Part III course "Stochastic Processes in Biology"

Maria Bruna & Tomislav Plesa

Lent Term 2024

Contents

File last modified March 12, 2024

Lecture 10: Brownian motion

10.0.1 Historical perspective

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 [4]. 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 [3] 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' [23]), 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 [24]. 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 [8]. 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 [15]. 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 [16, 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.

10.1 Brownian motion as the limit of one-dimensional random walks

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. Consider the following stochastic process in $\mathbb{Z}$ at discrete times $n = 0, 1, 2, \dots$

$$X_n = \sum_{j=1}^{n} \xi_j, \qquad X_0 = 0. \tag{10.1}$$

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 shits).
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 $X_m - X_n \sim X_{m-m}$.

(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 on a computer (see Figure 10.1). We notice that, if we take a large number of steps, the random walk starts looking like a continuous time process with continuous paths.

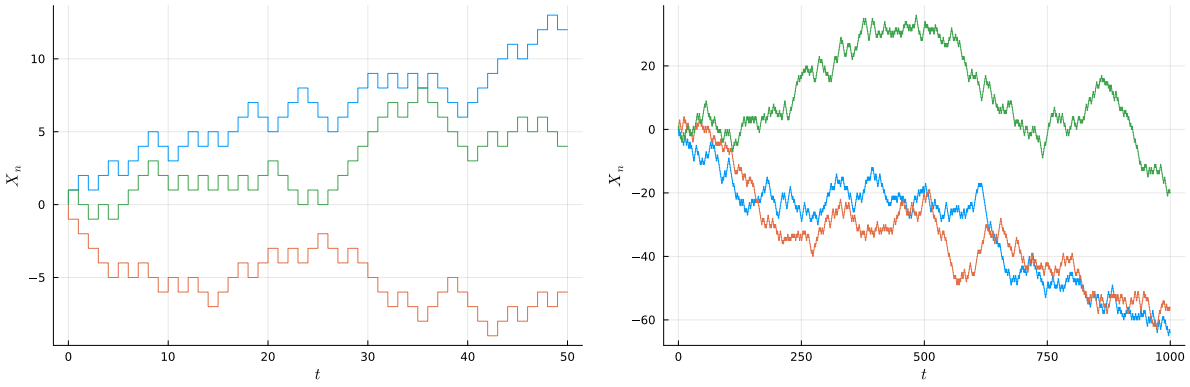

Figure 10.1: Three paths of the random walk (10.1) X_n of length $n = 50$ (left) and $n = 1000$ (right).

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 *continuous time* rescaled process is

$$\widetilde{X}(t) = \Delta x X_{\lfloor t/\Delta t \rfloor} = \Delta x \left(\xi_1 + \cdots + \xi_{\lfloor t/\Delta t \rfloor}\right), \tag{10.2}$$

where $\lfloor t/\Delta t \rfloor$ means the largest integer less than or equal to $t/\Delta t$. 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:

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

This tells us that we should pick $\Delta x^2 \sim \Delta t$ as $\Delta t, \Delta x \to 0$. The time and space scaling $t \sim x^2$ is known as the *diffusion or parabolic scaling*. This is an important point: in Brownian motion,

space scales as the square root of time. (Note also that this scaling could be directly inferred from point (v) above.)

This motivates the choice $\Delta t = 1/N$ and $\Delta x = 1/\sqrt{N}$ and the definition

$$W^N(t) = \frac{X_{\lfloor Nt \rfloor}}{\sqrt{N}}. \tag{10.3}$$

The *Donsker's Theorem* or *Donsker's Invariance Principle* says that as $N \to \infty$, $W^N(t)$ converges (in some appropriate sense) to a Wiener process or Brownian motion $W(t)$. (This is like the Central Limit Theorem extended to functions.)

10.2 Properties of Brownian motion

Definition 10.1 (Brownian motion)**.** *A stochastic process $W(t) : [0, \infty) \to \mathbb{R}$ is a one-dimensional Wiener process (standard Brownian motion) if $W(t)$ depends continuously on t, and the following three conditions hold:*

(i) $W(0) = 0$.

(ii) *$W(t)$ 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.*

(iii) *$W(t)$ has Gaussian stationary increments: for $0 \leq t_1 < t_2 < \infty$, $W(t_2) - W(t_1)$ is normally distributed with mean zero and variance $t_2 - t_1$; that is, $W(t_2) - W(t_1) \sim \mathcal{N}(0, t_2 - t_1)$.*

Note that Gaussianity (iii) implies that

$$\mathbb{P}(W(t) \in [x_1, x_2)) = \int_{x_1}^{x_2} \rho(x, t)\mathrm{d}x, \tag{10.4}$$

$$\mathbb{P}(W(t) \in [x, x + \mathrm{d}x) | W(s) = y) = \rho(x - y, t - s)\mathrm{d}x, \tag{10.5}$$

with

$$\rho(x, t) = \frac{e^{-x^2/(2t)}}{\sqrt{2\pi t}}. \tag{10.6}$$

Mean and covariance of $W(t)$ The standard Brownian motion has

- Mean
$$m(t) = 0.$$
Since $W(t) \sim \mathcal{N}(0, t)$, we have that $m(t) = \mathbb{E}W(t) = 0$.

- Covariance function
$$C(t, s) = \min(t, s). \tag{10.7}$$
Suppose that $s < t$
$$\begin{aligned} C(t, s) &= \mathbb{E}\left\{[W(t) - \mathbb{E}W(t)][W(s) - \mathbb{E}W(s)]\right\} = \mathbb{E}[W(t)W(s)] \\ &= \mathbb{E}[(W(t) - W(s) + W(s))W(s)] = \mathbb{E}[(W(t) - W(s))W(s)] + \mathbb{E}[W(s)^2] = s. \end{aligned}$$
In the second line, we have used that $W(t) - W(s)$ is independent of $W(s)$ (property (ii)) in the second equality and property (iii) in the third.

- It is also useful to remember that
$$\mathbb{E}(W(t) - W(s))^2 = |t - s|. \tag{10.8}$$
Note that $\mathbb{E}(W(t) - W(s))^2 = C(t, t) + C(s, s) - 2C(t, s)$ and use (10.7).

Differentiability of $W(t)$ An important property of Brownian motion is its lack of differentiability: $W(t)$ is continuous everywhere but nowhere differentiable. There is a theorem stating this, but for this course, we will just give some heuristic explanations to see why the derivative does not exist:

- Suppose we try to calculate the derivative as

$$\xi(t) = \frac{dW(t)}{dt} = \lim_{h\to 0} \frac{W(t+h) - W(t)}{h}. \tag{10.9}$$

  But $W(t+h) - W(t) \sim \mathcal{N}(0, h)$ so $\frac{W(t+h)-W(t)}{h} \sim N(0, 1/h)$, which does not converge in any way 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 $t \to s^-$. Therefore, $W(t)$ "cannot remember" how to leave y in such as way that there will be a tangent there.

10.3 Probabilistic description of Brownian motion

Einstein studied the Brownian motion phenomena describing the fluctuations in the particle's position probabilistically. Here we consider a variant of Einstein's argument starting from a discrete-time discrete-space stochastic process. To this end, let's consider the rescaled process $\widetilde{X}(t)$ (10.2) on the state space $\{i\Delta x\}_{i\in\mathbb{Z}}$ at times $\{n\Delta t\}_{n\in\mathbb{N}}$ again. In other words, we consider a two-dimensional rectangular lattice comprising the sites

$$\{(i\Delta x, n\Delta t) | i = 0, \pm 1, \pm, \dots; n = 0, 1, 2, \dots\}$$

for some given spacing $\Delta x > 0$ and time step $\Delta t > 0$.

Consider a particle starting at $x = 0$ at time $t = 0$, and at each time step, the particle moves Δx to the left with probability $1/2$, and Δx to the right with probability $1/2$. Define

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

Then

$$P(i, 0) = \begin{cases} 0, & i \neq 0 \\ 1, & i = 0, \end{cases}$$

and P satisfies the recurrence relation

$$P(i, n+1) = \frac{1}{2}P(i-1, n) + \frac{1}{2}P(i+1, n) \tag{10.10}$$

or

$$P(i, n+1) - P(i, n) = \frac{1}{2}\left[P(i-1, n) - 2P(i, n) + P(i+1, n)\right]]$$

Now assume the diffusion time and space scaling

$$\frac{(\Delta x)^2}{\Delta t} = 2D, \tag{10.11}$$

for some positive constant D. This implies

$$\frac{P(i,n+1)-P(i,n)}{\Delta t}=D\frac{[P(i-1,n)-2P(i,n)+P(i+1,n)]]}{(\Delta x)^2}. \tag{10.12}$$

Letting $\Delta t, \Delta x \to 0$, $i\Delta x \to x$, $n\Delta t \to t$ with $D=\Delta x^2/(2\Delta t)$, we have $P(i,n)\to p(x,t)$, which now represents the probability density that the particle is at x at time t (a continuous-time and continuous state process). The difference equation above becomes formally in the limit

$$\frac{\partial p}{\partial t}=D\frac{\partial^2 p}{\partial x^2}. \tag{10.13}$$

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 (1.14)). Note that in §10.1 we had $\Delta x=\sqrt{\Delta t}$, which corresponds to $D=1/2$. Therefore, the Fokker–Planck equation of Brownian motion is a diffusion equation with $D=1/2$. The solution to (10.13) with $D=1/2$ and the initial condition $p(x,0)=\delta_0(x)$ (the Dirac point mass at the origin) is (10.6).

Lecture 11: Markov processes and Kolmogorov equations

11.1 Continuous Markov processes

Definition 11.1 (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 $\epsilon > 0$,*

$$\lim_{\Delta t \to 0^+} \frac{1}{\Delta t} \int_{|x-y|>\epsilon} p(x, t + \Delta t|y, t)\mathrm{d}x = 0, \tag{11.1}$$

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; see §10.2. 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.

Exercise. *Check that the transition probability density $p(x, t + \Delta t|y, t)$ of Brownian motion satisfies (11.1), thus confirming it has continuous sample paths. Compare this with the Cauchy process, which is also a continuous state-space Markov process with transition probability density*

$$p(x, t + \Delta t|y, t) = \frac{\Delta t}{\pi\left[(x-y)^2 + \Delta t^2\right]}.$$

Simulate one trajectory of each of these two processes for $t = [0, 1]$ and observe if the results are consistent with what you expect (you should get something like Figure 11.1).

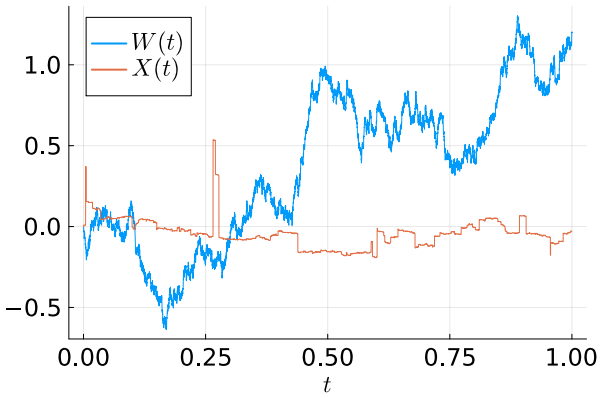

Figure 11.1: Illustration of sample paths of the Cauchy process $X(t)$ and Brownian motion $W(t)$.

11.2 The forward Kolmogorov equation

Recall the Chapman–Kolmogorov equation satisfied by a Markov process in $\mathbb{R}$

$$p(x,t|y,s) = \int_{\mathbb{R}} p(x,t|z,u)p(z,u|y,s)\mathrm{d}z, \qquad s \leq u \leq t, \forall x,y \in \mathbb{R}, \tag{11.2}$$

which is valid for any time interval $\Delta t = t - s$. One can derive a differential form of this equation by taking the limit $\Delta t \to 0^+$. The result is a differential equation known as the forward Kolmogorov equation that determines the transition probability from its local in-time behaviour under appropriate assumptions.

Consider a Markov process $X(t)$ and assume that its transition probability density $p(x,t|y,s)$ is a smooth function of x, y and that, additionally, it satisfies the following conditions *for all* $\epsilon > 0$ [1]:

(i) (**Jump rate**) For every x, y such that $|x - y| \geq \epsilon$

$$\lim_{\Delta t \to 0^+} \frac{p(x, t + \Delta t|y,t)}{\Delta t} = W(x|y,t). \tag{11.3}$$

(ii) (**Drift coefficient**)

$$\lim_{\Delta t \to 0^+} \frac{1}{\Delta t} \int_{|x-y|\leq\epsilon} (x-y)p(x,t+\Delta t|y,t)\mathrm{d}x = a(y,t) + O(\epsilon), \tag{11.4}$$

(iii) (**Diffusion coefficient**)

$$\lim_{\Delta t \to 0^+} \frac{1}{\Delta t} \int_{|x-y|\leq\epsilon} (x-y)^2 p(x,t+\Delta t|y,t)\mathrm{d}x = b(y,t) + O(\epsilon). \tag{11.5}$$

Comparing (11.1) with (11.3), the process can only have continuous paths if $W(x|y,t)$ is zero for all $x \neq y$. Conversely, if $W(x|y,t)$ is finite, the process has jumps of at least size ϵ.

Exercise. *Show that all higher-order coefficients of the form in* (11.4)-(11.5)

$$C_\delta = \lim_{\Delta t \to 0^+} \frac{1}{\Delta t} \int_{|x-y|\leq\epsilon} (x-y)^{2+\delta} p(x,t+\Delta t|y,t)\mathrm{d}x,$$

for $\delta > 0$ *vanish. [*Hint: consider $|C_\delta|$ for, e.g., $\delta = 1$.*]*

Consider a test function $f \in C_0^2(\mathbb{R})$ and its expectation conditional on $X(s) = y$:

$$u(y,t) = \mathbb{E}[f(X(t))|X(s) = y] = \int_{\mathbb{R}} f(x)p(x,t|y,s)\mathrm{d}x.$$

The time derivative of the above is

$$\begin{aligned}\partial_t u(y,t) &= \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int f(x) \left[p(x,t+\Delta t|y,s) - p(x,t|y,s)\right] \mathrm{d}x \\ &= \lim_{\Delta t \to 0} \frac{1}{\Delta t} \left[\iint f(x)p(x,t+\Delta t|z,t)p(z,t|y,s)\mathrm{d}z\mathrm{d}x - \int f(z)p(z,t|y,s)\mathrm{d}z \right]. \end{aligned} \tag{11.6}$$

In the last line, we have used the Chapman–Kolmogorov equation (11.2) in the first term and renamed the integration variable in the second. Now divide the integral over x over two regions $|x - z| \geq \epsilon$ and $|x - z| < \epsilon$ and in the latter insert

$$f(x) = f(z) + f'(z)(x-z) + \frac{1}{2}f''(z)(x-z)^2(1+\alpha_\epsilon), \qquad |x-z| < \epsilon,$$

with $\lim_{\epsilon\to 0}\alpha_\epsilon = 0$. Then

$$\begin{aligned}\partial_t u = \lim_{\Delta t\to 0}\frac{1}{\Delta t}\Bigg\{ &\iint_{|x-z|<\epsilon}\left[f'(z)(x-z)+\frac{1}{2}f''(z)(x-z)^2(1+\alpha_\epsilon)\right]p(x,t+\Delta t|z,t)p(z,t|y,s)\mathrm{d}x\mathrm{d}z\\ &+\iint_{|x-z|<\epsilon} f(z)p(x,t+\Delta t|z,t)p(z,t|y,s)\mathrm{d}x\mathrm{d}z\\ &+\iint_{|x-z|\geq\epsilon} f(x)p(x,t+\Delta t|z,t)p(z,t|y,s)\mathrm{d}x\mathrm{d}z\\ &-\iint f(z)p(x,t+\Delta t|z,t)p(z,t|y,s)\mathrm{d}x\mathrm{d}z\Bigg\}.\end{aligned}$$

In the last line, we have inserted $\int p(x,t+\Delta t|z,t)\mathrm{d}x$ in the negative term of (11.6) using the fact that it integrates to one. Now using conditions (11.3), (11.4) and (11.5) we obtain

$$\begin{aligned}\partial_t u = &\int\left[f'(z)a(z,t)+\frac{1}{2}f''(z)b(z,t)(1+\alpha_\epsilon)\right]p(z,t|y,s)\mathrm{d}z\\ &+\iint_{|x-z|\geq\epsilon} f(z)\left[W(z|x,t)p(x,t|y,s)-W(x|z,t)p(z,t|y,s)\right]\mathrm{d}x\mathrm{d}z.\end{aligned}$$

Taking the limit $\epsilon\to 0$ we obtain[1]

$$\begin{aligned}\int f(z)\partial_t p(z,t|y,s)\mathrm{d}z = &\int\left[f'(z)a(z,t)+\frac{1}{2}f''(z)b(z,t)\right]p(z,t|y,s)\mathrm{d}z\\ &+\iint f(z)\left[W(z|x,t)p(x,t|y,s)-W(x|z,t)p(z,t|y,s)\right]\mathrm{d}x\mathrm{d}z.\end{aligned}$$

Performing two integrations by parts and using the fact that, since the test function f has compact support, the boundary terms vanish, we obtain

$$\begin{aligned}\int f(z)\partial_t p(z,t|y,s)\mathrm{d}z = \int f(z)\Bigg\{ &-\frac{\partial}{\partial z}[a(z,t)p(z,t|y,s)]+\frac{1}{2}\frac{\partial}{\partial z^2}[b(z,t)p(z,t|y,s)]\\ &+\int\left[W(z|x,t)p(x,t|y,s)-W(x|z,t)p(z,t|y,s)\right]\mathrm{d}x\Bigg\}\mathrm{d}z.\end{aligned}$$

Since, in the equation above, the test function is arbitrary, we conclude that

$$\begin{aligned}\frac{\partial}{\partial t}p(z,t|y,s) = &\frac{\partial}{\partial z}\left\{-a(z,t)p(z,t|y,s)+\frac{1}{2}\frac{\partial}{\partial z}[b(z,t)p(z,t|y,s)]\right\}\\ &+\int\left[W(z|x,t)p(x,t|y,s)-W(x|z,t)p(z,t|y,s)\right]\mathrm{d}x. \quad (11.7)\end{aligned}$$

The partial differential equation (11.7) is the *forward Kolmogorov equation* for a Markov process satisfying conditions (i)-(iii) above. Each of the conditions (11.3), (11.4), (11.5) can be seen to give rise to a specific part of the forward Kolmogorov equation (11.7), namely a jump process, a drift and a random part, respectively. It is called forward as it tells us about the evolution of $p(x,t|y,s)$ forward in time or involves the *forward variables* x,t. Later, we will consider the backward Kolmogorov equation, which is a PDE involving derivatives with respect to the *backward variables* y,s.

[1]Here we assume that $W(x|y,t)$ is not singular at $x=y$; otherwise one must replace the integrals with W by principal value integrals [10, §3.4.1]

11.3 Special cases of the forward Kolmogorov equation

There are two important special cases of the Kolmogorov equation (11.7):

- When $a = b = 0$, equation (11.7) reduces to the *master equation*

$$\frac{\partial}{\partial t}p(x,t|y,s) = \int [W(x|z,t)p(z,t|y,s) - W(z|x,t)p(x,t|y,s)]\,\mathrm{d}z. \tag{11.8}$$

  In the case where the state space is discrete, say the process can only take integer values, then the master equation takes the form

$$\frac{\partial}{\partial t}P(i,t|j,s) = \sum_k [W(i|k,t)P(k,t|j,s) - W(k|i,t)P(i,t|j,s)]\,. \tag{11.9}$$

- When the Markov process has no jumps, $W(x|y,t) = 0$, and conditions (11.4) and (11.5) hold, then we call it a *diffusion process* and the forward Kolmogorov equation, in this case, is known as the *Fokker–Planck equation.*

Definition 11.2 (Diffusion process)**.** *A Markov process $X(t)$ in $\mathbb{R}$ with transition probability $p(x,t|y,s)$ is called a* diffusion process *if* (11.1)*,* (11.4) *and* (11.5) *are satisfied.*

In other words, a diffusion process is a Markov process with continuous sample paths and additional properties regarding its infinitesimal mean and variance.

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 Fokker–Planck equation

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}[a(x,t)p] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[b(x,t)p], \qquad \lim_{t\to s} p(x,t|y,s) = \delta(x-y). \tag{11.10}$$

An example of a jump process arises in modelling the motion of organisms just as flagellated bacteria (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 [19]. In the simplest model, one assumes that bacteria move at a constant speed during a run, leading to the forward Kolmogorov equation

$$\frac{\partial p}{\partial t}(\mathbf{x},\mathbf{v},t) + \mathbf{v}\cdot\nabla_{\mathbf{x}}p = -\lambda p + \lambda\int_V W(\mathbf{v}|\mathbf{w},t)p(\mathbf{x},\mathbf{w},t)\mathrm{d}\mathbf{w}, \tag{11.11}$$

in state space $(\mathbf{x},\mathbf{v}) \in \mathbb{R}^2 \times V$. Here $W(\mathbf{v}|\mathbf{w},t)$ is called the turning kernel; it denotes 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 $W = 1/|V|$.

Another example of a Markov process described by a continuous part superimposed to a jump process is when modelling stock prices. The typical model for stock pricing is the *Geometric Brownian motion*

$$\mathrm{d}S(t)/S(t) = \mu\mathrm{d}t + \sigma\mathrm{d}W(t), \tag{11.12}$$

where $S(t) \geq 0$ is the stock price. Now assume that important pieces of information about the stock arrive periodically, which cause a jump in the stock price. This was considered in [17], modelling the news arrival times by a Poisson process.

Lecture 12: Diffusion processes and stochastic differential equations

12.1 Diffusion processes with finite first and second moments

In (11.4), (11.5), we truncated the domain of integration as we did not make any assumptions on whether the first and second moments of $X(t)$ are finite. However, if we assume that there exists $\delta > 0$ such that

$$\lim_{\Delta t \to 0^+} \frac{1}{\Delta t} \int_{\mathbb{R}} |x-y|^{2+\delta} p(x, t+\Delta t|y,t) \mathrm{d}x = 0, \tag{12.1}$$

then it can be shown that (see [20, §2.5])

$$\lim_{\Delta t \to 0^+} \frac{1}{\Delta t} \int_{|x-y|>\epsilon} |x-y|^{k} p(x, t+\Delta t|y,t) \mathrm{d}x = 0, \qquad k = 0, 1, 2.$$

This implies that (12.1) is sufficient for the sample paths to be continuous ($k = 0$) and for replacing the truncated integrals in (11.4), (11.5) to the whole domain (for $k = 1, 2$). This means that drift and diffusion coefficients can then be expressed in terms of expectations:

$$\lim_{\Delta t \to 0} \mathbb{E}\left(\frac{X(t+\Delta t) - X(t)}{\Delta t} \Big| X(t) = y \right) = a(y,t), \tag{12.2}$$

$$\lim_{\Delta t \to 0} \mathbb{E}\left(\frac{|X(t+\Delta t) - X(t)|^2}{\Delta t} \Big| X(t) = y \right) = b(y,t). \tag{12.3}$$

These are often easier to check than the truncated (weaker) conditions.

Exercise. *Use the transition probability density of Brownian motion (see (10.6)),*

$$p(x,t|y,s) = \frac{1}{\sqrt{2\pi(t-s)}} \exp\left(-\frac{(x-y)^2}{2(t-s)} \right), \tag{12.4}$$

to show that it satisfies conditions (11.1), (11.4) and (11.5). Further show that the stronger conditions on the infinitesimal mean and variance (12.1), (12.2), (12.3) with $\delta = 2$ are also satisfied.

A diffusion process is characterised by the (almost sure) continuity of its paths and by specifying the first two moments. A natural question is whether other types of stochastic processes can be defined by specifying a fixed number of moments higher than two. It turns out that this is not possible: we either need to retain two or all (i.e. infinitely many) moments. Specifying a finite number of moments greater than two leads to inconsistencies [20, §2.6]. A more or less equivalent definition, and the more important one for applications, is that a diffusion process is a process that satisfies a stochastic differential equation (SDE); see §12.4 below.

12.2 The backward Kolmogorov equation

The backward Kolmogorov equation for a diffusion process $X(t)$ can be derived based on assumptions (12.1), (12.2), (12.3) (or the weaker assumptions (11.1), (11.4) and (11.5)) similarly to the forward Kolmogorov equation, by considering $p(x,t|y,s-\Delta s)$ and Taylor expanding around y [instead of $p(x,t+\Delta t|y,s)$ and expanding around x]. We will leave the derivation as an exercise. The result is

$$-\frac{\partial}{\partial s} p(x,t|y,s) = a(y,s) \frac{\partial p(x,t|y,s)}{\partial y} + \frac{1}{2} b(y,s) \frac{\partial^2 p(x,t|y,s)}{\partial y^2}, \qquad s < t, \tag{12.5}$$

and *final condition* $\lim_{s\to t} p(x,t|y,s) = \delta(x-y)$.

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, $a(y,t) = a(y)$ and $b(y,t) = b(y)$. 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 $t > 0$

$$\frac{\partial p}{\partial t} = a(y)\frac{\partial p}{\partial y} + \frac{1}{2}b(y)\frac{\partial^2 p}{\partial y^2}, \tag{12.6}$$

with *initial condition* $\lim_{t\to 0} p(x,t|y,0) = \delta(x-y)$.

12.3 Kolmogorov equations for multivariate diffusion processes

It is common to see the forward and backward Kolmogorov equations (11.10) and (12.5) written using operator notation. We will use the following notation:

Kolmogorov equations for a diffusion process

Let $X(t)$ be diffusion process in $\mathbb{R}$ and assume that $p(x,t|y,s), a(x,t), b(x,t) \in C^{2,1}(\mathbb{R}\times\mathbb{R}_>)$. Then the transition probability density satisfies the backward Kolmogorov equation

$$-\frac{\partial p}{\partial s} = \mathcal{L}_{y,s}\, p,$$

where

$$\mathcal{L}_{y,s} p := a(y,s)\frac{\partial p}{\partial y} + \frac{1}{2}b(y,s)\frac{\partial^2 p}{\partial y^2}, \tag{12.7}$$

and the forward Kolmogorov equation (or Fokker–Planck equation)

$$\frac{\partial p}{\partial t} = \mathcal{L}^*_{x,t}\, p,$$

where

$$\mathcal{L}^*_{x,t}\, p := -\frac{\partial}{\partial x}[a(x,t)p] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[b(x,t)p]. \tag{12.8}$$

The forward and backward Kolmogorov equations can be derived for multidimensional diffusion processes following similar calculations we used for the one-dimensional case. Let $\mathbf{X}(t)$ be a diffusion process in $\mathbb{R}^d$. We will assume its first and second moments exist, so we can define its drift vector and diffusion matrix as[2]

$$\mathbf{a}(\mathbf{y},t) = \lim_{\Delta t\to 0} \mathbb{E}\left(\frac{\mathbf{X}(t+\Delta t)-\mathbf{X}(t)}{\Delta t}\middle|\mathbf{X}(t)=\mathbf{y}\right), \tag{12.9}$$

$$\mathbf{B}(\mathbf{y},t) = \lim_{\Delta t\to 0} \mathbb{E}\left(\frac{[\mathbf{X}(t+\Delta t)-\mathbf{X}(t)]\otimes[\mathbf{X}(t+\Delta t)-\mathbf{X}(t)]}{\Delta t}\middle|\mathbf{X}(t)=\mathbf{y}\right). \tag{12.10}$$

Here $\mathbf{a} = (a_1,\ldots,a_d)$ is a d-dimensional vector and that the diffusion coefficient $\mathbf{B} = (B_{ij})$ is a $d\times d$ *symmetric* nonnegative matrix.

[2]Of course, it is still possible to use the analogous weaker assumptions for the drift and diffusion coefficients for processes without finite first and second moments.

Kolmogorov equations for a multidimensional diffusion process

Let $\mathbf{X}(t)$ be diffusion process in $\mathbb{R}^d$ with drift vector $\mathbf{a}(\mathbf{x},t) \in \mathbb{R}^d$ in (12.9) and diffusion matrix $B(\mathbf{x},t) \in \mathbb{R}^{d\times d}$ in (12.10). Then the transition probability density $p(\mathbf{x},t|\mathbf{y},s)$ satisfies the backward Kolmogorov equation

$$\begin{aligned} -\frac{\partial p}{\partial s} &= \mathbf{a}(\mathbf{y},s)\cdot\nabla p + \frac{1}{2}\mathbf{B}(\mathbf{y},s) : \nabla\nabla p \\ &= \sum_{i=1}^{d} a_i(\mathbf{y},s)\frac{\partial p}{\partial y_i} + \frac{1}{2}\sum_{i,j=1}^{d} B_{ij}(\mathbf{y},s)\frac{\partial^2 p}{\partial y_i \partial y_j}. \end{aligned} \tag{12.11}$$

The corresponding forward Kolmogorov equation (or Fokker–Planck equation) is

$$\begin{aligned} \frac{\partial p}{\partial t} &= \nabla\cdot\left[-\mathbf{a}(\mathbf{x},t)p + \frac{1}{2}\nabla\cdot(\mathbf{B}(\mathbf{x},t)p)\right] \\ &= -\sum_{i=1}^{d}\frac{\partial}{\partial x_i}[a_i(\mathbf{x},t)p] + \frac{1}{2}\sum_{i,j=1}^{d}\frac{\partial^2}{\partial x_i\partial x_j}[B_{ij}(\mathbf{x},t)p]. \end{aligned} \tag{12.12}$$

12.4 Itô stochastic integral

As mentioned above, we can define a diffusion process as the solution of a stochastic differential equation (SDE). An ordinary differential equation specifies the rate of change of $x(t)$ as a function of x and t and has a differential update formula $\mathrm{d}x = a(x(t),t)\mathrm{d}t$. An SDE for a one-dimensional diffusion process $X(t)$ specifies the infinitesimal mean, or drift, and the infinitesimal variance, or noise, with an update formula of the form

$$\mathrm{d}X(t) = a(X(t),t)\mathrm{d}t + \big(\text{random term of mean zero and variance } b(X(t),t)\mathrm{d}t\big).$$

The last term needs to have mean zero and variance $b(X(t),t)\mathrm{d}t$ even conditional of the path or history $\{X(s), s\le t\}$. This may be achieved by taking the last term to be $\sqrt{b(X(t),t)}\mathrm{d}W(t)$, recalling that $\mathbb{E}(W(t)-W(s))^2 = |t-s|$ (see also (12.18) and (12.19) further below). We thus formally arrive at the *stochastic differential equation* or SDE

$$\mathrm{d}X(t) = a(X(t),t)\mathrm{d}t + \sqrt{b(X(t),t)}\mathrm{d}W(t), \tag{12.13}$$

where $W(t)$ is a standard Brownian motion as given in Definition 10.1. We say that $X(t)$ is *locally* a Gaussian process.

The correct interpretation of the SDE (12.13) is as a *stochastic integral equation*

$$X(t) = X_0 + \int_0^t a(X(s),s)\mathrm{d}s + \int_0^t \sqrt{b(X(s),s)}\mathrm{d}W(s). \tag{12.14}$$

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 this, suppose that we wanted to compute

$$I(t) = \int_0^t W(s)\mathrm{d}W(s),$$

using Riemann integration. Consider a partition of $[0,t]$ with equally spaced points $\{0,t_1,t_2,\dots,t_n\}$ with timestep $\Delta t = t/n$. Then the Riemann sum to approximate $I(t)$ is

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

Depending on the choice of τ_k, we get different Riemann sums:

- Left Riemann sum: $I_L(t)$ with $\tau_k = t_k$.
- Middle Riemann sum: $I_M(t)$ with $\tau_k = (t_k + t_{k+1})/2$.
- Right Riemann sum: $I_R(t)$ with $\tau_k = t_{k+1}$.

We leave as an exercise to show that $\mathbb{E}I_L(t) = 0$, $\mathbb{E}I_M(t) = t/2$ and $\mathbb{E}I_R(t) = t$. Hence, the limits of the Riemann sum approximations as $n \to \infty$ depend on the choice of τ_k and differ by an $O(1)$ amount (since we get $0, t/2, t$ independent of n!). Therefore, we conclude that the Riemann integral 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 two most well-known definitions of a stochastic integral are *Itô* and *Stratonovich*.

Definition 12.1 (Itô stochastic integral)**.** *The Itô integral is the mean-square limit*[3] *of the left Riemann sums,*

$$\int_0^t f(s)\mathrm{d}W(s) = \lim_{n\to\infty} \sum_{k=0}^{n-1} f(t_k)[W(t_{k+1}) - W(t_k)], \tag{12.15}$$

where $f(t)$ is a square-integrable random function

$$\int_0^T \mathbb{E}[f^2(t)]\mathrm{d}t < \infty. \tag{12.16}$$

Definition 12.2 (Stratonovich stochastic integral)**.** *The Stratonovich integral is the mean-square limit of the mid-point Riemann sums,*

$$\int_0^t f(s) \circ \mathrm{d}W(s) = \lim_{n\to\infty} \sum_{k=0}^{n-1} f\left(\frac{t_k + t_{k+1}}{2}\right)[W(t_{k+1}) - W(t_k)], \tag{12.17}$$

for f satisfying (12.16).

Exercise. *Applying the definitions above, show that*

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

Each definition leads to a different "stochastic calculus"; therefore, when speaking of stochastic integrals, it is important to specify whether we are using the Itô or Stratonovich definition. The Stratonovich integral (12.17) leads to the standard chain rule, while the Itô integral (12.15) requires a correction (Itô's formula, see below). On the other hand, the Itô integral is a martingale and is well-defined for non-anticipating processes.[4] If $f(\cdot)$ depends on $W(t)$, we do not know what it will do on $[t_k, t_{k+1}]$, so it is best to use the known value of $f(t_k)$ in the approximation. These reasons make the Itô integral (12.15) desirable and worth accepting the slight inconvenience of the modified chain rule. We will only use the Itô integral in this course and hence Itô SDEs. However, note that an Itô SDE can be converted into a Stratonovich SDE and conversely (see [20, p.61]).

Two important properties of the Itô integral are:

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

[4] Not expected in this course, but for completeness: A function $f(t)$ is called non anticipating if $f(t)$ is statistically independent of the Wiener increment $W(t+s) - W(t)$ for $s > 0$. A stochastic process $X(t)$ is a martingale is $\mathbb{E}[X(t)|\mathcal{F}_s] = X(s)$ for all $t \geq s$, where $\mathcal{F}_s$ is the filtration generated by $X(s)$ up to time s (think of it as the path $\{X(s), s \leq t\}$).

(i) Martingale property:

$$\mathbb{E}\left[\int_0^t f(t)\mathrm{d}W(t)\right] = 0. \tag{12.18}$$

(ii) Itô isometry:

$$\mathbb{E}\left[\left(\int_0^t f(t)\mathrm{d}W(t)\right)^2\right] = \int_0^t \mathbb{E}(f^2(t))\mathrm{d}t. \tag{12.19}$$

12.5 Itô stochastic differential equation

We are now ready to define Itô SDEs. Given a stochastic process $X(t)$ in $\mathbb{R}$, it is said to satisfy an Itô stochastic differential equation (SDE)

$$\mathrm{d}X(t) = a(X(t),t) + \sigma(X(t),t)\mathrm{d}W(t), \tag{12.20}$$

if, for $t \geq 0$, it is a solution of the integral equation

$$X(t) = X_0 + \int_0^t a(X(s),s)\mathrm{d}s + \int_0^t \sigma(X(s),s)\mathrm{d}W(s), \tag{12.21}$$

where the first integral is a Riemann integral and the second integral is an Itô stochastic integral.

The conditions needed for the existence and uniqueness of SDEs on the drift and diffusion coefficients in (12.20) are similar to the Lipschitz continuity and linear growth assumptions from the theory of ODEs [20, p.64]. This is out of the scope of this course, where we will assume that the coefficients are "nice" such that a unique solution to the SDE (12.20) exists.

Suppose that the diffusion process $X(t)$ is the solution to (12.20) for $t \in [0,T]$. What SDE does

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

solve? It turns out that $Y(t)$ is *again* a diffusion process. From (12.20), 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 (12.20) 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

$$\mathrm{d}Y(t) = \frac{\partial f}{\partial t}(X(t),t)\mathrm{d}t + \frac{\partial f}{\partial x}(X(t),t)\mathrm{d}X(t) + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(X(t),t)(\mathrm{d}X(t))^2, \tag{12.22}$$

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

$$\begin{aligned}\mathrm{d}Y(t) = &\left[\frac{\partial f}{\partial t}(X(t),t) + a(X(t),t)\frac{\partial f}{\partial x}(X(t),t) + \frac{1}{2}\sigma^2(X(t),t)\frac{\partial^2 f}{\partial x^2}(X(t),t)\right]\mathrm{d}t \\ &+ \sigma(X(t),t)\frac{\partial f}{\partial x}(X(t),t)\mathrm{d}W(t)\end{aligned} \tag{12.23}$$

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 $\int_0^t W(s)\mathrm{d}W(s) = [W^2(t) - t]/2$ in the exercise above. Consider the SDE for geometric Brownian motion (11.12), that is,

$$\mathrm{d}S(t) = \mu S(t)\mathrm{d}t + \sigma S(t)\mathrm{d}W(t), \qquad S(0) = S_0. \tag{12.24}$$

We can readily solve (12.24) 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

$$S(t) = S_0 e^{\sigma W(t) + \left(\mu - \frac{1}{2}\sigma^2\right)t}. \tag{12.25}$$

Now consider a multidimensional diffusion process $\mathbf{X}(t)$ in $\mathbb{R}^d$

$$\mathrm{d}\mathbf{X}(t) = \mathbf{a}(\mathbf{X}(t), t)\mathrm{d}t + \boldsymbol{\sigma}(\mathbf{X}(t), t)\mathrm{d}\mathbf{W}(t), \tag{12.26}$$

where $\mathbf{a}(\mathbf{x}, t) \in \mathbb{R}^d, \boldsymbol{\sigma}(\mathbf{x}, t) \in \mathbb{R}^{d\times m}$ and $\mathbf{W}(t) = (W_1(t), \ldots, W_m(t)) \in \mathbb{R}^m$, with $m \leq d$. Note that the diffusion matrix (12.10) associated to (12.26) is $\mathbf{B}(\mathbf{x}, t) = \boldsymbol{\sigma}\boldsymbol{\sigma}^\top$. (In particular, this means that more than one $\boldsymbol{\sigma}$ is consistent with the same $\mathbf{B}$. That is, any two processes with the same $B(\mathbf{x}, t)$ are identical from Kolmogorov's equations point of view.)

Multivariate Itô's formula

Let $\mathbf{X}(t) \in \mathbb{R}^d$ be a solution to (12.26). Let $Y(t) = f(\mathbf{X}(t), t)$, 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

$$\mathrm{d}Y(t) = \frac{\partial f}{\partial t}\mathrm{d}t + \sum_{i=1}^{d}\frac{\partial f}{\partial x_i}\mathrm{d}X_i(t) + \frac{1}{2}\sum_{i,j=1}^{d}\frac{\partial^2 f}{\partial x_i \partial x_j}\mathrm{d}X_i(t)\mathrm{d}X_j(t), \tag{12.27}$$

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 $i, j = 1, \ldots, d$.

Lecture 13: Numerical integration of SDEs

While we saw that we could use Itô's formula to obtain an explicit solution of Geometric Brownian motion (c.f. (12.24) and (12.25)), this is not the case for most SDEs, and one must resort to numerical techniques. One may be tempted to think that, since "anyway I'm solving something with noise, numerical errors do not matter in the same way as in numerically solving ODEs or PDEs". But of course, this is false, and care should be given to the choice of numerical integration of SDEs.

Suppose that we want to solve the scalar time-homogeneous SDE

$$\mathrm{d}X(t) = a(X(t))\mathrm{d}t + \sigma(X(t))\mathrm{d}W(t), \qquad X(0) = X_0, \tag{13.1}$$

and write it in integral form

$$X(t) = X_0 + \int_0^t a(X(s))\mathrm{d}s + \int_0^t \sigma(X(s))\mathrm{d}W(s). \tag{13.2}$$

If we had an ODE ($\sigma \equiv 0$), there exist many well-established methods to solve (13.2), such as

- Taylor expansions (they require derivatives $a^{(n)}(x)$),
- Runge–Kutta methods (they do not require derivatives),
- Multistep methods.

These methods can be generalised with appropriate modifications to SDEs. We will show a systematic method to obtain Taylor methods for (13.1).

Consider a function $f \in C^2(\mathbb{R})$ and recall Itô's formula (12.23), 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

$$f(X(t)) = f(X_0) + \int_0^t L_0 f(X(s))\mathrm{d}s + \int_0^t L_1 f(X(s))\mathrm{d}W(s), \tag{13.3}$$

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 (13.3) to each of the integrands in (13.2), that is, with $f = a$ in the first integral and $f = \sigma$ in the second, we obtain

$$X(t) = X_0 + a(X_0)\int_0^t \mathrm{d}s + \sigma(X_0)\int_0^t \mathrm{d}W(s) + R_1, \tag{13.4}$$

with reminder

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

13.1 Euler–Maruyama scheme

If we ignore the reminder R_1, we obtain the Euler–Maruyama (EM) scheme we already saw in the first lecture. Suppose we wish to solve (13.1) on some interval of time $[0, T]$. We denote by X_n the approximation to $X(t_n)$, where $t_n = n\Delta t$ for $\Delta t > 0, n \in \mathbb{N}$. Then neglecting R_1 in (13.4) we have

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

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

SSA 13.1: Euler–Maruyama scheme

Partition the interval $[0, T]$ into N equal subintervals of length $\Delta t = T/N$). Set $X_0 = x_0$. For $n = 0$ to $n = N - 1$

(a1) Generate a random number $\xi \sim \mathcal{N}(0, 1)$ (the standard normal distribution).

(b1) Set

$$X_{n+1} = X_n + a(X_n)\Delta t + \sigma(X_n)\sqrt{\Delta t}\,\xi. \tag{13.5}$$

end

13.2 Milstein scheme

We can carry on with the expansion of R_1 to obtain more accurate schemes. 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

$$R_1 = \int_0^t \int_0^s L_1\sigma(X(s'))\mathrm{d}W(s')\mathrm{d}W(s) + o(\sqrt{ts}). \tag{13.6}$$

Applying (13.3) with $f = L_1\sigma$ gives

$$X(t) = X_0 + a(X_0)\int_0^t \mathrm{d}s + \sigma(X_0)\int_0^t \mathrm{d}W(s) + L_1\sigma(X_0)\int_0^t \int_0^s \mathrm{d}W(s')\mathrm{d}W(s) + R_2, \tag{13.7}$$

with $R_2 = o(\sqrt{ts})$. We already know that the last integral in (13.7) is

$$\int_0^t \int_0^s \mathrm{d}W(s')\mathrm{d}W(s) = \int_0^t W(s)\mathrm{d}W(s) = \frac{1}{2}(W(t)^2 - t).$$

Ignoring the reminder R_2 leads to the so-called *Milstein scheme*

$$X_{n+1} = X_n + a(X_n)\Delta t + \sigma(X_n)\Delta W_n + \frac{1}{2}\sigma(X_n)\sigma'(X_n)(\Delta W_n^2 - \Delta t),$$

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

SSA 13.2: Milstein scheme

Partition the interval $[0, T]$ into N equal subintervals of length $\Delta t = T/N$). Set $X_0 = x_0$. For $n = 0$ to $n = N - 1$

(a1) Generate a random number $\xi \sim \mathcal{N}(0, 1)$ (the standard normal distribution).

(b1) Set

$$X_{n+1} = X_n + a(X_n)\Delta t + \sigma(X_n)\sqrt{\Delta t}\,\xi + \frac{1}{2}\sigma(X_n)\sigma'(X_n)\Delta t(\xi^2 - 1). \tag{13.8}$$

end

- For *additive-noise processes*, that is, $\sigma(x) \equiv \sigma$ constant, the Milstein scheme reduces to the Euler–Maruyama scheme.
- We could, in theory, continue with the expansion to derive more accurate Taylor schemes. But, in practice, they are not used.
- For multivariate SDEs (see (12.26)), the Milstein scheme would require calculating integrals such as

$$\int_0^t \int_0^s \mathrm{d}W_i(s')\mathrm{d}W_j(s), \qquad i \neq j,$$

where $W_i(t), W_j(t)$ are independent Brownian motions. These integrals arise from non-diagonal terms in the diffusion tensor $\boldsymbol{\sigma}(\mathbf{x})$. There are no known analytic expressions for such integrals. While these multiple integrals can be approximated, the EM scheme SSA 13.1 is used instead in practice for multivariate diffusion processes.

13.3 Two-step Runge–Kutta scheme

Another common scheme, which is derivative-free, is the 2-step Runge–Kutta scheme:

SSA 13.3: 2-step Runge–Kutta scheme

Partition the interval $[0, T]$ into N equal subintervals of length $\Delta t = T/N$). Set $X_0 = x_0$. For $n = 0$ to $n = N - 1$

(a1) Generate a random number $\xi \sim \mathcal{N}(0, 1)$ (the standard normal distribution).

(b1) Do

$$\begin{aligned} \hat{X}_n &= X_n + \sigma(X_n)\sqrt{\Delta t}, \\ X_{n+1} &= X_n + a(X_n)\Delta t + \sigma(X_n)\sqrt{\Delta t}\,\xi + \frac{\sqrt{\Delta t}}{2}[\sigma(\hat{X}_n) - \sigma(X_n)](\xi^2 - 1). \end{aligned} \tag{13.9}$$

end

Note that

$$\sigma(\hat{X}_n) = \sigma(X_n + \sigma(X_n)\sqrt{\Delta t}) \approx \sigma(X_n) + \sigma(X_n)\sigma'(X_n)\sqrt{\Delta t}.$$

Inserting this expression into (13.9), we recover the Milstein scheme (13.8).

13.4 Strong and weak convergence

How good is a numerical scheme in approximating the original problem? As in the deterministic case, common notions are:

- Consistency: the local discretisation error converges to zero with order (at least) $O(\Delta t)$.
- Convergence: the global discretisation error in a time interval $[0, T]$ converges to zero with order $O(\Delta t^\alpha)$ for some $\alpha > 0$.
- Stability: this concerns whether errors are damped out, and the numerical solution reproduces the long-time behaviour of the exact solution.

Since both the exact path $X(t)$ and the numerical approximation X_n are random, there are different ways to quantify these notions: *strong forms* consider pathways results, while *weak forms* deal with probability distributions. Here we consider only the weak and strong convergence of numerical schemes for SDEs.

Consider still the Itô SDE

$$\mathrm{d}X(t) = a(X(t)) + \sigma(X(t))\mathrm{d}W(t), \qquad X(0) = X_0, \tag{13.10}$$

for $t \in [0, T]$, an equal partition into N subintervals of size Δt and X_n the numerical approximation of $X(t)$ at times $t_n = n\Delta t$.

Definition 13.1 (Strong convergence)**.** *Given an approximation X_n to $X(t)$ in $[0, T]$, we define the* strong error *as*

$$e_{\Delta t}^{strong} = \sup_{0 \le t_n \le T} \mathbb{E}|X(t_n) - X_n|. \tag{13.11}$$

We say a scheme converges strongly if $e_{\Delta t}^{strong} \to 0$ as $\Delta t \to 0$. Furthermore, we say that the scheme has strong order of convergence α if there exist $C > 0$ and $\Delta t^* > 0$ such that

$$e_{\Delta t}^{strong} \le C\Delta t^\alpha, \qquad \forall 0 < \Delta t < \Delta t^*.$$

Definition 13.2 (Weak convergence)**.** *Given an approximation X_n to $X(t)$ in $[0, T]$, we define the* weak error *as*

$$e_{\Delta t}^{weak} = \sup_{0 \le t_n \le T} |\mathbb{E}\Phi(X(t_n)) - \mathbb{E}\Phi(X_n)|, \tag{13.12}$$

where Φ is, e.g., the set of polynomials of degree at most k. We say a scheme converges weakly if $e_{\Delta t}^{weak} \to 0$ as $\Delta t \to 0$. Furthermore, we say that the scheme has weak order of convergence β if there exists $C > 0$ and $\Delta t^* > 0$ (both depending on k)

$$e_{\Delta t}^{weak} \le C\Delta t^\beta, \qquad \forall 0 < \Delta t < \Delta t^*.$$

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 ($\beta \ge \alpha$).

13.5 Numerical convergence tests

Note that strong convergence requires convergence as $\Delta t \to 0$ but for a *fixed realisation* of the solution $X(t)$ to (13.10). Therefore, in numerical convergence tests, the same realisation of $W(t)$ must be used for all approximations (see Figure 13.1). In contrast, for weak convergence, we can use different paths. Finally, we must also be aware of other sources of error [12], which are implicitly assumed negligible when monitoring the errors $e_{\Delta t}^{\text{strong}}$ and $e_{\Delta t}^{weak}$:

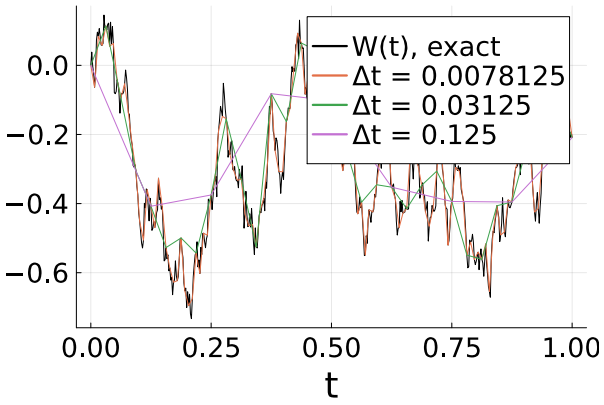

Figure 13.1: Exact path of Brownian motion $W(t)$ and evaluation at coarser time grids.

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

In general, when the solution of the SDE is not known analytically, we compute the order of convergence by comparing two approximate solutions obtained with different time steps. For example, suppose that $X_n^{\Delta t}$ is the approximation to $X(t)$ using a scheme with strong order α (which we are trying to find) and timestep Δt. Then

$$\mathbb{E}|X(t_n) - X_n^{\Delta t}| \sim C\Delta t^\alpha \quad \text{and} \quad \mathbb{E}|X(t_n) - X_n^{\Delta t/2}| \sim C(\Delta t/2)^\alpha.$$

Then the mean of the error between the coarse and finer approximations is

$$\mathbb{E}|X_n^{\Delta t} - X_n^{\Delta t/2}| = \mathbb{E}|(X_n^{\Delta t} - X(t_n)) - (X_n^{\Delta t/2} - X(t_n))| \sim C\Delta t^\alpha,$$

and

$$\frac{\mathbb{E}\left|X_n^{\Delta t} - X_n^{\Delta t/2}\right|}{\mathbb{E}\left|X_n^{\Delta t/2} - X_n^{\Delta t/4}\right|} \sim \frac{C\Delta t^\alpha}{C(\Delta t/2)^\alpha} = 2^\alpha.$$

To test the different numerical schemes we have presented, we need a process with *multiplicative noise*, $\sigma'(x) \neq 0$. A good choice is Geometric Brownian motion (12.24) as we have the explicit solution in hand, that is,

$$X(t) = X_0 e^{\sigma W(t) + \left(\mu - \frac{1}{2}\sigma^2\right)t}. \tag{13.13}$$

In Figure 13.2, we show the result of numerical tests of the strong and the weak convergence of approximate solutions $X(t)$ using the Euler–Maruyama (13.5), Milstein (13.8) and 2-step Runge–Kutta (13.9) schemes to the exact solution $X(t)$ in (13.13). Instead of taking the supremum in (13.11) and (13.12), we consider the endpoint errors at time $T = 1$. The numerical tests suggest that

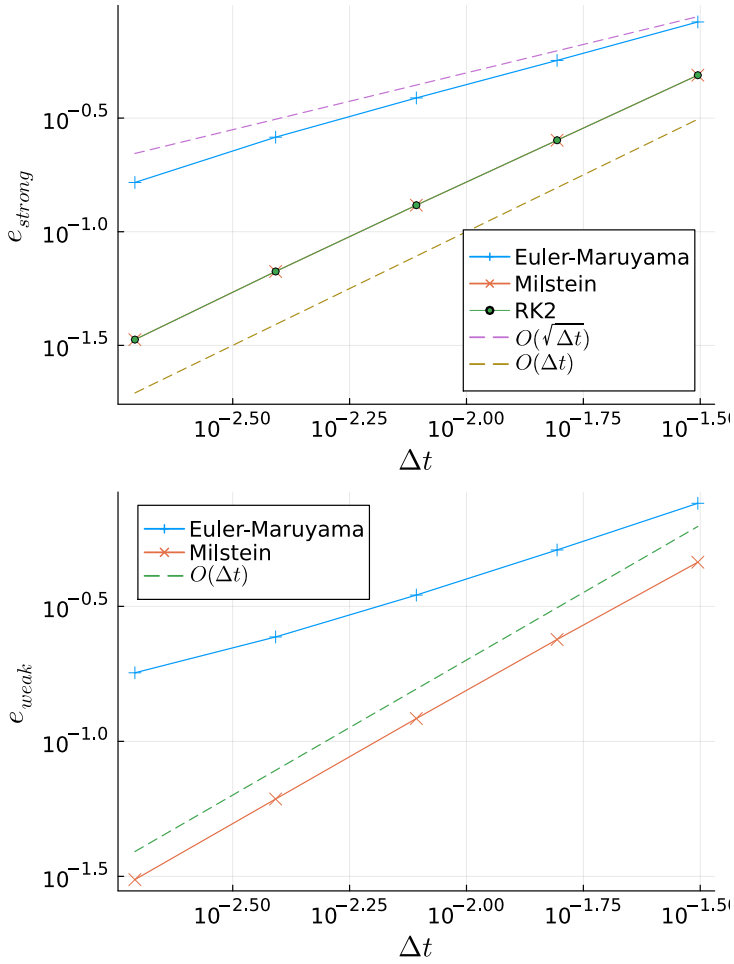

Figure 13.2: Strong (left) and weak (right) endpoint errors of convergence of the Euler–Maruyama (13.5), Milstein (13.8) and 2-step Runge–Kutta (13.9) schemes to approximate Geometric Brownian Motion (13.13). Parameters used: $\mu = 2$, $\sigma = 1$, $X_0 = 1$. Errors measured at the final time $T = 1$. For the weak error (13.12) we use $\Phi(x) = x$.

	Strong order	Weak order
Euler–Maruyama (SSA 13.1)	1/2	1
Milstein (SSA 13.2)	1	1
2-step Runge–Kutta (SSA 13.3)	1	1

These orders of convergence are valid for a general SDE (13.10) and can be proven rigorously. Note that for SDEs with additive noise, $\sigma(x)$ constant, all three schemes are equivalent; hence, for diffusions with additive noise, the Euler–Maruyama scheme has a strong order of convergence equal to one.

Lecture 14: Second-order models for diffusion

In Lecture 10, 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 $V(t)$.

14.1 Velocity-jump process

Recall the run-and-tumble model (11.11), used to describe the motion and search behaviour of organisms such as bacteria. While the particle's position is continuous, the velocity is subject to a jump process governed by the turning kernel $W(\mathbf{v}|\mathbf{w},t)$. Hence the name velocity-jump process [9].

Let's consider a one-dimensional version of (11.11). We assume that a particle with position $X(t) \in \mathbb{R}$ moves along the real line with velocity $V(t) \in \{-s, s\}$ where $s > 0$ is a constant:

$$\mathrm{d}X(t) = V(t)\mathrm{d}t, \qquad V(t) \in \{-s, s\}. \tag{14.1}$$

The particle reverses its velocity $V(t)$ according to a Poisson process with rate λ. Recall from the example in §1.2 that this means that

- $\mathbb{P}\{\text{particle changes direction in the time interval } (t, t+\Delta t)\} = \lambda\Delta t + O(\Delta t)$,
- $\mathbb{P}\{\text{particle does not change direction in the time interval } (t, t+\Delta t)\} = 1 - \lambda\Delta t + O(\Delta t)$.

The $V(t)$ is called a *telegraph process*. Note that the joint process $(X(t), V(t))$ in (14.1) is a continuous-time Markov process (see Definition (A.2)) but that $X(t)$ alone is not Markovian. This is because if $X(t + \mathrm{d}t) > X(t)$ (particle has moved right), then it will likely move right again between $t + \mathrm{d}t$ and $t + 2\mathrm{d}t$.

In principle, λ could depend on $X(t)$ (higher chance of turning depending on the particle's position) or $V(t)$ (asymmetric turning rates from left to right and right to left). For example, a common model for bacterial chemotaxis is

$$\lambda(x, v) = \lambda_0 - \mathrm{sign}(v) b c'(x),$$

where $c(x)$ is the (fixed) concentration of a chemical, such that $|c'(x)| < \lambda_0/b$ for all x. For simplicity, let's take λ constant here.

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:

$$p^{\pm}(x,t)\mathrm{d}x = \mathbb{P}\{X(x) \in [x, x+\mathrm{d}x), V(t) = \pm s\}.$$

The densities $p^{\pm}$ satisfy the following system of forward Kolmogorov equations

$$\frac{\partial p^+}{\partial t} + s\frac{\partial p^+}{\partial x} = \lambda p^- - \lambda p^+, \tag{14.2a}$$

$$\frac{\partial p^-}{\partial t} - s\frac{\partial p^-}{\partial x} = \lambda p^+ - \lambda p^-. \tag{14.2b}$$

We may complement (14.2) 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 $t = 0$.

Let's now introduce the probability $p(x,t)\mathrm{d}x$ that the particle is at $[x, x+\mathrm{d}x)$ at time t, and the probability flux $j(x,t) = s[p^+(x,t) - p^-(x,t)]$. The density satisfies

$$p(x,t) = p^+(x,t) + p^-(x,t). \tag{14.3}$$

Adding and subtracting equations (14.2) we obtain

$$\frac{\partial p}{\partial t} + \frac{\partial j}{\partial x} = 0, \tag{14.4}$$

and

$$\frac{\partial j}{\partial t} + s^2\frac{\partial p}{\partial x} = -2\lambda j \tag{14.5}$$

Differentiating (14.5) with respect to x and using (14.4) we obtain

$$\frac{1}{2\lambda}\frac{\partial^2 p}{\partial t^2} + \frac{\partial p}{\partial t} = \frac{s^2}{2\lambda}\frac{\partial^2 p}{\partial x^2}. \tag{14.6}$$

Equation (14.6) is a hyperbolic PDE known as the *telegraph equation*.

We now consider the parabolic scaling of (14.6) 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 §10.1). Then (14.6) becomes (dropping asterisks)

$$\frac{\delta^2}{2\lambda}\frac{\partial^2 p}{\partial t^2} + \frac{\partial p}{\partial t} = \frac{s^2}{2\lambda}\frac{\partial^2 p}{\partial x^2}, \tag{14.7}$$

and hence at leading order in δ, we obtain

$$\frac{\partial p}{\partial t} = \frac{s^2}{2\lambda}\frac{\partial^2 p}{\partial x^2}. \tag{14.8}$$

Equation (14.8) is valid for long times $\gg 1/2\lambda$ (corresponding to $\delta^2 \ll 2\lambda$ so that the first term in (14.7) is negligible). Therefore, we have seen that for long times, a bacterium undergoing the velocity-jump process (14.1) satisfies the diffusion equation (10.13) with diffusion coefficient

$$D = \frac{s^2}{2\lambda}. \tag{14.9}$$

Figure (14.1) shows a simulation of the two-dimensional version of the velocity-jump process (14.1) (see (11.11)), 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 (14.8).

14.2 Langevin's equation

We now consider another second-order model, which Langevin proposed to describe Brownian motion. 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:

$$F(t) = -\gamma V(t) + R(t).$$

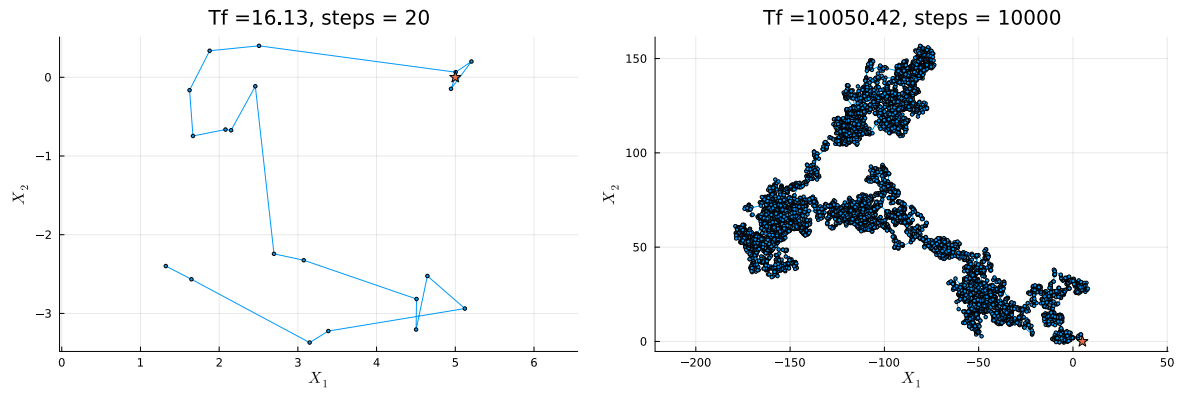

Figure 14.1: Simulation of the two-dimensional version of the velocity-jump process (14.1) (see (11.11)), with $\mathbf{V}(t) \in s\cdot\mathbb{S}^1 = \{\mathbf{v} \in \mathbb{R}^2 : \|\mathbf{v}\| = s\}$ and a uniform turning kernel $W(\mathbf{v}|\mathbf{w}, t) \equiv 1/2\pi$. The trajectories are simulated 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 $T_f = 10050.42$.

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. Suppose the Brownian particle is in a thermal bath of solvent. In that case, it is reasonable to assume that the collisions it feels are independent of its position $X(t)$ and that R is a white noise, with $\mathbb{E}R(t) = 0$ and covariance $C(t, s) = \mathbb{E}[R(t)R(s)] = \Gamma\delta(t - s)$, where Γ a constant to be determined. Hence we can identify the equation for $V(t)$ (14.10b) as the Ornstein–Uhlenbeck SDE (1.12) from Lecture 1:

$$\mathrm{d}V(t) = -\frac{\gamma}{m}V(t)\mathrm{d}t + \sqrt{\Gamma/m^2}\mathrm{d}W(t),$$

The Einstein–Smoluchowski relation We may rewrite (14.10) as

$$m\frac{\mathrm{d}^2X(t)}{\mathrm{d}t} = -\gamma\frac{\mathrm{d}X(t)}{\mathrm{d}t} + R(t). \tag{14.11}$$

Multiplying both sides by $X(t)$ and rearranging gives

$$m\frac{\mathrm{d}^2X^2(t)}{\mathrm{d}t} + \gamma\frac{\mathrm{d}X^2(t)}{\mathrm{d}t} = 2mV^2(t) + 2R(t)X(t).$$

Taking expectations, we arrive at

$$m\frac{\mathrm{d}^2}{\mathrm{d}t}\mathbb{E}X^2(t) + \gamma\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}X^2(t) = 2k_BT,$$

where k_B is the Boltzmann constant and T the absolute temperature. In the line above, we have used the assumptions of $R(t)$ $\mathbb{E}[R(t)X(t)] = \mathbb{E}R(t)\mathbb{E}X(t) = 0$ and the equipartition theorem for the kinetic energy in one dimension, namely that the average kinetic energy in thermal equilibrium of the particles is $\mathbb{E}[\frac{1}{2}mV^2(t)] = k_BT/2$. Integrating once we arrive at

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}X^2(t) = \frac{2k_BT}{\gamma} + Ce^{-\gamma t/m},$$

where C is an arbitrary constant. The second term on the right-hand side decays rapidly with time. So for times $t \gg m/\gamma$ we may neglect it and integrate again, setting $X(0) = 0$,

$$\mathbb{E}X^2(t) = \frac{2k_BT}{\gamma}t. \tag{14.12}$$

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 $2D$, as his result was derived using a different method (c.f. §10.3).[5] Equating the two expressions for the mean-squared distance leads to

$$D = \frac{k_B T}{\gamma}. \tag{14.13}$$

Equation (14.13) is known as the *Einstein–Smoluchowski relation* or simply the Einstein relation.

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 (14.13) leads to the *Stokes–Einstein relation*

$$D = \frac{k_B T}{3\pi\eta a}. \tag{14.14}$$

The fluctuation–dissipation theorem Note that in the previous calculation, the random force $R(t)$ dropped. How come $\mathbb{E}X^2(t)$ is independent of the noise? In fact, it does depend on it, as a certain magnitude of the noise is required to assume that the system is in thermal equilibrium. The relationship between friction and noise is known as the *fluctuation-dissipation theorem* (see Sheet 3):

$$\mathbb{E}[R(t)R(s)] = 2k_B T\gamma\delta(t-s). \tag{14.15}$$

Any system in thermal equilibrium experiences both friction and fluctuations, and the magnitudes of the two effects are directly linked. You can only get one with the other. Hence, we have that Γ above is $\Gamma = 2k_B T\gamma$ and the Langevin SDE is

$$m\mathrm{d}V(t) = -\gamma V(t)\mathrm{d}t + \sqrt{2\gamma k_B T}\mathrm{d}W(t), \qquad V(0) = 0, \tag{14.16a}$$

$$\mathrm{d}X(t) = V(t)\mathrm{d}t, \qquad X(0) = 0. \tag{14.16b}$$

We denote by $p(x, v, t)$ the probability density function of $\{X(t), V(t)\}$ in (14.16). From (12.12), we have that p satisfies the Fokker–Planck equation

$$\frac{\partial p}{\partial t} + v\frac{\partial p}{\partial x} = \frac{\gamma}{m}\frac{\partial}{\partial v}\left(vp + \frac{k_B T}{m}\frac{\partial p}{\partial v}\right), \quad p(x, v, 0) = \delta(x)\delta(v). \tag{14.17}$$

The overdamped limit of Langevin's equation Consider the regime when $m/\gamma \ll 1$, corresponding to a system with large friction. Dividing (14.16a) by γ, we may formally set the left-hand side to zero, giving:

$$V(t)\mathrm{d}t = \sqrt{2k_B T/\gamma}\mathrm{d}W(t) = \sqrt{2D}\mathrm{d}W(t),$$

using (14.13). Substituting this into (14.16b), we arrive at

$$\mathrm{d}X(t) = \sqrt{2D}\mathrm{d}W(t). \tag{14.18}$$

Hence, we see that in the limit $m/\gamma \to 0$, Langevin's model of diffusion (14.16) reduces to Einstein's model. Large friction leads to a fast relaxation of momentum and overdamped motion. The process $\{X(t), V(t)\}$ is Markovian, but note that $X(t)$ alone is not. This is a case where we can make the process Markovian by extending the dimension to include an additional variable that models the memory of the non-Markovian process, in this case $V(t)$. In the overdamped limit, $X(t)$ loses memory quickly.

[5] Recall the diffusive scaling (10.11), or the fact that $\mathbb{E}W^2(t) = t$ (see (10.7)) and $D = 1/2$ for Brownian motion.

We may also consider the overdamped limit $m/\gamma \ll 1$ at the level of the Fokker–Planck equation (14.17). To this end, we set $\epsilon := m/\gamma$ and consider large times $t \gg \epsilon^{-1}$ in the limit of ϵ small. We rescale time

$$\epsilon t = \bar{t}, \qquad p(x, v, t) = \bar{p}(x, v, \bar{t})$$

to give

$$\epsilon^2 \frac{\partial \bar{p}}{\partial \bar{t}} + \epsilon v \frac{\partial \bar{p}}{\partial x} = \frac{\partial}{\partial v}\left(v\bar{p} + \beta^{-1}\frac{\partial \bar{p}}{\partial v}\right) = L\bar{p}, \tag{14.19}$$

where we have introduced $\beta^{-1} = k_B T/m$ 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 (14.19) and equating powers of ϵ we obtain the hierarchy of equations

$$0 = L\bar{p}^{(0)}, \tag{14.20}$$

$$v\frac{\partial \bar{p}^{(0)}}{\partial x} = L\bar{p}^{(1)}, \tag{14.21}$$

$$\frac{\partial \bar{p}^{(0)}}{\partial \bar{t}} + v\frac{\partial \bar{p}^{(1)}}{\partial x} = L\bar{p}^{(2)}. \tag{14.22}$$

Integrating (14.20) with respect to v and using $\bar{p}, \partial_v\bar{p} \to 0$ for $|v| \to \infty$, we get

$$\bar{p}^{(0)}(x, v, \bar{t}) = \bar{\rho}(x, \bar{t})f(v), \tag{14.23}$$

where $\bar{\rho}(x, \bar{t})$ is the constant of integration (to be determined) and $f(v)$ is the *Maxwellian distribution*

$$f(v) = \frac{1}{\sqrt{2\pi\beta^{-1}}}\exp\left(-\frac{\beta v^2}{2}\right) = \sqrt{\frac{m}{2\pi k_B T}}\exp\left(-\frac{mv^2}{2k_B T}\right). \tag{14.24}$$

So, at leading order, we see that the particle's position and velocity are independent. Inserting (14.23) into (14.21) we have

$$L\bar{p}^{(1)} = \frac{\partial \bar{\rho}}{\partial x} v f(v), \tag{14.25}$$

with solution

$$\bar{p}^{(1)} = -\frac{\partial \bar{\rho}}{\partial x} v f(v). \tag{14.26}$$

Substituting (14.23) and (14.26) into (14.22) 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 (14.24). Going back to the original time $t = \epsilon^{-1}\bar{t}$ and writing $\bar{\rho}(x, \bar{t}) = \rho(x, t)$ gives

$$\frac{\partial \rho}{\partial t} = D\frac{\partial^2 \rho}{\partial x^2}. \tag{14.27}$$

where D is given in (14.13). Using (14.23) we see that the *spatial probability density*

$$\int_{\mathbb{R}} p(x, v, t)\mathrm{d}v, \tag{14.28}$$

satisfies the diffusion equation (14.27) for large γ/m. This equation is consistent with the overdamped limit (14.18) formally obtained, taking $m/\gamma \to 0$ at the level of the SDE.

In Figure 14.2, we show trajectories of the two-dimensional version of the Langevin dynamics (14.16) for increasing values of the damping coefficient γ. In particular, we consider $\gamma = 1, 10, 100, 1000$ while keeping the diffusivity D (14.13) fixed. We observe that for small γ, the position $\mathbf{X}(t)$ 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 (14.24).

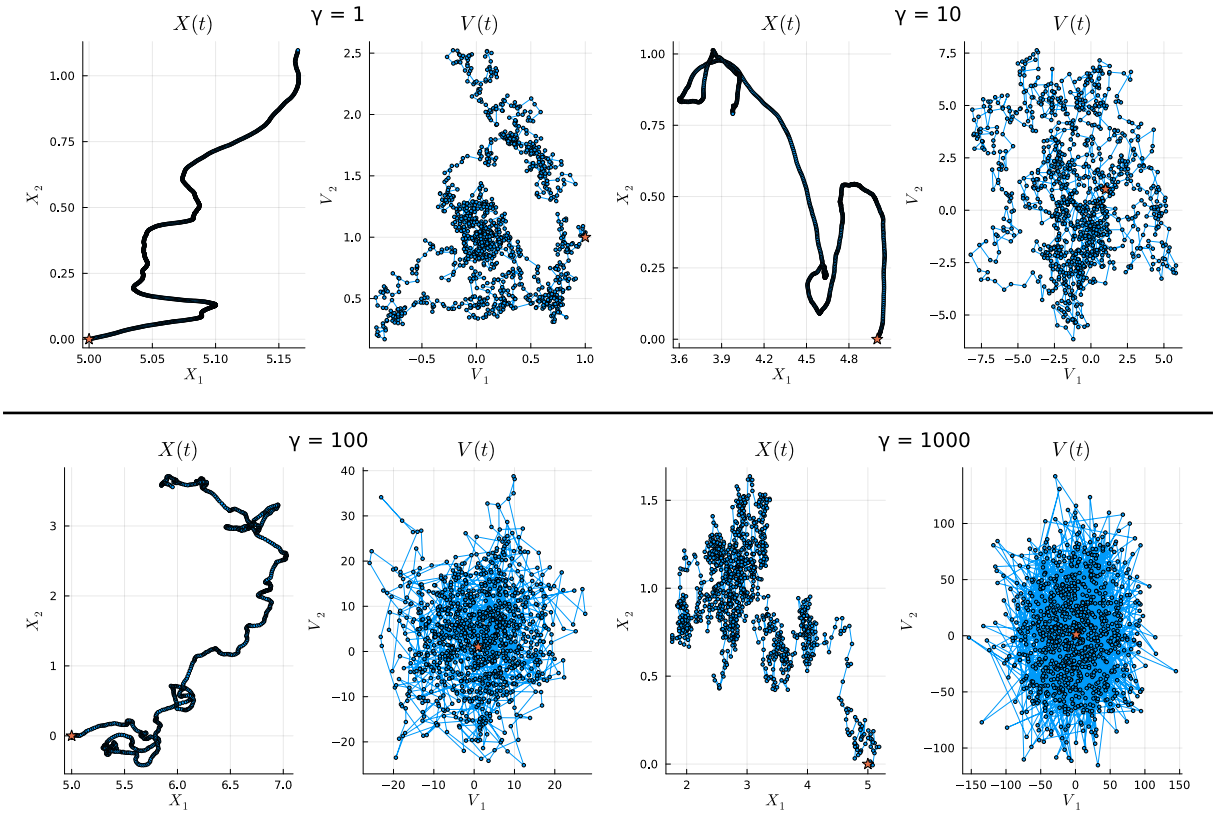

Figure 14.2: Simulation of the two-dimensional version of the Langevin dynamics (14.16) for different values of the damping coefficient γ ($\gamma = 1, 10, 100, 1000$) at a fixed diffusivity D (14.13). The trajectories of the position $\mathbf{X}(t)$ and velocity $\mathbf{V}(t)$ are simulated using the Euler–Maruyama SSA 13.1 with timestep $\Delta t = 10^{-3}$ and final time $T = 1$.

We have seen that the two second-order models (14.1) and (14.10) reduce to Brownian motion in the large time limit. We will next consider the stationary solution of the Langevin equation with confining potential.

Lecture 15: Stationary processes

Recall the concept of a time-homogeneous Markov process, whose transition probabilities are invariant to time shifts, see Definition A.5. Here we look at whether these may settle into a steady state or *stationary process* as time becomes large.

15.1 Stationary solution of the Fokker–Planck equation

Consider the Langevin SDE (14.16) with an additional potential $U(x)$. Setting $m = 1$ for simplicity, we have

$$\mathrm{d}V(t) = -U'(X(t))\mathrm{d}t - \gamma V(t)\mathrm{d}t + \sqrt{2\gamma k_B T}\mathrm{d}W(t), \qquad \mathrm{d}X(t) = V(t)\mathrm{d}t, \tag{15.1}$$

and the corresponding Fokker–Planck equation becomes

$$\frac{\partial p}{\partial t} = -v\frac{\partial p}{\partial x} + \frac{\partial}{\partial v}\left[U'(x)p\right] + \gamma\frac{\partial}{\partial v}\left(vp + k_B T\frac{\partial p}{\partial v}\right). \tag{15.2}$$

We can identify two distinct processes on the right-hand side of (15.2): the first two terms correspond to (deterministic) Hamiltonian dynamics[6] with *Hamiltonian*

$$\mathcal{H}(x, v) = \frac{v^2}{2} + U(x), \tag{15.3}$$

while the last two terms indicate the presence of noise and dissipation.

We ask under which circumstances the solution to (15.2) approaches a time-invariant limit as $t \to \infty$, so that we may talk about the *stationary probability density* (or invariant density) p_∞,

$$p_\infty(x, v) = \lim_{t\to\infty} p(x, v, t). \tag{15.4}$$

Setting the left-hand side of (15.2) to zero, we find that

$$p_\infty(x, v) = \frac{1}{Z}\exp\left(-\frac{\mathcal{H}(x, v)}{k_B T}\right), \tag{15.5}$$

where Z is the normalisation constant,

$$Z = \int_{\mathbb{R}}\int_{\mathbb{R}} \exp\left(-\frac{\mathcal{H}(x, v)}{k_B T}\right)\mathrm{d}x\mathrm{d}v = \sqrt{2\pi k_B T}\int_{\mathbb{R}} \exp\left(-\frac{U(x)}{k_B T}\right)\mathrm{d}x$$

Suppose the potential is given by $U(x) = \alpha x^2$. If there is no potential ($\alpha = 0$), it is clear that $Z = +\infty$ and hence the solution (15.5) does not exist. The same is true for $\alpha < 0$ since the integral in x above does not converge. Conversely, $Z < +\infty$ for potentials with $\alpha > 0$. We see that the Fokker–Planck equation (15.2) for $(x, v) \in \mathbb{R} \times \mathbb{R}$ admits a stationary solution for potentials that are integrable in the sense above; these are called confining potentials.

Definition 15.1 (Confining potential)**.** *A function $U(x) : \mathbb{R} \to \mathbb{R}$ is called a confining potential if it satisfies the following two conditions:*

$$\lim_{|x|\to\infty} U(x) = +\infty, \qquad \int_{\mathbb{R}} e^{-cU(x)}\mathrm{d}x < +\infty,$$

for all $c > 0$.

[6] This means that $x'(t) = \partial_v \mathcal{H} = v$ and $v'(t) = -\partial_x \mathcal{H} = -U'(x)$.

Note that the stationary density (15.5) is separable in x and v, with the velocity component corresponding to the Maxwellian (14.24). This is consistent with what we found in (14.23) for the case without a potential. Also, observe that the stationary density p_∞ is independent of γ and D (as long as $\gamma, D > 0$); it only depends on the temperature T.

Let's now consider the stationary spatial density associated with (15.5),

$$\rho_\infty(x) = \int_{\mathbb{R}} p_\infty(x, v)\mathrm{d}v = \frac{1}{\tilde{Z}} \exp\left(-\frac{U(x)}{k_B T}\right), \tag{15.6}$$

where $\tilde{Z}$ is the normalisation constant. Recall the Ornstein–Uhlenbeck process from Lecture 1,

$$\mathrm{d}X(t) = -\alpha X(t)\mathrm{d}t + \sqrt{2D}\mathrm{d}W(t), \qquad \alpha > 0, \tag{15.7}$$

with associated stationary density (see (1.17))

$$\rho_\infty(x) = \sqrt{\frac{\alpha}{2\pi D}} e^{-\frac{\alpha x^2}{2D}}. \tag{15.8}$$

Comparing (15.6) and (15.8), we see that they coincide if $U(x) = x^2/2$ and $(k_B T)^{-1} = \alpha/D$, that is, $\alpha = 1/\gamma$. This makes sense as, had we taken the overdamped limit of the Langevin dynamics with a confining potential (15.2), we would find that the corresponding limit is

$$\mathrm{d}X(t) = -\frac{U'(X(t))}{\gamma}\mathrm{d}t + \sqrt{2D}\mathrm{d}W(t), \tag{15.9}$$

Therefore, for long times, both the Langevin and the overdamped Langevin dynamics converge to the same spatial stationary density (and the velocities of the Langevin model equilibrate to the Maxwellian (14.24)).

15.2 Ergodic properties of stationary processes

The stationary density exists and is unique for Langevin dynamics (15.1) and the Ornstein–Uhlenbeck process (15.7). This is, in fact, a characterisation of *ergodic* diffusion processes at the level of the Fokker–Planck equation. We say a diffusion process is ergodic if the stationary probability density

$$\mathcal{L}^* p_\infty = 0, \tag{15.10}$$

where $\mathcal{L}^*$ is the forward Kolmogorov operator (12.8), exists and is unique.

Examples We have already seen that Langevin dynamics and Brownian motion under a confining potential are ergodic processes in $\mathbb{R}^2$ and $\mathbb{R}$, respectively. On the other hand, Brownian motion in $\mathbb{R}^d$ is not an ergodic Markov process as p_∞ does not exist (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 for the particle's positions, e.g., $X(t) \in [0, 1]$, and appropriate boundary conditions. We will discuss processes in bounded domains in §16.

Ergodic processes also have the nice property that we may calculate expectation values of an observable, $\mathbb{E}f(X(t))$, by calculating time averages, provided that the correlation between values of the process at different times decays sufficiently fast [20]. First, we give a characterisation of ergodicity at the level of the process $X(t)$.

Definition 15.2 (Second-order stationary process)**.** *A stochastic process $X(t)$ is second-order stationary (or weakly stationary) if the first moment $\mathbb{E}X(t)$ is a constant and the covariance function $C(t, s)$ depends only on $t - s$:*

$$\mathbb{E}X(t) = \mu, \qquad C(t, s) = \mathbb{E}[(X(t) - \mu)(X(s) - \mu)] = C(t - s).$$

We saw that both Brownian motion and the Ornstein–Uhlenbeck process are time-homogeneous (their transition probabilities are invariant to time shifts, see Definition A.5). Second-order stationarity asks for more: that their statistics (specifically their first two moments) remain invariant under time translations. This is not the case for Brownian motion, since $C(t,s) = \min(t,s)$ (see (10.7)), but it is for the Ornstein–Uhlenbeck process (15.7), if initialised in a suitable way or one waits for long enough. In particular, the OU process with initial condition $X(0) = X_0$ satisfies

$$\begin{aligned} \mathbb{E}X(t) &= \mathbb{E}X_0 e^{-\alpha t}, \\ C(t,s) &= \left[\mathrm{Var}(X_0) - \frac{D}{\alpha}\right] e^{-\alpha(t+s)} + \frac{D}{\alpha} e^{-\alpha|t-s|}. \end{aligned} \tag{15.11}$$

If $X_0 \sim \mathcal{N}(0, D/\alpha)$, then the OU is already stationary. We may also compare (15.11) with the explicit probability density (1.15) for $X_0 = x_0$ (deterministic), so that $\mathbb{E}X(t) = x_0 e^{-\alpha t}$ and

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

Theorem 15.1 (Ergodic theorem for stationary processes, see [20])**.** *Let $X(t)$ be a second-order stationary process in $\mathbb{R}$ with mean μ and covariance $C(t)$. If $\int_0^\infty C(t)\mathrm{d}t < +\infty$, then*

$$\lim_{T\to\infty} \mathbb{E}\left|\frac{1}{T}\int_0^T X(s)\mathrm{d}s - \mu\right|^2 = 0. \tag{15.12}$$

In general, for an observable f, we have

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T f(X(s))\mathrm{d}s = \int_{\mathbb{R}} f(x)p_\infty(x)\mathrm{d}x, \tag{15.13}$$

where p_∞ is the stationary density (15.10). Hence, for an ergodic process, we can either simulate trajectories for a very long time and compute the time average or average over many short independent simulations. This fact is used by Markov Chain Monte Carlo (MCMC) methods to sample the stationary density function of ergodic processes. We see an example in the next section.

15.3 Numerical simulation of the stationary state

Suppose we have an SDE for $X(t)$, and we are interested in the behaviour of $X(T)$ for a long time T. If the process is ergodic, we may estimate its stationary density using (15.13). To fix some ideas, let's consider the Ornstein–Uhlenbeck process (15.7), so we may compare the numerically estimated stationary density with its exact value (15.8).

SSA 15.1: Stationary density estimation of the OU process (15.7) via ensemble averaging of the Euler–Maruyama

Choose a final time T (large), a timestep Δt with $N\Delta t = T$, and a number of paths M.
for $s = 1$ to M
    Generate an initial condition X_0.
    for $n = 0$ to $n = N-1$
        (a1) Generate a random number $\xi \sim \mathcal{N}(0,1)$.
        (b1) Set $X_{n+1} = X_n - \alpha X_n \Delta t + \sqrt{2D\Delta t}\,\xi$.
    end
    Store X_N in a histogram.
end

Using the computational approximation for X_{n+1} in SSA 15.1 one may show that

$$\mathbb{E}X_{n+1} = (1-\alpha\Delta t)\mathbb{E}X_n \to 0,$$
$$\mathbb{E}X_{n+1}^2 = (1-\alpha\Delta t)^2\mathbb{E}X_n^2 + 2D\Delta t \to \frac{2D}{\alpha(2-\alpha\Delta t)},$$

as $n \to \infty$ 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

$$\sigma_{\Delta t}^2 = \frac{2D}{\alpha(2-\alpha\Delta t)}. \tag{15.14}$$

This approaches the exact value D/α (see (15.8)) as $\Delta t \to 0$, but we expect an error for finite Δt. We test this in Figure 15.1, 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 [11, §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, (15.8) (shown as a solid black line), and that of the Euler–Maruyama SSA (shown as a black dash line). The histograms obtained from running the SSA 15.1 with $M = 10^4$ samples at times $T = 0, 1, 6$ are also shown.

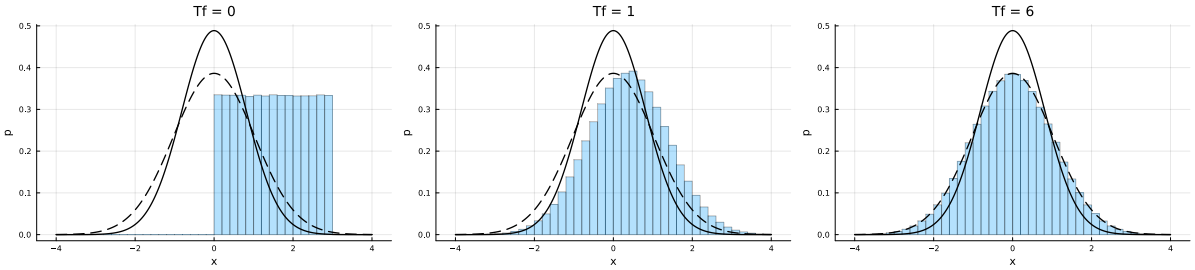

Figure 15.1: Histograms to estimate p_∞ from the Euler–Maruyama method SSA 15.1 using $\Delta t = 1$ and $M = 10^4$ on an Ornstein–Uhlenbeck process (15.7) at times $T = 0, 1, 6$, with initial condition X_0 uniform over $(1, 3)$. Solid curve: stationary probability density (15.8) of the SDE. Dashed curve: stationary probability density of the EM method with variance (15.14).

In Figure 15.2, we repeat the experiment of Figure 15.1 still with the EM SSA 15.1 but with a finer timestep $\Delta t = 10^{-3}$. The results are consistent with the fact that $\sigma_{\Delta t}^2 \approx 1.0005D/\alpha$ in this case so that the curves corresponding to the exact and approximated stationary densities overlap and the result of the simulation for long enough time agrees with (15.7). We note three possible sources of error:

- Number of samples M.
- Numerical integration of the SDE (related to the time-step Δt).
- Final time T should be large enough so that the process has become stationary.

Since the OU process (15.7) is an ergodic stationary process, we may also estimate its stationary density by doing a time average, see (15.13). We may converge faster to the stationary state if we choose to initialise the process appropriately, so it is already stationary ($\mathbb{E}X_0 = 0$ and $\mathrm{Var}(X_0) = D/\alpha$, see (15.11)) or close to stationary for more complex processes, or if we throw away the transient part of the trajectory (when the process has not yet reached stationarity). The pseudocode for this method is given below (compare it with SSA 15.1).

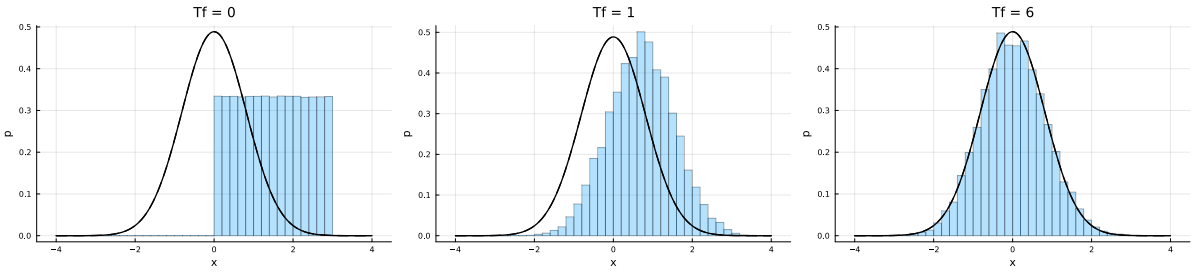

Figure 15.2: Histograms to estimate p_∞ from the Euler–Maruyama method SSA 15.1 using $\Delta t = 10^{-3}$ and $M = 10^4$ on an Ornstein–Uhlenbeck process (15.7) at times $T = 0, 1, 6$, with initial condition X_0 uniform over $(1, 3)$. Solid curve: stationary probability density (15.8) of the SDE. Dashed curve: stationary probability density for the EM method with variance (15.14) (not visible).

SSA 15.2: Stationary density estimation of the OU process (15.7) via time-averaging of the Euler–Maruyama

Choose a final time T (large), a timestep Δt with $N\Delta t = T$.
Generate an initial condition X_0.
for $n = 0$ to $n = N - 1$
(a1) Generate a random number $\xi \sim \mathcal{N}(0, 1)$.
(b1) Set $X_{n+1} = X_n - \alpha X_n \Delta t + \sqrt{2D\Delta t}\,\xi$.
(c1) Store X_{n+1} in the histogram.
end

SSA 15.2 is an example of the Markov Chain Monte Carlo method. The time-averaging approach of SSA 15.2 is generally more efficient than the ensemble averaging SSA 15.1 since it uses a large proportion of the computed values towards the histogram, as opposed to only using the final value of the trajectory. However, this requires the process (if not already stationary) to reach its stationary regime. When using SSA 15.2, the time it takes to reach this stage depends on the dynamical properties of the process. We next consider another MCMC method that samples from the stationary density efficiently.

15.4 The Metropolis–Hastings algorithm

The Metropolis–Hastings (MH) algorithm is an MCMC method that stems out of work by Metropolis *et al.* in 1953 [18]. In their work, the method is used to compute the equilibrium properties of a system with many rigid spheres, but the method is quite general and versatile. While the algorithm is particularly effective for high-dimensional problems, we explain it with the simple one-dimensional Ornstein–Uhlenbeck process (15.7). Suppose that we want to generate many points $X_i \in \mathbb{R}$, $i = 1, 2, 3, \ldots$, according to the stationary density of the OU process (15.8), which is of the form $p_\infty(x) = C\pi(x)$, where

$$\pi(x) = \exp(-H(x)), \qquad H(x) = \frac{\alpha x^2}{2D}, \tag{15.15}$$

and C is the normalisation constant. The MH algorithm computes the sequence $X_i, i = 1, 2, 3, \ldots$, iteratively as follows. Given X_i, we generate a candidate y for X_{i+1} according to some distribution $q(x, X_i)$. For example, let's use $y = X_i + \delta\xi$, where δ is the size of the step

and $\xi \sim \mathcal{N}(0,1)$. Then the proposed move is accepted with probability

$$p_{\text{acc}} = \min[1, \exp(-\Delta H)], \qquad \Delta H = H(x) - H(X_i). \tag{15.16}$$

That is, if the candidate x 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 accepted, we set $X_{i+1} = y$; otherwise we set $X_{i+1} = X_i$. We write the pseudocode implementing this method in SSA 15.3.[7]

SSA 15.3: Metropolis–Hastings algorithm for the OU process (15.7)

Set the initial condition X_0, the number of steps M, and the spread of the move δ.
for $i = 1$ to M
- **(a1)** Generate random numbers $\xi \sim \mathcal{N}(0,1)$ and $r \sim U(0,1)$.
- **(b1)** Set the proposed move $y = X_i + \delta\xi$ and compute $\Delta H = H(y) - H(X_i)$ from (15.15).
- **(c1)** If $r < \exp(-\Delta H)$, set $X_{i+1} = y$ (accept the move); otherwise set $X_{i+1} = X_i$ (reject the move).

end

It is worth noting that the algorithm does not require knowledge of the normalisation constant C in (15.15), and this is one of its greatest strengths when it comes to very high-dimensional systems (see §17.4). We now present a more general version of the MH algorithm. Suppose again that we want to generate samples from a probability density $C\pi(x)$ and consider an auxiliary function $q(x,y) : \mathbb{R} \times \mathbb{R} \to (0,\infty)$ to generate the candidate moves. The candidate-generating function $q(x,\cdot)$ is a probability density for every $x \in \mathbb{R}$ and is symmetric, $q(x,y) = q(y,x)$. The function q should be such that, given x, it is straightforward to generate a random value y according to $q(x,y)$. For example, above, we had $q(x,y) \equiv q_1(x-y)$ with q_1 the normal distribution with mean zero and variance δ. The algorithm proceeds as follows: given X_i, generate the candidate y for X_{i+1} according to the probability density $q(X_i,\cdot)$. The move is accepted with probability

$$p_{\text{acc}} = \min\left[1, \frac{\pi(y)}{\pi(X_i)}\right]. \tag{15.17}$$

The choice of the spread or scale of the candidate-generating density (δ in SSA 15.3) has important implications for the efficiency of the MH algorithm [7]. It affects the "acceptance rate" (the percentage of accepted moves) and the region of sample space covered by the chain. The optimal acceptance rate depends on the target distribution, but it has been shown theoretically that for an N-dimensional Gaussian target distribution, it approaches 0.234 [21]. Intuitively, if δ is too small, the acceptance rate is high, but the chain will move around the configuration space slowly, so it will take many steps to sample the whole space. On the other extreme, if δ is too high, the proposed move y is likely to have a considerably lower probability than X_i, and hence the acceptance rate will be very low. In practice, an acceptance rate in the 0.1 order of magnitude gives the desired results. It should be clear that the MH algorithm generates a sequence of configurations which correctly samples the stationary density, but this sequence is not related to the time evolution of the corresponding dynamical system [9].

[7] The MH algorithm is written in its most basic form in the SSA box. One could optimise the code to avoid throwing away unused random numbers. For example, we only require to generate $r \sim U(0,1)$ if $\Delta H > 0$.

Lecture 16: Transport in bounded domains

So far, we have considered SDEs of the form

$$\mathrm{d}X(t) = a(X(t),t)\mathrm{d}t + \sqrt{b(X(t),t)}\mathrm{d}W(t), \qquad X(0) \sim p_0(x) \tag{16.1}$$

and its associated Fokker–Planck equation

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}[a(x,t)p] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[b(x,t)p], \qquad p(x,0) = p_0(x), \tag{16.2}$$

in $\mathbb{R}$ (or $\mathbb{R}^d$), with p decaying fast enough at $x = \pm\infty$. But, in applications, we are often interested in processes in a bounded domain $\Omega \subset \mathbb{R}$ with *appropriate* boundary conditions.

16.1 Boundary conditions on the Fokker–Planck equation ($\mathcal{L}^*$)

For simplicity, let's consider a time-homogeneous process (so that the explicit time dependence of the drift and diffusion coefficients drops) in a one-dimensional domain. We can rewrite (16.2) as

$$\frac{\partial p}{\partial t} + \frac{\partial J}{\partial x} = 0, \qquad x \in \Omega, t \geq 0, \tag{16.3}$$

where the *probability flux*

$$J(x,t) = a(x)p - \frac{1}{2}\partial_x[b(x)p] \tag{16.4}$$

describes how the probability moves. In particular,

$$\frac{\partial}{\partial t}\int_\Omega p\mathrm{d}x = -\int_\Omega \frac{\partial J}{\partial x}\mathrm{d}x = -\int_{\partial\Omega} J\cdot n\mathrm{d}s,$$

where n is the outward normal on the boundary $\partial\Omega$. Thus, we see that for the total probability to be conserved,

$$\int_\Omega p(x,t)\mathrm{d}x = \int_\Omega p_0(x)\mathrm{d}x = 1,$$

we require that $\int J\cdot n|_{\partial\Omega}\mathrm{d}s = 0$. This is automatically satisfied if $\Omega \equiv \mathbb{R}$ using the decay at infinity, but in a bounded domain, it will depend on the boundary conditions on $\partial\Omega$.

Consider a diffusion process (16.1) in $\Omega = [0, L]$. Possible boundary conditions are:

(i) Reflecting boundary conditions (e.g. the particle hits a wall and is reflected back into the domain):

$$J(x,t) = 0, \quad \text{for} \quad x \in \partial\Omega.$$

(ii) Periodic boundary conditions (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 of a possibly nonzero flux at the boundaries but still satisfying the constraint $\int J\cdot n|_{\partial\Omega}\mathrm{d}s = 0$ so that the total probability is conserved. For this boundary condition to make sense, we require the coefficients a and b also to be L-periodic.

(iii) Adsorbing boundary conditions (e.g. the particle hits a reactive wall and is removed):

$$p(x,t) = 0, \quad \text{for} \quad x \in \partial\Omega.$$

In this case, note that the probability is *not conserved*.

(iv) Partially reflecting / adsorbing boundaries (e.g. the particle hits an imperfect reactive wall, with a certain probability is adsorbed and otherwise is reflected):

$$J(x,t) = \kappa p(x,t), \quad \text{for} \quad x \in \partial\Omega,$$

where 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 adsorbing boundary above.

(v) Mixed boundary conditions: these involve a combination of the above on different boundaries.

16.2 Boundary conditions on the backward Kolmogorov equation ($\mathcal{L}$)

Recall the backward Kolmogorov equation in (12.7). We justified the notation $\mathcal{L}$ for the backward and $\mathcal{L}^*$ for the forward (or Fokker–Planck) operators by saying that the backward operator is formally the adjoint of the forward operator. What's missing to make the definition rigorous? Suitable boundary conditions. We have that

$$\langle \mathcal{L}f, g\rangle = \langle f, \mathcal{L}^* g\rangle, \tag{16.5}$$

for all f, g in a suitable function space (with boundary conditions attached). Here $\langle f, g\rangle = \int fg\mathrm{d}x$ is the L^2-inner product.

Boundary conditions for the backward equation can be derived from those for the forward equation using integration by parts. To this end, one computes $\langle \mathcal{L}f, p\rangle$ and manipulates it to obtain $\langle f, \mathcal{L}^* p\rangle$ plus additional boundary terms. By integration by parts, one can show (see Example Sheet 3) that

$$\langle \mathcal{L}f, p\rangle = \langle f, \mathcal{L}^* p\rangle + \int_{\partial\Omega} \left[fJ \cdot n + \frac{1}{2} b(x) p \frac{\partial f}{\partial x} \cdot n \right] \mathrm{d}s,$$

with J given in (16.4) and n the outward normal. Using the expression above, we find that

(i) Reflecting boundary conditions: we have that $J \cdot n = 0$ on $\partial\Omega$, so the first term on the right-hand side vanishes. We require the second to vanish, leading to the boundary condition on $\mathcal{L}$:

$$b(y) \frac{\partial f}{\partial y} \cdot n = 0, \quad \text{for} \quad y \in \partial\Omega.$$

(ii) Adsorbing boundary conditions: in this case we have that $p = 0$ on $\partial\Omega$, so the second term on the right-hand side drops. We must choose

$$f(y,t) = 0, \quad \text{for} \quad y \in \partial\Omega.$$

Other boundary conditions for $\mathcal{L}$ can be derived similarly.

16.3 Boundary conditions for SDEs

It is challenging to write down boundary conditions explicitly at the level of the SDE. For this reason, it is most common to consider a discretised version of the SDE that approximates the solution of the SDE. This could be a continuous-time Markov chain (like the random walk in §10.1) or a discrete-time approximation (like the numerical schemes seen in Lecture 13). Let's do the latter and consider the Euler–Maruyama discretisation of (16.1),

$$X_{n+1} = X_n + a(X_n)\Delta t + \sqrt{b(X_n)\Delta t}\,\xi, \tag{16.6}$$

with $\xi \sim \mathcal{N}(0,1)$.

Periodic boundary Suppose that $\Omega = [0, L]$ and that we impose periodic boundary conditions. This means that if X_{n+1} is not in Ω, we must place it back in imposing periodicity:

$$X'_{n+1} = \quad \bmod (X_{n+1}, L). \tag{16.7}$$

Reflective boundary Take $\Omega = [0, \infty)$ and suppose that $x = 0$ is a reflective boundary. How do we modify the SSA to deal with cases when $X_{n+1} < 0$? Corrections such as not accepting the move (similar to what was done in the Metropolis–Hastings algorithm when the move leads to very large/infinite energy) or placing the particle at contact ($X'_{n+1} = 0$) can lead to wrong dynamics and statics [22]. A better approximation if $X_{n+1} < 0$ is to reflect it back:

$$X'_{n+1} = -X_{n+1}.$$

A way to think about this is that we are approximating $X(t)$ by a linear interpolation at points X_n, so that

$$X(t) = X_n + (t - t_n)V_n, \quad \text{for } t \in [t_n, t_{n+1}).$$

with "velocity"

$$V_n = \frac{X_{n+1} - X_n}{\Delta t}, \tag{16.8}$$

and performs ballistic dynamics (in particular, when the particle hits the wall, the velocity flips, $V_n \to -V_n$).

This method does not introduce any further approximation for an isotropic diffusion ($a = 0$ and b constant). To see this, note that for every trajectory that reaches $x = 0$, there is an equal probability for $X(t)$ to go to y or $-y$.

SSA 16.1: Euler–Maruyama scheme of (16.6) with a reflective boundary

Set $X_0 = x_0 > 0$ and $\Delta t = T/N$.
For $n = 0, \dots, N-1$

(a1) Generate a random number $\xi \sim \mathcal{N}(0, 1)$ (the standard normal distribution).

(b1) Set $X_{n+1} = X_n - \alpha X_n \Delta t + \sqrt{2D\Delta t}\,\xi$.

(c1) If $X_{n+1} < 0$, set $X_{n+1} \leftarrow -X_{n+1}$.

end

Adsorbing boundary Still with $\Omega = [0, \infty)$, now assume that $x = 0$ is an adsorbing boundary. If after one step of (16.6), $X_{n+1} < 0$, this means that the particle has crossed $x = 0$ since the process is continuous. Hence, we must remove it from the system.

However, as Andrews and Bray point out in [2], we are not quite done. Even if $X_n, X_{n+1} > 0$, there is still a non-zero probability that the trajectory has hit $x = 0$ during the time interval $[t_n, t_{n+1})$. For example, in Figure (16.1), the red trajectory should have been killed between t_n and t_{n+1}. Consider a Brownian motion with $a(x) \equiv 0$ and $b(x) = 2D$ constant. Take $x_n, x_{n+1} > 0$. What is the probability of having hit $x = 0$ (A) given that the transition $x_n \to x_{n+1}$ occurred (B)? That is, what is $\mathbb{P}(A/B)$? We may compute it using

$$\mathbb{P}(A/B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)},$$

where (see (12.4))

$$\mathbb{P}(B) = p(x_{n+1}, \Delta t | x_n, 0) = \frac{1}{\sqrt{4\pi D\Delta t}} \exp\left(-\frac{(x_{n+1} - x_n)^2}{4D\Delta t}\right).$$

and

$$\mathbb{P}(A \cap B) = \mathbb{P}(x_n \to -x_{n+1}) = p(-x_{n+1}, \Delta t | x_n, 0),$$

using the symmetry and continuity of the process. Hence we find that

$$\mathbb{P}(A/B) = \frac{p(-x_{n+1}, \Delta t | x_n, 0)}{p(x_{n+1}, \Delta t | x_n, 0)} = \exp\left(-\frac{x_{n+1}x_n}{D\Delta t}\right). \tag{16.9}$$

Thus, 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. The pseudocode to simulate (16.6) with an adsorbing boundary at $x = 0$ including (16.9) is given in SSA 16.2. We will denote the correction **(c1)** in SSA 16.2 as the *Andrews and Bray correction* [2].

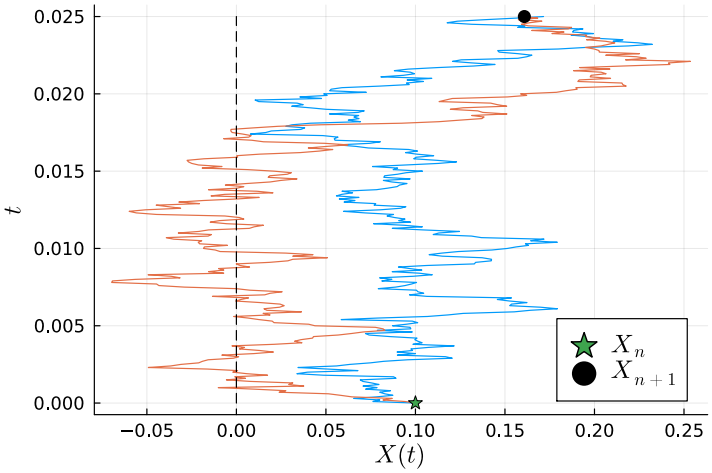

Figure 16.1: Two trajectories of (16.6) with $a = 0, b = 2$. With an adsorbing boundary at $x = 0$, the red trajectory should have been terminated.

SSA 16.2: Euler–Maruyama scheme of (16.6) with an adsorbing boundary

Set $X_0 = x_0 > 0$ and $\Delta t = T/N$.
For $n = 0, \ldots, N-1$
(a1) Generate a random number $\xi \sim \mathcal{N}(0, 1)$ (the standard normal distribution).
(b1) Set $X_{n+1} = X_n - \alpha X_n \Delta t + \sqrt{2D\Delta t}\,\xi$.
(c1) 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

16.4 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 [13]:

- Channel transport (ion channels, porins in bacteria): how long it takes 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 virial 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 (12.6). To see this, let $X(t)$ be a time-homogeneous diffusion process

$$\mathrm{d}X(t) = a(X(t))\mathrm{d}t + \sqrt{b(X(t))}\mathrm{d}W(t),$$

in a bounded domain $\Omega = (0, L)$ with a reflecting boundary at $x = 0$ and an adsorbing boundary at $x = L$. Given an initial condition $X(0) = y \in \Omega$, we define the *survival probability* $Q(y, t)$ as the probability that the particle has not left Ω by time t:

$$Q(y,t) = \int_0^L p(x,t|y,0)\mathrm{d}x. \tag{16.10}$$

Since y is in Ω, we have that $Q(y,0) = \int \delta(x-y)\mathrm{d}x = 1$. That is, at $t = 0$, we know for sure that the particle is still "alive". However, as time progresses, the adsorbing boundary condition $p(L,t|y,0) = L$ means that there is a leak of probability through $x = L$ and that $Q(y,t)$ will decrease.

Definition 16.1 (First passage time)**.** *We denote by first passage time $T(y)$ the first time that $X(t)$ reaches the boundary $x = L$ having started at $y \in \Omega$:*

$$T(y) = \inf\{t \geq 0 : X(t) = L\}. \tag{16.11}$$

Note that $T(y)$ is a random variable. The definition can be generalised to include multiple boundaries or cases where the particle is not removed when it hits the boundary (e.g., if there is a reflective or partially adsorbing boundary). Hence the *first* in its definition. Let $p_T(y,t)$ be the probability density function associated with $T(y)$. Noting that $\mathbb{P}(T(y) \geq t) = Q(y,t)$, we have that

$$p_T(y,t) = \frac{\partial}{\partial t}[1 - Q(y,t)] = -\frac{\partial Q(y,t)}{\partial t}. \tag{16.12}$$

Definition 16.2 (Mean first passage time)**.** *The mean first massage time (MFPT) $\tau(y)$ is*

$$\tau(y) = \mathbb{E}T(y) = \int_0^\infty t p_T(y,t)\mathrm{d}t = \int_0^\infty Q(y,t)\mathrm{d}t. \tag{16.13}$$

We may use that Q satisfies (16.10) to find an expression for $\tau(y)$. Since $Q(y,t)$ is a function of the initial position y, we consider the homogeneous backward Kolmogorov equation for $p(y,t)$

$$\frac{\partial p}{\partial t} = \mathcal{L}_y\, p,$$

where $\mathcal{L}_y$ is given in (12.6), with boundary conditions (see §16.2)

$$\frac{\partial p}{\partial y}(0,t) = 0, \qquad p(L,t) = 0,$$

(assuming that $b(0) \neq 0$). Integrating the equation for p with respect to x, we find that the survival probability satisfies

$$\frac{\partial Q}{\partial t} = \mathcal{L}_y Q, \tag{16.14}$$

$$\frac{\partial Q}{\partial y} = 0, \qquad \text{on } y = 0, \tag{16.15}$$

$$Q = 0, \qquad \text{on } y = L, \tag{16.16}$$

$$Q = 1, \qquad \text{at } t = 0. \tag{16.17}$$

Using the definition of $\tau(y)$ (16.13) in (16.14) together with (16.17), we have that

$$\mathcal{L}_y \tau(y) = \int_0^\infty \frac{\partial Q}{\partial t} \mathrm{d}t = Q(\infty) - Q(0) = -1, \tag{16.18}$$

with boundary conditions

$$\tau'(0) = 0, \qquad \tau(L) = 0. \tag{16.19}$$

In (16.18), we have used that $Q(\infty) = 0$ since, eventually, the particle will hit $x = L$.

For example, consider a standard Brownian motion, $a(x) = 0$ and $b(x) = 1$, in the domain $\Omega = (0, L)$. Then with $x = 0$ reflecting and $x = L$ adsorbing, we have that

$$\frac{1}{2}\tau''(y) = -1, \qquad \tau'(0) = \tau(L) = 0,$$

leading to $\tau(y) = L^2 - y^2$. We may also consider the MFPT if both $x = 0$ and $x = L$ are adsorbing boundaries. In this case, $\tau(y)$ satisfies

$$\frac{1}{2}\tau''(y) = -1, \qquad \tau(0) = \tau(L) = 0,$$

with solution $\tau(y) = y(L-y)$. The derivation above may be repeated for more general boundaries in $\mathbb{R}^d$. We give the result in the box below.

Mean first passage time in $\Omega \subset \mathbb{R}^d$

Let $X(t)$ be a homogeneous diffusion process in $\Omega \subset \mathbb{R}^d$

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

with $\partial\Omega = \partial\Omega_A \cup \partial\Omega_B$ (part of the boundary is adsorbing, and part of the boundary is reflecting). Then the MFPT $\tau(y)$ to leave Ω through $\partial\Omega_A$ having started at $\mathbf{y} \in \Omega$, $\mathbf{X}(0) = \mathbf{y}$, satisfies the following problem

$$\mathcal{L}\tau = -1, \qquad y \in \Omega, \tag{16.20}$$

$$\tau = 0, \qquad y \in \partial\Omega_A, \tag{16.21}$$

$$(\mathbf{B} \cdot \nabla\tau) \cdot \mathbf{n} = 0, \qquad y \in \partial\Omega_R, \tag{16.22}$$

where $\mathbf{n}$ is the outward normal vector on $\partial\Omega$ and $\mathbf{B}(\mathbf{x}, t) = \boldsymbol{\sigma}\boldsymbol{\sigma}^\top$. The backward Kolmogorov operator is

$$\mathcal{L}\tau = \mathbf{a}(\mathbf{y}) \cdot \nabla\tau + \frac{1}{2}\mathbf{B}(\mathbf{y}, s) : \nabla\nabla\tau = \sum_{i=1}^{d} a_i(\mathbf{y}) \frac{\partial \tau}{\partial y_i} + \frac{1}{2} \sum_{i,j=1}^{d} B_{ij}(\mathbf{y}) \frac{\partial^2 \tau}{\partial y_i \partial y_j}.$$

The reflecting boundary condition (16.22) is, in components, $\sum_{i,j} n_i B_{ij} \partial_{y_j} \tau$.

Lecture 17: Interacting particle systems [non-examinable]

We have already seen a few examples of multivariate SDEs, for instance, the Langevin process (14.16) or the two-dimensional run-and-tumble (11.11). The former is an example of a process in one-dimensional physical space but where the stochastic particle is described by multiple variables (its position $X(t)$ and its velocity $V(t)$). The latter is an example of a process describing a single particle in a higher-dimensional physical space. A third class of multivariate SDEs, which is particularly relevant in biological applications, concerns processes involving many particles whose motions are coupled in some sense.

Consider a system of N particles (e.g., bacteria, ions, animals) with positions $X_i(t) \in \Omega \subseteq \mathbb{R}$, $i = 1, \dots, N$. The whole system is described by the N-dimensional configuration vector $\mathbf{X}(t) = (X_1(t), X_2(t), \dots, X_N(t))$. Suppose that $\mathbf{X}(t)$ satisfies the SDE

$$\mathrm{d}\mathbf{X}(t) = \mathbf{a}(\mathbf{X}(t), t)\mathrm{d}t + \boldsymbol{\sigma}(\mathbf{X}(t), t)\mathrm{d}\mathbf{W}(t), \tag{17.1}$$

where $\mathbf{a}(\mathbf{x}, t) \in \mathbb{R}^N, \boldsymbol{\sigma}(\mathbf{x}, t) \in \mathbb{R}^{N\times M}$ and $\mathbf{W}(t) = (W_1(t), \dots, W_M(t)) \in \mathbb{R}^M$, with $M \leq N$. Componentwise, we have

$$\mathrm{d}X_i(t) = a_i(\mathbf{X}(t), t)\mathrm{d}t + \sum_{j=1}^{M} \sigma_{ij}(\mathbf{X}(t), t)\mathrm{d}W_j(t), \qquad \text{for } i = 1, \dots, N. \tag{17.2}$$

Hence, each particle's position $X_i(t)$ is a diffusion process coupled with the other particles' positions $X_j(t)$ through its dependence of a_i and σ_{ij} on the whole process $\mathbf{X}(t)$ and/or the nondiagonal terms in $\boldsymbol{\sigma}$. In particular, if $\mathbf{W}(t) \in \mathbb{R}^N$ and $\sigma_{ij}(\mathbf{x}) = \sigma_i(\mathbf{x})\delta_{ij}$, then (17.2) reduces to

$$\mathrm{d}X_i(t) = a_i(\mathbf{X}(t), t)\mathrm{d}t + \sigma_i(\mathbf{X}(t), t)\mathrm{d}W_i(t), \qquad \text{for } i = 1, \dots, N,$$

and the positions become uncorrelated if

- the initial positions $X_i(0)$ are independent.
- the drift on particle i only depends on its own position, $a_i(\mathbf{x}) \equiv a_i(x_i)$.
- the variance is $\sigma_i(\mathbf{x}) \equiv \sigma_i(x_i)$.

We say that particles are *indistinguishable* if a_i and σ_i for $i = 1, \dots, N$ are all equal (independent of i) and the initial positions $X_i(0)$ are identically distributed.

17.1 Pairwise interactions

A common assumption in biological applications is that $\boldsymbol{\sigma}(\mathbf{x}) \equiv \sigma I_N$, where I_N is the N-dimensional identity matrix, and that the correlation between particles' occurs through the drift. In particular, the simplest model for interactions is to assume particles interact with a potential force. In particular, suppose that the N particles evolve according to

$$\mathrm{d}X_i(t) = f_i(\mathbf{X}(t))\mathrm{d}t + \sqrt{2D}\mathrm{d}W_i(t), \qquad \text{for } i = 1, \dots, N, \tag{17.3}$$

where D constant is their diffusion coefficient, and f_i is the drift or "force" acting on the ith particle.[8] Suppose that the drift is given by a pairwise interacting potential $u(r)$ that depends on the separation distance $r = |x_i - x_j|$ between two particles. Then we have that

$$f_i(\mathbf{x}) = -\chi \sum_{j=1, j\neq i}^{N} u'(|x_i - x_j|/\ell). \tag{17.4}$$

[8] Since we consider a first-order model, f_i in (17.3) does not have units of force, but we formally identify it as a reduced force. To see this, recall the Langevin SDE (15.1) with a potential force $-U'(x)$ and note that, in the overdamped limit, we obtained a diffusion for $X(t)$ with a drift $-U'(x)/\gamma$.

In (17.4), χ is the strength of the interaction potential, and ℓ is its range (which may depend on the number of particles N). Inserting (17.4) into (17.3) we arrive at

$$\mathrm{d}X_i(t) = -\chi \sum_{j=1, j\neq i}^{N} u'(|X_i(t) - X_j(t)|/\ell)\mathrm{d}t + \sqrt{2D}\mathrm{d}W_i(t), \qquad \text{for } i = 1, \dots, N. \tag{17.5}$$

The interaction potential is *attractive* at r if $u'(r) > 0$ and *repulsive* if $u'(r) < 0$. Common potentials include

- *Power-law potentials:* $u(r) = r^{-\nu}$. Examples include the Coulomb interaction ($\nu = 1$) or the *Lennard–Jones potential*
$$u(r) = (\epsilon/r)^{12} - (\epsilon/r)^6,$$
often used to model molecular or atomic interactions. This corresponds to short-range repulsion (power 12) and long-range attraction (power 6).

- *Exponential potentials:* $u(r) = \exp(-r/\epsilon)$. For example, the *Morse potential*
$$u(r) = -C_a \exp(-r/\ell_a) + C_r \exp(-r/\ell_r), \tag{17.6}$$
is commonly used in flocking models. In (17.6), C_a, C_r and ℓ_r,ℓ_r are the strengths and the typical lengths of attraction and repulsion, respectively [6]. Birds in a flock or fish in a school want to stay in the group while avoiding colliding with each other. This again corresponds to the regime of long-range attraction and short-range repulsion (which requires $C_r/C_a > 1$ and $\ell_r/\ell_a < 1$).

- *Hard-core potentials:* in this case, particles either do not feel each other or they repel each other with infinity force
$$u(r) = \begin{cases} +\infty, & r < \epsilon, \\ 0, & r > \epsilon, \end{cases} \tag{17.7}$$
Here ϵ may represent the diameter or typical size of particles. This way, if the distance between two particles is less than ϵ, this corresponds to particles overlapping each other. The infinite force ensures that no overlaps are allowed (as in the hard-sphere system (17.22)). For this reason, hard-core potentials are useful to model biological organisms (for which models that assume they are point particles would be poor).

We may use (12.12) to write down the Fokker–Planck equation associated with (17.5). To this end, we consider the joint probability density $P_N(x_1, x_2, \dots, x_N, t) = P_N(\mathbf{x}, t)$ of the N particle system. Using (12.12), we have that it satisfies

$$\frac{\partial P_N}{\partial t} = \sum_{i=1}^{N} \frac{\partial}{\partial x_i}\left[D\frac{\partial P_N}{\partial x_i} - f_i(\mathbf{x})P_N\right] = \nabla_{\mathbf{x}} \cdot \left[D\nabla_{\mathbf{x}} P_N + \nabla_{\mathbf{x}} U_N(\mathbf{x}) P_N\right], \tag{17.8}$$

where

$$U_N(x_1, \dots, x_N) = \chi \sum_{1 \leq i < j \leq N} u(|x_i - x_j|/\ell) \tag{17.9}$$

is the total interaction potential of the system. If the physical space for the particles' positions is $X_i(t) \in \Omega$, then the Fokker–Planck equation (17.8) is defined in Ω^N. If the number of particles N is large, as is often the case in real applications, (17.8) becomes computationally and analytically intractable. We must therefore find ways to simplify it and obtain a more manageable model.

We define the kth probability density marginal as

$$P_N^k(x_1,\dots,x_k,t) = \int_{\Omega^{N-k}} P_N(\mathbf{x},t)\,\mathrm{d}x_{k+1}\dots\mathrm{d}x_N. \tag{17.10}$$

We proceed to reduce the dimensionality of (17.8) by looking at the marginal probability density function of one particle (the first particle, say) given by

$$p(x_1,t) := P_N^1(x_1,t) = \int_{\Omega^{N-1}} P_N(\mathbf{x},t)\,\mathrm{d}x_2\dots\mathrm{d}x_N. \tag{17.11}$$

Note that the particle choice is unimportant since P_N is invariant with respect to permutations of particle labels. Integrating (17.8) over $x_2,\dots,x_N$ and applying the divergence theorem gives

$$\frac{\partial p}{\partial t}(x_1,t) = \frac{\partial}{\partial x_1}\left[D\frac{\partial p}{\partial x_1} + G(x_1,t)\right], \tag{17.12}$$

where G is given by

$$\begin{aligned} G(x_1,t) &= \chi\int_{\Omega^{N-1}} P_N(x_1,x_2,\dots,x_N,t)\sum_{j=2}^{N} u'(|x_1-x_j|/\ell)\,\mathrm{d}x_2\dots\mathrm{d}x_N \\ &= \chi(N-1)\int_{\Omega} P_N^2(x_1,x_2,t)u'(|x_1-x_2|/\ell)\,\mathrm{d}x_2, \end{aligned} \tag{17.13}$$

where

$$P_N^2(x_1,x_2,t) = \int_{\Omega^{N-2}} P_N(\mathbf{x},t)\,\mathrm{d}x_3\dots\mathrm{d}x_N, \tag{17.14}$$

is the two-particle density function, which gives the joint probability density of particle 1 being at position x_1 and particle 2 being at x_2. An equation for P_2 can be written from (17.8), but this depends on P_N^3, the three-particle density function. This results in a hierarchy (the BBGKY hierarchy [14]) of N equations for the set of k-particle density functions ($k=1,\dots,N$), the last of which is (17.8) itself.

17.2 The mean-field approximation

To obtain a practical model, a common approach is to truncate this hierarchy at a certain level to obtain a closed system. In particular, closure approximations in which the kth particle density function P_N^k is replaced by an expression involving lower marginals P_N^s, $s<k$, are commonly used. The simplest and most common closure approximation is to assume that particles are not correlated at all in evaluating the interaction term $\mathbf{G}$, that is,

$$P_N^2(x_1,x_2,t) = p(x_1,t)p(x_2,t). \tag{17.15}$$

It can be shown that (17.15) holds in the limit of $N\to\infty$ under the so-called *mean-field scaling* of the potential [14], which corresponds to

$$\chi = \frac{1}{N}, \qquad \ell = O(1). \tag{17.16}$$

This means that each particle interacts weakly with all others and that the interaction strength goes to zero as $N\to\infty$. This result is known as the *propagation of chaos*; it holds under suitable conditions on the potential u (e.g., Lipschitz).

Substituting (17.15) and (17.16) into (17.13) gives

$$G(x_1,t) = \frac{N-1}{N}\,p(x_1,t)\int_{\Omega} p(x_2,t)u'(|x_1-x_2|)\,\mathrm{d}x_2. \tag{17.17}$$

Combining this with the equation for p in (17.12) gives, taking $N \to \infty$,

$$\begin{aligned}\frac{\partial p}{\partial t}(x_1,t) &= \frac{\partial}{\partial x_1}\left[D\frac{\partial p}{\partial x_1} + p(x_1,t)\int_\Omega p(x_2,t)u'(|x_1-x_2|)\,\mathrm{d}x_2\right] \\ &= \frac{\partial}{\partial x_1}\left[D\frac{\partial p}{\partial x_1} + p\frac{\partial}{\partial x_1}(u*p)\right],\end{aligned} \tag{17.18}$$

where $u*p$ denotes the convolution $\int u(|x-y|)p(y,t)\mathrm{d}y$. The mean-field closure is often used implicitly with (17.18) written down directly rather than being derived from (17.8). The reasoning goes as follows: if $p(x,t)$ is the probability of finding a particle at x, the force on a particle at x_1 is given by multiplying the force due to another particle at x_2 by the density of particles at x_2 and integrating over all positions x_2.

17.3 Local interactions

In the following, we reintroduce the general scaling of the potential of strength χ and range ℓ. Suppose that the interaction potential $\chi u(r/\ell)$ is repulsive and short-ranged. In particular, we assume that it decays as $u = O(r^{-1-\delta}$ for some $\delta > 0$ as $r \to \infty$. Introducing the change of variable $x_2 = x_1 + \ell y$ with $\ell \ll 1$ in (17.17) and expanding $p(x_2,t)$ about x_1 gives[9]

$$G(x_1,t) = -\chi(N-1)p(x_1,t)\int_{\mathbb{R}}\left[p(x_1,t) + \ell y\partial_{x_1}p(x_1,t)\right]u'(|y|)\,\mathrm{d}y + \cdots, \tag{17.19}$$

where we can extend the integral with respect to variable y to the whole space since the potential u is localised near the origin and decays at infinity. Noting that the potential is a symmetric function, the leading-order term in the integral vanishes, and after integrating by parts in the next term, we obtain

$$G(x_1,t) \sim \chi(N-1)\ell p(x_1,t)\partial_{x_1}p(x_1,t)\int_{\mathbb{R}} u(|y|)\,\mathrm{d}y. \tag{17.20}$$

Inserting (17.20) into (17.12), we find that the marginal density function satisfies the following nonlinear diffusion equation

$$\frac{\partial p}{\partial t}(x_1,t) = \frac{\partial}{\partial x_1}\left[\bar{D}(p)\frac{\partial p}{\partial x_1}\right], \qquad \bar{D}(p) = D + \ell\chi(N-1)\|u\|_1 p, \tag{17.21}$$

where $\|u\|_1 = \int u(|y|)\mathrm{d}y$.

A few comments are in order. Starting from the mean-field equation (17.18) (which is rigorous as $N \to \infty$ under the mean-field scaling (17.16)), we relaxed the conditions on the interaction potential u and considered what happens to the mean-field model as the range of the potential ℓ is reduced (i.e., the interaction force becomes localised). The result is that the nonlocal term in the mean-field model (17.18) becomes a local nonlinearity; namely, it appears as a density-dependent term in the diffusivity in (17.21). In particular, we find that the effective diffusivity increases with the density and that the increase is proportional to the magnitude of the interactions in the system: each pairwise interaction is of order $\ell\chi\|u\|_1$, and particle one "sees" $N-1$ of such interactions. Note that this derivation is only valid for potentials $u \in L^1(\mathbb{R}_+)$, so, for example, it is not valid for power-law potentials or the hard-sphere potentials (17.7). There are alternative approximate methods for dealing with singular repulsive interactions [5].

It may be slightly counterintuitive that repulsive interactions between particles lead to enhanced diffusion instead of restricted diffusion. The key is that while an individual particle

[9] We have used that $u'(|x_1-x_2|/\ell) \to -u'(|y|)/\ell$ and $\mathrm{d}x_2 \to \ell\mathrm{d}y$.

does indeed have a reduced *self-diffusion* coefficient due to repulsive interactions, the *collective* diffusivity of the system increases. In deriving the equation for the marginal $p(x,t)$ from that of $P^N(\mathbf{x},t)$, we obtain an approximation to the collective diffusion coefficient $\bar{D}(p)$, which describes the evolution of the concentration of the system [5]. To extract the self-diffusion coefficient, which describes the evolution of a single tagged particle, one needs to make one particle distinguishable from the rest (hence deriving a system of PDEs for the tagged particle and the remaining $N-1$ particles).

17.4 Metropolis–Hastings SSA for a system of Brownian hard-spheres

We present an application of the MH algorithm SSA 15.3 to a high-dimensional system. In particular, consider a system of N hard spheres of diameter ϵ in $\mathbb{R}^d$, with positions $\mathbf{X}_i(t)$, each evolving according to an OU process with independent Brownian motions but with non-overlapping constraints:

$$\mathrm{d}\mathbf{X}_i(t) = -\nabla U(\mathbf{X}_i(t))\mathrm{d}t + \sqrt{2D}\mathrm{d}\mathbf{W}_i(t), \qquad i=1,\dots,N, \tag{17.22a}$$

$$\|\mathbf{X}_i(t) - \mathbf{X}_j(t)\| \geq \epsilon, \qquad i \neq j, \tag{17.22b}$$

where $U:\mathbb{R}^d \to \mathbb{R}$ is a confining potential. The Fokker–Planck equation for the probability density $p(\vec{x},t)$, $\vec{x}=(\mathbf{x}_1,\mathbf{x}_2,\dots,\mathbf{x}_N)\in\mathbb{R}^{dN}$, associated to (17.22) is given by

$$\frac{\partial p}{\partial t}(\vec{x},t) = \sum_{i=1}^{N} \nabla_{\mathbf{x}_i}\cdot\left\{p\nabla_{\mathbf{x}_i}\left[\log p + U(\mathbf{x}_i)\right]\right\}. \tag{17.23}$$

The non-overlapping constraints between hard spheres mean that the joint probability density p takes values in the configuration domain

$$\mathbb{R}^{dN}_{\epsilon} = \left\{\vec{x}\in\mathbb{R}^{dN} : \|\mathbf{x}_i - \mathbf{x}_j\| \geq \epsilon, \forall j\neq i\right\}. \tag{17.24}$$

The stationary density of (17.23) is given by

$$p_\infty(\vec{x}) = C\exp\left[\sum_{i=1}^{N} U(\mathbf{x}_i)\right] \qquad \text{in} \qquad \mathbb{R}^{dN}_{\epsilon} \tag{17.25}$$

where C is the normalisation constant. At this point, one may be tempted to say that we are done since we have found the stationary density p_∞. However, the normalisation constant C is still unknown and, because p_∞ is defined in a high-dimensional and complicated domain $\mathbb{R}^{dN}_{\epsilon}$, a direct calculation approach (that is, integrate (17.25) over $\mathbb{R}^{dN}_{\epsilon}$ and impose its integral is equal to one) is not feasible. We may extend the domain of definition of p_∞ to $\mathbb{R}^{dN}$ by writing

$$p_\infty(\vec{x}) = C\exp\left[-H(\vec{x})\right] \qquad \text{in} \qquad \mathbb{R}^{dN}, \tag{17.26}$$

with

$$H(\vec{x}) = \begin{cases} \sum_{i=1}^{N} U(\mathbf{x}_i), & \vec{x}\in\mathbb{R}^{dN}_{\epsilon}, \\ \infty, & \text{otherwise.} \end{cases} \tag{17.27}$$

The MH algorithm samples configurations according to the density (17.26) by constructing a Markov chain over the configuration space as follows:

- Select a particle i at random and generate a candidate $\mathbf{x}_i'$ for x_i according to $\mathbf{x}_i' = \mathbf{x}_i + \delta\boldsymbol{\xi}$, with $\boldsymbol{\xi}=(\xi_1,\dots,\xi_d)$ and $\xi_k\sim\mathcal{N}(0,1)$ for a given δ.
- Compute the energy difference ΔH between the current and modified configurations: $\Delta H = H(\vec{x}\,') - H(\vec{x})$, where $\vec{x}\,' = (\mathbf{x}_1,\dots,\mathbf{x}_i',\dots,\mathbf{x}_N)$.

- Accept the move from $\vec{x}$ to $\vec{x}\,'$ with probability $p_{acc} = \min[1, \exp(-\Delta H)]$.

Observe that the second step only requires considering potential overlaps between $\mathbf{x}_i'$ and $\mathbf{x}_j$ for $j \neq i$ and computing $U(\mathbf{x}_i') - U(\mathbf{x}_i)$. Also note that, in the case of hard spheres, the chain generated by the MH algorithm satisfies the non-overlapping constraints (17.22b) since an overlap would result in $H(\vec{x}') = +\infty$ and thus $\exp(\Delta H) = 0$. Here the candidate move is a single particle local move, but other options are possible and may lead to faster convergence. Examples include a single particle global move or moving multiple particles at once.

Appendix A: Markov processes: Definitions and notation

We begin the second half of the course by collating some important definitions that have already come out in the course in the context of discrete processes. 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. $S = \mathbb{Z}_{\geq}$ in chemical reaction networks) or continuous (e.g. $S = \mathbb{R}$ in Brownian motion).

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

$$\mathbb{P}\{X(t) \in [a,b]\} = \int_a^b p(x,t)\mathrm{d}x.$$

Note the capital letter to denote the PMF and distinguish it from a PDF; we will stick to this notation for the remainder of the course.

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 $h \geq 0$.

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. A Markov process with discrete state space is called a *continuous-time Markov chain*.

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 $X(t)$ [20, §2.2]. We will see an example of this in Example Sheet 3.

Definition A.3 (Transition probability)**.** *We define the* transition probability *for a continuous-time Markov chain in* $\mathbb{Z}$ *as*

$$P(i,t|j,s) = \mathbb{P}\{X(t) = i | X(s) = j\}, \quad \text{for all } i,j \in \mathbb{Z}, s < t. \tag{A.2}$$

Definition A.4 (Transition probability density)**.** *We define the* transition probability density function $p(x,t|y,s)$ *of a continuous-time Markov process in* $\mathbb{R}$ *as the conditional probability*

$$\mathbb{P}(X(t) \in \Gamma | X(s) = y) = \int_\Gamma p(x,t|y,s)\mathrm{d}x, \quad \text{for all } y, \Gamma \in \mathbb{R}, s \leq t. \tag{A.3}$$

In other words, $p(x,t|y,s)\mathrm{d}x$ *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:*

$$\begin{aligned} P(i,t|j,s) &= P(i,t-s|j,0), \qquad \text{for all } i,j \in \mathbb{Z}, s \leq t, \\ p(x,t|y,s) &= p(x,t-s|y,0), \qquad \text{for all } x,y \in \mathbb{R}, s \leq t. \end{aligned}$$

Examples of time-homogeneous processes are the Ornstein–Uhlenbeck process and Brownian motion.

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*

$$P(i,t|j,s) = \sum_{k\in\mathbb{Z}} P(i,t|k,u)P(k,u|j,s), \qquad s \leq u \leq t, \forall i,j \in \mathbb{Z}. \tag{A.4}$$

Similarly, the transition probability density of a Markov process in $\mathbb{R}$ satisfies the CK equation

$$p(x,t|y,s) = \int_{\mathbb{R}} p(x,t|z,u)p(z,u|y,s)\mathrm{d}z, \qquad s \leq u \leq t, \forall x,y \in \mathbb{R}. \tag{A.5}$$

The derivation of the CK equation (A.5) is based on the Markovian assumption and properties of conditional probability (see [20, p.35]).

Consider a Markov process $X(t)$ in $\mathbb{R}$ with initial density $p(x,0) := \rho_0(x)$. We have that

$$p(x,t) = \int_{\mathbb{R}} p(x,t|y,0)p(y,0)\mathrm{d}y = \int_{\mathbb{R}} p(x,t|y,0)\rho_0(y)\mathrm{d}y. \tag{A.6}$$

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 $p(x,t|x_0,0)$.

References

[1] L. J. Allen, *An Introduction to Stochastic Processes with Applications to Biology*, CRC Press, 2nd ed., 2010.

[2] S. S. Andrews and D. Bray, *Stochastic simulation of chemical reactions with spatial resolution and single molecule detail*, Physical biology, 1 (2004), p. 137.

[3] L. Bachelier, *Théorie de la spéculation*, Ann. Sci. Ecole Norm. S., 3 (1900), pp. 21–86.

[4] 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.

[5] M. Bruna and S. J. Chapman, *Diffusion of multiple species with excluded-volume effects*, The Journal of Chemical Physics, 137 (2012), p. 204116.

[6] J. A. Carrillo, M. Fornasier, G. Toscani, and F. Vecil, *Particle, kinetic, and hydrodynamic models of swarming*, in Mathematical Modeling of Collective Behavior in Socio-Economic and Life Sciences, G. Naldi, L. Pareschi, and G. Toscani, eds., Birkhäuser Boston, MA, 2010, pp. 297–336.

[7] S. Chib and E. Greenberg, *Understanding the Metropolis–Hastings algorithm*, Am. Stat., 49 (1995), pp. 327–335.

[8] 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.

[9] R. Erban and S. J. Chapman, *Stochastic modelling of reaction–diffusion processes*, vol. 60, Cambridge University Press, 2020.

[10] C. Gardiner, *Stochastic Methods*, Springer Berlin, Heidelberg, 2009.

[11] D. Higham and P. Kloeden, *An introduction to the numerical simulation of stochastic differential equations*, SIAM, 2021.

[12] D. J. Higham, *An algorithmic introduction to numerical simulation of stochastic differential equations*, SIAM Review, 43 (2001), pp. 525–546.

[13] S. Iyer-Biswas and A. Zilman, *First-passage processes in cellular biology*, Advances in Chemical Physics, 160 (2016), pp. 261–306.

[14] P.-E. Jabin and Z. Wang, *Mean Field Limit for Stochastic Particle Systems*, in Active Particles, Volume 1, Birkhäuser, Cham, Cham, 2017, pp. 379–402.

[15] P. Langevin, *Sur la théorie du mouvement Brownien*, C. R. Acad. Sci. Paris, 146 (1908).

[16] R. M. Mazo, *Brownian Motion: Fluctuations, Dynamics, and Applications*, Clarendon Press, Oxford, 2002.

[17] R. C. Merton, *Option pricing when underlying stock returns are discontinuous*, Journal of Financial Economics, 3 (1976), pp. 125–144.

[18] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *Equation of state calculations by fast computing machines*, J. Chem. Phys., 21 (1953), pp. 1087–1092.

[19] H. G. Othmer, S. R. Dunbar, and W. Alt, *Models of dispersal in biological systems*, Journal of Mathematical Biology, 26 (1988), pp. 263–298.

[20] G. A. Pavliotis, *Stochastic processes and applications*, vol. 60, Springer, 2014.

[21] G. O. Roberts, A. Gelman, and W. R. Gilks, *Weak convergence and optimal scaling of random walk Metropolis algorithms*, Ann. Appl. Probab., 7 (1997), pp. 110–120.

[22] P. Strating, *Brownian dynamics simulation of a hard-sphere suspension*, Physical Review E, 59 (1999), p. 2175.

[23] 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.

[24] M. von Smoluchowski, *Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen*, Ann. Phys., 326 (1906), pp. 756–780.