Week 6
Ito's lemma for Brownian motion

Jonathan Goodman

October 21, 2013

1 Introduction to the material for the week

Ito's lemma is the big thing this week. It plays the role in stochastic calculus that the fundamental theorem of calculus plays in ordinary calculus. Most actual calculations in stochastic calculus use some form of Ito's lemma. Ito's lemma is one of a family of facts that make up the *Ito calculus*. We use it both as a language for expressing models, and as a set of tools for reasoning about models.

We give a standard example from ordinary calculus to illustrate reasoning and modeling with differentials. Let N_t denote the number of bacteria in a dish. We model N_t as growing with simple rate r. In a small increment of time dt, the model is that N increases by an amount

$$dN_t = rN_t dt \ . \tag{1}$$

The increase in the number of bacteria, in a small amount of time, is proportional to the number and the time interval, with r as the constant of proportionality. Here we used calculus as a modeling tool. We next use calculus as an analysis tool, and calculate $d\left(Ae^{rt}\right) = rAe^{rt}dt$. So, if $N_t = Ae^{rt}$, then N_t satisfies the model equation (1). We pin down the constant A by matching initial conditions. If N_0 is given, then $N_0 = A$, so we have $N_t = N_0 e^{rt}$. This allows us to estimate the doubling time, which is T so that $N_T = 2N_0$. Some algebra leads to $T = \frac{1}{r}\log(2) = \frac{.693}{r}$. This is a non-trivial prediction from the model (1).

Here is a similar example for a stochastic process X_t that could model a stock price. We suppose that in the time interval dt that X_t changes by a random amount whose size is proportional to X_t. In stock terms, the probability to go from 100 to 102 is the same as the probability to go from 10 to 10.2. A simple way to do this is to make dX proportional to X_t and dW_t, as in

$$dX_t = \sigma X_t dW_t \ . \tag{2}$$

Here, we used stochastic calculus to create a mathematical model. Ito's lemma is an analysis tool that allows us to solve this model with the formula $X_t =$

$X_0 e^{\sigma W_t - \frac{\sigma^2}{2}t}$. This implies, for example, that

$$\mathrm{P}(X_T > X_0) = \mathrm{P}\left(W_t > \frac{\sigma t}{2}\right) \, .$$

This may be expressed in terms of the cumulative normal using the trick $W_t \sim \sqrt{t}Z$, where $Z \sim \mathcal{N}(0,1)$, since both W_t and $\sqrt{t}Z$ are Gaussian with mean zero and variance t. Therefore $\mathrm{P}\left(W_t > \frac{\sigma t}{2}\right) = \mathrm{P}\left(Z > \frac{\sigma\sqrt{t}}{2}\right)$, and

$$\mathrm{P}(X_T > X_0) = 1 - N\left(\frac{\sigma\sqrt{t}}{2}\right) \approx \frac{1}{\sqrt{2\pi\sigma^2 t}} e^{-\sigma^2 t/2} \, .$$

The last approximation comes from a useful approximation to $N(z)$ when z is large. We conclude that it is exponentially unlikely to have $X_T > X_0$ in the stock "growth" model (2).

There is more than one version of Ito's lemma. The one discussed this week is about the time derivative of stochastic processes $f(W_t, t)$, where $f(w,t)$ is a differentiable function of its arguments. In future weeks we will discuss more functions, such as $f(X_t, t)$, where X_t is a more general diffusion process.

Informally, the Ito differential is

$$df = f(W_{t+dt}, t + dt) - f(W_t, t) \, .$$

This is the change in f over a small increment of time dt. If you add together (integrate) the values of df over a range of times, you get the total change in f. If X_t is any process, then

$$X_b - X_a = \int_a^b dX_s \, . \tag{3}$$

If you want to call something dX_t, it should have this property. In particular, the differential formula $dX_t = \mu_t dt + \sigma_t dW_t$, means that

$$X_b - X_a = \int_a^b \mu_s \, ds + \int_a^b \sigma_s \, dW_s \, . \tag{4}$$

The first integral on the right is an ordinary integral. The second is the Ito integral from last week. The Ito integral is well defined provided σ_t is an adapted process.

Ito's lemma for Brownian motion is

$$df(W_t, t) = \partial_w f(W_t, t) dW_t + \frac{1}{2}\partial_w^2 f(W_t, t) dt + \partial_t f(W_t, t) dt \, . \tag{5}$$

An informal derivation starts by expanding df in Taylor series in dW and dt up to second order in dW and first order in dt,

$$df = \partial_w f \, dW + \frac{1}{2}\partial_w^2 f \, (dW)^2 + \partial_t f \, dt \, .$$

We get (5) from this using $(dW_t)^2 = dt$. The formula $(dW_t)^2 = dt$ cannot be exactly true, because $(dW_t)^2$ is random and dt is not random. It is true that $\mathrm{E}\left[(dW_t)^2 | \mathcal{F}_t \right] = dt$, but Ito's lemma is about more than expectations.

The real theorem of Ito's lemma, in the spirit of (4), is

$$\begin{aligned} f(W_b, b) - f(W_a, a) & \\ &= \int_a^b \partial_w f(W_t, t) dW_t + \int_a^b \left(\frac{1}{2} \partial_w^2 f(W_t, t) + \partial_t f(W_t, t) \right) dt \end{aligned} \qquad (6)$$

Everything here is has been defined. The second integral on the right is an ordinary Riemann integral. The first integral on the right is the Ito integral defined last week. We give an informal proof of this in Section 2.

You see the convenience of Ito's lemma by re-doing the example from last week

$$X_t = \int_0^t W_s dW_s \ .$$

A first guess from ordinary calculus might be $X_t = \frac{1}{2} W_t^2$. Let us take the Ito differential of $\frac{1}{2} W_t^2$. For this, we take $f(w, t) = \frac{1}{2} w$, and we calculate the terms in (5):

$$\partial_w f = w \ \ , \quad \partial_w^2 f = 1 \ \ , \quad \partial_t f = 0 \ .$$

Therefore, (5) gives

$$d\left(\frac{1}{2} W_t^2 \right) = W_t dW_t + \frac{1}{2} dt \ .$$

Therefore,

$$\begin{aligned} \frac{1}{2} W_t^2 - \frac{1}{2} W_0^2 &= \int_0^t W_s dW_s + \frac{1}{2} \int_0^t ds \\ &= \int_0^t W_s dW_s + \frac{1}{2} t \ . \end{aligned}$$

You just rearrange this and recall that $W_0 = 0$, and you get the formula from Week 5:

$$X_t = \int_0^t W_s dW_s = \frac{1}{2} W_t^2 - \frac{1}{2} t \ .$$

This is quicker than the telescoping sum stuff from Week 5.

Another example is the solution of the stochastic differential equation, or , *SDE*, (2). A first guess might be $X_t = X_0 e^{\sigma W_t}$. This is the solution you would get, using ordinary calculus, if W_t were a differentiable function of t. So let us try $f(w, t) = X_0 e^{\sigma w}$. This time, the calculations are

$$\partial_w f = \sigma f \ \ , \quad \partial_w^2 f = \sigma^2 f \ \ , \quad \partial_t f = 0 \ .$$

We then get

$$d\left(X_0 e^{\sigma W_t} \right) = \sigma \left(X_0 e^{\sigma W_t} \right) dW_t + \frac{1}{2} \sigma^2 \left(X_0 e^{\sigma W_t} \right) dt \ .$$

We see that $X_t = X_0 e^{\sigma W_t}$ does not satisfy the SDE (2). We can get rid of the second term on the right by trying a more complicated solution

$$X_t = X_0 e^{\sigma W_t - \frac{1}{2}\sigma^2 t} . \tag{7}$$

We take the Ito differential of this expression using $f(w,t) = X_0 e^{\sigma w - \frac{1}{2}\sigma^2 t}$. The result is

$$\partial_w f = \sigma f \quad , \quad \partial_w^2 f = \sigma^2 f \quad , \quad \partial_t f = -\frac{1}{2}\sigma^2 f \ .$$

Now, if we set $X_t = f(W_t, t)$, we get

$$\begin{aligned} dX_t &= \partial_w f(W_t,t) dW_t + \frac{1}{2}\partial_w^2 f(W_t,t) dt + \partial_t f(W_t,t) dt \\ dX_t &= \sigma X_t dW_t + \frac{1}{2}\sigma^2 X_t dt - \frac{1}{2}\sigma^2 X_t dt \\ &= \sigma X_t dW_t \end{aligned}$$

This is the SDE (2). The Ito calculus has produced the solution formula (7).

Ito's lemma gives a convenient way to figure out the backward equation for many problems. Ito's lemma and the martingale (mean zero) property of Ito integrals work together to tell you how to evaluate conditional expectations. Consider the Ito integral

$$X_T = \int_0^T g_s dW_s \ .$$

Then

$$\mathrm{E}[\, X_T \mid \mathcal{F}_t] = \mathrm{E}\left[\int_0^t g_s dW_s \mid \mathcal{F}_t\right] + \mathrm{E}\left[\int_t^T g_s dW_s \mid \mathcal{F}_t\right]$$

The first term is completely known at time t, so the expectation is irrelevant. The second term is zero, because dW_s is in the future of g_s and $\mathcal{F}_t$. Therefore

$$\mathrm{E}\left[\int_0^T g_s dW_s \mid \mathcal{F}_t\right] = \int_0^t g_s dW_s \ .$$

Now suppose $f(w,t)$ is the value function

$$f(w,t) = \mathrm{E}[\, V(W_T) \mid W_t = w] \ . \tag{8}$$

The integral form of Ito's lemma (6)

$$\begin{aligned} V(W_T) - f(W_t,t) &= \int_t^T df(W_s,s) \\ &= \int_t^T \partial_w f(W_s,s) dW_s + \int_t^T \left(\partial_t f(W_s,s) + \frac{1}{2}\partial_w^2 f(W_s,s)\right) ds \end{aligned}$$

Take the conditional expectation in $\mathcal{F}_t$. In view of the definition (8), the left side gives

$$\mathrm{E}[\, V(W_T) \mid \mathcal{F}_t] - f(W_t,t) = 0 \ .$$

The conditional expectation of the Ito integral on the right also vanishes, as we just said. Therefore

$$\mathrm{E}\left[\int_t^T \left(\partial_t f(W_s,s) + \frac{1}{2}\partial_w^2 f(W_s,s)\right) ds \mid \mathcal{F}_t\right] = 0 \ .$$

The simplest way for this to happen is for the integrand to vanish identically. The equation you get by setting the integrand to zero is

$$\partial_t f + \frac{1}{2}\partial_w^2 f = 0 \ . \tag{9}$$

We can turn the logic around to a form mathematicians like better. The above calculations show the following. If you solve the backward equation (9) for $t \leq T$ with final conditions $f(w,t) = V(w)$, then $f(W_t,t) = \mathrm{E}[\,V(W_T) \mid \mathcal{F}_t]$, which is the same as (8). That is to say, the function you get by solving the backward equation is the value function.

The derivation here is quicker than the one in Week 4. Ito's lemma does all the heavy lifting.

2 Informal proof of Ito's lemma

This section explains how to prove Ito's lemma in the integral formula (6). We will prove it under the assumption that $f(w,t)$ is a differentiable function of its arguments up to third derivatives. We assume all mixed partial derivatives up to that order exist and are bounded. That means $\left|\partial_w^3 f(w,t)\right| \leq C$, and $\left|\partial_t^2 f(w,t)\right| \leq C$, and $\left|\partial_w^2 \partial_t f(w,t)\right| \leq C$, and so on.

We use the notation of Week 5, with $\Delta t = 2^{-m}$, and $t_j = j\Delta t$. The change in any quantity, Q, from t_j to t_{j+1} is ΔQ_j. We use the subscript j for t_j, as in W_j instead of W_{t_j}. For example, $\Delta f_j = f(W_j + \Delta W_j, t_j + \Delta t) - f(W_j, t_j)$. In this notation, the left side of (6) is

$$f(W_b,b) - f(W_a,a) \approx \sum_{a \leq t_j < b} \Delta f_j \ . \tag{10}$$

The right side is a telescoping sum, which is equal to the left side if $b = n\Delta t$ and $a = m\Delta t$ for some integers $m < n$. When Δt and ΔW are small, there is a Taylor series approximation of Δf_j. The leading order terms in the Taylor series combine to form the integrals on the right of (6). The remainder terms add up to something that goes to zero as $\Delta t \to 0$.

Suppose w and t are some numbers and Δw and Δt are some small changes. Define $\Delta f = f(w + \Delta w, t + \Delta t) - f(w,t)$. The Taylor series, up to the order we need, is

$$\Delta f = \partial_w f(w,t)\Delta w + \tfrac{1}{2}\partial_w^2 f(w,t)\Delta w^2 + \partial_t f(w,t)\Delta t \tag{11}$$

$$+ O\left(\left|\Delta w^3\right|\right) + O\left(\left|\Delta w\right|\Delta t\right) + O\left(\left|\Delta t^2\right|\right) \ . \tag{12}$$

The big O quantities on the second line refer to things bounded by a multiple of what's in the big O, so $O\left(\left|\Delta w^3\right|\right)$ means: "some quantity Q so that there is a C with $|Q| \leq C\left|\Delta w^3\right|$". The error terms on the second line correspond to the highest order neglected terms in the Taylor series. These are (constants omitted) $\partial_w^3 f(w,t)\Delta w^3$, and $\partial_w\partial_t f(w,t)\Delta w\Delta t$, and $\partial_t^2 f(w,t)\Delta t^2$. The Taylor remainder theorem tells us that if the derivatives of the appropriate order are bounded (third derivatives in this case), then the errors are on the order of the neglected terms.

The sum on the right of (10) now breaks up into six sums, one for each term on the right of (11) and (12):

$$\sum_{a\leq t_j<b} \Delta f_j = S_1 + S_2 + S_3 + S_4 + S_5 + S_6 \ .$$

We consider them one by one. It does not take long.

The first is

$$S_1 = \sum_{a\leq t_j<b} \partial_w f(W_j, t_j)\Delta W_j \ .$$

In the limit $\Delta t \to 0$ (more precisely, $m \to \infty$ with $\Delta t = 2^{-m}$), this converges to

$$\int_a^b \partial_w f(W_s, s) dW_s \ .$$

The second is

$$S_2 = \sum_{a\leq t_j<b} \tfrac{1}{2}\partial_w^2 f(W_j, t_j)\Delta W_j^2 \ . \tag{13}$$

This is the term in the Ito calculus that has no analogue in ordinary calculus. We come back to it after the others. The third is

$$S_3 = \sum_{a\leq t_j<b} \partial_t f(W_j, t_j)\Delta t \ .$$

As $\Delta t \to 0$ this one converges to

$$\int_a^b \partial_t f(W_s, s)\, ds \ .$$

The first error sum is

$$|S_4| \leq C \sum_{a\leq t_j<b} \left|\Delta W_j^3\right| \ .$$

This is random, so we evaluate its expected value. We know from experience that $\mathrm{E}\left[\,\left|\Delta W_j^3\right|\right]$ scales like $\Delta t^{3/2}$, which is one half power of Δt for each power of ΔW. Therefore

$$\mathrm{E}[\,S_4] \leq C \sum_{a\leq t_j<b} \Delta t^{3/2} = C\Delta t^{1/2} \sum_{a\leq t_j<b} \Delta t = C(b-a)\Delta t^{1/2} \ .$$

The second error term goes the same way, as $\mathrm{E}[\,|\Delta W_j|\,\Delta t]$ also scales as $\Delta t^{3/2}$. The last error term has

$$|S_6| \leq C \sum_{a \leq t_j < b} \Delta t^2 = C(b-a)\Delta t \; .$$

Last week we showed that if

$$\sum_{m=1}^{\infty} \mathrm{E}[\,|S_{4,m}|] < \infty \; , \tag{14}$$

then $S_{4,m} \to 0$ as $m \to \infty$, almost surely. We just showed that

$$\mathrm{E}[\,|S_{4,m}|] \leq C_t \Delta t_m^{1/2} = C_t \left(\frac{1}{\sqrt{2}}\right)^{-m} \; ,$$

which bounds the sum (14) by a convergent geometric series. The same arguments apply to S_5 and S_6. Together they show that the total contribution of the error terms (12) vanishes in the limit $\Delta t \to 0$.

The trick in this sort of analysis is distinguishing between *small* terms and *tiny* terms. Small terms, such as $f_j \Delta W_j^2$, add up to something important. Tiny terms, such as $f_j \Delta W_j \Delta t$ add up to something small. Suppose, for simplicity, that $t = 1$. Then the number of times $t_j < t$ is $1/\Delta t$. Therefore, if H_j is a term of order Δt, then

$$\sum_{t_j < 1} H_j$$

can be $O(\Delta t)/\Delta t$, which is order 1. The sum does not go to zero as $\Delta t \to 0$. This makes H_j a small term. Terms like $f_j \Delta W_j^2$ are small in this sense, because $\left| f_j \Delta W_j^2 \right|$ is order Δt. Terms that are smaller, such as $f_j \Delta W_j \Delta t$, are tiny.

It comes now to the sum (13). The $\left(\Delta W_j^2\right) \leftrightarrow \Delta t$ connection suggests we write

$$\left(\Delta W_j\right)^2 = \Delta t + R_j \; ,$$

Clearly

$$\mathrm{E}[\, R_j \mid \mathcal{F}_j] = 0 \;\; , \quad \text{and} \;\; \mathrm{E}\left[\, R_j^2 \mid \mathcal{F}_j\right] = \mathrm{var}(R_j \mid \mathcal{F}_j) = 2\Delta t^2 \; .$$

Now,

$$\begin{aligned} S_2 = & \sum_{a \leq t_j < b} \tfrac{1}{2} \partial_w^2 f(W_j, t_j) \Delta t \; + \sum_{a \leq t_j < b} \tfrac{1}{2} \partial_w^2 f(W_j, t_j) R_j \\ = & \qquad\quad S_{2,1} \qquad\qquad + \qquad\qquad S_{2,2} \; . \end{aligned}$$

The first term converges to the Riemann integral

$$\int_a^b \tfrac{1}{2} \partial_w^2 f(W s, s) \, ds \; .$$

The second term converges to zero almost surely. We see this using the now familiar trick of calculating $\mathrm{E}\left[S_{2,2}^2 \right]$. This becomes a double sum over t_j and t_k. The off diagonal terms, the ones with $j \neq k$ vanish. If $j > k$, we see this as usual:

$$\begin{aligned}
&\mathrm{E}\left[\left(\tfrac{1}{2}\partial_w^2 f(W_j,t_j)R_j\right)\left(\tfrac{1}{2}\partial_w^2 f(W_k,t_j)R_k\right) \mid \mathcal{F}_j \right] \\
&= \mathrm{E}[\, R_j \mid \mathcal{F}_j]\ \frac{1}{4}\partial_w^2 f(W_j,t_j)\partial_w^2 f(W_k,t_j)R_k \ ,
\end{aligned}$$

and the right side vanishes. The conditional expectation of a diagonal term is

$$\begin{aligned}
\frac{1}{4}\mathrm{E}\left[\left(\partial_w^2 f(W_j,t_j)R_j\right)^2 \mid \mathcal{F}_j \right] &= \frac{1}{4}\left(\partial_w^2 f(W_j,t_j)\right)^2 \mathrm{E}\left[R_j^2 \mid \mathcal{F}_j \right] \\
&= \frac{1}{2}\left(\partial_w^2 f(W_j,t_j)\right)^2 \Delta t^2
\end{aligned}$$

These calculations show that in $\mathrm{E}\left[S_{2,2}^2 \right]$, the diagonal terms, which are the only non-zero ones, sum to $\leq C(b-a)\Delta t$.

3 Backward equations

Suppose $W_{[0,T]}$ is a standard Brownian motion path for $0 \leq t \leq T$. A function of Brownian motion, $\Phi(W_{[0,T]})$, is a number whose value is determined by the path. A simple example is $\Phi(W_{[0,T]}) = V(W_T)$, which we discussed above. There are many many others (In each case, V is some function of one variable.):

$$\Phi(W_{[0,T]}) = \int_0^T V(W_t)\, dt \tag{15}$$

$$\Phi(W_{[0,T]}) = \exp\left(\int_0^T V(W_t)\, dt \right) \tag{16}$$

$$\Phi(W_{[0,T]}) = V(\max_{0\leq t\leq T} |W_t|) \tag{17}$$

$$\Phi(W_{[0,T]}) = V(\tau_a \wedge T) \tag{18}$$

$$\Phi(W_{[0,T]}) = \left(\int_0^{\tau_a \wedge T} V(W_t)\, dt \right) \tag{19}$$

A function of a path is often called a *functional.* A functional that depends on more than just W_T is called *path dependent.* The five examples above are path dependent functionals.

An important part of stochastic calculus is finding $\mathrm{E}\left[\Phi(W_{[0,T]}) \right]$ for these and other functionals. In specific applications, the quantity of interest often is one of these. Moreover, evaluating expectation values is how we we learn about Brownian motion paths. For example, calculating values of (17) helps prove that Brownian motion paths are continuous functions of t.

There are backward equation approaches to many of these functionals. The trick is to find the right value function. Ito's lemma will tell you the PDE (partial differential equation). The rest, the final conditions, and possible boundary conditions, are usually obvious.

The functional (15) is an *additive* function (integrate means add up over time). An engineer might call it a “running cost” and a finance person a “running reward”. An appropriate value function for this problem is

$$f(W_t, t) = \mathrm{E}\left[\int_t^T V(W_s) ds \mid \mathcal{F}_t \right] . \tag{20}$$

Ito’s lemma gives

$$f(W_T, T) - f(W_t, t) = \int_t^T f_w(W_s, s) dW_s + \int_t^T \left(\frac{1}{2} f_{ww}(W_s, s) + f_t(W_s, s) \right) ds .$$

The definition (20) gives $f(W_T, T) = 0$. Therefore, as in the Introduction,

$$f(W_t, t) = -\mathrm{E}\left[\int_t^T \left(\frac{1}{2} f_{ww}(W_s, s) + f_t(W_s, s) \right) ds \mid \mathcal{F}_t \right] .$$

We set the two expressions for f equal:

$$\mathrm{E}\left[\int_t^T V(W_s) ds \mid \mathcal{F}_t \right] = -\mathrm{E}\left[\int_t^T \left(\frac{1}{2} f_{ww}(W_s, s) + f_t(W_s, s) \right) ds \mid \mathcal{F}_t \right] .$$

The natural way to achieve this is to set the integrands equal to each other, which gives

$$\frac{1}{2} f_{ww}(w, s) + f_t(w, s) + V(w) = 0 . \tag{21}$$

The final condition for this PDE is $f(w, T) = 0$. The PDE then determines the values $f(w, s)$ for $s < T$. Now that we have guessed the backward equation, we can show that it is right by Ito differentiation once more. If $f(w, s)$ satisfies the backward equation (21), then $f(W_t, t)$ satisfies (20).

Here is a slightly better way to say this. From ordinary calculus, we get

$$d\left(\int_t^T V(W_s) ds \mid \mathcal{F}_t \right) = -V(W_t) dt .$$

We pause to consider this. The stochastic process

$$X_t = \int_t^T V(W_s) ds$$

is a differentiable function of t. Its derivative with respect to t follows from the ordinary rules of calculus, the fundamental theorem in this case

$$\frac{dX}{dt} \int_t^T V(W_s) ds = -V(W_t) .$$

This is true for any continuous function W_t whether or not it is random. Conditioning on $\mathcal{F}_t$ just ties down the value of W_t. From Ito's lemma, any function $f(w,s)$ satisfies

$$\mathrm{E}[\, df(W_t,t) \mid \mathcal{F}_t] = \left(\frac{1}{2} f_{ww}(W_s,s) + f_t(W_s,s)\right) dt \; .$$

Taking expectations on both sides of (20) gives

$$\left(\frac{1}{2} f_{ww}(W_s,s) + f_t(W_s,s)\right) dt = -V(W_t) dt \; ,$$

which is the backward equation (21).

Consider the specific example

$$f(w,t) = \mathrm{E}_{w,t}\left[\int_t^T W_s^2 dt \right] \; .$$

We could find the solution by direct calculations, since there is a simple formula $\mathrm{E}_{w,t}\left[\, W_s^2 \right] = \mathrm{E}_{w,t}\left[\, W_t^2 + (W_s - W_t)^2 \right] = w^2 + (s-t)$. Instead we use the *ansatz* method. Suppose the solution has the form $f(w,t) = A(t)w^2 + B(t)$. It is easy to plug into the backward equation

$$\frac{1}{2} f_{ww} + f_t + w^2 = 0$$

and get

$$2A + \dot{A}w^2 + \dot{B} + w^2 = 0 \; .$$

This gives $\dot{A} = -1$. Since $f(w,T) = 0$, we have $A(T) = 0$ and therefore $A(t) = T - t$. Next we have $\dot{B} = 2T - 2t$, so $B = 2Tt - t^2 + C$. The final condition $B(T) = 0$ gives $C = -T^2$. The simplified form is $B(t) = 2Tt - t^2 - T^2 = -(T-t)^2$. The solution is $f(w,t) = (T-t)w^2 - -(T-t)^2$.

3.1 Doing without Ito's lemma, Feynman Kac

Here is a direct derivation of a backward equation for (16). The value function is given by either of these equivalent expressions

$$f(w,t) = \mathrm{E}_{w,t}\left[e^{\int_t^T V(W_s) ds} \right] \quad , \quad f(W_t,t) = \mathrm{E}\left[e^{\int_t^T V(W_s) ds} \mid \mathcal{F}_t \right] \; . \tag{22}$$

In a future week we will a derivation of the backward equation using Ito calculus. But we do it here using just Taylor series and the tower property.

Consider the time interval between t and $t + \Delta t$. During that time, suppose a Brownian motion path goes from $W_t = w$ to $W_{t+\Delta t} = w + \Delta W$. The tower property relates $f(w,t)$ to the value function at time $t + \Delta t$, as follows:

$$f(w,t) = \mathrm{E}_{w,t}\left[\mathrm{E}\left[e^{\int_t^T V(W_s) ds} \mid \mathcal{F}_{t+\Delta t} \right]\right] \; . \tag{23}$$

Of course, at time $t+\Delta t$, the first part of the path $W_{[t,T]}$ is known. Therefore the inner expectation can be rewritten as

$$\begin{aligned}
\mathrm{E}\left[e^{\int_t^T V(W_s)ds} \mid \mathcal{F}_{t+\Delta t}\right] &= \mathrm{E}\left[e^{\int_t^{t+\Delta t} V(W_s)ds + \int_{t+\Delta t}^T V(W_s)ds} \mid \mathcal{F}_{t+\Delta t}\right] \\
&= e^{\int_t^{t+\Delta t} V(W_s)ds}\, \mathrm{E}\left[e^{\int_{t+\Delta t}^T V(W_s)ds} \mid \mathcal{F}_{t+\Delta t}\right] \\
&= e^{\int_t^{t+\Delta t} V(W_s)ds}\, f(W_{t+\Delta t}, t+\Delta t)\ .
\end{aligned}$$

Restoring the outer expectation in (23) then transforms this to

$$f(w,t) = \mathrm{E}_{w,t}\left[e^{\int_t^{t+\Delta t} V(W_s)ds} f(w+\Delta W, t+\Delta t)\right]\ . \tag{24}$$

Some Taylor expansions turn this into the backward equation.

The following notation simplifies Taylor remainder computations. If Q is a quantity bounded by a power of Δt, we might say $Q = O(\Delta t^p)$. That means that there is a C do that as $|Q| \leq C\Delta t^p$ as $\Delta t \to 0$. Equivalently, we can say $Q = A\Delta t^p$, where A is not constant, but is bounded as $\Delta t \to 0$. (Those who have taken mathematical analysis can translate "bounded as $\Delta t \to 0$" to "there is a C and a $\Delta t_0 > 0$ so that if $\Delta t < \Delta t_0$ then $|A| \leq C$.") We use this also for random quantities. For example, $\Delta W = B\sqrt{\Delta t}$ implies that the distribution of B is bounded as $\Delta t \to 0$.

You will see how this convention system works as we go through the expression (24). If $0 \leq s \leq \Delta t$, then $W_s - w = A\sqrt{\Delta t}$. Therefore

$$\begin{aligned}
V(W_s) &= V(w) + V'(w)A\sqrt{\Delta t} \\
&= V(w) + A\sqrt{\Delta t}\ .
\end{aligned}$$

We use the A like we use C before; it is just "some bounded function", so the A on the first line is not the same as the A on the second line, and neither is the same as the A above. We continue in this way:

$$\begin{aligned}
\int_0^{\Delta t} V(W_s)ds &= \Delta t V(w) + \sqrt{\Delta t}\int_0^{\Delta t} A ds \\
&= \Delta t V(w) + A\Delta t^{3/2}\ .
\end{aligned}$$

Continuing with the expansion of the exponential,

$$\begin{aligned}
e^{\int_0^{\Delta t} V(W_s)ds} &= e^{V(w)\Delta t + A\Delta t^{3/2}} \\
&= 1 + V(w)\Delta t + A\Delta t^{3/2} + B\left(V(w)\Delta t + A\Delta t^{3/2}\right)^2 \\
&= 1 + V(w)\Delta t + A\Delta t^{3/2}\ .
\end{aligned} \tag{25}$$

In going from the second to the third lines, we combined terms with powers of Δt, all powers $3/2$ or higher. The A on the third line depends on the A and B on the second line. We then expand the f term on the right of (24). We need to to compute all terms of size $A\Delta t$. Since $\Delta W = A\sqrt{\Delta t}$, this means keeping terms up to ΔW^2 and treating terms ΔW^3 or higher as Taylor remainders. The expansion, as far as we need it, and with remainder bounds in our style, is

$$\begin{aligned} f(w+\Delta w, t+\Delta t) =& f(w,t) + \partial_w f(w,t)\Delta W + \frac{1}{2}\partial_w^2 f(w,t)\Delta W^2 + \partial_t f(w,t)\Delta t \\ &+ \left[(\partial_w \partial_t f \text{ term }) + (\partial_w^3 f \text{ term }) + (\partial_t^2 f \text{ term }) \right] \\ &\left[\quad A\Delta t^{1/2}\Delta t \quad + \quad B\Delta t^{3/2} \quad + \quad C\Delta t^2 \quad \right] . \end{aligned}$$

We now write f for $f(w,t)$, etc., to save space. The above Taylor expansion is

$$f(w+\Delta W, t+\Delta t) = f + \partial_w f \Delta W + \frac{1}{2}\partial_w^2 f \Delta W^2 + \partial_t f \Delta t + A\Delta t^{3/2} \ .$$

Now put this and (25) back into (24). In the first line, the things that are random once w and t are given are ΔW with $\mathrm{E}_{w,t}[\,\Delta W] = 0$ and $\mathrm{E}_{w,t}\left[\,\Delta W^2\right] = \Delta t$, and A and B with $\mathrm{E}_{w,t}[\,A] = A$ and $\mathrm{E}_{w,t}[\,B] = B$ (our convention: A is a bounded function or a bounded number).

$$\begin{aligned} f &= \mathrm{E}_{w,t}\left[\left(1 + V(w)\Delta t + A\Delta t^{3/2}\right)\left(f + \partial_w f\Delta W + \frac{1}{2}\partial_w^2 f \Delta W^2 + \partial_t f\Delta t + B\Delta t^{3/2}\right)\right] \\ &= f + V(w) f \Delta t + \partial_w \mathrm{E}_{w,t}[\,\Delta W] + \frac{1}{2}\mathrm{E}_{w,t}\left[\,\Delta W^2\right] + \partial_t f \Delta t + \mathrm{E}_{w,t}\left[\,(A+B)\,\Delta t^{3/2}\right] \ . \end{aligned}$$

We simplify this by canceling the common $f(w,t)$ from both sides, then dividing both sides by Δt. The terms are rearranged into a traditional order:

$$0 = \partial_t f + \frac{1}{2}\partial_2^2 f + V(w) f + A\sqrt{\Delta t} \ .$$

Finally we take the limit $\Delta t \to 0$ and restore the w, t arguments:

$$0 = \partial_t f(w,t) + \frac{1}{2}\partial_2^2 f(w,t) + V(w) f(w,t) \ . \tag{26}$$

This is the backward equation for (16). The final condition, clearly, is $f(w,T) = 1$ for all w.

The names Feynman and Kac (pronounced "cats") are associated with the PDE (26). Richard Feynman was studying the Schrödinger equation, which is similar to (26) and proposed a non-rigorous "path integral" solution formula for it. Mark Kac noticed that Feynman's reasoning could be applied to the PDE (26), and that it would be rigorous. The formula (22) gives the solution of the PDE (26), so (22) is called the *Feynman Kac formula.* Alas, there is lots of lazy attribution. Therefore, you will see (26) or even (21) called the Feynman Kac formula.

3.2 How to use the backward equation

Backward equations are important even when they do not have exact solutions. Numerical solution of the backward equation can be faster and more accurate than evaluating the value function using Monte Carlo. There is no random statistical error in a finite difference or finite element approximation of the backward equation. There are analytic methods for PDEs, including series and asymptotic approximations.

We sometimes go the other way, using Monte Carlo evaluation of the value functions defined by (15), etc. This is particularly common in high dimensions, where finite difference methods are impractical.