Chapter 20

More on Stochastic Differential Equations

Section 20.1 shows that the solutions of SDEs are diffusions, and how to find their generators. Our previous work on Feller processes and martingale problems pays off here. Some other basic properties of solutions are sketched, too.

Section 20.2 explains the "forward" and "backward" equations associated with a diffusion (or other Feller process). We get our first taste of finding invariant distributions by looking for stationary solutions of the forward equation.

Section 20.3 makes sense of the idea of white noise. This topic will be continued in the next lecture, forming one of the bridges to ergodic theory.

For the rest of this lecture, whenever I say "an SDE", I mean "an SDE satisfying the requirements of the existence and uniqueness theorem", either Theorem 215 (in one dimension) or Theorem 216 (in multiple dimensions). And when I say "a solution", I mean "a strong solution". If you are really curious about what has to be changed to accommodate weak solutions, see Rogers and Williams (2000, ch. V, sec. 16–18).

20.1 Solutions of SDEs are Diffusions

Solutions of SDEs are diffusions: i.e., continuous, homogeneous strong Markov processes.

Theorem 217 *The solution of an SDE is non-anticipating, and has a version with continuous sample paths. If $X(0) = x$ is fixed, then $X(t)$ is $\mathcal{F}_t^W$-adapted.*

PROOF: Every solution is an Itô process, so it is non-anticipating by Lemma 198. The adaptation for non-random initial conditions follows similarly. (Informally: there's nothing else for it to depend on.) In the proof of the existence of solutions, each of the successive approximations is continuous, and we bound the maximum deviation over time, so the solution must be continuous too. □

Theorem 218 *Let X_x be the process solving a one-dimensional SDE with non-random initial condition $X(0) = x$. Then X_x forms a homogeneous strong Markov family.*

PROOF: By Exercise 19.4, for every C^2 function f,

$$f(X(t))-f(X(0))-\int_0^t \left[a(X(s))\frac{\partial f}{\partial x}(X(s)) + \frac{1}{2}b^2(X(s))\frac{\partial^2 f}{\partial x^2}(X(s))\right] ds \quad (20.1)$$

is a martingale. Hence, for every x_0, there is a unique, continuous solution to the martingale problem with operator $G = a(x)\frac{\partial}{\partial x} + \frac{1}{2}b^2(x)\frac{\partial^2}{\partial x^2}$ and function class $\mathcal{D} = C^2$. Since the process is continuous, it is also cadlag. Therefore, by Theorem 137, X is a homogeneous strong Markov family, whose generator equals G on C^2. □

Similarly, for a multi-dimensional SDE, where a is a vector and b is a matrix, the generator extends[1] $a_i(x)\partial_i + \frac{1}{2}(bb^T)_{ij}(x)\partial^2_{ij}$. Notice that the coefficients are *outside* the differential operators.

Corollary 219 *Solutions of SDEs are diffusions.*

PROOF: Obvious from Theorem 218, continuity, and Definition 177. □

Remark: To see what it is like to try to prove this without using our more general approach, read pp. 103–114 in Øksendal (1995).

Theorem 220 *Solutions of SDEs are Feller processes.*

PROOF: We need to show that (i) for every $t > 0$, $X_y(t) \xrightarrow{d} X_x(t)$ as $y \to x$, and (ii) $X_x(t) \xrightarrow{P} x$ as $t \to 0$. But, since solutions are a.s. continuous, $X_x(t) \to x$ with probability 1, automatically implying convergence in probability, so (ii) is automatic.

[1] Here, and elsewhere, I am going to freely use the Einstein conventions for vector calculus: repeated indices in a term indicate that you should sum over those indices, ∂_i abbreviates $\frac{\partial}{\partial x_i}$, ∂^2_{ij} means $\frac{\partial^2}{\partial x_i \partial x_j}$, etc. Also, $\partial_t \equiv \frac{\partial}{\partial t}$.

To get (i), prove convergence in mean square (i.e. in L_2), which implies convergence in distribution.

$$\begin{aligned}
&\mathbf{E}\left[|X_x(t) - X_y(t)|^2\right] & (20.2)\\
&= \mathbf{E}\left[\left|x - y + \int_0^t a(X_x(s)) - a(X_y(s))ds + \int_0^t b(X_x(s)) - b(X_y(s))dW\right|^2\right] \\
&\leq |x-y|^2 + \mathbf{E}\left[\left|\int_0^t a(X_x(s)) - a(X_y(s))ds\right|^2\right] & (20.3)\\
&\quad + \mathbf{E}\left[\left|\int_0^t b(X_x(s)) - b(X_y(s))dW\right|^2\right] \\
&= |x-y|^2 + \mathbf{E}\left[\left|\int_0^t a(X_x(s)) - a(X_y(s))ds\right|^2\right] & (20.4)\\
&\quad + \int_0^t \mathbf{E}\left[|b(X_x(s)) - b(X_y(s))|^2\right] ds \\
&\leq |x-y|^2 + K\int_0^t \mathbf{E}\left[|X_x(s) - X_y(s)|^2\right] ds & (20.5)
\end{aligned}$$

for some $K \geq 0$, using the Lipschitz properties of a and b. So, by Gronwall's Inequality (Lemma 214),

$$\mathbf{E}\left[|X_x(t) - X_y(t)|^2\right] \leq |x-y|^2 e^{Kt} \tag{20.6}$$

This clearly goes to zero as $y \to x$, so $X_y(t) \to X_x(t)$ in L_2, which implies convergence in distribution. □

Corollary 221 *For a given SDE, convergence in distribution of the initial condition implies convergence in distribution of the trajectories: if* $Y \stackrel{d}{\to} X_0$*, then* $X_Y \stackrel{d}{\to} X_{X_0}$*.*

PROOF: For every initial condition, the generator of the semi-group is the same (Theorem 218, proof). Since the process is Feller for every initial condition (Theorem 220), and a Feller semi-group is determined by its generator (Theorem 153), the process has the same evolution operator for every initial condition. Hence, condition (ii) of Theorem 170 holds. This implies condition (iv) of that theorem, which is the stated convergence. □

20.2 Forward and Backward Equations

You will often seen probabilists, and applied stochastics people, write about "forward" and "backward" equations for Markov processes, sometimes with the eponym "Kolmogorov" attached. We have already seen a version of the "backward" equation for Markov processes, with semi-group K_t and generator G, in Theorem 125:

$$\partial_t K_t f(x) = G K_t f(x) \tag{20.7}$$

Let's unpack this a little, which will help see where the "backwards" comes from. First, remember that the operators K_t are really just conditional expectation:

$$\partial_t \mathbf{E}\left[f(X_t)|X_0 = x\right] = G\mathbf{E}\left[f(X_t)|X_0 = x\right] \tag{20.8}$$

Next, turn the expectations into integrals with respect to the transition probability kernels:

$$\partial_t \int \mu_t(x, dy) f(y) = G \int \mu_t(x, dy) f(y) \tag{20.9}$$

Finally, assume that there is some reference measure $\lambda \gg \mu_t(x, \cdot)$, for all $t \in T$ and $x \in \Xi$. Denote the correspond transition densities by $\kappa_t(x, y)$.

$$\partial_t \int d\lambda \kappa_t(x, y) f(y) = G \int d\lambda \kappa_t(x, y) f(y) \tag{20.10}$$

$$\int d\lambda f(y) \partial_t \kappa_t(x, y) = \int d\lambda f(y) G\kappa_t(x, y) \tag{20.11}$$

$$\int d\lambda f(y) \left[\partial_t \kappa_t(x, y) - G\kappa_t(x, y)\right] = 0 \tag{20.12}$$

Since this holds for arbitrary nice test functions f,

$$\partial_t \kappa_t(x, y) = G\kappa_t(x, y) \tag{20.13}$$

The operator G alters the way a function depends on x, the *initial* state. That is, this equation is about how the transition density κ depends on the starting point, "backwards" in time. Generally, we're in a position to know $\kappa_0(x, y) = \delta(x - y)$; what we want, rather, is $\kappa_t(x, y)$ for some positive value of t. To get this, we need the "forward" equation.

We obtain this from Lemma 122, which asserts that $GK_t = K_t G$.

$$\partial_t \int d\lambda \kappa_t(x, y) f(y) = K_t G f(x) \tag{20.14}$$

$$= \int d\lambda \kappa_t(x, y) G f(y) \tag{20.15}$$

Notice that here, G is altering the dependence on the y coordinate, i.e. the state at time t, not the initial state at time 0. Writing the adjoint[2] operator as $G^\dagger$,

$$\partial_t \int d\lambda \kappa_t(x, y) f(y) = \int d\lambda G^\dagger \kappa_t(x, y) f(y) \tag{20.16}$$

$$\partial_t \kappa_t(x, y) = G^\dagger \kappa_t(x, y) \tag{20.17}$$

[2] Recall that, in a vector space with an inner product, such as L_2, the adjoint of an operator A is another operator, defined through $\langle f, Ag\rangle = \langle A^\dagger f, g\rangle$. Further recall that L_2 is an inner-product space, where $\langle f, g\rangle = \mathbf{E}\left[f(X)g(X)\right]$.

N.B., $G^{\dagger}$ is acting on the y-dependence of the transition density, i.e., it says how the probability density is going to change *going forward from t*.

In the physics literature, this is called the Fokker-Planck equation, because Fokker and Planck (independently, so far as I know) discovered it, at least in the special case of Langevin-type equations, in 1913, about 20 years before Kolmogorov's work on Markov processes. Notice that, writing ν_t for the distribution of X_t, $\nu_t = \nu_0 \mu_t$. Assuming ν_t has density ρ_t w.r.t. λ, one can get, by integrating the forward equation over space,

$$\partial_t \rho_t(x) = G^{\dagger} \rho_t(x) \tag{20.18}$$

and this, too, is sometimes called the "Fokker-Planck equation".

We saw, in the last section, that a diffusion process solving an equation with drift terms $a_i(x)$ and diffusion terms $b_{ij}(x)$ has the generator

$$Gf(x) = a_i(x)\partial_i f(x) + \frac{1}{2}(bb^T)_{ij}(x)\partial^2_{ij} f(x) \tag{20.19}$$

You can show — it's an exercise in vector calculus, integration by parts, etc. — that the adjoint to G is the differential operator

$$G^{\dagger} f(x) = -\partial_i a_i(x) f(x) + \frac{1}{2}\partial^2_{ij}(bb^T)_{ij}(x) f(x) \tag{20.20}$$

Notice that the space-dependence of the SDE's coefficients now appears *inside* the derivatives. Of course, if a and b are independent of x, then they simply pull outside the derivatives, giving us, in that special case,

$$G^{\dagger} f(x) = -a_i \partial_i f(x) + \frac{1}{2}(bb^T)_{ij}\partial^2_{ij} f(x) \tag{20.21}$$

Let's interpret this physically, imagining a large population of independent tracer particles wandering around the state space Ξ, following independent copies of the diffusion process. The second derivative term is easy: diffusion tends to smooth out the probability density, taking probability mass away from maxima (where $f'' < 0$) and adding it to minima. (Remember that bb^T is positive semi-definite.) If a_i is positive, then particles tend to move in the positive direction along the i^{th} axis. If $\partial_i \rho$ is also positive, this means that, on average, the point x sends more particles up along the axis than wander down, against the gradient, so the density at x will tend to decline.

Example 222 (Wiener process, heat equation) *Notice that (for diffusions produced by SDEs) $G^{\dagger} = G$ when $a = 0$ and b is constant over the state space. This is the case with Wiener processes, where $G = G^{\dagger} = \frac{1}{2}\nabla^2$. Thus, the heat equation holds both for the evolution of observable functions of the Wiener process, and for the evolution of the Wiener process's density. You should convince yourself that there is no non-negative integrable ρ such that $G\rho(x) = 0$.*

Example 223 (Ornstein-Uhlenbeck process) *For the one-dimensional Ornstein-Uhlenbeck process, the generator may be read off from the Langevin equation,*

$$Gf(p) = -\gamma p \partial_p f(p) + \frac{1}{2} D^2 \partial^2_{pp} f(p)$$

and the Fokker-Planck equation becomes

$$\partial_t \rho(p) = \gamma \partial_p (p\rho(p)) + D^2 \frac{1}{2} \partial^2_{pp} f(p)$$

It's easily checked that $\rho(p) = \mathcal{N}(0, D^2/2\gamma)$ gives $\partial_t \rho = 0$. That is, the long-run invariant distribution can be found as a stationary solution of the Fokker-Planck equation. See also Exercise 20.1.

20.3 White Noise

Scientists and engineers are often uncomfortable with the SDEs in the way probabilists write them, because they want to divide through by dt and have the result mean something. The trouble, of course, is that dW/dt does not, in any ordinary sense, exist. They, however, are often happier ignoring this inconvenient fact, and talking about "white noise" as what dW/dt ought to be. This is not totally crazy. Rather, one can define $\xi \equiv dW/dt$ as a *generalized* derivative, one whose value at any given time is a random real linear functional, rather than a random real number. Consequently, it only really makes sense in integral expressions (like the solutions of SDEs!), but it can, in many ways, be formally manipulated like an ordinary function.

One way to begin to make sense of this is to start with a standard Wiener process $W(t)$, and a C^1 non-random function $u(t)$, and to use integration by parts:

$$\frac{d}{dt}(uW) = u\frac{dW}{dt} + \frac{du}{dt}W \quad (20.22)$$

$$= u(t)\xi(t) + \dot{u}(t)W(t) \quad (20.23)$$

$$\int_0^t \frac{d}{dt}(uW)ds = \int_0^t \dot{u}(s)W(s) + u(s)\xi(s)ds \quad (20.24)$$

$$u(t)W(t) - u(0)W(0) = \int_0^t \dot{u}(s)W(s)ds + \int_0^t u(s)\xi(s)ds \quad (20.25)$$

$$\int_0^t u(s)\xi(s)ds \equiv u(t)W(t) - \int_0^t \dot{u}(s)W(s)ds \quad (20.26)$$

We can take the last line to *define* ξ, and time-integrals within which it appears. Notice that the terms on the RHS are well-defined *without* the Itô calculus: one is just a product of two measurable random variables, and the other is the time-integral of a continuous random function. With this definition, we can establish some properties of ξ.

Proposition 224 *$\xi(t)$ is a linear functional:*

$$\int_0^t (a_1u_1(s) + a_2u_2(s))\xi(s)ds = a_1\int_0^t u_1(s)\xi(s)ds + a_2\int_0^t u_2(s)\xi(s)ds \quad (20.27)$$

PROOF:

$$\int_0^t (a_1u_1(s) + a_2u_2(s))\xi(s)ds \quad (20.28)$$
$$= (a_1u_1(t) + a_2u_2(t))W(t) - \int_0^t (a_1\dot{u}_1(s) + a_2\dot{u}_2(s))W(s)ds$$
$$= a_1\int_0^t u_1(s)\xi(s)ds + a_2\int_0^t u_2(s)\xi(s)ds \quad (20.29)$$

□

Proposition 225 *For all t, $\mathbf{E}[\xi(t)] = 0$*

PROOF:

$$\int_0^t u(s)\mathbf{E}[\xi(s)]\,ds = \mathbf{E}\left[\int_0^t u(s)\xi(s)ds\right] \quad (20.30)$$
$$= \mathbf{E}\left[u(t)W(t) - \int_0^t \dot{u}(s)W(s)ds\right] \quad (20.31)$$
$$= \mathbf{E}[u(t)W(t)] - \int_0^t \dot{u}(s)\mathbf{E}[W(t)]\,ds \quad (20.32)$$
$$= 0 - 0 = 0 \quad (20.33)$$

Proposition 226 *For all $u \in C^1$, $\int_0^t u(s)\xi(s)ds = \int_0^t u(s)dW$.*

PROOF: Apply Itô's formula to the function $f(t,W) = u(t)W(t)$:

$$d(uW) = W(t)\dot{u}(t)dt + u(t)dW \quad (20.34)$$
$$u(t)W(t) = \int_0^t \dot{u}(s)W(s)ds + \int_0^t u(t)dW \quad (20.35)$$
$$\int_0^t u(t)dW = u(t)W(t) - \int_0^t \dot{u}(s)W(s)ds \quad (20.36)$$
$$= \int_0^t u(s)\xi(s)ds \quad (20.37)$$

□

This can be used to extend the definition of white-noise integrals to any Itô-integrable process.

Proposition 227 *ξ has delta-function covariance:* $\text{cov}(\xi(t_1), \xi(t_2)) = \delta(t_1 - t_2)$.

PROOF: Since $\mathbf{E}[\xi(t)] = 0$, we just need to show that $\mathbf{E}[\xi(t_1)\xi(t_2)] = \delta(t_1 - t_2)$. Remember (Eq. 17.14 on p. 95) that $\mathbf{E}[W(t_1)W(t_2)] = t_1 \wedge t_2$.

$$\int_0^t \int_0^t u(t_1)u(t_2)\mathbf{E}[\xi(t_1)\xi(t_2)]\,dt_1 dt_2 \tag{20.38}$$

$$= \mathbf{E}\left[\int_0^t u(t_1)\xi(t_1)dt_1 \int_0^t u(t_2)\xi(t_2)dt_2\right] \tag{20.39}$$

$$= \mathbf{E}\left[\left(\int_0^t u(t_1)\xi(t_1)dt_1\right)^2\right] \tag{20.40}$$

$$= \int_0^t \mathbf{E}\left[u^2(t_1)\right] dt_1 = \int_0^t u^2(t_1)dt_1 \tag{20.41}$$

using the preceding proposition, the Itô isometry, and the fact that u is non-random. But

$$\int_0^t \int_0^t u(t_1)u(t_2)\delta(t_1 - t_2)dt_1 dt_2 = \int_0^t u^2(t_1)dt_1 \tag{20.42}$$

so $\delta(t_1 - t_2) = \mathbf{E}[\xi(t_1)\xi(t_2)] = \mathrm{cov}(\xi(t_1), \xi(t_2))$. □

Proposition 228 *ξ is weakly stationary.*

PROOF: Its mean is independent of time, and its covariance depends only on $|t_1 - t_2|$, so it satisfies Definition 50. □

Proposition 229 *ξ is Gaussian, and hence strongly stationary.*

PROOF: To show that it is Gaussian, use Exercise 19.2. Strong stationarity follows from weak stationarity (Proposition 228) and the fact that it is Gaussian. □

20.4 Exercises

Exercise 20.1 *A conservative force is one derived from an external potential, i.e., there is a function $\phi(x)$ giving energy, and $F(x) = -d\phi/dx$. The equations of motion for a body subject to a conservative force, drag, and noise read*

$$dx = \frac{p}{m}dt \tag{20.43}$$

$$dp = -\gamma p dt + F(x)dt + \sigma dW \tag{20.44}$$

a *Find the corresponding forward (Fokker-Planck) equation.*

b *Find a stationary density for this equation, at least up to normalization constants. Hint: use separation of variables, i.e., $\rho(x, p) = u(x)v(p)$. You should be able to find the normalizing constant for the momentum density $v(p)$, but not for the position density $u(x)$. (Its general form should however be familiar from theoretical statistics: what is it?)*

c *Show that your stationary solution reduces to that of the Ornstein-Uhlenbeck process, if* $F(x) = 0$.