3 Stochastic differential equations

Stochastic differential equations (SDEs) come up naturally when we attempt to add diffusion into our models of chemical reactions. They also arise when we wish to approximate the stochastic dynamics of a CRN in the limit where there are large numbers of molecules, and reactions occur at high frequency.

3.1 Brownian motion

In order to understand SDEs, we need some basic understanding of Brownian motion. The study of Brownian motion takes us beyond the scope of this course, so the aim here will be to summarise key results, and provide some intution.

Brownian motion, which we will denote by $\{W(t)\}_{t\geq 0}$, is a continuous-time stochastic process which takes values in $\mathbb{R}$ (or $\mathbb{R}^n$ in the case of n-dimensional Brownian motion). Note that this is unlike the processes we have looked at so far which take values in subsets of $\mathbb{Z}^n$.

As always, we have two ways of thinking about a continuous-time stochastic process: (i) as a sequence of random variables (on the same probability space); or (ii) in terms of its sample paths.

There are various characterisations of Brownian motion which can all be shown to be equivalent. We will focus on 1D Brownian motion, as n-dimensional Brownian motion is simply a vector n independent Brownian motions. In one dimension, Brownian motion can be defined via the following properties:

1. It takes the value 0 at time 0, i.e., $W(0) = 0$ almost surely.
2. It has continuous sample paths almost surely. I.e., given our underlying probability space Ω, associated with each $\omega \in \Omega$ is the function $t \mapsto W_\omega(t)$, and these functions are continuous with probability 1.
3. $W(t)$ has stationary increments, i.e., $W(t+h) - W(s+h)$ has the same distribution as $W(t) - W(s)$ for each s, t and h such that all these quantities are defined.
4. $W(t)$ has independent increments, i.e., given any $0 \leq t_1 < t_2 \leq t_3 < t_4$, $W(t_2) - W(t_1)$ and $W(t_4) - W(t_3)$ are independent random variables.
5. $W(t) - W(s)$ is normally distributed with mean 0 and variance $t - s$ for any $0 \leq s < t$.

In fact, the final assumption can be replaced by $\mathbb{E}[W(1)] = 0$ and $\mathrm{Var}\,(W(1)) = 1$: we do not need to explicitly specify that increments are normally distributed – rather this comes naturally from the assumption of stationary, independent increments and continuous paths.

It is a nontrivial fact that there is, indeed, a stochastic process with these properties; and that these properties uniquely determine the process.

One astonishing feature of Brownian motion is that sample paths are, almost surely, *nowhere differentiable*. This means that if we fix some $\omega \in \Omega$, then with probability 1 the limit

$$\lim_{\Delta t \to 0} \frac{W_\omega(t + \Delta t) - W_\omega(t)}{\Delta t}$$

is undefined for all t.

This extreme "jaggedness" of the paths leads to problems with any limiting process where we look at Brownian motion over smaller and smaller time-intervals. But this is precisely what we need to do when we use Brownian motion as a model of diffusion, or of "noise" in some otherwise deterministic system. Resolving this problem is at the heart of the study of stochastic differential equations (SDEs).

While individual paths are badly behaved, the fact that increments are independent, stationary, and normally distributed means that taken collectively paths are well-behaved. It is this fact which forms the basis of stochastic integration and allows us to define solutions of SDEs.

3.2 SDEs: an informal introduction

Consider the ODE

$$\frac{dx}{dt} = f(x, t)\,. \tag{16}$$

Here x is an unknown function of t. If we are given the value of x at some time t_0, we may hope for a unique solution, and indeed, you have seen sufficient conditions for existence and uniqueness in previous courses. We could rewrite (16) as an integral equation

$$x(t) = x(t_0) + \int_{t_0}^{t} f(x(s), s)\, ds\,. \tag{17}$$

Recall that it is the integral form which is the most natural starting point for proofs of existence and uniqueness of solutions. Note, however, that for (16) to make sense x must be differentiable, whereas this is not automatic in (17). However, if we *assume* that x and f are continuous, then (17) indeed implies that x must be differentiable (by the fundamental theorem of calculus), and so with these assumptions (16) and (17) are equivalent.

We can also write (16) in the form

$$dx(t) = f(x(t), t)\, dt\,. \tag{18}$$

We can think of this as a convenient *shorthand* to indicate that we have not decided whether to interpret the equation as the differential equation (16) or the integral equation (17)! It is also a natural starting point for numerical approaches to solving (16). Noting that $dx(t) = x(t + dt) - x(t)$, and replacing the infinitesimal dt with the small Δt we have the approximation

$$x(t + \Delta t) \simeq x(t) + f(x(t), t)\, \Delta t\,,$$

which forms the basis of the Euler forward difference approach to solving ODEs numerically. Of course, the point here is that the error in this approximation is $o(\Delta t)$, i.e., approaches zero faster than Δt as $\Delta t \to 0$.

Let us now consider what happens if we were to add a random "noise" term to (18). Since the noise is acting over an infinitesimally small time-interval, we expect it to be infinitesimally small in some sense. Moreover the strength of the noise might depend on the current state of the system. We would like to write something like

$$\mathrm{d}\mathbf{x}(\mathrm{t}) = \mathrm{f}(\mathbf{x}(\mathrm{t}), \mathrm{t})\,\mathrm{dt} + \mathrm{g}(\mathbf{x}(\mathrm{t}), \mathrm{t})\,\mathrm{d}W(\mathrm{t})\,.$$

Here, if $\mathbf{X}(\mathrm{t})$ takes values in $\mathbb{R}^{\mathrm{n}}$, then g is in general some $\mathrm{n} \times \mathrm{m}$ matrix of functions, and $\mathrm{d}W(\mathrm{t})$ is meant to represent some m-vector of "small", indendent noise terms. In fact, for very good reasons connected with the nice way that increments of Brownian motion behave, we often would like to think of each component of $\mathrm{d}\,W$ as a "small" increment of Brownian motion. (Recall that Brownian motion is essentially the *only* stochastic process with continuous paths and stationary, independent increments.)

There are two issues we need to consider carefully about the last equation. Firstly, as $W(\mathrm{t})$ is a stochastic process, we now expect $\mathbf{x}(\mathrm{t})$ to be a stochastic process with, we hope, some nice properties like almost surely continuous sample paths. To remind us that the output is now a stochastic process we can write

$$\mathrm{d}\mathbf{X}(\mathrm{t}) = \mathrm{f}(\mathbf{X}(\mathrm{t}), \mathrm{t})\,\mathrm{dt} + \mathrm{g}(\mathbf{X}(\mathrm{t}), \mathrm{t})\,\mathrm{d}W(\mathrm{t})\,. \tag{19}$$

This is just a cosmetic change, but an important one, because it reminds us that a "solution" to an initial value problem is no longer a single function of time, but a stochastic process. But we still face the difficulty of how to interpret, and solve, (19). We might try dividing by dt and taking a limit to get

$$\frac{\mathrm{d}\mathbf{X}(\mathrm{t})}{\mathrm{dt}} = \frac{\mathbf{X}(\mathrm{t}+\mathrm{dt}) - \mathbf{X}(\mathrm{t})}{\mathrm{dt}} = \mathrm{f}(\mathbf{X}(\mathrm{t}), \mathrm{t}) + \mathrm{g}(\mathbf{X}(\mathrm{t}), \mathrm{t})\frac{\text{“}\mathrm{d}W(\mathrm{t})\text{”}}{\mathrm{dt}}\,.$$

But this is doomed to fail: although the expression above looks like a set of ordinary differential equations, one for each point ω in our probability space, we have already observed that the sample paths of Brownian motion are almost surely everywhere non-differentiable, and so for any fixed ω, $\frac{\mathrm{d}W(\mathrm{t})}{\mathrm{dt}}$ is almost surely undefined for every t.

On the other hand, we might still ask: is there a stochastic process $\mathbf{X}(\mathrm{t})$ which solves the integral version of (19), given some initial state? Namely, can we make sense of the following equation?

$$\mathbf{X}(\mathrm{t}) = \mathbf{X}(\mathrm{t}_0) + \int_{\mathrm{t}_0}^{\mathrm{t}} \mathrm{f}(\mathbf{X}(\mathrm{s}), \mathrm{s})\,\mathrm{ds} + \int_{\mathrm{t}_0}^{\mathrm{t}} \mathrm{g}(\mathbf{X}(\mathrm{s}), \mathrm{s})\,\mathrm{d}W(\mathrm{s})\,. \tag{20}$$

The good news is that this can be made to work, provided we are very careful in how we interpret (20). If we try to interpret it one sample path at a time we run into difficulties with the final integral in (20) because of the non-differentiability of sample paths of $W(\mathrm{t})$. But the theory of stochastic integration, in particular in the form developed by Itô, gives us a way of interpreting such integrals, and solving SDEs such as (19). Thus (19) does have a well-defined meaning, but we must not believe that the quantity $\mathrm{d}W(\mathrm{t})$ truly represents the differential of any quantity in the usual sense.

We won't go into any of the details of stochastic integration here, but instead consider the approximate form of (19):

$$\mathbf{X}(\mathrm{t}+\Delta\mathrm{t}) - \mathbf{X}(\mathrm{t}) \simeq \mathrm{f}(\mathbf{X}(\mathrm{t}), \mathrm{t})\,\Delta\mathrm{t} + \mathrm{g}(\mathbf{X}(\mathrm{t}), \mathrm{t})\,\Delta W(\mathrm{t})\,, \tag{21}$$

where Δt is small. Note that $\Delta W(t) = W(t+\Delta t) - W(t)$ is just an increment of (m-dimensional) Brownian motion, so each component is a normally distributed random variable with mean 0 and variance Δt. Moreover these components are independent of each other and of $\mathbf{X}(s)$ for all $s \leq t$. We may thus replace $\Delta W(t)$ in (21) with $\sqrt{\Delta t}\xi$, where ξ is a vector of m independent standard normal variables, which are also independent of $\mathbf{X}(t)$, and write:

$$\mathbf{X}(t + \Delta t) = \mathbf{X}(t) + f(\mathbf{X}(t), t)\,\Delta t + g(\mathbf{X}(t), t)\,\sqrt{\Delta t}\,\xi\,, \tag{22}$$

We can take (22) as the **computational definition of the SDE** (19). Note that it makes no sense to subtract $\mathbf{X}(t)$ from both sides, divide through by Δt, and let $\Delta t \searrow 0$. However, we expect, given some conditions on f and g, that the solutions of (22) converge in some sense to the "true" solution of the SDE as $\Delta t \searrow 0$. (Making this precise is the subject of stochastic calculus, and the solution that we converge to is a solution in the Itô sense.) What is important for us is that when we read (22) – and other equations derived from it below – we should remember that these are approximations which become, in a precise sense but one we won't study, exact in the limit $\Delta t \searrow 0$.

From (22), we can hope to compute approximate sample paths of the process $\mathbf{X}(t)$ by choosing choosing Δt to be sufficiently small, and sampling from a standard normal distribution at each time step. Moreover, we might hope to compute the evolution of moments of $\mathbf{X}(t)$. For example, taking expectations in (22) gives

$$\mathbb{E}(\mathbf{X}(t + \Delta t)) \simeq \mathbb{E}(\mathbf{X}(t)) + \mathbb{E}(f(\mathbf{X}(t), t))\,\Delta t\,,$$

because $\mathbb{E}(\xi) = 0$, and ξ is independent of $\mathbf{X}(t)$. From this we can recover an ODE for the evolution of $\mathbb{E}(\mathbf{X}(t))$, namely,

$$\frac{d}{dt}\mathbb{E}(\mathbf{X}(t)) = \mathbb{E}(f(\mathbf{X}(t), t))\,.$$

If f was a linear function of $\mathbf{X}$ for example, we would recover the original ODE $d\overline{\mathbf{X}}(t) = f(\overline{\mathbf{X}}(t), t)\,dt$ for the evolution of $\overline{\mathbf{X}}(t) := \mathbb{E}(\mathbf{X}(t))$. In this very special case, the stochastic component of the SDE (19) contributes nothing to the evolution of the mean.

In a similar vein, if we know the value of $\mathbf{X}(t)$ to be x, we can ask for the expectation of $\mathbf{X}(t+\Delta t)$ a short time later. Taking conditonal expectations in (22), we get

$$\mathbb{E}[\mathbf{X}(t + \Delta t) - x \,|\, \mathbf{X}(t) = x] = f(x, t)\,\Delta t + g(x, t)\,\sqrt{\Delta t}\,\mathbb{E}[\xi] = f(x, t)\,\Delta t\,. \tag{23}$$

We expect this equation to become exact in the limit $\Delta t \searrow 0$ in the sense that $\mathbb{E}[\mathbf{X}(t + \Delta t) - x \,|\, \mathbf{X}(t) = x] = f(x, t)\,\Delta t + o(\Delta t)\,.$

It turns out to be useful also to ask about higher moments of $\mathbf{X}(t + \Delta t) - x$ conditioned on $X(t) = x$. For example, restricting attention to the 1D case, we find that

$$\mathbb{E}[(X(t + \Delta t) - x)^2 \,|\, X(t) = x] = g(x, t)^2\,\Delta t + o(\Delta t)\,. \tag{24}$$

Thus the random term in the SDE contributes and $O(\Delta t)$ term to the evolution of the second moment, namely the variance, of X. The same holds in higher dimensions

For $n \geq 3$, $\mathbb{E}[(X(t + \Delta t) - x)^n \,|\, X(t) = x] = o(\Delta t)\,.$ We will use all of these observations in the derivation of the Fokker-Planck equation below.

3.3 The Fokker-Planck equation (forward Kolmogorov equation)

In fact, we can do better than simply writing down ODEs for the evolution of moments of $\mathbf{X}(t)$ where $\mathbf{X}(t)$ is the stochastic process which solves (19). We can try to write down a PDE for the evolution of the pdf of $\mathbf{X}(t)$. We will sketch the derivation of this PDE, termed the forward Kolmogorov equation or the **Fokker-Planck equation**, but first let's write it down and do some examples. Once we understand the meaning of this equation, we will understand why it is also called the forward Kolmogorov equation, even though it looks very different from (10).

Let $p(x,t)$ denote the probability density function (pdf) of $\mathbf{X}(t)$ conditional on $\mathbf{X}(s) = y$ at some time $s < t$. I.e., $p(x,t)\,dx = \mathbb{P}(\mathbf{X}(t) \in [x, x+dx] \,|\, \mathbf{X}(s) = y]$. Here, and in arguments to follow, in general dx and $[x, x+dx]$ refer to a products of infinitesimal intervals, i.e., $dx = dx_1 \cdots dx_n$, and $[x, x+dx] = [x_1, x_1 + dx_1] \times \cdots \times [x_n, x_n + dx_n]$.

In the one-dimensional case we find that p evolves according the equation,

$$\frac{\partial}{\partial t}p(x,t) = -\frac{\partial}{\partial x}[f(x,t)p(x,t)] + \frac{\partial^2}{\partial x^2}[D(x,t)p(x,t)]\,, \tag{F-P1}$$

where $D(x,t) := \frac{1}{2}g(x,t)^2$. In n dimensions, recall that $f(x,t)$ is a vector with n components, and $g(x,t)$ is an $n \times m$ matrix. Define the $\mathbf{D}(x,t) := \frac{1}{2}g(x,t)\,g(x,t)^t$. Then we have

$$\frac{\partial}{\partial t}p(x,t) = -\sum_{i=1}^{n}\frac{\partial}{\partial x_i}[f_i(x,t)p(x,t)] + \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial^2}{\partial x_i \partial x_j}[D_{ij}(x,t)p(x,t)]\,. \tag{F-P}$$

We see that the F-P equation defines the evolution of probabilities as we move forward in time, just as the forward Kolmogorov equation (10) did. The only difference is that this time we are considering a continuous random variable, and hence evolution of a pdf, rather than a discrete random variable, and hence a probability mass function.

Before we sketch the derivation of the F-P equation in 1D, let us consider its meaning. It tells us about the evolution of a probability density function (or more generally a measure) in time. In other words, if we take some distribution of initial conditions for the SDE (19) and follow these conditions for some time, the solution of the F-P equation gives us the corresponding distribution at a later time t. Note that we can write the right-hand side of the 1D F-P equation as

$$-\frac{\partial}{\partial x}\left(f(x,t)p(x,t) - \frac{\partial}{\partial x}[D(x,t)p(x,t)]\right)\,.$$

If we denote the quantity in brackets above by $Q(x,t)$, we can interpret it as the **probability flux**. The equation then reads

$$\frac{\partial}{\partial t}p(x,t) + \frac{\partial}{\partial x}Q(x,t) = 0\,,$$

which can be regarded as a **conservation equation** for the probability density: roughly, the change in probability in a small region of space is equal to the flux through the boundaries of this region. The interpretation in higher dimensions is similar. You will have seen the same idea in derivations of the heat equation which can be regarded as a conservation law for heat.

Let us now consider some basic examples.

Example 3.1 (**F-P equation for a deterministic system)**. Consider a 1D equation $dx = f(x,t)\,dt$, i.e., $g \equiv 0$. In this case (19) is a deterministic ODE, and the 1D Fokker-Planck equation reduces to:

$$\frac{\partial p(x,t)}{\partial t} = -\frac{\partial}{\partial x}[f(x,t)p(x,t)]\,.$$

This first order PDE is easily solved by the method of characteristics, at least in theory. The equations still make sense: even though the evolution is deterministic, we can choose initial conditions according to a probability distribution, and ask how this distribution evolves over time, as the initial conditions are transported by trajectories of the ODE.

Example 3.2 (**F-P equation for Brownian motion)**. Let us again consider the 1D case, this time in the special case where $f \equiv 0$ and $g \equiv 1$. In this case (19) just describes standard Brownian motion (provided we set the intial value to 0), and the 1D Fokker-Planck equation reduces to the simplest form of the diffusion equation, namely,

$$\frac{\partial p(x,t)}{\partial t} = \frac{1}{2}\frac{\partial^2 p(x,t)}{\partial x^2}\,.$$

(In fact, a stochastic process which solves (19) in the autonomous case, i.e., where f and g do not explicitly depend on time and satisfy some mild technical restrictions, is known as an Itô diffusion.)

If the motion describes the physical evolution of particles in space, then we can set boundary conditions by considering the conservation relation $\frac{\partial}{\partial t}p(x,t) + \frac{\partial}{\partial x}Q(x,t) = 0$. So, for example, if we have a reflecting boundary at $x = 0$, this would correspond to the no-flux condition $Q(0,t) = 0$ for all t.

For example, if we consider a domain $[0,1]$ with initial pdf uniform on $[0,1/2]$ and no-flux boundary conditions at 0 and 1 (i.e., homogeneous Neumann boudary conditions), then by Fourier series methods, you should check, as an exercise, that we obtain the solution

$$p(x,t) = 1 + \frac{4}{\pi}\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n+1)}\cos((2n+1)\pi x)\,\exp[-(2n+1)^2\pi^2 t/2]$$

Note that distribution rapidly approaches the uniform distribution.

Informal derivation of the F-P equation in 1D. We present an informal derivation of the F-P equation following [Erban and Chapman] and focussing only on the 1D case. All the steps can be made rigorous, provided some conditions on the coefficient functions f and g are satisfied.

Let us start by writing $p(x,t)$ in full form, $p(x,t\,|\,y,s)$. Next, we write down the corresponding Chapman-Kolmogorov equation:

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

Note that this equation makes sense as a statement about probabilities if we multiply both sides by dz and integrate over some set of z-values. In the end we will let $\Delta t \to 0$; and so we expect z to be close to x. It is tempting to Taylor expand $p(z, t+\Delta t \,|\, x, t)$ in $z - x$ and this is indeed what we will do later, when we derive the backward Kolmogorov equation. But in this case this is not helpful.

Instead, we take a smooth **test function** $\phi(z)$ with compact support, multiply and integrate both sides of the Chapman-Kolmogorov equation.

We get

$$\int_{\mathbb{R}} p(x, t+\Delta t \,|\, y, s)\phi(x)dx = \int_{\mathbb{R}} \left[\int_{\mathbb{R}} \phi(z)p(z, t+\Delta t \,|\, x, t)\, dz\right] p(x, t \,|\, y, s)\, dx\,. \tag{25}$$

where we have renamed the dummy variable of integration x instead of z on the left and switched the order of integration on the right.

Let us now Taylor expand ϕ about $z = x$ and consider the various terms in the inner integral on the right. Set $\phi(z) = \phi(x) + (z-x)\phi'(x) + \frac{(z-x)^2}{2}\phi''(x) + o((z-x)^2)$. We observe that:

$$\int_{\mathbb{R}} p(z, t+\Delta t \,|\, x, t)\, dz = 1\,.$$

(Given state x at time t, we must reach *some* state at time $t + \Delta t$.) Next, using (23),

$$\int_{\mathbb{R}} (z-x)p(z, t+\Delta t \,|\, x, t)\, dz = \mathbb{E}[X(t+\Delta t) - x \,|\, X(t) = x] = f(x, t)\Delta t + o(\Delta t)\,.$$

Similarly, using (24),

$$\int_{\mathbb{R}} (z-x)^2 p(z, t+\Delta t \,|\, x, t)\, dz = \mathbb{E}[(X(t) - x)^2 \,|\, X(t) = x] = (g(x, t))^2\Delta t + o(\Delta t)\,.$$

We now return to (25). Subtracting the zeroth order term on the right from both sides, and dividing through by Δt gives

$$\int_{\mathbb{R}} \phi(x)\frac{p(x, t+\Delta t \,|\, y, s) - p(x, t \,|\, y, s)}{\Delta t}dx = \int_{\mathbb{R}} [\phi'(x)f(x, t) + \frac{1}{2}\phi''(x)(g(x, t)^2) + \frac{o(\Delta t)}{\Delta t}]\, p(x, t \,|\, y, s)\, dx\,.$$

ϕ has been chosen to have compact support, so integrating by parts on the RHS (once for the first term and twice for the second) gives

$$\int_{\mathbb{R}} \phi(x) \left[-\frac{\partial}{\partial x}(f(x, t)p(x, t \,|\, y, s)) + \frac{\partial}{\partial x^2}\left(\frac{1}{2}g(x, t)^2\, p(x, t \,|\, y, s)\right) + \frac{o(\Delta t)}{\Delta t}\right] dx\,.$$

Putting it all together

$$\int_{\mathbb{R}} \phi(x) \left[\frac{p(x, t+\Delta t \,|\, y, s) - p(x, t \,|\, y, s)}{\Delta t} + \frac{\partial}{\partial x}(f(x, t)p(x, t \,|\, y, s)) - \frac{\partial}{\partial x^2}\left(\frac{1}{2}g(x, t)^2\, p(x, t \,|\, y, s)\right) \frac{o(\Delta t)}{\Delta t}\right] dx = 0\,.$$

In the limit $\Delta t \to 0$, since ϕ is arbitrary, we get

$$\frac{\partial}{\partial t}p(x,t\,|\,y,s) = -\frac{\partial}{\partial x}(f(x,t)p(x,t\,|\,y,s)) + \frac{\partial}{\partial x^2}\left(\frac{1}{2}g(x,t)^2\,p(x,t\,|\,y,s)\right)$$

which is indeed the 1D F-P equation.

Note that many of the steps in the above derivation are formal – we have frequently exchanged various limiting operations without stating any assumptions, and each of these steps would require more justification.

Stationary distributions from the F-P equation. Let us now consider how we can use the Fokker-Planck equation to find stationary distributions of SDEs in the special case of 1D, and assuming that the coefficients f and g do not depend on time. In this case, the Fokker-Planck equation reads

$$\frac{\partial}{\partial t}p(x,t) = -\frac{\partial}{\partial x}[f(x)p(x,t)] + \frac{\partial^2}{\partial x^2}[D(x)p(x,t)]\,,$$

where $D(x) = \frac{1}{2}g(x)^2$. Any stationary pdf, say $p_s(x)$, must satisfy the stationary equation obtained by setting the time-derivative to be zero, namely

$$0 = -\frac{d}{dx}[f(x)p_s(x)] + \frac{d^2}{dx^2}[D(x)p_s(x)]\,. \tag{26}$$

We can integrate once to get

$$\frac{d}{dx}[D(x)p_s(x)] - f(x)p_s(x) = C_0\,.$$

C_0 can be interpreted as the steady state probability flux, which we assume to be zero. Setting $C_0 = 0$, we can now use the integrating factor $\exp\left[\int_0^x \frac{f(y)}{D(y)}\,dy\right]$, and integrate again to obtain

$$p_s(x) = \frac{C}{D(x)}\exp\left[\int_0^x \frac{f(y)}{D(y)}\,dy\right]\,, \tag{27}$$

Here, the constant of integration, C, can be obtained by observing that $\int_{\mathbb{R}} p_s(x)\,dx = 1$. In this special case, the F-P equation allows us to calculate explicitly the stationary distribution, presuming, of course, that such a distribution exists and is unique.

3.4 The chemical Langevin equation and chemical Fokker-Planck equation

[At this point you may want to look back to Section 2.6 on the reaction counting process.]

We start with (6) for the evolution of a general stochastic CRN, namely

$$\mathbf{X}(t) = \mathbf{X}(0) + \Gamma\mathbf{N}(t)\,.$$

Recall that $N_j(t)$ counts how many times the jth reaction has fired by time t. Recall that over a small time-period Δt we have:

$$\mathbf{X}(t+\Delta t) - \mathbf{X}(t) \simeq \zeta_1 Y_1(\nu_1(\mathbf{X}(t))\,\Delta t) + \cdots + \zeta_m Y_m(\nu_m(\mathbf{X}(t))\,\Delta t)\,,$$

where ζ_j is the reaction vector of the jth reaction, $\nu_j(\cdot)$ is the intensity of the jth reaction, and Y_j are independent unit Poisson processes. In our discussion of τ-leaping, we already noted that it may make sense to hold intensities constant over the time-step Δt.

Although we know that $\mathbf{X}(t)$ is actually an *integer-valued* random variable, the form of the above equation tempts us to consider the increment $\mathbf{X}(t+\Delta t) - \mathbf{X}(t)$ in terms of a derivative. This would amount, effectively, to "interpolating" between integer values of the components of $\mathbf{X}$. To proceed, let us make the following approximate assumptions:

1. Each $\nu_j(\mathbf{X}(t))$ is constant over the time-interval Δt
2. $\nu_j(\mathbf{X}(t))\Delta t$ is large, so the expected number of times that each reaction fires in the time-interval Δt is large. In this situation, we can approximate the Poisson distribution by a normal distribution, i.e., $Y_j(\nu_j(\mathbf{X}(t))\,\Delta t)$ by $\mathcal{N}_j(\nu_j(\mathbf{X}(t))\,\Delta t, \nu_j(\mathbf{X}(t))\,\Delta t)$ where $\mathcal{N}_j$ are (independent) normal random variables with mean and variance both equal to $\nu_j(\mathbf{X}(t))\,\Delta t$ (the justification of this approximation, using the central limit theorem, is left as an exercise).
3. Now let $\mathcal{N}_j(0,1)$ be independent standard normal variables, so that

$$\mathcal{N}_j(\nu_j(\mathbf{X}(t))\,\Delta t, \nu_j(\mathbf{X}(t))\,\Delta t) = \nu_j(\mathbf{X}(t))\,\Delta t + \sqrt{\nu_j(\mathbf{X}(t))\,\Delta t}\;\mathcal{N}_j(0,1)\,.$$

   Replacing each $Y_j(\nu_j(\mathbf{X}(t))\,\Delta t)$ with this normal approximating random variable gives

$$\mathbf{X}(t+\Delta t) - \mathbf{X}(t) \simeq \sum_{j=1}^{m} \zeta_j Y_j(\nu_j(\mathbf{X}(t))\,\Delta t) \simeq \sum_{j=1}^{m} \zeta_j \left[\nu_j(\mathbf{X}(t))\,\Delta t + \sqrt{\nu_j(\mathbf{X}(t))\,\Delta t}\;\mathcal{N}_j(0,1)\right]$$

4. We can regard $\sqrt{\Delta t}\;\mathcal{N}_j(0,1)$ as an increment of Brownian motion, and formally write, in the limit,

$$\mathrm{d}\mathbf{X}(t) := \mathbf{X}(t+\mathrm{d}t) - \mathbf{X}(t) = \sum_{j=1}^{m} \zeta_j \left[\nu_j(\mathbf{X}(t))\,\mathrm{d}t + \sqrt{\nu_j(\mathbf{X}(t))}\;\mathrm{d}W_j\right]. \tag{28}$$

   The stochastic differential equation (28) is called the **Chemical Langevin Equation**. Note that it looks exactly like the deterministic equation for the evolution, with "noise terms" added on. There is one independent noise term associated with each reaction, corresponding to uncertainty in the number of times this reaction will fire in a given time interval.
5. We could rewrite (28) tidily in vector notation as follows

$$\mathrm{d}\mathbf{X}(t) = \Gamma\,\boldsymbol{\nu}(\mathbf{X}(t))\,\mathrm{d}t + \Gamma \mathrm{D}(\sqrt{\boldsymbol{\nu}(\mathbf{X}(t))})\;\mathrm{d}\boldsymbol{W}\,,$$

   where Γ is the stoichiometric matrix, $\boldsymbol{W}$ is a vector of m independent Brownian motions, and $\mathrm{D}(\sqrt{\boldsymbol{\nu}(\mathbf{X}(t))})$ is a diagonal $m \times m$ matrix whose (j,j)th entry is $\sqrt{\nu_j(\mathbf{X}(t))}$. This makes it clear that the chemical Fokker-Planck equation is an SDE with drift and diffusion terms, and both are easily written down for any given CRN.

6. We can now write down the Fokker-Planck equation corresponding to the chemical Langevin equation

$$\frac{\partial}{\partial t}p(x,t) = -\sum_{i=1}^{n}\frac{\partial}{\partial x_i}[f_i(x,t)p(x,t)] + \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial^2}{\partial x_i \partial x_j}[D_{ij}(x,t)p(x,t)]\,,$$

where

- f_i is the ith component of $\Gamma\boldsymbol{\nu}(x(t))$, i.e., $f_i(x,t) = \sum_{j=1}^{m}\Gamma_{ij}\nu_j(x(t))$, and
- $D_{ij}(x,t) = \frac{1}{2}\sum_{k=1}^{m}\Gamma_{ik}\Gamma_{jk}\nu_k(x(t))$.

Exercise 3.3 (**Normal approximation to the Poisson distribution)**. Consider a Poisson random variable X with parameter λ and a normal random variable Y with mean and variance λ. Show, with the help of the central limit theorem, that for large λ

$$\mathbb{P}(X = n) \simeq \mathbb{P}(n - \frac{1}{2} \le Y \le n + \frac{1}{2})\,.$$

[Note that a Poisson random variable with parameter λ has the same distribution as the sum of N independent Poisson random variables, each with parameter λ/N.]

Example 3.4 (Chemical Langevin equation and chemical F-P equation). Let us consider the production-degradation system

$$0 \xrightarrow{k_1} A\,, \quad A \xrightarrow{k_2} 0\,.$$

We have

$$\zeta_1 = 1,\ \zeta_2 = -1,\ \nu_1 = k_1 V,\ \nu_2 = k_2 a\,.$$

where V is the system volume, and we now write a for the number of molecules of A considered as a *real* variable. The corresponding chemical Langevin equation is

$$\begin{aligned} da &= \begin{pmatrix} 1 & -1 \end{pmatrix}\begin{pmatrix} k_1 V \\ k_2 a \end{pmatrix} dt + \begin{pmatrix} 1 & -1 \end{pmatrix}\begin{pmatrix} \sqrt{k_1 V} & 0 \\ 0 & \sqrt{k_2 a} \end{pmatrix}\begin{pmatrix} dW_1 \\ dW_3 \end{pmatrix}. \\ &= (k_1 V - k_2 a)\,dt + \sqrt{k_1 V}\,dW_1 - \sqrt{k_2 a}\,dW_2 \end{aligned}$$

In this case, we have the drift $k_1 V - k_2 a$ and the diffusion $\frac{k_1 V + k_2 a}{2}$, giving the F-P equation

$$\frac{\partial}{\partial t}p(a,t) = -\frac{\partial}{\partial a}\left((k_1 V - k_2 a)p(a,t)\right) + \frac{\partial^2}{\partial a^2}\left(\frac{k_1 V + k_2 a}{2}\,p(a,t)\right).$$

We saw earlier how to find stationary distributions from the F-P equation: in this case we get the stationary pdf:

$$p_s(a) = \frac{C}{D(a)}\exp\left[\int_0^a \frac{f(y)}{D(y)}\,dy\right],$$

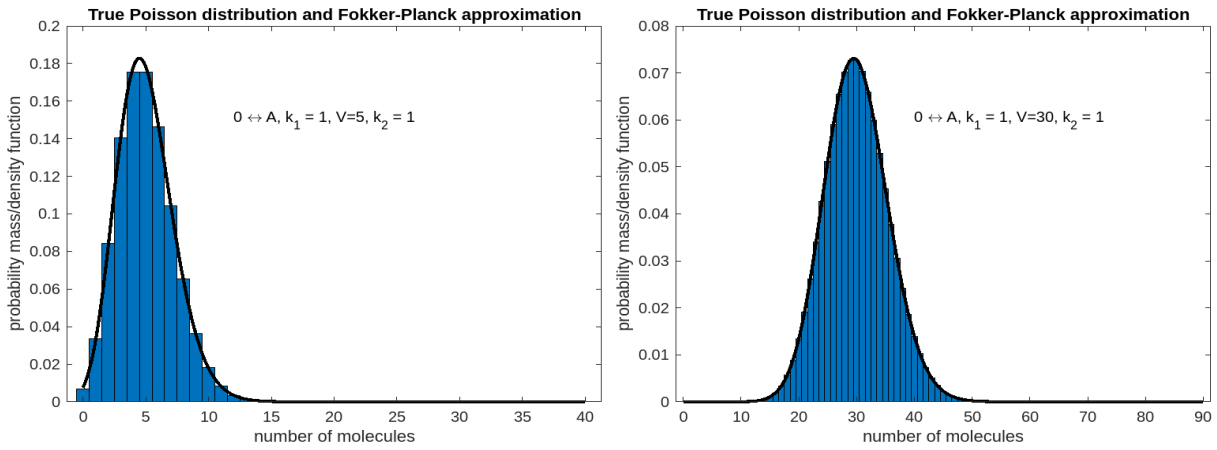

Figure 1: Stationary distribution for the reaction network $0 \xrightarrow{k_1} A$, $A \xrightarrow{k_2} 0$, for two values of the reactor volume V. The exact distribution from the chemical master equation as a histogram, along with the pdf computed numerically using the chemical Fokker-Planck equation as above. Even for a small reactor volume, i.e., for relatively small numbers of molecules, the PDF derived from the chemical Fokker-Planck equation matches the exact distribution well.

where $f(a) = k_1 V - k_2 a$ and $D(a) = \frac{k_1 V + k_2 a}{2}$ and C is the normalisation constant

$$C = \left(\int_{\mathbb{R}} \frac{1}{D(a)} \exp\left[\int_0^a \frac{f(y)}{D(y)} \, dy \right] da \right)^{-1}$$

We can exactly compute $p_s(a)$ upto normalisation, and compute the normalisation constant numerically (you should try this as an exercise). A comparison between the PDF estimated using the chemical Fokker-Planck equation, and the exact PDF for two choices of parameter values is shown in Figure 1. In this case, even when molecule numbers are quite small, the stationary distribution calculated using the chemical Fokker-Planck equation, matches the true distribution quite well.

3.5 The backward Kolmogorov equation (BKE)

The Fokker-Planck equation tells us how the probability density function $p(x, t \,|\, y, s)$ associated with an SDE evolves given some initial state y at some initial time s. Suppose, instead, that we are interested in some given final state(s), and allow the initial state to vary. Such a problem is called a "final value problem" rather than initial value problem. We can write down the corresponding PDE in much the same way as we did the F-P equation – in fact, the derivation is somewhat easier. We only present a sketch of the derivation in 1D, but the principle in higher dimensions is similar. Several steps are formal and strictly require further justification.

As before we start with the Chapman-Kolmogorov equation, this time in the following form

$$p(x, t \,|\, y, s - \Delta s) = \int_{\mathbb{R}} p(x, t \,|\, z, s) \, p(z, s \,|\, y, s - \Delta s) \, dz \,. \tag{29}$$

We are interested in letting $\Delta s \to 0$, in which case, we expect $z \to y$. Now we do not need to integrate against a test function – instead, we can simply Taylor expand $p(x,t\,|\,z,s)$ about the point $z = y$ to get

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

If we now substitute for $p(x,t\,|\,z,s)$ in (29), the first term in the Taylor expansion gives rise to a term of the form

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

where the integral is equal to 1 from the law of total probability. The next term gives

$$\frac{\partial p}{\partial y}(x,t\,|\,y,s)\int_{\mathbb{R}}(z-y)\,p(z,s\,|\,y,s-\Delta s)\,dz = \frac{\partial p}{\partial y}(x,t\,|\,y,s)\,f(y,s)\,\Delta s + o(\Delta s)\,.$$

The equality follows from (23) because the integral can be interpreted as $\mathbb{E}(X(s)-y\,|\,X(s-\Delta s) = y]$. Similarly, the third term gives

$$\frac{1}{2}\frac{\partial^2 p}{\partial y^2}(x,t\,|\,y,s)\int_{\mathbb{R}}(z-y)^2\,p(z,s\,|\,y,s-\Delta s)\,dz = \frac{\partial^2 p}{\partial y^2}(x,t\,|\,y,s)\,\frac{g(y,s)^2}{2}\,\Delta s + o(\Delta s)\,.$$

This follows from (24) and the interpretation of $\int_{\mathbb{R}}(z-y)^2\,p(z,s\,|\,y,s-\Delta s)\,dz$ as $\mathbb{E}((X(s)-y)^2\,|\,X(s-\Delta s)=y]$. Putting this all together, we get

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

Subtracting $p(x,t\,|\,y,s)$ from both sides, dividing through by Δs and letting $\Delta s \to 0$ we get the backward Kolmogorov equation in 1D:

$$-\frac{\partial p}{\partial s}(x,t\,|\,y,s) = f(y,s)\,\frac{\partial p}{\partial y}(x,t\,|\,y,s) + \frac{g(y,s)^2}{2}\,\frac{\partial^2 p}{\partial y^2}(x,t\,|\,y,s)\,. \tag{30}$$

Similar arguments in higher dimensions lead to

$$-\frac{\partial p}{\partial s}(x,t\,|\,y,s) = \sum_{i=1}^{n} f_i(y,s)\frac{\partial p(x,t\,|\,y,s)}{\partial y_i} + \sum_{i=1}^{n}\sum_{j=1}^{n} D_{ij}(y,s)\,\frac{\partial^2 p(x,t\,|\,y,s)}{\partial y_i\,\partial y_j}\,, \tag{31}$$

where, as before, $\mathbf{D} = \frac{1}{2}g\,g^t$.

3.5.1 First passage times

The BKE is particularly useful when we want to know how long we will have to wait for some event to occur. Let us consider an SDE of the form (19), but where we now assume that the coefficients f and g do not explicitly depend on time, i.e.,

$$d\mathbf{X}(t) = f(\mathbf{X}(t))\,dt + g(\mathbf{X}(t))\,dW(t)\,. \tag{32}$$

Given some open subset A of the state-space, and some initial state $y \in A$, we can use the BKE to find out about the first time that trajectories hit the boundary of A. We will refer to this random time as the **first passage time from** A (even though a trajectory hitting the boundary could immediately return to A). For any $y \in A$, we thus define the first passage time from A as

$$T_A(y) := \inf_{t \geq 0}(t : \mathbf{X}(t) \notin A \,|\, X(0) = y)\,.$$

To compute $T_A(y)$, first define the $p(x, t \,|\, y, 0)$ via:

$$p(x, t \,|\, y, 0)\,dx = \mathbb{P}(\mathbf{X}(t) \in [x, x + dx],\ \mathbf{X}(u) \in A \ \text{ for all } \ u \in [0, t) \,|\, \mathbf{X}(0) = y)\,,$$

Thus, roughly, $p(x, t \,|\, y, 0)\,dx$ is the probability that we reach x at time t having started at y at time 0 *and* having remained in A throughout $[0, t)$.

Note that we have the boundary conditions $p(x, t \,|\, y, s) = 0$ for $x \notin A$ or $y \notin A$ (and all s and t), corresponding to the fact that if either initial or final state is outside A, then the probability in question is 0. By examining the derivation of the BKE, we can confirm that p satisfies the BKE (31), which, in this case, reads

$$\frac{\partial p(x, t \,|\, y, 0)}{\partial t} = \sum_{i=1}^{n} f_i(y) \frac{\partial p(x, t \,|\, y, 0)}{\partial y_i} + \sum_{i=1}^{n}\sum_{j=1}^{n} D_{ij}(y)\, \frac{\partial^2 p(x, t \,|\, y, 0)}{\partial y_i\, \partial y_j}\,. \tag{33}$$

(To write the equation in this form, we have noted that $p(x, t \,|\, y, 0) = p(x, 0 \,|\, y, -t)$ as we have assumed that f and g do not explicitly depend on t; and further noted that $-\frac{\partial p(x, 0 \,|\, y, -t)}{\partial t} = \frac{\partial p(x, t \,|\, y, 0)}{\partial t}$.)

Now define

$$h(y, t) := \mathbb{P}(T_A(y) > t) = \int_{x \in A} p(x, t \,|\, y, 0)\,dx\,.$$

The equation simply tells us that if we have not at any point exited A in the interval $[0, t]$, then we have remained in A and reached some point $x \in A$ at time t. Clearly $h(y, 0) = 1$ and we make the assumption that $\lim_{t \to \infty} h(y, t) = 0$, i.e., given any fixed y, with probability 1 we exit A in finite time. (Note that this does *not* imply that the expected first passage time must be finite as we shall see in an example below.)

Integrating (33) w.r.t. x over A we get a PDE satisfied by h, namely,

$$\frac{\partial h(y, t)}{\partial t} = \sum_{i=1}^{n} f_i(y) \frac{\partial h(y, t)}{\partial y_i} + \sum_{i=1}^{n}\sum_{j=1}^{n} D_{ij}(y)\, \frac{\partial^2 h(y, t)}{\partial y_i\, \partial y_j}\,. \tag{34}$$

Consider now the expected (i.e., mean) first passage time $\tau(y)$, given initial state y. We have:

$$\tau(y) := \mathbb{E}(T_A(y)) = \int_0^\infty \mathbb{P}(T_A(y) > t)\,dt = \int_0^\infty h(y, t)\,dt\,.$$

Integrating (34) w.r.t. t over $[0, \infty)$ we get for the evolution of the mean first passage time $\tau(y)$, the PDE

$$-1 = \sum_{i=1}^{n} f_i(y) \frac{\partial \tau(y)}{\partial y_i} + \sum_{i=1}^{n}\sum_{j=1}^{n} D_{ij}(y)\, \frac{\partial^2 \tau(y)}{\partial y_i\, \partial y_j}\,, \tag{35}$$

where on the left we have noted that $h(y,\infty) - h(y,0) = -1$. This is known as the Andronov-Vitt-Pontryagin formula. To solve this equation, we need to impose some boundary conditions. First, we expect that $\tau(y) = 0$ if $y \in \partial A$. If the domain is unbounded, then we may also need to impose further boundary conditions in order to obtain τ.

In 1-D (35) becomes

$$-1 = f(y)\frac{\partial \tau(y)}{\partial y} + D(y)\,\frac{\partial^2 \tau(y)}{\partial y^2}\,, \tag{36}$$

where $D(y) = \frac{g(y)^2}{2}$. Multiplying by the integrating factor $\frac{1}{D(y)} \exp\left[\int_0^y f(s)/D(s)\,ds\right]$ gives

$$\frac{d}{dy}\left[\exp\left[\int_0^y f(s)/D(s)\,ds\right]\frac{d\tau}{dy}\right] = -\frac{1}{D(y)}\exp\left[\int_0^y f(s)/D(s)\,ds\right].$$

Note that the right hand side is, up to a sign and the normalisation constant, the stationary distribution $p_s(y)$ given in (27).

If, for example, we are interested in a region $A := (-\infty, x_u)$ for some x_u, then we have the boundary condition $\tau(x_u) = 0$, and we impose the condition that $\frac{d\tau}{dy}(-\infty) = 0$, i.e., the deterministic dynamics dominates far from x_u and the mean first passage time from A becomes insensitive to the precise value of y. Then we may integrate the last equation over $(-\infty, y)$ and rearrange to get

$$\frac{d\tau}{dy} = -\frac{1}{p_s(y)D(y)}\int_{-\infty}^{y} p_s(x)\,dx\,.$$

Integrating again from y to x_u (and noting that $\tau(x_u) = 0$) gives

$$\tau(y) = \int_y^{x_u} \frac{1}{p_s(z)D(z)}\int_{-\infty}^{z} p_s(x)\,dx\,dz\,. \tag{37}$$

Observe that we do not require the stationary distribution $p_s(x)$ to be normalised in order to use it in this formula.

Example 3.5 (**Mean first passage time of 1D Brownian motion)**. In this case, the SDE in question is

$$X(t+dt) = X(t) + dW\,.$$

The mean first passage time from some region satisfies (36) which, in this case, reads

$$-1 = \frac{1}{2}\,\frac{d^2\tau(y)}{dy^2}\,.$$

Suppose that we are interested in an interval $[a,b]$, and $y := X(0) \in (a,b)$. Then we have the boundary conditions $\tau(a) = \tau(b) = 0$, and we easily find by integration that for $y \in (a,b)$, $\tau(y) = (b-y)(y-a)$. For example, we expect Brownian motion beginning at $y = \frac{1}{2}$ to exit the interval $[0,1]$ in time $\frac{1}{4}$. If we let $b \to \infty$ then notice that $\tau(y) \to \infty$ for any $y \in (0,\infty)$, i.e., even though Brownian motion is recurrent in 1D, the expected first passage time from the positive half-line is infinite. (Can you remember an analogous result for a symmetric random walk in 1D?)

Example 3.6 (**Mean first passage time: example in Section 3.6 of Erban and Chapman)**. We consider the SDE:

$$X(t+dt) = X(t) + f(X(t))\,dt + g(X(t))\,dW,$$

where $f(x) = -k_1x^3 + k_2x^2 - k_3x + k_4$ and $g(X(t)) = k_5$. The conditions which led us to Equation 37 are satisfied in this case, so we get

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

where $D(z) = \frac{k_5^2}{2}$. We can compute the stationary distribution, which is, up to normalisation,

$$p_s(y) = \frac{2}{k_5^2} \exp\left[\frac{2}{k_5^2}\int_0^y (-k_1s^3 + k_2s^2 - k_3s + k_4)\,ds\right] = \frac{2}{k_5^2}\exp\left[\frac{-3k_1y^4 + 4k_2y^3 - 6k_3y^2 + 12k_4y}{6k_5^2}\right]$$

and then use it to compute $\tau(y)$ for any y by evaluating the integral

$$\tau(y) = \int_y^{x_u} \frac{1}{\exp\left[\frac{-3k_1z^4+4k_2z^3-6k_3z^2+12k_4z}{6k_5^2}\right]} \int_{-\infty}^{z} \frac{2}{k_5^2}\exp\left[\frac{-3k_1x^4 + 4k_2x^3 - 6k_3x^2 + 12k_4x}{6k_5^2}\right] dx\,dz.$$

numerically.