B5.1 Stochastic Modelling of Biological Processes

Maria Bruna

Hilary Term 2026

Aims

This course provides an introduction to stochastic methods for modelling biological systems, covering a number of applications, ranging in size from simulations of small biomolecules to stochastic modelling of groups of animals. The focus is on the underlying mathematics, i.e., it is not assumed that students have any prior knowledge in biology or chemistry.

Course synopsis

- Stochastic simulation of chemical reactions (birth-death processes).
- Brownian processes.
- Stochastic reaction-diffusion modelling.
- Stochastic models of dispersal in biological systems.

Algorithms and Computer Codes

The algorithms in the course text are all written in a general pseudocode. Although the final examination does not include computer practicals, you are encouraged to implement the algorithms contained in these notes as doing so will greatly aid in understanding. Because of this, some exercises will be computer-based, and you are welcome to implement these in any language of your choosing.

Reading list

- R. Erban & S. J. Chapman, Stochastic Modelling of Reaction-Diffusion Processes. Cambridge University Press, 2019.
- L. J. S. Allen, An Introduction to Stochastic Processes with Applications to Biology. CRC Press, 2010.
- G. A. Pavliotis, Stochastic Processes and Applications. Springer, 2015.
- C. W. Gardiner, Handbook of Stochastic Methods. Springer, 1985.

Pre-requisites

Part A Probability is essential.

Acknowledgements

The original version of this course was written by Radek Erban (as a Part C course). These notes are based on previous versions by Ruth Baker and Murad Banaji. I am very grateful to all of them. Any errors are my responsibility (please email them to me!).

Contents

Changelog (March 24, 2026)

- Example 12.1: clarified the definition of MFPT when two adsorbing boundaries are present (previous wording was ambiguous).

Lecture 1: Introduction

This course covers applied stochastic processes with a focus on biological applications. Stochastic processes are commonly used to model dynamic biological phenomena involving *agents* moving in space and/or interacting with each other and the environment. For example, agents could be:

- a grain of pollen in a glass of water,
- chemicals reacting with each other,
- a molecule diffusing inside a cell,
- cells moving within an organism using chemical sensing,
- an ant foraging for food,
- animals interacting competitively or cooperatively,
- trees competing for space and light in a forest,
- people being free, susceptible or infected with a virus.

A *stochastic process* is a random quantity which evolves in time, i.e., a collection $\{X(t) : t \in T\}$ of random variables indexed by an ordered time-set T and taking values in a state space S. The sets T and S can be discrete or continuous, and this determines the methods for formulating and analysing the stochastic process. This course will focus on continuous-time stochastic processes, $T \subseteq \mathbb{R}_+ = [0, \infty)$ and cover both discrete and continuous state spaces.

The nature of the models we study depends fundamentally on the scale at which we observe a system. Consider for example pollen particles in water. At the very detailed scale, we might consider studying deterministic models, where we track each individual pollen particle and water molecule in space and time as they move, collide with other particles/molecules, etc. The complexity rapidly becomes immense, and it is natural to simplify the picture by moving to stochastic models: we do not try to follow all the water molecules; rather the state of the pollen particle colliding with millions of molecules is treated as a collection of random variables – a stochastic process. If we increase the scale still further, we no longer focus on individual pollen particles, but rather on concentrations, gradients, etc., which we hope to study using differential equations: we end up back again with deterministic models.

A range of mathematical and numerical methods feature in studying continuous-time discrete- and continuous-space stochastic processes. These include continuous-time Markov chains, stochastic differential equations, stochastic simulation algorithms (SSAs) such as Monte Carlo methods and the Gillespie algorithm, and the Chapman–Kolmogorov equation. We will study forward and backward equations describing the evolution of the probability density associated with a stochastic process. These can either be an (infinite) system of ODEs in the case of discrete-space processes, or PDEs in the case of continuous-space processes. We will compare different models among themselves, as well as with their deterministic counterparts, and highlight their connections and differences. In this first lecture, we present two examples, namely a Brownian particle and simple chemical reaction networks, to illustrate and motivate some of the models and methods we will study.

1.1 A Brownian particle in a harmonic potential

Let $x(t) \in \mathbb{R}$ denote the (deterministic) position of a particle at time t under the action of a potential energy $\alpha x^2/2$, with $\alpha > 0$ constant. In the absence of any dissipative forces, we have that

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -U'(x) = -\alpha x, \qquad x(0) = x_0, \tag{1.1}$$

which is a gradient flow. Hence

$$x(t) = x_0 e^{-\alpha t}, \tag{1.2}$$

and $x_\infty := \lim_{t\to\infty} x(t) = 0$. That is, the particle ends up at the bottom of the potential well, which is the unique stable equilibrium of the system (see Figure 1.1).

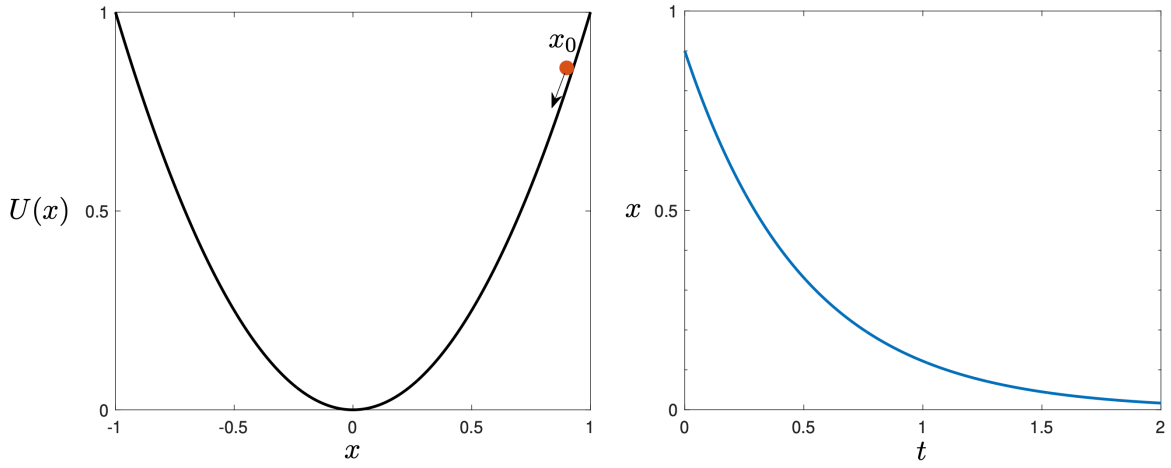

Figure 1.1: Left: Quadratic potential $U(x) = x^2$ and ball at $x_0 = 0.9$. Right: trajectory $x(t)$ given by (1.2) with $\alpha = 2$.

Stochastic model In many applications, experimentally measured trajectories are apparently subject to random perturbations. Hence it is reasonable to modify (1.1) to include a noise term $\xi(t)$

$$\frac{\mathrm{d}X}{\mathrm{d}t} = -\alpha X + \sqrt{2D}\xi(t), \qquad X(0) = X_0, \tag{1.3}$$

where $\sqrt{2D}$ sets the scale of the fluctuations. Let's assume that the noise $\xi(t)$ is modelled by "white noise".[1]

The white noise can be interpreted as the "time derivative" of Brownian motion or Wiener process $W(t)$, "$\xi(t) = \mathrm{d}W(t)/\mathrm{d}t$" (we will discuss this in detail in a later lecture). Formally multiplying (1.3) by $\mathrm{d}t$ we arrive at the *stochastic differential equation* (SDE)

$$\mathrm{d}X(t) = -\alpha X(t)\mathrm{d}t + \sqrt{2D}\mathrm{d}W(t), \qquad X(t) = x_0. \tag{1.4}$$

[1]Note the convention we will aim to stick to throughout the course: random variables are typically capitalised, while deterministic variables are in lower caps.

This is known as the *Ornstein–Uhlenbeck process*. The Ornstein–Uhlenbeck adds noise to the gradient flow (1.1). Ornstein and Uhlenbeck introduced it in 1930 as a model for the velocity of a Brownian particle.

The OU process can be sampled exactly as a scaled, time-transformed Wiener process. But for now, suppose we want to simulate (1.4) by numerical integration. The *Euler–Maruyama scheme* is analogous to the forward Euler scheme for SDEs instead of ODEs. We consider a fixed timestep $\Delta t = T/I$ for some number of steps I and final time t_f. Let's X_i denote the approximation to $X(i\Delta t)$.

Algorithm 1.1: Euler–Maruyama integration of the Ornstein–Uhlenbeck process

Set $X_0 = x_0$ and $\Delta t = T/I$.
For $i = 0$ to $i = I - 1$

(a) Generate a random number $\xi \sim \mathcal{N}(0, 1)$ (the standard normal distribution).

(b) Set

$$X_{i+1} = X_i - \alpha X_i \Delta t + \sqrt{2D\Delta t}\,\xi. \tag{1.5}$$

end

We plot ten trajectories generated with the Euler–Maruyama algorithm for two values of the diffusion coefficient in Figure 1.2. Observe how, unlike the ODE in Figure 1.1, each trajectory is different despite having the same initial condition. Also note the role of the diffusion coefficient: the larger value of D (right plot) leads to more disparate paths or, in other words, higher noise. Let $M(t)$ be the mean value of $X(t)$. Using the computational definition (13.16), we can compute

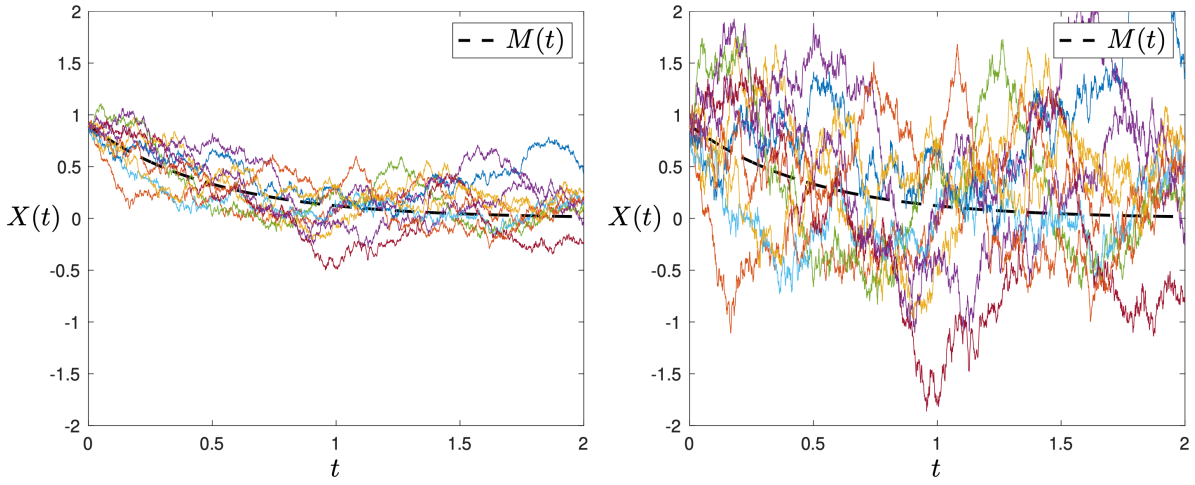

Figure 1.2: Ten realisations of (13.16) for $D = 0.1$ (left) and $D = 1$ (right), together with the ODE solution (1.1), which agrees with the mean $M(t)$ (black dashed line). Parameters used: $\alpha = 2, x_0 = 0.9, \Delta t = 10^{-3}$.

the average value of X_{i+1} by

$$M_{i+1} = \mathbb{E}[X_{i+1}] = \mathbb{E}[X_i - \alpha X_i \Delta t + \sqrt{2D\Delta t}\,\xi] = (1 - \alpha\Delta t)M_i,$$

since $\mathbb{E}[\xi] = 0$. We see that the mean M_i is given by the forward Euler approximation of the ODE (1.1) (see the dash-black line in Figure 1.2).

Probabilistic model Let $p(x,t)$ be the probability density of $X(t)$, where $X(t)$ satisfies (1.4). This means that $p(x,t)\mathrm{d}x$ gives the probability that $X(t) \in [x, x+\mathrm{d}x)$, i.e., $p(x,t)\mathrm{d}x = \mathbb{P}[X(t) \in [x, x+\mathrm{d}x)]$. We will see that the evolution of p is given by the forward Kolmogorov equation or *the Fokker–Planck* PDE

$$\frac{\partial p}{\partial t} = \alpha\frac{\partial (xp)}{\partial x} + D\frac{\partial^2 p}{\partial x^2}, \qquad p(x,0) = \delta(x - x_0). \tag{1.6}$$

We can use the Fokker–Planck equation (1.6) to describe the average evolution of the Ornstein–Uhlenbeck process (13.16). For a general continuous stochastic process, the Fokker–Planck cannot be solved explicitly, and one must resort to numerical methods for PDEs. But the solution of (1.6) can be obtained exactly as

$$p(x,t) = \frac{1}{\sqrt{2\pi\sigma^2(t)}} e^{-\frac{1}{2}\left(\frac{x-\mu(t)}{\sigma(t)}\right)^2}, \tag{1.7}$$

with mean and variance

$$\mu(t) = x_0 e^{-\alpha t}, \qquad \sigma^2(t) = \frac{D}{\alpha}\left(1 - e^{-2\alpha t}\right).$$

The long-time behaviour of $p(x,t)$ is given by the stationary probability density

$$p_\infty(x) = \lim_{t\to\infty} p(x,t) = \sqrt{\frac{\alpha}{2\pi D}} e^{-\frac{\alpha x^2}{2D}}. \tag{1.8}$$

We note that, as $D \to 0$, $p_\infty(x)$ converges to a delta function $\delta(x)$, that is, $\lim_{t\to\infty} X(t) = 0$ with probability one. This is consistent with the SDE (1.4) reducing to the ODE (1.1) as the noise strength goes to zero.

We can compare the probabilistic description (1.7) with a histogram obtained from multiple trajectories of the stochastic description (1.4).[2] In Figure 1.3 we plot the histogram and the solution (1.7) at three time instances. We observe how the Gaussian distribution shifts from around $x_0 = 0.9$ towards $x = 0$ as time progresses, and the variance increases from zero (since the initial data is deterministic) to its equilibrium value of $\sigma_s = D/\alpha = 1/2$.

1.2 A chemical species under degradation

Let A denote a chemical species inside a well-mixed compartment of unit volume, which is subject to degradation:

$$\mathrm{A} \xrightarrow{k} \emptyset, \tag{1.9}$$

[2] See [7, §2.4] for details on how to build a histogram to estimate a probability density.

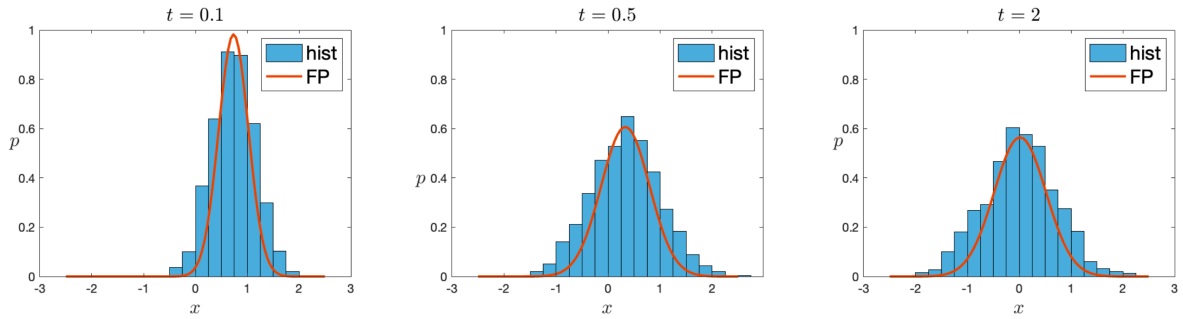

Figure 1.3: Histograms obtained from 1000 realisations of (13.16) and solution to the Fokker–Planck PDE (1.7) at times $t = 0.1, 0.5, 2$. Parameters used: $\alpha = 2, D = 1, x_0 = 0.9, \Delta t = 10^{-3}$.

where $k > 0$ is the rate constant of the reaction. It is defined so that $k\Delta t$ is the probability that a molecule of A degrades during the time interval $[t, t + \Delta t)$ where t is the time and Δt is a sufficiently small time step. In (1.9), $\emptyset$ represents some chemical species that are not explicitly modelled.

Stochastic model. Denote the number of molecules of A present at time t by $A(t)$. Then

$$\mathbb{P}(\text{no reactions in } [t, t+\Delta t)) = 1 - A(t)k\Delta t + o(\Delta t), \tag{1.10}$$

$$\mathbb{P}(\text{one reaction in } [t, t+\Delta t)) = A(t)k\Delta t + o(\Delta t), \tag{1.11}$$

$$\mathbb{P}(\text{more than one reaction in } [t, t+\Delta t)) = o(\Delta t). \tag{1.12}$$

This definition of the chemical reaction (1.9) can be directly used to design a "naive" simulation algorithm for simulating it. We denote by A_i the approximation to $A(i\Delta t)$ and choose a fixed small time step $\Delta t = tf/I$, where t_f is the final time and I the number of steps, such that $o(\Delta t)$ terms can be neglected.

Algorithm 1.2: Naive Algorithm for degradation (1.9)

Set $A_0 = N > 0$ and $\Delta t = t_f/I$.
For $i = 0$ to $I - 1$:

(a) Generate a random number $r \sim \mathcal{U}(0, 1)$.

(b) Set $A_{i+1} = A_i - 1$, if $r < k\Delta t A_i$, and $A_{i+1} = A_i$ otherwise.

end

In Figure 1.4(left), we plot as coloured lines five different trajectories generated by the naive Algorithm 1.2. This algorithm relies on choosing a sufficiently small time-step Δt as it introduces a $O(\Delta t^2)$ error at each iteration. This also make is very inefficient: in many steps, no reaction occurs and this "throws away" many of the generated random numbers. In the next lecture, we will introduce the *Gillespie SSA*, an exact and more computationally efficient algorithm.

Probabilistic model. Let p_n denote the probability that there are n molecules of A present

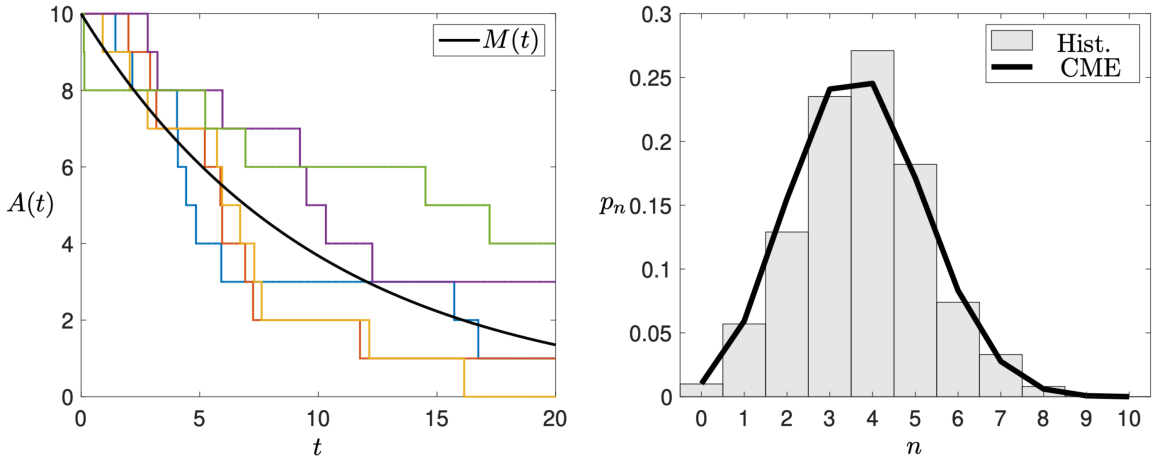

Figure 1.4: Degradation system (Left) Five sample trajectories (coloured curves) generated by using the Algorithm 1.2, together with the mean $M(t)$ (1.17) (thick black curve). (Right) Probability $p_i(t)$ of reaction (1.9) at time $t = 10$, obtained by 10^3 sample trajectories of the Algorithm 1.2 (grey histogram) and the exact solution (1.16) for the chemical master equation (black curve). Parameters are $N = 10$, $\Delta t = 0.001$, and $k = 0.1$.

in the system at time t. Then

$$p_n(t + \Delta t) = (1 - kn\Delta t)\, p_n(t) + k(n + 1)\Delta t\, p_{n+1}(t) + o(\Delta t), \tag{1.13}$$

so that, rearranging and taking the limit $\Delta t \to 0$, a system of differential equations is obtained, known as the *forward Kolmogorov differential equations*:

$$\frac{\mathrm{d}p_n}{\mathrm{d}t} = k(n + 1)p_{n+1}(t) - knp_n(t), \tag{1.14}$$

where $n = 0, 1, 2, \ldots, I - 1$. In physics and chemistry, the system (1.14) is often referred to as the *master equation* or the chemical master equation.

If the initial number of A molecules in the system is $A(0) = N$, this means $p_N(0) = 1$ and $p_n(0) = 0$ for all $n \neq N$. Hence, the equation for p_N is

$$\frac{\mathrm{d}p_N}{\mathrm{d}t} = -kNp_N(t) \quad \text{with} \quad p_N(0) = 1 \quad \Longrightarrow \quad p_N(t) = e^{-kNt}. \tag{1.15}$$

Inductively, one can show that

$$p_n(t) = \binom{N}{n} e^{-nkt} \left(1 - e^{-kt}\right)^{N-n}, \quad n = 0, 1, \ldots, N, \tag{1.16}$$

i.e., $A(t) \sim \mathrm{Bin}(N, e^{-kt})$ is a binomially distributed random variable. This means that the mean and variance of the number of A molecules present in the system at time t are, respectively, given by

$$M(t) = Ne^{-kt} \quad \text{and} \quad V(t) = Ne^{-kt}\left(1 - e^{-kt}\right). \tag{1.17}$$

The forward Kolmogorov equations (1.14) and its solution, (1.16), enable us to quantify the stochastic fluctuations around the mean. Evolution of the mean molecule number is plotted in Figure 1.4(left), alongside the distribution $p_i(t)$ at a given time t in Figure 1.4(right).

Deterministic model. Let $a(t)$ be the concentration of the species A in a voluem ν. The deterministic *reaction-rate equation* associated with the degradation reaction (1.9) is

$$\frac{\mathrm{d}a}{\mathrm{d}t} = -ka, \qquad a(0) = N/V. \tag{1.18}$$

Its solution is

$$a(t) = N\exp(-kt)/V, \tag{1.19}$$

which coincides with the mean number of molecules M up to the constant ν. That is, the deterministic solution times the volume coincides with the mean of the stochastic model. Later in this course, we will show that the deterministic solution is not always the average of the stochastic solutions. In fact, significant discrepancies can exist between the two models.

1.3 Links between different modelling approaches

In the two preceding subsections, we have seen two examples of continuous-time *Markov processes* taking either continuous values (in the Brownian particle example) or discrete values (in the degradation examples). The course will roughly consist of two parts, covering applications of continuous- and discrete-space Markov processes. In each of the parts, we will cover a range of modelling approaches or descriptions available such as deterministic, stochastic, computational and probabilistic models. A key learning output of this course is to identify and connect different descriptions for the same problem (see Table 1.1).

	Continuous process	**Discrete process**
Model example	Brownian motion	Degradation reaction
Stochastic model	Stochastic differential equations	Continuous-time Markov chains
Computational model	Euler–Maruyama (Algorithm 1.1)	Algorithm 1.2, Gillespie SSA
Probabilistic model	Fokker–Planck equation (1.6)	Chemical master equation (1.14)

Table 1.1: Examples of modelling approaches for continuous and discrete stochastic processes.

We will see that in some applications, it is convenient to approximate a truly discrete state space by a continuous space and vice-versa. For example, chemical reaction networks and population dynamics are often studied using tools from continuous stochastic processes for computational or analytical purposes. In the other direction, a typical example is the Drunkard's random walk, which assumes the drunkard makes discrete steps left and right as a model for Brownian motion.

Tasks

- Implement the Euler-Maruyama scheme in Algorithm 1.1 and plot some trajectories as in Figure 1.2.
- Implement a stochastic simulation algorithm that models the degradation reaction (1.9), and plot some sample paths generated using the algorithm and the parameter values in Figure 1.4.

- Use your algorithm to estimate the evolution of the mean and variance in the number of A molecules by averaging over a large number of sample paths, and compare them with the analytical expressions in (1.17).

Example codes

`Lect1.ipynb`: contains simulations of the Ornstein–Uhlenbeck process and degradation process.

Lecture 2: Production and degradation

The two examples from the previous lecture are instances of Markov processes. Simply put, a Markov process is a stochastic process that retains no memory of where it has been in the past: only the current state of a Markov process can influence where it will go next.

Recall the degradation reaction

$$\mathrm{A} \xrightarrow{k} \emptyset, \tag{2.1}$$

and the naive Algorithm 1.2. We observed that this is an inefficient (and inaccurate) method for simulating (2.1), as in many time steps, no degradation reaction occurs.

2.1 Waiting times are exponentially distributed

How long do we have to wait until the reaction (2.1) "fires" and the state of the system changes? Understanding how to compute this (random) time, is key both to understanding and to simulating stochastic models. It is remarkable that we can infer some very specific and useful information from the Markov assumptions alone.

Given $A(0) = n$, we define the *waiting time* associated with the state n,

$$T_n := \inf\{t > 0 \,|\, A(t) \neq n\}\,.$$

We may also refer to T_n as the first jump time. We also introduce the *survival probability*

$$S_n(t) = \mathbb{P}(T_n > t).$$

Using that $A(t)$ is a time-homogeneous Markov process, we can see that T_n has the *memoryless property*:[1]

$$\begin{aligned} S_n(t+u) &= \mathbb{P}(T_n > t+u \,|\, T_n > t) S_n(t) \\ &= \mathbb{P}(\text{no jump in } (t, t+u] \,|\, A(t) = n) S_n(t) \\ &= \mathbb{P}(\text{no jump in } (0, u] \,|\, A(0) = n) S_n(t) \\ &= S_n(u) S_n(t). \end{aligned}$$

In the first line, we have used conditional probability, in the second the Markov property of $A(t)$, and in the third the fact that it is time-homogeneous. Hence, T_n is exponentially distributed since the exponential is the only continuous memoryless distribution.

Remark 2.1. The following properties are equivalent for a memoryless waiting time T. Let $\lambda > 0$.

1. T has $\mathrm{Exp}(\lambda)$ distribution.
2. $\mathbb{P}(T \leq \Delta t) = \lambda\, \Delta t + o(\Delta t) \quad$ as $\Delta t \downarrow 0$.

[1] A random variable T has the memoryless property if $\mathbb{P}(T > a + b) = \mathbb{P}(T > a)\mathbb{P}(T > b)$ for any positive a and b.

In one direction this is obvious as $1 - e^{-\lambda\,\Delta t} = \lambda\,\Delta t + o(\Delta t)$. In the other direction, let $S(t) = \mathbb{P}(T > t)$. Clearly $S(0) = 1$. Then, by memorylessness, and our assumption,

$$S(t + \Delta t) = S(t)S(\Delta t) = S(t)(1 - \lambda\Delta t + o(\Delta t)),$$

or

$$\frac{S(t + \Delta t) - S(t)}{\Delta t} = -S(t)\lambda + S(t)\frac{o(\Delta t)}{\Delta t}\,.$$

Taking the limit $\Delta t \to 0$ gives the ODE

$$S'(t) = -\lambda S(t) \quad \text{with solution} \quad S(t) = e^{-\lambda t}\,,$$

where we have used the initial condition $S(0) = 1$.

Thus, for (2.1) and $A(t) = n$, the waiting time is $T_n \sim \text{Exp}(kn)$ and its probability density, given by $-S_n'(t)$, is

$$f_{T_n}(t) = kne^{-knt}. \tag{2.2}$$

We see that very general assumptions lead us to the conclusion that given any initial state, the waiting time until the state of the system changes (i.e., the first reaction fires) is exponentially distributed. We should bear in mind the special case that it may be infinite, if the current state of the system is an absorbing state.

2.1.1 Generating random numbers

To generate stochastic sample paths, we need a way to sample the waiting times T_n, which we have shown to be exponentially distributed with rate kn when the current state is $A(t) = n$. To this end, we consider the survival probability $S_n(t) : (0, \infty) \to (0, 1)$, which is monote decreasing for $n > 0$. For all $0 < a < b < 1$ we then have that

$$\begin{aligned}
\mathbb{P}\big(S_n(T_n) \in (a, b)\big) &= \mathbb{P}\big(T_n \in (S_n^{-1}(a), S_n^{-1}(b))\big) \\
&= \int_{S_n^{-1}(a)}^{S_n^{-1}(b)} kne^{-kns}\,\mathrm{d}s \\
&= S_n(S_n^{-1}(b)) - S_n(S_n^{-1}(a)) \\
&= b - a.
\end{aligned}$$

Hence $S_n(T_n)$ is uniformly distributed in $(0, 1)$. This means that, if r is a random number uniformly distributed on $(0, 1)$, we may generate a sample of the waiting time T_n by solving

$$r = S_n(T_n) = e^{-knT_n},$$

or

$$T_n = \frac{1}{kn}\,\ln\!\left(\frac{1}{r}\right). \tag{2.3}$$

This is an instance of the *inverse transform sampling*, which allows us to sample a random number from any probability distribution given its cumulative distribution function (or its survival probability).

2.2 Stochastic simulation algorithm of degradation

The stochastic simulation algorithm for the degradation reaction (2.1) is a particular instance of the Gillespie SSA.

Algorithm 2.1: Gillespie SSA for degradation (2.1)

Set $A(0) = N > 0$, $t = 0$ and final time t_f.
While $t < t_f$:

(a) Generate a random number $r \sim \mathcal{U}(0,1)$ and set $T = \frac{1}{kA(t)} \ln\left(\frac{1}{r}\right)$.

(b) Set $A(t+T) = A(t) - 1$ and $t = t + T$.

end

In contrast with the naive Algorithm 1.2, SSA 2.1 samples trajectories of $A(t)$ exactly.

2.3 Production and degradation

Consider the production and degradation of the chemical species A according to the following reactions:

$$\mathrm{A} \xrightarrow{k_1} \emptyset, \tag{2.4a}$$

$$\emptyset \xrightarrow{k_2} \mathrm{A}, \tag{2.4b}$$

where $k_1 > 0$ is the rate of degradation and k_2 is the rate of production of A per unit volume. This means that one molecule of A is produced during the time interval $[t, t+\Delta t)$ with probability $k_2\nu\Delta t$ where ν is the volume of the system.

As before, we denote the number of molecules of A present at time t by $A(t)$. Then we have

$$\begin{aligned}
\mathbb{P}(\text{no reactions in } [t, t+\Delta t)) &= 1 - (k_1 A(t) + k_2\nu)\Delta t + o(\Delta t), && (2.5)\\
\mathbb{P}(\text{one A molecule decays in } [t, t+\Delta t)) &= k_1 A(t)\Delta t + o(\Delta t), && (2.6)\\
\mathbb{P}(\text{one A molecule produced in } [t, t+\Delta t)) &= k_2\nu\Delta t + o(\Delta t), && (2.7)\\
\mathbb{P}(\text{more than one reaction in } [t, t+\Delta t)) &= o(\Delta t). && (2.8)
\end{aligned}$$

2.3.1 The chemical master equation

As before, let $p_n(t)$ denote the probability that there are n molecules of A present in the system at time t. Then, for $n > 0$, we have

$$p_n(t+\Delta t) = (1 - k_1 n\Delta t - k_2\nu\Delta t)\, p_n(t) + k_1(n+1)\Delta t p_{n+1}(t) + k_2\nu\Delta t p_{n-1}(t) + o(\Delta t), \tag{2.9}$$

so that, rearranging and taking the limit $\Delta t \to 0$, we arrive at the forward Kolmogorov equations or the master equation for the production-degradation system (2.4)

$$\frac{\mathrm{d}p_n}{\mathrm{d}t} = k_1(n+1)p_{n+1} - (k_1 n + k_2\nu)p_n + k_2\nu p_{n-1}, \qquad n = 0, 1, 2, \ldots, \tag{2.10}$$

where $p_{-1} \equiv 0$ and $p_n(0) = \delta_{nN}$ if $A(0) = N$ is the initial number of A molecules in the system.

We can use the master equation to derive equations for the mean and variance:

$$M(t) = \sum_{n=0}^{\infty} n\, p_n(t) \quad \text{and} \quad V(t) = \sum_{n=0}^{\infty} (n - M(t))^2\, p_n(t). \tag{2.11}$$

Exercise 2.1 (Mean and variance of molecule number)**.** Given $A(0) = N$, use (2.10) to show that for the chemical system (2.4)

$$M'(t) = -k_1 M + k_2 \nu, \qquad V'(t) = -2k_1 V + k_1 M + k_2 \nu.$$

The stationary mean and variance are:

$$M_s = \lim_{t\to\infty} M(t) = \frac{k_2\nu}{k_1} \quad \text{and} \quad V_s = \lim_{t\to\infty} V(t) = \frac{k_2\nu}{k_1} = M_s. \tag{2.12}$$

Evolution of $M(t)$ towards this steady state is plotted in Figure 2.1.

2.3.2 The stationary distribution

The variance, $V(t)$, gives us some information about the fluctuations in molecule number. However, we can fully characterise them by considering the *stationary distribution*:

$$\phi_n \coloneqq \lim_{t\to\infty} p_n(t), \quad n = 0, 1, 2, \ldots \tag{2.13}$$

We can compute ϕ_n by considering the steady states of the master equation:

$$0 = k_1(n+1)\phi_{n+1} - (k_1 n + k_2\nu)\phi_n + k_2\nu\phi_{n-1}; \tag{2.14}$$

for $n \geq 0$ (noting $\phi_{-1} = 0$), to arrive at the recursive definition

$$\phi_1 = \frac{k_2\nu}{k_1}\phi_0, \tag{2.15a}$$

$$\phi_{n+1} = \frac{1}{k_1(n+1)}\left(k_1 n\phi_n + k_2\nu\phi_n - k_2\nu\phi_{n-1}\right), \quad n \geq 1. \tag{2.15b}$$

We can eliminate the remaining constant, ϕ_0, by noting that $\sum_{n=0}^{\infty}\phi_n = 1$. The stationary distribution is plotted in Figure 2.1.

2.4 Stochastic simulation of production/degradation

As previously, we would like to be able to generate individual sample paths from the model using a computational algorithm. This time we have two decisions: when the next reaction occurs; and which reaction (production or degradation) occurs.

Waiting times Following the same arguments as before, the waiting time T_n given state $A(t) = n$ is exponentially distributed with rate $\alpha_0(n)$. Here $\alpha_0(n)$ is the stochastic reaction rate or propensity that *a* reaction will occur, that is, $\alpha_0(n) = k_1 n + k_2\nu$. Therefore, given $A(t)$ molecules at time t and $r_1 \sim U(0,1)$, the time of the next event can be obtained by solving

$$r_1 = e^{-\alpha_0(n)T_n}.$$

Choosing which reaction occurs Whether the reaction that occurs is a production or degradation reaction depends on the relative probabilities of the two reactions:

$$\mathbb{P}(\text{degradation reaction occurs}) = \frac{k_1 n}{k_1 n + k_2 \nu} = \frac{k_1 n}{a_0(n)};$$
$$\mathbb{P}(\text{production reaction occurs}) = \frac{k_2 \nu}{k_1 n + k_2 \nu} = \frac{k_2 \nu}{a_0(n)}.$$

We summarize this stochastic simulation algorithm in box SSA 2.2:

Algorithm 2.2: Gillespie SSA for production and degradation (2.4)

Set $A(0) = N > 0$, $t = 0$ and final time t_f.
While $t < T$:

(a) Generate two random numbers, r_1 and r_2, uniformly distributed in $(0, 1)$.

(b) Compute $\alpha_0(t) = k_1 A(t) + k_2 \nu$ and the time of the next reaction by $T = -\ln r_1/\alpha_0$.

(c) Set

$$A(t+T) = \begin{cases} A(t) - 1, & \text{if} \quad \alpha_0 r_2 < k_1 A(t), \\ A(t) + 1, & \text{if} \quad \alpha_0 r_2 \geq k_1 A(t), \end{cases}$$

and $t = t + T$.

end

Five sample paths generated using SSA 2.2 are plotted in Figure 2.1. A large number (10^7) of sample paths generated using the algorithm are also used to approximate the stationary distribution on the right of Figure 2.1.

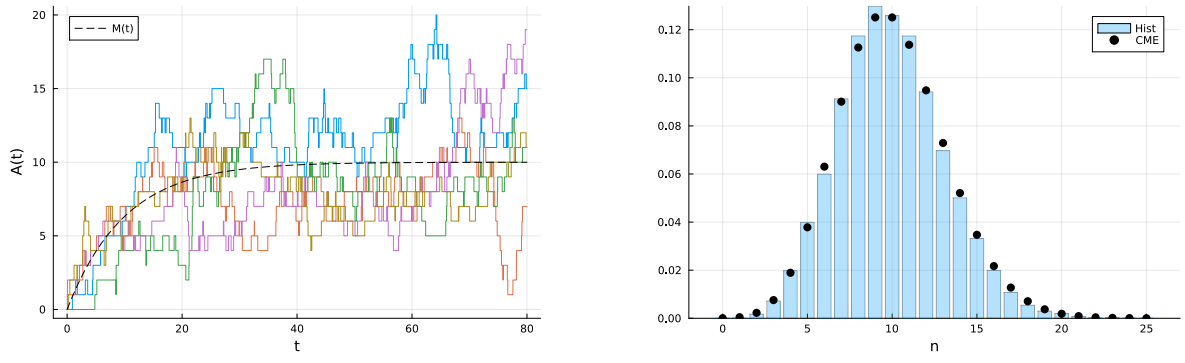

Figure 2.1: Sample paths from a production-degradation reaction system. Left: five different sample paths with the mean (black dashed line). Right: the stationary distribution ϕ_n calculated using both equations (2.15a) and (2.15b) (orange) and repeated stochastic simulation (grey). Parameters are $A(0) = 0$, $k_1 = 0.1$ and $k_2\nu = 1.0$.

Tasks

- Revise Appendix A.1.
- Do Exercise 2.1 and solve (2.15) to find that

$$\phi_n = \frac{C}{n!}\left(\frac{k_2\nu}{k_1}\right)^n \quad \text{with} \quad C = \exp\left(-\frac{k_2\nu}{k_1}\right). \tag{2.16}$$

- Implement the SSA 2.2 that models the production-degradation system (2.4) and reproduce Figure 2.1, including a measure of $V(t)$.

Example codes

`Lect2.ipynb`: contains implementations of the SSAs 2.1 for degradation and the SSA 2.2 for production-degradation and a long sample path to estimate the stationary distribution ϕ_n.

Lecture 3: Chemical reaction networks: sample paths

In Lecture 2 we saw how to simulate a one-species production–degradation system by sampling (i) an exponential waiting time and (ii) which of two reactions fires. In this lecture we generalise this to *arbitrary* chemical reaction networks (CRNs), where we may have many species and many possible reaction channels. We explain how both deterministic and stochastic models arise from the same underlying structure. Our focus in this lecture is on sample paths: how a system evolves along a single realisation of the stochastic dynamics.

3.1 Chemical reaction networks

Consider a system of m chemical reactions on n species evolving in a well-mixed container. A chemical reaction network consists of a finite set of species

$$\mathcal{S} = \{\mathrm{X}_1, \dots, \mathrm{X}_n\},$$

together with a finite set of reactions. Each reaction is written in the form

$$C_1 \to C_2,$$

where C_1 and C_2 are formal linear combinations of species with nonnegative integer coefficients, denoted *complexes*. In this reaction, C_1 is the *reactant complex* and C_2 is the *product complex*. Given a reactant complex $C_1 = c_1\mathrm{X}_1 + c_2\mathrm{X}_2 + \cdots + c_n\mathrm{X}_n$, the sum of its coefficients, $|c| := \sum_j c_j$, is called the *order* of the reaction. For example, the reaction

$$2\mathrm{X}_1 + \mathrm{X}_3 \xrightarrow{k} \mathrm{X}_2, \tag{3.1}$$

is of third order, with reactant complex $2\mathrm{X}_1 + \mathrm{X}_3$ and product complex X_2.

Each reaction is associated with a stoichiometric, or state-change, or simply reaction vector

$$\boldsymbol{\zeta} = C_2 - C_1 \in \mathbb{Z}^n,$$

whose components record the net change in the number of each species when the reaction occurs. Geometrically, the reaction vectors define allowed directions of motion in the state space. For example, reaction (3.1) has reaction vector $\boldsymbol{\zeta} = (-2, 1, 1)^\top$. Collecting all reaction vectors as columns gives the *stoichiometric matrix*

$$\Gamma = [\boldsymbol{\zeta}_1 \ \boldsymbol{\zeta}_2 \ \cdots \ \boldsymbol{\zeta}_m].$$

3.2 Deterministic models of CRNs

In deterministic models, the state of the system is a vector

$$\boldsymbol{x}(t) \in \mathbb{R}^n_{\geq 0},$$

whose components $x_i(t)$ represent concentrations of the species i. Let $v_j(\boldsymbol{x})$ be the rate of the jth reaction. Assuming the system is well mixed, the time evolution is governed by an ordinary differential equation of the form

$$\frac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t} = v_1(\boldsymbol{x})\boldsymbol{\zeta}_1 + \cdots + v_m(\boldsymbol{x})\boldsymbol{\zeta}_m \equiv \Gamma \mathbf{v}(\boldsymbol{x}(t)), \tag{3.2}$$

where $\mathbf{v}(x) = (v_1(x), \ldots, v_m(x))$ is the vector of reaction rates.

Given an initial condition at time 0, we can rewrite (3.2) as an integral equation

$$\boldsymbol{x}(t) = \boldsymbol{x}(0) + \Gamma \left[\int_0^t \mathbf{v}(\boldsymbol{x}(s))\mathrm{d}s \right], \tag{3.3}$$

which could also be written

$$\boldsymbol{x}(t) = \boldsymbol{x}(0) + \left[\int_0^t v_1(\boldsymbol{x}(s))\mathrm{d}s \right] \zeta_1 + \left[\int_0^t v_2(\boldsymbol{x}(s))\mathrm{d}s \right] \zeta_2 + \cdots + \left[\int_0^t v_m(\boldsymbol{x}(s))\mathrm{d}s \right] \zeta_m . \tag{3.4}$$

The *stoichiometric class* of the chemical reaction network is the set of all nonnegative states that you can reach from a given initial state $\boldsymbol{x}(0)$ using the reaction vectors of the network, that is,

$$(\boldsymbol{x} + \mathrm{im}\Gamma) \cap \mathbb{R}^n_{\geq 0}$$

Later on, we will see stochastic analogues of all of these equations, and it will be helpful to look back to compare and contrast the determistic and stochastic evolutions equations.

The formulation (3.4) makes clear that the deterministic dynamics is a superposition of the reaction vectors, weighted by their rates. Conservation laws arise naturally: each element of the kernel of $\Gamma^\top$ corresponds to a linear conserved quantity.

Conservation laws of a CRN

Given a CRN with stoichiometric matrix Γ, let $\boldsymbol{p} \in \ker \Gamma^\top$. Then, along any trajectory of the system (3.2),

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{p}^\top \boldsymbol{x} = \boldsymbol{p}^\top \Gamma \mathbf{v}(\boldsymbol{x}) = 0, \tag{3.5}$$

i.e., $\boldsymbol{p}^\top \boldsymbol{x}(t)$ is a conserved (linear) quantity.

Furthermore, the following are equivalent:

(i) There exists a positive vector $\boldsymbol{p} \in \ker \Gamma^\top$

(ii) all stoichiometric classes of the CRN are bounded.

We will see that this is also true in the stochastic case.

Exercise 3.1 (Necessary and sufficient conditions for bounded stoichiometric classes). We may assume the following result which is a straightforward corollary of Farkas' Lemma in convex geometry (it is sometimes called "Gordan's alternative"):

> Let S be a linear subspace of $\mathbb{R}^n$, and $S^\perp$ the orthogonal complement of S. Then one of the following two occurs:
>
> - $S \cap \mathbb{R}^n_{\geq 0} = \{0\}$; or

- $S^\perp \cap \mathbb{R}^n_+ \neq \emptyset$.

With the help of this result, prove that the following are equivalent for a CRN with stoichiometric matrix Γ. (i) There exists a positive vector $\boldsymbol{p} \in \ker \Gamma^\top$; (ii) all stoichiometric classes of the CRN are bounded. (It may help to note from basic linear algebra that $\operatorname{im}\Gamma$ and $\ker \Gamma^\top$ are orthogonal subspaces of $\mathbb{R}^n$.)

Mass action kinetics The function $v_j(x)$ specifies how fast the jth reaction occurs as a function of the current state. A common choice is *mass action kinetics*, where the rate is proportional to the product of concentrations appearing in the reactant complex. For example, if (3.1) has rate constant k, the mass action reaction rate is $v_j = kx_1^2 x_3$. More generally, a reaction with reactant complex

$$c_1\mathrm{X}_1 + c_2\mathrm{X}_2 + \cdots + c_n\mathrm{X}_n\,,$$

with mass action kinetics has rate

$$v_j(\boldsymbol{x}) = k\, x_1^{c_1} x_2^{c_2} \cdots x_n^{c_n}\,, \tag{3.6}$$

where x_i is the concentration of X_i, and k is the deterministic rate constant of the reaction j.

3.3 From deterministic rates to stochastic intensities

In stochastic models, we no longer track concentrations but instead track molecule numbers. The state of the system is now a random variable

$$\boldsymbol{X}(t) \in \mathbb{Z}^n_{\geq 0}.$$

The same reaction network underlies both the deterministic and stochastic models, with the same reaction vectors $\boldsymbol{\zeta}_j$ and stoichiometric matrix Γ. The difference between the two modelling frameworks lies in how reaction kinetics are specified:

- In the deterministic model, reaction j is assigned a reaction rate $v_j(\boldsymbol{x})$, where $\boldsymbol{x}$ denotes concentrations. This rate determines the contribution of reaction j to the time derivative of the state via (3.2).
- In the stochastic model, reaction j is assigned an *intensity* (or *propensity*) $\alpha_j(\boldsymbol{X})$, where $\boldsymbol{X}$ denotes molecule counts. The intensity is interpreted such that $\alpha_j(\boldsymbol{X})\Delta t$ is the probability that reaction j occurs in a short time interval of length Δt, given that the system is currently in state $\boldsymbol{X}$ (as already seen in Lecture 2 for two simple reactions).

Note that $\alpha_j(\boldsymbol{X})$ has units of 1/time, whereas the deterministic reaction rate $v_j(\boldsymbol{x})$ has units of concentration per unit time.

The function $\alpha_j(\boldsymbol{X})$ specifies the instantaneous rate at which reaction j occurs when the system is in state $\boldsymbol{X}$. Together with the reaction vectors ζ_j, the collection of intensities $\{\alpha_j\}_{j=1}^m$ completely specifies the stochastic model. In the next section, we show how these intensities determine the distribution of reaction times and reaction choices, leading to an explicit algorithm for simulating sample paths.

3.3.1 Stochastic mass action kinetics

Analogously to the deterministic mass action rates, a common choice is to consider stochastic mass action kinetics. Consider again the reaction (3.1) in a volume ν. For this reaction to occur, two molecules of X_1 and a molecule of X_3 must meet, and hence the rate should be proportional to the number of triplets available at any given time, namely

$$\alpha_j(\boldsymbol{X}) = \kappa_j \binom{X_1}{2} X_3 = \kappa_j \frac{X_1(X_1-1)X_3}{2},$$

where κ_j is the stochastic mass action rate constant. The rate constant κ_j scales differently with the container volume ν depending on the order of the reaction. For example, we already saw in Lecture **??** that for a zeroth order reaction, $\kappa_j = O(\nu)$, and for a first order reaction, $\kappa_j = O(1)$. For a second-order reaction, say $2\mathrm{X}_1 \to \emptyset$, the rate constant of two molecules of X_1 meeting scales with $1/\nu$: if we double the domain volume, the probability of meeting is halved. In general, we have that $\kappa_j = O(\nu^{1-|c|})$, where $|c|$ is the order of the reaction.

Another way to see the scaling of κ_j with ν is to compare the deterministic and stochastic mass action rates v_j and α_j. Let's consider the example (3.1) and a large volume ν such that $X_1(t) - 1 \approx X_1(t)$ for all times. Then the stochastic intensity satisfies

$$\alpha_j(\boldsymbol{X}) \approx \kappa_j \frac{X_1^2 X_3}{2}. \tag{3.7}$$

On the deterministic side, the mass action rate is written in terms of concentrations $\boldsymbol{x} = \boldsymbol{X}/\nu$ as

$$v_j(\boldsymbol{x}) = k\, x_1^2 x_3 = k \frac{X_1^2 X_3}{\nu^3}.$$

Since a single reaction event changes molecule numbers by an amount of order one, the quantity $\nu v_j(\boldsymbol{x})$ represents an event rate and can be directly compared with the stochastic intensity $\alpha_j(\boldsymbol{X})$:

$$\nu v_j(\boldsymbol{x}) = k \frac{X_1^2 X_3}{\nu^2} = \kappa_j \frac{X_1^2 X_3}{2} \quad \to \quad \kappa_j \approx \frac{2k}{\nu^2}. \tag{3.8}$$

With the above example in mind, it becomes easy to write down intensities in terms of deterministic mass action rate constants. The general rule is that a reaction with reactant complex $c_1\mathrm{X}_1 + \cdots + c_n\mathrm{X}_n$, with $|c| := \sum_i c_i$ and deterministic mass action rate constant k, has stochastic mass action intensity

$$\alpha_j(\boldsymbol{X}) \approx \frac{k}{\nu^{|c|-1}} \frac{(X_1)!}{(X_1-c_1)!} \frac{(X_2)!}{(X_2-c_2)!} \cdots \frac{(X_n)!}{(X_n-c_n)!}, \tag{3.9}$$

where X_k is the number of molecules of species X_k. It is important to remember that for reactions of order 2 or more, the correspondence only holds approximately, becoming more accurate for large numbers of molecules.

In general, when writing a mass action reaction between two complexes $C_1 \xrightarrow{\mu} C_2$, it is important to specify which rate is being used. Namely, whether μ denotes a deterministic rate constant (such as k above) or a stochastic rate constant (such as κ above). Throughout this course, we adopt the convention that the rate written above the reaction arrow refers to the *deterministic* mass action rate constant. The corresponding stochastic mass action intensity is then defined using (3.9).

3.4 The Gillespie algorithm

We are ready to discuss how to simulate a general stochastic CRN by constructing its sample paths using the most important algorithm for the stochastic simulation of CRNs: the so-called Gillespie algorithm (see also note on p. 21).

Fix a state $\boldsymbol{X} \in \mathbb{Z}^n_{\geq 0}$. We now explain how the intensities $\alpha_1, \ldots, \alpha_m$ fully determine the stochastic evolution of a chemical reaction network. One convenient way to think about the dynamics is to imagine that each reaction channel is equipped with an independent random clock whose ringing time is exponentially distributed with parameter $\alpha_j(\boldsymbol{X})$.

Lemma 3.1 (Competing exponentials (Q1 in Problem Sheet 1))**.** Let $T_1, \ldots, T_m$ be independent exponential random variables with parameters $\alpha_1, \ldots, \alpha_m$. Then:

1. $T := \min_j T_j$ is exponentially distributed with parameter $\alpha_0 = \sum_{j=1}^m \alpha_j$;
2. $\mathbb{P}(T = T_j) = \alpha_j/\alpha_0$;
3. T and the index of the minimiser are independent.

Suppose the system is currently in state $\boldsymbol{X}$ and the reaction intensities are $\alpha_1(\boldsymbol{X}), \ldots, \alpha_m(\boldsymbol{X})$. By Lemma 3.1, the time until the next reaction occurs is exponentially distributed with parameter

$$\alpha_0(\boldsymbol{X}) = \sum_{j=1}^{m} \alpha_j(\boldsymbol{X}),$$

and, conditional on a reaction occurring, reaction j is chosen with probability $\alpha_j(\boldsymbol{X})/\alpha_0(\boldsymbol{X})$. Moreover, the waiting time and the reaction index are independent.

These observations lead directly to the Gillespie algorithm:

Algorithm 3.1: Gillespie SSA

Set $\boldsymbol{X}(0) = \boldsymbol{X}_0$, $t = 0$ and final time t_f.
While $t < t_f$:

(a) At state $\boldsymbol{X}(t)$, compute $\alpha_j(\boldsymbol{X}(t))$ for all reactions and set $\alpha_0(\boldsymbol{X}(t)) = \sum_j \alpha_j(\boldsymbol{X}(t))$.

(b) Sample a waiting time with $r_1 \sim U(0,1)$ and $T = -\ln r_1/\alpha_0$.

(c) Choose a reaction index $J \in \{1, \ldots, m\}$ with $r_2 \sim U(0,1)$ and J such that

$$\sum_{i=1}^{J-1} \alpha_i \leq r_2\, \alpha_0 < \sum_{i=1}^{J} \alpha_i.$$

(d) Update $\boldsymbol{X}(t+T) = \boldsymbol{X}(t) + \zeta_J$ and set $t \leftarrow t + T$.

end

The Gillespie algorithm generates exact sample paths of the continuous-time Markov process associated with the stochastic CRN. Randomness enters through two independent mechanisms: the waiting time until the next reaction (T) and the identity of the reaction that fires (J).

3.5 The jump chain

If we ignore the waiting times and record only the sequence of states visited by the system, we obtain a discrete-time Markov chain called the jump chain. From a state $\boldsymbol{x}$, the jump chain moves to $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{\zeta}_j$ with probability

$$\hat{p}(\boldsymbol{x}, \boldsymbol{y}) = \frac{\alpha_j(\boldsymbol{x})}{\sum_k \alpha_k(\boldsymbol{x})}, \tag{3.10}$$

provided reaction j takes the system from $\boldsymbol{x}$ to $\boldsymbol{y}$. If, for some fixed state $\boldsymbol{x}$, $\hat{p}(\boldsymbol{x}, \boldsymbol{y}) = 0$ for all $\boldsymbol{y} \neq \boldsymbol{x}$, then $\boldsymbol{x}$ is an *absorbing state* and we set $\hat{p}(\boldsymbol{x}, \boldsymbol{x}) = 1$. We can put these jump probabilities $\hat{p}(\boldsymbol{x}, \boldsymbol{y})$ into a matrix, say $\widehat{P}$, to define a discrete-time Markov chain, termed the *jump chain* of the system, with transition matrix $\widehat{P} = (\hat{p}(\boldsymbol{x}, \boldsymbol{y}))_{\boldsymbol{x}, \boldsymbol{y} \in S}$ where $S \subseteq \mathbb{Z}^n_{\geq 0}$ is the state space.

The jump chain isolates the *state-to-state structure* of the stochastic CRN, independently of the random waiting times between reactions. By recording only the sequence of states visited by the system, and ignoring the times at which transitions occur, we obtain a discrete-time Markov chain whose transition probabilities encode which reactions are likely to fire from each state.

This viewpoint cleanly separates the *geometry* of the reaction network (which states can be reached, and how they are connected) from the *timing* of events. Many qualitative properties of the stochastic dynamics (such as reachability, communicating classes, absorbing states, and irreducibility) can be understood by studying the jump chain alone, without reference to the exponential waiting times.

In the next lecture, we move from individual sample paths to the evolution of probability distributions over the state space.

Tasks

- Do Exersice 3.1.
- Consider the one-species CRM

$$\emptyset \xrightarrow{k_1} \mathrm{A}, \qquad 2\mathrm{A} \xrightarrow{k_2} \emptyset, \tag{3.11}$$

  with mass action rates. Write the stoichiometric matrix Γ, the deterministic and stochastic mass action rates, v_j and α_j, respectively. Sample some trajectories of the stochastic model for several choices of $k_1, k_2, A(0)$ and compare them with the deterministic solution.

Example codes

`Lect3.ipynb`: contains an implementation of the Gillespie algorithm of (3.11).

Historical note. The terminology used in this course provides an example of *Stigler's law of eponymy*, which states that "no scientific discovery is named after its original discoverer". The Gillespie SSA builds on a long development, which we summarise briefly below:

- **1931:** Kolmogorov develops the analytic theory of Markov processes, including the forward and backward equations.
- **1940:** Feller establishes the exponential distribution of residence times for pure jump processes.
- **1942:** Doob introduces generalisations of Markov processes to include diffusion.
- **1950:** Kendall studies birth–death processes and implements simulations on the Manchester Mark I computer.
- **1953:** Bartlett applies stochastic process theory to models in epidemiology.
- **1977:** Gillespie introduces a stochastic simulation algorithm for chemical reaction networks.

The terminology introduced by Gillespie is now the most commonly used in the context of stochastic chemical kinetics, and is the one we will use in this course. The algorithm is also referred to as the *direct method*.

Lecture 4: Chemical reaction networks: distributions

In Lecture 3 we described stochastic chemical reaction networks (CRNs) at the level of *sample paths*, leading to the Gillespie SSA. In this lecture, we switch viewpoint: we study the time evolution of the *probability distribution* of the state. This leads to the transition probabilities, the infinitesimal generator, and the *chemical master equation* (CME), i.e. the forward Kolmogorov equation for a CRN. The generator also provides a direct route to evolution equations for moments.

4.1 Transition probabilities

Let $\boldsymbol{X} = (\boldsymbol{X}(t))_{t\geq 0}$ be the stochastic CRN from Lecture 3, with state space $S \subseteq \mathbb{Z}^n_{\geq 0}$ and autonomous[1] propensities $\alpha_j(\boldsymbol{x})$ for reactions $j = 1, \ldots, m$. For $\boldsymbol{x}, \boldsymbol{y} \in S$ and $t \geq 0$, define the transition probabilities

$$p_{\boldsymbol{x}\boldsymbol{y}}(t) := \mathbb{P}(\boldsymbol{X}(t) = \boldsymbol{y} \mid \boldsymbol{X}(0) = \boldsymbol{x}). \tag{4.1}$$

We have seen that the probability that the jth reaction occurs in a time interval Δt is $\alpha_j(\boldsymbol{x})\Delta t + o(\Delta t)$. For simplicity, assume that there is a single reaction j that takes the system from state $\boldsymbol{x}$ to state $\boldsymbol{y}$, that is, $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{\zeta}_j$.[2] Then, as $\Delta t \to 0^+$, we have

$$p_{\boldsymbol{x}\boldsymbol{y}}(\Delta t) = \alpha_j(\boldsymbol{x})\,\Delta t + o(\Delta t) \qquad \text{and} \qquad p_{\boldsymbol{x}\boldsymbol{x}}(\Delta t) = 1 - \sum_{j=1}^{m} \alpha_j(\boldsymbol{x})\Delta t + o(\Delta t). \tag{4.2}$$

We write $P(t) = (p_{\boldsymbol{x}\boldsymbol{y}}(t))_{\boldsymbol{x},\boldsymbol{y}\in S}$ for the matrix of transition probabilities at time t. It has the following properties:

- $P(0) = \mathbb{I}$, the identity matrix.
- $P(t)$ is a stochastic matrix, that is, $P(t)$ has non-negative entries and row sums 1.
- P satisfies the Chapman-Kolmogorov equation (A.4),

$$P(t+s) = P(t)P(s), \qquad t, s \geq 0, \tag{4.3}$$

  equivalently $p_{\boldsymbol{x}\boldsymbol{y}}(t+s) = \sum_{\boldsymbol{z}\in S} p_{\boldsymbol{x}\boldsymbol{z}}(t)p_{\boldsymbol{z}\boldsymbol{y}}(s)$. Eq. (4.3) is also known as the *semigroup property*.

The transition probabilities $P(t)$ describe the distribution of a continuous-time Markov chain at every time t, but they are more detailed than we usually need. Instead, we work with the generator Q, which captures the infinitesimal jump rates and determines $P(t)$ through the Kolmogorov equations.

Remark 4.1 (Multispecies)**.** For n species, $\boldsymbol{x}$ is an n-dimensional vector in $S \subseteq \mathbb{Z}^n_{\geq 0}$. The transition matrix P is indexed by states (and not species).

[1] Autonomous means that there is no explicit time dependence. Autonomous rates imply that the CRN is time-homogeneous.

[2] If there were two such reactions, we would simply add the two propensities into α_j.

4.2 The infinitessimal generator Q

The *generator* (or infinitessimal generator) is defined by[3]

$$Q := P'(0) = \lim_{h\to 0^+} \frac{P(h) - \mathbb{I}}{h}. \tag{4.4}$$

Writing its entries as $Q = (q_{\boldsymbol{xy}})_{\boldsymbol{x},\boldsymbol{y}\in S}$, the following properties follow immediately from (4.4):

$$\sum_{\boldsymbol{y}\in S} q_{\boldsymbol{xy}} = 0, \qquad q_{\boldsymbol{xy}} \geq 0 \quad (\boldsymbol{y} \neq \boldsymbol{x}), \qquad q_{\boldsymbol{xx}} \leq 0.$$

Or equivalently, for small h,

$$p_{\boldsymbol{xy}}(h) = \begin{cases} q_{\boldsymbol{xy}}\, h + o(h), & \boldsymbol{y} \neq \boldsymbol{x}, \\ 1 + q_{\boldsymbol{xx}}\, h + o(h), & \boldsymbol{y} = \boldsymbol{x}. \end{cases} \tag{4.5}$$

For a stochastic CRN, the only possible transitions are reaction jumps $\boldsymbol{x} \mapsto \boldsymbol{x} + \boldsymbol{\zeta}_j$ with rates given by the propensities $\alpha_j(\boldsymbol{x})$. Therefore, using (4.2), the entries of Q take the sparse form

$$q_{\boldsymbol{xy}} = \begin{cases} \alpha_j(\boldsymbol{x}), & \boldsymbol{y} = \boldsymbol{x} + \boldsymbol{\zeta}_j \text{ for some } j, \\ -\sum_{k=1}^m \alpha_k(\boldsymbol{x}), & \boldsymbol{y} = \boldsymbol{x}, \\ 0, & \text{otherwise.} \end{cases} \tag{4.6}$$

Remark 4.2 (Relationship between Q and the jump-chain matrix $\widehat{P}$)**.** Recalling the definition of the jump chain in (3.10), the corresponding transition matrix $\widehat{P}$ has entries

$$\hat{p}_{\boldsymbol{xy}} = \begin{cases} \frac{q_{\boldsymbol{xy}}}{-q_{\boldsymbol{xx}}}, & \boldsymbol{y} \neq \boldsymbol{x}, \\ 0, & \boldsymbol{y} = \boldsymbol{x}, \end{cases} \tag{4.7}$$

provided $q_{\boldsymbol{xx}} \neq 0$ (which is true if $\boldsymbol{x}$ is not an absorbing state). If $\boldsymbol{x}$ is absorbing, then $\hat{p}_{\boldsymbol{xx}} = 1$ and $\hat{p}_{\boldsymbol{xy}} = 0$ for $\boldsymbol{y} \neq \boldsymbol{x}$.

Example 4.1 (Q **matrix and jump chain of a CRN with finite state space).** Consider the CRN with four reactions

$$\mathrm{A} \xrightarrow{1} \mathrm{B} + \mathrm{C}, \quad \mathrm{B} + \mathrm{C} \xrightarrow{1} \mathrm{A}, \quad B \xrightarrow{1/2} \mathrm{C}, \quad \mathrm{C} \xrightarrow{1/2} \mathrm{B}\,.$$

Let us assume that all the reactions occur in a container of volume $\nu = 1$ and have deterministic mass action rate constants as shown. There is a strictly positive vector in $\ker \Gamma^\top$, where Γ is the stoichiometric matrix (check this!), and so, given any initial state, this CRN has finite state space (see Exercise 3.1).

[3] This assumes that the transition matrix $P(t)$ is right-differentiable at $t = 0$, which is true when the process is right-continuous. This means that $\boldsymbol{X}(t) = \lim_{s\to t^+} \boldsymbol{X}(s)$. This property is automatic when we consider systems with only a finite number of distinct reactions. It can also hold for systems with infinitely many reactions, but then we need to think more carefully about the intensities.

Suppose we initially have one molecule of A, $\boldsymbol{X}(0) = (1,0,0)^\top$. The system then has four states it can reach from this initial state:

$$S = \big\{\{\mathrm{A}\},\ \{\mathrm{B},\mathrm{C}\},\ \{\mathrm{B},\mathrm{B}\},\ \{\mathrm{C},\mathrm{C}\}\big\},$$

(or, as vectors, $S = \{(1,0,0),(0,1,1),(0,2,0),(0,0,2)\}$.) The Q-matrix and transition matrix of the corresponding jump-chain are:

$$Q = \begin{pmatrix} -1 & 1 & 0 & 0 \\ 1 & -2 & 1/2 & 1/2 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix} \quad \text{and} \quad \widehat{P} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1/2 & 0 & 1/4 & 1/4 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

Remark 4.3 (Infinite state space S)**.** It is technically convenient to assume that the state space is finite, i.e., $|S| < \infty$. Most of the results below, however, also extend to countably infinite state spaces provided the total jump rate is uniformly bounded:

$$\sup_{\boldsymbol{x}\in S}\big(-q(\boldsymbol{x},\boldsymbol{x})\big) = \sup_{\boldsymbol{x}\in S}\sum_{\boldsymbol{y}\neq\boldsymbol{x}} q(\boldsymbol{x},\boldsymbol{y}) < \infty.$$

This *non-explosion* condition rules out the possibility of infinitely many jumps occurring in finite time. We will therefore also consider examples with $|S| = \infty$, although we will omit the necessary extensions here.

4.3 Forward and backward Kolmogorov equations

Using the short-time behaviour derived in Section 4.1, we now derive a differential equation that determines the transition probabilities. This viewpoint is very useful in modelling, since dynamical systems are often observed and characterised through their short-time behaviour, e.g.,

$$m\,\mathrm{d}v = F\,\mathrm{d}t, \qquad \mathrm{d}x = v\,\mathrm{d}t.$$

Assuming that $|S| < \infty$ and using the Chapman–Kolmogorov property (4.3), we compute

$$P'(t) = \lim_{h\to 0^+}\frac{P(t+h)-P(t)}{h} = \lim_{h\to 0^+}\frac{P(t)P(h)-P(t)\mathbb{I}}{h} = P(t)\left(\lim_{h\to 0^+}\frac{P(h)-I}{h}\right) = P(t)Q.$$

We may translate (4.8) into an equation for the probabilities

$$p_{\boldsymbol{y}}(t) = \mathbb{P}(\boldsymbol{X}(t) = \boldsymbol{y}).$$

By conditioning on the initial state,

$$p_{\boldsymbol{y}}(t) = \sum_{\boldsymbol{x}\in S} p_{\boldsymbol{x}}(0)\, p_{\boldsymbol{x}\boldsymbol{y}}(t).$$

Differentiating in time and using (4.8) yields

$$\frac{d}{dt}p_{\boldsymbol{y}}(t) = \sum_{\boldsymbol{x}\in S} p_{\boldsymbol{x}}(t)\, q_{\boldsymbol{x}\boldsymbol{y}}.$$

We obtain two versions of the forward Kolmogorov equation

Forward Kolmogorov equation

For a time-homogeneous continuous-time Markov chain with generator Q, the transition probabilities $P(t)$ evolve as

$$\frac{d}{dt}P(t) = P(t)Q, \qquad P(0) = \mathbb{I}, \tag{4.8}$$

and the probability distribution $p_{\boldsymbol{y}}(t)$ evolves as

$$\frac{dp_{\boldsymbol{y}}}{dt} = \sum_{\boldsymbol{x}\in S} p_{\boldsymbol{x}}(t)\, q_{\boldsymbol{x}\boldsymbol{y}}, \qquad p_{\boldsymbol{y}}(0) = \delta_{\boldsymbol{y}\boldsymbol{y}_0}, \tag{4.9}$$

where $\boldsymbol{y}_0$ is the (deterministic) initial state of the system.

Returning to the initial calculation, we may instead factor $P(t)$ on the right:

$$P'(t) = \lim_{h\to 0^+} \frac{P(t+h) - P(t)}{h} = \lim_{h\to 0^+} \frac{P(h)P(t) - \mathbb{I}P(t)}{h} = \left(\lim_{h\to 0^+} \frac{P(h) - \mathbb{I}}{h}\right) P(t) = QP(t).$$

This form is naturally connected with expectations of functions of the process. Let $f : S \to \mathbb{R}$ be a test function and define

$$u(\boldsymbol{x}, t) = \mathbb{E}\big[f(\boldsymbol{X}(t)) \mid \boldsymbol{X}(0) = \boldsymbol{x}\big].$$

Then

$$u(\boldsymbol{x}, t) = \sum_{\boldsymbol{y}\in S} p_{\boldsymbol{x}\boldsymbol{y}}(t)\, f(\boldsymbol{y}).$$

Differentiating in time and using (4.10) gives

$$\frac{d}{dt}u(\boldsymbol{x}, t) = \sum_{\boldsymbol{y}\in S} q_{\boldsymbol{x}\boldsymbol{y}}\, u(\boldsymbol{y}, t).$$

We then obtain two versions of the backward Kolmogorov equation:

Backward Kolmogorov equation

For a time-homogeneous continuous-time Markov chain with generator Q, the transition probabilities $P(t)$ evolve as

$$\frac{d}{dt}P(t) = QP(t), \qquad P(0) = \mathbb{I}, \tag{4.10}$$

and the expectation $u(\boldsymbol{x}, t) = \mathbb{E}[f(\boldsymbol{X}(t)) \,|\, \boldsymbol{X}(0) = \boldsymbol{x}]$ evolves as

$$\frac{d}{dt}u(\boldsymbol{x}, t) = \sum_{\boldsymbol{y}\in S} q_{\boldsymbol{x}\boldsymbol{y}} u(\boldsymbol{y}, t), \qquad u(\boldsymbol{x}, 0) = f(\boldsymbol{x}). \tag{4.11}$$

Why "forward" and "backward"? Recall that $q_{\boldsymbol{xy}}$ is the rate at which the process jumps from state $\boldsymbol{x}$ to state $\boldsymbol{y}$. Therefore, a row index of Q gives the source state and a column index fixes the destination state. The asymmetry in Q is what gives rise to the distinction between forward and backward equations. The forward equation sums over a fixed *column* of Q and tracks probability flowing into a state, while the backward equation sums over a fixed *row* of Q and tracks how future transitions affect expectations. Both equations are generated by the same matrix Q, but they encode different viewpoints on the dynamics.

When the state space is finite, the forward Kolmogorov equation (4.8) is a linear system of ODEs with constant coefficients. Its unique solution is

$$P(t) = e^{Qt} = \sum_{n=0}^{\infty} \frac{Q^n t^n}{n!}.$$

Also, combining (4.8) and (4.10) shows that $QP(t) = P(t)Q$, that is, the transition probability matrix commutes with the generator. Finally, the backward equation suggests viewing the generator as an operator acting on functions, which we make precise next.

4.4 The chemical master equation (CME)

For a CRN, the forward Kolmogorov equation (4.9) is typically written in *inflow–outflow* form and is called the chemical master equation. Let $p_{\boldsymbol{x}}(t) = \mathbb{P}(\boldsymbol{X}(t) = \boldsymbol{x})$. Using (4.6), (4.9) reads

$$\begin{aligned} \frac{d}{dt} p_{\boldsymbol{x}}(t) &= \sum_{\boldsymbol{y} \in S} p_{\boldsymbol{y}}(t) q_{\boldsymbol{yx}} \\ &= \sum_{j=1}^{m} p_{\boldsymbol{x}-\boldsymbol{\zeta}_j}(t)\, \alpha_j(\boldsymbol{x} - \boldsymbol{\zeta}_j) - p_{\boldsymbol{x}}(t) \sum_{j=1}^{m} \alpha_j(\boldsymbol{x}), \end{aligned} \tag{4.12}$$

where terms with $\boldsymbol{x} - \boldsymbol{\zeta}_j \notin S$ are interpreted as zero (equivalently, α_j vanishes there). The gain term $p_{\boldsymbol{x}-\boldsymbol{\zeta}_j}(t)\alpha_j(\boldsymbol{x} - \boldsymbol{\zeta}_j)$ is the rate at which probability flows *into* $\boldsymbol{x}$ via reaction j, while the loss term $p_{\boldsymbol{x}}(t)\alpha_j(\boldsymbol{x})$ is the rate at which probability flows *out of* $\boldsymbol{x}$ via reaction j.

4.5 The generator as an operator and moment equations

So far, the generator Q has been introduced as a matrix describing transition rates between states. A complementary and very useful viewpoint is to regard the generator as an operator acting on functions of the state. This is the natural setting for studying expectations and moments of the process.

The generator acting on functions Let $f : S \to \mathbb{R}$ be a test function of the state. We define the generator $\mathcal{L}$ acting on f by

$$(\mathcal{L}f)(\boldsymbol{x}) := \sum_{\boldsymbol{y} \in S} q_{\boldsymbol{xy}} f(\boldsymbol{y}). \tag{4.13}$$

This definition is simply the action of the matrix Q on the vector f, evaluated at the state $\boldsymbol{x}$.

In the case of a stochastic chemical reaction network, using the sparse structure (4.6) of the generator, this can be written in the more explicit reaction-sum form

$$(\mathcal{L}f)(\boldsymbol{x}) = \sum_{j=1}^{m} \alpha_j(\boldsymbol{x})\big(f(\boldsymbol{x}+\boldsymbol{\zeta}_j) - f(\boldsymbol{x})\big). \tag{4.14}$$

Each term represents the contribution of reaction j: the rate at which it fires, multiplied by the change it induces in the observable f.

Connection with the backward Kolmogorov equation Recall from Section 4.3 that, for a given function f,

$$u(\boldsymbol{x},t) = \mathbb{E}\big[f(\boldsymbol{X}(t)) \mid \boldsymbol{X}(0) = \boldsymbol{x}\big]$$

satisfies the backward Kolmogorov equation (4.11). In operator form, this may be written compactly as

$$\frac{d}{dt}u(\cdot,t) = \mathcal{L}u(\cdot,t).$$

Thus, the generator $\mathcal{L}$ governs the time evolution of expectations of functions of the process.

A key consequence of this operator viewpoint is the following identity, often referred to as Dynkin's formula in differential form.

Generator identity (Dynkin's formula)

For a sufficiently regular function $f : S \to \mathbb{R}$,

$$\frac{d}{dt}\,\mathbb{E}[f(\boldsymbol{X}(t))] = \mathbb{E}\big[(\mathcal{L}f)(\boldsymbol{X}(t))\big]. \tag{4.15}$$

Equation (4.15) states that the time derivative of the expected value of $f(\boldsymbol{X}(t))$ is obtained by first applying the generator to f, and then taking the expectation. In practice, this allows us to derive evolution equations for moments by choosing f appropriately. For example,

- Choosing $f(\boldsymbol{x}) = x_i$ yields an equation for the mean $\mathbb{E}[\boldsymbol{X}_i(t)]$.
- Choosing $f(\boldsymbol{x}) = x_i x_j$ yields equations for second moments and covariances.

For nonlinear reaction networks, these equations typically do not close, leading to the moment-closure problem discussed later.

4.6 Running example: CME and mean equation

Consider the one-species CRN (as in Lecture 3)

$$2A \xrightarrow{k_1} \emptyset, \qquad \emptyset \xrightarrow{k_2} A, \tag{4.16}$$

with state $A(t) \in \mathbb{Z}_{\geq 0}$ and propensities (taking the volume $\nu \equiv 1$)

$$\alpha_1(n) = k_1 n(n-1), \qquad \alpha_2(n) = k_2.$$

The reaction "vectors" are $\zeta_1 = -2$ and $\zeta_2 = +1$.

Generator From (4.13), for any $f : \mathbb{Z}_{\geq 0} \to \mathbb{R}$,

$$(\mathcal{L}f)(n) = k_1 n(n-1)\big(f(n-2) - f(n)\big) + k_2\big(f(n+1) - f(n)\big). \tag{4.17}$$

Chemical master equation Let $p_n(t) := \mathbb{P}(A(t) = n)$. Then (4.12) gives, for $n \geq 0$,

$$\frac{dp_n}{dt} = k_1(n+2)(n+1)\, p_{n+2}(t) - k_1 n(n-1)\, p_n(t) + k_2\, p_{n-1}(t) - k_2\, p_n(t), \tag{4.18}$$

with the convention $p_{-1} \equiv 0$ (and similarly $p_{-2} \equiv 0$).

Evolution of the mean Let $M(t) = \mathbb{E}[A(t)]$ and choose $f(n) = n$. From (4.17),

$$(\mathcal{L}f)(n) = k_1 n(n-1)\big((n-2) - n\big) + k_2\big((n+1) - n\big) = -2k_1 n(n-1) + k_2.$$

Hence by (4.15),

$$M'(t) = k_2 - 2k_1\, \mathbb{E}[A(t)(A(t)-1)] = k_2 - 2k_1\big(\mathbb{E}[A(t)^2] - M(t)\big). \tag{4.19}$$

This does *not* close in terms of $\mathbb{E}[A(t)]$ alone, because the bimolecular reaction is nonlinear.

Note that we could also have obtained (4.19) by multiplying the chemical master equation (15.16) by n and summing over all states (as we did in Exercise 2.1). However, this tends to get algebraically messier. The fact that both approaches lead to the same result reflects the duality between the forward and backward Kolmogorov equations.

Comparison with the deterministic model. If $a(t)$ denotes the deterministic concentration of A, the reaction-rate equation is

$$\frac{da}{dt} = k_2 - 2k_1 a^2. \tag{4.20}$$

Comparing (4.19) and (4.20) highlights a key point: *the deterministic dynamics need not coincide with the stochastic mean dynamics.*

Exercise 4.1 (Mean versus deterministic solution)**.** Use the Gillespie SSA from Lecture 3 to estimate $\mathbb{E}[A(t)]$ for (4.16) by averaging many sample paths. Solve (4.20) numerically for the same parameters and compare. How does the discrepancy change as typical molecule numbers increase?

Tasks

- Write the generator $\mathcal{L}$ and the CME (4.12) of the CRN in Example 4.1.
- Use (4.15) with $f(\boldsymbol{x}) = A$ to derive the evolution of the mean $M_A(t) := \mathbb{E}[A(t)]$ for the CRN in Example 4.1.
- For the CRN (4.16), compare the simulated mean with the deterministic reaction-rate equation.

Example codes

`Lect4.ipynb`: empirical mean/variance of the CRN (4.16) from SSA; comparison with a deterministic ODE.

Lecture 5: Stochastic phenomena in Chemical Reaction Networks

In the previous lectures we analysed stochastic chemical reaction networks (CRNs) at the level of sample paths (Gillespie SSA) and distributions (generators and Kolmogorov equations). In this lecture we highlight stochastic phenomena that are *invisible* in the deterministic reaction-rate equations.

We distinguish two complementary sources of stochastic effects.

- **Linear CRNs:** moment equations close and the deterministic model agrees with the mean. Nevertheless, the stochastic model may exhibit qualitatively different long-time behaviour, for example because *absorbing states* allow extinction with positive probability.
- **Nonlinear CRNs:** moment equations do not close. Any attempt to approximate the mean dynamics by a finite-dimensional "moment closure" can introduce errors already at the mean-field level. Generating functions provide an alternative description of the full distribution.

5.1 Birth–death with extinction

Consider the linear CRN

$$\mathrm{A} \xrightarrow{\lambda} 2\mathrm{A}, \qquad \mathrm{A} \xrightarrow{\mu} \emptyset, \tag{5.1}$$

with state $A(t) \in \mathbb{Z}_{\geq 0}$ and propensities (taking the volume $\nu \equiv 1$)

$$\lambda_n = \lambda n, \qquad \mu_n = \mu n.$$

The reaction vectors are $\zeta_1 = +1$ and $\zeta_2 = -1$. The state $n = 0$ is absorbing: if no molecules of A are present then no reaction can occur, and $A(t) = 0$ for all future time.

In Q2 of Sheet 1 you show that the probability of extinction is positive for all $\lambda, \mu > 0$. The corresponding deterministic reaction-rate equation is

$$\frac{\mathrm{d}a}{\mathrm{d}t} = (\lambda - \mu)a,$$

so if $\lambda > \mu$ then the deterministic solution grows exponentially and never reaches zero. Thus, despite being a linear CRN for which the deterministic model agrees with the mean dynamics, the stochastic model may still reach an absorbing state due to random fluctuations.

Systems with absorbing states highlight a fundamental difference between deterministic and stochastic models: in the stochastic model, random fluctuations can take the system to an absorbing state with positive probability, and after this recovery is impossible; this cannot occur in the deterministic model.

5.2 Conservation laws and stoichiometric structure

We now place the extinction phenomenon of (5.1) into a more general structural framework. The key idea is that the long-time behaviour of a stochastic CRN is strongly constrained by its stoichiometric structure, independently of the precise values of the rate constants.

Let $\Gamma = [\boldsymbol{\zeta}_1 \cdots \boldsymbol{\zeta}_m] \in \mathbb{Z}^{n\times m}$ denote the stoichiometric matrix of a CRN with n species and m reactions. Recall from Lecture 3 that for the deterministic model with concentration vector $\boldsymbol{x}(t)$,

$$\frac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t} = \Gamma \mathbf{v}(\boldsymbol{x}(t)),$$

any vector $\boldsymbol{p} \in \ker \Gamma^\top$ defines a conserved quantity $\boldsymbol{p}^\top \boldsymbol{x}(t)$.

The same conservation law holds for the stochastic model. Each reaction event updates the state as $\boldsymbol{X} \mapsto \boldsymbol{X} + \boldsymbol{\zeta}_j$; if $\boldsymbol{p}^\top \boldsymbol{\zeta}_j = 0$ for all j then $\boldsymbol{p}^\top \boldsymbol{X}(t)$ is unchanged by every jump, hence

$$\boldsymbol{p}^\top \boldsymbol{X}(t) = \boldsymbol{p}^\top \boldsymbol{X}(0) \qquad \text{for all } t \geq 0.$$

In particular, if there exists a strictly positive vector $\boldsymbol{p} \in \ker \Gamma^\top$ then the process is confined to a finite stoichiometric class (recall Exercise 3.1). In particular, explosion is impossible and the restricted process always admits a stationary distribution (though it need not be unique).

Example 5.1 (Reversible dimerisation)**.** Consider

$$\mathrm{A}_1 + \mathrm{A}_1 \rightleftharpoons \mathrm{A}_2.$$

The reaction vectors are $\boldsymbol{\zeta}_1 = (-2, 1)^\top$ and $\boldsymbol{\zeta}_2 = (2, -1)^\top$, and $\boldsymbol{p} = (1, 2)^\top \in \ker \Gamma^\top$. Thus $A_1(t) + 2A_2(t)$ is conserved, and the reachable state space is finite for any initial condition.

Example 5.2 (No conservation and absorbing states)**.** For (5.1),

$$\Gamma = (+1 \quad -1), \qquad \ker \Gamma^\top = \{0\},$$

so there is no conserved quantity restricting the evolution. From any state $n > 0$ the chain can reach 0; once at 0 it cannot return. Hence the chain is not irreducible and extinction is always possible.

In summary, the absence of conservation laws means that nothing prevents the process from eventually reaching the boundary. In contrast, in the presence of a conservation law the stoichiometric class is finite, and the stochastic process cannot drift arbitrarily far in state space.

Exercise 5.1 (Conservation laws and stationarity)**.** Consider a stochastic CRN with stoichiometric matrix $\Gamma = [\boldsymbol{\zeta}_1 \cdots \boldsymbol{\zeta}_m] \in \mathbb{Z}^{n\times m}$ and associated Markov chain $\boldsymbol{X}(t)$ on $\mathbb{Z}^n_{\geq 0}$.

Assume that there exists a vector $\boldsymbol{p} \in \ker \Gamma^\top$ such that $p_i > 0$ for all $i = 1, \ldots, n$.

(i) Show that $\boldsymbol{p}^\top \boldsymbol{X}(t) = \boldsymbol{p}^\top \boldsymbol{X}(0)$ almost surely for all $t \geq 0$.

(ii) Deduce that the stoichiometric class

$$\mathcal{S}_c := \{\boldsymbol{x} \in \mathbb{Z}^n_{\geq 0} : \boldsymbol{p}^\top \boldsymbol{x} = c\}, \qquad c = \boldsymbol{p}^\top \boldsymbol{X}(0),$$

is finite.

(iii) Conclude that the stochastic process admits at least one stationary distribution in $\mathcal{S}_c$.

5.3 Mean time to extinction: backward viewpoint

For the birth–death system (5.1), extinction occurs with positive probability from any initial state. A natural quantitative question is: *if extinction occurs, how long does it typically take?*

Define the (random) extinction time

$$\tau_n \;:=\; \inf\{t \geq 0 : A(t) = 0\} \qquad \text{given } A(0) = n,$$

and its mean

$$u_n \;:=\; \mathbb{E}[\tau_n].$$

Clearly $u_0 = 0$.

The key idea is to condition on the first reaction event. Starting from $n > 0$, the process (5.1) waits an exponential time

$$T_n \sim \mathrm{Exp}(\lambda_n + \mu_n),$$

then jumps to $n+1$ (birth) or $n-1$ (death). By the Markov property,

$$u_n = \mathbb{E}[T_n] + \mathbb{E}[u_{A(T_n)}].$$

Since $\mathbb{E}[T_n] = 1/(\lambda_n + \mu_n)$ and

$$\mathbb{P}(A(T_n) = n+1) = \frac{\lambda_n}{\lambda_n + \mu_n}, \qquad \mathbb{P}(A(T_n) = n-1) = \frac{\mu_n}{\lambda_n + \mu_n},$$

we obtain

$$u_n = \frac{1}{\lambda_n + \mu_n} + \frac{\lambda_n}{\lambda_n + \mu_n}\, u_{n+1} + \frac{\mu_n}{\lambda_n + \mu_n}\, u_{n-1}.$$

Rearranging,

$$-1 = \lambda_n(u_{n+1} - u_n) + \mu_n(u_{n-1} - u_n), \qquad u_0 = 0.$$

This is the backward Kolmogorov equation for the mean hitting time. (Mean hitting and exit times for diffusion processes satisfy an analogous backward equation, as we will see later.)

Mean time to extinction

Let $\mathcal{L}$ be the generator of a stochastic CRN with state $X(t)$. Define $\tau_n = \inf\{t \geq 0 : X(t) = 0\}$ given $X(0) = n$, and $u(n) = \mathbb{E}_n[\tau_n]$. Then u satisfies

$$-1 = (\mathcal{L}u)(n) \quad \text{for } n \geq 1, \qquad u(0) = 0. \tag{5.2}$$

So far we have seen a stochastic effect in a linear CRN. We now turn to nonlinear CRNs, where even the moment equations fail to close.

5.4 Nonlinear CRNs and probability generating functions

For a stochastic CRN with generator $\mathcal{L}$ and state $\boldsymbol{X}(t) \in \mathbb{N}_0^n$, expectations evolve according to (cf. (4.15))

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[f(\boldsymbol{X}(t))] = \mathbb{E}\big[(\mathcal{L}f)(\boldsymbol{X}(t))\big].$$

Choosing polynomial test functions $f(\boldsymbol{x}) = x_i$, $x_i x_j$, etc. yields moment equations. For nonlinear reactions this hierarchy does *not* close, so a finite-dimensional moment description typically requires an additional approximation (a "closure"), which can introduce errors already at the level of the mean. We saw an example of this in Section 4.6.

A complementary approach is to encode the entire distribution in the *probability generating function* (PGF). Let $p_{\boldsymbol{x}}(t) = \mathbb{P}(\boldsymbol{X}(t) = \boldsymbol{x})$ and define

$$G(\boldsymbol{z}, t) = \sum_{\boldsymbol{x} \in \mathbb{N}_0^n} \boldsymbol{z}^{\boldsymbol{x}} p_{\boldsymbol{x}}(t), \qquad \boldsymbol{z}^{\boldsymbol{x}} := \prod_{i=1}^{n} z_i^{x_i}.$$

Equivalently, $G(\boldsymbol{z}, t) = \mathbb{E}[\boldsymbol{z}^{\boldsymbol{X}(t)}]$, with $\boldsymbol{z}^{\boldsymbol{X}(t)} := \prod_{i=1}^{n} z_i^{X_i(t)}$. In general, derivatives of G at $\boldsymbol{z} = \mathbf{1}$ recover factorial moments. To see concretely how derivatives arise, it is helpful to work through a one-species example.

Example 5.3 (Nonlinear birth–death CRN)**.** Consider the CRN (4.16) from the end of Lecture 4,

$$2\mathrm{A} \xrightarrow{k_1} \emptyset, \qquad \emptyset \xrightarrow{k_2} \mathrm{A}.$$

Let $p_n(t) = \mathbb{P}(A(t) = n)$ and define the one-variable PGF

$$G(z, t) = \sum_{n \geq 0} z^n p_n(t).$$

First, note the identities

$$z \partial_z G(z, t) = \sum_{n \geq 0} n \, p_n(t) \, z^n, \qquad z^2 \partial_z^2 G(z, t) = \sum_{n \geq 0} n(n-1) \, p_n(t) \, z^n.$$

Thus factors of n and $n(n-1)$ become differential operators acting on G.

With our convention for the mass-action propensity of $2A \to \emptyset$,

$$\alpha_1(n) = k_1 \, n(n-1), \qquad \alpha_2(n) = k_2,$$

and reaction increments $\zeta_1 = -2$, $\zeta_2 = +1$. Recall that the generator (4.17). Using $G(z, t) = \mathbb{E}[f(A(t))]$ with $f(n) = z^n$ in (4.15) yields

$$\begin{aligned} \frac{\partial}{\partial t} G(z, t) = \mathbb{E}\left[(\mathcal{L} f)(A(t))\right] &= k_1 \mathbb{E}[A(t)(A(t)-1) z^{A(t)-2}(1 - z^2)] + k_2 \mathbb{E}[z^A(t)(z-1)] \\ &= k_1 (1 - z^2) \partial_z^2 G(z, t) + k_2 (z - 1) G(z, t). \end{aligned} \tag{5.3a}$$

This is a parabolic PDE in $z \in [-1, 1]$ that degenerates at the boundaries. It is complemented by two boundary conditions [5, see Eq. (1.38)]:

$$G(1) = 1, \qquad G(-1, t) = G(-1, 0) e^{-2k_2 t}. \tag{5.3b}$$

The former comes from the normalisation condition and the latter from imposing that $\partial_z^2 G$ bounded at $z = -1$.

Equation (5.3a) is nothing else than the backward Kolmogorov equation above written in a different basis. Differentiating (5.3a) and evaluating at $z = 1$ recovers moment equations; for instance,

$$\partial_z G(1, t) = \mathbb{E}[A(t)].$$

Exercise 5.2 (Stationary moments and distribution of Example 5.3)**.** Use (5.3) to show that the stationary probability generating function $G_s(z)$ solves

$$\frac{\mathrm{d}^2 G_s}{\mathrm{d}z^2} = \frac{k}{1+z} G_s, \quad G_s(-1) = 0, \quad G_s(1) = 1, \qquad \text{with } k = \frac{k_2}{k_1}.$$

The solution of this equation is

$$G_s(z) = \frac{\sqrt{1+z}\, I_1(2\sqrt{k(1+z)})}{\sqrt{2} I_1(2\sqrt{2k})},$$

where I_1 is the modified Bessel function of the first kind.

Use this to show that

$$M_s = \frac{1}{4} + \sqrt{\frac{k}{2}} \frac{I_1'\left(2\sqrt{2k}\right)}{I_1\left(2\sqrt{2k}\right)}, \qquad V_s = \frac{k}{2} + M_s - M_s^2,$$

and that the stationary distribution is

$$\phi(n) = C \frac{k^n}{n!} I_{n-1}(2\sqrt{k}), \quad n = 1, 2, 3, \ldots$$

[Hint: use the series expansion of $I_1(x)$ and collect terms proportional to x^n.]

5.5 Metastability and large system size

Nonlinear CRNs with large typical molecule numbers often exhibit *metastability*: the stochastic dynamics spends very long times fluctuating near a deterministic equilibrium, with rare noise-driven transitions to a different long-time regime.

A classical mechanism is bistability. Consider the one-species CRN

$$3\mathrm{A} \xrightarrow{k_1} 2\mathrm{A}, \qquad 2\mathrm{A} \xrightarrow{k_2} 3\mathrm{A}, \qquad \mathrm{A} \xrightarrow{k_3} \emptyset, \qquad \emptyset \xrightarrow{k_4} \mathrm{A}. \tag{5.4}$$

The deterministic concentration $a(t)$ solves

$$\frac{\mathrm{d}a}{\mathrm{d}t} = -k_1 a^3 + k_2 a^2 - k_3 a + k_4, \tag{5.5}$$

a cubic reaction-rate equation which, for suitable parameters, has two stable equilibria separated by an unstable one (see Figure 5.1). Deterministically, trajectories converge to one stable equilibrium depending on the initial condition.

In the stochastic model, the state $A(t) \in \mathbb{Z}_{\geq 0}$ is a birth–death process with nonlinear propensities. Typical trajectories rapidly approach the vicinity of one stable deterministic equilibrium and fluctuate there for a long time, as seen in the right plot of Figure 5.1. Eventually, a rare sequence of reaction events pushes the process across the unstable region, after which it relaxes near the other stable equilibrium. Figure 5.2 shows the stationary distribution with the characteristic bimodal shape for parameter values in the bisable region. The residence times between such switches can be extremely long and typically grow rapidly with system size.

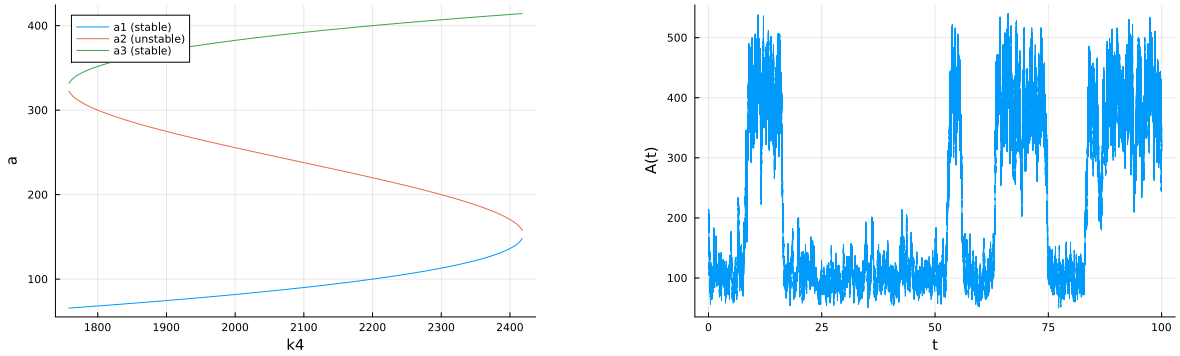

Figure 5.1: Left: Bifurcation diagram of the deterministic model (5.5) as a function of k_4. Right: Sample trajectory of the stochastic CRN (5.4). Parameters are $A(0) = 200$, $k_1 = 2.5 \times 10^{-4}$, $k_2 = 0.18$, $k_3 = 37.5$, and $k_4 = 2200$.

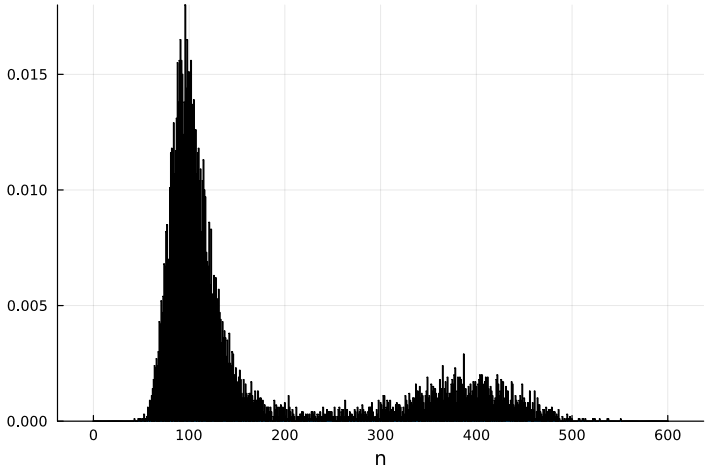

Figure 5.2: Stationary distribution of $X(t)$ for the same parameters as Figure 5.1. Obtained as a histogram from 10^4 samples, running the Gillespie SSA up to $T_f = 1000$ and storing the trajectory every step $\Delta t = 0.1$.

Conceptually, these long switching times are the same type of object as the mean time to extinction studied in Section 5.3. If D denotes a subset of state space corresponding to one metastable region, we define the exit time

$$\tau_D := \inf\{t \geq 0 : X(t) \notin D \mid X(0) \in D\}. \tag{5.6}$$

The associated mean exit time $u(x) = \mathbb{E}_x[\tau_D]$ satisfies the same backward equation as the mean extinction time,

$$-1 = (\mathcal{L}u)(x) \quad \text{for } x \in D, \qquad u(x) = 0 \quad \text{for } x \notin D. \tag{5.7}$$

In contrast to the definition of the mean extinction time in (5.2), we have deliberately been imprecise about the initial condition within D in (5.6). This imprecision is justified by the separation of time scales: the process relaxes rapidly toward the interior of the metastable region, whereas exits from D occur on much longer time scales. As a result, the mean exit time is largely insensitive to the precise starting point within D, provided it is not close to the

unstable state. Similarly, the precise choice of the boundary of D has little influence on the long switching times, as long as it is placed sufficiently beyond the unstable state: once the process reaches this region, it will typically either return to D or transition to the other basin of attraction with comparable probabilities.

Solving exit-time problems such as (5.7) directly in large nonlinear CRNs is typically infeasible, which motivates diffusion and large-deviation approximations developed later in the course.

Tasks

- Do Exercise 5.2 and compare the analytical predictions of M_s and V_s with simulations using the SSA in `Lect4.ipynb`.
- Consider (5.4) with $A(0) = 0$ and $k_1 = 2.5 \times 10^{-4}$, $k_2 = 0.18$ and $k_3 = 37.5$. Use the Gillespie SSA to generate long-time sample paths in the four cases $k_4 \in \{1750, 2100, 2200, 2450\}$ and estimate the stationary distribution.

Example codes

`Lect5.ipynb`: contains sample path simulations and mean first passage times for the birth–death process (5.1) and the metastable process (5.4).

Lecture 6: Reaction–diffusion systems as jump processes

6.1 Motivation: adding space

In Lectures 2–5 we modelled chemical reaction networks under a *well-mixed* assumption: only molecule numbers matter, not where molecules are located. In many applications this is not appropriate: molecules may be produced in a localised region, signals form gradients, or transport is slow compared to reaction times.

The simplest way to incorporate spatial effects while staying within the CTMC framework is:

- discretise space into well-mixed compartments,
- model *diffusion* as random jumps between neighbouring compartments,
- allow reactions *within* each compartment.

6.2 Compartment-based diffusion

We begin with diffusion only, for a single species A in a one-dimensional interval $[0, L]$. Fix an integer $K \geq 1$ and partition $[0, L]$ into K compartments of length

$$h := \frac{L}{K}.$$

Let $A_i(t)$ be the number of A molecules in compartment i at time t, and collect these into the state vector

$$\boldsymbol{A}(t) := (A_1(t), \ldots, A_K(t)) \in \mathbb{Z}_{\geq 0}^K.$$

Diffusion as jumps. Diffusion is modelled as jumps between neighbouring compartments, i.e., as the following chain of "chemical reactions":

$$\mathrm{A}_1 \underset{d}{\overset{d}{\rightleftharpoons}} \mathrm{A}_2 \underset{d}{\overset{d}{\rightleftharpoons}} \mathrm{A}_3 \underset{d}{\overset{d}{\rightleftharpoons}} \cdots \underset{d}{\overset{d}{\rightleftharpoons}} \mathrm{A}_K. \tag{6.1}$$

Each jump moves a single molecule, so the propensities are linear:

$$\alpha_{i \to i+1}(\boldsymbol{A}) = d\, A_i, \qquad \alpha_{i \to i-1}(\boldsymbol{A}) = d\, A_i, \tag{6.2}$$

and the corresponding reaction vectors are

$$\boldsymbol{\zeta}_{i \to i+1} = \boldsymbol{e}_{i+1} - \boldsymbol{e}_i, \qquad \boldsymbol{\zeta}_{i \to i-1} = \boldsymbol{e}_{i-1} - \boldsymbol{e}_i, \tag{6.3}$$

where $\boldsymbol{e}_i$ is the ith standard basis vector in $\mathbb{R}^K$. This defines a CTMC on $\mathbb{Z}_{\geq 0}^K$ and can be simulated by the Gillespie SSA 3.1 (Section 3.4).

Figure 6.1 illustrates two complementary viewpoints: (i) sample paths of a small number of tracked molecules (a microscopic viewpoint), and (ii) the distribution of molecule numbers across compartments at a fixed time (a mesoscopic viewpoint). Even for pure diffusion, individual trajectories are highly irregular, while the ensemble distribution is smooth and well-approximated by the diffusion PDE in the continuum limit discussed below.

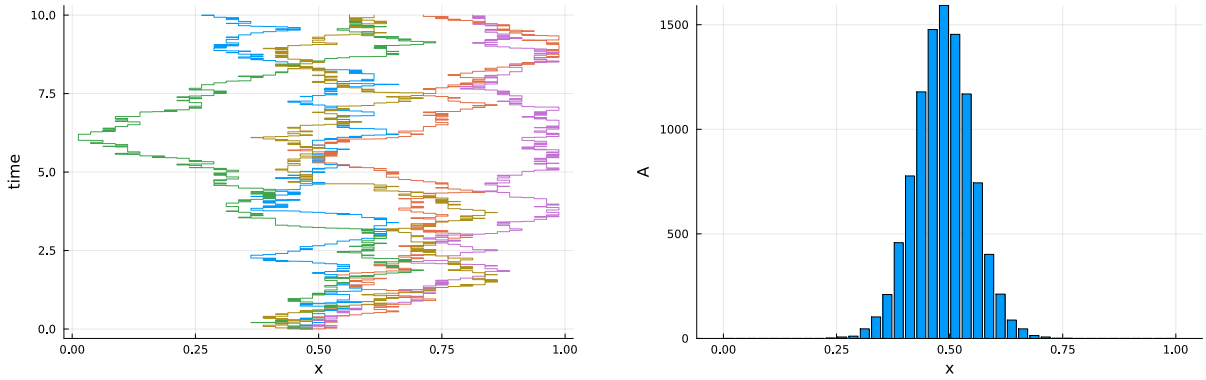

Figure 6.1: **Compartment-based diffusion in $[0, L]$.** *Left:* sample paths of 5 individual molecules. *Right:* histogram of molecule numbers $\boldsymbol{A}(t)$ at a fixed time $t = 0.2$. Parameters used: $D = 0.01$, $K = 40$, $\boldsymbol{A}(0) = 0.5$, $L = 1$.

Biased random walk. To model drift in addition to diffusion, we could choose different left/right jump rates:

$$\mathrm{A}_i \xrightarrow{d^+} \mathrm{A}_{i+1}, \qquad \mathrm{A}_i \xrightarrow{d^-} \mathrm{A}_{i-1},$$

so that molecules preferentially move in one direction when $d^+ \neq d^-$. In the continuum limit this produces an advection–diffusion equation (see Exercise 6.1).

The diffusion generator. Let $p(\boldsymbol{n}, t)$ denote the joint probability

$$p(\boldsymbol{n}, t) = \mathbb{P}(\boldsymbol{A}(t) = \boldsymbol{n}), \qquad \boldsymbol{n} = (n_1, \ldots, n_K) \in \mathbb{Z}^K_{\geq 0}.$$

The diffusion model defines a generator matrix $Q_{\text{diff}} = (q_{\boldsymbol{m}\boldsymbol{n}})_{\boldsymbol{m},\boldsymbol{n}\in S}$ whose rows and columns are indexed by all possible configurations of N molecules distributed among K compartments. Even for moderate values of N and K, this matrix is prohibitively large and never written down explicitly. For this reason, it is far more convenient to work directly with the generator in its operator form $\mathcal{L}_{\text{diff}}$. Substituting the diffusion propensities (6.2) and reaction vectors (6.3) into the general CRN generator formula (4.14), we obtain

$$\begin{aligned}(\mathcal{L}f)(\boldsymbol{n}) &= \sum_{\boldsymbol{m}\in S} q_{\boldsymbol{n}\boldsymbol{m}} f(\boldsymbol{m}) \\ &= \sum_{i=1}^{K-1} d\, n_i \big[f(\boldsymbol{n} + \boldsymbol{e}_{i+1} - \boldsymbol{e}_i) - f(\boldsymbol{n})\big] + \sum_{i=2}^{K} d\, n_i \big[f(\boldsymbol{n} + \boldsymbol{e}_{i-1} - \boldsymbol{e}_i) - f(\boldsymbol{n})\big],\end{aligned} \tag{6.4}$$

for any test function $f : \mathbb{Z}^K_{\geq 0} \to \mathbb{R}$. The first sum in the right-hand side corresponds to right jumps, while the second term tracks the left jumps.

The diffusion master equation. The forward equation (the *diffusion master equation*) is a special case of the chemical master equation (4.12):

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}p(\boldsymbol{n},t) = d\sum_{i=1}^{K-1}\Big[(n_i+1)p(\boldsymbol{n}+\boldsymbol{e}_i-\boldsymbol{e}_{i+1},t) - n_i p(\boldsymbol{n},t)\Big] \\ + d\sum_{i=2}^{K}\Big[(n_i+1)p(\boldsymbol{n}+\boldsymbol{e}_i-\boldsymbol{e}_{i-1},t) - n_i p(\boldsymbol{n},t)\Big].\end{aligned} \tag{6.5}$$

The (n_i+1) term premultiplying the terms with $p(\boldsymbol{n}-\boldsymbol{\zeta}_{i\to i\pm1},t)$ tells that if a reaction $i\to i\pm1$ occurred that led to state $\boldsymbol{n}=(n_1,\dots,n_i,\dots)$, then there must have been n_i+1 molecules in compartment i.

Mean dynamics and the diffusion PDE limit. Define the compartment means

$$M_i(t) := \mathbb{E}[A_i(t)] = \sum_{\boldsymbol{n}} n_i\, p_{\boldsymbol{n}}(t).$$

A standard calculation (as in Section 4.5, taking $f(\boldsymbol{n})=n_i$ in (6.4)) yields the discrete diffusion system

$$M_i'(t) = d\big(M_{i+1}-2M_i+M_{i-1}\big), \qquad i=2,\dots,K-1, \tag{6.6}$$

$$M_1'(t) = d\big(M_2-M_1\big), \tag{6.7}$$

$$M_K'(t) = d\big(M_{K-1}-M_K\big). \tag{6.8}$$

To connect with a deterministic diffusion descriptions, introduce a continuous concentration field $a(x,t)$ with $x\in[0,L]$ and identify M_i as its discretised approximation via

$$a(x_i,t)\approx a_i(t) := \frac{M_i(t)}{h}, \tag{6.9}$$

where x_i is the centre of compartment i. Then, for interior points,

$$a_i'(t) = d\,\frac{M_{i+1}-2M_i+M_{i-1}}{h} = d\,h^2\,\frac{a_{i+1}(t)-2a_i(t)+a_{i-1}(t)}{h^2}.$$

Choosing the jump rate to scale with h as

$$d = D/h^2, \tag{6.10}$$

the finite difference scheme converges, as $h\to0$ (with L fixed), to the diffusion equation with diffusion constant D.[1] Similarly, (6.7)–(6.8) converge to no-flux (reflecting) boundary conditions.[2] Thus the continuum limit is

$$\frac{\partial a}{\partial t} = D\,\frac{\partial^2 a}{\partial x^2} \quad \text{for } x\in(0,L), \qquad \left.\frac{\partial a}{\partial x}\right|_{x=0,L} = 0. \tag{6.11}$$

In summary, the diffusion-jump model (6.1) approximates one-dimensional diffusion with diffusion constant $D=dh^2$.

[1] To see this, Taylor expand $a(x_i\pm h,t) = a(x_i,t)\pm h\partial_x a(x_i,t)+\frac{h^2}{2}\partial_x^2 a(x_i,t)+O(h^3)$ and substitute into $a_{i\pm1}$. The even terms cancel out at we are left with $h^2\partial_x^2 a(x_i,t)+O(h^4)$.

[2] To recover the left boundary condition, write $a_1(t) = a(0,t)+\frac{h}{2}\partial_x a(0,t)+O(h^2)$ and $a_2(t) = a(0,t)+\frac{3h}{2}\partial_x a(0,t)+O(h^2)$. Inserting into (6.7), we find $\partial_t a(0,t)+O(h) = d[h\partial_x a(0,t)+O(h^2)]$, that is, $D\partial_x a(0,t)=0$ as $h\to0$.

Exercise 6.1 (Biased random walk and continuum limit)**.** Modify (6.1) to

$$A_i \xrightarrow{d^+} A_{i+1}, \qquad A_i \xrightarrow{d^-} A_{i-1},$$

with left and right jump rates $d^+ \neq d^-$. Determine how the rates $d^\pm$ should scale with h so that, in the limit $h \to 0$, the continous concentration field $a(x,t)$ satisfies the advection–diffusion equation

$$\frac{\partial a}{\partial t} = \frac{\partial}{\partial x}\left(D\frac{\partial a}{\partial x} - va\right)$$

with constant coefficients $D > 0$ and $v \in \mathbb{R}$.

6.3 A reaction–diffusion example

We now add reactions, still in K compartments. For clarity we keep a single reacting species A and consider local production and degradation as in (2.4):

$$\mathrm{A}_i \xrightarrow{k_1} \emptyset, \qquad \emptyset \xrightarrow{k_2} \mathrm{A}_i, \qquad i = 1, \dots, K, \tag{6.12}$$

together with diffusion jumps.

State and propensities. We keep the same state vector notation

$$\boldsymbol{A}(t) = (A_1(t), \dots, A_K(t)),$$

The propensities are

$$\alpha_i^{\mathrm{prod}}(\boldsymbol{n}) = k_2 h, \qquad \alpha_i^{\mathrm{deg}}(\boldsymbol{n}) = k_1 n_i, \qquad \alpha_{i\to i\pm1}^{\mathrm{jump}}(\boldsymbol{n}) = d n_i,$$

with reaction vectors $\pm\boldsymbol{e}_i$ for birth/death in compartment i, and (6.3) for diffusion. Note that here it is important to account explicitly for the compartment volume $\nu = h$, since a bigger compartment should make the production reaction occur more frequently. [We could also choose $\nu = h^3$ if we are thinking of three-dimensional compartments of side h. What matters is to be consistent in how we define the concentration field a, see (6.9).]

Generator form of the reaction-diffusion system. Let $\mathcal{L}$ be the generator of the resulting CRN on $\mathbb{Z}_{\geq 0}^K$. It is convenient to decompose it into two parts,

$$\mathcal{L} = \mathcal{L}_{\mathrm{diff}} + \mathcal{L}_{\mathrm{react}}, \tag{6.13}$$

where Q_{diff} contains the diffusion jumps (nearest-neighbour moves) as is given in (6.4) and Q_{react} contains the reactions. Concretely, acting on an observable $f : \mathbb{Z}_{\geq 0}^K \to \mathbb{R}$,

$$(\mathcal{L}_{\mathrm{react}} f)(\boldsymbol{n}) = \sum_{i=1}^{K} k_2\big[f(\boldsymbol{n}+\boldsymbol{e}_i) - f(\boldsymbol{n})\big] + \sum_{i=1}^{K} k_1 n_i\big[f(\boldsymbol{n}-\boldsymbol{e}_i) - f(\boldsymbol{n})\big]. \tag{6.14}$$

Mean dynamics and the reaction–diffusion PDE limit. The mean equation follows from the generator identity

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[f(\boldsymbol{A}(t))] = \mathbb{E}[(\{)(\boldsymbol{A}(t))],$$

applied to the coordinate functions $f(\boldsymbol{n}) = n_i$. Since both diffusion and the reactions (6.12) are linear, the mean closes exactly. Let $M_i(t) := \mathbb{E}[A_i(t)]$. For $i = 2, \ldots, K-1$ one finds

$$\frac{\mathrm{d}M_i}{\mathrm{d}t} = d\big(M_{i+1} - 2M_i + M_{i-1}\big) + k_2 h - k_1 M_i, \tag{6.15a}$$

and

$$\frac{\mathrm{d}M_1}{\mathrm{d}t} = d(M_2 - M_1) + k_2 h - k_1 M_1, \tag{6.15b}$$

$$\frac{\mathrm{d}M_K}{\mathrm{d}t} = d(M_{K-1} - M_K) + k_2 h - k_1 M_K, \tag{6.15c}$$

and the boundary compartments. Passing to concentrations $a(x_i, t) \approx M_i(t)/h$ and using the scaling $d = D/h^2$ gives, as $h \to 0$,

$$\frac{\partial a}{\partial t} = D\,\frac{\partial^2 a}{\partial x^2} + k_2 - k_1 a, \qquad \left.\frac{\partial a}{\partial x}\right|_{x=0,L} = 0, \tag{6.16}$$

closed by an initial condition $a(x, 0)$. Note how the chosen scaling of $\nu = h$ in the production rate leads to the right equation where we can take the limit $h \to 0$ without "losing" the production rate.

6.4 Limits and consistency checks

It is useful to keep in mind several limits, corresponding to different modelling regimes.

One compartment: recovering the well-mixed model. If $K = 1$ (so $h = L$), diffusion disappears and we recover a standard well-mixed CRN model from earlier lectures. In the present example (6.12), the state reduces to the single molecule count $A_1(t)$ and the CME reduces to the familiar birth–death master equation.

Fine mesh: continuum diffusion and reaction–diffusion. If $h \to 0$ with $d = D/h^2$, then the diffusion component of the dynamics converges, at the level of individual molecules, to Brownian motion with diffusion constant D (reflecting boundaries correspond to no-flux). At the level of the mean (and, for linear systems, more generally at the level of moments), the compartment model converges to the corresponding diffusion or reaction–diffusion PDE, as in (6.11) and (6.16).

Fast diffusion: effective well-mixed behaviour. If diffusion is much faster than reactions (heuristically, $d \gg k_1, k_2$ in the linear example), then the system quickly equilibrates spatially between reaction events. In that regime, the total molecule number

$$N(t) := \sum_{i=1}^{K} A_i(t)$$

behaves approximately like a well-mixed process with volume-averaged rates, and spatial variation is weak. This provides a practical check: increasing d at fixed K should make the reaction-diffusion model outputs closer to those of a well-mixed model for $N(t)$.

6.5 Limitations and choice of compartment size

Compartment-based models where molecules can only react if they are in the same compartment are attractive because they are simple and remain within the CRN framework. However, for higher-order reactions the approximation can become delicate: taking h too small can *reduce* the frequency of encounters, because reactions are only allowed between molecules in the same compartment. In the extreme limit $h \to 0$ (with diffusion implemented as nearest-neighbour jumps), two molecules may spend very little time in the same compartment, so bimolecular reactions can be artificially suppressed.

A simple bimolecular test case. To expose this issue, consider a bimolecular reaction in each compartment, for example

$$\mathrm{A}_i + \mathrm{B}_i \xrightarrow{k} \emptyset, \qquad i = 1, \ldots, K,$$

together with diffusion of A and B. In a standard implementation as discussed in previous lectures, the bimolecular propensity in compartment i scales like[3]

$$\alpha_i^{\mathrm{bi}}(\boldsymbol{A}, \boldsymbol{B}) \propto \frac{k}{h} A_i B_i,$$

reflecting that the same macroscopic rate constant corresponds to a *larger* per-compartment propensity as the compartment volume shrinks. Even with such scaling, the *encounter* statistics can still be distorted if h is taken too small compared to the typical reaction length-scale.

For bimolecular reactions, it is useful to distinguish two competing time scales.

- Mean time to meet (τ_{meet}): defined as the typical time it takes two diffusing molecules to occupy the same compartment. If moleculs A and B have diffusion coefficients D_A and D_B respectively, the we have that the mean relative displacement in some time τ scales like $\sqrt{(D_A + D_B)\tau}$ and hence that $\tau_{\mathrm{meet}} \sim h^2/(D_A + D_B)$, independently on the dimension.
- Mean time to react (τ_{react}): defined as the typical waiting time for the chemical reaction to occur once the molecules are co-located. This is given by the inverse propensity of the reaction, that is, $\tau_{\mathrm{react}} \sim h^d/k$, where d is the dimension.

For the reaction-diffusion CRN framework, to give a faithful approximation of the underlying reaction-diffusion process, these time scales should be comparable, in the sense that molecules should have a reasonable chance to react before diffusing away:

$$\frac{h^d}{k} = \tau_{\mathrm{react}} \sim \tau_{\mathrm{meet}} = \frac{h^2}{D_A + D_B} \qquad \to \qquad h^{d-2} \sim \frac{k}{D_A + D_B}. \tag{6.17}$$

We see that the interpretation of the matching condition depends crucially on the spatial dimension d:

[3] This scaling corresponds to one dimensons. In higher dimensions, the power of h changes accordingly.

- $d = 3$: taking h too small makes $\tau_{\text{react}} \gg \tau_{\text{meet}}$. In this regime, molecules typically diffuse out of a compartment many times before reacting, and bimolecular reactions are artificially suppressed. This is the standard breakdown mechanism of the RDME for bimolecular reactions on overly fine spatial meshes.
- $d = 2$: the matching condition (6.17) becomes independent of h. This marginal case reflects the fact that diffusion is recurrent in two dimensions. While reactions are not suppressed in the same way as in $d = 3$, the validity of the compartment-based approximation becomes delicate and depends on logarithmic corrections (since the mean time to meet diverges logarithmically).
- $d = 1$: the situation is reversed, since decreasing h makes $\tau_{\text{react}} \ll \tau_{\text{meet}}$. Once two molecules enter the same compartment they are very likely to react before separating. In this sense, reactions are not suppressed by refining the mesh; instead, the RDME tends to overestimate reaction efficiency in one dimension.

Tasks and example code

Tasks

- Implement the SSA for pure diffusion (6.1) on $[0, L]$ with K compartments and reflecting boundaries. Plot histograms of $X_i(t)$ at several times.
- Verify numerically that the empirical mean profile $\mathbb{E}[A_i(t)]/h$ approaches the solution of the diffusion PDE (6.11) when h is decreased and $d = D/h^2$.
- Modify the diffusion model to a biased random walk ($d^+ \neq d^-$). Identify the corresponding continuum limit (advection–diffusion) by comparing to a finite-difference PDE solver.
- Add production/degradation (6.12) and compare the mean profile with the PDE (6.16). Add a bimolecular reaction and explore the dependence on h; illustrate the "too small h" pathology.

Example code

`Lect6.ipynb`: contains the Gillespie SSA for the compartment-based diffusion in Section 6.2, both molecule- and comparment-centered. *It will* contain an example of a reaction-diffusion system.

Lecture 7: Tau-leaping and Poisson representations

In previous lectures we introduced stochastic chemical reaction networks (CRNs) as continuous-time Markov jump processes, and studied their sample-path description via the Gillespie algorithm (SSA 3.1). In this lecture we present an alternative but equivalent viewpoint based on *reaction counting processes* and *Poisson clocks*.

This representation makes explicit where randomness enters the dynamics, clarifies the structure underlying the Gillespie algorithm, and provides a natural route to approximate descriptions such as tau-leaping. It will also serve as the starting point for diffusion approximations discussed later.

7.1 Reaction counting processes

For a CRN with reaction channels $j = 1, \dots, m$, let $N_j(t)$ denote the number of times reaction j has fired up to time t. Each $N_j(t)$ is a counting process: it takes values in $\mathbb{Z}_{\geq 0}$ and increases only by jumps of size one.

Writing ζ_j for the reaction vectors, the stochastic evolution of the system can be written pathwise as

$$\boldsymbol{X}(t) = \boldsymbol{X}(0) + \sum_{j=1}^{m} \boldsymbol{\zeta}_j N_j(t). \tag{7.1}$$

Equation (7.1) is exact. It mirrors the deterministic integral formulation (3.4), with reaction counts replacing time integrals of rates.

7.2 Poisson time-change representation

If reaction intensities were constant, each $N_j(t)$ would be a Poisson process. In a CRN, however, the intensities depend on the current state.

Let $Y(t)$ denote a unit-rate Poisson process. An inhomogeneous Poisson process with rate $\lambda(t)$ can be written as

$$Y\left(\int_0^t \lambda(s)\,\mathrm{d}s\right).$$

This representation can be interpreted as a unit-rate Poisson process run on a clock whose speed at time s is $\lambda(s)$. In particular, if $\lambda(t) \equiv \lambda$, the clock advances at constant speed and $Y(\lambda t)$ is a Poisson process with rate λ.

Applying this to CRNs yields the *Poisson time-change representation*

$$N_j(t) = Y_j\left(\int_0^t \alpha_j(\boldsymbol{X}(s))\,\mathrm{d}s\right). \tag{7.2}$$

Substituting (7.2) into (7.1) gives

$$\boldsymbol{X}(t) = \boldsymbol{X}(0) + \sum_{j=1}^{m} \boldsymbol{\zeta}_j Y_j\left(\int_0^t \alpha_j(\boldsymbol{X}(s))\,\mathrm{d}s\right), \tag{7.3}$$

where $Y_1, \dots, Y_m$ are independent unit-rate Poisson processes.

Although the integrals $\int_0^t \alpha_j(X(s))\, ds$ are random (as they depend on the random process $X(s)$), they introduce no additional independent sources of noise; all randomness is driven by the independent Poisson processes Y_j, evaluated at random, state-dependent times.

A useful interpretation is that each reaction channel is equipped with an independent Poisson clock. All clocks tick at unit rate, but their internal time runs faster or slower depending on the current state. Whenever one clock rings, the state changes and all clock speeds are updated.

7.3 Tau-leaping as a mesoscopic approximation

While (7.3) is exact, it also suggests approximate simulation methods. Suppose the system is in state $\boldsymbol{X}(t)$ and consider a short time interval $[t, t+\tau)$. If τ is chosen so that none of the propensities $\alpha_j(\boldsymbol{X}(t))$ change significantly over this interval, then each reaction channel may be approximated as a Poisson process with constant rate $\alpha_j(\boldsymbol{X}(t))$.

Under this assumption, the number of firings of reaction j in $[t, t+\tau)$ is approximately Poisson distributed with parameter $\alpha_j(\boldsymbol{X}(t))\,\tau$. Let P_j be independent Poisson random variables with these parameters. The tau-leaping update is then

$$\boldsymbol{X}(t+\tau) = \boldsymbol{X}(t) + \sum_{j=1}^{m} P_j \boldsymbol{\zeta}_j.$$

These observations lead to an approximate stochastic simulation algorithm, known as the *tau-leaping method*.

Algorithm 7.1: Tau-leaping method

Set $\boldsymbol{X}(0) = \boldsymbol{X}_0$, $t = 0$ and final time t_f. Fix a time-step $\tau > 0$.
While $t < t_f$:

(a) Compute the propensities $\alpha_j(\boldsymbol{X}(t))$ for all reactions.

(b) Sample independent Poisson random variables

$$P_j \sim \text{Poisson}(\alpha_j(\boldsymbol{X}(t))\,\tau), \qquad j = 1, \dots, m.$$

(c) Update $\boldsymbol{X}(t+\tau) = \boldsymbol{X}(t) + \sum_{j=1}^{m} P_j\, \boldsymbol{\zeta}_j$.

(d) Set $t \leftarrow t + \tau$.

end

Remark 7.1. The tau-leaping method is *not exact*. It relies on the assumption that the propensities remain approximately constant over $[t, t+\tau)$. If τ is chosen too large, this assumption may fail and lead to poor approximations. Moreover, when molecule numbers are small, relative fluctuations are large and the Poisson increments in a single leap may exceed the available reactants, resulting in unphysical states such as negative molecule numbers; in contrast, the exact SSA enforces these state constraints event by event.

7.4 Comparison between tau-leaping and the 'naive' algorithms

At first sight, tau-leaping may appear similar to the 'naive' Algorithm 1.2 introduced in Lecture 1, since both methods advance the system using a fixed time step. However, the two algorithms are conceptually quite different.

Recall the naive algorithm for the degradation reaction

$$\mathrm{A} \xrightarrow{k} \emptyset, \tag{7.4}$$

in which, over each small time interval Δt, a single reaction is allowed to occur with probability $kA(t)\Delta t$. This algorithm is based on the short-time expansion

$$\mathbb{P}(\text{one reaction in } [t, t+\Delta t)) = kA(t)\Delta t + o(\Delta t),$$

and therefore neglects the possibility of two or more reactions occurring in the same interval.

From a probabilistic viewpoint, the naive algorithm replaces the Poisson distribution of reaction counts in $[t, t+\Delta t)$ by a Bernoulli random variable. This approximation is accurate only when $kA(t)\Delta t \ll 1$.

By contrast, tau-leaping preserves the Poisson statistics of reaction counts over a finite time interval. It approximates the evolution by freezing the propensities, not by suppressing multiple reaction events. As a result, tau-leaping provides a natural bridge to diffusion approximations, while the naive algorithm does not admit a comparable systematic scaling limit.

The results of using the tau-leaping method 7.1 to generate sample paths from the degradation reaction (7.4) are shown in Figure 7.1. Four sample paths generated using $\tau = 0.1$ are shown on the left-hand side. On the right-hand side the predicted mean molecule number is plotted for $\tau = 1.0$ and $\tau = 5.0$, alongside the analytical solution (1.17). We see that the accuracy of the method decreases as τ increases.

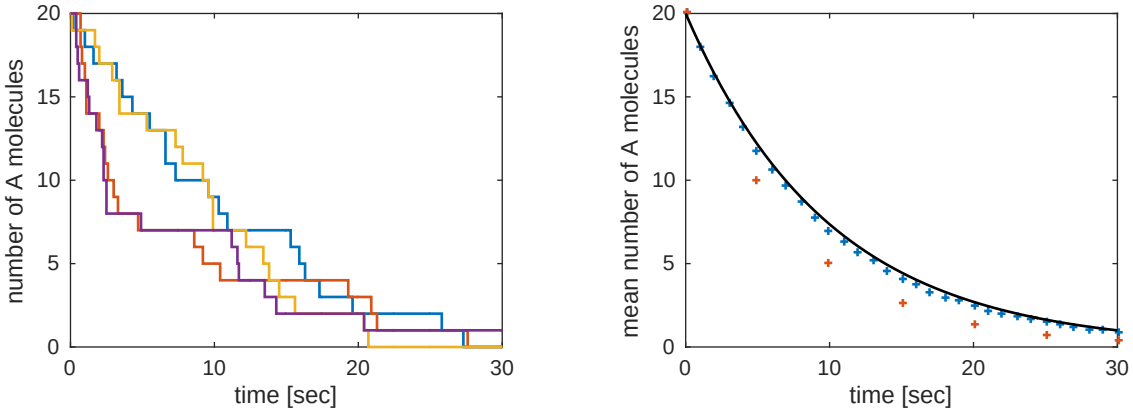

Figure 7.1: Sample paths from a degradation reaction system (7.4) generated using the tau-leaping algorithm 7.1. Left: four different sample paths each generated using $\tau = 0.1$. Right: mean number of A molecules predicted using $\tau = 1$ (blue) and $\tau = 5$ (orange). In each case 10^4 sample paths were used to estimate the mean. The exact solution (1.17) for the mean number of A molecules is plotted in black. Parameters are $A(0) = 20$, $k = 0.1\,\mathrm{s}^{-1}$.

We can also perform a qualitative comparison between the numerical methods by comparing the predicted mean number of molecules at a fixed final time T with the exact value

$$M(T) = \mathbb{E}[A(T)] = Ne^{-kT},$$

which is known explicitly for the linear degradation reaction (7.4).

We denote by $M^{\Delta t}_{\text{tau}}(T, K)$ and $M^{\Delta t}_{\text{naive}}(T, K)$ the empirical means obtained using the tau-leaping and naive algorithms, respectively, with time step Δt and K independent sample paths. Similarly, $M_{\text{G}}(T, K)$ denotes the empirical mean obtained using the Gillespie stochastic simulation algorithm (SSA) with K samples.

Figure 7.2 illustrates the qualitative differences between the empirical means produced by the naive and tau-leaping methods. The naive algorithm systematically *underestimates* the number of reactions, while the tau-leaping method *overestimates* them. This behaviour can be understood directly from the structure of the algorithms. The naive algorithm allows at most one reaction per time step, even when the step size is large, thereby suppressing reaction events. In contrast, the tau-leaping method assumes that the propensity $\alpha(n) = kn$ remains constant over each interval of length Δt. Since the reaction vector is negative, the true propensity satisfies $\alpha(A(t+\tau)) \leq \alpha(A(t))$ for all $\tau > 0$, and freezing the propensity therefore leads to an overestimation of the number of reactions.

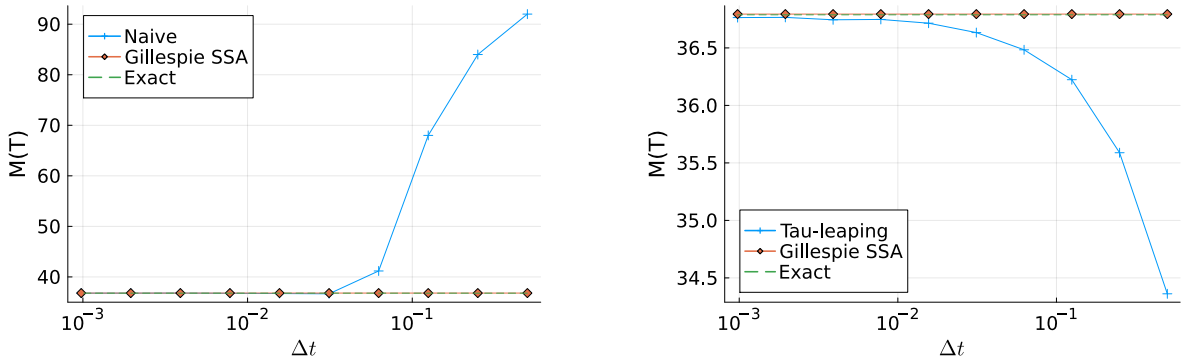

Figure 7.2: Empirical mean number of molecules as a function of time for the degradation reaction (7.4), computed using the naive and tau-leaping algorithms. Parameters used are $k = 1$, $N = 100$, final time $T_f = 1.0$, and $K = 10^5$ sample paths.

To quantify these observations, we define the errors

$$e^{\Delta t}_{\text{tau}} = \left|M^{\Delta t}_{\text{tau}}(T, K) - M(T)\right|, \qquad e^{\Delta t}_{\text{naive}} = \left|M^{\Delta t}_{\text{naive}}(T, K) - M(T)\right|, \qquad e_{\text{G}} = \left|M_{\text{G}}(T, K) - M(T)\right|.$$

These quantities are examples of *weak errors* of stochastic numerical methods (we will discuss this notion more formally later). The decay of the weak errors $e^{\Delta t}_{\text{tau}}$ and $e^{\Delta t}_{\text{naive}}$ as Δt decreases, together with the sampling error of the Gillespie SSA e_{G}, is shown in Figure 7.3.

For the Gillespie SSA there is no time-discretisation error; the error e_{G} is purely statistical. By the central limit theorem, $e_{\text{G}} = O(K^{-1/2})$, with a prefactor determined by $\text{Var}(A(T))$. This agrees with the value shown in Figure 7.3.

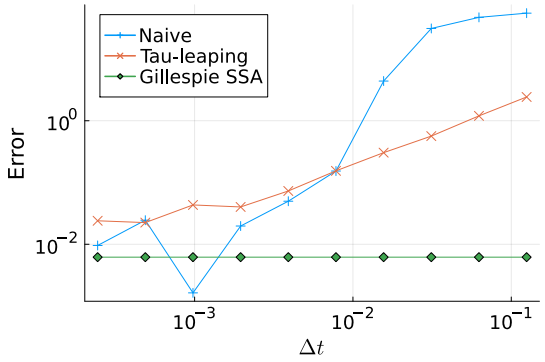

Figure 7.3: Weak errors for the degradation reaction (7.4) using the naive and tau-leaping algorithms, together with the sampling error of the Gillespie SSA. Parameters used are $k = 1$, $N = 100$, final time $T_f = 1.0$, and $K = 10^5$ sample paths.

Tasks

- Implement tau-leaping for the degradation reaction $\mathrm{A} \xrightarrow{k} \emptyset$ and compare sample paths and empirical means with the Gillespie SSA for several values of τ.
- Compare the naive SSA and tau-leaping for the same reaction at fixed final time T_f, and record the computational cost required to achieve a given accuracy in the mean.
- Implement tau-leaping for the bistable CRN

$$3\mathrm{A} \xrightarrow{k_1} 2\mathrm{A}, \qquad 2\mathrm{A} \xrightarrow{k_2} 3\mathrm{A}, \qquad \mathrm{A} \xrightarrow{k_3} \emptyset, \qquad \emptyset \xrightarrow{k_4} \mathrm{A}, \tag{7.5}$$

  introduced previously. Compare the long-time behaviour with Gillespie simulations. Identify regimes in which tau-leaping fails.

Example code

`Lect7.ipynb` contains the implementation of the tau-leaping method for the degradation reaction (7.4) and the bistable CRN (7.5). It also shows how to compare methods by measuring the weak error of an observable.

Lecture 8: From CME to diffusion

In the previous lecture we showed that stochastic chemical reaction networks (CRNs) can be represented exactly in terms of independent Poisson processes via the Poisson time-change representation. In this lecture we investigate what happens to this description when the system size becomes large.

Our goal is to understand how deterministic reaction-rate equations and Gaussian fluctuations emerge from the underlying jump process. The discussion is deliberately formal and heuristic. Rigorous justification requires the theory of Brownian motion and stochastic differential equations, which we will introduce later at an applied level.

8.1 System-size scaling

Throughout this lecture, we keep the one-species nonlinear CRN introduced earlier,

$$3\mathrm{A} \xrightarrow{k_1} 2\mathrm{A}, \qquad 2\mathrm{A} \xrightarrow{k_2} 3\mathrm{A}, \qquad \mathrm{A} \xrightarrow{k_3} \emptyset, \qquad \emptyset \xrightarrow{k_4} \mathrm{A}, \tag{8.1}$$

with stochastic mass-action propensities

$$\alpha_1(A) = \frac{k_1}{\nu^2} A(A-1)(A-2), \quad \alpha_2(A) = \frac{k_2}{\nu} A(A-1), \quad \alpha_3(A) = k_3 A, \quad \alpha_4(A) = k_4 \nu,$$

where ν is the reactor volume, and reaction vectors

$$\zeta_1 = -1, \quad \zeta_2 = +1, \quad \zeta_3 = -1, \quad \zeta_4 = +1.$$

It is reasonable to suppose that molecule numbers scale proportionally with the volume ν. Motivated by this, we write the stochastic state as

$$A(t) = \nu a(t) + \sqrt{\nu}\,\eta(t), \tag{8.2}$$

where $a(t)$ represents the macroscopic (deterministic) concentration and $\eta(t)$ represents fluctuations around the macroscopic state. The factor $\sqrt{\nu}$ reflects central-limit scaling: relative fluctuations are typically of order $\nu^{-1/2}$ when molecule numbers are large.

8.2 Law of large numbers: deterministic limit

Recall the Poisson time-change representation

$$A(t) = A(0) + \sum_{j=1}^{4} \zeta_j Y_j \left(\int_0^t \alpha_j(A(s))\,\mathrm{d}s \right),$$

where Y_j are independent unit-rate Poisson processes. A Poisson random variable with parameter λ has mean and variance equal to λ.

When ν is large, the propensities $\alpha_j(A)$ are large and the Poisson processes concentrate around their means.[1] Formally replacing each Poisson process by its expectation yields

$$A(t) \approx A(0) + \sum_{j=1}^{4} \zeta_j \int_0^t \alpha_j(A(s))\,\mathrm{d}s.$$

Dividing by ν and substituting $A(s) \approx \nu a(s)$ gives

$$a(t) \approx a(0) + \int_0^t \left(-k_1 a^3(s) + k_2 a^2(s) - k_3 a(s) + k_4\right)\,\mathrm{d}s.$$

Differentiating in time, we recover the deterministic reaction-rate equation

$$\frac{\mathrm{d}a}{\mathrm{d}t} = -k_1 a^3 + k_2 a^2 - k_3 a + k_4. \tag{8.3}$$

This argument illustrates a general principle: deterministic reaction-rate equations arise as law-of-large-numbers limits of stochastic CRNs.

8.3 Central-limit scaling of fluctuations

The law of large numbers captures the leading-order behaviour of the system but ignores fluctuations. To understand fluctuations, we examine deviations from the Poisson mean.

For large λ, a Poisson random variable satisfies

$$\mathrm{Poisson}(\lambda) = \lambda + \sqrt{\lambda}\,\xi, \qquad \xi \approx \mathcal{N}(0,1),$$

a direct manifestation of the central limit theorem.

Applying this approximation to the tau-leaping description over a time interval $[t, t+\tau)$ gives

$$A(t+\tau) - A(t) \approx \sum_{j=1}^{4} \zeta_j \alpha_j(A(t))\,\tau + \sum_{j=1}^{4} \zeta_j \sqrt{\alpha_j(A(t))\,\tau}\,\xi_j, \tag{8.4}$$

where ξ_j are independent standard normal random variables.

The first term corresponds to deterministic drift, while the second term represents stochastic fluctuations of order $\sqrt{\nu}$, consistent with the scaling ansatz (8.2). Substituting $A(t) = \nu a(t) + \sqrt{\nu}\eta(t)$ and collecting terms shows that the leading-order terms reproduce the deterministic equation (8.3), while the next-order terms describe Gaussian fluctuations around the deterministic trajectory.

Origin of Gaussian noise. The appearance of Gaussian noise in (8.4) is not an additional modelling assumption. It arises directly from the Poisson statistics of reaction counts when many reaction events occur over a short time interval. From this viewpoint, Poisson noise is the fundamental source of randomness in CRNs, Gaussian noise appears as a large-system approximation, and diffusion models are mesoscopic descriptions of jump processes.

[1] Although the variance of a Poisson random variable grows with its mean, the relative fluctuations are of order $1/\sqrt{\lambda}$ and vanish as $\lambda \to \infty$.

8.4 Preview of diffusion descriptions

Letting $\tau \to 0$ in (8.4) suggests a continuous-time description in which Poisson noise is replaced by Gaussian noise. Formally, this leads to diffusion models such as the *Chemical Langevin equation*

$$\mathrm{d}A(t) = \sum_{j=1}^{4} \zeta_j \alpha_j(A(t))\,\mathrm{d}t + \sum_{j=1}^{4} \zeta_j \sqrt{\alpha_j(A(t))}\,\mathrm{d}W_j(t), \tag{8.5}$$

where W_j are independent one-dimensional Brownian motions, and the associated *Chemical Fokker–Planck equation*

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}\left(p \sum_{j=1}^{4} \zeta_j \alpha_j(x)\right) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left(p \sum_{j=1}^{4} \zeta_j^2 \alpha_j(x)\right). \tag{8.6}$$

Both equations are now defined on a continuous state space $x \in \mathbb{R}$, in contrast to the original discrete state space $\mathbb{Z}_{\geq 0}$.

At this stage, we emphasize that these are *formal limits*. To interpret them correctly, we will first introduce Brownian motion and stochastic differential equations at an applied level, before returning to diffusion approximations of CRNs.

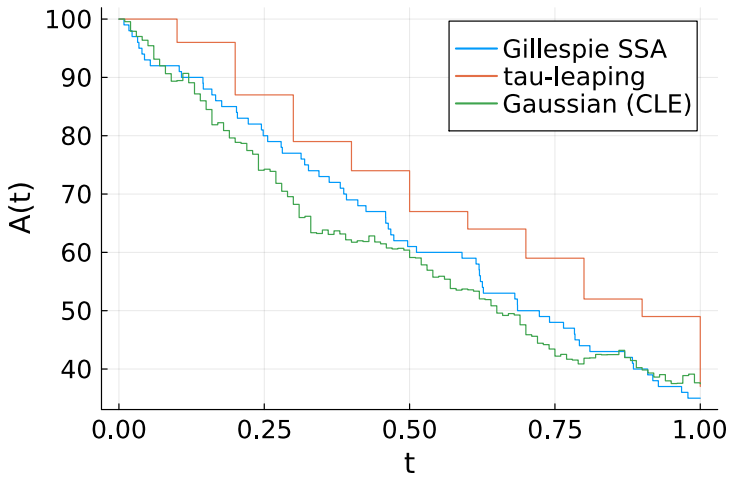

Figure 8.1: Comparison of gillespie, tau leaping and the Gaussian approximation for the simple degradation (7.4). Parameters used are $k = 1$, $A(0) = 100$, final time $T_f = 1.0$, $\tau = 0.1$ for the tau leaping and $\Delta t = 0.01$ for the Euler-maruyama.

Tasks

- For the degradation reaction $\mathrm{A} \xrightarrow{k} \emptyset$, compare the exact mean and variance obtained from the CME with those predicted by the Gaussian approximation for large initial copy numbers.
- Implement the Gaussian approximation (8.4) for the CRN (8.1) and compare sample paths with Gillespie simulations.

- Investigate numerically how the magnitude of fluctuations scales with ν by running simulations for increasing system size.

Example codes

`Lect8.ipynb`: simulations illustrating the law-of-large-numbers and central-limit scaling for simple CRNs, and comparisons between Gillespie SSA, tau-leaping, and Gaussian approximations.

Lecture 9: Brownian motion

9.1 Overview

In previous lectures we studied stochastic chemical reaction networks as jump processes and showed how diffusion-type behaviour emerges from many small random reaction events. In this lecture we focus on the simplest and most fundamental diffusion process: *Brownian motion*.

Brownian motion plays a dual role in this course. First, it arises as the scaling limit of random jump processes, when many small, independent displacements are observed on large space and time scales. Second, once this limit has been established, Brownian motion is used as the driving noise in stochastic differential equations.

9.2 Brownian motion as a limit of random walks

Consider a simple random walk on the set $S = \mathbb{Z}$. Let $\xi_1, \xi_2, \dots$ be a collection of independent, identically distributed (i.i.d.) random variables with mean 0 and variance 1. For simplicity, suppose that $\xi_j = \pm 1$ with equal probability. Then consider the following discrete-time and state space Markov process or random walk $\{X_n\}_{n\in\mathbb{N}}$

$$X_n = \sum_{j=1}^{n} \xi_j, \qquad X_0 = 0. \tag{9.1}$$

Consider the properties of $\{X_n\}_{n\in\mathbb{N}}$:

(i) $\mathbb{E}X_n = 0$.

(ii) $\mathrm{Var}(X_n) = \mathbb{E}X_n^2 - (\mathbb{E}X_n)^2 = n$.

(iii) $\{X_n\}_{n\in\mathbb{N}}$ has independent increments. To see why, let $0 < n_1 < n_2 \leq n_3 < n_4$ and write

$$X_{n_2} - X_{n_1} = \xi_{n_1+1} + \cdots \xi_{n_2}, \qquad X_{n_4} - X_{n_3} = \xi_{n_3+1} + \cdots \xi_{n_4}.$$

Each increment is a sum of distinct, independent random variables, so they are independent.

(iv) $\{X_n\}_{n\in\mathbb{N}}$ has stationary increments (invariant to time shifts).
To see why, note that, for $m > n$

$$X_m - X_n = \xi_{n+1} + \cdots + \xi_m, \qquad X_{m-n} = \xi_1 + \cdots + \xi_{m-n}.$$

Each of these terms is a sum of $m - n$ i.i.d. random variables, so $X_m - X_n \sim X_{m-n}$.

(v) The Central Limit Theorem asserts that, for large N,

$$\frac{X_N}{\sqrt{N}} \sim \mathcal{N}(0, 1),$$

that is, a Gaussian variable with mean 0 and variance 1.

We can simulate the random walk (see Figure 9.1). We notice that, if we take a large number of steps, the random walk starts looking like a continuous time process with continuous paths.

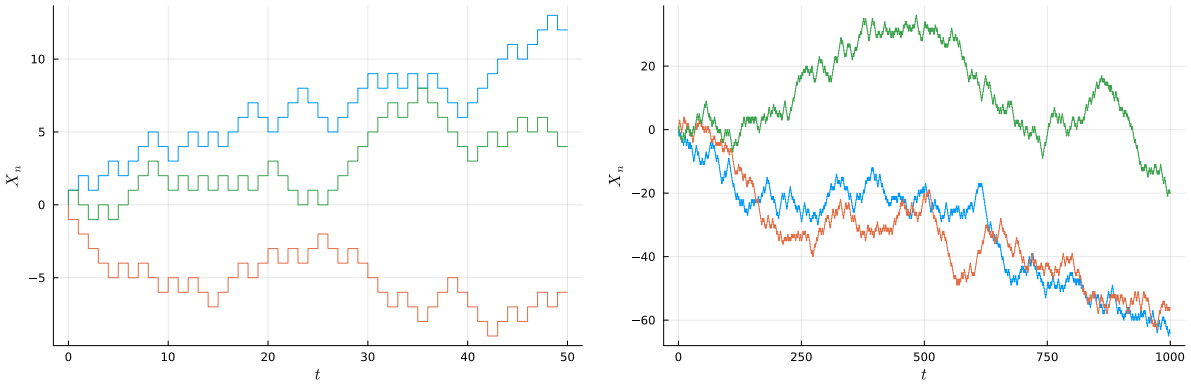

Figure 9.1: Three paths of the random walk (9.1) X_n of length $n = 50$ (left) and $n = 1000$ (right).

Given the properties of $\{X_n\}_{n\in\mathbb{N}}$, we might wonder if there is a way to scale it so it approaches a Brownian motion in some limit. To this end, let's scale space with Δx and time with Δt, $t = \Delta t n$, such that the rescaled process is

$$\widetilde{X}(t) = \Delta x X_{t/\Delta t} = \Delta x \left(\xi_1 + \cdots + \xi_{t/\Delta t}\right) \tag{9.2}$$

We would like to consider the process $\{\widetilde{X}(t)\}_{t\in\mathbb{N}/\Delta t}$ as $\Delta x, \Delta t \to 0$. For the limit to make sense, the variance of the limiting process should be finite too:

$$\mathrm{Var}(\widetilde{X}(t)) = \Delta x^2 \mathrm{Var}(X_{t/\Delta t}) = \Delta x^2 \frac{t}{\Delta t}.$$

This tells us that we should pick $\Delta x^2 \sim \Delta t$ as $\Delta t, \Delta x \to 0$.

Let's write $\Delta t = 1/N$ and $\Delta x = 1/\sqrt{N}$ and define a sequence of continuous-time processes $W^N(t)$ in terms of the parameter $N \in \mathbb{N}$:

$$W^N(t) = \frac{X_{\lfloor Nt \rfloor}}{\sqrt{N}}, \tag{9.3}$$

where $\lfloor x \rfloor$ denotes the largest integer less than or equal to x. Then *Donsker's Theorem* or *Donsker's Invariance Principle* says that as $N \to \infty$, $W^N(t)$ converges in distribution to a stochastic process $W(t)$, termed the Wiener process or Brownian motion. (This is like the Central Limit Theorem extended to functions.)

Diffusive scaling

The time and space scaling $t \sim x^2$ is known as the *diffusion or parabolic scaling*. This is an important point: for a diffusion process, space scales as the square root of time. (Note also that this scaling could be directly inferred from point (v) above.)

9.2.1 Probabilistic description

Einstein studied Brownian motion by describing the fluctuations in the particle's position probabilistically. Here we consider a variant of Einstein's argument starting from the discrete-time, discrete-space stochastic process (9.2).

Consider the rescaled process $\widetilde{X}(t)$ on the state space $\{i\Delta x\}_{i\in\mathbb{Z}}$ at times $\{n\Delta t\}_{n\in\mathbb{N}}$ starting at the origin at time $t = 0$. Define

$$P(i, n) = \mathbb{P}\{\widetilde{X}(n\Delta t) = i\Delta x\}.$$

Then initially $P(i, 0) = \delta_{i0}$ and P satisfies the recurrence relation

$$P(i, n + 1) = \frac{1}{2}P(i - 1, n) + \frac{1}{2}P(i + 1, n). \tag{9.4}$$

Assume the diffusion scaling

$$\frac{(\Delta x)^2}{\Delta t} = 2D, \tag{9.5}$$

for some constant $D > 0$. Then

$$\frac{P(i, n + 1) - P(i, n)}{\Delta t} = D\frac{P(i - 1, n) - 2P(i, n) + P(i + 1, n)}{(\Delta x)^2}.$$

Letting $\Delta t, \Delta x \to 0$ with $i\Delta x \to x$, $n\Delta t \to t$, we obtain in the limit

$$\frac{\partial p}{\partial t} = D\frac{\partial^2 p}{\partial x^2}, \tag{9.6}$$

where the limiting $p(x, t)$ is a probability density for $t, x \in \mathbb{R}$. Thus we see that the probability density of the limiting process satisfies a diffusion PDE, which is another instance of a Fokker–Planck equation as seen in Lecture 1, see (1.6).

Note that choosing $\Delta x = \sqrt{\Delta t}$ corresponds to $D = 1/2$, which explains why the diffusion coefficient $1/2$ appears in many textbook definitions of Brownian motion.[1] The solution of (9.6) with $p(x, 0) = \delta_0(x)$ is the Gaussian heat kernel (9.8).

Remark 9.1 (Paths versus densities: Donsker and Einstein)**.** Donsker's theorem describes convergence at the level of sample paths, while Einstein's argument describes convergence at the level of probability densities (forward equation limit). Both lead to Brownian motion, but emphasise different objects.

9.3 Properties of Brownian motion

Definition 9.1 (Brownian motion)**.** A stochastic process $W = \{W(t) : t \geq 0\}$, starting from $W(0) = 0$, is called a standard Brownian motion (or Wiener process) if:

(i) W has independent increments: for $0 \leq t_0 < t_1 < t_2 < \infty$, the random variables $W(t_1) - W(t_0)$ and $W(t_2) - W(t_1)$ are independent.

(ii) W has Gaussian increments: $W(t_2) - W(t_1) \sim \mathcal{N}(0, t_2 - t_1)$ for all $t_2 \geq t_1 \geq 0$.

(iii) W has continuous sample paths.

[1] An alternative common convention in probability textbooks is to define the diffusion coefficient as $D = \Delta^2/\Delta t$, and thus $\partial_t p = \frac{1}{2}D\partial_x^2 p$ as a diffusion equation with diffusion coefficient D.

Let us examine these conditions. Point (ii) implies stationary increments since $W(t_2)-W(t_2)$ depends only on the difference $t_2 - t_1$ alone. Gaussianity (ii) implies that

$$\mathbb{P}(W(t) \in [x_1, x_2)) = \int_{x_1}^{x_2} p(x,t)\mathrm{d}x, \tag{9.7}$$

with

$$p(x,t) = \frac{e^{-x^2/(2t)}}{\sqrt{2\pi t}}. \tag{9.8}$$

The autocovariance of $W(t)$, for $s < t$ is given by

$$\begin{aligned} C(t,s) &= \mathbb{E}\left[(W(t) - \mathbb{E}W(t))(W(s) - \mathbb{E}W(s))\right] = \mathbb{E}[W(t)W(s)] \\ &= \mathbb{E}[(W(t) - W(s) + W(s))W(s)] = \mathbb{E}[(W(t) - W(s))W(s)] + \mathbb{E}[W(s)^2] = s. \end{aligned}$$

In the second line, we have used independence in the second equality and the variance formula (ii) in the last step. Thus,

$$C(t,s) = \min(t,s). \tag{9.9}$$

It is also useful to remember that

$$\mathbb{E}(W(t) - W(s))^2 = |t - s|. \tag{9.10}$$

To see this, $\mathbb{E}(W(t) - W(s))^2 = C(t,t) + C(s,s) - 2C(t,s)$ and use (9.9). Note how this is consistent with the diffusive scaling discussed above.

Brownian motion is the unique continuous-time continous process with stationary independent Gaussian increments. A Brownian motion with diffusion coefficient D is the scaled Wiener process $X(t) = \sqrt{2D}W(t)$, which satisfies $\mathrm{Var}(X(t)) = 2Dt$ and whose density solves (9.6).

Differentiability of $W(t)$ An important property of Brownian motion is its lack of differentiability: $W(t)$ is continuous everywhere but nowhere differentiable. Here we just give some heuristic explanations to see why the derivative does not exist:

- Suppose we try to calculate the derivative as

$$\xi(t) = \frac{dW(t)}{dt} = \lim_{h\to 0} \frac{W(t+h) - W(t)}{h}. \tag{9.11}$$

  But $W(t+h) - W(t) \sim \mathcal{N}(0,h)$ so $\frac{W(t+h)-W(t)}{h} \sim \mathcal{N}(0, 1/h)$. Hence the variance diverges as $h \to 0$. Despite this, in the physics literature, it is common to speak of "the derivative of Brownian motion" or *white noise* $\xi(t)$.

- By self-similarity: for $a > 0$,

$$\frac{W(at)}{\sqrt{a}} \sim W(t),$$

  so zooming in takes you to the same place, that is, the same level of irregularity.

- By the Markov property of Brownian motion: the future of $W(t)$ for $t > s$ depends only on $W(s) = y$ and not how it got there. In particular, we have no idea how $W(t)$ approached y as $t- > s^-$. Therefore, $W(t)$ "cannot remember" how to leave y in such as way that there would be a tangent there.

9.4 Connection to compartment-based diffusion

Recall the compartment-based diffusion model introduced in Lecture 6, where we considered many molecules of species A and studied the evolution of compartment occupancies $A_i(t)$, $i = 1, \dots, K$. Here suppose that we only have one of such molecules undergoing a lattice jump process. Then in this case it makes sense to take the 'molecule-centered' description, tracking the position of the molecule in time.[2]

Thus, we consider a single molecule moving on a one-dimensional lattice with spacing $h > 0$. Let $X_h(t) \in h\mathbb{Z}$ denote its position at time t. The molecule jumps to the left or right nearest neighbour, with each jump direction occurring at rate $d > 0$. Equivalently, this motion is described by the continuous-time Markov chain with transitions

$$x \xrightarrow{d} x+h, \qquad x \xrightarrow{d} x-h,$$

defined on the state space $h\mathbb{Z}$. Note that this is the continuous-time analogue of the random walk in Section 9.2.

The generator $\mathcal{L}_h$ of this process acts on test functions $f : h\mathbb{Z} \to \mathbb{R}$ as

$$(\mathcal{L}_h f)(x) = d\big[f(x+h) - f(x)\big] + d\big[f(x-h) - f(x)\big]. \tag{9.12}$$

As in (6.10), we introduce the diffusive scaling

$$d = \frac{D}{h^2}, \tag{9.13}$$

where $D > 0$ is fixed. Expanding $f(x \pm h)$ in Taylor series gives

$$f(x \pm h) = f(x) \pm hf'(x) + \frac{h^2}{2} f''(x) + O(h^3).$$

Substituting into the generator yields

$$(\mathcal{L}_h f)(x) = d\left[\frac{h^2}{2} f''(x) + \frac{h^2}{2} f''(x)\right] + O(h) = Df''(x) + O(h).$$

Thus, as $h \to 0$,

$$\mathcal{L}_h \longrightarrow \mathcal{L}, \qquad (\mathcal{L}f)(x) = D\frac{\mathrm{d}^2 f}{\mathrm{d}x^2}.$$

The above argument shows that, under the diffusive scaling (9.13), the generators $\mathcal{L}_h$ converge to the second-order differential operator $\mathcal{L} = D\partial_{xx}$ when tested against smooth functions (so that the Taylor expansion is valid). It turns out that, for this simple process,[3] the limit operator uniquely determines the limiting process, namely, Brownian motion with diffusion coefficient D.

[2] This is in constrast to the 'compartment-centered' description $A_i(t)$, which instead tracks how many molecules we have in each compartment.

[3] In the present nearest-neighbour lattice random walk setting, the convergence of generators holds automatically for all test functions $f \in C_c^\infty(\mathbb{R})$ (or equivalently $f \in C_b^2(\mathbb{R})$): for such functions the Taylor expansion is valid with a uniformly controlled remainder. In addition, the jump sizes are of order h and the total jump rate is of order h^{-2}, so in a finite time you see $O(h^{-2})$ jumps of size h, giving a total displacement $O(1)$. This prevents explosions and ensures tightness of the family of processes $\{X_h(t)\}$. Under these conditions, the limiting second-order operator uniquely characterises Brownian motion.

Tasks

- Formally verify properties (i)–(iii) in Definition 9.1 using the random walk of Section 9.2 and (9.3).
- Construct a modification of the random walk $\widetilde{X}(t)$ (9.2) such that in limit $\Delta t, \Delta x \to 0$, it converges to the Fokker–Planck equation for the Ornstein–Uhlenbeck process (1.6).
- Simulate the discrete-time random walk and its rescaled version $W^N(t)$ for increasing N, and compare the paths with a Brownian motion simulated via Euler–Maruyama.

Example code

`Lect9.ipynb`: contains simulations of random walk process, the rescaled process and Brownian motion.

Historical perspective

Historical note. In 1827, the prominent botanist Robert Brown was studying the structure of pollen grains when, suspending some pollen particles in water, he noticed that these were constantly in motion, performing rapid oscillatory motion without ever stopping. His first assumption was that pollen grains are living matter. He ruled this out by repeating the experiment with (other) inanimate particles, such as dust. While Brown was not the first to observe such microscopic movement, he was the first to study it meticulously and show that it was not due to moving particles being alive [3]. Thereafter his name became associated with this phenomenon, which came to be known as *Brownian* motion.

It took until the beginning of the 20th century before Bachelier, Einstein, Smoluchowski, and Langevin developed the theoretical approaches to Brownian motion and the French physicist Perrin performed the experiments confirming their theoretical results. While Bachelier's PhD thesis in 1900 [2] concerned the analysis of the stock and option markets (and was largely ignored despite being a pioneer to Einstein's until rediscovered in the 1950s' [15]), Einstein, Smoluchowski and Langevin brought Brownian motion to the attention of the scientific community.

Smoluchowski worked on the molecular kinetic approach to Brownian motion, using a detailed kinetic model of the collision of hard spheres and thus treating the solvent particles explicitly [16]. In contrast, Einstein's approach was based on statistical assumptions (so it did not include explicit solvent molecules) and neglected the inertia of the Brownian particle. That is, he never introduced its velocity and only considered its position [4]. In 1905 Einstein obtained a diffusion equation for the Brownian particle and a relation between the diffusion coefficient and measurable physical quantities now known as the *Stokes–Einstein relation*.

The link between the finer approach of Smoluchowski and the coarser approach of Einstein was provided by Langevin in 1908 [10]. His work built on the observation that a particle suspended in a fluid is under a force due to the solvent molecules. This force can be written as a sum of its mean value and a fluctuation about this mean value. His description is on a finer scale than Einstein's, as it considers both the position and velocity of the Brownian particle (the

space of positions and velocities is known as the *phase space*). An important consequence of the works described above was to provide an indirect method to confirm the existence of atoms and molecules. Perrin's experimental verification in 1908 of the Stokes–Einstein relation finally persuaded most anti-atomists that atoms did exist. An excellent historical account of the early stages of the theory of Brownian motion can be found in [11, Chap. 1].

Finally, the rigorous construction of Brownian motion as a continuous stochastic process is due to Norbert Wiener in 1923. Later Wiener proved the non-differentiability of the Brownian paths, which Perrin had conjectured.

Lecture 10: Stochastic differential equations

10.1 From Brownian motion to stochastic differential equations

In Lecture 9 we showed that Brownian motion arises as the diffusive scaling limit of random walks and lattice jump processes. In particular, under the scaling $t \sim x^2$, the generator of a nearest-neighbour random walk converges to

$$\mathcal{L}f = D\frac{\mathrm{d}^2 f}{\mathrm{d}x^2}.$$

We now extend this idea. Suppose that, in addition to random fluctuations, a particle experiences a systematic drift. For example:

- a molecule in a potential field,
- a population relaxing towards equilibrium,
- a chemical species evolving near a deterministic rate equation.

In a small time interval Δt, we may expect an increment of the form

$$X(t+\Delta t) - X(t) \approx a(X(t),t)\Delta t + \sigma(X(t),t)\Delta W,$$

where $\Delta W \sim \mathcal{N}(0,\Delta t)$. Formally, this leads to a stochastic differential equation (SDE)

$$\mathrm{d}X(t) = a(X(t),t)\mathrm{d}t + \sigma(X(t),t)\mathrm{d}W(t), \tag{10.1}$$

where $W(t)$ is a standard Brownian motion.

10.2 Itô stochastic differential equation

The correct interpretation of (10.1) is as a *stochastic integral equation*

$$X(t) = X_0 + \int_0^t a(X(s),s)\mathrm{d}s + \int_0^t \sigma(X(s),s)\mathrm{d}W(s). \tag{10.2}$$

While the first integral can be dealt with in the standard way, the second is a *stochastic integral* that needs to be defined. Because of the nature of $W(t)$ (non-differentiable and with infinite variation), it cannot be understood as an ordinary integral.

To see why a new notion of integral is needed, consider formally computing

$$I(t) = \int_0^t W(s)\,dW(s)$$

using Riemann sums. For a partition $0 = t_0 < \cdots < t_n = t$ with $\Delta t = t/n$, consider

$$I(t) \approx \sum_{k=0}^{n-1} W(\tau_k)\big(W(t_{k+1}) - W(t_k)\big), \qquad \tau_k \in [t_k, t_{k+1}].$$

One can show that, depending on the choice of τ_k:

(left points)	$\tau_k = t_k$,	$\mathbb{E}I_L(t) = 0$,
(midpoints)	$\tau_k = (t_k + t_{k+1})/2$,	$\mathbb{E}I_M(t) = \dfrac{t}{2}$,
(right points)	$\tau_k = t_{k+1}$,	$\mathbb{E}I_R(t) = t$.

Thus different choices of evaluation point lead to different limits as $n \to \infty$. The classical Riemann integral therefore does not exist.

How do we get around this? We must decide ahead of time which point to use to approximate the integrand, which gives rise to different definitions of the stochastic integral. The most common choices are the *Itô* integral (left points) and the *Stratonovich* integral (middle points). Each definition leads to a different "stochastic calculus". The Stratonovich integral leads to the standard chain rule, while the Itô integral requires a correction (Itô's formula (10.12), see below). In this course, we will use Itô.

Definition 10.1 (Itô stochastic integral)**.** The Itô integral is the mean-square limit[1] of the left Riemann sums,

$$\int_0^t f(s)\mathrm{d}W(s) = \lim_{n\to\infty} \sum_{k=0}^{n-1} f(t_k)[W(t_{k+1}) - W(t_k)], \tag{10.3}$$

where $f(t)$ is a square-integrable adapted process[2]

$$\int_0^t \mathbb{E}[f^2(s)]\mathrm{d}s < \infty. \tag{10.4}$$

Two important properties of the Itô integral are:

(i) Martingale property:

$$\mathbb{E}\left[\int_0^t f(s)\mathrm{d}W(s)\right] = 0. \tag{10.5}$$

(ii) Itô isometry:

$$\mathbb{E}\left[\left(\int_0^t f(s)\mathrm{d}W(s)\right)^2\right] = \int_0^t \mathbb{E}\left[f^2(s)\right]\mathrm{d}s. \tag{10.6}$$

Exercise 10.1. Applying the definition, show that

$$\int_0^t W(s)\mathrm{d}W(s) = \frac{1}{2}W^2(t) - \frac{1}{2}t.$$

10.3 Itô stochastic differential equation

We are now ready to define Itô SDEs.

Definition 10.2 (Itô SDE)**.** A stochastic process $X(t)$ is said to satisfy the Itô SDE

$$\mathrm{d}X(t) = a(X(t), t)\mathrm{d}t + \sigma(X(t), t)\mathrm{d}W(t) \tag{10.7}$$

if, for $t \geq 0$,

$$X(t) = X_0 + \int_0^t a(X(s), s)\mathrm{d}s + \int_0^t \sigma(X(s), s)\mathrm{d}W(s), \tag{10.8}$$

where the first integral is a standard Riemann integral and the second is an Itô stochastic integral (Definition 10.1).

[1] X_n converges to X in mean square if $\lim_{n\to\infty} \mathbb{E}(X_n - X)^2 = 0$

[2] Here adapted means that $f(t)$ depends only on information available up to time t.

We will assume throughout that the coefficients are sufficiently regular so that a unique solution to (10.8) exists.

The coefficients $a(x,t)$ and $\sigma(x,t)$ in (10.8) can be interpreted via infinitesimal moments:

(i) The **drift coefficient** $a(x,t)$ (infinitesimal mean):

$$\lim_{\Delta t\to 0}\mathbb{E}\left[\frac{X(t+\Delta t)-X(t)}{\Delta t}\,\middle|\, X(t)=x\right]=a(x,t). \tag{10.9}$$

(ii) The **diffusion coefficient** $\sigma^2(x,t)$ (infinitesimal variance):

$$\lim_{\Delta t\to 0}\mathbb{E}\left[\frac{(X(t+\Delta t)-X(t))^2}{\Delta t}\,\middle|\, X(t)=x\right]=\sigma^2(x,t). \tag{10.10}$$

This interpretation directly connects SDEs with the diffusion limits of CRNs studied earlier in the course.

Suppose that the process $X(t)$ is the solution to (10.7) for $t\in[0,T]$. What SDE does

$$Y(t)=f(X(t)),$$

solve? It turns out that $Y(t)$ satisfies *again* an Itô SDE. From (10.7), we might guess

$$\mathrm{d}Y=f'\mathrm{d}X=f'a\mathrm{d}t+f'\sigma\mathrm{d}W,$$

according to the usual chain rule. *However, this is wrong!* Since $\mathrm{d}W(t)\approx(\mathrm{d}t)^{1/2}$, to compute $\mathrm{d}Y$ we must keep all the terms of order $\mathrm{d}t$ and $(\mathrm{d}t)^{1/2}$. This is the content of Itô's chain rule or Itô's formula.

Itô's chain rule/Itô's formula

Let $X(t)$ solve (10.7) for $t\in[0,T]$, where $X(t),a,\sigma,W(t)\in\mathbb{R}$. Assume that $f=f(x,t)$ is continuous and that its partial derivatives $\partial_t f,\partial_x f,\partial_x^2 f$ exist and are continuous. Then $Y(t)=f(X(t),t)$ satisfies

$$\mathrm{d}Y(t)=\frac{\partial f}{\partial t}(X(t),t)\mathrm{d}t+\frac{\partial f}{\partial x}(X(t),t)\mathrm{d}X(t)+\frac{1}{2}\frac{\partial^2 f}{\partial x^2}(X(t),t)(\mathrm{d}X(t))^2, \tag{10.11}$$

where $(\mathrm{d}X(t))^2$ is computed according to the rules $\mathrm{d}t\cdot\mathrm{d}t=\mathrm{d}t\cdot\mathrm{d}W(t)=\mathrm{d}W(t)\cdot\mathrm{d}t=0$ and $\mathrm{d}W(t)\cdot\mathrm{d}W(t)=\mathrm{d}t$. That is

$$\begin{aligned}\mathrm{d}Y(t)=&\left[\frac{\partial f}{\partial t}(X(t),t)+a(X(t),t)\frac{\partial f}{\partial x}(X(t),t)+\frac{1}{2}\sigma^2(X(t),t)\frac{\partial^2 f}{\partial x^2}(X(t),t)\right]\mathrm{d}t\\&+\sigma(X(t),t)\frac{\partial f}{\partial x}(X(t),t)\mathrm{d}W(t)\end{aligned} \tag{10.12}$$

For example, let $X(t)=W(t)$ and $f(x)=x^2$. Then

$$\mathrm{d}(W^2)=2W\mathrm{d}W+\mathrm{d}t.$$

Integrating this gives the identity $\int_0^t W(s)\mathrm{d}W(s)=[W^2(t)-t]/2$. Consider the SDE for *geometric Brownian motion*

$$\mathrm{d}S(t)=\mu S(t)\mathrm{d}t+\sigma S(t)\mathrm{d}W(t),\qquad S(0)=S_0. \tag{10.13}$$

We can readily solve (10.13) using Itô's formula with $f(S) = \ln(S)$. It follows that

$$\mathrm{d}\ln(S(t)) = \left(\mu - \frac{1}{2}\sigma^2\right)\mathrm{d}t + \sigma \mathrm{d}W(t),$$

or

$$S(t) = S_0 e^{\sigma W(t) + \left(\mu - \frac{1}{2}\sigma^2\right)t}. \tag{10.14}$$

This example illustrates how multiplicative noise modifies the deterministic growth rate.

Multivariate Itô's formula

Consider the multidimensional process $\boldsymbol{X}(t)$ in $\mathbb{R}^d$

$$\mathrm{d}\boldsymbol{X}(t) = \boldsymbol{a}(\boldsymbol{X}(t), t)\mathrm{d}t + \boldsymbol{\sigma}(\boldsymbol{X}(t), t)\mathrm{d}\mathbf{W}(t), \tag{10.15}$$

where $\boldsymbol{a}(\boldsymbol{x}, t) \in \mathbb{R}^d, \boldsymbol{\sigma}(\boldsymbol{x}, t) \in \mathbb{R}^{d\times m}$ and $\mathbf{W}(t) = (W_1(t), \ldots, W_m(t)) \in \mathbb{R}^m$.
Let $Y(t) = f(\boldsymbol{X}(t), t)$, where $f : \mathbb{R}^d \times [0, T] \to \mathbb{R}$, $f \in C^{2\times 1}(\mathbb{R}^d \times [0, T])$. The multidimensional Itô formula states

$$\mathrm{d}Y(t) = \frac{\partial f}{\partial t}\mathrm{d}t + \sum_{i=1}^{d} \frac{\partial f}{\partial x_i}\mathrm{d}X_i(t) + \frac{1}{2}\sum_{i,j=1}^{d} \frac{\partial^2 f}{\partial x_i \partial x_j}\mathrm{d}X_i(t)\mathrm{d}X_j(t), \tag{10.16}$$

with the convention $\mathrm{d}W_i(t)\mathrm{d}W_j(t) = \delta_{ij}\mathrm{d}t$, $\mathrm{d}W_i(t)\mathrm{d}t = 0$ for $i, j = 1, \ldots, d$.

10.4 Numerical simulation of SDEs

In most cases, SDEs cannot be solved explicitly and must be simulated numerically. Consider the scalar time-homogeneous SDE

$$\mathrm{d}X(t) = a(X(t))\mathrm{d}t + \sigma(X(t))\mathrm{d}W(t), \qquad X(0) = X_0, \tag{10.17}$$

Partition $[0, T]$ into steps of size Δt, define $t_n = n\Delta t$, and denote by X_n the approximation to $X(t_n)$. The Euler–Maruyama approximation is

$$X_{n+1} = X_n + a(X_n)\Delta t + \sigma(X_n)\Delta W_n,$$

where $\Delta W_n = W(t_{n+1}) - W(t_n) \sim \mathcal{N}(0, \Delta t)$.

Algorithm 10.1: Euler–Maruyama scheme

Set $X_0 = x_0$.
For $n = 0$ to $n = N - 1$

(a) Generate a random number $\xi \sim \mathcal{N}(0, 1)$ (the standard normal distribution).

(b) Set

$$X_{n+1} = X_n + a(X_n)\Delta t + \sigma(X_n)\sqrt{\Delta t}\,\xi. \tag{10.18}$$

end

The Euler–Maruyama scheme (10.18) is the simplest of the so-called Taylor schemes. Next we show a systematic method to obtain higher-order methods. First we write (10.17) in integral form

$$X(t) = X_0 + \int_0^t a(X(s))\mathrm{d}s + \int_0^t \sigma(X(s))\mathrm{d}W(s). \tag{10.19}$$

Consider a function $f \in C^2(\mathbb{R})$ and recall Itô's formula (10.12), which in this case reads

$$\mathrm{d}f(X(t)) = \left[a(X(t))\frac{\partial f}{\partial x}(X(t)) + \frac{1}{2}\sigma^2(X(t))\frac{\partial^2 f}{\partial x^2}(X(t))\right]\mathrm{d}t + \sigma(X(t))\frac{\partial f}{\partial x}(X(t))\mathrm{d}W(t).$$

Let's write it as an Itô integral equation

$$f(X(t)) = f(X_0) + \int_0^t L_0 f(X(s))\mathrm{d}s + \int_0^t L_1 f(X(s))\mathrm{d}W(s), \tag{10.20}$$

with

$$L_0 = a(x)\frac{\partial}{\partial x} + \frac{1}{2}\sigma^2(x)\frac{\partial^2}{\partial x^2}, \qquad L_1 = \sigma(x)\frac{\partial}{\partial x}.$$

Applying (10.20) to each of the integrands in (10.19), that is, with $f = a$ in the first integral and $f = \sigma$ in the second, we obtain

$$X(t) = X_0 + a(X_0)\int_0^t \mathrm{d}s + \sigma(X_0)\int_0^t \mathrm{d}W(s) + R_1, \tag{10.21}$$

with reminder

$$\begin{aligned} R_1 = &\int_0^t\int_0^s L_0 a(X(s'))\mathrm{d}s'\mathrm{d}s + \int_0^t\int_0^s L_1 a(X(s'))\mathrm{d}W(s')\mathrm{d}s \\ &+ \int_0^t\int_0^s L_0\sigma(X(s'))\mathrm{d}s'\mathrm{d}W(s) + \int_0^t\int_0^s L_1\sigma(X(s'))\mathrm{d}W(s')\mathrm{d}W(s). \end{aligned}$$

We see that the Euler–Maruyama consists of neglecting R_1 altoghether. But we can carry on with the expansion of R_1 to obtain more accurate schemes. The expression for R_1 looks complicated, but we only need to understand its scaling in Δt. Noting that, for $\alpha, \beta \geq 0$,

$$(\Delta t)^\alpha(\Delta W)^\beta = O(\Delta t^{\alpha+\beta/2}),$$

we see that the leading-order term in R_1 is

$$R_1 = \int_0^t\int_0^s L_1\sigma(X(s'))\mathrm{d}W(s')\mathrm{d}W(s) + o(\sqrt{ts}). \tag{10.22}$$

Exercise 10.2. Apply (10.20) with $f = L_1\sigma$ to compute the leading-order contribution in R_1 and derive the Milstein scheme (10.23).

Algorithm 10.2: Milstein scheme

Set $X_0 = x_0$.

For $n = 0$ to $n = N - 1$

(a) Generate a random number $\xi \sim \mathcal{N}(0, 1)$ (the standard normal distribution).

(b) Set

$$X_{n+1} = X_n + a(X_n)\Delta t + \sigma(X_n)\sqrt{\Delta t}\,\xi + \frac{1}{2}\sigma(X_n)\sigma'(X_n)\Delta t(\xi^2 - 1). \tag{10.23}$$

end

Note that, for *additive-noise processes* ($\sigma(x) \equiv \sigma$ constant), the Milstein scheme reduces to the Euler–Maruyama scheme. One could, in theory, continue with the expansion to derive more accurate Taylor schemes. But, in practice, they are not used.

10.4.1 Strong and weak convergence

Because both $X(t)$ and its approximation X_n are random, there are two natural notions of convergence: *strong convergence* considers pathways results, while *weak convergence* deals with probability distributions.

Definition 10.3 (Strong convergence)**.** The scheme converges strongly with order α if

$$e_{\Delta t}^{\text{strong}} := \sup_{t_n \leq T} \mathbb{E}|X(t_n) - X_n| \leq C\Delta t^{\alpha}. \tag{10.24}$$

Definition 10.4 (Weak convergence)**.** The scheme converges weakly with order β if, for suitable test functions Φ,

$$e_{\Delta t}^{\text{weak}} := \sup_{t_n \leq T} |\mathbb{E}[\Phi(X(t_n))] - \mathbb{E}[\Phi(X_n)]| \leq C\Delta t^{\beta}. \tag{10.25}$$

It is typical to restrict Φ to be a member of a class of functions, such as polynomials of degree at most k [7, Chap. 8].

Thus, the strong convergence is the mean of the error and considers how accurately the approximation follows the paths. In contrast, the weak convergence is the error of the means and measures how well the approximation captures the average behaviour. It is hence clear the weak order of convergence of a scheme is always greater or equal to its strong order ($\beta \geq \alpha$).

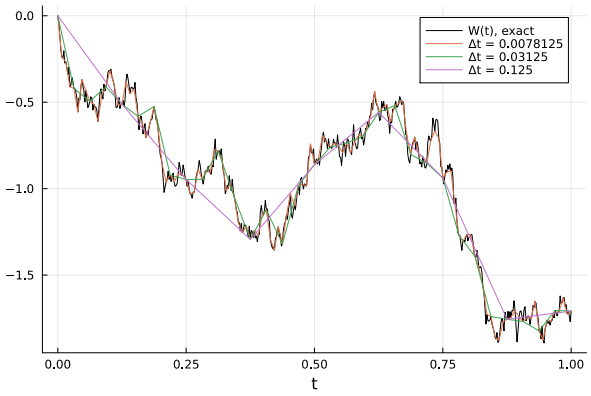

Figure 10.1: Exact path of Brownian motion $W(t)$ and evaluation at coarser time grids.

Strong convergence requires convergence as $\Delta t \to 0$ but for a *fixed realisation* of the solution $X(t)$ to (10.17). Therefore, in numerical convergence tests, the same realisation of $W(t)$ must be used for all approximations (see Figure 10.1). In contrast, for weak convergence, we can use different paths. Finally, we must also be aware of other sources of error [8], which are implicitly assumed negligible when monitoring the errors $e_{\Delta t}^{\text{strong}}$ and $e_{\Delta t}^{weak}$:

- Sampling error: the error arising from approximating an expected value by a sample mean. It decays as $1/\sqrt{M}$, where M is the number of sample paths used.

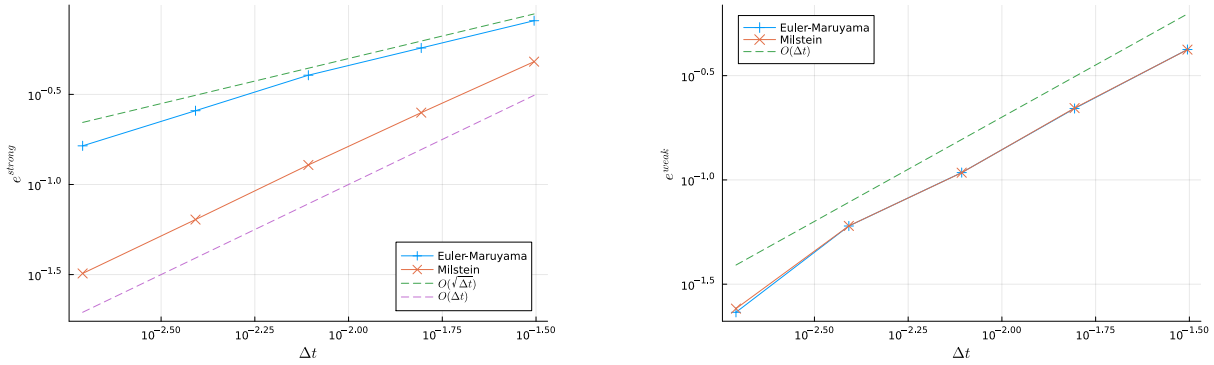

Figure 10.2: Strong (left) and weak (right) endpoint errors of convergence of the Euler–Maruyama (10.18) and Milstein (10.23) schemes to approximate Geometric Brownian Motion (10.26). Parameters used: $\mu = 2$, $\sigma = 1$, $X_0 = 1$. Errors measured at the final time $T = 1$. For the weak error (10.25), we use $\Phi(x) = x$.

- Random number bias: errors in the random number generator.
- Rounding error: floating-point roundoff errors.

Typically the sampling error will be the most significant of these three, so we should take M large enough.

Example: numerical convergence test with Geometric Brownian Motion To test the two numerical schemes, we need a process with *multiplicative noise*, $\sigma'(x) \neq 0$. A good choice is Geometric Brownian motion (10.13) as we have the explicit solution in hand, that is,

$$X(t) = X_0 e^{\sigma W(t) + \left(\mu - \frac{1}{2}\sigma^2\right)t}. \tag{10.26}$$

In Figure 10.2, we show the result of numerical tests of the strong and the weak convergence of approximate solutions $X(t)$ using the Euler–Maruyama (10.18) and Milstein (10.23) schemes to the exact solution $X(t)$ in (10.26). Instead of taking the supremum in (10.24) and (10.25), we consider the endpoint errors at time $T = 1$. The numerical tests suggest that:

	Strong order	Weak order
Euler–Maruyama (SSA 10.1)	1/2	1
Milstein (SSA 10.2)	1	1

These orders of convergence are valid for a general SDE (10.17) and can be proven rigorously. Note that for SDEs with additive noise, $\sigma(x)$ constant, the two schemes are equivalent; hence, for diffusions with additive noise, the Euler–Maruyama scheme has a strong order of convergence equal to one.

Lecture 11: Kolmogorov equations of diffusion processes

In Lecture 10, we described stochastic dynamics at the level of sample paths via stochastic differential equations (SDEs):

$$\mathrm{d}X(t) = a(X(t), t)\mathrm{d}t + \sigma(X(t), t)\mathrm{d}W(t).$$

We now take a complementary viewpoint:

How does the probability distribution of $X(t)$ evolve in time?

This parallels the first half of the course:

- For continuous-time Markov chains, we derived the *master equation*.
- For diffusion processes, we will derive the *Fokker–Planck equation*.

As in the discrete case, the bridge between the two viewpoints is the *generator*.

11.1 Infinitesimal classification of Markov processes

Consider a continuous-time scalar Markov process $X(t)$ with associated transition probability density $p(x, t \mid y, s)$.[1]

Jump processes revisited. In the first half of the course we studied stochastic chemical reaction networks with state $X(t) \in \mathbb{Z}^d$ evolving through jumps

$$X \longrightarrow X + \zeta_j,$$

with propensity $\alpha_j(X)$. The corresponding chemical master equation (see (4.12)) reads

$$\partial_t p(x, t) = \sum_j \Big[\alpha_j(x - \zeta_j)\, p(x - \zeta_j, t) - \alpha_j(x)\, p(x, t)\Big].$$

Equivalently, in terms of the short-time transition probabilities, this corresponds to *jump rates from y to x*

$$W(x \mid y, t) = \sum_j \alpha_j(y)\, \delta_{x, y+\zeta_j},$$

so that, from (4.2)

$$p(x, t + \Delta t \mid y, t) = \begin{cases} 1 - \sum_{z \neq y} W(z \mid y, t)\Delta t + o(\Delta t), & x = y, \\ W(x \mid y, t)\Delta t + o(\Delta t), & x \neq y. \end{cases}$$

[1] Here we loosely use p for both the probability density and the probability mass function when X has a discrete state space.

Continuous-state processes. Let $X(t)$ be a continuous-state Markov process with transition density $p(x,t|y,s)$. As in the discrete case, the short-time behaviour may include a jump kernel. For every $\varepsilon > 0$ and all x, y such that $|x-y| \geq \varepsilon$,

$$W(x|y,t) = \lim_{\Delta t \to 0^+} \frac{p(x, t+\Delta t|y,t)}{\Delta t}. \tag{11.1}$$

In addition, the process may also possess a continuous (diffusive) component. To separate these two contributions, we define the drift and diffusion coefficients via *truncated* moments:

$$a(y,t) = \lim_{\varepsilon \to 0} \lim_{\Delta t \to 0^+} \frac{1}{\Delta t} \int_{|x-y| \leq \varepsilon} (x-y)\, p(x, t+\Delta t|y,t) \mathrm{d}x, \tag{11.2}$$

$$b(y,t) = \lim_{\varepsilon \to 0} \lim_{\Delta t \to 0^+} \frac{1}{\Delta t} \int_{|x-y| \leq \varepsilon} (x-y)^2\, p(x, t+\Delta t|y,t) \mathrm{d}x. \tag{11.3}$$

These coefficients measure the contribution of arbitrarily small increments of the process. Large increments (jumps of finite size) are described separately by the kernel $W(x|y,t)$.

If $W(x|y,t) = 0$ for all $x \neq y$, the process has no jump component and its sample paths are almost surely continuous. In this case, the truncation can be removed, and the drift and diffusion coefficients reduce to the full infinitesimal moments (10.9) and (10.10)

$$\begin{aligned} a(y,t) &= \lim_{\Delta t \to 0^+} \frac{1}{\Delta t} \int_{\mathbb{R}} (x-y)\, p(x, t+\Delta t|y,t) \mathrm{d}x \equiv \lim_{\Delta t \to 0^+} \mathbb{E}\left[\frac{X(t+\Delta t) - X(t)}{\Delta t} \middle| X(t) = x \right] \\ b(y,t) &= \lim_{\Delta t \to 0^+} \frac{1}{\Delta t} \int_{\mathbb{R}} (x-y)^2\, p(x, t+\Delta t|y,t)\, dx \equiv \lim_{\Delta t \to 0^+} \mathbb{E}\left[\frac{(X(t+\Delta t) - X(t))^2}{\Delta t} \middle| X(t) = x \right] \end{aligned} \tag{11.4}$$

Definition 11.1 (Diffusion process)**.** A diffusion process $X(t)$ is a continuous-state Markov process with continuous sample paths and mean $a(x,t)$ and variance $b(x,t)$ (11.4). In this case, it admits an SDE representation

$$\mathrm{d}X(t) = a(X(t),t)\mathrm{d}t + \sigma(X(t),t)\mathrm{d}W(t), \qquad b = \sigma^2. \tag{11.5}$$

11.2 The generator of a diffusion process

Let $f \in C^2(\mathbb{R})$. Given $X(t)$ satisfying (11.5), by Itô's formula (10.12)

$$\mathrm{d}f(X(t)) = \left(a(X(t),t) f'(X(t)) + \frac{1}{2}\sigma^2(X(t),t) f''(X(t)) \right) \mathrm{d}t + \sigma(X(t),t) f'(X(t)) \mathrm{d}W(t).$$

Taking expectations and using the martingale property of the Itô integral (10.5),

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}[f(X(t))] = \mathbb{E}\left[a(X(t),t) f'(X(t)) + \frac{1}{2} b(X(t),t) f''(X(t)) \right].$$

Definition 11.2 (Generator)**.** The generator $\mathcal{L}$ of the diffusion process (11.5) is the differential operator

$$\mathcal{L} f(x) = a(x,t) f'(x) + \frac{1}{2} b(x,t) f''(x). \tag{11.6}$$

We occasionally write $\mathcal{L}_{x,t}$ or $\mathcal{L}_{y,s}$ to emphasize which variables the operator acts on.

Therefore,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[f(X(t))] = \mathbb{E}[(\mathcal{L}f)(X(t))]. \tag{11.7}$$

This relation plays the same role as the generator for continuous-time Markov chains (compare with (4.15)).

11.3 The Fokker–Planck equation

Let $p(x,t \mid y,s)$ denote the transition probability density of the diffusion process. Fix y and s and define

$$u(y,t) := \mathbb{E}[f(X(t)) \mid X(s) = y] = \int_{\mathbb{R}} f(x)\, p(x,t \mid y,s)\mathrm{d}x.$$

From the generator relation (11.7),

$$\frac{\partial}{\partial t}u(y,t) = \mathbb{E}[(\mathcal{L}f)(X(t)) \mid X(s) = y]$$

or

$$\frac{\partial}{\partial t}\int f(x)\, p(x,t \mid y,s)\mathrm{d}x = \int (\mathcal{L}f)(x)\, p(x,t \mid y,s)\mathrm{d}x.$$

Substituting the expression for $\mathcal{L}$ (11.6),

$$\int f(x)\partial_t p(x,t \mid y,s)\mathrm{d}x = \int \left(a(x,t)f'(x) + \frac{1}{2}b(x,t)f''(x)\right) p(x,t \mid y,s)\mathrm{d}x.$$

Integrating by parts and assuming suitable decay at infinity,

$$\int_{\mathbb{R}} f(x)\partial_t p\,\mathrm{d}x = \int f(x)\left[-\partial_x(ap) + \frac{1}{2}\partial_{xx}(bp)\right]\mathrm{d}x.$$

Since this holds for all smooth test functions f, we conclude:

$$\partial_t p(x,t \mid y,s) = -\partial_x\left[a(x,t)p(x,t \mid y,s)\right] + \frac{1}{2}\partial_{xx}\left[b(x,t)p(x,t \mid y,s)\right].$$

The Fokker–Planck equation

Let $X(t)$ be diffusion process in $\mathbb{R}$ and assume that $p(x,t|\cdot,\cdot), a(x,t), b(x,t) \in C^{2,1}(\mathbb{R} \times \mathbb{R}_>)$. Then the transition probability density satisfies the forward Kolmogorov or Fokker–Planck equation

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}[a(x,t)p] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[b(x,t)p] := \mathcal{L}^*_{x,t}p, \tag{11.8}$$

with $\lim_{t\to s} p(x,t|y,s) = \delta(x-y)$.

Notes

- The Fokker–Planck PDE (11.8) is the continuous analogue of the master equation (4.12).
- If the process is time-homogeneous, then

$$p(x,t \mid y,s) = p(x,t-s \mid y,0),$$

and the coefficients a, b are independent of t.

- The Fokker–Planck equation can be written as a conservation law

$$\partial_t p + \partial_x J = 0,$$

where the *probability flux* is

$$J(x,t) = a(x,t)p(x,t) - \frac{1}{2}\partial_x\big[b(x,t)p(x,t)\big].$$

- In (11.8), we write $\mathcal{L}^*_{x,t}$ to emphasize that, formally, the differential operator in this equation is the adjoint of the generator $\mathcal{L}$ in (11.6). The subscripts (x,t) emphasize that it is the *forward variables* x, t that we are using to evaluate the coefficients and derivatives.

Remark 11.1 (Why only first and second moments?)**.** A diffusion process is characterised by the (almost sure) continuity of its sample paths and by specifying its first two infinitesimal moments, namely the drift a and diffusion coefficient b.

The key reason that the Fokker–Planck equation contains at most second derivatives is the scaling of the increments of a diffusion process. For small time steps Δt, we have

$$X(t+\Delta t) - X(t) = O(\sqrt{\Delta t}).$$

Consequently,

$$\frac{\mathbb{E}[(X(t+\Delta t) - X(t))^n]}{\Delta t} = O(\Delta t^{\frac{n}{2}-1}).$$

For $n \geq 3$, this quantity vanishes as $\Delta t \to 0$. Thus only the first and second infinitesimal moments survive in the limit, and the generator is necessarily a second-order differential operator.

A natural question is whether one could define stochastic processes by specifying a fixed finite number of moments higher than two. It turns out that this is not possible: either one retains only the first two moments (leading to diffusion processes), or one must retain all higher-order moments. In the latter case, the process necessarily has jumps and the generator contains a nonlocal integral term involving the jump kernel $W(x \mid y,t)$, see (11.16). Specifying a finite number of moments greater than two leads to inconsistencies (see [13, §2.6]).

11.4 The backward Kolmogorov equation

The Fokker-Planck equation tells us how the probability density function $p(x,t \,|\, y,s)$ associated with an SDE evolves given some initial state y at some initial time s. Suppose we fix the terminal time t and view the transition density as a function of the initial variables (y,s). We now derive the evolution equation in the initial time s. Such a problem is called a *final value problem* rather than an initial value problem.

In contrast to the forward equation, which differentiates with respect to the terminal time t, to obtain the backward equation, we now differentiate with respect to the initial time s. We start with the Chapman-Kolmogorov equation

$$p(x,t \mid y, s-\Delta s) = \int_{\mathbb{R}} p(x,t \mid z,s)\, p(z,s \mid y, s-\Delta s)\mathrm{d}z, \tag{11.9}$$

for small $\Delta s > 0$ such that $s - \Delta s < s < t$. Taylor expanding $p(x, t \,|\, z, s)$ about the point $z = y$, we get

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

If we now substitute for $p(x, t \,|\, z, s)$ in (11.9), the first term in the Taylor expansion gives rise to a term of the form

$$p(x,t \,|\, y,s) \int_{\mathbb{R}} p(z,s \,|\, y, s-\Delta s)\,\mathrm{d}z = p(x,t \,|\, y,s)\,,$$

where the integral is equal to 1 from the law of total probability. The next term gives, using (**??**)

$$\frac{\partial p}{\partial y}(x,t \,|\, y,s) \int_{\mathbb{R}} (z-y)\, p(z,s \,|\, y, s-\Delta s)\,\mathrm{d}z = \frac{\partial p}{\partial y}(x,t \,|\, y,s)\, a(y,s)\,\Delta s + o(\Delta s)\,.$$

Similarly, using (**??**), the third term gives

$$\frac{1}{2}\frac{\partial^2 p}{\partial y^2}(x,t \,|\, y,s) \int_{\mathbb{R}} (z-y)^2\, p(z,s \,|\, y, s-\Delta s)\,\mathrm{d}z = \frac{\partial^2 p}{\partial y^2}(x,t \,|\, y,s)\,\frac{b(y,s)}{2}\,\Delta s + o(\Delta s).$$

Putting this all together, we get

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

Subtracting $p(x, t \,|\, y, s)$ from both sides, dividing through by Δs and letting $\Delta s \to 0$ we get the backward Kolmogorov equation:

$$-\frac{\partial p}{\partial s}(x,t \,|\, y,s) = a(y,s)\,\frac{\partial p}{\partial y}(x,t \,|\, y,s) + \frac{b(y,s)}{2}\,\frac{\partial^2 p}{\partial y^2}(x,t \,|\, y,s) = \mathcal{L}_{y,s} p,$$

with *final condition* $\lim_{s\to t} p(x,t|y,s) = \delta(x-y)$. Note that the right-hand side coincides exactly with the generator of the process (11.6); we include the subscripts (y, s) to emphasize that the backward equation involves the initial variables y, s.

This derivation mirrors that of the forward equation. In the forward case, Itô's formula effectively performs a second-order Taylor expansion in the terminal variable $X(t)$, leading to the generator acting on test functions. Here, instead, we perform a Taylor expansion in the initial variable y via the Chapman–Kolmogorov equation. Both derivations rely on the same infinitesimal structure of the process and lead to dual (adjoint) PDEs.

Time-homogeneous diffusion process Note that the backward Kolmogorov equation is a final value problem for a parabolic PDE. But we can turn it into an initial value problem if the process is time-homogeneous (see Definition A.5). In this case, $p(x,t|y,s) = p(x,t-s|y,0)$ and the drift and diffusion coefficients are time-independent, $a(s,t) = a(y)$ and $b(y,s) = b(y)$. Then we have that $\partial_s p = -\partial_t p$ and can set the initial time $s = 0$, leading to the *time-homogeneous backward Kolmogorov equation* for $t > 0$

$$\frac{\partial p}{\partial t} = a(y)\frac{\partial p}{\partial y} + \frac{1}{2}b(y)\frac{\partial^2 p}{\partial y^2},$$

with *initial condition* $\lim_{t\to 0} p(x,t|y,0) = \delta(x-y)$.

The backward Kolmogorov equation

Let $X(t)$ be diffusion process in $\mathbb{R}$ and assume that $p(x,t|\cdot,\cdot), a(x,t), b(x,t) \in C^{2,1}(\mathbb{R}\times\mathbb{R}_>)$. Then the transition probability density satisfies the backward Kolmogorov equation

$$-\frac{\partial p}{\partial s} = \mathcal{L}_{y,s}\, p, \qquad \mathcal{L}_{y,s}p := a(y,s)\frac{\partial p}{\partial y} + \frac{1}{2}b(y,s)\frac{\partial^2 p}{\partial y^2}, \tag{11.10}$$

If $X(t)$ is time-homogeneous, then the backward Kolmogorov equation reads

$$\frac{\partial p}{\partial t} = \mathcal{L}_{y}\, p, \qquad \mathcal{L}_{y}p := a(y)\frac{\partial p}{\partial y} + \frac{1}{2}b(y)\frac{\partial^2 p}{\partial y^2}. \tag{11.11}$$

11.5 Multi-dimensional diffusion processes

Let $\boldsymbol{X}(t)$ be a diffusion process in $\mathbb{R}^d$ satisfying the SDE (10.15), that is, with drift vector and diffusion matrix given by

$$\boldsymbol{a}(\boldsymbol{y},t) = \lim_{\Delta t\to 0} \mathbb{E}\left(\frac{\boldsymbol{X}(t+\Delta t)-\boldsymbol{X}(t)}{\Delta t}\middle|\boldsymbol{X}(t)=\boldsymbol{y}\right), \tag{11.12}$$

$$\mathbf{B}(\boldsymbol{y},t) = \lim_{\Delta t\to 0} \mathbb{E}\left(\frac{[\boldsymbol{X}(t+\Delta t)-\boldsymbol{X}(t)]\otimes[\boldsymbol{X}(t+\Delta t)-\boldsymbol{X}(t)]}{\Delta t}\middle|\boldsymbol{X}(t)=\boldsymbol{y}\right). \tag{11.13}$$

Here $\boldsymbol{a} = (a_1,\dots,a_d)$ is a d-dimensional vector and the diffusion coefficient $\mathbf{B} = (B_{ij})$ is a $d\times d$ *symmetric* nonnegative matrix. Note the relation $\mathbf{B} = \boldsymbol{\sigma}\boldsymbol{\sigma}^\top$ (in particular, this means that more than one $\boldsymbol{\sigma}$ is consistent with the same $\mathbf{B}$).

The transition probability density $p(\boldsymbol{x},t|\boldsymbol{y},s)$ associated to $\boldsymbol{X}(t)$ satisfies the backward Kolmogorov equation

$$-\frac{\partial p}{\partial s} = \sum_{i=1}^{d} a_i(\boldsymbol{y},s)\frac{\partial p}{\partial y_i} + \frac{1}{2}\sum_{i,j=1}^{d} B_{ij}(\boldsymbol{y},s)\frac{\partial^2 p}{\partial y_i \partial y_j}. \tag{11.14}$$

The corresponding forward Kolmogorov equation (or Fokker–Planck equation) is

$$\frac{\partial p}{\partial t} = -\sum_{i=1}^{d}\frac{\partial}{\partial x_i}[a_i(\boldsymbol{x},t)p] + \frac{1}{2}\sum_{i,j=1}^{d}\frac{\partial^2}{\partial x_i \partial x_j}[B_{ij}(\boldsymbol{x},t)p]. \tag{11.15}$$

11.6 Jump vs. diffusion processes

We now have the continuous-state analogue of the structure developed earlier in the course:

Continuous-time Markov chain	Diffusion process
Jump rates W	Drift a, diffusion b
Generator matrix Q	Differential generator $\mathcal{L}$
Master equation	Fokker–Planck equation
Backward equation	Backward Kolmogorov equation

Remark 11.2 (Jump-diffusion processes)**.** A continuous-time Markov process may possess *both* a jump component and a diffusion component. Such processes are often called *Lévy-type processes* or *jump–diffusions*. In this case, the infinitesimal generator has the form

$$(\mathcal{L}f)(x) = a(x,t)f'(x) + \frac{1}{2}b(x,t)f''(x) + \int_{\mathbb{R}} [f(y) - f(x)]\, W(y|x,t)\mathrm{d}y, \tag{11.16}$$

where the drift and diffusion coefficients are given in (11.2) and (11.3), respectively, and the jump rate in (11.1). The integral term reduces to a sum of the form given in (4.14) for discrete-state spaces.

The forward Kolmogorov equation then contains both a differential part and an integral (master equation) part:

$$\partial_t p = -\partial_x(ap) + \frac{1}{2}\partial_{xx}(bp) + \int_{\mathbb{R}} [W(x|y,t)p(y,t) - W(y|x,t)p(x,t)]\, \mathrm{d}y.$$

Thus, pure jump processes (CME) and pure diffusion processes (Fokker–Planck) are two special cases of (11.16).

Lecture 12: Diffusion in bounded domains and first-exit problems

So far we have assumed that the state space is the whole of $\mathbb{R}$ (or $\mathbb{R}^d$), with densities decaying sufficiently fast at infinity. In many applications, this is not appropriate. Instead, the dynamics takes place in a bounded region, and interactions with the boundary play a crucial role.

12.1 Diffusion in a bounded domain

Consider the time-homogeneous SDE

$$\mathrm{d}X(t) = a(X(t))\mathrm{d}t + \sqrt{b(X(t))}\,\mathrm{d}W(t), \tag{12.1}$$

in a bounded interval $\Omega = (0, L)$. The probability density $p(x, t)$ satisfies the Fokker–Planck equation for $x \in \Omega$

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}[a(x)p] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[b(x)p], \qquad p(x, 0) = p_0(x). \tag{12.2}$$

It is convenient to rewrite this equation in conservation form:

$$\frac{\partial p}{\partial t} + \frac{\partial J}{\partial x} = 0, \tag{12.3}$$

where the probability flux is defined by

$$J(x, t) = a(x)p - \frac{1}{2}\frac{\partial}{\partial x}[b(x)p]. \tag{12.4}$$

Integrating (12.3) over Ω gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega p(x, t)\,\mathrm{d}x = J(0, t) - J(L, t). \tag{12.5}$$

Thus, total probability changes only through flux across the boundary. In contrast to the case $\Omega = \mathbb{R}$, the boundary conditions now determine whether probability is conserved or lost.

Common boundary conditions

(i) Reflecting boundary (e.g., the particle hits a wall and is reflected back into the domain). A reflecting boundary corresponds to zero probability flux:

$$J(x, t) = 0, \quad \text{at the boundary.}$$

In this case, probability is conserved.

(ii) Periodic boundary. (e.g., the particle is moving in a circular arena or $X(t)$ describes an intrinsically periodic quantity such as a particle orientation):

$$J(L^+, t) = J(0^-, t), \qquad p(L^+, t) = p(0^-, t).$$

This is an example where probability is conserved, but flux may be non-zero. For this boundary condition to make sense, we require the coefficients a and b also to be L-periodic.

(iii) Absorbing boundary (e.g., the particle hits a reactive wall and is removed):

$$p(x,t) = 0, \quad \text{at the boundary.}$$

Here, the total probability decreases in time.

(iv) Partially reflecting / absorbing boundaries (e.g., the particle hits an imperfect reactive wall, with a certain probability, is absorbed, and otherwise is reflected):

$$J(x,t)n = \kappa p(x,t), \quad \text{at the boundary,}$$

where n is the unit outward ($n = \pm 1$ in one dimension) and κ is the reactivity of the boundary: $\kappa = 0$ corresponds to the pure reflective boundary condition, and $\kappa = \infty$ corresponds to the perfect sink or absorbing boundary above.

(v) Mixed boundary conditions: these involve a combination of the above on different boundaries.

We now examine how they appear in the backward formulation.

12.2 Boundary conditions on the backward operator

Recall that the backward generator associated with (12.1) is

$$\mathcal{L}f = a(x)f'(x) + \frac{1}{2}b(x)f''(x). \tag{12.6}$$

Formally, $\mathcal{L}$ is the adjoint of the Fokker–Planck operator $\mathcal{L}^*$. To make this precise, consider the L^2 inner product

$$\langle f, p\rangle = \int_\Omega f(x)p(x)\,\mathrm{d}x.$$

A straightforward integration by parts yields

$$\langle \mathcal{L}f, p\rangle = \langle f, \mathcal{L}^* p\rangle + \left[fJ + \frac{1}{2}b(x)p(x)f'(x)\right]_{x=0}^{x=L}. \tag{12.7}$$

To ensure that the adjoint relation holds without boundary terms, the function f must satisfy boundary conditions compatible with those imposed on p.

(i) Reflecting boundary. Since $J = 0$ at the boundary, the first term vanishes. To eliminate the second term we require

$$b(x)f'(x) = 0 \quad \text{at the boundary.} \tag{12.8}$$

Thus, reflecting boundaries correspond to homogeneous Neumann-type conditions for the backward operator.

(ii) Absorbing boundary. Since $p = 0$ at the boundary, the second term vanishes. To eliminate the first term we require

$$f(x) = 0 \quad \text{at the boundary.} \tag{12.9}$$

Thus absorbing boundaries correspond to Dirichlet conditions for the backward problem. Other boundary conditions for $\mathcal{L}$ can be derived similarly.

12.3 Boundary conditions at the level of the SDE

So far, we have imposed boundary conditions at the level of the Fokker–Planck and backward equations. It is natural to ask how these conditions should be interpreted in terms of the sample paths of the SDE, and how to incorporate them in numerical simulations.

To make this concrete, consider the Euler–Maruyama discretisation of (12.1):

$$X_{n+1} = X_n + a(X_n)\Delta t + \sqrt{b(X_n)\Delta t}\,\xi_n, \qquad \xi_n \sim \mathcal{N}(0,1). \tag{12.10}$$

Periodic boundary If $\Omega = [0, L]$ with periodic boundary conditions, any proposed step leaving the domain is wrapped back into it:

$$X_{n+1} \leftarrow \mathrm{mod}(X_{n+1}, L).$$

This corresponds to identifying the endpoints 0 and L.

Reflecting boundary Suppose $\Omega = [0, \infty)$ and $x = 0$ is reflecting.

A naive approach would be to reject any step with $X_{n+1} < 0$. However, this introduces bias and does not correctly approximate the continuous reflecting diffusion. A better approximation is to reflect the step:

$$X_{n+1} \leftarrow -X_{n+1} \quad \text{if } X_{n+1} < 0. \tag{12.11}$$

Heuristically, this corresponds to reversing the normal component of the motion at the boundary. For pure diffusion ($a = 0$, b constant), this reflection procedure introduces no additional approximation beyond the time discretisation.

It is important to note that reflecting boundary conditions define a different stochastic process, known as *reflecting Brownian motion*. For example, Brownian motion on $[0, \infty)$ with reflection at 0 can be characterised as the solution of

$$\mathrm{d}X(t) = \sqrt{2D}\,\mathrm{d}W(t) + \mathrm{d}L(t),$$

where $L(t)$ is a non-decreasing process that increases only when $X(t) = 0$. The additional term $\mathrm{d}L(t)$ prevents the process from crossing the boundary. The generator of reflecting Brownian motion is

$$\mathcal{L}f = Df'',$$

with Neumann boundary condition

$$f'(0) = 0.$$

Thus, the Neumann condition derived earlier for the backward equation corresponds precisely to this reflected diffusion. The reflection rule (12.11) in the Euler–Maruyama scheme provides a discrete approximation of this process.

Algorithm 12.1: Euler–Maruyama (12.10) with reflective boundary at $x = 0$

Set $X_0 = x_0 > 0$ and $\Delta t = T/N$.
For $n = 0, \dots, N-1$
- **(a)** Generate a random number $\xi \sim \mathcal{N}(0,1)$.
- **(b)** Set $X_{n+1} = X_n + a(X_n)\Delta t + \sqrt{b(X_n)\Delta t}\,\xi$.
- **(c)** If $X_{n+1} < 0$, set $X_{n+1} \leftarrow -X_{n+1}$.

end

Absorbing boundary Still with $\Omega = [0, \infty)$, now assume that $x = 0$ is an absorbing boundary. If after one step of (12.10), $X_{n+1} < 0$, this means that the particle has crossed $x = 0$ since the process is continuous. Hence, the trajectory must be terminated.

But even if $X_n > 0$ and $X_{n+1} > 0$, there remains a non-zero probability that the continuous trajectory crossed the boundary during the time step (see Figure 12.1). For Brownian motion with diffusion coefficient D ($b = 2D$ constant), this crossing probability can be computed explicitly as

$$\mathbb{P}(X(t) = 0 \text{ for some } t \in (t_n, t_{n+1}) | X(t_n) = X_n, X(t_{n+1}) = X_{n+1}) = \exp\left(-\frac{X_n X_{n+1}}{D\Delta t}\right).$$

It is known as the Andrews–Bray correction [1]. We see that the further away from the boundary X_n, X_{n+1} are, the less likely it is that the trajectory hit the boundary in the Δt time interval. On the other hand, a larger diffusion coefficient D (more noise) makes having hit the boundary more likely. In practice, choosing Δt sufficiently small reduces the importance of this correction. The pseudocode is given in Algorithm 12.2.

Algorithm 12.2: Euler–Maruyama (12.10) with absorbing boundary at $x = 0$

Set $X_0 = x_0 > 0$ and $\Delta t = T/N$.
For $n = 0, \dots, N-1$
- **(a)** Generate a random number $\xi \sim \mathcal{N}(0,1)$ (the standard normal distribution).
- **(b)** Set $X_{n+1} = X_n + a(X_n)\Delta t + \sqrt{b(X_n)\Delta t}\,\xi$.
- **(c)** if $X_{n+1} < 0$, terminate the trajectory.
  else generate a random number $r \sim U(0,1)$
    if $r < \exp[-X_n X_{n+1}/(D\Delta t)]$, terminate the trajectory
  end

end

12.4 First passage times

An important characteristic of a diffusion process is the first passage time for a particle to reach a given target or boundary or become extinct. These types of problems are prevalent in biological applications of stochastic processes, such as:

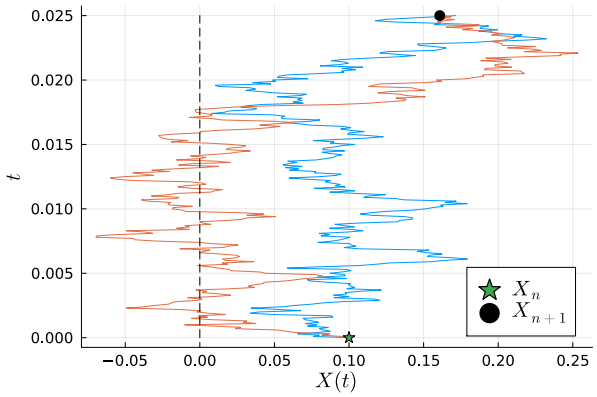

Figure 12.1: Two trajectories of (12.10) with $a = 0, b = 2$. With an absorbing boundary at $x = 0$, the red trajectory should have been terminated.

- Channel transport (ion channels, porins in bacteria): how long does it take to reach the end of the channel?
- Translocations of long chains, such as DNA and RNA, through nanopores (as it occurs in the sequencing of DNA/RNA). What is the expected time to finish the translocation?
- Receptor binding (e.g., receptor signalling and viral cell entry). How long does it take to find the receptor and bind?
- Single-cell growth and division: what is the mean time to reach a specific size and divide?

The mean first passage time (MFPT) to a given event can be calculated from the backward Kolmogorov equation. To see this, consider (12.1) in $\Omega = (0, L)$ with $x = 0$ reflecting and $x = L$ absorbing. Given $X(0) = y \in \Omega$, we define the *first passage time*

$$T(y) = \inf\{t \geq 0 : X(t) = L\}. \tag{12.12}$$

Note that $T(y)$ is a random variable. The *survival probability* $Q(y, t)$, or the probability that the particle has not left Ω by time t, is

$$Q(y,t) = \int_0^L p(x,t|y,0)\mathrm{d}x = \mathbb{P}(T(y) > t). \tag{12.13}$$

Since y is in Ω, we have that $Q(y, 0) = \int \delta(x - y)\mathrm{d}x = 1$. As time progresses, the absorbing boundary condition implies there is a "probability leak" through $x = L$, and that $Q(y, t)$ will decrease.

Writing the (time-homogeneous) backward Kolmogorov equation for p, together with the corresponding boundary conditions (see §12.2),

$$\frac{\partial p}{\partial t}(x,t|y,0) = \mathcal{L}_y p(x,t|y,0), \qquad \frac{\partial p}{\partial y}(x,t|y,0)|_{y=0} = 0, \qquad p(x,y|L,t) = 0,$$

and integrating with respect to x, we see that the survival probability Q satisfies

$$\frac{\partial Q}{\partial t} = \mathcal{L}_y Q, \tag{12.14}$$

$$\frac{\partial Q}{\partial y} = 0, \qquad \text{on } y = 0, \tag{12.15}$$

$$Q = 0, \qquad \text{on } y = L, \tag{12.16}$$

$$Q = 1, \qquad \text{at } t = 0. \tag{12.17}$$

Definition 12.1 (Mean first passage time)**.** The mean first passage time (MFPT) $\tau(y)$ is

$$\tau(y) = \mathbb{E}T(y). \tag{12.18}$$

Using

$$\tau(y) = \int_0^\infty Q(y,t)\mathrm{d}t,$$

and integrating (12.14) in time, we have that

$$\mathcal{L}_y \tau(y) = \int_0^\infty \frac{\partial Q}{\partial t}\mathrm{d}t = Q(\infty) - Q(0) = -1, \tag{12.19}$$

where we have used that, since absorption at $x = L$ occurs with probability one in this setting, we have $Q(\infty) = 0$. The boundary conditions are

$$\tau'(0) = 0, \qquad \tau(L) = 0. \tag{12.20}$$

Example 12.1 (Brownian motion)**.** Consider a standard Brownian motion, $a(x) = 0$ and $b(x) = 1$, in $\Omega = (0, L)$. Then

$$\frac{1}{2}\tau''(y) = -1.$$

- Reflecting at $x = 0$, absorbing at $x = L$: this means $\tau'(0) = \tau(L) = 0$, and

$$\tau(y) = L^2 - y^2.$$

- Absorbing at both ends: in this case we define the first passage time $T(y)$ as

$$T(y) = \inf\{t \geq 0 : X(t) \in \partial\Omega\},$$

and the boundary conditions on $\tau = \mathbb{E}T$ become $\tau(0) = \tau(L) = 0$. This results in

$$\tau(y) = y(L - y).$$

The derivation above may be repeated for more general boundaries in $\mathbb{R}^d$. We give the result in the box below. Note this is the continuous analogue of the extinction times (5.2) in finite Markov chains.

Mean first passage time in $\Omega \subset \mathbb{R}^d$

Let $X(t)$ be a homogeneous diffusion process in $\Omega \subset \mathbb{R}^d$

$$\mathrm{d}\boldsymbol{X}(t) = \boldsymbol{a}(\boldsymbol{X}(t))\mathrm{d}t + \boldsymbol{\sigma}(\boldsymbol{X}(t))\mathrm{d}\mathbf{W}(t),$$

with $\partial\Omega = \partial\Omega_A \cup \partial\Omega_R$ (part of the boundary is absorbing, and part of the boundary is reflecting). Then the MFPT $\tau(y)$ to leave Ω through $\partial\Omega_A$ having started at $\boldsymbol{y} \in \Omega$, $\boldsymbol{X}(0) = \boldsymbol{y}$, satisfies the following problem

$$\mathcal{L}\tau = -1, \qquad y \in \Omega, \tag{12.21}$$

$$\tau = 0, \qquad y \in \partial\Omega_A, \tag{12.22}$$

$$(\mathbf{B} \cdot \nabla\tau) \cdot \boldsymbol{n} = 0, \qquad y \in \partial\Omega_R, \tag{12.23}$$

where $\boldsymbol{n}$ is the outward normal on $\partial\Omega$, $\mathbf{B}(\boldsymbol{x}, t) = \boldsymbol{\sigma}\boldsymbol{\sigma}^\top$, and the backward Kolmogorov operator is given in (11.14). The boundary condition (12.23) is $\sum_{i,j} n_i B_{ij} \partial_{y_j} \tau = 0$.

Remark 12.1 (Large exit times and rare-events simulation)**.** When first-exit times are very large, naive Monte Carlo simulation becomes computationally inefficient. If trajectories are stopped at a finite time $T_{\max}$, the resulting estimator of the mean first passage time is biased downward, since paths with $T > T_{\max}$ are right-censored and contribute too little to the average. More systematic approaches either account for censoring at the level of the survival probability $Q(y, t)$, or modify the dynamics so that exits occur more frequently and compensate by appropriate reweighting (for example, via importance sampling or splitting methods). In metastable regimes, where the exit time is approximately exponentially distributed, the mean first passage time can alternatively be recovered from the effective escape rate, which is closely related to the principal eigenvalue of the generator in the domain.

Lecture 13: Stationary densities and detailed balance

In this lecture, we study stationary solutions of the Fokker–Planck equation, give explicit formulae in one dimension, and discuss the important subclass of gradient systems and their relation to detailed balance. The goal is to provide practical tools for identifying invariant measures, to explain when an invariant density exists and is normalisable, and to relate these properties to potential landscapes and metastability phenomena encountered in the course.

13.1 Stationary density

Recall from Section 11.5 that the multidimensional Ito diffusion in $\mathbb{R}^d$

$$\mathrm{d}\boldsymbol{X}(t) = \boldsymbol{a}(\boldsymbol{X}(t))\,\mathrm{d}t + \boldsymbol{\sigma}(\boldsymbol{X}(t))\,\mathrm{d}\mathbf{W}(t) \tag{13.1}$$

has generator

$$\mathcal{L}f(\boldsymbol{x}) = \boldsymbol{a}(\boldsymbol{x})\cdot\nabla f(\boldsymbol{x}) + \frac{1}{2}\sum_{i,j} B_{ij}(\boldsymbol{x})\frac{\partial^2 f}{\partial x_i \partial x_j}(\boldsymbol{x}), \tag{13.2}$$

where $\mathbf{B}(\boldsymbol{x}) = (B_{ij}) = \boldsymbol{\sigma}(\boldsymbol{x})\boldsymbol{\sigma}(\boldsymbol{x})^{\top}$. If $p(\boldsymbol{x},t)$ denotes the density of $\boldsymbol{X}(t)$, then the Fokker–Planck equation reads

$$\frac{\partial}{\partial t}p(\boldsymbol{x},t) = \mathcal{L}^{*}p(\boldsymbol{x},t) = -\nabla\cdot\mathbf{J}(\boldsymbol{x},t), \tag{13.3}$$

where $\mathbf{J}(\boldsymbol{x},t)$ is the probability flux whose components are

$$J_i(\boldsymbol{x},t) = \boldsymbol{a}_i(\boldsymbol{x})p(\boldsymbol{x},t) - \frac{1}{2}\sum_{j}\frac{\partial}{\partial x_j}\big(\mathbf{B}_{ij}(\boldsymbol{x})p(\boldsymbol{x},t)\big). \tag{13.4}$$

An *invariant or stationary density* $p_\infty(\boldsymbol{x})$ is a time-independent solution of (13.3), that is,

$$0 = \mathcal{L}^{*}p_\infty(\boldsymbol{x}), \tag{13.5}$$

or equivalently

$$\nabla\cdot\mathbf{J}_\infty(\boldsymbol{x}) = 0. \tag{13.6}$$

Thus, stationarity means that the stationary flux is divergence-free.

13.2 Detailed balance and reversibility

We now strengthen the notion of invariance. A stationary density p_∞ is called *reversible* if the generator is symmetric in $L^2(p_\infty)$, that is, for all smooth compactly supported test functions f and g,

$$\int_{\mathbb{R}^d} f(\boldsymbol{x})\,\mathcal{L}g(\boldsymbol{x})\,p_\infty(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \int_{\mathbb{R}^d} g(\boldsymbol{x})\,\mathcal{L}f(\boldsymbol{x})\,p_\infty(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}. \tag{13.7}$$

This condition is the diffusion analogue of *detailed balance* for continuous-time Markov chains. For a finite-state Markov chain with generator matrix $Q = (q_{ij})$ and stationary distribution ϕ_i, the detailed balance condition reads

$$\phi_i q_{ij} = \phi_j q_{ji} \quad \text{for all } i,j.$$

It expresses the fact that, at stationarity, the probability flow from state i to state j is exactly balanced by the reverse flow from j to i. Equation (13.7) is the continuous-state counterpart of this local balance condition.

Integrating by parts shows that (13.7) is equivalent to

$$\mathbf{J}_\infty(\boldsymbol{x}) \equiv 0. \tag{13.8}$$

Thus, detailed balance is equivalent to the absence of stationary probability currents. Clearly, (13.8) is stronger than (13.6) (d equations vs. 1 equation, except for one dimension, in which they are equivalent when taking the boundary condition into account).

If $p_\infty(\boldsymbol{x}) > 0$ and $\mathbf{B}(\boldsymbol{x})$ is invertible, the identity (13.8) can be rearranged to give

$$\boldsymbol{a}_i(\boldsymbol{x}) = \frac{1}{2}\sum_j \frac{\partial}{\partial x_j}\mathbf{B}_{ij}(\boldsymbol{x}) + \frac{1}{2}\sum_j \mathbf{B}_{ij}(\boldsymbol{x})\frac{\partial}{\partial x_j}\log p_\infty(\boldsymbol{x}). \tag{13.9}$$

Equation (13.9) characterises reversible diffusions with invariant density p_∞.

The flux $\mathbf{J}_\infty(\boldsymbol{x})$ defined in (13.4) measures the local net transport of probability at stationarity. Its ith component gives the steady current of probability in the x_i-direction. When $\mathbf{J}_\infty(\boldsymbol{x}) \equiv 0$, there is no net transport anywhere, so the forward and backward microscopic transitions balance locally. This characterises *equilibrium (reversible) systems*. If instead $J_\infty(\boldsymbol{x})$ is nonzero but divergence-free, probability circulates persistently even though the density is time-independent. This corresponds to a *nonequilibrium steady state*.

Now consider the one-dimensional setting $d = 1$ with drift by $a(x)$ and diffusion coefficient $b(x)$, see (12.1). The stationary Fokker-Planck equation (13.5) is

$$0 = -\frac{\mathrm{d}}{\mathrm{d}x}\big(a(x)\,p_\infty(x)\big) + \frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}x^2}\big(b(x)\,p_\infty(x)\big) = \frac{\mathrm{d}}{\mathrm{d}x}J_\infty(x).$$

Hence, the stationary flux $J_\infty(x)$ is constant. Under the zero-flux (reversible) condition $J = 0$, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}x}\big(b(x)\,p_\infty(x)\big) = 2\,a(x)\,p_\infty(x).$$

We rewrite it as

$$\frac{\mathrm{d}}{\mathrm{d}x}\log\big(b(x)\,p_\infty(x)\big) = \frac{2\,a(x)}{b(x)}.$$

Integrating and rearranging gives the explicit formula

$$p_\infty(x) = \frac{C}{b(x)}\exp\left(2\int^x \frac{a(y)}{b(y)}\,\mathrm{d}y\right), \tag{13.10}$$

where C is the normalisation constant provided the right-hand side is integrable.

13.3 Gradient systems and Gibbs measures

A particularly important special case arises when diffusion is constant and isotropic. Suppose $b(x) = 2D$ with constant $D > 0$. Then (13.10) simplifies to

$$p_\infty(x) = C\exp\left(\int^x \frac{a(y)}{D}\,\mathrm{d}y\right).$$

If the drift is of gradient form, $a(x) = -V'(x)$ for some potential function V, then

$$p_\infty(x) = C \exp\left(-\frac{V(x)}{D}\right). \tag{13.11}$$

In the multidimensional setting, the same structure appears. If $\mathbf{B}(\boldsymbol{x}) = 2D\mathbb{I}$ and $\boldsymbol{a}(\boldsymbol{x}) = -\nabla V(\boldsymbol{x})$, then equation (13.8) implies

$$p_\infty(\boldsymbol{x}) = C \exp\left(-\frac{V(\boldsymbol{x})}{D}\right). \tag{13.12}$$

Equation (13.12) is called a *Gibbs measure* (or Boltzmann distribution). Here, V plays the role of an energy function and D acts as a temperature parameter. Such densities arise naturally in statistical mechanics as equilibrium distributions of systems at temperature D. Note that for (13.12) to make sense, it must be normalizable. This can be made precise with the following definition.

Definition 13.1 (Confining potential)**.** A function $V(\boldsymbol{x}) : \mathbb{R}^d \to \mathbb{R}$ is called a confining potential if it satisfies the following two conditions:

$$\lim_{|\boldsymbol{x}|\to\infty} V(\boldsymbol{x}) = +\infty, \qquad \int_{\mathbb{R}^d} e^{-cV(\boldsymbol{x})} \mathrm{d}\boldsymbol{x} < +\infty,$$

for all $c > 0$.

The function V is sometimes called the *potential landscape*. When the noise level D is small, p_∞ concentrates near the minima of V. If the landscape contains multiple wells separated by barriers, the process exhibits metastability: it spends long periods near a local minimum before making a rare transition to another well.

To make this more precise, let T denote the first exit time from a potential well,

$$T = \inf\{t > 0 : \boldsymbol{X}(t) \notin \mathcal{W}\},$$

where $\mathcal{W}$ is a neighbourhood of a local minimum of V. The quantity of interest is the mean first-passage time (MFPT),

$$\tau = \mathbb{E}[T].$$

In the small-noise regime $D \ll 1$, one finds that (see Sheet 4)

$$\tau \sim C \exp\left(\frac{\Delta V}{D}\right),$$

where ΔV is the height of the energy barrier separating neighbouring wells. Thus, even a moderate decrease in the noise strength D leads to an exponential increase in the typical residence time within a well.

13.4 Ergodicity

So far we have characterised stationary densities through the Fokker–Planck equation

$$\mathcal{L}^* p_\infty = 0.$$

However, the existence of a stationary density does not automatically imply that the process converges to it from arbitrary initial data.

Definition 13.2 (Ergodic diffusion)**.** A diffusion process $\boldsymbol{X}(t)$ is called ergodic if

(i) a stationary density p_∞ exists,

(ii) it is unique, and

(iii) for any initial density p_0, the solution of the Fokker–Planck equation satisfies

$$p(\boldsymbol{x}, t) \longrightarrow p_\infty(\boldsymbol{x}) \quad \text{as } t \to \infty.$$

In other words, ergodicity means that the long-time behaviour of the process is independent of its initial condition.

Example 13.1 (Ergodic processes)**.** The Ornstein–Uhlenbeck process

$$\mathrm{d}X(t) = -\alpha X(t)\,\mathrm{d}t + \sqrt{2D}\,\mathrm{d}W(t), \qquad \alpha > 0, \tag{13.13}$$

is ergodic. Its stationary density is

$$p_\infty(x) = \sqrt{\frac{\alpha}{2\pi D}} \exp\left(-\frac{\alpha x^2}{2D}\right), \tag{13.14}$$

which fits the Gibbs structure (13.11) with potential $V(x) = \alpha x^2/2$ and diffusion D. In contrast, Brownian motion on $\mathbb{R}$ has no stationary density and is therefore not ergodic (as one has that $p \to 0$ as $t \to \infty$). Note that the situation is different if we consider Brownian motion in a bounded domain with appropriate boundary conditions.

Ergodicity has an important computational consequence. If $\boldsymbol{X}(t)$ is ergodic and f is an observable, then

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T f(\boldsymbol{X}(s))\,\mathrm{d}s = \int_{\mathbb{R}^d} f(\boldsymbol{x})\, p_\infty(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}. \tag{13.15}$$

Thus, expectations with respect to the stationary density p_∞ can be computed as long-time averages along a single trajectory. In practice, this means that one may either simulate a single trajectory for a sufficiently long time or average over many independent shorter simulations. This principle forms the basis of many sampling methods, including Markov Chain Monte Carlo (MCMC), which generate sequences whose long-time behaviour approximates the stationary density.

13.5 Numerical simulation of the stationary state

Suppose we wish to estimate p_∞ numerically for the OU process (13.13). We may use the Euler–Maruyama discretisation

$$X_{n+1} = X_n - \alpha X_n \Delta t + \sqrt{2D\Delta t}\,\xi, \qquad \xi \sim \mathcal{N}(0,1). \tag{13.16}$$

As discussed above, there are two natural approaches.

Ensemble averaging. We simulate many independent trajectories up to a large time T and store only the final values.

Algorithm 13.1: Stationary density estimation via ensemble averaging

Choose final time T, timestep Δt with $N\Delta t = T$, and number of paths M.
for $s = 1$ to M
    Generate initial condition X_0.
    for $n = 0$ to $N - 1$
        (a) Generate $\xi \sim \mathcal{N}(0, 1)$.
        (b) Set $X_{n+1} = X_n - \alpha X_n \Delta t + \sqrt{2D\Delta t}\,\xi$.
    Store X_N in histogram.
end

Time averaging. Alternatively, we simulate a single long trajectory and store all values after transients have decayed.

Algorithm 13.2: Stationary density estimation via time averaging

Choose final time T and timestep Δt with $N\Delta t = T$.
Generate initial condition X_0.
for $n = 0$ to $N - 1$
    (a) Generate $\xi \sim \mathcal{N}(0, 1)$.
    (b) Set $X_{n+1} = X_n - \alpha X_n \Delta t + \sqrt{2D\Delta t}\,\xi$.
    (c) Store X_{n+1} in histogram.
end

For ergodic processes, both approaches converge to the same stationary density (13.14) provided that:

- the final time T is large enough,
- the timestep Δt is sufficiently small,
- the number of samples is sufficiently large.

Using the computational approximation for X_{n+1} (13.16) one may show that, as $n \to \infty$,

$$\mathbb{E}X_{n+1} = (1 - \alpha\Delta t)\mathbb{E}X_n \to 0,$$
$$\mathbb{E}X_{n+1}^2 = (1 - \alpha\Delta t)^2\mathbb{E}X_n^2 + 2D\Delta t \to \frac{2D}{\alpha(2 - \alpha\Delta t)},$$

provided that $\alpha\Delta t < 2$ and $\alpha > 0$. In this case, the numerical steady state is $\mathcal{N}(0, \sigma_{\Delta t}^2)$, with variance

$$\sigma_{\Delta t}^2 = \frac{2D}{\alpha(2 - \alpha\Delta t)}. \tag{13.17}$$

This approaches the exact value D/α (see (13.14)) as $\Delta t \to 0$, but we expect an error for finite Δt. We test this in Figure 13.1, where we take $\alpha = 0.75, D = 0.5$ and $X_0 \sim U([0, 3])$ for $\Delta t = 1$ (numerical example mostly follows from [7, §14.3]). Hence the constraint $\alpha\Delta t < 2$ is satisfied, and both the SDE and the EM SSA have normally distributed stationary densities with mean

zero and variances D/α and $1.6D/\alpha$, respectively. With such large Δt, we observe noticeable differences between the exact stationary density of the process, (13.14) (shown as a solid black line), and that of the Euler–Maruyama (shown as a black dashed line). The histograms obtained from running Algorithm 13.1 with $M = 10^4$ samples at times $T = 0, 1, 6$ are also shown.

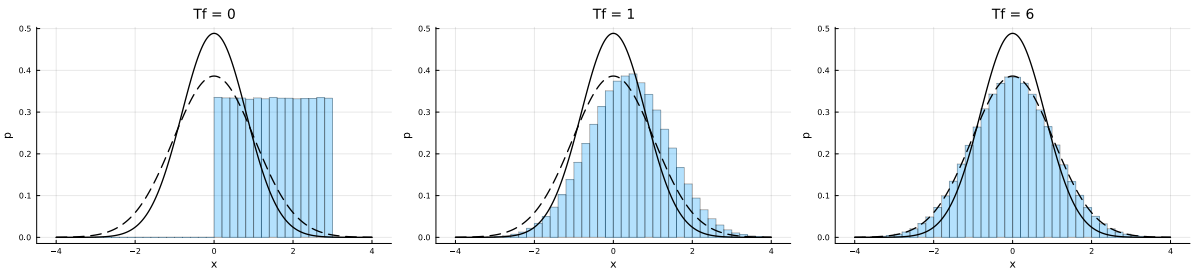

Figure 13.1: Histograms to estimate p_∞ of the Ornstein–Uhlenbeck process using Algorithm 13.1 with $\Delta t = 1$ and $M = 10^4$ at times $T = 0, 1, 6$, with initial condition X_0 uniform over $(1, 3)$. Solid curve: stationary probability density (13.14). Dashed curve: theoretical stationary density of the EM method with variance (13.17).

13.6 The Metropolis–Hastings algorithm

The time-averaging method in Algorithm 13.2 provides a way of sampling the stationary density p_∞ by simulating the underlying SDE. It generates a Markov chain whose long-time distribution is p_∞, and is thus an example of a Markov Chain Monte Carlo (MCMC) method. More generally, one may construct Markov chains that have a prescribed target density $p_\infty(\boldsymbol{x})$ as their stationary distribution without explicitly simulating the original diffusion. One of the most important and widely used such methods is the Metropolis–Hastings algorithm.

To fix ideas, consider again the Ornstein–Uhlenbeck process (13.13), whose stationary density (13.14) can be written as

$$p_\infty(x) = C\pi(x), \qquad \pi(x) = \exp(-H(x)), \tag{13.18}$$

where

$$H(x) = \frac{\alpha x^2}{2D},$$

and C is the normalisation constant. The Metropolis–Hastings algorithm generates a sequence $X_0, X_1, X_2, \ldots$ that samples from p_∞ as follows. Given X_i, propose a candidate

$$y = X_i + \delta\xi, \qquad \xi \sim \mathcal{N}(0, 1),$$

where $\delta > 0$ controls the proposal size. The proposed move is accepted with probability

$$p_{\text{acc}} = \min\big[1, \exp(-\Delta H)\big], \qquad \Delta H = H(y) - H(X_i). \tag{13.19}$$

That is, if the candidate y has a lower energy than the previous step, the move is always accepted, whereas if it leads to a higher energy, the probability of accepting it depends on the

energy difference (the higher ΔH, the less likely the move is to be accepted). If the move is accepted, set $X_{i+1} = y$; otherwise set $X_{i+1} = X_i$.

Algorithm 13.3: Metropolis–Hastings algorithm for sampling p_∞

Set initial value X_0, number of steps M, and proposal scale δ.
for $i = 0$ to $M - 1$
(a) Generate $\xi \sim \mathcal{N}(0, 1)$ and $r \sim U(0, 1)$.
(b) Set $y = X_i + \delta\xi$ and compute $\Delta H = H(y) - H(X_i)$.
(c) If $r < \exp(-\Delta H)$, set $X_{i+1} = y$ (accept); otherwise set $X_{i+1} = X_i$ (reject).
end

A key feature of the algorithm is that it does not require knowledge of the normalisation constant C in (13.18). Only the ratio $\pi(y)/\pi(X_i)$ appears in (13.19). This makes the method particularly useful in high-dimensional problems, where computing normalisation constants is typically infeasible.

To see that the Metropolis–Hastings algorithm samples from p_∞, observe that the resulting Markov chain satisfies detailed balance with respect to p_∞. If $q(x, y)$ denotes the proposal density (here symmetric, $q(x, y) = q(y, x)$ with $q(x, \cdot)$ a normal distribution), then the acceptance probability (13.19) ensures that

$$p_\infty(x)\, q(x, y)\, p_{\text{acc}}(x, y) = p_\infty(y)\, q(y, x)\, p_{\text{acc}}(y, x). \tag{13.20}$$

This is the discrete analogue of the zero-flux condition $J_\infty(\boldsymbol{x}) = 0$ in (13.8), expressing equality of forward and backward probability flows. Consequently, p_∞ is the stationary distribution of the chain.

The efficiency of the method depends strongly on the choice of the proposal scale δ. If δ is too small, most proposals are accepted but the chain explores the state space slowly. If δ is too large, most proposals are rejected and the chain barely moves. For high-dimensional Gaussian targets, the optimal acceptance rate is approximately 0.234, although in practice acceptance rates in the range 0.1–0.4 are often satisfactory.

Unlike the time-discretised SDE simulation in Algorithm 13.2, the Metropolis–Hastings chain does not represent the physical time evolution of the diffusion process; it is a sampling device designed to have p_∞ as its stationary distribution.

Although each Metropolis–Hastings step requires two random numbers (one for the proposal and one for the acceptance test), this does not significantly increase the computational cost compared to Euler–Maruyama. The relevant notion of efficiency is not the number of random numbers per iteration, but how rapidly the method produces effectively independent samples. Euler–Maruyama requires a sufficiently small timestep Δt to control discretisation error, which can lead to many highly correlated steps. In contrast, Metropolis–Hastings avoids time-discretisation bias and, with an appropriate choice of δ, can explore the state space efficiently despite the acceptance test.

Tasks

Exercise 13.1. Show that $\mathbf{J}_\infty = 0$ if and only if (13.7) holds, that is

$$\langle \mathcal{L}g, f\rangle_{p_\infty} = \langle g, \mathcal{L}^* f\rangle_{p_\infty}$$

Exercise 13.2. Show that the detailed balance condition (13.20) holds for the Metropolis–Hastings transition probability from x to y $P(x,y) = q(x,y)p_{\text{acc}}(x,y)$. [*Hint: write* $p_{acc}(x,y) = \min(1, R(x,y))$ *with* $R(x,y) = \frac{p_\infty(y)q(y,x)}{p_\infty(x)q(x,y)}$ *and split into two cases depending on* $R \leq 1$ *or* $R > 1$.

Example codes

`Lect13.ipynb` contains simulations to sample the stationary distribution of the Ornstein–Uhlenbeck process, via ensemble averaging, time averaging, and the Metropolis–Hastings.

Lecture 14: Second-order models

In Lecture 9, we saw how we may model the position $X(t)$ of a diffusive particle by Brownian motion. In this lecture, we will consider alternative models for diffusion, which reduce to Brownian motion in certain limits. In particular, we will consider so-called second-order models where, in addition to the particle's position $X(t)$, we also consider its velocity $V(t)$.

These models provide a richer description of motion. We will see that, under appropriate conditions, they reduce to Brownian motion, but their microscopic structure is fundamentally different.

14.1 Velocity-jump process

We begin with the run-and-tumble model, used to describe the motion of bacteria such as *E. coli*. The particle position is continuous, but the velocity undergoes random jumps. It is this an example of a jump-diffusion process (see Remark 11.2).

The run-and-tumble model arises in modelling the motion of organisms with flagellum (e.g., *E. coli*). To search for food or escape an unfavourable environment, *E. coli* alternates between a more or less linear motion called a *run* and a highly erratic motion called a *tumble*, after which the cell reorients itself [12]. In the simplest model, one assumes that bacteria move at a constant speed during a run, leading to the forward Kolmogorov equation

$$\frac{\partial p}{\partial t}(\boldsymbol{x}, \mathbf{v}, t) + \mathbf{v} \cdot \nabla_{\boldsymbol{x}} p = -\lambda p + \lambda \int_V W(\mathbf{v}|\mathbf{w}, t) p(\boldsymbol{x}, \mathbf{w}, t) \mathrm{d}\mathbf{w}, \tag{14.1}$$

in state space $(\boldsymbol{x}, \mathbf{v}) \in \mathbb{R}^2 \times V$. The turning kernel $W(\mathbf{v}|\mathbf{w}, t)$ describes the probability of choosing $\mathbf{v} \in V$ as the new velocity, given that the old velocity was $\mathbf{w} \in V$. It is common to consider the state space for velocity $V = s\mathbb{S}$ with s constant and an unbiased turning kernel of the form $W = 1/|V|$.

One-dimensional model. Let us consider the one-dimensional version of (14.1). A particle with position $X(t) \in \mathbb{R}$ moves with velocity

$$V(t) \in \{-s, s\}, \qquad s > 0,$$

according to

$$\mathrm{d}X(t) = V(t)\, \mathrm{d}t. \tag{14.2}$$

The velocity switches sign at rate $\lambda > 0$. The velocity process $V(t)$ is called a *telegraph process*. The joint process $(X(t), V(t))$ is Markovian, but $X(t)$ alone is not: knowledge of the current direction influences future motion.

From the perspective of the first half of the course, the velocity process is simply a two-state continuous-time Markov chain with transition

$$V \longrightarrow -V \quad \text{at rate } \lambda.$$

Equivalently, this is a one-species reaction network with reaction $V \to -V$ and propensity $\alpha(V) = \lambda$. Hence the waiting times between velocity flips are exponentially distributed with parameter λ, and successive waiting times are independent.

Forward equations. We define by $p^+(x,t)$ the probability density of the particle moving right, and by $p^-(x,t)$ the probability density of the particle moving left:

$$p^\pm(x,t)\,\mathrm{d}x = \mathbb{P}\{X(t) \in [x, x+\mathrm{d}x),\, V(t) = \pm s\}.$$

Then the densities $p^\pm$ satisfy the following system of forward Kolmogorov equations

$$\partial_t p^+ + s\partial_x p^+ = \lambda p^- - \lambda p^+, \tag{14.3a}$$
$$\partial_t p^- - s\partial_x p^- = \lambda p^+ - \lambda p^-. \tag{14.3b}$$

We may complement (14.3) with initial conditions

$$p^+(x,0) = p_0^+(x), \qquad p^-(x,0) = p_0^-(x).$$

For example, $p_0^+(x) = \delta(x - x_0)$ and $p_0^-(x) = 0$ would correspond to initialising the system deterministically by setting the particle at x_0 and velocity s at $t = 0$.

Defining the total density and the probability flux as

$$p = p^+ + p^-, \qquad J = s(p^+ - p^-),$$

and adding and substracting equations (14.2), we obtain

$$\partial_t p + \partial_x J = 0, \tag{14.4}$$
$$\partial_t J + s^2 \partial_x p = -2\lambda J. \tag{14.5}$$

Eliminating J yields the *telegraph equation*

$$\frac{1}{2\lambda}\partial_t^2 p + \partial_t p = \frac{s^2}{2\lambda}\partial_x^2 p. \tag{14.6}$$

This is a hyperbolic PDE. Unlike the diffusion equation, it incorporates finite propagation speed.

Long-time behaviour. For times much larger than $1/\lambda$, the motion resembles diffusion with effective diffusivity

$$D = \frac{s^2}{2\lambda}. \tag{14.7}$$

We will revisit this connection more systematically when discussing diffusion limits in the next lecture.

Figure (14.1) shows a simulation of the two-dimensional version of the velocity-jump process (14.1), with $\mathbf{V}(t) \in s \cdot \mathbb{S}^1 = \{\mathbf{v} \in \mathbb{R}^2 : \|\mathbf{v}\| = s\}$ and a uniform turning kernel $W(\mathbf{v}|\mathbf{w}, t) \equiv 1/2\pi$. The left panel corresponds to a short trajectory ($T_f = 26.39$) with 20 velocity jumps. The right panel shows a trajectory with 10^4 velocity jumps, which looks much like a Brownian trajectory. This is consistent with (14.7).

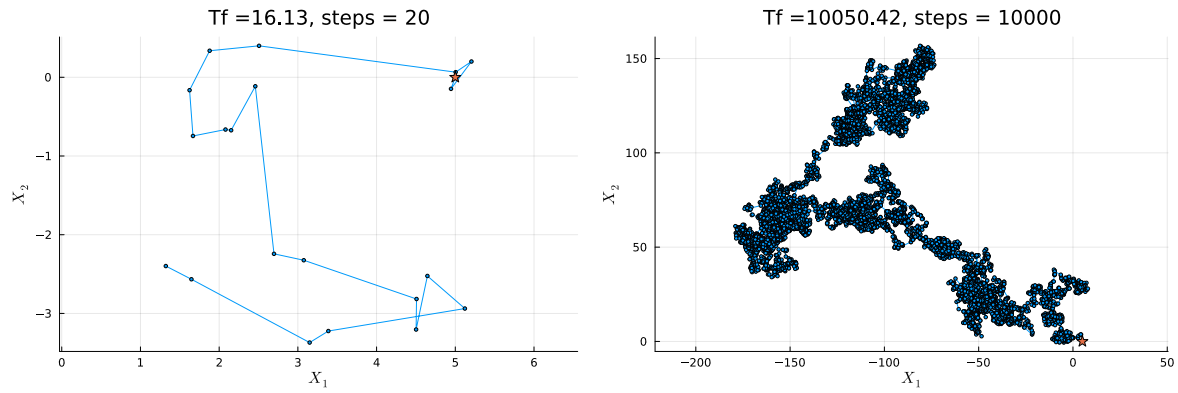

Figure 14.1: Simulation of the two-dimensional version of the velocity-jump process (14.1), with $\mathbf{V}(t) \in s \cdot \mathbb{S}^1 = \{\mathbf{v} \in \mathbb{R}^2 : \|\mathbf{v}\| = s\}$ and a uniform turning kernel $W(\mathbf{v}|\mathbf{w}, t) \equiv 1/2\pi$. The trajectories are simulated using the Gillespie SSA (details in forthcoming lecture) and parameters $\lambda = s = 1$. Left: 20 steps of the Gillespie SSA, reaching a final time $T_f = 26.39$. Right: 10^4 steps of the Gillespie SSA, reaching $T_f = 10050.42$.

Chemotaxis. The turning rate in a velocity-jump process may depend on position and velocity. A classical model for bacterial chemotaxis assumes that the turning frequency depends on the local gradient of a chemoattractant concentration $c(x)$. In one dimension, a simple choice is

$$\lambda(x, v) = \lambda_0 - \operatorname{sign}(v)\, \lambda_1 \frac{\mathrm{d}c}{\mathrm{d}x}, \tag{14.8}$$

where $\lambda_0, \lambda_1 > 0$ are constants. If $c'(x) > 0$, then a bacterium moving up the gradient turns less frequently, while one moving down the gradient turns more frequently. This creates a net drift toward higher concentrations of c.

In $\mathbb{R}^d$, an analogous bias can be introduced by letting the turning kernel depend on $\nabla c(\boldsymbol{x})$. Under suitable scaling assumptions (discussed in the next lecture), the probability density $p(x, t)$ of particle positions satisfies the macroscopic equation

$$\frac{\partial p}{\partial t} = \nabla \cdot \big(D \nabla p - \chi p \nabla c\big), \qquad x \in \mathbb{R}^d, \tag{14.9}$$

where D is the effective diffusion coefficient and χ is the chemotactic sensitivity. Both parameters depend on the microscopic speed s and turning rate λ.

Equation (14.9) is known as the *Keller–Segel equation*. It provides a macroscopic description of biased cell movement arising from the underlying velocity-jump dynamics. The parameter χ measures the strength of the chemotactic drift relative to diffusion. If χ is sufficiently small, diffusion dominates and solutions remain smooth and dispersed. However, when χ exceeds a critical threshold (depending on the dimension and total mass), chemotactic attraction can overcome diffusion and lead to aggregation of the population. In certain settings, this aggregation manifests as finite-time blow-up of solutions.

14.2 Langevin's equation

We now consider a different second-order model, originally proposed by Langevin to describe Brownian motion. Unlike the velocity-jump process, here the velocity evolves continuously under the influence of friction and random forcing.

Microscopic model. Consider a particle of mass m subject to Newton's law

$$\frac{\mathrm{d}X(t)}{\mathrm{d}t} = V(t), \tag{14.10a}$$

$$m\frac{\mathrm{d}V(t)}{\mathrm{d}t} = F(t), \tag{14.10b}$$

where the force $F(t)$ is made up of two components:

$$F(t) = -\gamma V(t) + R(t).$$

The first term on the right-hand side is the frictional force with drag coefficient γ, while $R(t)$ is a "random" force describing the rapidly fluctuating interactions between the particle and solvent molecules. Assuming the particle is immersed in a thermal bath, we model $R(t)$ as white noise satisfying

$$\mathbb{E}[R(t)R(s)] = 2k_BT\gamma\delta(t-s). \tag{14.11}$$

This relation expresses the fact that in thermal equilibrium friction and fluctuations are inseparably linked: the strength of the noise is determined by the temperature and the drag coefficient. Hence we can identify the equation for $V(t)$ (14.10b) as the Ornstein–Uhlenbeck SDE (13.13):

$$\mathrm{d}V(t) = -\frac{\gamma}{m}V(t)\mathrm{d}t + \sqrt{\frac{2\gamma k_BT}{m^2}}\mathrm{d}W(t). \tag{14.12}$$

Thus velocity relaxes exponentially toward the Maxwellian equilibrium distribution

$$f(v) = \sqrt{\frac{m}{2\pi k_BT}}\exp\left(-\frac{mv^2}{2k_BT}\right).$$

Einstein relation. A direct computation (see Sheet 4) shows that for large times $t \gg m/\gamma$,

$$\mathbb{E}X^2(t) = \frac{2k_BT}{\gamma}t. \tag{14.13}$$

Hence, Langevin predicts that as time passes, the mean squared distance travelled by the particle increases at a constant rate. Einstein obtained the same result but with a different constant, namely $2D$, see (9.5). Equating the two expressions for the mean-squared distance leads to

$$D = \frac{k_BT}{\gamma}. \tag{14.14}$$

This is the *Einstein–Smoluchowski relation*. It links macroscopic diffusion to microscopic friction and temperature.

The value of γ depends on the particle's geometry and the medium's properties. If the Brownian particle is spherical with diameter a, Stokes' law states $\gamma = 3\pi\eta a$, where η is the viscosity of the medium. Substituting this into (14.14) leads to the *Stokes–Einstein relation*

$$D = \frac{k_BT}{3\pi\eta a}. \tag{14.15}$$

Overdamped regime. When $m/\gamma \ll 1$, velocity relaxes rapidly. Formally neglecting inertial effects in (14.12) yields

$$V(t)\mathrm{d}t = \sqrt{2k_BT/\gamma}\mathrm{d}W(t) \quad \longrightarrow \quad \mathrm{d}X(t) = \sqrt{2D}\,\mathrm{d}W(t),$$

which is precisely Brownian motion. Again, the full process $(X(t), V(t))$ is Markovian, but $X(t)$ alone has memory. In the overdamped regime, this memory decays rapidly.

Fokker–Planck description. Let $p(x, v, t)$ denote the joint density. From (11.15), p satisfies

$$\partial_t p + v\partial_x p = \frac{\gamma}{m}\partial_v\left(vp + \frac{k_BT}{m}\partial_v p\right). \tag{14.16}$$

In the absence of an external potential, there is no confinement in space, so no normalisable stationary density exists in x. However, the velocity component admits the Maxwellian equilibrium

$$f(v) = \sqrt{\frac{m}{2\pi k_BT}}\exp\left(-\frac{mv^2}{2k_BT}\right). \tag{14.17}$$

If the particle is confined by a potential $U(x)$ so that the position equation becomes

$$\mathrm{d}X(t) = V(t)\,\mathrm{d}t, \qquad m\mathrm{d}V(t) = -\gamma V(t)\,\mathrm{d}t - U'(X(t))\,\mathrm{d}t + \sqrt{2\gamma k_BT}\,\mathrm{d}W(t), \tag{14.18}$$

then the full process admits the stationary density

$$p_\infty(x, v) = Z^{-1}\exp\left(-\frac{1}{k_BT}\left(\frac{1}{2}mv^2 + U(x)\right)\right), \tag{14.19}$$

where Z is a normalisation constant. This invariant distribution is known as *the Maxwell–Boltzmann distribution*, factorising into kinetic and potential energy contributions.

In Figure 14.2, we show trajectories of the two-dimensional version of the Langevin dynamics (14.12) for increasing values of the damping coefficient γ. In particular, we consider $\gamma = 1, 10, 100, 1000$ while keeping the diffusivity D (14.14) fixed. We observe that for small γ, the position $\boldsymbol{X}(t)$ has a lot of memory, while as γ increases, it looks more and more like Brownian motion. Also, we observe that for large γ, the velocity of the Brownian particle decouples from its position and equilibrates to its stationary density, the Maxwellian (14.17).

Exercise 14.1 (Fluctuation–dissipation and the Einstein relation)**.** Consider the Langevin equation (14.10) with $R(t)$ a mean-zero Gaussian white noise with covariance

$$\mathbb{E}[R(t)R(s)] = \Gamma\delta(t - s).$$

(i) Solve the equation for $V(t)$ and compute $\mathbb{E}[V(t)^2]$. Assuming the system reaches thermal equilibrium, where

$$\frac{1}{2}m\mathbb{E}[V_\infty^2] = \frac{1}{2}k_BT,$$

show that this determines Γ and hence derive the fluctuation–dissipation relation

$$\Gamma = 2\gamma k_BT.$$

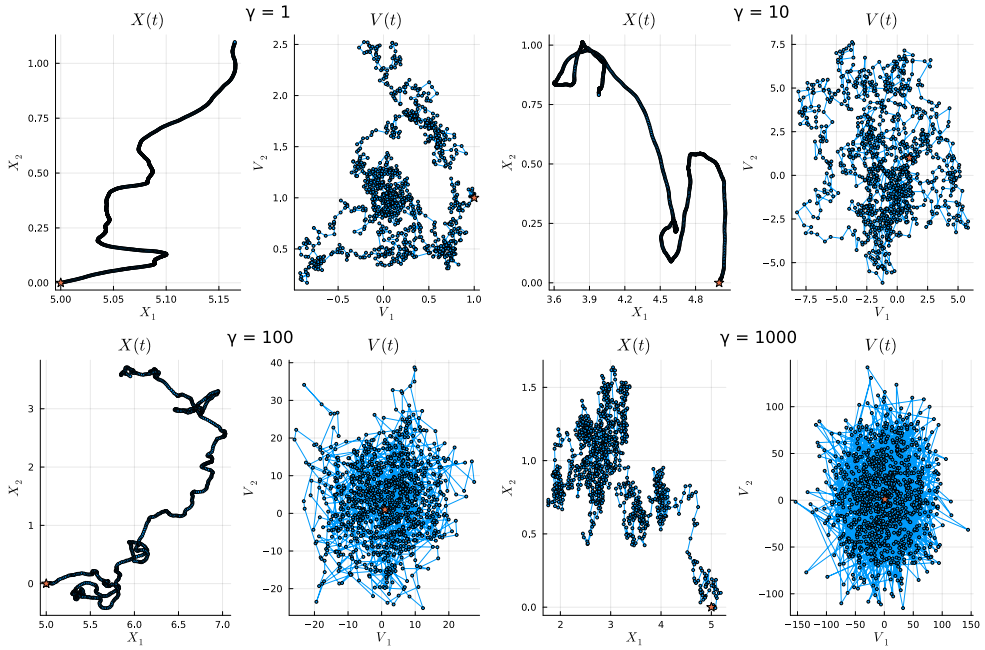

Figure 14.2: Simulation of the 2D version of the Langevin dynamics (14.12) for different values of the damping coefficient γ at a fixed diffusivity D given by (14.14). The trajectories of the position $\boldsymbol{X}(t)$ and velocity $\mathbf{V}(t)$ are simulated via the Euler–Maruyama with $\Delta t = 10^{-3}$ and $T_f = 1$.

(ii) Multiply the Langevin equation by $X(t)$ and take expectations to derive an ODE for $\mathbb{E}[X^2(t)]$. Using part (i), show that for large times $t \gg m/\gamma$,

$$\mathbb{E}[X^2(t)] = \frac{2k_BT}{\gamma}t.$$

Conclude the Einstein–Smoluchowski relation (14.14).

Tasks

- Write the forward Kolmogorov equations for the run-and-tumble model with the position-dependent turning rate in (14.8). Obtain the conservation law for the total density $p = p^+ + p^-$ and the equation for the flux $J = s(p^+ - p^-)$.
- Do Exercise 14.1.
- Write the Fokker–Planck equation for the Langevin dynamics with a confining potential $U(x)$,

$$\mathrm{d}X(t) = V(t)\,\mathrm{d}t, \qquad m\mathrm{d}V(t) = -\gamma V(t)\,\mathrm{d}t - U'(X(t))\,\mathrm{d}t + \sqrt{2\gamma k_BT}\,\mathrm{d}W(t),$$

and verify that the stationary density is (14.19).

Example codes

`Lect14.ipynb` contains simulations of the two second-order models discussed in this lecture: the run-and-tumble and the Langevin models. I suggest exploring how changing the parameters of the models makes the simulations look more "diffusion-like".

Lecture 15: Diffusion approximations

15.1 From second-order models for motility to Brownian motion

In the last lecture we discussed two common models for motility: the run-and-tumble and the Langevin equation. Here we consider limits of these models to Brownian motion.

Run-and-tumble Recall the coupled system of forward Kolmogorov equations (14.3) for $p^{\pm}(x,t)$ for a run-and-tumble particle in one dimension, doing runs at speed s and tumbling at rate λ. We also showed how the total probability density $p(x,t) = p^{+}(x,t) + p^{-}(x,t)$ satisfies the hyperbolic PDE (14.6), that is,

$$\frac{1}{2\lambda}\partial_t^2 p + \partial_t p = \frac{s^2}{2\lambda}\partial_x^2 p. \tag{15.1}$$

We now consider the parabolic scaling of (15.1) by rescaling space and time as $x = x^*/\delta$ and $t = t^*/\delta^2$ with $\delta \ll 1$ (recall that this is the same scaling we used in the random walk converging to Brownian motion of §9.2). Then (15.1) becomes (dropping asterisks)

$$\frac{\delta^2}{2\lambda}\frac{\partial^2 p}{\partial t^2} + \frac{\partial p}{\partial t} = \frac{s^2}{2\lambda}\frac{\partial^2 p}{\partial x^2}, \tag{15.2}$$

and hence at leading order in δ, we obtain

$$\frac{\partial p}{\partial t} = \frac{s^2}{2\lambda}\frac{\partial^2 p}{\partial x^2}. \tag{15.3}$$

Equation (15.3) is valid for times $t \gg 1/\lambda$, that is, after many reorientation events when the memory term in (15.2) becomes negligible. Therefore, for long times, a bacterium undergoing the velocity-jump process (14.2) satisfies the diffusion equation (15.3) with diffusion coefficient

$$D = \frac{s^2}{2\lambda}. \tag{15.4}$$

Langevin's equation Similarly, recall the Langevin model for the velocity of a Brownian motion. Multiplying by m/γ on both sides of (14.12) we have

$$\frac{m}{\gamma}\mathrm{d}V(t) = -V(t)\mathrm{d}t + \sqrt{2D}\mathrm{d}W(t), \qquad \mathrm{d}X(t) = V(t)\mathrm{d}t, \tag{15.5}$$

where we have used the Einstein relation (14.14), $D = k_B T/\gamma$. The large friction limit $m/\gamma \ll 1$ means that formally we may set the left-hand side of the SDE for V to zero and substitute $V(t)$ in the equation for $X(t)$, leading to Brownian motion with diffusivity D.

Let $\epsilon = m/\gamma$. We may also consider the overdamped limit $\epsilon \ll 1$ at the level of the associated Fokker–Planck equation (14.16). Since the velocity relaxes on the fast time scale $O(\epsilon)$, we introduce the slow time variable

$$\bar{t} = \epsilon t,$$

which captures the long-time evolution of the position occurring on time scales $t = O(\epsilon^{-1})$. Writing $p(x, v, t) = \bar{p}(x, v, \bar{t})$ gives

$$\epsilon^2 \frac{\partial \bar{p}}{\partial \bar{t}} + \epsilon v \frac{\partial \bar{p}}{\partial x} = \frac{\partial}{\partial v}\left(v\bar{p} + \beta^{-1} \frac{\partial \bar{p}}{\partial v} \right) =: L\bar{p}, \tag{15.6}$$

where we have introduced β as the inverse temperature, $\beta^{-1} = k_B T/m$, and the linear operator L. We look for a solution of the form $\bar{p}(x, v, \bar{t}) = \bar{p}^{(0)}(x, v, \bar{t}) + \epsilon \bar{p}^{(1)}(x, v, \bar{t}) + \epsilon^2 \bar{p}^{(2)}(x, v, \bar{t}) + \cdots$. Substituting this into (15.6) and equating powers of ϵ we obtain the hierarchy of equations

$$0 = L\bar{p}^{(0)}, \tag{15.7}$$

$$v \frac{\partial \bar{p}^{(0)}}{\partial x} = L\bar{p}^{(1)}, \tag{15.8}$$

$$\frac{\partial \bar{p}^{(0)}}{\partial \bar{t}} + v \frac{\partial \bar{p}^{(1)}}{\partial x} = L\bar{p}^{(2)}. \tag{15.9}$$

Integrating (15.7) with respect to v and using $\bar{p}, \partial_v \bar{p} \to 0$ for $|v| \to \infty$, we get

$$\bar{p}^{(0)}(x, v, \bar{t}) = \bar{\rho}(x, \bar{t}) f(v), \tag{15.10}$$

where $\bar{\rho}(x, \bar{t})$ is the constant of integration (to be determined) and $f(v)$ is the *Maxwellian distribution*

$$f(v) = \frac{1}{\sqrt{2\pi\beta^{-1}}} \exp\left(-\frac{\beta v^2}{2} \right) = \sqrt{\frac{m}{2\pi k_B T}} \exp\left(-\frac{m v^2}{2 k_B T} \right). \tag{15.11}$$

So, at leading order, the velocity distribution is Maxwellian and independent of position. Inserting (15.10) into (15.8) we have

$$L\bar{p}^{(1)} = \frac{\partial \bar{\rho}}{\partial x} v f(v), \tag{15.12}$$

with solution

$$\bar{p}^{(1)} = -\frac{\partial \bar{\rho}}{\partial x} v f(v). \tag{15.13}$$

Substituting (15.10) and (15.13) into (15.9) gives

$$L\bar{p}^{(2)} = f(v)\left(\frac{\partial \bar{\rho}}{\partial \bar{t}} - v^2 \frac{\partial^2 \bar{\rho}}{\partial x^2} \right).$$

Now integrating over v and using again that $\bar{p}$ decays to 0 as $|v| \to \infty$ leads to the solvability condition

$$\begin{aligned} 0 &= \frac{\partial \bar{\rho}}{\partial \bar{t}} \int_{\mathbb{R}} f(v) \mathrm{d}v - \frac{\partial^2 \bar{\rho}}{\partial x^2} \int_{\mathbb{R}} v^2 f(v) \mathrm{d}v \\ &= \frac{\partial \bar{\rho}}{\partial \bar{t}} - \beta^{-1} \frac{\partial^2 \bar{\rho}}{\partial x^2}, \end{aligned}$$

after one integration by parts and using (15.11). Going back to the original time $t = \epsilon^{-1}\bar{t}$ and writing $\bar{\rho}(x, \bar{t}) = \rho(x, t)$ gives

$$\frac{\partial \rho}{\partial t} = D \frac{\partial^2 \rho}{\partial x^2}. \tag{15.14}$$

where D is given in (14.14). Using (15.10) we see that the *spatial probability density*

$$\int_{\mathbb{R}} p(x, v, t) \mathrm{d}v, \tag{15.15}$$

satisfies the diffusion equation (15.14) for large γ/m. This equation is consistent with the overdamped limit formally obtained, taking $m/\gamma \to 0$ at the level of the SDE.

The overdamped limit should be interpreted as the singular limit $\epsilon = m/\gamma \to 0$ with temperature T fixed. The mass m controls the fast relaxation time of the velocity, while the friction coefficient γ determines the spatial diffusivity $D = k_B T/\gamma$. Sending $m \to 0$ eliminates the fast kinetic time scale, whereas sending $\gamma \to \infty$ without rescaling time would simply suppress diffusion.

Exercise 15.1. Consider the overdamped limit $m/\gamma \ll 1$ of the Langevin model with a confining potential U (14.18). Show that the spatial density $\rho(x,t)$ satisfies

$$\frac{\partial \rho}{\partial t} = \frac{\partial}{\partial x}\left(D\frac{\partial \rho}{\partial x} + \frac{U'(x)}{\gamma}\rho\right).$$

15.2 The chemical Langevin SDE revisited

In the first part of the course, we introduced three levels of description for well-mixed chemical reaction networks (CRNs):

- exact stochastic simulation algorithms (SSA),
- the chemical master equation (CME),
- the chemical Langevin equation (CLE) / chemical Fokker–Planck equation.

The first two are different descriptions of the same process (trajectories vs. probabilities), while the third is another example of a diffusion approximation. Let's look at when the chemical Langevin provides a valid approximation to the CRN. First, consider the following example to refresh the concepts we covered for CRNs.

Example 15.1. Consider the CRN (4.16) with two reactions

$$2A \xrightarrow{k_1} \emptyset, \qquad \emptyset \xrightarrow{k_2} A,$$

with state $A(t) \in \mathbb{Z}_{\geq 0}$ in a volume ν.

The Gillespie SSA for this system is

(a) At state $A(t)$, compute $\alpha_1(A) = k_1 A(A-1)/\nu$ and $\alpha_2(A) = k_2\nu$.

(b) The next reaction occurs at time $T \sim \mathrm{Exp}(\alpha_0)$, where $\alpha_0 = \alpha_1 + \alpha_2$.

(c) Set $A(t+T) = A(t) + \zeta$, where $\zeta = \zeta_1 = -2$ with probability α_1/α_0 and $\zeta = \zeta_2 = +1$ with probability α_2/α_0.

The chemical master equation (CME): Let $p_n(t) := \mathbb{P}(A(t) = n)$. Then, for $n \geq 0$,

$$\frac{\mathrm{d}p_n}{\mathrm{d}t} = (k_1/\nu)(n+2)(n+1)\,p_{n+2}(t) - (k_1/\nu)n(n-1)\,p_n(t) + k_2\nu\,p_{n-1}(t) - k_2\nu\,p_n(t), \quad (15.16)$$

Poisson representation and tau-leaping: The evolution of $A(t)$ may be written as

$$A(t) = A(0) - 2Y_1\left(\int_0^t \alpha_1(A(s))\mathrm{d}s\right) + Y_2\left(\int_0^t \alpha_2(A(s))\mathrm{d}s\right),$$

where Y_j are unit-rate Poisson processes. Taking fixed time steps τ, the tau-leaping approximation is

$$A(t+\tau) = A(t) - 2P_1 + P_2,$$

where $P_i \sim \text{Poisson}(\alpha_i(A(t))\tau)$. Using that $\text{Poisson}(\lambda) \sim \lambda + \sqrt{\lambda}\mathcal{N}(0,1)$ for $\lambda \gg 1$, we wrote the Chemical Langevin SDE (see (8.5))

$$\mathrm{d}A(t) = [-2\alpha_1(A(t)) + \alpha_2(A(t))]\,\mathrm{d}t + 2\sqrt{\alpha_1(A)}\mathrm{d}W_1(t) + \sqrt{\alpha_2(A)}\mathrm{d}W_2(t).$$

Consider a general CRN with n species and m reactions. Let $\boldsymbol{X}(t) \in \mathbb{N}^n$ denote the vector of molecule numbers, $\Gamma \in \mathbb{Z}^{n\times m}$ the stoichiometric matrix, and $\alpha_j(\boldsymbol{x})$ the propensity functions. The chemical master equation reads (see (4.12))

$$\frac{\partial}{\partial t}p(\boldsymbol{x},t) = \sum_{j=1}^{m}\left[\alpha_j(\boldsymbol{x}-\boldsymbol{\zeta}_j)\,p(\boldsymbol{x}-\boldsymbol{\zeta}_j,t) - \alpha_j(\boldsymbol{x})\,p(\boldsymbol{x},t)\right]. \tag{15.17}$$

Equivalently, we saw that the process admits the random time-change representation, see (7.3)

$$\boldsymbol{X}(t) = \boldsymbol{X}(0) + \sum_{j=1}^{m}\boldsymbol{\zeta}_j Y_j\left(\int_0^t \alpha_j(\boldsymbol{X}(s))\mathrm{d}s\right), \tag{15.18}$$

where Y_j are independent unit-rate Poisson processes. For this representation, in Lecture 8 we saw that as we scale the volume of the system up, we can approximate the Poisson noise in (15.18) as Gaussian noise. Here we derive the diffusion approximation directly at the level of the generator via the Kramers–Moyal expansion.

15.2.1 The Kramers-Moyal expansion

An alternative route from the chemical master equation to the chemical Fokker–Planck equation is via the Kramers–Moyal expansion. This makes explicit that the CLE corresponds to a second-order approximation of the generator of the jump process.

Consider the CME (15.17). We assume that the state $\boldsymbol{x}$ is large (high-copy-number regime) and treat it as a continuous variable in $\mathbb{R}^n$. For each reaction channel j, we expand the gain term around $\boldsymbol{x}$ using a Taylor expansion:

$$\begin{aligned}\alpha_j(\boldsymbol{x}-\boldsymbol{\zeta}_j)p(\boldsymbol{x}-\boldsymbol{\zeta}_j,t) = \alpha_j(\boldsymbol{x})p(\boldsymbol{x},t) &- \sum_{i=1}^{n}\zeta_{ij}\frac{\partial}{\partial x_i}\left[\alpha_j(\boldsymbol{x})p(\boldsymbol{x},t)\right] \\ &+ \frac{1}{2}\sum_{i,k=1}^{n}\zeta_{ij}\zeta_{kj}\frac{\partial^2}{\partial x_i\partial x_k}\left[\alpha_j(\boldsymbol{x})p(\boldsymbol{x},t)\right] + \cdots.\end{aligned}$$

Substituting this expansion into (15.17), the zeroth-order terms cancel, and we obtain formally

$$\frac{\partial p}{\partial t} = -\sum_{i=1}^{n}\frac{\partial}{\partial x_i}\left[\sum_{j=1}^{m}\zeta_{ij}\alpha_j(\boldsymbol{x})\,p\right] + \frac{1}{2}\sum_{i,k=1}^{n}\frac{\partial^2}{\partial x_i\partial x_k}\left[\sum_{j=1}^{m}\zeta_{ij}\zeta_{kj}\alpha_j(\boldsymbol{x})\,p\right] + \cdots. \tag{15.19}$$

Truncating this expansion at second order yields a Fokker–Planck equation

$$\frac{\partial p}{\partial t} = -\sum_{i=1}^{n}\frac{\partial}{\partial x_i}\left[a_i(\boldsymbol{x})p\right] + \frac{1}{2}\sum_{i,k=1}^{n}\frac{\partial^2}{\partial x_i\partial x_k}\left[B_{ik}(\boldsymbol{x})p\right], \tag{15.20}$$

with drift and diffusion coefficients

$$a_i(\boldsymbol{x}) = \sum_{j=1}^{m} \zeta_{ij}\alpha_j(\boldsymbol{x}), \qquad B_{ik}(\boldsymbol{x}) = \sum_{j=1}^{m} \zeta_{ij}\zeta_{kj}\alpha_j(\boldsymbol{x}).$$

In matrix form,

$$\boldsymbol{a}(\boldsymbol{x}) = \Gamma\boldsymbol{\alpha}(\boldsymbol{x}), \qquad \boldsymbol{B}(\boldsymbol{x}) = \Gamma\mathrm{diag}(\alpha_1(\boldsymbol{x}), \ldots, \alpha_m(\boldsymbol{x}))\Gamma^\top.$$

These coincide with the drift vector and diffusion matrix obtained from the chemical Langevin equation, see (8.4). Note that $\mathrm{rank}(\boldsymbol{B}) \leq m$, so $\boldsymbol{B}$ may be degenerate. This reflects conserved combinations of species and a lower-dimensional stochastic manifold.

The Kramers–Moyal expansion is formal: retaining all higher-order terms recovers the original jump process, while truncation at second order yields a diffusion approximation. This dichotomy is captured by a classical result known as Pawula's theorem [14], which states that a finite truncation of the Kramers–Moyal expansion is consistent only up to second order. In other words, either all coefficients of order higher than two vanish identically, leading to a diffusion process, or infinitely many higher-order terms must be retained. There is no consistent truncation at any finite order greater than two.

Probabilistic interpretation. In Lecture 8, we derived the same diffusion approximation heuristically by applying tau-leaping to the random time-change representation and approximating Poisson increments by Gaussian increments. Consider the system-size scaling,

$$\boldsymbol{X}^\nu(t) = \frac{1}{\nu}\boldsymbol{X}(t),$$

where ν denotes the system volume, and mass-action kinetics,

$$\alpha_j(\boldsymbol{X}) = \nu v_j(\boldsymbol{X}^\nu) + o(1),$$

where v_j are independent of ν and coincide with the deterministic rates, see (3.6). Under system-size scaling $\nu \gg 1$, reaction counts over short time intervals have large Poisson parameters and the Poisson representation (15.18) becomes

$$\begin{aligned}
\boldsymbol{X}^\nu(t+\tau) &= \boldsymbol{X}^\nu(t) + \frac{1}{\nu}\sum_{j=1}^{m} \boldsymbol{\zeta}_j Y_j\left(\nu\int_0^\tau v_j(\boldsymbol{X}^\nu(t+s))\mathrm{d}s\right) \\
&\approx \boldsymbol{X}^\nu(t) + \frac{1}{\nu}\sum_{j=1}^{m} \boldsymbol{\zeta}_j Y_j\left(\nu v_j(\boldsymbol{X}^\nu(t))\tau\right) \\
&\approx \boldsymbol{X}^\nu(t) + \frac{1}{\nu}\sum_{j=1}^{m} \boldsymbol{\zeta}_j \left[\nu v_j(\boldsymbol{X}^\nu)\tau + \sqrt{\nu v_j(\boldsymbol{X}^\nu)\tau}\,\xi_j\right]
\end{aligned}$$

with $\xi_j \sim \mathcal{N}(0,1)$. Hence, the central-limit approximation leads to the Chemical Langevin equation:

$$\mathrm{d}\boldsymbol{X}^\nu(t) = \Gamma\boldsymbol{v}(\boldsymbol{X}^\nu(t))\mathrm{d}t + \frac{1}{\sqrt{\nu}}\Gamma\mathrm{diag}\left(\sqrt{v_1(\boldsymbol{X}^\nu(t))}, \ldots, \sqrt{v_m(\boldsymbol{X}^\nu(t))}\right)\mathrm{d}\boldsymbol{W}(t), \tag{15.21}$$

where $\boldsymbol{v} = (v_1, \dots, v_m)^\top$ and $\boldsymbol{W}$ is an m-dimensional standard Brownian motion. In terms of the original variables $\boldsymbol{X} = \nu \boldsymbol{X}^\nu$, the *chemical Langevin SDE* reads

$$\mathrm{d}\boldsymbol{X}(t) = \Gamma \boldsymbol{\alpha}(\boldsymbol{X}(t))\mathrm{d}t + \Gamma \mathrm{diag}\left(\sqrt{\alpha_1(\boldsymbol{X}(t))}, \dots, \sqrt{\alpha_m(\boldsymbol{X}(t))}\right) \mathrm{d}\boldsymbol{W}(t), \tag{15.22}$$

The Fokker–Planck associated to this SDE is exactly (15.20) derived above.

As $\nu \to \infty$, the noise term in (15.21) vanishes, $\boldsymbol{X}^\nu \to \boldsymbol{x}$ and we recover the deterministic reaction-rate equation (RRE) (3.2)

$$\dot{\boldsymbol{x}} = \Gamma \boldsymbol{v}(\boldsymbol{x}).$$

Thus we obtain the hierarchy

$$\text{CME} \longrightarrow \text{CLE} \longrightarrow \text{RRE}.$$

The first arrow corresponds to a diffusion (central limit) approximation of the jump process described by the CME. Formally, the CLE is obtained as the second-order truncation of the Kramers–Moyal expansion of the CME, that is by approximating the generator by a second-order differential operator [6, §11.2.2]. The second arrow corresponds to a law-of-large-numbers limit: as the system size $\nu \to \infty$, fluctuations vanish and the CLE converges to the deterministic reaction-rate equation (RRE).

15.2.2 Validity regime of the chemical Langevin equation (CLE)

While the Gillespie SSA, the chemical master equation and the Poisson time-change representation are exact descriptions of a continuous-time Markov chain on $\mathbb{N}^n$, the CLE is a diffusion approximation to the jump process. In particular, the approximation relies on replacing Poisson increments by Gaussian increments. This approximation is valid when, on the time scale of interest, the expected number of firings for every reaction channel is large:

$$\nu v_j(\boldsymbol{X}^\nu(t))\,\tau \gg 1 \qquad \text{for all } j,$$

so that Poisson counts may be approximated by Gaussians. This happens in the high-copy-number regime. In this regime, relative fluctuations are $O(\nu^{-1/2})$ and hence small when the system volume ν is large. There are important situations where it fails qualitatively.

Low copy numbers. If some species have small molecule numbers, propensities may be small and the Gaussian approximation is no longer valid. Moreover, the diffusion term in (15.22) is continuous and unbounded, so trajectories may become negative. The chemical master equation, by contrast, preserves the discrete state space $\mathbb{N}^N$ exactly.

Boundary behaviour and extinction. If the reaction network admits absorbing states (for instance extinction states), the CME respects them exactly. The CLE replaces the jump process by a diffusion on $\mathbb{R}^N$, and boundary hitting behaviour may be distorted. Extinction times may therefore be poorly approximated.

Multistability and rare transitions. In bistable systems, the deterministic dynamics admits multiple stable equilibria. For finite ν, the CME exhibits noise-induced transitions between metastable states. These are rare events whose mean transition times typically scale exponentially in ν. We saw an example of this in Section 5.5.

While the CLE captures small fluctuations around equilibria, it may not reproduce the correct large-deviation structure of the jump process. Consequently, switching times predicted by the CLE may differ significantly from those of the CME.

15.3 Summary

To sum up, both the motility examples (telegraph/run-and-tumble and Langevin) and the chemical reaction network example are instances of diffusion approximations, but they arise from slightly different microscopic mechanisms.

- For velocity-jump models (e.g., run-and-tumble), the diffusion limit is obtained by a parabolic (hydrodynamic) scaling of space and time, yielding a diffusion equation for the spatial density.
- The Langevin overdamped limit proceeds by separation of fast (velocity) and slow (position) variables and a solvability condition for the associated Fokker–Planck operator.
- The chemical diffusion approximation (CME $\to$ CLE) is a central-limit-type limit in which discrete Poisson noise is approximated by Gaussian noise. We derived this heuristically via tau-leaping, but it has been shown rigorously in [9].

In all cases, the unifying principle is that many small, frequent, and approximately independent microscopic events lead, at macroscopic scales, to Gaussian fluctuations and a second-order (diffusion) description.

Lecture 16: Reaction–diffusion SDEs

In Lecture 6 we incorporated spatial effects into chemical reaction networks (CRNs) by dividing space into compartments and modelling

- diffusion as jumps between neighbouring compartments,
- reactions within each compartment.

For n species and K compartments, this led to the reaction–diffusion master equation (RDME), a continuous-time Markov chain on $\mathbb{Z}_{\geq 0}^{nK}$.

The spatial discretisation was artificial: in most applications, space is continuous, but we imposed a lattice. Now that we have developed stochastic differential equations, we are ready to formulate spatial stochastic models directly in continuous space.

Recall the hierarchy for the well-mixed case from the last lecture:

$$\text{CME} \longrightarrow \text{CLE} \longrightarrow \text{RRE}.$$

In this lecture we show that, introducing space, it becomes

$$\text{RDME} \longrightarrow \text{reaction–diffusion SPDE} \longrightarrow \text{reaction–diffusion PDE}.$$

16.1 Brownian particles with zeroth- and first-order reactions

Instead of compartments, we now track individual molecules in continuous space. Let $\boldsymbol{X}_i(t) \in \Omega \subset \mathbb{R}^d$ denote the position of the ith molecule. Between reactions,

$$\mathrm{d}\boldsymbol{X}_i(t) = \sqrt{2D}\,\mathrm{d}\mathbf{W}_i(t), \tag{16.1}$$

with independent Brownian motions W_i.

Reactions that do not require molecular encounters are straightforward to incorporate in the Brownian particle framework. In particular, zeroth- and first-order reactions act independently on each molecule and therefore do not involve spatial proximity.

As an example, consider diffusion together with production and degradation

$$\mathrm{A} \xrightarrow{k_1} \emptyset, \qquad \emptyset \xrightarrow{k_2} \mathrm{A}, \quad \text{for } x \in \Omega_0 \subset \Omega. \tag{16.2}$$

Each molecule of A diffuses according to (16.1) and degrades independently with rate k_1. New molecules are created in the subdomain Ω_0 at rate k_2 per unit volume. This is the continuous-space analogue of the linear reaction–diffusion model studied in Section 6.3, see (6.12). We may simulate this system using Euler–Maruyama for the spatial dynamics and the naive algorithm 1.2 for the zeroth- and first-order reactions; see Algorithm 16.1. Since the spatial evolution is already discretised in time with a fixed timestep Δt, we do not use the Gillespie SSA for the reactions.

Algorithm 16.1: Brownian diffusion with production and degradation (16.2)

Set final time T, timestep Δt, the initialise the number of molecules ($A(0) = N$) and their positions, $\boldsymbol{X}_i(0) \sim p_0(\boldsymbol{x})$ for $i = 1, \ldots, N$.
For $n = 0, 1, \ldots$ while $t_n < T$:

(a) For each molecule i, generate $\boldsymbol{\xi}_i \sim \mathcal{N}(\mathbf{0}, \mathbb{I}_d)$ and update position by Euler–Maruyama:

$$\boldsymbol{X}_i^{n+1} = \boldsymbol{X}_i^n + \sqrt{2D\Delta t}\,\boldsymbol{\xi}_i.$$

(b) (Boundary conditions) If $\boldsymbol{X}_i^{n+1} \notin \Omega$, reflect, eliminate, or wrap according to the chosen boundary condition.

(c) (Degradation) For each molecule generate $r \sim \mathcal{U}(0,1)$. If $r < k_1\Delta t$, remove the molecule.

(d) (Production) Generate $r \sim \mathcal{U}(0,1)$. If $r < k_2|\Omega_0|\Delta t$, create a new molecule at a uniformly random position in Ω_0.

Set $t_{n+1} = t_n + \Delta t$ and repeat.

In principle, one could construct a hybrid scheme in which the spatial motion is advanced with fixed timesteps (Euler–Maruyama), while reaction times are generated using the Gillespie SSA. However, when reactions are proximity-based (as we will see next), the continuous motion of particles changes the reaction propensities between events. For this reason, in Brownian dynamics simulations it is often more practical to use fixed-timestep methods for both diffusion and reactions.

16.2 Brownian particles with bimolecular reations

We now consider reactions that require molecular encounters. For example,

$$\mathrm{A} + \mathrm{B} \xrightarrow{k} \emptyset. \tag{16.3}$$

In continuous space, such reactions are modelled using a *reaction radius* ρ. Let $\boldsymbol{X}_A(t)$ and $\boldsymbol{X}_B(t)$ denote the positions of two molecules. Whenever

$$|\mathbf{X}_A(t) - \mathbf{X}_B(t)| < \rho,$$

the molecules are sufficiently close to react. This is the continuous analogue of being in the same compartment in the RDME.

Perfect versus imperfect reactions. The simplest assumption is that the reaction is *perfect*: whenever the separation is less than ρ, the molecules react with probability one. More generally, we may allow an *imperfect reaction*: whenever

$$|\mathbf{X}_A(t) - \mathbf{X}_B(t)| < \rho,$$

the molecules react with probability $P \in (0, 1]$. If the reaction does not occur, one may either

- allow the particles to continue diffusing freely, or
- reflect them at contact (as if they were hard spheres).

This introduces an additional modelling degree of freedom: the length scale ρ may represent particle size, while the probability P controls the chemical reactivity. In particular, the second option models excluded volume and treats ρ as the physical particle diameter.

Diffusion-limited reaction rate in $d = 3$. To connect this microscopic rule to macroscopic mass-action kinetics, consider the relative coordinate

$$\mathbf{R}(t) = \boldsymbol{X}_A - \boldsymbol{X}_B.$$

If A and B diffuse independently with diffusion coefficients D_A and D_B, then $\mathbf{R}(t)$ diffuses with effective diffusion coefficient

$$D = D_A + D_B.$$

To see this, recall that if $\mathbf{W}_i(t), i = 1, 2$ are independent standard Brownian motions, then for constants $a, b \in \mathbb{R}$ the linear combination $a\mathbf{W}_1(t) + b\mathbf{W}_2(t)$ is again a Brownian motion with variance $(a^2 + b^2)t$.

Assume first a perfect reaction ($P = 1$). Let $c(\boldsymbol{r}, t)$ denote the radial concentration of B around a fixed A particle. At steady state, by spherical symmetry,

$$D\left[c''(r) + \frac{1}{r}c'(r)\right] = 0 \qquad \text{for } r > \rho,$$

with boundary conditions

$$c(\rho) = 0, \qquad c(r) \to c_\infty \quad \text{as } r \to \infty.$$

The solution is given by

$$c(r) = c_\infty \left(1 - \frac{\rho}{r}\right).$$

The inward diffusive flux at $r = \rho$ is

$$J_r = D\frac{\mathrm{d}c}{\mathrm{d}r} = D\frac{c_\infty}{\rho}.$$

The total inward flux into the sphere is

$$J = 4\pi\rho^2 J_r = 4\pi D\rho c_\infty.$$

Identifying $J = kc_\infty$ gives the Smoluchowski rate constant

$$k = 4\pi(D_A + D_B)\rho \qquad (d = 3). \tag{16.4}$$

In other words, this gives the appropriate reaction radius ρ corresponding to a perfect reaction ($P = 1$) with rate k.

Choice of reaction radius and timestep restrictions. Although the Smoluchowski formula (16.4) provides a direct link between the reaction radius ρ and the macroscopic rate constant k, this raises two modelling issues.

First, molecules are treated as points in the Brownian dynamics description, whereas in reality they have finite size. Using the Stokes–Einstein relation (14.15)

$$a = \frac{k_B T}{6\pi\eta D},$$

we may estimate the physical molecular diameter a. For typical biochemical parameters, one finds that achieving $\rho > (a_A + a_B)/2$ would require reaction rate constants of order $10^9\,\mathrm{M}^{-1}\mathrm{s}^{-1}$. Since many protein–protein reactions have rate constants around $10^6\,\mathrm{M}^{-1}\mathrm{s}^{-1}$, the model typically requires $\rho \ll a$. Thus the reaction radius should be interpreted as an effective interaction length scale rather than the true molecular size.

Second, a small reaction radius imposes restrictions on the timestep Δt in simulations. For the Smoluchowski description to remain valid, the typical diffusive displacement during one timestep must be much smaller than ρ. Since the root-mean-square displacement of the relative coordinate is

$$s = \sqrt{2(D_A + D_B)\Delta t},$$

we require $s \ll \rho$. For typical diffusion coefficients of order $10\,\mu\mathrm{m}^2\mathrm{s}^{-1}$ and biochemical rate constants, this can force Δt to be extremely small.

One way to relax these constraints is via the imperfect reaction mechanism. If reactions occur with probability $P < 1$ upon contact, then ρ may be chosen to represent the physical particle size a, while P controls the intrinsic chemical reactivity. In this case, the diffusion-limited value (16.4) represents the maximal possible rate, and the intrinsic reactivity P provides an additional degree of freedom that modulates the effective macroscopic rate constant.

In well-mixed systems, reaction rates depend only on concentrations. In spatial systems, two regimes arise:

- Reaction-limited: diffusion fast, well-mixed approximation valid.
- Diffusion-limited: search time dominates reaction time.

In low dimensions ($d = 1, 2$), diffusion is recurrent. Spatial correlations build up and decay rates differ from mass-action predictions. Such effects may not be captured by coarse RDME discretisations, but arise naturally in Brownian particle models.

Brownian dynamics simulation. We may simulate diffusion and bimolecular reactions using a fixed timestep Δt. Algorithm 16.2 prescribes a reaction radius ρ and a reaction probability P. If $P = 1$, the model corresponds to a perfectly absorbing reaction at contact, and the effective macroscopic rate constant k is given by the Smoluchowski formula (16.4). If $P < 1$, the effective reaction rate is reduced. In this case there is no simple closed-form expression for k in terms of ρ and P. Approximate relationships can be derived using the so-called Doi or the λ–$\bar{\rho}$ models; see, for example, [5, Section 6.6].

Algorithm 16.2: Brownian diffusion with bimolecular reaction (16.3)

Set final time T, timestep Δt, the initialise the number of molecules ($A(0) = N$) and their positions, $\boldsymbol{X}_i(0) \sim p_0(\boldsymbol{x})$ for $i = 1, \dots, N$.
For $n = 0, 1, \dots$ while $t_n < T$:

(a) For each molecule i, generate $\boldsymbol{\xi}_i \sim \mathcal{N}(\mathbf{0}, \mathbb{I}_d)$ and update position:

$$\boldsymbol{X}_i^{n+1} = \boldsymbol{X}_i^n + \sqrt{2D_i \Delta t}\, \boldsymbol{\xi}_i.$$

(b) Impose boundary conditions on Ω.

(c) For each pair (i, j) of molecules of types A and B, if

$$|\boldsymbol{X}_i^{n+1} - \boldsymbol{X}_j^{n+1}| < \rho,$$

then:

- generate $r \sim \mathcal{U}(0, 1)$,
- if $r < P$, remove both molecules; otherwise either allow them to continue diffusing or reflect them (hard-sphere interaction).

Set $t_{n+1} = t_n + \Delta t$ and repeat.

16.3 From particles to stochastic fields

We continue with the microscopic Brownian model for two species A and B. Particle positions are denoted by $\boldsymbol{X}_i(t) \in \Omega \subset \mathbb{R}^3$ and evolve according to

$$\mathrm{d}\boldsymbol{X}_i(t) = \sqrt{2D_i}\, \mathrm{d}\mathbf{W}_i(t),$$

with independent Brownian motions $\mathbf{W}_i$. Here $D_i = D_A$ if $\boldsymbol{X}_i$ is of type A, and $D_i = D_B$ otherwise. Particles undergo the bimolecular reaction (16.3) with macroscopic rate constant k.

We now pass formally from particle descriptions to continuum fields. Define the empirical number densities

$$u_A(\boldsymbol{x}, t) = \sum_{i \in \mathcal{I}_A(t)} \delta(\boldsymbol{x} - \boldsymbol{X}_i(t)), \qquad u_B(\boldsymbol{x}, t) = \sum_{j \in \mathcal{I}_B(t)} \delta(\boldsymbol{x} - \boldsymbol{X}_j(t)).$$

The densities u_A and u_B represent local particle densities (particles per unit volume), since for example $\int_\Omega u_A(\boldsymbol{x}, t)\mathrm{d}\boldsymbol{x} = A(t)$ gives precisely the number of molecules of A in Ω at time t. Note that they are random measures as $\boldsymbol{X}_i(t)$ is random, and that the sets $\mathcal{I}_A(t), \mathcal{I}_B(t)$ depend on time due to the reaction. Heuristically:

- Independent Brownian motion of particles produces Laplacians at the field level. So we expect diffusions with constants D_A and D_B.
- Reaction events occurring in small space–time cells are random and, when many occur, are approximately Gaussian with variance proportional to the local reaction rate. The well-mixed case has propensity $kA(t)B(t)/\nu$, and writing $A/\nu \to u_A, B/\nu \to u_B$ gives propensity $\nu k u_A u_B$. Thus the local rate per unit volume is $k u_A u_B$.

By direct analogy with the chemical Langevin equation, we are led formally to the reaction–diffusion SPDE system

$$\begin{aligned} \partial_t u_A(\boldsymbol{x},t) &= D_A \Delta u_A(\boldsymbol{x},t) - f(u_A,u_B) + \sqrt{f(u_A,u_B)}\,\eta(\boldsymbol{x},t), \\ \partial_t u_B(\boldsymbol{x},t) &= D_B \Delta u_B(\boldsymbol{x},t) - f(u_A,u_B) + \sqrt{f(u_A,u_B)}\,\eta(\boldsymbol{x},t), \end{aligned} \tag{16.5}$$

where

$$f(u_A,u_B) = k\,u_A u_B,$$

and $\eta(\boldsymbol{x},t)$ denotes a (formal) space–time white noise field with covariance

$$\mathbb{E}[\eta(\boldsymbol{x},t)\eta(\boldsymbol{x}',t')] = \delta(\boldsymbol{x}-\boldsymbol{x}')\delta(t-t').$$

Numerically, we may approximate η by i.i.d. Gaussian variables scaled like

$$\eta_{i,j} \sim \mathcal{N}\left(0, \frac{1}{\Delta x \Delta t}\right),$$

where Δx and Δt are the discretization steps in space and time, respectively. [Compare this with white noise in time only (9.11).] An example of simulated space-time white noise in one dimension is shown in Figure 16.1.

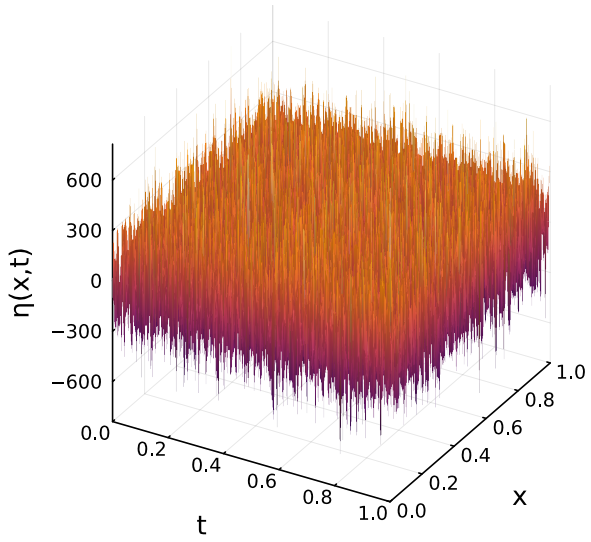

Figure 16.1: Space-time white noise $\eta(x,t)$ in $[0,1]\times[0,1]$ with discretisation $\Delta x = \Delta t = 0.005$.

The multiplicative noise amplitude $\sqrt{f(u_A,u_B)}$ reflects that the instantaneous variance of reaction fluctuations is proportional to the local reaction rate $k\,u_A u_B$. Both equations share the same noise term, since each $A+B$ reaction removes one A and one B simultaneously, inducing perfectly correlated fluctuations between the two species at a given location.

Equation (16.5) is formal. Space–time white noise and distribution-valued densities require careful interpretation and regularisation, which are well beyond the scope of this course. Here, the equation should be understood as a modelling limit rather than a rigorous object.

Neglecting the noise terms in (16.5) recovers a deterministic system of reaction–diffusion PDEs

$$\partial_t u_A = D_A \Delta u_A - k\, u_A u_B, \qquad \partial_t u_B = D_B \Delta u_B - k\, u_A u_B. \tag{16.6}$$

Connection to the RDME. In Lecture 6, we introduced the reaction–diffusion master equation (RDME) by discretising space into compartments of size h and modelling diffusion as nearest-neighbour jumps and reactions within each compartment. At the level of expectations, letting $h \to 0$ led to the deterministic reaction–diffusion PDE of the form (16.6), see for example (6.16). Thus we recover the same modelling hierarchy in space that we previously saw in the well-mixed setting:

$$\text{RDME} \longrightarrow \text{reaction–diffusion SPDE} \longrightarrow \text{reaction–diffusion PDE}.$$

16.4 Unifying perspective

The full modelling hierarchy is now:

	Well-mixed systems	**Spatial systems**
Microscopic	Poisson jump processes (CME)	Brownian dynamics + reactions / RDME
Mesoscopic	Chemical Langevin SDE (CLE)	Reaction–diffusion SPDE
Macroscopic	Reaction-rate ODE (RRE)	Reaction–diffusion PDE

Tasks

Exercise 16.1. Consider again the reaction (16.3), $A + B \to \emptyset$ in $\mathbb{R}^2$ with reaction radius ρ, between two diffusing species. How does the argument to derive the Smoluchowski rate constant (16.4) change in $d = 2$?
Hint: we may want to introduce a finite outer cutoff R where to impose $c(R) = c_\infty$. Why?

Chapter A Markov processes

This appendix collects a number of standard results from Markov processes and Markov chain theory. The material here is *not examinable*, but provides useful background.

A.1 Markov processes: definitions

This section collates important definitions that will be useful for this course. A stochastic process is essentially a random function of a single variable t, often taken to represent time.

Definition A.1 (Stochastic process)**.** A stochastic process is a collection of random variables $\{X(t), t \in T\}$. The set T is the *index set*. The random variables take values in a *state space* S.

The set T can be either discrete, for example, the set of positive integers $\mathbb{Z}_{\geq}$, or continuous, $T = \mathbb{R}_{\geq}$. Again the state space can be discrete (e.g. $S = \mathbb{Z}_{\geq}$) or continuous (e.g. $S = \mathbb{R}$).

A stochastic process $X(t)$ with a discrete state space has an associated probability distribution $\{p_i(t)\}_{i\in\mathbb{Z}}$, known as the *probability mass function* (PMF) of $X(t)$, where

$$p_i(t) = \mathbb{P}(X(t) = i).$$

Similarly, associated with a process $X(t)$ with a continuous state space, say $S = \mathbb{R}$, is a *probability density function* (PDF), $p(x, t)$, such that

$$\mathbb{P}\{X(t) \in [a, b]\} = \int_a^b p(x, t)\mathrm{d}x.$$

Definition A.2 (Markov process)**.** Denote by $\{X(s), s \leq t\}$ the collection of values of the stochastic process up to time t. We say that $X(t)$ is a Markov process if

$$\mathbb{P}(X(t + h) = x \,|\, \{X(s), s \leq t\}) = \mathbb{P}(X(t + h) = x \,|\, X(t)), \tag{A.1}$$

for all $h \geq 0$.

A Markov process with continuous time and discrete state space is called a *continuous-time Markov chain*. Brownian motion is an example of a Markov process with continuous time and continuous state space.

Note that sometimes it is possible to describe a non-Markovian process $X(t)$ in terms of a Markov process $Y(t)$ in a higher-dimensional state space, where the additional variables account for the memory of $X(t)$ (see for example [13, §2.2]).

Definition A.3 (Transition probability)**.** We define the *transition probability* for a continuous-time Markov chain in $\mathbb{Z}$ as

$$p_{ij}(s, t) = \mathbb{P}\{X(t) = j \mid X(s) = i\}, \quad \text{for all } i, j \in \mathbb{Z},\ s < t. \tag{A.2}$$

Definition A.4 (Transition probability density)**.** We define the *transition probability density function* $p(x, t \mid y, s)$ of a continuous-time Markov process in $\mathbb{R}$ as the conditional probability

$$\mathbb{P}(X(t) \in \Gamma \mid X(s) = y) = \int_\Gamma p(x, t \mid y, s)\mathrm{d}x, \quad \text{for all } y, \Gamma \in \mathbb{R},\ s \leq t. \tag{A.3}$$

In other words, $p(x, t \mid y, s)\mathrm{d}x$ is the probability that $X(t) \in [x, x+\mathrm{d}x)$ conditioned on $X(s) = y$.

Definition A.5 (Time-homogeneous Markov process)**.** If the transition probability of a Markov process (discrete or continuous) is invariant to time shifts, that is, does not depend on s or t but only on the length of the time interval, $t-s$, it is called a *time-homogeneous* Markov process:

$$p_{ij}(s,t) = p_{ij}(0,t-s), \qquad \text{for all } i,j \in \mathbb{Z},\ s \le t,$$
$$p(x,t \mid y,s) = p(x,t-s \mid y,0), \qquad \text{for all } x,y \in \mathbb{R},\ s \le t.$$

The Ornstein–Uhlenbeck process and Brownian motion are examples of time-homogeneous Markov processes.

The Markov property (A.1) enables us to obtain an evolution equation for the transition probability (A.2) or the transition probability density (A.3), called the *Chapman–Kolmogorov equation*.

Definition A.6 (Chapman–Kolmogorov equations)**.** The transition probability of a continuous-time Markov chain in $\mathbb{Z}$ satisfies the Chapman–Kolmogorov (CK) equation

$$p_{ij}(s,t) = \sum_{k\in\mathbb{Z}} p_{kj}(u,t)\, p_{ik}(s,u), \qquad s \le u \le t,\ \forall i,j \in \mathbb{Z}. \tag{A.4}$$

Similarly, the transition probability density of a Markov process in $\mathbb{R}$ satisfies the CK equation

$$p(x,t \mid y,s) = \int_{\mathbb{R}} p(x,t \mid z,u)\, p(z,u \mid y,s)\mathrm{d}z, \qquad s \le u \le t,\ \forall x,y \in \mathbb{R}. \tag{A.5}$$

The derivation of the CK equation (A.5) is based on the Markovian assumption and properties of conditional probability (see [13, p.35]).

Consider a Markov process $X(t)$ in $\mathbb{R}$ with initial density $p(x,0) := \rho_0(x)$. We have that

$$p(x,t) = \int_{\mathbb{R}} p(x,t \mid y,0)\, p(y,0)\mathrm{d}y = \int_{\mathbb{R}} p(x,t \mid y,0)\, \rho_0(y)\mathrm{d}y. \tag{A.6}$$

In applications, it is typical for the initial condition to be deterministic, $X(0) = x_0$, so that the initial density is a Dirac delta function $\rho_0(x) = \delta(x - x_0)$. In this case, $p(x,t)$ coincides with the transition probability density $p(x,t \mid x_0, 0)$.

Definition A.7 (Continuity of a Markov process)**.** The sample paths of a Markov process $X(t)$ taking values in $\mathbb{R}$ are continuous functions of t with probability one if, for any $\epsilon > 0$,

$$\lim_{\Delta t \to 0^+} \frac{1}{\Delta t} \int_{|x-y|>\epsilon} p(x, t+\Delta t \mid y, t)\mathrm{d}x = 0, \tag{A.7}$$

uniformly in y, t and Δt.

Continuous state space vs continuous sample paths Whether a stochastic process $X(t)$ has a continuous state space, namely a continuous range of possible values, is an entirely different question from whether the sample path of $X(t)$ is a continuous function of t. For example, Brownian motion is an example of a continuous state space Markov process with continuous sample paths. However, suppose that we model the molecules of a gas as hard spheres undergoing instantaneous elastic collisions (as often done in kinetic theory) and denote their velocities by $\mathbf{V}(t)$. Then all possible values of $\mathbf{V}(t)$ are, in principle, realisable so that the range or state-space of velocities is continuous. But the velocity of a molecule involved in a collision will change from its pre-collision value to its post-collision value instantaneously at the time of the collision, so the sample path of $\mathbf{V}(t)$ is not continuous.

A.2 Markov chain theory

This section records a small number of standard definitions and facts from Markov chain theory. Unless stated otherwise, we consider a time-homogeneous continuous-time Markov chain (CTMC) $X(t)$ with discrete state space S.

Definition A.8 (Finite and infinite state spaces)**.** The state space S is *finite* if $|S| < \infty$, and *infinite* otherwise.

Many qualitative properties of Markov chains depend crucially on whether S is finite. In particular, in the finite-state setting, explosion is impossible and stationary distributions always exist. In contrast, in the infinite-state setting, explosion may occur and stationary distributions may fail to exist unless additional recurrence assumptions are satisfied.

Definition A.9 (Jump times and explosion)**.** Let $T_0 := 0$ and define recursively the *nth jump time* by

$$T_n := \inf\{t > T_{n-1} : X(t) \neq X(T_{n-1})\}, \qquad n \geq 1.$$

We say that the CTMC *explodes* if it performs infinitely many jumps in finite time, i.e. if

$$\mathbb{P}\Big(\lim_{n\to\infty} T_n < \infty\Big) > 0.$$

Definition A.10 (Communication)**.** Two states $x, y \in S$ are said to *communicate* if

$$\mathbb{P}\big(X(t) = y \text{ for some } t > 0 \,\big|\, X(0) = x\big) > 0 \quad \text{and} \quad \mathbb{P}\big(X(t) = x \text{ for some } t > 0 \,\big|\, X(0) = y\big) > 0.$$

We write $x \leftrightarrow y$.

Communication is an equivalence relation on S; its equivalence classes are called *communicating classes*.

Definition A.11 (Closed communicating class)**.** A communicating class $C \subset S$ is said to be *closed* if

$$\mathbb{P}\left(X(t) \in C \quad \forall t \geq 0 \mid X(0) = x\right) = 1 \qquad \forall x \in C,$$

that is, no transitions leave C.

Definition A.12 (Irreducibility)**.** A CTMC is *irreducible* if its state space S consists of a single communicating class.

In particular, a system with more than one state and an absorbing state cannot be irreducible.

Definition A.13 (Generator)**.** If S is finite, the generator of a CTMC is an $N \times N$ matrix $Q = (q_{xy})_{x,y\in S}$, where $N = |S|$, satisfying

$$q_{xy} \geq 0 \ \text{ for } x \neq y, \qquad q_{xx} = -\sum_{y\neq x} q_{xy}.$$

In particular, each row of Q sums to zero:

$$\sum_{y\in S} q_{xy} = 0 \qquad \forall x \in S, \tag{A.8}$$

or equivalently $Q\mathbf{1} = 0$, where $\mathbf{1}$ denotes the vector with all entries equal to one.

Definition A.14 (Stationary distribution)**.** Assume S is finite with generator Q. A probability distribution ϕ on S is called *stationary* if

$$\phi^\top Q = 0, \qquad \sum_{x \in S} \phi(x) = 1. \tag{A.9}$$

If $X(0)$ is distributed according to a stationary distribution, then $X(t)$ has the same distribution for all $t \geq 0$. Note that since P and Q commute (from the backward and forward Kolmogorov equations (4.8), (4.10)), then the stationary distribution also satisfies $Q\phi = 0$.

Existence and uniqueness (finite state space). On a finite state space, at least one stationary distribution always exists. If the CTMC is irreducible, the stationary distribution is unique. If the CTMC is not irreducible, stationary distributions are not unique: each closed communicating class admits a stationary distribution, and any convex combination of these stationary distributions is again stationary.

Linear-algebraic characterisation ($|S| = N < \infty$). Let $|S| = N$ and let Q be the generator. The row-sum property (A.8) implies

$$\operatorname{rank}(Q) \leq N - 1.$$

Moreover,

$$\phi^\top Q = 0 \quad \Longleftrightarrow \quad \phi \in \ker Q^\top,$$

so stationary distributions are precisely normalised elements of the left nullspace of Q.

The dimension of $\ker Q^\top$ has a direct probabilistic interpretation: it equals the number of closed communicating classes. In particular,

$$\dim \ker Q^\top = 1 \quad \Longleftrightarrow \quad X(t) \text{ is irreducible,}$$

and in this case

$$\operatorname{rank}(Q) = N - 1.$$

Uniqueness via normalisation. Assume that the CTMC is irreducible, so that $\operatorname{rank}(Q) = N - 1$. Then the linear system

$$\phi^\top Q = 0$$

has a one-dimensional space of solutions. The stationary distribution is obtained by imposing the normalisation condition $\mathbf{1}^\top \phi = 1$.

Equivalently, one may replace one of the equations in $Q\phi = 0$ by the normalisation condition. For instance, if the last row of Q is linearly dependent on the first $N - 1$ rows, one may solve

$$\begin{pmatrix} q_{11} & \cdots & q_{1N} \\ \vdots & & \vdots \\ q_{(N-1)1} & \cdots & q_{(N-1)N} \\ 1 & \cdots & 1 \end{pmatrix} \phi = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

By construction, this linear system has a unique solution, which is the unique stationary distribution of the irreducible CTMC.

Bibliography

[1] S. S. Andrews and D. Bray, *Stochastic simulation of chemical reactions with spatial resolution and single molecule detail*, Physical biology, 1 (2004), p. 137.

[2] L. Bachelier, *Théorie de la spéculation*, Ann. Sci. Ecole Norm. S., 3 (1900), pp. 21–86.

[3] R. Brown, *A brief description of microscopical observations made in the months of June, July and August 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies*, Edinburgh Philos. J., (1828), pp. 358–371.

[4] A. Einstein, *Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen*, Ann. Phys., 322 (1905), pp. 549–560.

[5] R. Erban and S. J. Chapman, *Stochastic modelling of reaction–diffusion processes*, vol. 60, Cambridge University Press, 2020.

[6] C. Gardiner, *Stochastic Methods*, Springer Berlin, Heidelberg, 2009.

[7] D. Higham and P. Kloeden, *An introduction to the numerical simulation of stochastic differential equations*, SIAM, 2021.

[8] D. J. Higham, *An algorithmic introduction to numerical simulation of stochastic differential equations*, SIAM Review, 43 (2001), pp. 525–546.

[9] T. G. Kurtz, *Limit theorems for sequences of jump markov processes approximating ordinary differential processes*, Journal of Applied Probability, 8 (1971), pp. 344–356.

[10] P. Langevin, *Sur la théorie du mouvement Brownien*, C. R. Acad. Sci. Paris, 146 (1908).

[11] R. M. Mazo, *Brownian Motion: Fluctuations, Dynamics, and Applications*, Clarendon Press, Oxford, 2002.

[12] H. G. Othmer, S. R. Dunbar, and W. Alt, *Models of dispersal in biological systems*, Journal of Mathematical Biology, 26 (1988), pp. 263–298.

[13] G. A. Pavliotis, *Stochastic processes and applications*, vol. 60, Springer, 2014.

[14] R. Pawula, *Approximation of the linear boltzmann equation by the fokker-planck equation*, Physical review, 162 (1967), p. 186.

[15] G. G. Szpiro, *Pricing the Future: Finance, Physics, and the 300-Year Journey to the Black–Scholes Equation: A Story of Genius and Discovery*, Basic Books, New York, 2011.

[16] M. von Smoluchowski, *Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen*, Ann. Phys., 326 (1906), pp. 756–780.