Chapter 19

Stochastic Differential Equations

Section 19.1 gives two easy examples of Itô integrals. The second one shows that there's something funny about change of variables, or if you like about the chain rule.

Section 19.2 explains how to do change of variables in a stochastic integral, also known as "Itô's formula".

Section 19.3 defines stochastic differential equations.

Section 19.4 sets up a more realistic model of Brownian motion, leading to an SDE called the Langevin equation, and solves it to get Ornstein-Uhlenbeck processes.

19.1 Some Easy Stochastic Integrals, with a Moral

19.1.1 $\int dW$

Let's start with the easiest possible stochastic integral:

$$\int_a^b dW \tag{19.1}$$

i.e., the Itô integral of the function which is always 1, $\mathbf{1}_{\mathbb{R}^+}(t)$. If this is any kind of integral at all, it should be W — more exactly, because this is a definite integral, we want $\int_a^b dW = W(b) - W(a)$. Mercifully, this works. Pick any set of time-points t_i we like, and treat 1 as an elementary function with those times as its break-points. Then, using our definition of the Itô integral for elementary functions,

$$\int_a^b dW = \sum_{t_i} W(t_{i+1}) - W(t_i) \tag{19.2}$$

$$= W(b) - W(a) \tag{19.3}$$

as required. (This would be a good time to convince yourself that adding extra break-points to an elementary function doesn't change its integral.)

19.1.2 $\int W dW$

Tradition dictates that the next example be $\int W dW$. First, we should convince ourselves that $W(t)$ is Itô-integrable: it's clearly measurable and non-anticipative, but is it square-integrable? Yes; by Fubini's theorem,

$$\mathbf{E}\left[\int_0^t W^2(s)ds\right] = \int_0^t \mathbf{E}\left[W^2(s)\right] ds \tag{19.4}$$

$$= \int_0^t s ds \tag{19.5}$$

which is clearly finite on finite intervals $[0,t]$. So, this integral should exist. Now, if the ordinary rules for change of variables held — equivalent, if the chain-rule worked the usual way — we could say that $W dW = \frac{1}{2}d(W^2)$, so $\int W dW = \frac{1}{2}\int dW^2$, and we'd expect $\int_0^t W dW = \frac{1}{2}W^2(t)$. But, alas, this can't be right. To see why, take the expectation: it'd be $\frac{1}{2}t$. But we know that it has to be zero, and it has to be a martingale in t, and this is neither. A bone-head would try to fix this by subtracting off the non-martingale part, i.e., a bone-head would guess that $\int_0^t W dW = \frac{1}{2}W^2(t) - t/2$. Annoyingly, in this case the bone-head is correct. The demonstration is fundamentally straightforward, if somewhat long-winded.

To begin, we need to approximate W by elementary functions. For each n, let $t_i = i\frac{t}{2^n}$, $0 \leq i \leq 2^n - 1$. Set $\phi_n(t) = \sum_{i=0}^{2^n-1} W(t_i)\mathbf{1}_{[t_i,t_{i+1})}$. Let's check that this converges to $W(t)$ as $n \to \infty$:

$$\begin{aligned}
\mathbf{E}\left[\int_0^t (\phi_n(s) - W(s))^2 ds\right] &= \mathbf{E}\left[\sum_{i=0}^{2^n-1} \int_{t_i}^{t_{i+1}} (B(t_i) - B(s))^2 ds\right] && (19.6)\\
&= \sum_{i=0}^{2^n-1} \mathbf{E}\left[\int_{t_i}^{t_{i+1}} (B(t_i) - B(s))^2 ds\right] && (19.7)\\
&= \sum_{i=0}^{2^n-1} \int_{t_i}^{t_{i+1}} \mathbf{E}\left[(B(t_i) - B(s))^2\right] ds && (19.8)\\
&= \sum_{i=0}^{2^n-1} \int_{t_i}^{t_{i+1}} (s - t_i) ds && (19.9)\\
&= \sum_{i=0}^{2^n-1} \int_0^{2^{-n}} s ds && (19.10)\\
&= \sum_{i=0}^{2^n-1} \left[\frac{t^2}{2}\right]_0^{2^{-n}} && (19.11)\\
&= \sum_{i=0}^{2^n-1} 2^{-2n-1} && (19.12)\\
&= 2^{-n-1} && (19.13)
\end{aligned}$$

which $\to 0$ as $n \to \infty$. Hence

$$\begin{aligned}
\int_0^t W(s)dW &= \lim_n \int_0^t \phi_n(s)dW && (19.14)\\
&= \lim_n \sum_{i=0}^{2^n-1} W(t_i)(W(t_{i+1}) - W(t_i)) && (19.15)\\
&= \lim_n \sum_{i=0}^{2^n-1} W(t_i)\Delta W(t_i) && (19.16)
\end{aligned}$$

where $\Delta W(t_i) \equiv W(t_{i+1}) - W(t_i)$, because I'm getting tired of writing both subscripts. Define $\Delta W^2(t_i)$ similarly. Since $W(0) = 0 = W^2(0)$, we have that

$$\begin{aligned}
W(t) &= \sum_i \Delta W(t_i) && (19.17)\\
W^2(t) &= \sum_i \Delta W^2(t_i) && (19.18)
\end{aligned}$$

Now let's re-write ΔW^2 in such a way that we get a $W\Delta W$ term, which is what we want to evaluate our integral.

$$\Delta W^2(t_i) = W^2(t_{i+1}) - W^2(t_i) \tag{19.19}$$
$$= (\Delta W(t_i) + W(t_i))^2 - W^2(t_i) \tag{19.20}$$
$$= (\Delta W(t_i))^2 + 2W(t_i)\Delta W(t_i) + W^2(t_i) - W^2(t_i) \tag{19.21}$$
$$= (\Delta W(t_i))^2 + 2W(t_i)\Delta W(t_i) \tag{19.22}$$

This looks promising, because it's got $W \Delta W$ in it. Plugging in to Eq. 19.18,

$$W^2(t) = \sum_i (\Delta W(t_i))^2 + 2W(t_i)\Delta W(t_i) \tag{19.23}$$
$$\sum_i W(t_i)\Delta W(t_i) = \frac{1}{2}W^2(t) - \frac{1}{2}\sum_i (\Delta W(t_i))^2 \tag{19.24}$$

Now, it is possible to show (Exercise 19.1) that

$$\lim_n \sum_{i=0}^{2^n-1} (\Delta W(t_i))^2 = t \tag{19.25}$$

in L^2, so we have that

$$\int_0^t W(s)dW = \lim_n \sum_{i=0}^{2^n-1} W(t_i)\Delta W(t_i) \tag{19.26}$$
$$= \frac{1}{2}W^2(t) - \lim_n \sum_{i=0}^{2^n-1} (\Delta W(t_i))^2 \tag{19.27}$$
$$= \frac{1}{2}W^2(t) - \frac{t}{2} \tag{19.28}$$

as required.

Clearly, something weird is going on here, and it would be good to get to the bottom of this. At the very least, we'd like to be able to use change of variables, so that we can find functions of stochastic integrals.

19.2 Itô's Formula

Integrating $\int W dW$ has taught us two things: first, we want to avoid evaluating Itô integrals directly from the definition; and, second, there's something funny about change of variables in Itô integrals. A central result of stochastic calculus, known as *Itô's formula*, gets us around both difficulties, by showing how to write functions of stochastic integrals as, themselves, stochastic integrals.

Definition 197 (Itô Process) *If A is a non-anticipating measurable process, B is Itô-integrable, and X_0 is an L_2 random variable independent of W, then $X(t) = X_0 + \int_0^t A(s)ds + \int_0^t B(s)dW$ is an* Itô process. *This is equivalently written $dX = Adt + BdW$.*

Lemma 198 *Every Itô process is non-anticipating.*

PROOF: Clearly, the non-anticipating processes are closed under linear operations, so it's enough to show that the three components of any Itô process are non-anticipating. But a process which is always equal to $X_0 \perp\!\!\!\perp W(t)$ is clearly non-anticipating. Similarly, since $A(t)$ is non-anticipating, $\int A(s)ds$ is too: its natural filtration is smaller than that of A, so it cannot provide more information about $W(t)$, and A is, by assumption, non-anticipating. Finally, Itô integrals are always non-anticipating, so $\int B(s)dW$ is non-anticipating. □

Theorem 199 (Itô's Formula (One-Dimension)) *Suppose X is an Itô process with $dX = Adt + BdW$. Let $f(t, x) : \mathbb{R}^+ \times \mathbb{R} \mapsto \mathbb{R}$ be a function with continuous partial time derivative $\frac{\partial f}{\partial t}$, and continuous second partial space derivative, $\frac{\partial^2 f}{\partial x^2}$. Then $F(t) = f(t, X(t))$ is an Itô process, and*

$$dF = \frac{\partial f}{\partial t}(t, X(t))dt + \frac{\partial f}{\partial x}(t, X(t))dX + \frac{1}{2}B^2(t)\frac{\partial^2 f}{dx^2}(t, X(t))dt \tag{19.29}$$

That is,

$$F(t) - F(0) = \tag{19.30}$$
$$\int_0^t \left[\frac{\partial f}{\partial t}(s, X(s)) + A(s)\frac{\partial f}{\partial x}(s, X(s)) + \frac{1}{2}B^2(s)\frac{\partial^2 f}{\partial x^2}(s, X(s))\right] dt + \int_0^t B(s)\frac{\partial f}{\partial x}(s, X(s))dW$$

PROOF: I will suppose first of all that f, and its partial derivatives appearing in Eq. 19.29, are all bounded. (You can show that the general case of C^2 functions can be uniformly approximated by functions with bounded derivatives.) I will further suppose that A and B are elementary processes, since in the last chapter we saw how to use them to approximate general Itô-integrable functions. (If you are worried about the interaction of all these approximations and simplifications, I commend your caution, and suggest you step through the proof in the general case.)

For each n, let $t_i = i\frac{t}{2^n}$, as in the last section. Define $\Delta t_i \equiv t_{i+1} - t_i$, $\Delta X(t_i) = X(t_{i+1}) - X(t_i)$, etc. Thus

$$F(t) = f(t, X(t)) = f(0, X(0)) + \sum_{i=0}^{2^n-1} \Delta f(t_i, X(t_i)) \tag{19.31}$$

Now we'll approximate the increments of F by a Taylor expansion:

$$\begin{aligned} F(t) &= f(0, X(0)) + \sum_{i=0}^{2^n-1} \frac{\partial f}{\partial t}\Delta t_i \\ &+ \sum_{i=0}^{2^n-1} \frac{\partial f}{\partial x}\Delta X(t_i) \\ &+ \frac{1}{2}\sum_{i=0}^{2^n-1} \frac{\partial^2 f}{\partial t^2}(\Delta t_i)^2 \\ &+ \sum_{i=0}^{2^n-1} \frac{\partial^2 f}{\partial t \partial x}\Delta t_i \Delta X(t_i) \\ &+ \frac{1}{2}\sum_{i=0}^{2^n-1} \frac{\partial^2 f}{\partial x^2}(\Delta X(t_i))^2 \\ &+ \sum_{i=0}^{2^n-1} R_i \end{aligned} \tag{19.32}$$

Because the derivatives are bounded, all the remainder terms R_i are $o((\Delta t_i)^2 + (\Delta X(t_i))^2)$. We will come back to showing that the remainders are harmless, but for now let's concentrate on the leading-order components of the Taylor expansion.

First, as $n \rightarrow \infty$,

$$\sum_{i=0}^{2^n-1} \frac{\partial f}{\partial t}\Delta t_i \rightarrow \int_0^t \frac{\partial f}{\partial t} ds \tag{19.33}$$

$$\sum_{i=0}^{2^n-1} \frac{\partial f}{\partial x}\Delta X(t_i) \rightarrow \int_0^t \frac{\partial f}{\partial x} dX \tag{19.34}$$

$$\equiv \int_0^t \frac{\partial f}{\partial x} A(s)dt + \int_0^t \frac{\partial f}{\partial x} B(s)dW \tag{19.35}$$

[You can use the definition in the last line to build up a theory of stochastic integrals with respect to arbitrary Itô processes, not just Wiener processes.]

$$\sum_{i=0}^{2^n-1} \frac{\partial^2 f}{\partial t^2}(\Delta t_i)^2 \rightarrow 0 \int_0^t \frac{\partial^2 f}{\partial t^2} ds = 0 \tag{19.36}$$

Next, since A and B are (by assumption) elementary,

$$\sum_{i=0}^{2^n-1} \frac{\partial^2 f}{\partial x^2} (\Delta X(t_i))^2 = \sum_{i=0}^{2^n-1} \frac{\partial^2 f}{\partial x^2} A^2(t_i)\,(\Delta t_i)^2 + 2\sum_{i=0}^{2^n-1} \frac{\partial^2 f}{\partial x^2} A(t_i)B(t_i)\Delta t_i \Delta W(t_i) + \sum_{i=0}^{2^n-1} \frac{\partial^2 f}{\partial x^2} B^2(t_i)(\Delta W(t_i))^2 \tag{19.37}$$

The first term on the right-hand side, in $(\Delta t)^2$, goes to zero as n increases. Since A is square-integrable and $\frac{\partial^2 f}{\partial x^2}$ is bounded, $\sum \frac{\partial^2 f}{\partial x^2} A^2(t_i)\Delta t_i$ converges to a finite value as $\Delta t \to 0$, so multiplying by another factor Δt, as $n \to \infty$, gives zero. (This is the same argument as the one for Eq. 19.36.) Similarly for the second term, in $\Delta t \Delta X$:

$$\lim_n \sum_{i=0}^{2^n-1} \frac{\partial^2 f}{\partial x^2} A(t_i)B(t_i)\Delta t_i \Delta W(t_i) = \lim_n \frac{t}{2^n} \int_0^t \frac{\partial^2 f}{\partial x^2} A(s)B(s)dW \tag{19.38}$$

because A and B are elementary and the partial derivative is bounded. Now apply the Itô isometry:

$$\mathbf{E}\left[\left(\frac{t}{2^n}\int_0^t \frac{\partial^2 f}{\partial x^2} A(s)B(s)dW\right)^2\right] = \frac{t^2}{2^{2n}} \mathbf{E}\left[\int_0^t \left(\frac{\partial^2 f}{\partial x^2}\right)^2 A^2(s)B^2(s)ds\right]$$

The time-integral on the right-hand side is finite, since A and B are square-integrable and the partial derivative is bounded, and so, as n grows, both sides go to zero. But this means that, in L_2,

$$\sum_{i=0}^{2^n-1} \frac{\partial^2 f}{\partial x^2} A(t_i)B(t_i)\Delta t_i \Delta W(t_i) \rightarrow 0 \tag{19.39}$$

The third term, in $(\Delta X)^2$, does *not* vanish, but rather converges in L_2 to a time integral:

$$\sum_{i=0}^{2^n-1} \frac{\partial^2 f}{\partial x^2} B^2(t_i)(\Delta W(t_i))^2 \rightarrow \int_0^t \frac{\partial^2 f}{\partial x^2} B^2(s)ds \tag{19.40}$$

You will prove this in part b of Exercise 19.1.

The mixed partial derivative term has no counterpart in Itô's formula, so it needs to go away.

$$\sum_{i=0}^{2^n-1} \frac{\partial^2 f}{\partial t \partial x} \Delta t_i \Delta X(t_i) = \sum_{i=0}^{2^n-1} \frac{\partial^2 f}{\partial t \partial x}\left[A(t_i)(\Delta t_i)^2 + B(t_i)\Delta t_i \Delta W(t_i)\right] \tag{19.41}$$

$$\sum_{i=0}^{2^n-1} \frac{\partial^2 f}{\partial t \partial x} A(t_i)(\Delta t_i)^2 \quad \rightarrow \quad 0 \tag{19.42}$$

$$\sum_{i=0}^{2^n-1} \frac{\partial^2 f}{\partial t \partial x} B(t_i)\Delta t_i \Delta W(t_i) \quad \rightarrow \quad 0 \tag{19.43}$$

where the argument for Eq. 19.43 is the same as that for Eq. 19.36, while that for Eq. 19.43 follows the pattern of Eq. 19.39.

Let us, as promised, dispose of the remainder term. Clearly,

$$(\Delta X)^2 = A^2(\Delta t)^2 + 2AB\Delta t\Delta W + B^2(\Delta W)^2 \tag{19.44}$$

$$= A^2(\Delta t)^2 + 2AB\Delta t\Delta W + B^2\Delta t \tag{19.45}$$

so, from the foregoing, it is clear that this goes to zero as $\Delta t \rightarrow 0$. Hence the remainder term will vanish as n increases.

Putting everything together, we have that

$$F(t) - F(0) = \int_0^t \left[\frac{\partial f}{\partial t} + \frac{\partial f}{\partial x}A + \frac{1}{2}B^2\frac{\partial^2 f}{\partial x^2}\right] dt + \int_0^t \frac{\partial f}{\partial x} B dW \tag{19.46}$$

exactly as required. This completes the proof, under the stated restrictions on f, A and B; approximation arguments extend the result to the general case. □

Remark 1. Our manipulations in the course of the proof are often summarized in the following multiplication rules for differentials: $dtdt = 0$, $dWdt = 0$, $dtdW = 0$, and, most important of all,

$$dWdW = dt$$

This last is of course related to the linear growth of the variance of the increments of the Wiener process.

Remark 2. Re-arranging Itô's formula a little yields

$$dF = \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial x}dX + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}dt \tag{19.47}$$

The first two terms are what we expect from the ordinary rules of calculus; it's the third term which is new and strange. Notice that it disappears if $\frac{\partial^2 f}{\partial x^2} = 0$. When we come to stochastic differential equations, this will correspond to state-independent noise.

Remark 3. One implication of Itô's formula is that *Itô processes are closed under the application of C^2 mappings.*

Example 200 *The integral $\int WdW$ is now trivial. Let $X(t) = W(t)$ (by setting $A = 0$, $B = 1$ in the definition of an Itô process), and $f(t,x) = x^2/2$. Applying*

Itô's formula,

$$dF = \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial x}dW + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}dt \tag{19.48}$$

$$\frac{1}{2}dW^2 = WdW + \frac{1}{2}dt \tag{19.49}$$

$$\frac{1}{2}\int dW^2 = \int WdW + \frac{1}{2}\int dt \tag{19.50}$$

$$\int_0^t W(s)dW = \frac{1}{2}W^2(t) - \frac{t}{2} \tag{19.51}$$

All of this extends naturally to higher dimensions.

Definition 201 (Multidimensional Itô Process) *Let A by an n-dimensional vector of non-anticipating processes, B an $n \times m$ matrix of Itô-integrable processes, and W an m-dimensional Wiener process. Then*

$$X(t) = X(0) + \int_0^t A(s)ds + \int_0^t B(s)dW \tag{19.52}$$

$$dX = A(t)dt + B(t)dW \tag{19.53}$$

*is an n-*dimensional Itô process*.*

Theorem 202 (Itô's Formula (Multidimensional)) *Let $X(t)$ be an n-dimensional Itô process, and let $f(t,x) : \mathbb{R}^+ \times \mathbb{R}^n \mapsto \mathbb{R}^m$ have a continuous partial time derivative and continuous second partial space derivatives. Then $F(t) = f(t, X(t))$ is an m-dimensional Itô process, whose k^{th} component F_k is given by*

$$dF_k = \frac{\partial g_k}{\partial t}dt + \frac{\partial g_k}{\partial x_i}dX_i + \frac{1}{2}\frac{\partial^2 g_k}{\partial X_i \partial X_j}dX_i dX_j \tag{19.54}$$

summing over repeated indices, with the understanding that $dW_i dW_j = \delta_{ij}dt$, $dW_i dt = dt dW_i = dt dt = 0$.

Proof: Entirely parallel to the one-dimensional case, only with even more algebra. □

It is also possible to define Wiener processes and stochastic integrals on arbitrary curved manifolds, but this would take us way, way too far afield.

19.2.1 Stratonovich Integrals

It is possible to make the extra term in Eq. 19.47 go away, and have stochastic differentials which work just like the ordinary ones. This corresponds to making stochastic integrals limits of sums of the form

$$\sum_i X\left(\frac{t_{i+1} + t_i}{2}\right)\Delta W(t_i)$$

rather than the Itô sums we are using,

$$\sum_i X(t_i)\Delta W(t_i)$$

That is, we could evade the Itô formula if we evaluated our test function in the middle of intervals, rather than at their beginnning. This leads to what are called *Stratonovich integrals*. However, while Stratonovich integrals give simpler change-of-variable formulas, they have many other inconveniences: they are not martingales, for instance, and the nice connections between the form of an SDE and its generator, which we will see and use in the next chapter, go away. Fortunately, every Stratonovich SDE can be converted into an Itô SDE, and vice versa, by adding or subtracting the appropriate noise term.

19.2.2 Martingale Representation

One property of the Itô integral is that it is always a square-integrable martingale. Remarkably enough, the converse is also true. In the interest of time, I omit the proof of the following theorem; there is one using only tools we've seen so far in Øksendal (1995, ch. 4).

Theorem 203 *Let $M(t)$ be a martingale, with $\mathbf{E}\left[M^2(t)\right] < \infty$ for all $t \geq 0$. Then there exists a unique process $M'(t)$, Itô-integrable for all finite positive t, such that*

$$M(t) = \mathbf{E}\left[M(0)\right] + \int_0^t M'(t)dW \ a.s. \tag{19.55}$$

19.3 Stochastic Differential Equations

Definition 204 (Stochastic Differential Equation, Solutions) *Let $a(x) : \mathbb{R}^n \mapsto \mathbb{R}^n$ and $b(x) : \mathbb{R}^n \mapsto \mathbb{R}^{nm}$ be measurable functions (vector and matrix valued, respectively), W an m-dimensional Wiener process, and X_0 an L_2 random variable in $\mathbb{R}^n$, independent of W. Then an $\mathbb{R}^n$-valued stochastic process X on $\mathbb{R}^+$ is a solution to the* autonomous stochastic differential equation

$$dX = a(X)dt + b(X)dW, \ X(0) = X_0 \tag{19.56}$$

when, with probability 1, it is equal to the corresponding Itô process,

$$X(t) \quad = \quad X_0 + \int_0^t a(X(s))ds + \int_0^s b(X(s))dW \ a.s. \tag{19.57}$$

The a term is called the drift*, and the b term the* diffusion*.*

Remark 1: A given process X can fail to be a solution either because it happens not to agree with Eq. 19.57, or, perhaps more seriously, because the integrals on the right-hand side don't even exist. This can, in particular, happen if $b(X(t))$ is anticipating. For a *fixed* choice of Wiener process, there are circumstances where otherwise reasonable SDEs have no solution, for basically this reason — the Wiener process is constructed in such a way that the class of Itô processes is impoverished. This leads to the idea of a *weak solution* to Eq. 19.56, which is a *pair* X, W such that W is a Wiener process, with respect to the appropriate filtration, and X then is given by Eq. 19.57. I will avoid weak solutions in what follows.

Remark 2: In a non-autonomous SDE, the coefficients would be explicit functions of time, $a(t, X)dt + b(t, X)dW$. The usual trick for dealing with non-autonomous n-dimensional ODEs is turn them into autonomous $n+1$-dimensional ODEs, making $x_{n+1} = t$ by decreeing that $x_{n+1}(t_0) = t_0$, $x'_{n+1} = 1$ (Arnol'd, 1973). This works for SDEs, too: add time as an extra variable with constant drift 1 and constant diffusion 0. Without loss of generality, therefore, I'll only consider autonomous SDEs.

Let's now prove the existence of unique solutions to SDEs. First, recall how we do this for ordinary differential equations. There are several approaches, most of which carry over to SDEs, but one of the most elegant is the "method of successive approximations", or "Picard's method" (Arnol'd, 1973, SS30–31)). To construct a solution to $dx/dt = f(x)$, $x(0) = x_0$, this approach uses functions $x_n(t)$, with $x_{n+1}(t) = x_0 + \int_0^t f(x_n(s)ds$, starting with $x_0(t) = x_0$. That is, there is an operator P such that $x_{n+1} = Px_n$, and x solves the ODE iff it is a fixed point of the operator. Step 1 is to show that the sequence x_n is Cauchy on finite intervals $[0, T]$. Step 2 uses the fact that the space of continuous functions is complete, with the topology of uniform convergence of compact sets — which, for $\mathbb{R}^+$, is the same as uniform convergence on finite intervals. So, x_n has a limit. Step 3 is to show that the limit point must be a fixed point of P, that is, a solution. Uniqueness is proved by showing that there cannot be more than one fixed point.

Before plunging in to the proof, we need some lemmas: an algebraic triviality, a maximal inequality for martingales, a consequent maximal inequality for Itô processes, and an inequality from real analysis about integral equations.

Lemma 205 *For any real numbers a and b, $(a+b)^2 \leq 2a^2 + 2b^2$.*

Proof: No matter what a and b are, a^2, b^2, and $(a-b)^2$ are non-negative, so

$$
\begin{aligned}
(a-b)^2 &\geq 0 && (19.58)\\
a^2 + b^2 - 2ab &\geq 0 && (19.59)\\
a^2 + b^2 &\geq 2ab && (19.60)\\
2a^2 + 2b^2 &\geq a^2 + 2ab + b^2 = (a+b)^2 && (19.61)
\end{aligned}
$$

□

Definition 206 (Maximum Process) *Given a stochastic process $X(t)$, we define its* maximum process $X^*(t)$ *as* $\sup_{0 \leq s \leq t} |X(s)|$.

Remark: Example 78 was of course designed with malice aforethought.

Definition 207 *Let $\mathcal{QM}(T)$, $T > 0$, be the space of all non-anticipating processes, square-integrable on $[0,T]$, with norm $\|X\|_{\mathcal{QM}(T)} \equiv \|X^*(T)\|_2$.*

(Technically, this is only a norm on equivalence classes of processes, where the equivalence relation is "is a version of". You may make that amendment mentally as you read what follows.)

Lemma 208 *$\mathcal{QM}(T)$ is a complete normed space for each T.*

PROOF: Identical to the usual proof that L_p spaces are complete.

Lemma 209 (Doob's Martingale Inequalities) *If $M(t)$ is a continuous martingale, then, for all $p \geq 1$, $t \geq 0$ and $\epsilon > 0$,*

$$\mathbb{P}\left(M^*(t) \geq \epsilon\right) \leq \frac{\mathbf{E}\left[|M(t)|^p\right]}{\epsilon^p} \quad (19.62)$$

$$\|M^*(t)\|_p \leq q\|M(t)\|_p \quad (19.63)$$

where $q^{-1} + p^{-1} = 1$. In particular, for $p = q = 2$,

$$\mathbf{E}\left[(M^*(t))^2\right] \leq 4\mathbf{E}\left[M^2(t)\right]$$

PROOF: See Propositions 7.15 and 7.16 in Kallenberg (pp. 128 and 129). □

These can be thought of as versions of the Markov inequality, only for martingales. They accordingly get used *all the time*.

Lemma 210 *Let $X(t)$ be an Itô process, $X(t) = X_0 + \int_0^t A(s)ds + \int_0^t B(s)dW$. Then there exists a constant C, depending only on T, such that, for all $t \in [0,T]$,*

$$\|X\|^2_{\mathcal{QM}(t)} \leq C\left(\mathbf{E}\left[X_0^2\right] + \mathbf{E}\left[\int_0^t A^2(s) + B^2(s)ds\right]\right) \quad (19.64)$$

PROOF: Clearly,

$$X^*(t) \leq |X_0| + \int_0^t |A(s)|ds + \sup_{0 \leq s \leq t}\left|\int_0^s B(s)dW\right| \quad (19.65)$$

$$(X^*(t))^2 \leq 2X_0^2 + 2\left(\int_0^t |A(s)|ds\right)^2 + 2\left(\sup_{0 \leq s \leq t}\left|\int_0^s B(s')dW\right|\right)^2 \quad (19.66)$$

by Lemma 205. By Jensen's inequality[1],

$$\left(\int_0^t |A(s)|ds\right)^2 \quad \leq \quad t\int_0^t A^2(s)ds \qquad (19.67)$$

Writing $I(t)$ for $\int_0^t B(s)dW$, and noticing that it is a martingale, we have, from Doob's inequality (Lemma 209), $\mathbf{E}\left[(I^*(t))^2\right] \leq 4\mathbf{E}\left[I^2(t)\right]$. But, from Itô's isometry (Corollary 196), $\mathbf{E}\left[I^2(t)\right] = \mathbf{E}\left[\int_0^t B^2(s)ds\right]$. Putting all the parts together, then,

$$\mathbf{E}\left[(X^*(t))^2\right] \quad \leq \quad 2\mathbf{E}\left[X_0^2\right] + 2\mathbf{E}\left[t\int_0^t A^2(s)ds + \int_0^t B^2(s)ds\right] \qquad (19.68)$$

and the conclusion follows, since $t \leq T$. □

Remark: The lemma also holds for multidimensional Itô processes, and for powers greater than two (though then the Doob inequality needs to be replaced by a different one: see Rogers and Williams (2000, Ch. V, Lemma 11.5, p. 129)).

Definition 211 *Given an SDE $dX = a(X)dt + b(X)dW$ with initial condition X_0, the corresponding integral operator $P_{X_0,a,b}$ is defined for all Itô processes Y as*

$$P_{X_0,a,b}Y(t) = X_0 + \int_0^t a(Y(s))ds + \int_0^t b(Y(s))dW \qquad (19.69)$$

Lemma 212 *Y is a solution of $dX = a(X)dt + b(X)dW$, $X(0) = X_0$, if and only if $P_{X_0,a,b}Y = Y$ a.s.*

PROOF: Obvious from the definitions. □

Lemma 213 *If a and b are uniformly Lipschitz continuous, with constants K_a and K_B, then, for some positive D depending only on T, K_a and K_b,*

$$\|P_{X_0,a,b}X - P_{X_0,a,b}Y\|^2_{\mathcal{QM}(t)} \leq D\int_0^t \|X - Y\|_{\mathcal{QM}(s)}ds \qquad (19.70)$$

PROOF: Since the SDE is understood to be fixed, abbreviate $P_{X_0,a,b}$ by P. Let X and Y be any two Itô processes. We want to find the $\mathcal{QM}(t)$ norm of

[1] Remember that Lebesgue measure isn't a probability measure on $[0,t]$, but $\frac{1}{t}ds$ is a probability measure, so we can apply Jensen's inequality to that. This is where the t on the right-hand side will come from.

$PX - PY$.

$$\begin{aligned} & |PX(t) - PY(t)| && (19.71) \\ = \; & \left| \int_0^t a(X(s)) - a(Y(s))dt + \int_0^t b(X(s)) - b(Y(s))dW \right| \\ \leq \; & \int_0^t |a(X(s)) - a(Y(s))|\, ds + \int_0^t |b(X(s)) - b(Y(s))|\, dW && (19.72) \\ \leq \; & \int_0^t K_a\, |X(s) - Y(s)|\, ds + \int_0^t K_b\, |X(s) - Y(s)|\, dW && (19.73) \\ & \|PX - PY\|^2_{\mathcal{QM}(t)} && (19.74) \\ \leq \; & C(K_a^2 + K_b^2)\mathbf{E}\left[\int_0^t |X(s) - Y(s)|^2 ds\right] \\ \leq \; & C(K_a^2 + K_b^2)t \int_0^t \|X - Y\|^2_{\mathcal{QM}(s)} ds && (19.75) \end{aligned}$$

which, as $t \leq T$, completes the proof. □

Lemma 214 (Gronwall's Inequality) *If f is continuous function on $[0,T]$ such that $f(t) \leq c_1 + c_2 \int_0^t f(s)ds$, then $f(t) \leq c_1 e^{c_2 t}$.*

Proof: See Kallenberg, Lemma 21.4, p. 415. □

Theorem 215 (Existence and Uniquness of Solutions to SDEs in One Dimension) *Let X_0, a, b and W be as in Definition 204, and let a and b be uniformly Lipschitz continuous. Then there exists a square-integrable, non-anticipating $X(t)$ which solves $dX = a(X)dt + b(X)dW$ with initial condition X_0, and this solution is unique (almost surely).*

Proof: I'll first prove existence, along with square-integrability, and then uniqueness. That X is non-anticipating follows from the fact that it is an Itô process (Lemma 198). For concision, abbreviate $P_{X_0,a,b}$ by P.

As with ODEs, iteratively construct approximate solutions. Fix a $T > 0$, and, for $t \in [0,T]$, set

$$\begin{aligned} X_0(t) &= X_0 && (19.76) \\ X_{n+1}(t) &= PX_n(t) && (19.77) \end{aligned}$$

The first step is showing that X_n is Cauchy in $\mathcal{QM}(T)$. Define $\phi_n(t) \equiv \|X_{n+1}(t) - X_n(t)\|^2_{\mathcal{QM}(t)}$. Notice that $\phi_n(t) = \|PX_n(t) - PX_{n-1}(t)\|^2_{\mathcal{QM}(t)}$, and that, for each n, $\phi_n(t)$ is non-decreasing in t (because of the supremum embedded in its definition). So, using Lemma 213,

$$\begin{aligned}
\phi_n(t) &\leq D\int_0^t \|X_n - X_{n-1}\|^2_{\mathcal{QM}(s)} ds && (19.78)\\
&\leq D\int_0^t \phi_{n-1}(s) ds && (19.79)\\
&\leq D\int_0^t \phi_{n-1}(t) ds && (19.80)\\
&= Dt\phi_{n-1}(0) && (19.81)\\
&\leq \frac{D^n t^n}{n!}\phi_0(t) && (19.82)\\
&\leq \frac{D^n t^n}{n!}\phi_0(T) && (19.83)
\end{aligned}$$

Since, for any constant c, $c^n/n! \rightarrow 0$, to get the successive approximations to be Cauchy, we just need to show that $\phi_0(T)$ is finite, using Lemma 210.

$$\begin{aligned}
\phi_0(T) &= \|P_{X_0,a,b,}X_0 - X_0\|^2_{\mathcal{QM}(T)} && (19.84)\\
&= \left\|\int_0^t a(X_0)ds + \int_0^t b(X_0)dW\right\|^2_{\mathcal{QM}(T)} && (19.85)\\
&\leq C\mathbf{E}\left[\int_0^T a^2(X_0) + b^2(X_0)ds\right] && (19.86)\\
&\leq CT\mathbf{E}\left[a^2(X_0) + b^2(X_0)\right] && (19.87)
\end{aligned}$$

Because a and b are Lipschitz, this will be finite if X_0 has a finite second moment, which, by assumption, it does. So X_n is a Cauchy sequence in $\mathcal{QM}(T)$, which is a complete space, so X_n has a limit in $\mathcal{QM}(T)$, call it X.

The next step is to show that X is a fixed point of the operator P. This is true because PX is also a limit of the sequence X_n.

$$\begin{aligned}
\|PX - X_{n+1}\|^2_{\mathcal{QM}(T)} &= \|PX - PX_n\|^2_{\mathcal{QM}(T)} && (19.88)\\
&\leq DT\|X - X_n\|^2_{\mathcal{QM}(T)} && (19.89)
\end{aligned}$$

which $\rightarrow 0$ as $n \rightarrow \infty$. So PX is the limit of X_{n+1}, which means it is the limit of X_n, and, since X is also a limit of X_n and limits are unique, $PX = X$. Thus, by Lemma 212, X is a solution.

To prove uniqueness, suppose that there were another solution, Y. By Lemma 212, $PY = Y$ as well. So, with Lemma 213,

$$\begin{aligned}
\|X - Y\|^2_{\mathcal{QM}(t)} &= \|PX - PY\|^2_{\mathcal{QM}(t)} && (19.90)\\
&\leq D\int_0^t \|X - Y\|^2_{\mathcal{QM}(s)} ds && (19.91)
\end{aligned}$$

So, from Gronwall's inequality (Lemma 214), we have that $\|X - Y\|_{\mathcal{QM}(t)} \leq 0$ for all t, implying that $X(t) = Y(t)$ a.s. □

Remark: For an alternative approach, based on Euler's method (rather than Picard's), see Fristedt and Gray (1997, §33.4). It has a certain appeal, but it also involves some uglier calculations. For a side-by-side comparison of the two methods, see Lasota and Mackey (1994).

Theorem 216 *Theorem 215 also holds for multi-dimensional stochastic differential equations, provided a and b are uniformly Lipschitz in the appropriate Euclidean norms.*

PROOF: Entirely parallel to the one-dimensional case, only with more algebra. □

The conditions on the coefficients can be reduced to something like "locally Lipschitz up to a stopping time", but it does not seem profitable to pursue this here. See Rogers and Williams (2000, Ch. V, Sec. 12).

19.4 Brownian Motion, the Langevin Equation, and Ornstein-Uhlenbeck Processes

The Wiener process is not a realistic model of Brownian motion, because it implies that Brownian particles do not have well-defined velocities, which is absurd. Setting up a (somewhat) more realistic model will eliminate this absurdity, and illustrate how SDEs can be used as models. I will first need to summarize classical mechanics in one paragraph.

Classical mechanics starts with Newton's laws of motion. The zeroth law, implicit in everything, is that the laws of nature are differential equations in position variables with respect to time. The first law says that they are not first-order differential equations. The second law says that they are second-order differential equations. The usual trick for higher-order differential equations is to introduce supplementary variables, so that we have a higher-dimensional system of first-order differential equations. The supplementary variable here is momentum. Thus, for particle i, with mass m_i,

$$\frac{d\vec{x}_i}{dt} = \frac{\vec{p}_i}{m_i} \tag{19.92}$$

$$\frac{d\vec{p}_i}{dt} = \frac{F(\mathbf{x}, \mathbf{p}, t)}{m_i} \tag{19.93}$$

constitute the laws of motion. All the physical content comes from specifying the force function $F(\mathbf{x}, \mathbf{p}, t)$. We will consider only autonomous systems, so we do not need to deal with forces which are explicit functions of time. Newton's third law says that total momentum is conserved, when all bodies are taken into account.

Consider a large particle of (without loss of generality) mass 1, such as a pollen grain, sitting in a still fluid at thermal equilibrium. What forces act on it? One is drag. At a molecular level, this is due to the particle colliding with the molecules (mass m) of the fluid, whose average momentum is zero. This typically results in momentum being transferred from the pollen to the fluid molecules, and the amount of momentum lost by the pollen is proportional to what it had, i.e., one term in $d\vec{p}/dt$ is $-\gamma\vec{p}$. In addition, however, there will be fluctuations, which will be due to the fact that the fluid molecules are not all at rest. In fact, because the fluid is at equilibrium, the momenta of the molecules will follow a Maxwell-Boltzmann distribution,

$$f(\vec{p}_{\mathrm{molec}}) = (2\pi m k_B T)^{-3/2} e^{-\frac{1}{2}\frac{p^2_{\mathrm{molec}}}{m k_B T}}$$

where which is a zero-mean Gaussian with variance mk_BT. Tracing this through, we expect that, over short time intervals in which the pollen grain nonetheless collides with a large number of molecules, there will be a random impulse (i.e., random change in momentum) which is Gaussian, but uncorrelated over shorter sub-intervals (by the functional CLT). That is, we would like to write

$$d\vec{p} = -\gamma\vec{p}dt + DIdW \quad (19.94)$$

where D is the *diffusion constant*, I is the 3×3 identity matrix, and W of course is the standard three-dimensional Wiener process. This is known as the *Langevin equation* in the physics literature, as this model was introduced by Langevin in 1907 as a correction to Einstein's 1905 model of Brownian motion. (Of course, Langevin didn't use Wiener processes and Itô integrals, which came much later, but the spirit was the same.) If you like time-series models, you might recognize this as a continuous-time version of an mean-reverting AR(1) model, which explains why it also shows up as an interest rate model in financial theory.

We can consider each component of the Langevin equation separately, because they decouple, and solve them easily with Itô's formula:

$$d(e^{\gamma t}p) = De^{\gamma t}dW \quad (19.95)$$

$$e^{\gamma t}p(t) = p_0 + D\int_0^t e^{\gamma s}dW \quad (19.96)$$

$$p(t) = p_0 e^{-\gamma t} + D\int_0^t e^{-\gamma(t-s)}dW \quad (19.97)$$

We will see in the next chapter a general method of proving that solutions of equations like 19.94 are Markov processes; for now, you can either take that on faith, or try to prove it yourself.

Assuming p_0 is itself Gaussian, with mean 0 and variance σ^2, then (using Exercise 19.2), $\vec{p}$ always has mean zero, and the covariance is

$$\mathrm{cov}\left(\vec{p}(t), \vec{p}(s)\right) = \sigma^2 e^{-\gamma(s+t)} + \frac{D^2}{2\gamma}\left(e^{-\gamma|s-t|} - e^{-\gamma(s+t)}\right) \quad (19.98)$$

If $\sigma^2 = D^2/2\gamma$, then the covariance is a function of $|s-t|$ alone, and the process is weakly stationary. Such a solution of Eq. 19.94 is known as a *stationary*

Ornstein-Uhlenbeck process. (Ornstein and Uhlenbeck provided the Wiener processes and Itô integrals.)

Weak stationarity, and the fact that the Ornstein-Uhlenbeck process is Markovian, allow us to say that the distribution $\mathcal{N}(0, D^2/2\gamma)$ is invariant. Now, if the Brownian particle began in equilibrium, we expect its energy to have a Maxwell-Boltzmann distribution, which means that its momentum has a Gaussian distribution, and the variance is (as with the fluid molecules) $k_B T$. Thus, $k_B T = D^2/2\gamma$, or $D^2 = 2\gamma k_b T$. This is an example of what the physics literature calls a *fluctuation-dissipation relation*, since one side of the equation involves the magnitude of fluctuations (the diffusion coefficient D) and the other the response to fluctuations (the frictional damping coefficient γ). Such relationships turn out to hold quite generally at or near equilibrium, and are often summarized by the saying that "systems respond to forcing just like fluctuations". (Cf. 19.97.)

Oh, and that story I told you before about Brownian particles following Wiener processes? It's something of a lie told to children, or at least to probability theorists, but see Exercise 19.5.

For more on the physical picture of Brownian motion, fluctuation-dissipation relations, and connections to more general thermodynamic processes in and out of equilibrium, see Keizer (1987).[2]

19.5 Exercises

Exercise 19.1 *Use the notation of Section 19.1 here.*

a *Show that* $\sum_i (\Delta W(t_i))^2$ *converges on* t *(in* L_2*) as* n *grows.* Hint: *Show that the terms in the sum are IID, and that their variance shrinks sufficiently fast as* n *grows. (You will need the fourth moment of a Gaussian distribution.)*

b *If* $X(t)$ *is measurable and non-anticipating, show that*

$$\lim_n \sum_{i=0}^{2^n-1} X(t_i)(\Delta W(t_i))^2 = \int_0^t X(s)ds$$

in L_2.

Exercise 19.2 *For any fixed, non-random cadlag function* f *on* $\mathbb{R}^+$*, let* $I_f(t) = \int_0^t f(s)dW$.

a *Show that* $\mathbf{E}\left[I_f(t)\right] = 0$ *for all* t.

b *Show* $\operatorname{cov}\left(I_f(t), I_f(s)\right) = \int_0^{t \wedge s} f^2(u)du$.

[2] Be warned that he perversely writes the probability of event A conditional on event B as $\mathbb{P}(B|A)$, not $\mathbb{P}(A|B)$.

c Show that $I_f(t)$ is a Gaussian process.

Exercise 19.3 *Consider*

$$dX = \frac{1}{2}X dt + \sqrt{1+X^2}dW \tag{19.99}$$

a Show that there is a unique solution for every initial value $X(0) = x_0$.

b It happens (you do not have to show this) that, for fixed x_0, the the solution has the form $X(t) = \phi(W(t))$, where ϕ is a C^2 function. Use Itô's formula to find the first two derivatives of ϕ, and then solve the resulting second-order ODE to get ϕ.

c Verify that, with the ϕ you found in the previous part, $\phi(W(t))$ solves Eq. 19.99 with initial condition $X(0) = x_0$.

Exercise 19.4 *Let X be an Itô process given by $dX = Adt + BdW$. Use Itô's formula to prove that*

$$f(X(t)) - f(X(0)) - \int_0^t \left[A\frac{\partial f}{\partial x} + \frac{1}{2}B^2\frac{\partial^2 f}{\partial x^2}\right] dt$$

where f is an C^2 function, is a martingale.

Exercise 19.5 (Brownian Motion and the Ornstein-Uhlenbeck Process) *Consider a Brownian particle whose momentum follows a stationary Ornstein-Uhlenbeck process, in one spatial dimension (for simplicity). Assume that its initial position $x(0)$ is fixed at the origin, and then $x(t) = \int_0^t p(t)dt$. Show that as $D \to \infty$ and $D/\gamma \to 1$, the distribution of $x(t)$ converges to a standard Wiener process. Explain why this limit is a physically reasonable one.*