Introduction to stochastic differential equations

An alternative approach to studying stochastic systems uses stochastic differential equations. You may have seen one written in the Langevin notation, e.g.:

$$\frac{dx}{dt} = \mu(x,t) + \sigma(x,t)\eta(t)$$

where $\eta(t)$ represents "random fluctuations." For such equations, we must specify characteristics of the deviations η, e.g., the type of distribution they are drawn from, its parameters, and their time correlation:

$$C(\tau) = \langle \eta(0)\eta(\tau)\rangle = \lim_{T\to\infty} \frac{1}{T}\int_0^T \eta(t)\eta(t+\tau)\,dt$$

C should be a decreasing function of τ[1]: a fast fall-off tells us that the fluctuations change quickly. For example, fluctuations in temperature might occur on the timescale of hours (illumination), months (seasons), and eons (ice ages). If the fluctuations are occurring much faster than our timescale of interest, we might make the simplifying assumption that they are not correlated at all in time:

$$\langle \eta(0)\eta(\tau)\rangle = c\delta(\tau)$$

Here δ represents the Kronecker delta and c is the second moment of η, which will depend on the distribution of η[2]. Fluctuations with zero time correlation are often called *white noise*. The term *noise* comes by analogy to measurement noise: a major difference is that noise in measurement does not affect the future evolution of the system. When $\sigma(x,t)$ is a constant, we say that the noise is additive (just "tacked on" to the deterministic solution); otherwise, we say that the noise is multiplicative.

White noise can be approximated in an otherwise deterministic simulation by taking small time steps and drawing a random variable to plug in for η at each time step. We can use this to simulate sample realizations of an SDE, but maybe not in the way you'd expect naïvely. For example, you might suspect that the analog of Euler's method for a stochastic differential equation with (normal) Gaussian white noise is:

$$x(t+\Delta t) = x(t) + [\mu(x,t) + \sigma(x,t)\eta]\,\Delta t, \qquad \eta \sim \mathcal{N}(0,1)$$

This is close, but not quite right. The actual equivalent, given by the Euler-Maruyama method, is:

$$x(t+\Delta t) = x(t) + \mu(x,t)\,\Delta t + \sigma(x,t)\eta\sqrt{\Delta t}, \qquad \eta \sim \mathcal{N}(0,1)$$

Later in this lecture we will discuss why this is so; however, I emphasize that there is a very rich literature on analysis of SDEs and that if this topic interests you, a proper course on stochastic processes or stochastic calculus (e.g. Stats 171) would be an appropriate next direction.

Our prior discussions of stochastic modeling have been motivated by the common-sense notion that molecule numbers in real biological systems are discrete and sometimes too low to legitimize continuous approximations. By contrast this *Langevin approach* restores us to the realm of continuity and imposes a distribution on noise which may be patently unrealizable. Motivating this treatment of noise requires context-dependent explanation, and for both historical and epistemic

[1] Note that time correlations for variables in a system may have very different shapes: consider, for example, the correlation function for an oscillator.

[2] If $\eta \sim \mathcal{N}(0,1)$, $c = 1$.

reasons there is no better place to start than with Brownian motion. What we learn about diffusion will be helpful later in our study of development and Turing patterns. We will begin by studying Brownian motion with more familiar approaches and eventually formulate it in terms of a Wiener process.

Brownian motion

Robert Brown, observing pollen grains (and inorganic material) suspended in water in the mid 1820s, noticed that they appeared to move erratically and without proximate cause. Einstein correctly asserted that individual water molecules imparted momentum on colliding with the pollen granule: many such collisions might sum to give a net momentum in a random direction, in which the pollen granule would briefly travel before additional collisions changed its trajectory once again. The fluctuations in position under *Brownian motion* occur on such short time scales relative to what we are able to observe physically that we can think of them as white noise. By symmetry, the distribution of fluctuations should be even, that is, $f(x) = f(-x)$ and therefore have an average value of zero.

During Einstein's miracle year (1905), he published a paper establishing a formula for the mean square displacement of a particle undergoing Brownian motion via a derivation of Fick's second law and a description of the diffusion coefficient D. He started from nothing but the assumption that the probability distribution for a displacement y in a small time period τ would be given by some $p(y)$, with p even and normalized. His method assumes a large ensemble of n particles: the number changing position by a distance between y and $y + dy$ units in a time τ would therefore be:

$$dn = np(y)dy$$

Denoting the particle concentrations at a time t by $c(x,t)$, we can calculate the concentrations a short while later by:

$$c(x, t+\tau) = \int_{-\infty}^{\infty} c(x-y, t)\, p(y)\, dy \tag{1}$$

Particles can move to point x from any other position (though the probability of moving there from far away is lower): this convolution integral sums all of those contributions to the concentration at point x after a time τ has passed. This general concept should be reminiscent of the master equation and the convolution integrals you have studied in this course. To get somewhere interesting, we perform some Taylor expansions and plug in to Equation 1:

$$\begin{aligned}
c(x,t+\tau) &\approx c(x,t) + \tau\frac{\partial c}{\partial t} \qquad (\tau \text{ small}) \\
c(x-y,t) &= c(x,t) - y\frac{\partial c}{\partial x} + \frac{y^2}{2}\frac{\partial^2 c}{\partial x^2} - \frac{y^3}{3!}\frac{\partial^3 c}{\partial x^3} + \dots \\
c(x,t) + \tau\frac{\partial c}{\partial t} &= \int_{-\infty}^{\infty}\left[c(x,t) - y\frac{\partial c}{\partial x} + \frac{y^2}{2!}\frac{\partial^2 c}{\partial x^2} - \frac{y^3}{3!}\frac{\partial^3 c}{\partial x^3} + \dots\right] p(y)\, dy \\
&= c(x,t)\int_{-\infty}^{\infty} p(y)\, dy - \frac{\partial c}{\partial x}\int_{-\infty}^{\infty} y\, p(y)\, dy + \frac{1}{2!}\frac{\partial^2 c}{\partial x^2}\int_{-\infty}^{\infty} y^2\, p(y)\, dy + \dots
\end{aligned}$$

Since $p(y)$ is normalized, we can cancel $c(x,t)$ from both sides of the equation. Moreover, since $p(y)$ is odd, the value of the integrals will be zero for all odd powers of y, so that we are left with:

$$\tau\frac{\partial c}{\partial t} = \frac{1}{2!}\frac{\partial^2 c}{\partial x^2}\int_{-\infty}^{\infty} y^2\, p(y)\, dy + \frac{1}{4!}\frac{\partial^4 c}{\partial x^4}\int_{-\infty}^{\infty} y^4\, p(y)\, dy + \dots$$

Noting that the terms at right are decreasing rapidly in value, and defining $D = \frac{1}{2\tau}\int_{-\infty}^{\infty} y^2 p(y)\,dy$, we arrive at Fick's second law (so named for the person who discovered it empirically half a century prior), otherwise known as the diffusion equation:

$$\frac{\partial c}{\partial t} = D\frac{\partial^2 c}{\partial x^2}$$

We can now introduce by ansatz[3] a solution to this equation for the initial condition where the concentration is initially one at the origin:

$$c(x,t) = \frac{1}{\sqrt{4\pi Dt}}e^{-x^2/4Dt}$$

Although we did not bother to derive it, let's pause to celebrate the existence of this solution, which is diffusion's equivalent of an impulse response function. (In some disciplines, the term "Green's function" is more common than "impulse response function." The underlying concepts are equivalent but you are more likely to find examples incorporating spatial dimensions if you look up Green's functions.) Since you know how particles that start diffusing from a certain point at time 0 will behave, you can use a convolution integral to describe how *any* initial spatial configuration of particles will evole with time. On problem set eight, you'll use this concept to explore how fluorescence recovery after photobleaching (FRAP) can be used to measure the diffusion constant.

The mean of this even function is clearly zero; the second moment (and thus the variance) can be found using a simple trick based on the normalization

$$\begin{aligned}
-\int_{\infty}^{\infty} e^{-x^2/4Dt}\,dx &= -\sqrt{4\pi Dt} \qquad \text{Letting } y = 1/4Dt\text{:} \\
-\frac{\partial}{\partial y}\int_{\infty}^{\infty} e^{-x^2/4Dt}\,dx = \int_{\infty}^{\infty} x^2 e^{-x^2/4Dt}\,dt &= -\frac{\partial}{\partial y}\frac{\sqrt{\pi}}{\sqrt{y}} = 2Dt\sqrt{4\pi Dt} \\
\left\langle x^2\right\rangle = \int_{\infty}^{\infty} x^2 c(x,t)\,dx &= 2Dt \\
\sigma_x^2 = \left\langle x^2\right\rangle - \left\langle x\right\rangle^2 &= 2Dt
\end{aligned}$$

Through an independent derivation of the diffusion coefficient of a hard sphere of radius r in media with viscosity η, $D = k_B T/6\pi r\eta$, Perrin was able to use this equation to calculate Boltzmann's constant and thus Avogadro's number. Perrin's experiment was done with small uniform spheres in saline solution on a cover slip: he used a camera lucida to record the position of these spheres at time intervals and thus to calculate $\langle x^2\rangle$. (A 2006 reiteration of this experiment is described in Newburgh et al.'s paper on the course website.)

Let's reflect briefly on how we arrived here: Einstein did not make any unjustified assumptions about $p(y)$ – indeed, never calculated it - but still gave us a deterministic formula for the time evolution of the concentration profile. Focusing on the concentration profile was the key: anything unusual that might befall an individual molecule is "averaged out" in the ensemble. Intuition and frequent exposure to the ergodic hypothesis accurately suggest that the probability distribution of a single molecule should reflect the normalized concentration profile, we we shall soon see.

[3] I have posted a derivation for those interested; it is omitted here for employing the residue theorem and some gnarly Gaussian integrals.

Langevin's derivation

Langevin's alternative derivation (1908) of the variance in a particle's position was one of the first attempts to employ stochastic differential equations to understand a natural phenomenon: some concerns with the rigor of his approach drove the development of modern stochastic analysis. He envisioned a single pollen granule buffetted by a "random force" $f_r(t)$ due to impacts with molecules of the fluid and also experiencing the force of viscous (Stokes) drag:

$$F_{\text{net}} = m\frac{d^2x}{dt^2} \quad = \quad -b\frac{dx}{dt} + f_r \tag{2}$$

Like Einstein, Langevin makes no assumptions about $f_r(t)$ other than "that it is indifferently positive and negative and that its magnitude is such that it maintains the agitation of the particle, which the viscous resistance would stop without it." (b is a Stokes drag coefficient with units of kg/s.) In this way he set about proving the single-particle equivalent of Einstein's result concerning the mean square displacement.

Starting from equation 2, multiply both sides by x and rearrange[4] to find:

$$m\left(\frac{d}{dt}\left[x\frac{dx}{dt}\right] - \left[\frac{dx}{dt}\right]^2\right) \quad = \quad -bx\frac{dx}{dt} + f_r x$$

Take the average of both sides, exploiting the commutativity of derivatives and averages:

$$m\frac{d}{dt}\left\langle x\frac{dx}{dt}\right\rangle - \left\langle mv^2\right\rangle \quad = \quad -b\left\langle x\frac{dx}{dt}\right\rangle + \left\langle f_r x\right\rangle$$

Langevin stipulated that $\langle x f_r\rangle$ is zero. The equipartition theorem describes the relationship between the average kinetic energy of a particle and the temperature: $\left\langle \frac{1}{2}mv^2\right\rangle = \frac{1}{2}k_b T$. Inserting theese values, we obtain:

$$\frac{d}{dt}\left\langle x\frac{dx}{dt}\right\rangle + \frac{b}{m}\left\langle x\frac{dx}{dt}\right\rangle \quad = \quad \frac{k_B T}{m}$$

There is nothing left to do but integrate. We use an integrating factor method, multiplying both sides by $e^{bt/m}$ and rearranging:

$$\frac{d}{dt}\left[e^{bt/m}\left\langle x\frac{dx}{dt}\right\rangle\right] \quad = \quad \frac{k_B T e^{bt/m}}{m} \implies \left\langle x\frac{dx}{dt}\right\rangle = \frac{k_B T}{b}\left(1 + e^{-bt/m}\right)$$

assuming the particle is initially at the origin for convenience. Finally, noting that $x\dot{x} = \frac{1}{2}d(x^2)/dt$:

$$\frac{d}{dt}\left\langle x^2\right\rangle = \frac{2k_b T}{b}\left(1 - e^{-bt/m}\right) \implies \left\langle x^2\right\rangle = \frac{2k_b T}{b}\left(t + \frac{m}{b}\left[1 - e^{-bt/m}\right]\right)$$

Compare two limiting cases:

$$\left\langle x^2\right\rangle = \begin{cases} k_B T t^2/b & : t \ll m/b \\ 2k_B T t/b & : t \gg m/b \end{cases}$$

In other words, at short times the particle behaves like it's traveling in a straight line: its rms distance from the origin is increasing linearly with time. At longer timescales the particle is properly

[4] Notice that we are not following Langevin's derivation precisely: Langevin's solution ignores the behavior at short timescales and requires ignoring an exponential term after the first integration. Otherwise all assumptions are Langevin's.

thought of as diffusing, with $D = k_B T/b$[5]. Any assumptions made about the nature of the random force have not deprived us of the expected behavior at long and short timescales. For a sense of proportion, $m/b \approx 10^8$ Hz for a pollen granule, so the crossover time is on the order of nanoseconds.

While we may be satisfied with the sense these results seem to make, and certainly their consistency with Einstein's result, we should pause to reflect on whether Langevin's treatment of $f_r(t)$ was justified. Is $x(t)$ twice differentiable, and if not, how can we rationalize writing a differential equation in terms of d^2x/dt^2? What exactly is f_r and is it fair to assume that it behaves as claimed, e.g. $\langle x f_r \rangle = 0$? (How do you propose to evaluate the integral of $x f_r$ in order to calculate this average?) What does an integral of such a random function look like? These questions and Langevin's cavalier treatment gave rise to the field of stochastic analysis and to such jabs as:

> "A stochastic differential equation is introduced in a rigorous fashion to give a precise meaning to the Langevin equation for the velocity function. This will avoid the usually embarrassing situation in which the Langevin equation, involving the second- derivative of x, is used to find a solution $x(t)$ not having a second-derivative."
> – Doob, 1942.

Wiener processes

The integral of $\eta(t)$ is an unusual construct which we begin by defining as the *Wiener process* W_t:

$$W_t = \int_0^t \eta(s)\, ds, \qquad \eta(s) \sim \mathcal{N}(0,1) \text{ and } \langle \eta(0)\eta(\tau) \rangle = \delta(\tau)$$

Heuristically, a Wiener process W_t may be envisioned as a random walk where the step at time t is η_t, a random variable drawn from $\mathcal{N}(0,1)$. The process is named after Norbert Wiener, who described its properties in the 1920s before going on to found cybernetics, the antecedent of systems biology. Norbert contributed to both World Wars, formalizing the notion of feedback in application to the development of anti-aircraft guns. He turned his attention to the frontier of neuroscience and robotics, recruiting Pitts and McCullogh (whose artifical neuron we have already studied) to MIT where they contributed substantially to the fields of computer science and artificial intelligence. (After artificial intelligence split off as its own field, cybernetics was downgraded to a social science and prompty began to circle the drain.)

The process is continuous everywhere but differentiable nowhere[6]. Two defining properties are that $W(0) = 0$ and that $W(t+\tau) - W(t)$ is normally distributed with mean zero and variance τ[7]. Notice that this probability distribution has exactly the same form as the Green's function found above for a concentration profile, if we chose $D = 1/2$:

$$P(W(t+\tau) - W(\tau) = x \mid t) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t}$$

Now that we know these properties we can revisit the Euler-Maruyama method for simulating systems of the form:

$$\frac{dx}{dt} = \mu(x,t) + \sigma(x,t)\eta(t) \implies dx = \mu(x,t)dt + \sigma(x,t)\, dW_t \qquad \left(\frac{dW_t}{dt} = \eta(t)\right)$$

[5] The latter relation, $D = k_B T/b$, is an instantiation of the fluctuation-dissipation theorem, developed two decades later by our old friend Nyquist.

[6] Continuity can be shown using Kolmogorov's continuity theorem; the latter can be shown from the scaling of variance with time for a Wiener process.

[7] To see why, consider the Wiener process's current displacement from the original position to be the limit of the sum $\sum_{i=1}^{N} \eta_i\, dt$ where $dt = \tau/N$. Then by the Central Limit Theorem, $W(t+\tau) - W(t) \sim \mathcal{N}(0 \cdot N\, dt, 1 \cdot N\, dt) = \mathcal{N}(0,\tau)$.

Since $W(t+\Delta t) - W(t) = \Delta W_t \sim \mathcal{N}(0,\tau) = \sqrt{\tau}\mathcal{N}(0,1)$,

$$x(t+\Delta t) = x(t) + \mu(x,t)\,\Delta t + \sigma(x,t)\,\Delta W_t = x(t) + \mu(x,t)\,\Delta t + \sqrt{\Delta t}\,\sigma(x,t)\,\eta(t),$$

just as we claimed earlier in the lecture.

Stochastic integrals

Now that we have defined the integral of $\eta(t)$ as a Wiener process, consider that for the same system,

$$x(t) = x(0) + \int_0^t \mu(x,\tau)\,d\tau + \int_0^t \sigma(x,\tau)\,dW_\tau$$

If we could take these integrals we'd find a closed form for $x(t)$ in terms of W_t, which would be very useful. Consider two different ways of approximating the second integral by sums of N terms with $t_j = j \times \frac{t}{N}$:

$$\int_0^t \sigma(x,\tau)\,dW_\tau \approx \sum_{k=0}^{N} \sigma(x,t_j)\left[W_{t_{j+1}} - W_{t_j}\right] \text{ or } \sum_{k=0}^{N}\left[\frac{\sigma(x,t_{j+1}) + \sigma(x,t_j)}{2}\right]\left[W_{t_{j+1}} - W_{t_j}\right]$$

The only difference is in whether we choose to evaluate σ at the beginning or in the middle of the time step – a distinction that would not make a lick of difference as $N \to \infty$ with your standard-issue Riemann integral. However, the sum is different for stochastic integrals (try simulating it!). The sum at left is the Itō formulation of a stochastic integral; the one at right is the Stratonovich formulation. As to which is more appropriate, van Kampen notes that for radioactive decay[8]:

$$dx = -kx\,dt + \sqrt{x}\,dW_t$$

and other reactions, it makes sense to evaluate σ at the beginning of the interval, and in fact we get an incorrect mean if we use the Stratonovich formulation. In principle the Itō formulation is appropriate when the time correlation in noise terms is precisely zero.

Why does it matter which formulation is chosen? As the inequality of the sums above suggests, your favorite rules from calculus do not always apply, though they have counterparts. For example, if $Y(W_t) = (W_t)^2$, then is $dY = 2W_t\,dW_t$? In the Stratonovich formulation, yes; in the Itō formulation, $dY = 2W_t\,dW_t + dt$. This implies another mind-bender in Itō calculus:

$$W_t^2 = Y = Y(0) + 2\int_0^t W_t\,dW_t + \int_0^t ds \implies \int_0^t W_t\,dW_t = \frac{1}{2}W_t^2 - \frac{t}{2}$$

In the words of Ramon van Handel:

> "The reader is probably left wondering at this point whether we did not get a little carried away. We started from the intuitive idea of an ordinary differential equation driven by noise. We then concluded that we can not make sense of this as a true differential equation, but only as an integral equation. Next, we concluded that we didn't really know what this integral is supposed to be, so we proceeded to make one up. Now we have finally reduced the notion of a stochastic differential equation to a mathematically meaningful form, but it is unclear that the objects we have introduced bear any resemblance to the intuitive picture of a noisy differential equation."
> – Caltech ACM 217 lecture notes (2007)

[8] Note that we have chosen $\sigma(x) = \sqrt{x}$ since we know this is a Poisson process and thus the variance in decays within a time interval is equal to the mean number of decays in that time interval.

If you begin to question whether we are on to something worthwhile, consider an application of the method to a stock's price, x:

$$dx = \mu x\, dt + \sigma x\, dW_t = x\left[\mu\, dt + \sigma\, dW_t\right] \qquad \mu, \sigma \in \mathbb{R}$$

The intuition captured here is that stock prices tend to increase exponentially, but with a variable rate[9]. (The formula is only realistic when the stock price stays positive, of course.) The solution to this equation is:

$$x(t) = x(0)e^{\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W_t}$$

from which we might find the probability that $x(t)$ is less than some specified value. We can use Itō calculus to e.g. construct a "riskless portfolio" and many other lucrative-sounding things. Of course, analytical solutions to stohastic differential equations are useful in biology as well.

On the other hand, much can be learned about the behavior of a stochastic differential equation without having to study stochastic calculus, and this will be our emphasis at the beginning of the next lecture. You have seen, when we studied the discrete case, that the chemical master equation allows us to predict the rate of change in the probability distribution of states: we will present the Fokker-Planck equation, which does the same for SDEs. We will also introduce Wright's formula for calculating stationary solutions to the Fokker-Planck equation. Both methods can be wielded using ordinary calculus.

[9] For Poisson processes we expect mean and variance to be equal, and thus often choose $\sigma(x,t) \sim \sqrt{x}$. Here we have $\sigma(x) \sim x$ implying that standard deviation and mean are equal: this is a property of the exponential distribution.