Stochastic Calculus, Courant Institute, Fall 2018
http://www.math.nyu.edu/faculty/goodman/teaching/StochCalc2018/index.html
Jonathan Goodman, October, 2018

Lesson 5, SDE and PDE

Revised, corrected and extended

1 Introduction

There is a close relation between stochastic differential equations (SDE) and partial differential equations (PDE). We learn about SDE paths using PDE solutions. We learn about PDE solutions by finding a related SDE. Computer solution of the PDE is a good way to calculate expected values of quantities related to an SDE.

Value functions and probability densities are two major SDE quantities that satisfy PDEs. Let X_t be a diffusion process with infinitesimal mean $a(x)$ and infinitesimal variance $v(x)$. Let $u(x,t)$ be the PDF of X_t. Then

$$\partial_t u = \frac{1}{2}\partial_x^2 \left(v(x)u(x,t) \right) - \partial_x \left(a(x)u(x,t) \right) \ . \tag{1}$$

This is called a *forward equation* because it determines $u(x,t)$ from $u(x,0)$ if $t > 0$, or $u(x,T_2)$ from $u(x,T_1)$ if $T_2 > T_1$. To be clear, the function of x at time T_2 is determined from the function at time T_1. You need to know the whole function $u(x,T_1)$ for all x to determine a single value $u(x,t_2)$. The function of x at time t is written $u(\cdot,t)$. The evolution goes forward in time from smaller t to larger t. The reverse problem, finding $u(\cdot,T_1)$ from $u(\cdot,T_2)$ with $T_1 < T_2$ is *ill posed*. Informally, an ill posed problem cannot be solved or should not be solved or cannot be solved. The formal definition is subtle and not given completely in this course.

A simple value function (there are more complicated ones) is defined for $t \leq T$ by

$$f(x,t) = \mathrm{E}[\, V(X_T) \mid X_t = x] \ . \tag{2}$$

This is the expected value of a payout function $V(X_T)$ if you start at time $t < T$ at a point x. We have seen that

$$\partial_t f(x,t) + a(x)\partial_x f(x,t) + \frac{1}{2}v(x)\partial_x^2 f(x,t) = 0 \ . \tag{3}$$

Earlier we gave a straightforward derivation using Taylor expansions and the tower property. Now we give a slick derivation using Ito's lemma. This is called a *backward equation* because $f(\cdot,T_1)$ is determined by $f(\cdot,T_2)$ for $T_2 > T_1$. The PDE moves f backward in time from T_2 to T_1. The reverse problem of going from T_1 to T_2 is ill posed.

The backward equation (3) and the forward equation (1) seem related. There is a *duality* relation connecting them. This duality allows us to derive the

forward equation from the backward equation. This is an efficient way to derive the forward equation. It seems complicated to derive it from first principles as we did for the backward equation.

There is a *Green's function*, or *transition density*, or *fundamental solution* $G(x, y, s)$ that may used to move u forward in time or to move f backward by time $s > 0$. This is the conditional density for y to x transitions in time s:

$$X_{t+s} \sim G(\cdot, y, s) \text{ if } X_t = y \, . \tag{4}$$

The forward evolution $u(\cdot, t) \to u(\cdot, t + s)$ is given in terms of G through

$$u(x, t + s) = \int_{-\infty}^{\infty} G(x, y, s) u(y, t) \, dy \, . \tag{5}$$

The backward evolution $f(\cdot, t) \to f(\cdot, t - s)$ satisfies

$$f(y, t - s) = \int_{-\infty}^{\infty} G(x, y, s) f(x, t) \, dx \, . \tag{6}$$

The *qualitative* properties of the evolution (smoothness, possible growth in time) are consequences of properties of G.

The backward equation (3) may be written in a more abstract way as

$$\partial_t f + \mathcal{L} f = 0 \, . \tag{7}$$

Here, $\mathcal{L}$ represents an *operator* that is the *generator* of the process X_t. An operator is a function of a function. That means that if $g(x)$ is a function of x, then $h = \mathcal{L} g$ is another function of x. The $\mathcal{L}$ that appears in (7) is a *differential operator* because it involves derivatives of the function it operates on. Specifically,

$$h = \mathcal{L} g \text{ has } h(x) = \frac{1}{2} v(x) \partial_x^2 g(x) + a(x) \partial_x g(x) \, . \tag{8}$$

The generator is a convenient way to define a stochastic process. Most of the PDEs related to the process involve the generator in some way. For example, the forward equation may be written as

$$\partial_t u = \mathcal{L}^* u \, . \tag{9}$$

The operator

$$\mathcal{L}^* u = \frac{1}{2} \partial_x^2 \left(v(x) u \right) - \partial_x \left(a(x) u \right) \, .$$

is the *adjoint* of the generator.

The forward and backward evolution formulas (7) and (9) may be thought of as evolution operators. These are $\mathcal{G}(s)$ for the backward evolution (7) and $\mathcal{G}^*(s)$ for the forward evolution (9). The *integral operator* formula for backward evolution is written abstractly as

$$f(\cdot, t - s) = \mathcal{G}(s) f(\cdot, t) \, . \tag{10}$$

The forward equation is written abstractly as

$$u(\cdot, t+s) = \mathcal{G}^*(s)u(\cdot, t) \; . \tag{11}$$

It may seem odd that the function of a function is written $\mathcal{L}g$ instead of normal function notation such as $\mathcal{L}(g)$. This is because $\mathcal{L}$ is linear. *Linear* means that $\mathcal{L}(g_1 + g_2) = \mathcal{L}g_1 + \mathcal{L}g_2$ and $\mathcal{L}(cg) = c\mathcal{L}g$. Linear operators on functions are similar to linear transformations or matrices operating on vectors and are written in matrix/vector notation. For example, $\mathcal{L}(g_1 + g_2)$ is the operator $\mathcal{L}$ applied to the function $g_1 + g_2$. The letter "L" (the fancy $\mathcal{L}$) is probably for "linear".

Like linear transformations or matrices, linear operators can be "multiplied" (or *composed*). The backward and forward evolution operators form a family indexed by the positive parameter s. They satisfy the relation, called the *semi-group property*,

$$\mathcal{G}(s_1 + s_2) = \mathcal{G}(s_1)\,\mathcal{G}(s_2) \; . \tag{12}$$

As with linear transformations, you compose operators by first doing one then the other. Thus

$$[\mathcal{G}(s_1)\,\mathcal{G}(s_2)]\, f(\cdot, t) = \mathcal{G}(s_1)\,[\,\mathcal{G}(s_2)f(\cdot, t)\,] \; .$$

The right side is the operator $\mathcal{G}(s_1)$ applied to the function $\mathcal{G}(s_2)f(\cdot, t)$. The left side is the composite operator $\mathcal{G}(s_1)\mathcal{G}(s_2)$ applied to the function $f(\cdot, t)$. The semigroup relation (12) is natural. The left side evolves the value function backward in time by $s_1 + s_2$. The right side evolves f backward first by s_2 and then by s_1. The "G" (the fancy $\mathcal{G}$) is for "Green" or for "group". A family of operators forms a *group* (a mathematical definition) if they are defined for positive and negative s and if (12) is satisfied regardless of the signs of s_1 and s_2. A *semi-group* is a family defined for positive s that satisfied the semi-group relation (12) for positive s_1 and s_2.

There are many equations related to basic forward (1) and backward (3) equations. Most of these include $\mathcal{L}$ or $\mathcal{L}^*$ together with more terms that represent more complicated expectations (for value functions) or dynamics (for probabilities). A *splitting* principle is that a more complicated situation is modeled by adding one term to the PDE for each new feature of the model. "Splitting" refers to a way of thinking of the features of a model acting one at a time, taking turns that last for time dt. For example, the forward equation (1) is often called an *advection diffusion* equation. The first derivative term $-\partial_x(av)$ represents *advection*, which means being carried (advected) by a velocity field $a(x)$. If the particle is just advected and has no noise, the probability density evolved according to the pure advection equation

$$\partial_t u + \partial_x(a(x)u) = 0 \; .$$

If a particle has zero drift, then it satisfies a pure diffusion equation

$$\partial_t u = \frac{1}{2}\partial_t^2(v(x)u) \; .$$

The forward equation has terms for advection and diffusion.

2 Basic equations

If you already know the backward equation, you can "derive" it using Ito's lemma. Suppose X_t satisfies the SDE

$$dX_t = a(X)dt + b(X_t)dW_t \ . \tag{13}$$

This gives the correct infinitesimal variance if $v(x) = b^2(x)$. Suppose $f(x,t)$ satisfies (3). Ito's lemma gives

$$df(X_t,t) = \partial_t f dt + \partial_x f dX + \frac{1}{2}v(x)\partial_x^2 f dt \ .$$

We substitute dX from the SDE (13) and get

$$df(X_t,t) = \left(\partial_t f dt + a(x)\partial_x f + \frac{1}{2}v(x)\partial_x^2 f\right)dt + b(X_t)dW_t \ .$$

If f satisfies the backward equation the stuff in parentheses multiplying dt is zero. The integral form of Ito's lemma gives

$$V(X_T) - f(X_t,t) = \int_t^T b(X_s)dW_s \ .$$

Take the expected value conditional on knowing $X_t = x$, or conditional on the whole path $X_{[0,t]}$, and you get (the Ito integral of a martingale is a martingale)

$$\mathrm{E}[\,V(X_T) \mid \mathcal{F}_t] - f(X_t,t) = \mathrm{E}\left[\int_t^T b(X_s)\,dW_s \mid \mathcal{F}_t\right] = 0$$
$$\mathrm{E}[\,V(X_T) \mid X_t = x] = f(x,t) \ .$$

This shows that if you solve the backward equation (3) with final condition $f(x,T) = V(x)$, then the solution is the value function.

Here is a derivation of the forward equation from the backward equation. This argument is even less rigorous than other arguments in this course, though it can be made rigorous using ideas from functional analysis that would be a big distraction for most students. The idea was used already in Lesson 1, but here is a version said a little more abstractly.

The notion of *adjoint operator* for functions is like the notion of transpose matrix in linear algebra. There would be a conflict of notation, using x for a real number and also using x for a vector in n dimensions. Therefore, we write $\vec{x} \in \mathbb{R}^n$ for a vector. The *inner product* of $\vec{x}$ and $\vec{y}$ can be written

$$\langle \vec{x}, \vec{y} \rangle = \sum_{i=1}^{n} x_i y_i \ .$$

The inner product is "linear in each factor". For example,

$$\langle [\vec{x}_1 + \vec{x}_2] \, , \vec{y} \rangle = \langle \vec{x}_1, \vec{y} \rangle + \langle \vec{x}_2, \vec{y} \rangle \ .$$

Also, if $\langle \vec{x}, \vec{y} \rangle = 0$ for "enough" vectors $\vec{y}$, then $\vec{x} = 0$. It could be all $\vec{y} \in \mathbb{R}^n$, or it could be just n linearly independent vectors.

If A is an $n \times n$ real matrix, the *adjoint* (also called *transpose* if A is real) is A^*. The entries are

$$(A^*)_{ij} = A_{ji} \; . \tag{14}$$

Abstractly, we say that A^* is the adjoint of A if, for all $\vec{x}$ and $\vec{y}$,

$$\langle \vec{x}, A\vec{y} \rangle = \langle A^* \vec{x}, \vec{y} \rangle \; . \tag{15}$$

You should check that the abstract definition (15) and the concrete definition (14) are equivalent.

For functions, we can define the inner product as

$$\langle u, g \rangle = \int_{-\infty}^{\infty} u(x) g(x) \, dx \; . \tag{16}$$

This inner product is also linear in each factor. There are many theorems that say that if $\langle u, g \rangle = 0$ for "enough" functions g, then $u = 0$. The proper theorem for the present purpose (deriving the forward equation from the backward equation) is more technical that is appropriate. If $\mathcal{L}$ is an operator, we say $\mathcal{L}^*$ is the adjoint if, for "every" pair of functions u and g,

$$\langle \mathcal{L}^* u, g \rangle = \langle u, \mathcal{L} g \rangle \; . \tag{17}$$

We can check that the operator $\mathcal{L}^*$ in the forward equation (9) is the adjoint in this sense of the operator $\mathcal{L}$ in the backward equation (7). For this, we assume that g and u have two derivatives and that one or the other (or both) goes to zero as $|x| \to \infty$ fast enough so that the "boundary terms" in integration by parts are zero. The calculation is

$$\begin{aligned}
\langle u, \mathcal{L} g \rangle &= \int_{-\infty}^{\infty} u(x) \left[\frac{1}{2} v(x) \partial_x^2 g(x) + a(x) \partial_x g(x) \right] \, dx \\
&= \frac{1}{2} \int_{-\infty}^{\infty} u(x) v(x) \partial_x^2 g(x) \, dx + \int_{-\infty}^{\infty} u(x) a(x) \partial_x g(x) \, dx \\
&= \frac{1}{2} \int_{-\infty}^{\infty} \partial_x^2 \left[v(x) u(x) \right] g(x) \, dx - \int_{-\infty}^{\infty} \partial_x \left[a(x) u(x) \right] g(x) \, dx \\
&= \int_{-\infty}^{\infty} \left[\frac{1}{2} \partial_x^2 \left(v(x) u(x) \right) - \partial_x \left(a(x) u(x) \right) \right] g(x) \, dx \\
&= \langle \mathcal{L}^* u, g \rangle \; .
\end{aligned}$$

Note that $\mathcal{L}$ has the multiplication outside the derivatives, so v and a are not differentiated, while $\mathcal{L}^*$ has the coefficients v and a inside both derivatives. Also, the advection term in $\mathcal{L}^*$, which is $-\partial_x(au)$, has the opposite sign from the advection term in $\mathcal{L}$, which is $a \partial_x g$. This implies that if "stuff" is advected to the right in the forward equation, then stuff is advected to the left in the

backward equation. We will see that this is related to the direction time is supposed to move.

Suppose $X_t \sim u(x,t)$, then if $t < T$,

$$\mathrm{E}[\, V(X_T)] = \int_{-\infty}^{\infty} u(x,t)\, f(x,t)\, dx \; .$$

The left side is the unconditional expectation. The right side is the integral of the expectation conditional on $X_t = x$, integrated over all possible x values and weighted with the probability density for $X_t = x$. It is important here that the left side is independent of t (because t does not appear on the left), while t does appear on the right. Therefore, the variation of u and f on the right must be coordinated in some way so that the integral does not change as t changes. We calculate the relation between u variation and f variation. We assume that f satisfies the backward equation (7) and see what this implies about the time dependence of u:

$$\begin{aligned}
0 &= \frac{d}{dt} \langle u(\cdot,t), f(\cdot,t) \rangle \\
&= \langle \partial_t u(\cdot,t), f(\cdot,t) \rangle + \langle u(\cdot,t), \partial_t f(\cdot,t) \rangle \\
&= \langle \partial_t u(\cdot,t), f(\cdot,t) \rangle - \langle u(\cdot,t), \mathcal{L} f(\cdot,t) \rangle \\
&= \langle \partial_t u(\cdot,t), f(\cdot,t) \rangle - \langle \mathcal{L}^* u(\cdot,t), f(\cdot,t) \rangle \\
0 &= \langle \left[\partial_t u(\cdot,t) - \mathcal{L}^* u(\cdot,t) \right], f(\cdot,t) \rangle \; .
\end{aligned}$$

This is supposed to be true for "every" function f. This implies that $\partial_t u(\cdot,t) - \mathcal{L}^* u(\cdot,t) = 0$, which is the forward equation (9).

2.1 Ornstein Uhlenbeck

The Ornstein Uhlenbeck SDE is

$$dX = -\gamma X dt + \sigma dW \; . \tag{18}$$

Its backward equation is

$$\partial_t f - \gamma x \partial_x f + \frac{\sigma^2}{2} \partial_x^2 f = 0 \; . \tag{19}$$

The generator is often written as a differential operator in terms of the basic derivative operators ∂_x and ∂_x^2:

$$\mathcal{L}_{\mathrm{OU}} = -\gamma \partial_x + \frac{\sigma^2}{2} \partial_x^2 \; . \tag{20}$$

This means that for any function g, the generator "acts on" g by

$$\mathcal{L}_{\mathrm{OU}} g = -\gamma \partial_x g + \frac{\sigma^2}{2} \partial_x^2 g \; .$$

It is harder to put differential operators in parentheses, so the adjoint is often written in terms of its action on a function, such as

$$\mathcal{L}^*_{\rm OU} u = \gamma \partial_x \left(x u \right) + \frac{\sigma^2}{2} \partial_x^2 u \; . \tag{21}$$

The forward equation is

$$\partial_t u = \mathcal{L}^*_{\rm OU} u = \gamma \partial_x \left(x u \right) + \frac{\sigma^2}{2} \partial_x^2 u \; . \tag{22}$$

There are PDE based methods for finding solutions of the forward equation. One is the *ansatz* method. An *ansatz* is a guess of a functional form that might satisfy the PDE. The ansatz has unknown functions or parameters that must be chosen correctly in order that the target PDE is satisfied. If it works, you find a solution. If it doesn't work, you try to make a getter guess ("refine the ansatz"), or give up. We believe that the OU process is Gaussian with a time varying mean and variance. This suggests an ansatz

$$u(x,t) = \frac{1}{\sqrt{2\pi v(t)}} e^{-\frac{(x-m(t))^2}{2v(t)}} \; . \tag{23}$$

We calculate[1] the first and second derivatives with respect to x:

$$\begin{aligned}
& \frac{1}{\sqrt{2\pi}} v(t)^{-\frac{1}{2}} e^{-\frac{(x-m(t))^2}{2v(t)}} \\
\xrightarrow{\partial_x} \; & -\frac{1}{\sqrt{2\pi}} \frac{x-m(t)}{v(t)} v(t)^{-\frac{1}{2}} e^{-\frac{(x-m(t))^2}{2v(t)}} = -\frac{1}{\sqrt{2\pi}} (x-m(t)) \, v(t)^{-\frac{3}{2}} e^{-\frac{(x-m(t))^2}{2v(t)}} \\
\xrightarrow{\partial_x} \; & \frac{1}{\sqrt{2\pi}} \left[(x-m(t))^2 \, v(t)^{-\frac{5}{2}} - v(t)^{-\frac{3}{2}} \right] e^{-\frac{(x-m(t))^2}{2v(t)}} \; .
\end{aligned}$$

For the advection term, we calculate

$$x u = (x - m(t)) \, u + m(t) u \; .$$

We have already calculated the derivative of the second term, and the first term gives

$$\begin{aligned}
& \frac{1}{\sqrt{2\pi}} (x-m(t)) \, v(t)^{-\frac{1}{2}} e^{-\frac{(x-m(t))^2}{2v(t)}} \\
\xrightarrow{\partial_x} \; & \frac{1}{\sqrt{2\pi}} \left[-v(t)^{-\frac{1}{2}} - (x-m(t))^2 \, v(t)^{-\frac{3}{2}} \right] e^{-\frac{(x-m(t))^2}{2v(t)}} \; .
\end{aligned}$$

The time derivative is the most "fun". We use a dot for time derivatives of v

[1] The ansatz method taxes one's ability and patience with algebra.

and m, as $\dot{v} = \frac{dv}{dt}$.

$$\begin{aligned} & \frac{1}{\sqrt{2\pi}} v(t)^{-\frac{1}{2}} e^{-\frac{(x-m(t))^2}{2v(t)}} \\ \xrightarrow{\partial_t} & \frac{1}{\sqrt{2\pi}} \left[-\frac{1}{2} v(t)^{-\frac{3}{2}} \dot{v}(t) + v(t)^{-\frac{3}{2}} (x - m(t)) \dot{m}(t) + \frac{1}{2} v(t)^{-\frac{5}{2}} (x - m(t))^2 \dot{v}(t) \right] \\ & \cdot \; e^{-\frac{(x-m(t))^2}{2v(t)}} \; . \end{aligned}$$

The next step is to substitute these calculations into the forward equation. I will let you do this. Then you cancel common factors $\frac{1}{\sqrt{2\pi}}$ and $e^{\cdots}$. Then you equate coefficients of $(x-m)^2$, coefficients of $(x-m)$ and the "constant" terms (independent of x). The resulting equations are

$$\frac{1}{2}\dot{v} = -\gamma v + \frac{1}{2}\sigma^2 \;\; (\text{ coefficients of } (x-m)^2 \;) \tag{24}$$

$$\dot{m} = -\gamma m \;\; (\text{ coefficients of } (x-m) \;) \tag{25}$$

$$\frac{1}{2}\dot{v} = -\gamma v + \frac{1}{2}\sigma^2 \;\; (\text{ coefficients of constant }) \; .$$

If all these equations are satisfied than the ansatz solution (23) satisfies the forward equation (22). The first equation (24) determines the variance at time t. You should take $v(0) = 0$ if you want the transition density/Green's function. Without solving the equation, you can see that $v_\infty = \frac{\sigma^2}{2\gamma}$ is the steady state. This is consistent with our direct calculation of the steady state variance using Ito's lemma. The second equation (25) says that the mean ignores the noise and just does what the noise free ($\sigma = 0$) equation does, it converges to zero with rate γ. The third equation is consistent with the first. If it weren't, the ansatz form would have had too few free parameters (two) to satisfy the three equations.

The variance equation has solution $v(t) = \frac{\sigma^2}{2\gamma}\left(1 - e^{-2\gamma t}\right)$. We derived this formula before from the SDE solution. The method here is possibly simpler. The behavior for small t is $v(t) = \sigma^2 t + O(t^2)$. This is the same as $v(0) = 0$ and $\dot{v}(0) = \sigma^2$, which you can see from the v equation (24) with $v = 0$. Without the stabilizing term $-\gamma x$, the OU process is just Brownian motion without drift but variance $\sigma^2 t$. This shows that the "drift term" $-\gamma x$ is irrelevant for small t. The short time (small t) behavior of the Green's function/transition density depends on the diffusion part only.

3 Probability flux, current, boundary conditions

It is useful to interpret the forward equation (1) as describing the "flow" of probability from one place to another. The *probability flux* (called *probability current* by physicists) describes the rate of flow at a point. This interpretation is based on a formula you derive by integrating the forward equation with respect

to x over an interval

$$\frac{d}{dt}\int_a^b u(x,t)\,dx = -\mathcal{F}(b,t) + \mathcal{F}(a,t)\ . \tag{26}$$

where

$$\mathcal{F}(x,t) = -\frac{1}{2}\partial_x\left(v(x)\,u(x,t)\right) + a(x)u(x,t)\ . \tag{27}$$

Physicists use J for current, and would replace $\mathcal{F}$ with $\mathcal{J}$ above. The physical interpretation is that $\mathcal{F}$ (or $\mathcal{J}$) represents the amount of flow across the point x per unit time.

It may help to imagine n (a large number) independent particles all evolving according to the stochastic dynamics (13). The number of particles inside $[a,b]$ is $N_t(a,b)$. For large n, it is approximately

$$N_t(a,b) \approx n\int_a^b u(x,t)\,dx\ .$$

According to (26), the rate of change is approximately

$$\frac{dN_t}{dt} \approx n\left[-\mathcal{F}(b,t) + \mathcal{F}(a,t)\right]\ .$$

The first term on the right represents the rate at which particles leave $[a,b]$ at $x = b$. Particle motion is random in stochastic dynamics, so particles at b cross in both directions. The flux $\mathcal{F}(b,t)$ is the net, the number (per unit time) crossing from $x < b$ to $x > b$ minus the number going the other way.

The flux/current formula (27) is the sum of a *diffusive* flux (the first term on the right) and an *advective* flux (the second term on the right). In keeping with the "splitting" point of view, the advective flux corresponds to the advective term $a(X_t)dt$ in the SDE, while the diffusive flux corresponds to the noise term $b(X_t)dW_t$. In pure advection ($v(x) = 0$), the rate of particles crossing $x = b$ is proportional to the speed, $a(x)$, and the density of particles near $x = b$, which is $u(b,t)$. The precise form of the diffusive flux is harder to justify by elementary arguments. However, if $v(x)$ is constant then the diffusive flux formula

$$\mathcal{F}_{\text{diff}} = -C\partial_x u$$

is called the *Fourier law* for heat conduction and *Fick's law* for ordinary diffusion in a uniform medium (such as the diffusion of ink die in still water). In pure diffusion, stuff moves from where it is more concentrated to where it is less concentrated. That's why the flux is proportional to the density gradient, but with the opposite sign.

Many problems involve diffusion processes with boundaries. For example, when we did hitting times, we said that $u(b,t) = 0$ if $x = b$ is an absorbing boundary. *Absorbing* means that a diffusing particle is removed the first time $X_t = b$. The condition $u(b,t) = 0$ is called an absorbing boundary condition, or a *Dirichlet* boundary condition. Other problems have simple constraints

(walls or other boundaries) that prevent a particle from crossing $x = b$. These are called *reflecting* boundaries, because a particle that hits $x = b$ reflects or bounces back. The appropriate boundary condition is that the $\mathcal{F}(b,t) = 0$. That is the condition that particles cannot go from one side of $x = b$ to the other. If $a = 0$ (pure advection), this is equivalent to $\partial_x u(b,t) = 0$. This is called a *Neumann*[2] boundary condition.

4 Other terms in the backward equation

Suppose instead of a payout $V(X_T)$, there was a time varying payout rate depending on X_t. This would be

$$R = \int_0^T r(X_t)\, dt \; . \tag{28}$$

In control theory, one often formulates a *running cost* function, where $r(x)$ is the cost rate for x. The goal is to find a control (more about this in a future Lesson) to minimize $\mathrm{E}[R]$. An appropriate value function is

$$f(x,t) = \mathrm{E}_{x,t}\left[\int_t^T r(X_t)\, dt \right] \; . \tag{29}$$

The backward equation for this f has the generator, $\mathcal{L}$ and (in the spirit of splitting) an extra term that corresponds to the running reward. When you go backward from time $t + \Delta t$ to time t you get the reward $r(X_t)$ dt. But the derivative $\partial_t f$ is in the forward direction, so this should have a minus sign. The result is

$$\partial_t f + \mathcal{L} f = -r(x) f(x,t) \; . \tag{30}$$

If we write this out more directly, it is

$$\partial_t f(x,t) + \frac{v(x)}{2} \partial_x^2 f(x,t) + a(x) \partial_x f(x,t) + r(x) f(x,t) = 0 \; .$$

It is possible to justify this using the tower property for time Δt as we did in an earlier Lesson. The final condition is $f(x,T) = 0$ because the integral in (**??**) runs from T to T and is equal to zero.

Once you conjecture the backward equation, you can prove/verify that it is true using Ito's lemma. Calculate $df(X_t,t)$, integrate, take expectations, use the fact that an Ito integral with respect to Brownian motion has expectation

[2]This is pronounced in the German way. The first syllable rhymes with "toy" and the second syllable rhymes with "wand".

zero, and you get

$$\begin{aligned}
f(X_T, T) - f(X_t, t) &= \int_t^T d(f(X_s, s) \\
&= \int_t^T \left(\partial_t f(X_s, s) + \frac{v(X_s)}{2} \partial_x^2 f(X_s, s) + a(X_s) \partial_x f(X_s, s) \right) ds \\
&\quad + \int_t^T \partial_x f(X_s, s) b(X_s) \, dW_s \\
&= \int_t^T -r(X_s, s) \, ds + \int_t^T \partial_x f(X_s, s) b(X_s) dW_s \\
-f(x, t) &= \mathrm{E}_{x,t} \left[\int_t^T -r(X_s, s) ds \right] .
\end{aligned}$$

This shows that the solution to the PDE (30) is the value function (29). In this calculation it would have been less writing to use Ito's lemma in the form

$$df(X_t, t) = \left[\partial_t f(X_t, t) + \mathcal{L} f(X_t, t) \right] \, dt + \frac{v(x)}{2} \partial_x f(X_t, t) \, dW_t \, . \tag{31}$$

A path-dependent quantity of interest in finance is the expected result of a variable and stochastic interest rate. The value function is

$$f(x, t) = \mathrm{E}_{x,t} \left[e^{\int_t^T r(X_s, ds)} \right] \, . \tag{32}$$

The backward equation for this includes the generator and a term corresponding to the effect of the interest at x. The interest at x in time dt increases the expected value by a factor of $1 + r(x)dt$. As with the running reward (29), this gets a negative sign because of the direction of time. We are led to the possible backward equation

$$\partial_t f + \mathcal{L} f = -r(x) f \, . \tag{33}$$

This also can be verified using Ito's lemma. The final condition is $f(x, T) = 1$. The conditional expectation (32) is a way to express the solution of the backward equation in terms of a random process. In that context it is called the *Feynman Kac*[3] formula. Feynman in the 1940's proposed a formula for the solution of the *Schrödinger* equation as a formal integral that is called the *Feynman integral*. Mathematicians quickly pointed out that the Feynman integral doesn't make sense mathematically. But Kac showed that reasoning similar to Feynman's suggests that (32) is the solution of (33). The expectation in (32) may be thought of as a fancy kind of integral. We will return to that point in a later Lesson.

[3]For physicist Richard Feynman and later mathematician Marc Kac. Kac was Polish, though he worked much of his life at Rockefeller University in New York City. His name is pronounced "cats".

5 Other terms in the forward equation

Consider a process that "dies" with probability $r(x)dt$ in time dt if $X_t = x$. Let $u(x,t)$ be the density (the probability density or the particle density) of particles that have not died yet. In the spirit of splitting, we imaging evolving u first for time dt using the stochastic process (13) and then evolving u for time dt with just the death ("killing") process. The result is

$$\partial_t u = \mathcal{L}^* u - r(x)u \; . \tag{34}$$

In a financial application, X_t could be the value of a company and $r(x)$ could be the probability of default (stopping payments) per unit time. In a physical application, we could consider a random process with some probability of absorption.

Suppose a process switches "regimes" at random times. If it's in regime L (for left moving), then it satisfies

$$dX_t = a_L dt + \sigma dW \; .$$

If it's moving right, then it satisfies

$$dX_t = a_R dt + \sigma dW \; .$$

Presumably, $a_L < 0$ and $a_R > 0$. Suppose it switches between left and right moving "completely at random" with rates

$$\begin{aligned} \text{L} &\xrightarrow{\lambda} \text{R} \\ \text{R} &\xrightarrow{\mu} \text{L} \; . \end{aligned}$$

This means that a particle in state L goes to state R in time dt with probability λdt. From the point of view of a state L particle, this is the same as killing. Suppose the probability densities for L and R particles are $u_L(x,t)$ and $u_R(x,t)$. The dynamics are (obviously?)

$$\begin{aligned} \partial_t u_L(x,t) &= -a_L \partial_x u_L(x,t) + \frac{\sigma^2}{2}\partial_x^2 u_L(x,t) - \lambda u_L(x,t) + \mu u_R(x,t) \\ \partial_t u_R(x,t) &= -a_R \partial_x u_R(x,t) + \frac{\sigma^2}{2}\partial_x^2 u_R(x,t) + \lambda u_L(x,t) - \mu u_R(x,t) \; . \end{aligned}$$

The $-\lambda u_L(x,t)$ in the first equation represents state L particles "dying" and making a transition to state R. The same term occurs in the u_R equation with a plus sign, because particles make the transition into the R state. Similarly, the $R \to L$ transition term, which is μu_R, has a plus sign in the u_L equation and a minus sign in the u_R equation.

You can check that total probability is preserved. The probability of the particle having some x and having some state is

$$P(t) = \int_{-\infty}^{\infty} u_L(x,t)\, dx + \int_{-\infty}^{\infty} u_R(x,t)\, dx \; .$$

If you differentiate and use the equations, you see that

$$\dot{P}(t) = 0 \ .$$

There is a good chance that this would not have happened if we had modeled the process incorrectly.

Suppose you want to model a process with absorption at a boundary but with some probability less than 1. Recall that a purely reflecting boundary has zero probability flux at the boundary. To model a particle with some probability of absorption, set the flux (the absorption rate) equal to a multiple of the probability density at the absorption point.