Week 9
Stochastic differential equations

Jonathan Goodman

November 19, 2012

1 Introduction to the material for the week

The material this week is all about the expression

$$dX_t = a_t dt + b_t dW_t \ . \tag{1}$$

There are two distinct ways to interpret this, which we will call *strong* and *weak*. The strong interpretation is more literally true, that X is a function of W, that X, a and b are adapted, and (1) is true in the sense of the Ito calculus. For example, we saw that if $X_t = e^{\sigma W_t}$, then

$$dX_t = \sigma e^{\sigma W_t} dW_t + \frac{1}{2}\sigma^2 e^{\sigma W_t} dt$$

$$= \frac{1}{2}\sigma^2 X_t dt + \sigma X_t dW_t \ .$$

In this case, (1) is satisfied with $a_t = \frac{1}{2}\sigma^2 X_t$ and $b_t \sigma X_t$. Roughly speaking, this is the sense we usually assume when doing analysis.

The weak interpretation is not literal. It does not require that we have a Brownian motion path W in mind. In the weak sense, (1) means that at time t you know X_t, a_t and b_t, and

$$\mathrm{E}[\, dX_t \mid \mathcal{F}_t] = a_t dt \ , \tag{2}$$

and

$$\mathrm{E}\left[\, (dX_t)^2 \mid \mathcal{F}_t\right] = b_t^2 dt \ . \tag{3}$$

In this view, (1) just says that for $dt > 0$, the corresponding dX is the sum of a deterministic and a random piece, $a_t dt$ and $b_t dW_t$. "Deterministic" means "known at time t". The example with $a_t = \frac{1}{2}\sigma^2 X_t^2$ shows that a_t need not be known at times earlier than t. The random piece models the part of dX_t that cannot be predicted at time t. We assume that this noise component has mean zero, because if the mean were not zero, we would put the mean into a_t instead. The modeling assumption is that the noise *innovation*, dW_t, not only has mean

zero, but is independent of anything known in $\mathcal{F}_t$. The strength of the noise at time t is b_t. This is assumed known at time t. An unknown part b_t would be part of W_t instead.

The point of all this philosophy is that we can create models of stochastic processes by writing expressions for a_t and b_t in (1). If we think we know the (conditional at time t) mean and variance, we use (1) to create a model process X_t. The Black and Scholes model of the evolution of a stock price is a great example of this kind of reasoning. Suppose S_t is the price of a stock at time t. Then dS_t should be proportional to S_t so that dS is measured as percentage of S_t (OK, a little tautology there). If S_t were replaced by, say, $2S_t$, then dS_t would be replaced by $2dS_t$. If there were a "doublestock" that consisted of two shares of stock, the price change of a doubleshare would be twice the change of a single share. Therefore, we think both the deterministic and random parts of dS_t should be proportional to S_t. The constants are traditionally called μ and σ, the expected rate of return and the volatility respectively:

$$dS_t = \mu S_t dt + \sigma S_t dW_t \ . \tag{4}$$

An equation of the form

$$dX_t = a(X_t, t)dt + b(X_t, t)dW_t \tag{5}$$

is a *stochastic differential equation*, or *SDE*. The difference between an SDE and a general process that satisfies (1) is that here a_t and b_t are required to be known deterministic functions. For example, the function $a(x, t)$ is a function of two variables that is completely known at time 0. For this reason, solution to an SDE is a Markov process. The probability distribution of the path starting at time t, which is $X_{[t,T]}$, is completely determined by X_t. If a and b are independent of t, then the SDE (5) and the corresponding Markov process are *homogeneous*. Otherwise they are *heterogeneous*. If b is independent of X we have *additive noise*. Otherwise, the noise is *multiplicative*. There are many problems with additive noise. These are simpler from both theoretical and practical points of view.

Most SDE's do not have closed form solutions. But solutions may be computed numerically. The *Euler* method, also called *EulerMaruyama*, is a way to create approximate sample paths. It is possible to get information about solutions by solving the forward or backward Kolmogorov equation. These, also, would generally be solved numerically. But that is impractical for SDE systems with more than a few components.

2   Geometric Brownian motion

Solutions to the SDE (4) are called *geometric* Brownian motion, or *GBM*. There are several ways to find the solution. One is to try an ansatz of the form $S_t = A_t e^{\sigma W_t}$. Here, A_t is a deterministic function of t. We do an Ito calculation

with $f(w, t) = A_t e^{\sigma w}$, so that $\partial_t f = \dot{A} e^{\sigma w}$, $\partial_w f = \sigma f$, and $\partial_w^2 f = \sigma^2 f$. The result is

$$d\left(A_t e^{\sigma W_t}\right) = \dot{A}_t e^{\sigma W_t} dt + \sigma A_t e^{\sigma W_t} dW_t + \frac{1}{2}\sigma^2 A_t e^{\sigma W_t} dt$$

$$= \left(\frac{\dot{A}_t}{A_t} + \frac{1}{2}\sigma^2\right) S_t dt + \sigma S_t dW_t .$$

This satisfies (4) if

$$\frac{\dot{A}_t}{A_t} + \frac{1}{2}\sigma^2 = \mu .$$

That implies that

$$\dot{A}_t = \left(\mu - \frac{1}{2}\sigma^2\right) A_t .$$

The solution is a simple exponential:

$$A_t = A_0 e^{\left(\mu - \frac{1}{2}\sigma^2\right)t} .$$

If we set $t = 0$ and use a standard Brownian motion with $W_0 = 0$, we find $A_0 = s_0$. The full solution is

$$S_t = s_0 e^{\sigma W_t + \left(\mu - \frac{1}{2}\sigma^2\right)t} . \tag{6}$$

Here is a related way to find the solution formula (6). Define

$$X_t = \log(S_t) \quad , \qquad S_t = e^{X_t} . \tag{7}$$

We use Ito's lemma for the process S_t. If $f(s) = \log(s)$, then $\partial_s f = \frac{1}{s}$ and $\partial_s^2 f = \partial_s \frac{1}{s} = -\frac{1}{s^2}$. Then

$$dX_t = df(S_t)$$

$$= \partial_s f(S_t) dS_t + \frac{1}{2}\partial_s^2 f(S_t)(dS_t)^2$$

$$= \frac{1}{S_t} dS_t - \frac{1}{2}\frac{1}{S_t^2}(dS_t)^2$$

$$= \frac{1}{S_t}\left(\mu S_t dt + \sigma S_t dW_t\right) - \frac{1}{2}\frac{1}{S_t^2}\sigma^2 S_t^2 dt$$

$$dX_t = \left(\mu - \frac{1}{2}\sigma^2\right) dt + \sigma dW_t .$$

The solution of this is ordinary *arithmetic* Brownian motion (there are geometric series and arithmetic series).

$$X_t = x_0 + \left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W_t .$$

3

An arithmetic Brownian motion has constant drift and Brownian motion parts. This one has drift $\mu - \frac{1}{2}\sigma^2$ and noise coefficient σ. A *standard* Brownian motion has zero drift and unit noise coefficient. The solution from this is

$$S_t = e^{X_t} = e^{x_0} e^{\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W_t} \ .$$

This is the same as before, with $s_0 = e^{x_0}$.

There is another way to arrive at the log variable transformation. Suppose $V(S_T)$ is a payout and we consider the value function

$$f(s,t) = \mathrm{E}_{s,t}[\,V(S_t)\,] \ .$$

This satisfies the backward equation

$$\partial_t f + \frac{\sigma^2 s^2}{2}\partial_s^2 f + \mu s \partial_s f = 0 \ . \tag{8}$$

This is because $a(s,t) = \mu s$, and $b(s,t) = \sigma s$. The Black Scholes PDE is similar to this. This PDE has *coefficients* $\frac{\sigma^2 s^2}{2}$ and μs, which are functions of the independent variable s. It is a *linear* PDE with *variable coefficients*. We can simplify the PDE by a change of variable to make it into a *constant coefficient* PDE. This change of variable is

$$x = \log(s) \quad , \qquad s = e^x \ .$$

There is an obvious sense in which this is the same as (7). But there is a sense in which it is different. Here, the substitution is about simple variables s and x, not stochastic processes S_t and X_t.

We rewrite (8) in the x variable. The chain rule from calculus gives

$$\partial_s f = \frac{\partial f}{\partial s} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial s} = \frac{1}{s}\partial_x f \ .$$

The next derivative is

$$\begin{aligned}
\partial_s^2 f &= \partial_s\left(\partial_s f\right) \\
&= \partial_s\left(\frac{1}{s}\partial_x f\right) \\
&= \left(\partial_s \frac{1}{s}\right)\partial_x f + \frac{1}{s}\partial_s\left(\partial_x f\right) \\
&= -\frac{1}{s^2}\partial_x f + \frac{1}{s^2}\partial_x^2 f
\end{aligned}$$

These derivatives go back into (8) to give

$$\partial_t f + \frac{\sigma^2}{2}\partial_x^2 f + \left(\mu - \frac{\sigma^2}{2}\right)\partial_x f = 0 \ . \tag{9}$$

This PDE is the backward equation for the SDE $dX_t = \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma dW_t$. We can transform (9) into the standard heat equation with some simple normalizations. The first is to get rid of the drift term, the term involving $\partial_x f$, using a coordinate that moves with the drift:

$$x = y + \left(\mu - \frac{\sigma^2}{2}\right) t \ , \qquad y = x - \left(\mu - \frac{\sigma^2}{2}\right) t \ .$$

It can be tricky to calculate what happens to the equation (9) in the new coordinates. Since $\partial_y x = \partial_x y = 1$, calculating *space* derivatives (x or y) does not change the equation. The change has to do with the time derivative, which may not seem to have changed. One way to do it is to recall the definition of *partial* derivative. If the variables are x and t, the partial derivative of f with respect to t with x fixed is

$$\partial_t f \Big|_x = \lim_{\Delta t \to 0} \frac{f(t + \Delta t) - f(t)}{\Delta t} \Bigg|_x$$

This is the rate of change of the quantity f when t is changed and x is held fixed. The definition of ∂_t changes if we use the y variable in place of x. It becomes the rate of change of f when t changes and y is held fixed. Suppose we have changes Δt, Δx and Δy. We fix Δt and calculate Δy:

$$\Delta y = \Delta x - \left(\mu - \frac{\sigma^2}{2}\right) \Delta t \ .$$

If $\Delta y = 0$, then $\Delta x = \left(\mu - \frac{\sigma^2}{2}\right) \Delta t$. Therefore

$$
\begin{aligned}
\partial_t f \Big|_y &= \lim_{\Delta t \to 0} \frac{f(x + \Delta x, t + \Delta t) - f(x, t)}{\Delta t} \Bigg|_y \\
&= \lim_{\Delta t \to 0} \frac{1}{\Delta t} \left[\left(\partial_x f \Big|_t\right) \Delta x + \left(\partial_t f \Big|_x\right) \Delta t + O\left(\Delta x^2\right) + O\left(\Delta t^2\right) \right] \\
&= \lim_{\Delta t \to 0} \frac{1}{\Delta t} \frac{\left[\left(\partial_x f \Big|_t\right)\left(\mu - \frac{\sigma^2}{2}\right) + \left(\partial_t f \Big|_x\right)\right] \Delta t}{\Delta t} \\
&= \left(\mu - \frac{\sigma^2}{2}\right) \left(\partial_x f \Big|_t\right) + \left(\partial_t f \Big|_x\right) \ .
\end{aligned}
$$

We express the backward equation (9) in the y and t variables as

$$\partial_t f + \frac{\sigma^2}{2} \partial_y^2 f = 0 \ .$$

The difference is that ∂_t refers to the derivative with y fixed rather than with x fixed as in (9). One more step will transform this to the standard heat

equation. We rescale the space variable to get rid of the coefficient σ^2. The change of variables that does that is

$$\frac{\sigma}{y} = \frac{1}{z} \quad , \qquad z = \frac{y}{\sigma} \quad , \qquad y = \sigma z \; .$$

In these variables,

$$\partial_t f + \frac{1}{2} \partial_z^2 f = 0 \; .$$

The payoff for these manipulations is that we know a lot about solutions of the heat equation. We know the Green's function and the Fourier transform. All of these tools now apply to the equation (8) too.

3   Simulating an SDE

There are three primary ways to get information about the solution of an SDE. The first is to solve it exactly in closed form. That option is limited to very few SDE's. The second is to solve the backward equation numerically. We discuss in a future class how to do that. For now it suffices to say that this approach is impractical for SDE systems with more than a few components. The third is direct numerical simulation of the SDE. We discuss that option here.

Consider an SDE (5). We want to approximate the path $X_{[0,T]}$. Choose a small Δt, define $t_n = n\Delta t$, and define the approximate sample path to be $X_n^{(\Delta t)} \approx X_{t_n}$. There are two forms of the approximation algorithm. One is

$$X_{n+1}^{(\Delta t)} = X_n^{(\Delta t)} + a(X_n^{(\Delta t)}, t_n)\Delta t + b(X_n^{(\Delta t)}, t_n)\Delta W_n \; . \tag{10}$$

Here W_t is a Brownian motion path, and $\Delta W_n = W_{t_{n+1}} - W_{t_n}$, is the increment for time Δt. In practice, we have to generate the Brownian motion increments using a random number generator. The properties of ΔW_n are that they are independent, and that they are Gaussian with mean zero and variance Δt. In case it is a multi-variate Brownian motion, the covariance is ΔI. We generate such random variables starting with independent standard normals Z_n and multiplying by $\Delta t^{1/2}$:

$$\Delta W_n = \Delta t^{1/2} Z_n \quad , \qquad Z_n \sim \mathcal{N}(0, I) \; . \tag{11}$$

This algorithm is motivated by the strong form of the SDE. When we integrate (5) over the time increment $[t_n, t_n + \Delta t]$, we find

$$X_{t_{n+1}} = X_{t_n} + \int_{t_n}^{t_{n+t}} a(X_t, t)dt + \int_{t_n}^{t_{n+t}} b(X_t, t)dW_t \; . \tag{12}$$

If Δt is small, then $X_t \approx X_{t_n}$ for $t \in [t_n, t_{n+1}]$. If we replace X_t by the approximate value X_{t_n} in (12), the integrals simplify to

$$\int_{t_n}^{t_{n+t}} a(X_t, t)dt \approx a(X_{t_n}, t_n)\Delta t \; ,$$

and

$$\int_{t_n}^{t_{n+t}} b(X_t,t)dW_t \approx b(X_{t_n},t_n) \int_{t_n}^{t_{n+t}} dW_t = b(X_{t_n},t_n)\Delta W_n \ .$$

This gives the approximate formula

$$X_{t_{n+1}} \approx X_{t_n} + a(X_{t_n},t_n)\Delta t + b(X_{t_n},t_n)\Delta W_n \ .$$

The approximate formula for the exact path motivates the exact formula (10) for the approximate path.

The other form of the approximation algorithm just looks for approximate sample paths that have the right mean and variance over a step of size Δt. For that purpose, let Z_n be a family of independent random variables with mean zero and variance 1, or covariance I. You want

$$\mathrm{E}\Big[X_{n+1}^{(\Delta t)} \mid \mathcal{F}_n \Big] = X_n^{(\Delta t)} + a(X_n^{(\Delta t)},t_n)\Delta t \ ,$$

and

$$\mathrm{var}\Big(X_{n+1}^{(\Delta t)} \mid \mathcal{F}_n \Big) = b^2(X_n^{(\Delta t)},t_n)\Delta t \ .$$

These formulas are not exact for SDE paths, but they will hold exactly for approximate sample paths $X^{(\Delta t)}$. The formula

$$X_{n+1}^{(\Delta t)} = X_n^{(\Delta t)} + a(X_n^{(\Delta t)},t_n)\Delta t + b(X_n^{(\Delta t)},t_n)\Delta t^{1/2}Z_n \qquad (13)$$

makes these true. The actual algorithms (10) and (13) are identical, if we use a Gaussian Z_n in (13). The difference is only in interpretation. In the strong form (10) we are generating an approximate path that is a function of the driving Brownian motion. In the weak form (13), we are just making a path with approximately the right statistics. In either case, generating an approximate sample path might be called *simulating* the SDE.

We are usually interested in more than simulations. Instead we want to know expected values of functions of the path. These could be expected payouts or hitting probabilities or something more complicated. A single path or a small number of paths may not represent the entire "population" of paths. The only way to learn about population properties is to do a large number of simulations. The main way to digest a large collection of sample paths is to compute statistics of them. Most such statistics correspond the expected value of some function of a path.

This brings up an important distinction, between *simulation* and *Monte Carlo*. Simulation is described above. Monte Carlo[1] it the process of using random numbers to compute something that itself is not random. A hitting probability is not a random number, for example, though it is defined in terms of a random process. The distinction is important because it suggests that there may be more than one way to estimate the same number. We will discuss this a little later when we talk about applications of Girsanov's theorem.

---

[1]This thoughtful definition may be found in the book *Monte Carlo Methods* by Malvin Kalos and Paula Whitlock.

For now, let $V(x_{[0,T]})$ be some function of a path on the interval $[0, T]$. Let its expected value for the SDE be

$$A = \mathrm{E}\big[V(X_{[0,T]}) \big] \ .$$

For example, to estimate a hitting probability you might take $V = \mathbf{1}_{\tau \leq T}$. To estimate A, we generate a large number of independent approximate sample paths $X_{[0,T],k}^{(\Delta t)}$, $k = 1, \ldots, L$. Here L is the *sample size*, which is the number of paths. The estimate of A is

$$\widehat{A} = \frac{1}{L} \sum_{k=1}^{L} V\left(X_{[0,T],k}^{(\Delta t)} \right) \ . \tag{14}$$

The error is $\widehat{A} - A$. This error is composed of *bias* and *statistical error*. Bias comes from the fact that sample paths are not exact. You reduce bias by letting $\Delta t \to 0$.

$$\mathrm{bias} = \mathrm{E}\left[\widehat{A} \right] - A = \mathrm{E}\left[V\left(X_{[0,T],k}^{(\Delta t)} \right) \right] - \mathrm{E}\big[V(X_{[0,T]}) \big] \ .$$

In statistics, we say a statistic is *unbiased* if the expected value of the statistic is the true parameter value. There are unbiased statistics, but it is very rare than the estimate of a quantity like A is unbiased. Statistical error comes from the fact that we use a finite number of sample paths. You reduce statistical error by taking $L \to \infty$. The definition is

$$\mathrm{statistical\ error} = \widehat{A} - \mathrm{E}\left[\widehat{A} \right] \ .$$

Neither the bias nor the statistical error goes to zero very fast as $\Delta t \to 0$ or $L \to \infty$. The bias typically is proportional to Δt or $\Delta t^{1/2}$, depending on the problem. The statistical error typically is proportional to $L^{-1/2}$, which comes from the central limit theorem. For this reason Monte Carlo estimation either is very expensive, or not very accurate, or both.

You could ask about the scientific justification for using SDE models if all we do with them is discretize and simulate. Couldn't we just have simulated the original process? There are several justifications for the SDE approach. One has to do with time scales. We saw on an old assignment that the diffusion process may be an approximation to another process that operates on a very fast time scale, T_m. It is possible that the time step Δt needed to simulate the SDE is much larger than T_m (the "microscopic" time scale). The SDE model also can be simpler than the microscopic model it approximates. For example, the Ornstein-Uhlenbeck/Einstein model of Brownian motion replaces a process in (X, V) space with a simpler process in X space alone.

4 Existence of solutions

The first question asked in a graduate course on differential equations is: Do differential equations have solutions? You can ask the same question about

stochastic differential equations, and the answer is similar. If the coefficients $a(x,t)$ and $b(x,t)$ are *Lipschitz continuous*, then a simple iteration argument shows that solutions exist. If the coefficients are not Lipschitz continuous, all questions get harder and more problem specific.

A function $f(x)$ is Lipschitz continuous with *Lipschitz constant C* if

$$|f(x) - f(y)| \le C|x - y| \ . \tag{15}$$

For example, $f(x) = \sin(2x)$ is Lipschitz with Lipschitz constant $C = 2$. The functions $f(x) = e^x$, $f(x) = x^2$, and $f(x) = \sin(x^2)$ are not Lipschitz continuous. If f is differentiable and $|f'(x)| \le C$ for all x, then f is Lipschitz continuous with constant C. We say that f is *locally* Lipschitz near the point x_0 if there is an R so that (15) holds whenever $|x - x_0| \le R$ and $|y - x_0| \le R$. Any differentiable function is locally Lipschitz. The examples show that many nice seeming functions are not *globally* Lipschitz.

In the easy part of the existence theory for differential equations, locally Lipshitz equations have local solutions (solutions defined for a finite but possibly not infinite range of t). Globally Lipschitz equations have solutions defined globally in time, which is to say, for all $t \in [0, \infty)$. In particular, if a and b are (globally) Lipschitz then the SDE (5) has a solution defined globally in time. More precisely, there is a function $W_{[0,t]} \to X_t$ so that the process X_t satisfies (5). This function may be constructed as a limit of the *Picard iteration* process. This defines a sequence of approximate solutions $X_{t,k}$ and recovers the exact solution in the limit $k \to \infty$. The iteration is

$$dX_{t,k+1} = a(X_{t,k}, t)dt + b(X_{t,k}, t)dW_k \ .$$

In integral form, this is

$$X_{t,k+1} = x_0 + \int_0^t a(X_{s,k}, s)ds + \int_0^t b(X_{s,k}, s)dW_s \ .$$

We must prove that these approximations converge to something.

Here is a quick version of the argument for the simpler case $a = 0$. We subtract the k and $k-1$ equations to get

$$X_{t,k+1} - X_{t,k} = \int_0^t \left(b(X_{s,k}, s) - b(X_{s,k-1}, s) \right) dW_s \ .$$

The object is to prove that $X_{t,k+1} - X_{t,k}$ is smaller than $X_{t,k} - X_{t,k-1}$ eventually as $k \to \infty$. You can use the Ito isometry formula to get

$$\mathrm{E}\left[(X_{t,k+1} - X_{t,k})^2 \right] = \int_0^t \mathrm{E}\left[(b(X_{s,k}, s) - b(X_{s,k-1}, s))^2 \right] ds \ .$$

Since b is Lipschitz continuous,

$$(b(X_{s,k}, s) - b(X_{s,k-1}, s))^2 \le C^2 (X_{s,k} - X_{s,k-1})^2 \ .$$

9

Therefore,

$$\mathrm{E}\left[(X_{t,k+1} - X_{t,k})^2 \right] \leq C^2 \int_0^t \mathrm{E}\left[(X_{s,k} - X_{s,k-1})^2 \right] ds$$

Now define

$$M_{t,k} = \mathrm{E}\left[(X_{t,k+1} - X_{t,k})^2 \right] .$$

Our integral inequality is

$$M_{t,k} \leq C^2 \int_0^t M_{s,k-1} ds .$$

It is easy to derive from this

$$M_{t,k} \leq M_{0,t} \frac{e^{C^t t}}{k!} .$$

For any fixed t, this implies that the $X_{k,t}$ are a Cauchy sequence almost surely (our Borel Cantelli lemma again).