Stochastic Calculus, Courant Institute, Fall 2021
http://www.math.nyu.edu/faculty/goodman/teaching/StochCalc2021/index.html

Week 3

Jonathan Goodman, November, 2021

1 Introduction value functions

The goal of probability model may be to compute expected values of some random variable. Financial and engineering control strategies are chosen based on expectation values of performance measures. A performance measure is a function of the random path that quantifies goodness or badness of the path, from the a specific point of view. In financial applications, this often takes the form of maximizing expected utility. In engineering, cost measures may include expected costs such as fuel or damage. If you are interested in the probability that something bad happens, you can consider the performance measure to be the *characteristic function* (also called *indicator function*) of the bad event. The performance measure is equal to one (say) if the event happens and zero if it does not.

The *value function* approach to expectation values it to consider also conditional expectations of the same or related quantities. The value function is the conditional expectation, as a function of what you're conditioning on. The conditional expectations may be related to each other by a partial differential equations (PDEs) called *backward equations*. There are different backward equations for different diffusion processes and different performance measures. Solving the PDE, which means finding all the conditional expectations, is a way to evaluate the expectation you were originally interested in. Moreover, as is explained in Week 5, backward equations may be used to design strategies that optimize performance measures. This is the *dynamic programming* (also called *Hamilton Jacobi Bellman*) approach to *optimal stochastic control*.

There is a two way relation between diffusion processes and partial differential equations. In one direction, we learn about diffusion processes by solving some associated partial differential equations called diffusion equations. In the other direction, we find solutions of diffusion equations by expressing the solution as the expected value of some quantity related to a diffusion process. The solution of the differential may be found by simulating a random process.

This class explores the relation between backward equations and diffusion process connection when the diffusion process is Brownian motion, and the PDE is a variant of the backward heat equation. The Week 4 class explains that these ideas apply to general diffusion processes (processes described by their infinitesimal mean and variance, or, equivalently, by an SDE) and general *backward diffusion equations*. In some sense, this week is just motivation for next week.

Notation. In this class, X_t will be Brownian motion. The CDF (cumulative distribution function) of a random variable with PDF $u(\cdot)$ is

$$F(x) = \Pr(\, X < x) = \int_{-\infty}^{x} u(y)\, dy \; .$$

The *standard normal* random variable Z has PDF $u(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$. Its CDF is called the *cumulative normal* distribution function, and is given b

$$\begin{aligned} N(x) &= \Pr(Z < x) \; , \quad Z \sim \mathcal{N}(0,1) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}}\, dz \; . \end{aligned} \tag{1}$$

This has the values

$$\begin{aligned} N(x) &\to 0 \text{ as } x \to -\infty \\ N(x) &\to 1 \text{ as } x \to +\infty \\ N(0) &= \frac{1}{2} \; . \end{aligned}$$

This is closely related to the *error function*

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-z^2}\, dz \; .$$

These are related by the scaling formula $N(x) = \frac{1}{2}\left(1 + \operatorname{erf}(2x)\right)$. If X is a general one component normal $X \sim \mathcal{N}(\mu, \sigma^2)$, then its CDF is

$$F(x) = N\left(\frac{x - \mu}{\sigma} \right) \; .$$

This is "the" cumulative normal (1), with the mean subtracted and scaled by the standard deviation.

2 A value function and its backward equation

Let X_t be Brownian motion. A simple value function is

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

You might be interested in the expected value of X_T^2 ($V(x) = x^2$), or the probability that $X_T > 1$ (in clumsy indicator function notation: $V(x) = \mathbf{1}_{x>1}(x)$), or the exponential ($V(x) = e^{ax}$, needed to study geometric Brownian motion). The value function measures progress toward the final value. Suppose you start at $t = 0$ and know (or specify) that $X_t = x$ at some time before T ($t < T$). What, then, is the expected value? It is given by the conditional expectation

(2). This is defined for any $t \leq T$ and for any x. We will see that this value function satisfies a PDE called the *backward heat equation*

$$\partial_t f + \frac{1}{2}\partial_x^2 f = 0 \; . \tag{3}$$

Warning: The conditional expectation (2) makes sense if $t < T$ or if $t > T$, but f satisfies the backward equation (3) only if $t < T$.

The backward equation, by itself, does not determine $f(x,t)$, even for $t < T$. For that, you also need *final conditions*

$$f(x,T) = V(x) \; . \tag{4}$$

The PDE (3) may not be obvious, but the final conditions are obvious. If you put $t = T$ in the value function definition (2), you find yourself asking what is the expected value of $V(X_T)$ conditional on $X_T = x$.

Once the final condition is given, the backward heat equation determines $f(x,t)$ for $t < T$. Roughly speaking, the PDE says that $f(x, t - \Delta t)$ is a local average of $f(\cdot, t)$ near the point x. In this way, the function $f(\cdot, t)$ determines the function at a slightly earlier time $f(\cdot, t - \Delta t)$. We can start with the final condition $f(\cdot, T) = V(\cdot)$ and work backwards in time using local averages to find $f(\cdot, t)$ for any $t < T$. The relation between $\partial_{xx}(x,t)$ and local averages may be/should be mysterious now. Hopefully, Section (**??**) and the more in-depth discussion in Week 4 will make this more clear. Section 10 presents a computing algorithm based on the Δt and local average point of view.

The conditional expectation (2) may be expressed as an integral involving the conditional probability density $G(\cdot, x, T - t)$. This G is the conditional density of X_T, conditional on $X_t = x$. We use y to represent the value of X_T, and get

$$f(x,t) = \int V(y) G(y, x, T - t)\, dt \; . \tag{5}$$

Note that the distribution of X_T depends on $X_t = x$ and the amount of time between t and T, which is $T - t$. That's why G depends on $T - t$ and not on T and t separately. The term *Green's function* comes from the way it appears in the solution formula (5). Functions that express the solution of a PDE in terms of the "data" (final condition in this case) are called "Green's functions" ("Green's theorem" is named for the same Brit). They are called *fundamental solution* because they satisfy the PDE (the backward equation in this case) with specific "data". A function that appears as G does in an integral like (5) is called an *integral kernel*. The integral kernel that occurs in the solution of the heat equation (forward or backward) is called the *heat kernel*. Lastly, as a function of y, $G(\cdot, x, T - t)$ is the probability density for transitions from $X_t = x$ to $X_T = y$, hence the name *transition density*.

We can derive the backward equation (3) from the integral representation of f (5) because there is an explicit formula for G. Representations like Gf exist for general dynamic stochastic models, but there usually isn't an explicit formula for the transition density. The difference here, $X_T - X_t$, is an increment of

Brownian motion. Such increments are Gaussian, with mean zero and variance equal to the time difference – $T - t$. If you specify that $X_t = x$, then X_T becomes (in the conditional distribution) Gaussian with mean x and variance $T - t$:

$$X_T \sim \mathcal{N}(x, T-t) \implies G(y, x, T-t) = \frac{1}{\sqrt{2\pi(T-t)}} e^{-\frac{(y-x)^2}{2(T-t)}} \; . \tag{6}$$

This puts the abstract solution formula (5) in the explicit form

$$f(x,t) = \int_{-\infty}^{\infty} V(y) \frac{1}{\sqrt{2\pi(T-t)}} e^{-\frac{(y-x)^2}{2(T-t)}} \, dy \; . \tag{7}$$

This leads to a derivation of the backward equation, as stated above. It also allows us to learn some things about the value function f that might not be directly obvious from the definition (2) or the backward equation.

We verify the PDE (3) by calculating the t and x derivatives of the integral on the right of (7). The integration domain does not depend on x or t, so the derivatives act "under the integral sign" directly on the integrand. The $V(y)$ in the integrand does not depend on x of t, which is why any value function for any V satisfies the same PDE. Thus,

$$\partial_t f(x,t) = \partial_t \int V(y) G(y,x,T-t) \, dy = \int V(y) [\, \partial_t G(y,x,T-t)] \; dy \; ,$$

and

$$\partial_x^2 f(x,t) = \int V(y) \left[\, \partial_x^2 G(y,x,T-t)\right] \; dy \; ,$$

We add these together and get

$$\partial_t f + \frac{1}{2}\partial_x^2 f = \int V(y) \left[\partial_t G(y,x,T-t) + \frac{1}{2}\partial_x^2 G(y,x,T-t) \right] \; dy \; . \tag{8}$$

This shows that f satisfies the backward heat equation if G does.

We can calculate the combination of derivatives in $[\cdots]$ of (8) by differentiating the formula (6). Here is the time derivative calculation. You have to do this slowly and carefully to get the signs and the powers of $(T-t)$ right. The first line is the product rule of differentiation. The next line is the chain rule applied to each of the terms.

$$\begin{aligned}
\partial_t \left[(2\pi(T-t))^{-\frac{1}{2}} \, e^{-\frac{(y-x)^2}{2(T-t)}} \right] &= \left[\frac{1}{\sqrt{2\pi}} \partial_t (T-t)^{-\frac{1}{2}} \right] e^{-\frac{(y-x)^2}{2(T-t)}} + \frac{1}{\sqrt{2\pi(T-t)}} \left[\partial_t e^{-\frac{(y-x)^2}{2(T-t)}} \right] \\
&= \frac{1}{2}\frac{1}{\sqrt{2\pi}} (T-t)^{-\frac{3}{2}} e^{-\frac{(y-x)^2}{2(T-t)}} - \frac{1}{\sqrt{2\pi(T-t)}} \frac{(y-x)^2}{2(T-t)^2} e^{-\frac{(y-x)^2}{2(T-t)}} \\
&= \frac{1}{2}\frac{1}{\sqrt{2\pi}} \left[\frac{1}{(T-t)^{\frac{3}{3}}} - \frac{(y-x)^2}{(T-t)^{\frac{5}{2}}} \right] e^{-\frac{(y-x)^2}{2(T-t)}} \; .
\end{aligned}$$

We have to do two ∂_x calculations, but they are not as tricky. First

$$\begin{aligned} \partial_x G(y,x,T-t) &= \frac{1}{\sqrt{2\pi(T-t)}} \frac{y-x}{T-t} e^{-\frac{(y-x)^2}{2(T-t)}} \\ &= \frac{1}{\sqrt{2\pi}} \frac{y-x}{(T-t)^{\frac{3}{2}}} e^{-\frac{(y-x)^2}{2(T-t)}} \ . \end{aligned}$$

Then

$$\partial_x^2 G(y,x,T-t) = \frac{1}{\sqrt{2\pi}} \left[\frac{-1}{(T-t)^{\frac{3}{2}}} e^{-\frac{(y-x)^2}{2(T-t)}} + \frac{(y-x)^2}{(T-t)^{\frac{5}{2}}} e^{-\frac{(y-x)^2}{2(T-t)}} \right] \ .$$

Now, compare the two results and you see that the $(T-t)^{-\frac{3}{2}}$ terms and the $(y-x)^2(T-t)^{-\frac{5}{2}}$ terms cancel in $\partial_t G + \frac{1}{2}\partial_x^2 G$. The calculation

$$\partial_t G(y,x,T-t) + \frac{1}{2}\partial_x^2 G(y,x,T-t) = 0 \tag{9}$$

implies that the value function satisfies the backward equation (3). It also shows that the transition density itself satisfies a PDE. It is called the *fundamental solution* because any other solution may be expressed in terms of this one

3 Examples

Here are a few examples of solutions to the backward equation. In some sense, they don't show off the PDE as a way to evaluate expected values. In each example, the solution may be found directly from the expected value definition of f, though the PDE might be a simpler route to the solution in some cases. The point of the SDE/PDE connection isn't really finding formulas for expectation values. More important uses are: (1) finding ways to solve a PDE using simulation, and (2) using finite differences (or other PDE methods) to find expectation values.

If you want to use the conditional expectation formula directly, it can help to denote the Brownian motion increment between t and T by some other letter, say, Z. Then, with the condition $X_t = x$, we have $X_T = x + Z$, and Z is normal with mean zero and variance $T-t$.

For example, suppose the payout function is linear: $V(x) = ax + b$, then

$$f(x,t) = \mathrm{E}[\, a(x+Z) + b] = ax + b \ .$$

If the payout is linear, then the value function is the same linear function for all t. You can verify that $f(x,t) = ax + b$ satisfies the backward equation: both $\partial_t f$ and $\partial_x^2 f$ are zero.

Next, suppose the payout is quadratic. For simplicity, take $V(x) = x^2$, then calculate, using the fact that x is not random:

$$\begin{aligned} f(x,t) &= \mathrm{E}\big[\,(x+Z)^2\big] \\ &= \mathrm{E}\big[\,x^2+2xZ+Z^2\big] \\ &= x^2+2x\mathrm{E}[\,Z]+\mathrm{E}\big[\,Z^2\big] \\ &= x^2+(T-t)\ . \end{aligned}$$

This satisfies the final condition (4) because $T-t=0$ when $t=T$. It satisfies the backward equation (3) because $\partial_t f=-1$ and $\frac{1}{2}\partial_x^2 f=1$. Note: this calculation is a special case of the fact that if U is any random variable, then

$$\mathrm{E}\big[\,U^2\big]=\mathrm{E}[\,U]^2+\mathrm{var}(U)\ .$$

A more complicated example, just to illustrate the process, is payout $V(x)=x^4$. In this case we have to take the expectation of

$$(x+Z)^4=x^4+4x^3Z+6x^2Z^2+4xZ^3+Z^4\ .$$

The terms with Z and Z^3 have zero expected value because Z has zero mean and the Gaussian has a symmetric PDF. The Z^4 term satisfies the Gaussian expected value formula

$$\mathrm{E}\big[\,Z^4\big]=3\mathrm{E}\big[\,Z^2\big]^2=3\left(\sigma_Z^2\right)^2=3(T-t)^2.$$

The final solution formula is

$$f(x,t)=x^4+6x^2(T-t)+3(T-t)^2\ .$$

This satisfies the final condition $f(x,T)=x^4$. It satisfies the PDE because

$$x^4+6x^2(T-t)+3(T-t)^2\ \xrightarrow{\partial_t}\ -6x^2-6(T-t)\ ,$$

and

$$x^4+6x^2(T-t)+3(T-t)^2\ \xrightarrow{\frac{1}{2}\partial_x}\ 2x^3+6x(T-t)\ \xrightarrow{\partial_x}\ 6x^2+6(T-t)\ .$$

A final example is the "payout" is a sinusoidal oscillation with *wave number* k, which is $V(x)=\sin(kx)$. This may not happen in finance, but the example comes up in other fields, and it says a lot about Brownian motion. You can find the answer by working the integral

$$f(x,t)=\frac{1}{\sqrt{2\pi(T-t)}}\int_{-\infty}^{\infty}\sin(kx+kz)e^{-\frac{z^2}{2(T-t)}}\,dz\ .$$

It would be a long exercise in integral calculus, but one that you could do. You can find the answer using the ansatz "method" of Section 7 (guess the answer, check that it works). The ansatz is $f(x,t) = A(t)\sin(kx)$. The solution, using the method of Section 7, is

$$f(x,t) = e^{-\frac{1}{2}k^2(T-t)}\sin(kx) \ .$$

Notice that this function goes to zero exponentially as $T-t$ increases. If t is long before the payout time T, then the value function expected value is tiny. The exponential decay rate has a factor of the wave number k^2. Faster oscillation makes the expectation go to zero faster, a lot faster. You can understand this in a qualitative way by thinking of the Gaussian distribution of Z as a "bell shaped curve" with width proportional to $\sqrt{T-t}$. If this width is large enough to include several cycles of the sin function, then the positive and negative parts roughly cancel. The cancellation is never completely perfect, because $f(x,t)$ is not exactly zero for any t, but is is extremely accurate when the Gaussian "bump" is wide.

4 Derivation via Ito's lemma and martingales

We can use Ito's lemma (W_t last week is X_t here) to see that certain processes are martingales. A process Y_t is a martingale if, for $t < T$,

$$\mathrm{E}[\, Y_T \mid \text{path up to time } t] = Y_t \ .$$

Suppose we have a function $h(x,t)$ and define the process Y_t in terms of h and the Brownian motion path X_t as $Y_t = h(X_t,t)$. Ito's lemma implies that Y is a martingale if the dt part of the Ito differential is equal to zero:

$$\partial_t h(x,t) + \tfrac{1}{2}\partial_x^2 h(X_t,t) = 0 \quad \Longrightarrow \quad Y_t \text{ is a martingale.} \tag{10}$$

Here is the explanation.

Ito's lemma is a chain rule formula involving $h(x,t)$:

$$dh(X_t,t) = \partial_x h(X_t,t)\, dX_t + \left[\partial_t h(X_t,t) + \tfrac{1}{2}\partial_x^2 h(X_t,t)\right] \, dt \ . \tag{11}$$

The dX term on the right determines the infinitesimal variance of Y_t while the dt term determines the infinitesimal mean. A diffusion process is a martingale if the infinitesimal mean is zero. That's a way to understand the implication (10).

You can say sort of the same thing using the integral version of Ito's lemma is (here $t \leq T$)

$$h(X_T,t) = h(X_t,t) + \int_t^T \partial_x h(X_s,t)\, dX_s \ + \int_t^T \left[\partial_t h(X_s,s) + \tfrac{1}{2}\partial_x^2 h(X_s,s)\right] \, ds \ .$$

The Ito integral on the right is not zero, but its expected value is zero. More precisely, the increment dX_s has mean value zero and is independent of anything before time s. In particular, the expected value of dX_s is zero conditioned on

$X_t = x$. As explained in Week 2, this implies that the conditional expectation of the Ito integral is zero:

$$\mathrm{E}\left[\int_t^T \partial_x h(X_s,t)\, dX_s \,\middle|\, X_t = x \right] = 0 \; .$$

Suppose the integrand in the ds integral is zero, which is

$$\partial_t h(X_s,s) + \tfrac{1}{2}\partial_x^2 h(X_s,s) = 0 \; .$$

Take the conditional expectation (with condition $X_t = x$) of both sides and you get

$$\mathrm{E}[\, h(X_T) \mid x_t = x] = h(x,t) \; .$$

That is, Y_t is a martingale.

Now, suppose $f(x,t)$ satisfies the backward equation (3), for $t < T$, and has the final values (4). Then $Y_t = f(X_t,t)$ is a martingale. This implies that it satisfies the conditional expectation formula (2). This would verify that the conditional expectation (2) satisfies the backward equation (3), except that the logic is exactly backwards. What's missing is *existence* and *uniqueness*. The backward equation with given final condition has a solution, *existence*, and it has only one solution, *uniqueness*. If you show that a solution of the backward equation, a solution having the right final condition, satisfies the conditional expectation formula (2), then the function defined by the conditional expectation formula must be that unique solution of the backward equation.

5 Derivation via the tower property

This derivation to me seems the most natural and fundamental. It explicitly asks how the conditional expectation changes over a short interval of time. The key fact is the tower property, which is that the expectation of the conditional expectation is the expectation. In this case, suppose $t < s < T$. Use z for the value of X_s. We can find the expected value at time t by integrating over all possible values of X_s and using their conditional probabilities

$$f(x,t) = \int \mathrm{E}[\, V(X_T) \mid X_s = z] \; \Pr(X_s = z | X_t = x)\, dz \; .$$

This formula depends on the fact that Brownian motion has the Markov property. The expectation on the right does not depend on the value of $X_t - x$ because $t < s$. If $X_s = z$, then the Brownian motion path after s evolves in the same way regardless of the earlier value $X_t - x$. This expectation is (and this is the main point) $f(z,s)$. The probability density on the right is the transition density from time t to time s. Therefore, we may write

$$f(x,t) = \int f(z,s) G(z,x,s-t) dz \; . \tag{12}$$

This formula may be understood as an instance of the *law of total probability*, which pretty much the same thing as the tower property.

The backward equation is derived from the tower property formula (12) by taking $s = t + \Delta t$, then expanding in Taylor series to the proper order, then letting Δt go to zero. When $s = t + Deltat$, then $X_s = z$ is approximately equal to $X_t = x$. Therefore, we expand in Taylor series. In these calculations, we write f for $f(x,t)$, and f_x for $\partial_x f(x,t)$, etc.

$$f(x+Z, t+\Delta t) \approx f(x,t) + f_x Z + \frac{1}{2} f_{xx} Z^2 + f_t \Delta t$$
$$+ \text{ terms that don't matter as } \Delta t \to 0 \; .$$

Of course, $\mathrm{E}[Z] = 0$ and $\mathrm{E}[Z^2] = \Delta t$. These facts may be assembled to give

$$f(x,t) = f(x,t) + \frac{1}{2} f_{xx} \Delta t + f_t \Delta t + \cdots \; .$$

We cancel $f(x,t)$ from both sides, then divide what's left by Δt, then check that the $\cdots$ terms still go to zero when $\Delta t \to 0$. The result is the backward equation (3).

6 Digital options, smoothing

A *digital* option is one that pays all or nothing depending on some criterion. A digital payout would be

$$V(x) = \begin{cases} 1 \text{ if } x > x_0 \\ 0 \text{ if } x \leq x_0 \end{cases} \; .$$

Corresponding to this is the value function (2). In this case, the value function may be written as a probabability

$$f(x,t) = \Pr(\, X_T > x_0 \mid X_t = x) \; .$$

In fact, the expected value of any $0,1$ function (a function that takes values $V = 0$ or $V = 1$ only) is the probability that the value is 1.

The value function may be expressed in terms of the cumulative normal distribution function. One way to derive the formula uses the fact that, conditional on $X_t = x$, the final position is $X_T \sim \mathcal{N}(x, T-t)$. You can represent such a random variable $Y \sim \mathcal{N}(x, T-t)$ in terms of the standard normal $Z \sim \mathcal{N}(0,1)$ as

$$Y = x + \sqrt{T-t}\, Z \; .$$

This is Gaussian with mean x and variance $T - t$. The condition $X_T > x_0$ has the same probability as $Y > x_0$ (because X_T and Y have the same distribution).

Therefore

$$
\begin{aligned}
f(x,t) &= \Pr(Y > x_0) \\
&= \Pr\Big(x + \sqrt{T-t}\, Z > x_0 \Big) \\
&= \Pr\left(Z > \frac{x_0 - x}{\sqrt{T-t}} \right) \\
&= 1 - \Pr\left(Z < \frac{x_0 - x}{\sqrt{T-t}} \right) \\
f(x,t) &= 1 - N\left(\frac{x_0 - x}{\sqrt{T-t}} \right) \; . \qquad (13)
\end{aligned}
$$

This has the feature than $f(x,t) \to 0$ as $x \to -\infty$ and $f(x,t) \to 1$ as $x \to \infty$. This is clear from the definition of f, and you can see it in the solution formula. Write

$$
z = \frac{x_0 - x}{\sqrt{T-t}} \; .
$$

Then $f(x,t) = 1 = N(z)$. For example, we see that $z \to -\infty$ as $x \to \infty$, so $1 - N(z) \to 1 - 1 = 0$. The solution formula (13) implies that for a fixed t, f makes a transition from 0 to 1 as x goes from $-\infty$ to ∞.

The specific formula (13) tells us that the transition from $f \approx 0$ to $f \approx 1$ happens quickly with t is close to T. The "length scale" of the transition is $\sqrt{T-t}$. This means that when x goes from $x_0 - \sqrt{T-t}$ to $x_0 + \sqrt{T-t}$, the value function $f(x,t)$ goes from a value close to zero to a value close to 1. We say that the solution of the backward equation is "smoothing". The sharp discontinuity is the final condition is smoothed into a rapid but smooth transition.

7 Quadratic exponential and the ansatz method

Suppose the payout function is a *quadratic exponential*

$$
V(x) = e^{-rx^2} \; .
$$

This is called "quadratic exponential" rather than "Gaussian" because it is not a probability density. Still, everything related to Brownian motion seems to turn Gaussians into Gaussians. Therefore, we *guess* that the value function has the form

$$
f(x,t) = A(t) e^{-s(t)x^2} \; . \qquad (14)
$$

A mathematical guess like this is called an *ansatz* (German word that means this). You guess the form and then see whether you can find formulas for $A(t)$ and $s(t)$ so that the ansatz (14) satisfies the backward equation (3) and the final condition. The ansatz "method" is to make an ansatz like (14) and then show it works. It's hard to call it a method because it's really just a guess. Experienced people may be led to specific guesses in specific ways, but even for them it's guessing. The final condition is easy, it gives final conditions for A and s, which are

$$A(T) = 1 \ , \quad s(T) = r \ . \tag{15}$$

The ansatz method requires you to put the ansatz (14) into the backward equation (3) and see what this says about A and s. We use a dot for time derivatives, so $\dot{q}(t) = \frac{d}{dt}q(t)$.

$$\partial_t f = \dot{A}(t)e^{-s(t)x^2} - \dot{s}x^2 A(t)e^{-s(t)x^2} \ .$$

Then

$$\partial_x f = -2s(t)xA(t)e^{-s(t)x^2} \ ,$$

and

$$\partial_x^2 f = -2s(t)Ae^{-s(t)x^2} + 4s(t)^2 A(t)x^2 e^{-s(t)x^2} \ .$$

You put this into the backward equation and find

$$\dot{A}(t)e^{-s(t)x^2} - \dot{s}x^2 A(t)e^{-s(t)x^2} + \frac{1}{2}\left[-2s(t)Ae^{-s(t)x^2} + 4s(t)^2 A(t)x^2 e^{-s(t)x^2}\right] = 0 \ .$$

The exponential factor $e^{-s(t)x^2}$ appears in every term and may be cancelled. The rest may be re-arranged to the form

$$x^2\left[-\dot{s}(t)A + 2s(t)^2 A\right] + \left[\dot{A}(t) - s(t)A(t)\right] = 0 \ .$$

The quantities in square brackets are functions of t alone. Therefore, the expression on the left is a quadratic function of x for each fixed t. A polynomial that is equal to zero, as this one is, must have all coefficients equal to zero. This gives two equations

$$\dot{s}(t) = 2s(t)^2 \tag{16}$$

$$\dot{A}(t) = s(t)A(t) \ . \tag{17}$$

It is "easy" to solve these differential equations with the final conditions given. Exercise 1 asks you to do the algebra and interpret the results.

The ansatz method is used in quantitative finance in several places. There are *affine* interest rate models in which the exponent is a linear function of the x variable with a time dependent coefficient and pre-factor.

8 Hitting probabilities and boundary conditions

A *hitting time* is the first time a stochastic process X_t "hits" a specific value or satisfies a given condition. There are hitting time problems in finance that come from contracts with conditions that depend on stochastic market prices. Among these are *knock-out options*, that pay nothing if the price ever exceeds a specified knock-out price.

Suppose X_t is a Brownian motion with $X_0 = x_0$ in the range $a \leq x_0 \leq b$. Suppose you get a payout $V(X_T)$ if $a \leq X_t \leq b$ for all t in the range $0 \leq t \leq T$. Otherwise, you get zero. The value function for this payout satisfies the backward equation (3) if $a < x < b$, but clearly $f = 0$ if $x = a$ or $x = b$. These are *absorbing boundary conditions* (because the Brownian motion is "absorbed" and stopped if it ever touches a boundary point). They are also called *Dirichlet* boundary conditions.

9 Running payouts

A *running payout* is a payout that you get continuously in time rather than just at the final time. A running payout might take the form

$$Y = \int_0^T V(X_t)\, dt \ .$$

A value function approach to this uses a value function that only "sees" the coming reward after time t, not the reward that has arrived (accrued, in financial language) so far. That is

$$f(x,t) = \mathrm{E}\left[\int_t^T V(X_s)\, ds \mid X_t = x \right] \ . \tag{18}$$

An Ito's lemma derivation of a backward equation uses the observation that when time goes from t to $t + dt$, the integral decreases by $V(x)dt$. Therefore

$$\mathrm{E}[\, df(X_t,t)] = -V(X_t)dt \ .$$

The Ito calculation from before (look at the quantity in square braces) implies that

$$\partial_t f(x,t) + \frac{1}{2}\partial_x^2 f(x,t) = -V(x) \ . \tag{19}$$

Of course, the final condition is $f(x,T) = 0$ because the payout stops at the final time T.

10 Finite difference methods

Finite difference methods are numerical algorithms for solving (approximately) PDEs. They apply to a vast range of PDEs of all types and from all fields. This section describes some finite difference methods for solving the backward equation. The derivation uses the convergence of random walk to Brownian motion (Week 1). The finite difference approximation is the backward equation that the random walk satisfies. There are ways to derive these and other finite difference methods that do not rely on probability.

Consider the value function that satisfies the simple backward equation (3). We often call x the *space variable* and t *time variable*. We consider a random walk approximation to Brownian motion. There is a *space step* Δx and space *grid points* $x_j = j\Delta x$. There is a *time step* Δt and discrete times $t_k = k\Delta t$. The random walk (notation from Week 1) has $X^{\Delta t}_{t_k} = x_j$ for some integer j. In one step, the walk can go left, or not move, or move right. The probabilities to move left, right, or not move are a, c, and b respectively.

$$X^{\Delta t}_{t_{k+1}} = \begin{cases} X^{\Delta t}_{t_k} - \Delta x & \text{with probability } a \\ X^{\Delta t}_{t_k} & \text{with probability } b \\ X^{\Delta t}_{t_k} + \Delta x & \text{with probability } c \end{cases} \tag{20}$$

The discrete value function will be called F. [Be careful when writing by hand to make the continuous value function f look different than the discrete value function F.] Assume that the final time T is one of the discrete times. There is an n with $T = t_n$. We may have to adjust Δt to make this happen. The values of F are

$$F_{kj} = \mathrm{E}\left[\, V\left(X^{\Delta t}_{t_n}\right) \mid X^{\Delta t}_{t_k} = x_j \right] \; . \tag{21}$$

This is like the definition of the continuous value function (2), but applied to the random walk $X^{\Delta t}$ instead of the Brownian motion X.

We have to relate the probabilities a, b, and c to the space step Δx and the time step Δt. The relationship comes from the fact that the random walk increment in one time step should have the mean and variance of the Brownian motion increment over a time Δt, which is Δt. The expected value of the discrete increment should be zero:

$$\mathrm{E}\Big[X^{\Delta t}_{t_k} - X^{\Delta t}_{t_{k-1}} \Big] = 0 \; .$$

The possible values of the increment are $\pm\Delta x$ and 0, so we get

$$0 = a\,(-\Delta x) + b\,(0) + c\,(\Delta x) \; .$$

This gives

$$a = c \; .$$

The random walk is symmetric. The variance calculation is similar

$$\Delta t = a\,(\Delta x^2) + b(0) + c(\Delta x^2) = 2a\Delta x^2 \; .$$

This leads to the *CFL* ratio formula

$$a = \frac{1}{2}\,\frac{\Delta t}{\Delta x^2} \; . \tag{22}$$

CFL is for the mathematicians Richard Courant (founder of the Courant Institute), Kurt Friedrichs (one of its first faculty) and Hans Lewy. Their 1928 paper laid the foundations for finite difference solution of PDEs. The fraction is the *CFL ratio*

$$\lambda = \frac{\Delta t}{\Delta x^2} \; . \tag{23}$$

The coefficients have to add up to one because they are probabilities. This leads to a formula for b

$$
\begin{aligned}
a + b + c &= 1 \\
\frac{1}{2}\lambda + b + \frac{1}{2}\lambda &= 1 \\
b = 1 - \lambda &= 1 - \frac{\Delta t}{\Delta x^2} \; .
\end{aligned}
\tag{24}
$$

The fact that $b \geq 0$ implies that

$$
1 - \frac{\Delta t}{\Delta x^2} \geq 0 \; .
$$

This may be written as

$$
\lambda = \frac{\Delta t}{\Delta x^2} \leq 1 \; . \tag{25}
$$

This is the famous CFL *stability limit.* People usually want a large CFL number λ so that fewer time steps are required. The formulas (26) with (22) and (24) make sense even if $\lambda > 1$. But the code will "blow up" if you do.

The code `FiniteDifference.py` uses these formulas. The number of grid points in space, n, is specified, along with the length of the interval, L. This determines the space step Δx. The CFL ratio λ is used to find Δt. This time step is then adjusted down slightly so that T is an integer number of time steps from $t = 0$, which is $T = n_t \Delta t$. Most of the work of the code is the time step calculation (26).

The discrete value function satisfies a discrete recursion relation relation. The expected values $F_{k-1,j}$ may be computed from the values F_{kj} using the fact that if $X^{\Delta t}_{t_{k-1}} = x_j$, then $X^{\Delta t}_{t_k}$ is one of the values $x_j - \Delta t = x_{j-1}$ or x_j or $x_j + \Delta t = x_{j+1}$, and the probabilities are a, b, and c. The calculations we're about to do simplify because $X^{\Delta t}$ is a Markov process (definition in Week 1). This implies that, for example, that if $X^{\Delta t}$ steps from x_{j-1} to x_j, then the expected value going forward from x_j doesn't depend on the fact that it came from x_{j-1}. In formulas, this is

$$
\mathrm{E}\Big[V\big(X^{\Delta t}_{t_n}\big) \mid X^{\Delta t}_{t_{k-1}} = x_j \text{ and } X^{\Delta t}_{t_k} = x_{j-1} \Big] = \mathrm{E}\big[V\big(X^{\Delta t}_{t_n}\big) \mid X^{\Delta t}_{t_k} = x_j \big]
$$

The conditional expectation calculation using these ideas is

$$
\begin{aligned}
F_{k-1,j} &= \mathrm{E}\Big[V\big(X^{\Delta t}_{t_n}\big) \mid X^{\Delta t}_{t_{k-1}} = x_j \Big] \\
&= a\,\mathrm{E}\Big[V\big(X^{\Delta t}_{t_n}\big) \mid X^{\Delta t}_{t_{k-1}} = x_{j-1} \Big] \\
&\quad + b\,\mathrm{E}\Big[V\big(X^{\Delta t}_{t_n}\big) \mid X^{\Delta t}_{t_{k-1}} = x_j \Big] \\
&\quad + c\,\mathrm{E}\Big[V\big(X^{\Delta t}_{t_n}\big) \mid X^{\Delta t}_{t_{k-1}} = x_{j+1} \Big] \\
F_{k-1,j} &= a\,F_{k,j-1} + b\,F_{k,j} + c\,F_{k,j+1} \; .
\end{aligned}
\tag{26}
$$

This calculation starts with given final values $F_{n_t,j} = V(x_j)$. Then it loops over k doing *time steps* going backwards from n_t. Each time step is a loop over j. The boundary conditions in the code are that $f(0,t) = f(L,t) = 0$. This translates into $F_{k,0} = 0$ and $F_{k,n+1} = 0$. That leaves n "interior" grid points $x_1, \cdots, x_n$, which are separated by Δx. Therefore,

$$\Delta x = \frac{L}{n+1} .$$

The calculations (26) are done for $j = 1, \cdots, n$. In principle you don't have to store the boundary values because they are known and don't have to be computed. Storing them makes the code simpler. You can do the formula (26) for every j value without writing special code for the end cases $j = 1$ and $j = n$. Values used in this way are *ghost values*.

11 Exercises

1. Carry out the ansatz analysis of Section 7

   (a) Solve the differential equation (16). *Hint*. It may be written

   $$\frac{ds}{s^2} = 2dt .$$

   The integral of the left side is $-\frac{1}{s} + C$. The integral of the right side is $2t + C$. The constant is determined by the final condition $s(T) = r$. If you want the Wikipedia solution, it might help to know this is an example of a *Riccati* equation.

   (b) Solve the differential equation (17) and use the final condition to find a formula for the *prefactor* $A(t)$.

   (c) Is $s(t)$ an increasing or decreasing function of t? What does this say about the "width" of the payout and the width of the value function? Intuitively, why should one be wider than the other?

   (d) Is $A(t)$ increasing or decreasing? Why should the maximum of $f(x,t)$ for $t < T$ be larger/smaller (you pick) than the maximum of V?

   (e) Show that

   $$\frac{d}{dt} \int_{-\infty}^{\infty} f(x,t)\, dx = 0 .$$

   Are your formulas for s and A consistent with this? Could you derive the formula for A from this identity and the formula for s?

2. Suppose $x > 0$ and $t < T$. Define the *survival probability* starting from x between times t and T to be

   $$f(x,t,T) = \Pr(\, X_s > 0 \ \text{ for all } s \in [t,T]) \ .$$

This is the probability that the Brownian motion does not hit $x = 0$ at any time between t and T. We put in the dependence on the final time T to enable the calculations below. Consider the seemingly different problem with payout $V(x) = 1$ if $x > 0$ and $V(x) = -1$ if $x \leq 0$. The corresponding value function is

$$g(x, t, T) = \mathrm{E}[\, V(X_T) \mid X_t = x] \ .$$

This is defined for any x and $t \leq T$.

(a) Show that $g(-x, t, T) = g(x, t, T)$ for all x and $t < T$. Show that this implies that $g(0, t) = 0$ if $t < T$.

(b) Show that $g(x, T, T) = f(x, T, T)$ if $x > 0$. Assuming that the solution to the problem f satisfies is unique, show that $g(x, t, T) = f(x, t, T)$ if $x > 0$. (The finite difference approximation suggests that the solution is unique because it gives an algorithm for computing it. A course on PDE typically has a real mathematical proof.

(c) Define the hitting time to be the first time the Brownian motion touches the boundary, $x = 0$:

$$\tau = \min \{\, s \mid X_s = 0 \} \ .$$

Let $u(s)$ be the PDF of τ. Show that, conditional on $X_t = x$,

$$u(T) = -\partial_T f(x, t, T) \ .$$

(d) Find a formula for g in terms of the cumulative normal. This is similar to the formula in Section 6.

(e) Find a formula for $u(T)$. We used this formula in Exercise 6 of Week 1. This exercise fulfills the promise made there.

3. Download and run the posted code `FiniteDifference.py`. Check that the resulting plot matches the posted plot. Modify the code to compute the expected running payout function (18) with payout $V(x) = e^{-r(x-\frac{L}{2})^2}$. Choose $L = 10$, $r = 1$, and $T = 2$. Plot a series of computations in the same figure, as `FiniteDifference.py` does, to see how many grid points are needed to get an accurate solution. Explore the grid spacing needed for accurate solution when r is larger – in a qualitative way (larger r needs more/fewer/a lot more/a lot fewer points. Part of this exercise is to derive a finite difference method for the backward equation (19). You can do this by making a random walk approximation to the process and a corresponding finite sum approximation to the running payout.

4. Consider the digital option of Section 6. This exercise asks you to *replicate* the option payout using a given initial *endowment* (amount of money) and a trading strategy on Brownian motion. The trading strategy is a function

$a(x,t)$. The initial endowment is a number g. The Brownian motion starts at $X_0 = 0$.

$$U_T = g + \int_0^T a(X_t,t)dX_t \ .$$

This random variable replicates the payout if $U_T = V(X_T)$. Find a way to replicate the digital payout in this way. *Hint.* Use Ito's lemma, the value function from Section 6, and the formula

$$\partial_x N(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \ .$$

5. Write a simulation code to verify the trading strategy of Exercise 4. Much of the code can be taken from Week 2. Choose a time step Δt and make the proper Ito approximation to the Ito integral of Exercise 4. Estimate the mean square replication error, which is

$$\mathrm{E}\Big[\,(U_T - V(X_T))^2\Big] \ .$$

This should decrease to zero as $\Delta t \to 0$. You need to make many paths to estimate the expected value accurately.