Basics of Stochastic Calculus

The goal of these notes is to introduce some basic notions of stochastic calculus. We will see that it is necessary to be very careful when we define the meaning of the noise term in a stochastic differential equation (SDE), particularly when this noise term is has a variance that depends on the dynamical variable. We will make use of numerical simulations to supplement our intuition.

1 An Example

Physicists are often rather cavalier about apparent mathematical niceties, with many results being independent of various details and careful definitions. However this is not always the case. Let us say that we want to model a population growing in a time-varying environment. One approach might be to write down the equation

$$\frac{dn}{dt} = (r + \xi(t))n \tag{1}$$

where $\mathrm{E}[\xi(t)\xi(t')] = 2D\delta(t - t')$. Here the term $\xi(t)$ is a gaussian random variable that depends on time, saying that the growth rate of the population varies from one instant to the next. In fact taking the correlation function of $\xi(t)$ with itself to be a δ function means that no matter how closely in time we measure two values of $\xi(t)$, the measurements we take are independent. This actually may seem a bit pathological — and indeed it is the root of the difficulties that we will address in these notes, but it is also in some ways a simplifying assumption that makes possible much of the analysis of stochastic differential equations.

As a starting point, let us ask what $\xi(t)$ actually *means*. One way to force ourselves to understand what it means is to implement it numerically. In order to numerically solve this equation, it is of course necessary to discretize in time. How might we do this?

1.1 An update rule

The first way we might try to numerically solve our equation is to use the update rule

$$n(t + \Delta t) = n(t) + \left[r\Delta t + \sqrt{2D\Delta t}Z\right] n(t), \tag{2}$$

where Z is a standard normal gaussian random variable. Note the scaling of the variance of Z with Δt: this comes from the fact that over an interval of length Δt, the integral of $\xi(t)$ is a gaussian with variance

$$\mathrm{E}\left[\left(\int_0^{\Delta t} \xi(t)dt\right)^2\right] = \mathrm{E}\left[\int_0^{\Delta t} dtdt'\xi(t)\xi(t')\right] = \int_0^{\Delta t} dtdt' 2D\delta(t - t') = 2D\Delta t. \tag{3}$$

This is a perfectly well defined update rule, but it is not the only choice available to us. The same is true when we integrate deterministic equations: there are many numerical integrations schemes: however they all converge to the same answer then the time step is taken to be infinitesimally small. In this case, we will see that different choices can converge to different solutions even when the time step become infinitesimal.

1.2 A general parameterization

Another update rule we might propose is

$$n(t + \Delta t) = n(t) + \left[r\Delta t + \sqrt{2D\Delta t}Z_1\right] n(t + \alpha\Delta t), \tag{4}$$

where $n(t + \alpha\Delta t)$ is computed as $n(t + \alpha\Delta t) = n(t) + \left[r\alpha\Delta t + \sqrt{2D\alpha\Delta t}Z_2\right] n(t)$. Note that $\alpha \in [0, 1]$, since the interpretation here is that we take a fraction α of a time step with the stochasticity over this step determined by the value of $n(t)$ at the beginning of the time step. Then we use the value of $n(t + \alpha\Delta t)$ after this fractional time step to calculate the variance of the noise term applied over the entire time step Δt. Note that $\alpha = 0$ recovers Equation 2; therefore we can adopt this general parameterization to describe the range of integration methods indexed by α.

What is the joint distribution of Z_1 and Z_2 as defined in Equation 4 and the line below? They are both gaussian with unit variance, but they are crucially not independent! This is because Z_1 describes the random forcing integrated over the entire interval $[t, t + \Delta t]$ and Z_2 described the random forcing integrated over just the first increment $[t, t + \alpha\Delta t]$. These two things are clearly correlated. What is their correlation? Note that $Z_1 = \sqrt{\alpha}Z_2 + \sqrt{1 - \alpha}Z'$ where Z_2 and Z'

are independent: this comes from the fact that the first and second parts of the time step have uncorrelated noises, and are of relative lengths α and $1-\alpha$. So therefore we know that $\mathrm{E}[Z_1 Z_2] = \sqrt{\alpha}$.

What is the difference between the formulations of Equation 4, as we vary α (with $\alpha = 0$ recovering Equation 2)? Let us directly numerically integrate Equation 1 using a few different values of α, and see if there is any difference. To check that any differences we see are due to the integration scheme and not due to the realization of the noise drawn from our random number generator, we can do all the integrations for the same realization of the noise. We will also set $r = 0$ and $D = 1$ for simplicity.

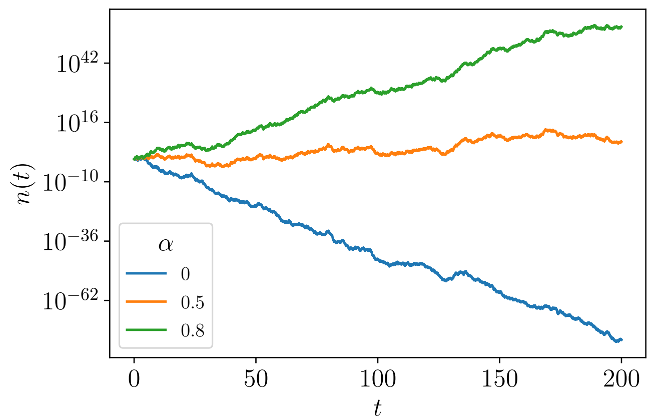

Figure 1: A comparison of the proposed integration schemes for $\alpha = 0, 0.5, 0.8$ for *the same realization of the noise*. We see that they produce vastly different results. In all cases $r = 0$ and $D = 1$.

Figure 1 shows that there is a large different between the behavior for different choices of α. Note that the plot is on a logarithmic scale, and the trajectories differ by many orders of magnitude. Furthermore, this behavior does not go away if we decreases Δt: it is baked into our choice of α. What is going on here? We can try to see how our numerical scheme depends on α by calculating the conditional expectation $\mathrm{E}[n(t+\Delta t)|n(t)]$ for general α.

Using Equation 4, we find that

$$\mathrm{E}\left[n(t+\Delta t)|n(t)\right] = n(t)\left(r\Delta t + 2D\alpha\Delta t + \mathcal{O}(\Delta t^2)\right). \tag{5}$$

From this it is clear that changing α affects our solution *in a deterministic way*. Therefore Equation 1 on its own is ill-defined. It is necessary to specify a value of α along with the equation. Two standard choices for α are included in those that we used in Figure 1: $\alpha = 0$ and $\alpha = 1/2$. These are known respectively as the *Ito* and *Stratonovich* prescriptions for stochastic calculus. They each have their own conveniences and inconveniences, which we will explore in the next section. However (and this is crucial) they are equivalent in the sense that it is always possible to transform from one picture to another.

There is another puzzle arising from Figure 1. Namely, for $\alpha = 0$ we have just shown that the conditional mean of $n(t)$ does not change. Therefore the expected value of n should remain at 1 since we started at $n = 1$. However our simulation of $n(t)$ is steadily decaying, and, judging by eye without doing an ensemble average over many trajectories, it looks unlikely that $n(t)$ will have mean 1 for late t. We will come back to this puzzle later.

1.3 Transforming our equation

We have seen that the interpretation of an SDE is complicated by the presence of a time-dependent variance multiplying the stochasticity. Our lives would be easier if we could transform an SDE of interest into a simpler equation in which the noise term has a prefactor that does not depend on time. This can be done by constructing an SDE for a *function* of our original dynamical variable: we will explore this in the next section. We already know how this can be done for an ordinary differential equation, such as

$$\frac{dx}{dt} = r(t)\sigma(x). \tag{6}$$

If we want the "forcing term" $r(t)$ not to be multiplied by a function of x, we can define a variable f such that

$$\frac{df}{dt} = \frac{1}{\sigma(x)}\frac{dx}{dt} \implies \frac{dx}{df} = \frac{1}{\sigma(x)} \implies f = \int \frac{dx}{\sigma(x)}. \tag{7}$$

Then our original differential equation will take the form

$$\frac{df}{dt} = r(t). \tag{8}$$

Now we want to do the same thing where $r(t)$ is a stochastic function of time with δ-function autocorrelation. This involves a few subtleties that we discuss below.

2 Converting between different noise prescriptions

We will now switch to using the mathematician's notation for SDEs. Instead of the physicist's $\xi(t)$ representing δ-correlated random forcing, mathematicians tend to define the *Wiener process* as continuous-time Brownian motion, so that dW is the random force, equivalent to $\xi(t)$, that gives rise to this Brownian motion when integrated. Mathematicians also tend to write SDEs as expressions for dX where X is a dynamical variable, instead of expressions for $\frac{dX}{dt}$ as we were doing before. This can be useful because we can think of the SDE evolution as being the limit as dt gets very small. In fact, as we have seen, thinking of SDEs as an infinitesimal limit of stochastic difference equations is actually important, and allows us to see why the choice of α is important for our solution.

Suppose we have a stochastic process

$$dX = \mu(X)dt + \sigma(X)dW \tag{9}$$

where W is a Wiener process and we interpret the noise for as-yet undetermined α, which will come into our subsequent expressions. In fact, to make this interpretation explicit, we can express our SDE as

$$dX = \mu(X)dt + \sigma\left(X + \alpha\mu(X)dt + \sigma(X)dW^{(1)}\right)\left(dW^{(1)} + dW^{(2)}\right), \tag{10}$$

where $dW^{(1)}$ is the noise over the first fraction α of the timestep and $dW^{(2)}$ is the noise over the second fraction $1-\alpha$ of the timestep. Crucially, now we can interpret both the stochastic terms in the Ito sense, with the noise acting at the beginning of the timestep. Expanding $\sigma(X + \alpha\mu(X)dt + \sigma(X)dW^{(1)})$ and keeping terms up to order dt, as well as using the fact that $\left(dW^{(1)}\right)^2 \approx \alpha dt$, we obtain

$$dX \approx [\mu(X) + \alpha\sigma'(X)\sigma(X)]dt + \sigma(X)dW. \tag{11}$$

Then if we want to find how a function $f(X)$ changes over a timestep dt, we can write

$$df = dXf'(X) + \frac{(dX)^2}{2}f''(X) + \cdots \tag{12}$$

Substituting in the expression that we know is obeyed by dX, we get

$$df = dXf'(X) + \frac{1}{2}\left([\mu(X) + \alpha\sigma'(X)\sigma(X)]dt + \sigma(X)dW\right)^2 f''(X). \tag{13}$$

Now, keeping only terms of order dt, and using the fact that $(dW)^2 \approx dt$, we are left with

$$df = dXf'(X) + \frac{1}{2}\sigma^2(X)f''(X)dt. \tag{14}$$

Plugging in our known expression for dX (Equation 11), we obtain

$$df = f'(X)\left([\mu(X) + \alpha\sigma'(X)\sigma(X)]dt + \sigma(X)dW\right) + \frac{1}{2}\sigma^2(X)f''(X)dt. \tag{15}$$

Recall from our original equation (Equation 9), that what it would mean for $f(X)$ to transform simply under the chain rule would be that

$$df \stackrel{?}{=} f'(X)[\mu(X)dt + \sigma(X)dW]. \tag{16}$$

Therefore, using Equation 15, we can see that our SDE transforms simply under the chain rule, if

$$2\alpha\sigma'(X)\sigma(X)f'(X) + \sigma^2(X)f''(X) = 0 \implies 2\alpha\sigma'(X)f'(X) + \sigma(X)f''(X) = 0. \tag{17}$$

One way to satisfy this equality is to set $\alpha = 1/2$ and have $\sigma(X)f'(X)$ independent of X (since for $\alpha = 1/2$ the expression in Equation 17 is an X derivative of $\sigma(X)f'(X)$). This is precisely condition needed for the dW in our transformed equation to have a constant prefactor (Equation 7). Therefore we see that if we make a transformation from X to $f(X)$ with $f(X)$ chosen in order to make the noise term have an X-independent prefactor, then our differential equation transforms simply under the chain rule *when* $\alpha = 1/2$. This is a very nice property of the Stratonovich calculus, that arises because we pick the *midpoint* of the interval to evaluate $\sigma(X)$. This would not in general be true if we evaluated $\sigma(X)$ at arbitrary α. In particular we see that for the Ito prescription, the standard chain rule does not hold, and we get an extra contribution to dX of $\frac{1}{2}\sigma^2(X)f''(X)dt$.

3 Back to our example

Does this explain why the numerical integration of the equation we wrote down by the Ito method causes n to decay exponentially in Figure 1, even though we *calculated* that this integration scheme with $\alpha = 0$ preserves the expectation of $n(t)$? It is difficult to understand what the equation is saying when there is n dependence multiplying the noise term $\xi(t)$, but we can get rid of this noise dependence by making the transformation via new variable $u = \log n$.

Let us think about exactly what the computer is doing when we integrate our equation by the Ito method. Each timestep, we multiply n by a gaussian random variable with mean $1 + r\Delta t$ and variance $2D\Delta t$. This means that

$$\log n(t + \Delta t) = \log n(t) + \log(1 + r\Delta t + \sqrt{2D\Delta t} Z), \tag{18}$$

where Z is a standard normal gaussian variable. It is tempting to expand out the logarithm to calculate the expectation of $\log n(t + \Delta t)$. But if we do so, we need to keep *all terms up to order* Δt. This is means that we need to take a second order expansion! Doing this out gives

$$\log n(t + \Delta t) \approx \log n(t) + r\Delta t + \sqrt{2D\Delta t} Z - \frac{1}{2}\left(r\Delta t + \sqrt{2D\Delta t} Z\right)^2 \approx \log n(t) + (r - D)\Delta t + \sqrt{2D\Delta t} Z. \tag{19}$$

Therefore we see that the *logarithm* of n has a drift term in Equation 1, even though $\mathrm{E}[n(t)] = n(0)$. Although in Figure 1 it does not look as though $\mathrm{E}[n(t)]$ for large t is in fact 1, it woul dbe iv we averaged over enough trajectories, due to rare trajectories that would dominate the average by going above $n = 1$. A slight variation of the above calculation shows that for arbitrary α, we have

$$\log n(t + \Delta t) \approx \log n(t) = (r - D + 2\alpha)\Delta t + \sqrt{2D\Delta Z}. \tag{20}$$

This is an explicit example of the basic idea that we worked out above. It we take $\alpha = 1/2$ and $D = 1$ then the drift terms from the Stratonovich prescription (which contributes positively) and the transformation to log variables (which contributes negatively) exactly cancel out, and $u = \log n$ does an unbiased random walk as seen in Figure 1. However, in this case $\mathrm{E}[n(t)]$ is exponentially increasing in time, as we showed previously from our conditional expectation calculation.

In Figure 2 we show the predicted dynamics of $\mathrm{E}[\log n(t)]$ together with our simulations, to check that we understand what is going on.

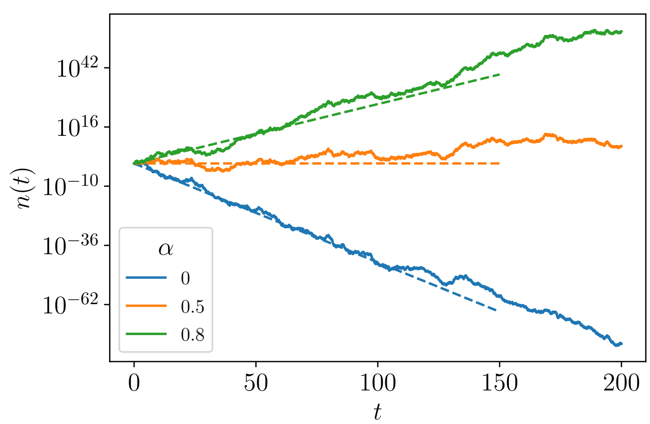

Figure 2: A comparison of simulations with Equation 20 for the mean $n(t)$ over time. In all cases $r = 0$ and $D = 1$.

4 The Ito Integral

We are now poised to make sense of expressions of the type

$$\int f(W(t'))dW(t'), \tag{21}$$

interpreted either in the Ito or Stratonovich sense. Since we have shown that the two are equivalent up to a deterministic drift term, we can work in the Ito prescription, in which integrals of the type in Equation 21 are called *Ito integrals*. However it is a good exercise to work out how things would be different if we carried through the following steps with arbitrary α.

As an example, let us see how we might evaluate

$$\int_0^t W(t')dW(t') \tag{22}$$

for some Wiener process W. The Ito prescription tells us that the interval dW is applied without any knowledge of the function it is multiplying over the corresponding time interval. So if we think about evaluating such an integral in discrete time from 0 to t, to increment our integral from $\int_0^t W(t')dt'$ to $\int_0^{t+dt} W(t')dt'$, we would multiply $W(t)$ by $W(t+dt)-W(t)$, which is independent of $W(t)$, and add this to our running sum. Note that the independence of dW from W, which comes from the Ito prescription, gives us the nice property that the expectation of the value of the integral (which is a random variable) is always 0.

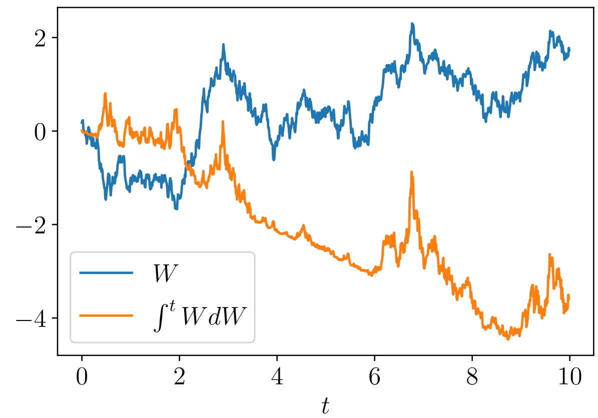

Figure 3: An Ito integral of a Wiener process integrated against itself.

We can carry out this procedure as in Figure 3. How can we interpret the result? In particular, can we analytically calculate this integral? One way to make progress, as before, is to approach the problem from the discrete perspective. Imagine that we have

$$W_n = \sum_{i=1}^{n} X_i \tag{23}$$

where the X_i are i.i.d. gaussian with variance dt. In this case, one can see that our prescription for the integral that we wish to evaluate is

$$\int_0^t W dW \approx \sum_{i=1}^{t/dt} W_i X_{i+1} = X_1X_2 + (X_1 + X_2)X_3 + (X_1 + X_2 + X_3)X_4 + \cdots \tag{24}$$

By staring at this a bit one should realize that this is that same as

$$\int_0^t W dW \approx \frac{1}{2}\left(W_{t/dt}^2 - \sum_{i=1}^{t/dt} X_i^2\right). \tag{25}$$

Finally, making use of $X_i^2 \approx dt$, we have

$$\int_0^t W dW \approx \frac{1}{2}(W^2 - t), \tag{26}$$

which is in fact the correct answer (and explains the linear downward slope in Figure 3 when W is small)! It is not as clear how to make progress for a general integral of form $\int^t f(W)dW$, but it is likely that discretization would help. I do not know if there is a general way to do integrals of this kind.

5 Noise with nonzero correlation time

The above discussion has totally avoided the crucial point that in the real world, there is no such thing as δ-correlated noise dW. Variables change continuously, and so the idealization of discontinuous random forcing with which we have been working is just that: an idealization. Therefore you might be concerned about how seemingly finicky this idealization is. In particular how do we know what α to choose when modeling some real world situation? Well, from what we have shown above, we can say confidently that *it does not matter*. Although the dynamics of our model change depending on how we choose α, this just changes the interpretation of the parameters in the equations that we work with.

In fact, if we imagine an SDE where the random forcing is no longer δ-correlated but instead has some nonzero correlation time $\tau > 0$, the distinction between different choices of α for our integration scheme breaks down. No matter what α we choose, all integration schemes converge to the same solution as $dt \to 0$. This is because in this case the random forcing is a continuous function of time, and so its values at two moments in time (e.g. the beginning and middle of a timestep) become the same as the two moments approach each other.

6 References

These slides were quite helpful.
This blog post has an excellent discussion of many of the similar issues.