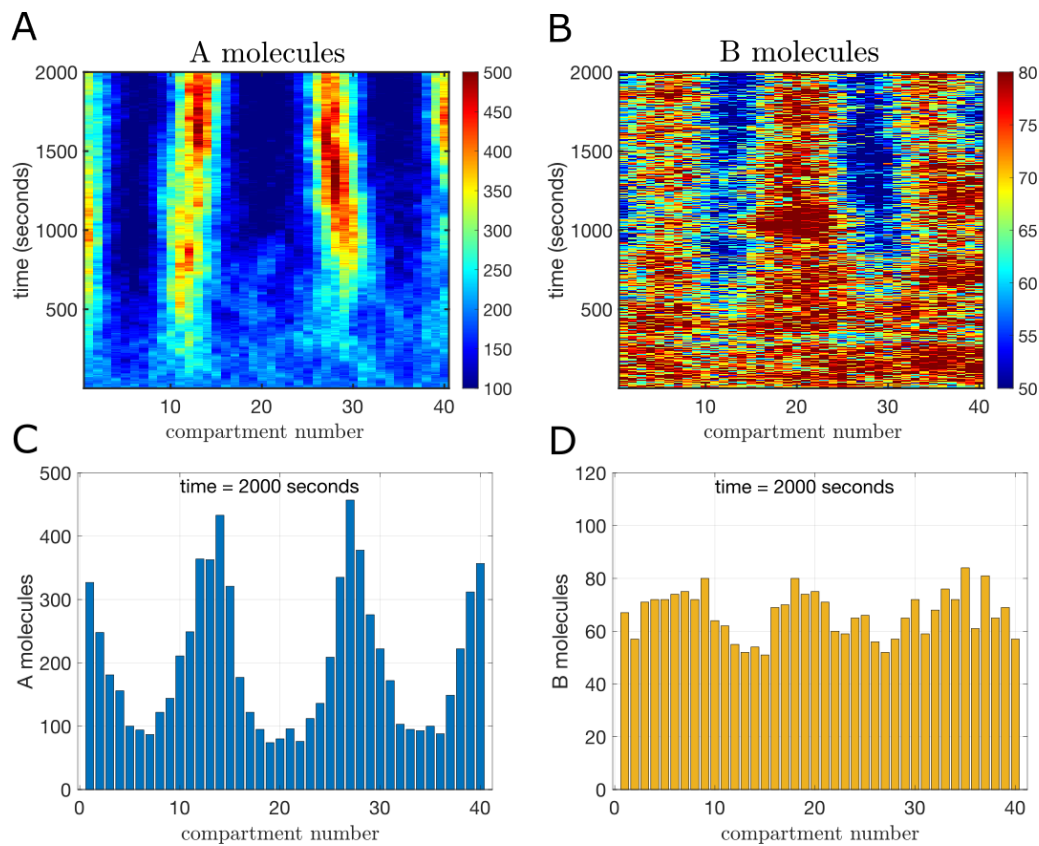
Figure 4.7: A/B: Space-time plots of the dynamics of the process Equation 4.24 (single realisation). C/D: Plots of the distribution of A and B molecules at time $t = 2000$ (in seconds). [Code: CH4_Schnakenberg.m]

$$\frac{\partial}{\partial t}a = k_1 a^2 b + k_2 - k_3 a + D_A \frac{\partial^2}{\partial x^2}a, \quad x \in (0, L),$$

$$\frac{\partial}{\partial t}b = -k_1 a^2 b + k_4 + D_B \frac{\partial^2}{\partial x^2}b, \qquad x \in (0, L), \tag{4.26}$$

$$\left.\frac{\partial}{\partial x}a(x,t)\right|_{x=0} = \left.\frac{\partial}{\partial x}a(x,t)\right|_{x=L} = \left.\frac{\partial}{\partial x}b(x,t)\right|_{x=0} = \left.\frac{\partial}{\partial x}b(x,t)\right|_{x=L} = 0.$$

Linear stability analysis of the approximate deterministic model Equation 4.26 shows that, for our chosen parameter set, it has a spatially homogeneous equilibrium solution corresponding to

$$\bar{A} = 200, \quad \bar{B} = 75,$$

molecules per volume h^3. The spatially homogeneous solution is stable for $D_A = D_B = 0$ but unstable for the diffusion coefficient values in Equation 4.25. Hence the results shown in Figure 4.7 really do represent a classical example of a spatially homogeneous equilibrium that is destabilised by a diffusion driven instability.

> **Exercise 4.1**
>
> 1. Solve the mass action PDE Equation 4.26 numerically using this VisualPDE link. Do the PDE dynamics qualitatively agree with those of the stochastic diffusion model?
>
> 2. Open the course Github page and use the MATLAB script `CH4_Schnakenberg.m` to compare the dynamics of the stochastic compartmental diffusion process with the dynamics of the PDE Equation 4.26 in VisualPDE for a range of parameter values.
>
> 3. Can you numerically identify the smallest value of D_A at which we no longer observe pattern formation in each model?
>
> **Bonus:** Can you analytically calculate the value of D_A at which the homogeneous solution of the PDE Equation 4.26 becomes unstable?
> **Note:** In VisualPDE notation, the species A is modelled as u and B is modelled as v. Similarly, the diffusion coefficients D_A and D_B are written as D_u and D_v in VisualPDE.

Knowledge checklist

Key topics:

1. Diffusion based on SDEs (SSAs and exact solutions)
2. Compartmental diffusion (SSAs and CME analysis)

3. Reaction-Diffusion processes (SSAs, higher-order reactions, tradeoffs)
4. Mass action approximations of reaction-diffusion processes.

Key skills:

- Represent spatial movement processes as chemical reaction processes.
- Use the Chemical Master Equations to analyse reaction-diffusion processes:

  – computing stationary distributions from CMEs.
  – calculating and solving moment equations.

- Write and analyse SSAs, and propose alternative SSAs for reaction-diffusion processes (e.g. Q3 on Problem Sheet 4).
- Write and analyse the mass-action (PDE) approximation for a given reaction-diffusion process.
- Discuss advantages and disadvantages of different approaches to modelling diffusion as a stochastic process (considering modelling implications and computational costs).