Stochastic Calculus Notes, Lecture 7

Last modified November 14, 2002

1 Ito Stochastic Differential Equations

1.1. Notation: We switch back to the notation W_t for Brownian motion. We use X_t to denote the solution of the stochastic differential equation (SDE). When we write forward and backward equations for X_t, the independent variable will still be x. Often we work in more than one dimension. In this case, W_t may be a vector of independent Brownian motion paths. As far as possible, we will use the same notation for the one dimensional (scalar) and multidimensional cases. The solution of the Ito differential equation will be X_t. We sometimes call these "diffusions".

1.2. The SDE: A stochastic differential equation is written

$$dX_t = a(X_t, t)dt + \sigma(X_t, t)dW_t \ . \tag{1}$$

A solution to (1) is a process $X_t(W)$ that is an adapted function of W ($X_t \in \mathcal{F}_t$, where $\mathcal{F}_t$ is generated by the values W_s for $s \leq t$), so that

$$X_T = X_0 + \int_0^T a(X_t, t)dt + \int_0^T \sigma(X_t, t)dW_t \ . \tag{2}$$

Because X_t is adapted, the Ito integral on the right of (2) makes sense. The term $a(X_t, t)dt$ is called the "drift" term. If $a \equiv 0$, X_t will be a martingale; any change in $E[X_t]$ is due to the drift term. The term $\sigma(X_t, t)dW_t$ is the "noise" term. the coefficient σ may be called the "diffusion" coefficient, or the "volitility" coefficient, though both of these are slight misnomers. The volitility coefficient determines the size of the small scale random motions that characterize diffusion processes. The form (1) is really just a shorthand for (2). It is traditionally written in differential notation (dX_t, dt, dW_t) as a reminder that Ito differentials are mor subtle than ordinary differentials from calculus with differentiable functions.

What separates diffusion processes from simple Brownian motions is that in diffusions the drift and volatility coefficients may depend on X and t. It might be, for example, that when X is large, it's fluctuation rate is also large. This would be modelled by having $\sigma(x, t)$ being an increasing function of x.

In the multidimensional case, we might have $X_t \in R^n$. Clearly, this calls for $a(x, t) \in R^n$ also. This might be called the "drift vector" or "velocity field" or "drift field". The volitility coefficient becomes an $n \times m$ matrix, with $W_t \in R^m$ being m independent sources of noise. The case $m < n$ is called "degenerate diffusion" and arises often in applications. The case $n = m$ and σ non singular is called "nondegenerate diffusion". The mathematical character of the forward and backward equations is far more subtle for degenerate diffusions than for nondegenerate diffusions. The case $m > n$ arises in practice only by mistake.

1.3. Existence and uniqueness of Ito solutions: Just as the Ito value of the stochastic integral is one of several possible values depending on details of the definition, we might expect the solution of (1) to be ambiguous. We will now see that this is not so as long as we use the Ito definition of the stochastic integral in (2). The main technical fact in the existence/uniqueness theory is a "short time contraction estimate": the mapping defined by (2) is a contraction for if t is small enough. Both the existence and uniqueness theorems follow quickly from this.

Suppose X_t and Y_t are two adapted stochastic processes with $X_0 = Y_0$. We define $\tilde{X}_t$ from X_t using (2) by

$$\tilde{X}_T = \int_{t=0}^{T} a(X_t, t)dt + \int_{t=0}^{T} \sigma(X_t, t)dW_t \; .$$

In the same way, $\tilde{Y}$ is defined from Y. We assume that a and σ are Lipshchitz continuous in the x arguments: $|a(x,t) - a(y,t)| \leq M\,|x-y|$, $|a(x,t) - a(y,t)| \leq M\,|x-y|$. The best possible constants in these inequalities are called the "Lipschitz constants" for a and σ. The mapping $X \mapsto \tilde{X}$ is a" contraction" if

$$\left\| \tilde{X} - \tilde{Y} \right\| \leq \alpha \left\| \tilde{X} - \tilde{Y} \right\| \; ,$$

for some $\alpha < 1$, that is, if the mapping shortens distances between objects by a definite ratio less than one. Of course, whether a mapping is a contraction might depend on the sense of distance, the norm $\|\cdot\|$. Because our tool is the Ito isometry formula, we use

$$\|X - Y\|_T^2 = \max_{0 \leq t \leq T} E\left[\left(X_t - Y_t \right)^2 \right] ; .$$

The contraction lemma is:

Lemma: If a and σ are Lipshcitz with Lipschitz constant M, then

$$\left\| \tilde{X} - \tilde{Y} \right\|_T^2 \leq 4M^2 T \left\| \tilde{X} - \tilde{Y} \right\|_T^2 \; . \tag{3}$$

For the proof, we first write

$$\tilde{X}_T - \tilde{Y}_T = \int_{t=0}^{T} \big(a(X_t,t) - a(Y_t,t) \big) dt + \int_{t=0}^{T} \big(\sigma(X_t,t) - \sigma(Y_t,t) \big) dW_t \; .$$

We have $E[(\tilde{X}_T - \tilde{Y}_T)^2] \leq 2A + 2B$ where

$$A = E\left[\left(\int_{t=0}^{T} \big(a(X_t,t) - a(Y_t,t) \big) dt \right)^2 \right] \; .$$

and

$$B = E\left[\left(\int_{t=0}^{T} \big(\sigma(X_t,t) - \sigma(Y_t,t) \big) dW_t \right)^2 \right] \; .$$

Bounding the B term is an application of the Ito isometry formula. Indeed,

$$B \leq \int_{t=0}^{T} E\left[\left(\sigma(X_t,t) - \sigma(Y_t,t)\right)^2\right] dt \ ,$$

Using the Lipschitz continuity of σ then gives

$$B \leq M^2 T \max_{0 \leq t \leq T} E[(X_t - Y_t)^2] \ ,$$

which is the sort of bound we need.

The A term is an application of the Cauchy Schwartz inequality

$$\left(\int_{t=0}^{T} \left(a(X_t,t) - a(Y_t,t)\right) dt\right)^2 \leq \int_{t=0}^{T} \left(a(X_t,t) - a(Y_t,t)\right)^2 dt \cdot \int_{t=0}^{T} 1 dt$$

If we now use the Lipschitz continuity of a and take expectations of both sides, we get

$$A \leq M^2 T \max_{0 \leq t \leq T} E[(X_t - Y_t)^2] \ ,$$

These two inequalities prove the contraction lemma estimate (3).

1.4. Uniqueness: The contraction inequality gives a quick proof of the uniqueness theorem. We will see that if X_0 is a random variable, then the solution up to some time T is unique. Of course, then X_T is a random variable and may be thought of as initial data for the next T time period. This give uniqueness up to time $2T$, and so on. Suppose X_t and Y_t were two solutions of (2). We want to argue that $E[(X_t - Y_t)^2] < \alpha E[(X_t - Y_t)^2]$ for $\alpha < 1$. This is impossible unless $[(X_t - Y_t)^2] = 0$, that is, unless $X_t = Y_t$. From (3), we will have $\alpha < 1$ if $T < 1/4M^2$.

The contraction lemma does not say $E[(\tilde{X}_t - \tilde{Y}_t)^2] < \alpha E[(X_t - Y_t)^2]$. To work with the information it actually gives, define $m_T = \max_{t<T} E[(X_t - Y_t)^2]$, and $\tilde{m}_T = \max_{t<T} E[(\tilde{X}_t - \tilde{Y}_t)^2]$. From the definitions, it is clear that m_t is an increasing function of t, so that (3) implies that $E[(\tilde{X}_t - \tilde{Y}_t)^2] \leq \alpha m_t \leq \alpha m_T$ if $T > t$. That is, (3) implies that $\tilde{m}_T \leq \alpha m_T$. This gives a contradiction as before: Since X and Y are solutions, we have $\tilde{m} = m$, so $\tilde{m}_T \leq \alpha m_T$ is impossible unless $m_T = 0$.

1.5. Existence of solutions via Picard iteration: The contraction inequality (3) allows us also to show that there is an X_T satisfying (2), at least for $T < 1/4M^2$. You might remember this construction, Picard iteration, from a class in ordinary differential equations. The first "iterate" does not come close to satisfying the equations but just gets the ball rolling: $X_t^{(0)} = X_0$ for all $t \leq T$. This $X_t^{(0)}$ does not depend on W_t, but it will still be random if X_0 is random. For $k > 0$, the iterates are defined by

$$X_t^{(k)} = \int_{s=0}^{t} a(X_s^{(k-1)}, t) dt + \int_{s=0}^{t} \sigma(X_s^{(k-1)}, t) dW_t \ . \tag{4}$$

The contraction inequality implies that the Picard iterates, $X^{(k)}$, converge as $k \to \infty$. In (3), take X to be $X^{(k-1)}$, and $Y = X^{(k)}$. Then $\tilde{X} = X^{(k)}$ and $\tilde{Y} = X^{(k+1)}$. If we define

$$m_T^{(k)} = \max_{0 \le t \le T} E\left[\left(X_t^{(k)} - X_t^{(k)} \right)^2 \right] ,$$

and use the ideas of the previous paragraph, (3) gives

$$m_T^{(k+1)} \le \alpha m_T^{(k)} .$$

This implies that, for any $t \le T$, the iterates $X_t^{(k)}$ have $E[(X_t^{(k+1)} - X_t^{(k)})^2] \le m_T^{(0)}$, which (as we saw in the previous lecture) implies that $\lim_k \to \infty X_t^{(k)}$ exists. The contraction inequality also shows (reader: think this through) that this limit, X_t satisfies (2) and therefore is what we are looking for.

2 Ito's Lemma

We want to work out the first few Picard iterates in an example. This leads to a large number of stochastic integrals. We could calculate any of them in an hour or so, but we would soon long for something like the Fundamental Theorem of calculus to make the calculations mechanical. That result is called "Ito's lemma". Not only is it helpful in working with stochastic integrals and SDE's, it is also a common interview question for young potential quants. Here is the answer.

2.1. The Fundamental Theorem of calculus: The following derivation of the Fundamental Theorem of ordinary calculus provides a template for the derivation of Ito's lemma. Let $V(t)$ be a differentiable function of t with $\partial_t V$ being Lipschitz continuous. The Fundamental Theorem states that (writing $\partial_t V$ for dV/dt although V depends only on t):

$$V(T) - V(0) = \int_0^T dV = \int_0^T \partial_t V(t) dt .$$

This exact formula follows from two approximate short time approximations, the first of which is

$$V(t + \Delta t) - V(t) = \partial_t V(t) \Delta t + O(\Delta t^2) .$$

The second approximation is (writing $\partial_t V(s)$ for $V'(s)$):

$$\int_t^{t+\Delta t} \partial_t V(s) ds = \partial_t V(t) \Delta t + O(\Delta t^2) .$$

Using our habitual notation ($\Delta t = T/n = T/2^L$, $t_k = k\Delta t$, $V_k = V(t_k)$), we have, using both approximations above,

$$V_n - V_0 \quad = \quad \sum_{k=0}^{n-1} \left(V_{k+1} - V_k \right)$$

$$
\begin{aligned}
&= \sum_{k=0}^{n-1}\big(\partial_t V(t_k)\Delta t + O(\Delta t^2)\big) \\
&= \sum_{k=0}^{n-1}\Big(\int_{t_k}^{t_{k+1}} \partial_t V(t_k)dt + O(\Delta t^2)\Big) \\
&= \int_0^T \partial_t V(t)dt + nO(\Delta t^2)\ .
\end{aligned}
$$

Because $n\Delta t = T$, $n\Delta t^2 = TO(\Delta t) \to 0$ as $n \to \infty$.

2.2. The Ito dV_t: The Fundamental Theorem may be stated $dV = \partial_t V dt$. This definition makes

$$\int_0^T dV_t = V(T) - V(0)\ . \tag{5}$$

We want to extend this to functions V_t that depend on W as well as t. For any adapted function, we define dV_t so that (5) holds. For example, if U_t is an adapted process and $V_T = \int_0^T U_t dW_t$, then $dV_t = U_t dW_t$ because that makes (5) hold. Ito's lemma is a statement of what makes (5) hold for specific adapted functions V_t.

2.3. First version: Our first version of Ito's lemma is a calculation of dV_t when $V_t = V(W_t, t)$ and V and W are one dimensional. The result is

$$dV_t = \partial_W V(W_t, t)dW_t + \frac{1}{2}\partial_W^2 V(W_t, t)dt + \partial_t V dt\ . \tag{6}$$

What's particular to stochastic calculus is the "Ito term" $\frac{1}{2}\partial_W^2 V(W_t, t)dt$. Even if we can't guess the precise form of the term, we know something has to be there. In the special case $V_t = V(W_t)$, the $\partial_t V dt$ term is missing. The guess $dV = \partial_W V dW_t$ would give (see (5)) $V(t) - V(0) = \int_0^T \partial_W V dW_t$. We know this cannot be correct: the right side is a martingale while the left side is not (see assignment 5, question 1). To make the martingale integral into the non martingale answer, we have to add a dt integral, which is why some term like $\frac{1}{2}\partial_W^2 V(W_t, t)dt$ is needed. A motivation for the specific form of the Ito term is the observation that it should vanish when V is a linear function of W.

2.4. Derivation, short time approximations: The defivation of Ito's lemma starts with the stochastic versions of the two short time approximations behind the Fundamental Theorem. For convenience, we drop all t subscripts and write ΔW for $W_{t+\Delta t} - W_t$. We have

$$
\begin{aligned}
&V(W_{t+\Delta t}, t+\Delta t) - V(W, t) = \\
&\qquad \partial_W V(W, t)\Delta W + \frac{1}{2}\partial_W^2 V(W, t)\Delta W^2 + \partial_t V(W, t)\Delta t + O(\Delta t^{3/2})\ .
\end{aligned}
$$

The other short time approximation is provided by assignment 7, question 3, applied to $\partial_W V$:

$$\int_t^{t+\Delta t} \partial_W V(W_s,s)dW_s = \partial_W V(W_t,t)\Delta W + \frac{1}{2}\partial_W^2 V(W,t)\big(\Delta W^2 - \Delta t\big) + O(\Delta t^{3/2}) \; .$$

For dt integrals, the result is simply

$$\int_t^{t+\Delta t} U(W_s,s)ds = U(W,t)\Delta t + O(\Delta t^{3/2} \; .$$

The error term is $O(\Delta t^{3/3})$ rather than $O(\Delta t^2)$ because W_t is not a Lipshcitz continuous function of t. We conbine these approximations with a little algebra ($\Delta W^2 = \Delta t + (\Delta W^2 - \Delta t)$, which might be considered the main idea of this section) gives

$$\begin{aligned}
\Delta V \quad = \quad & \int_t^{t+\Delta t} \partial_W V(W_s,t)dW_s \\
& + \int_t^{t+\Delta t} \left(\frac{1}{2}\partial_W^2 V(W_s,s) + \partial_t V(W_s,s)\right) ds \\
& + \partial_W^2 V(W,t)\big(\Delta W^2 - \Delta t\big) + O(\Delta t^{3/2}) \; .
\end{aligned}$$

As with the Fundamental Theorem, we apply this with $t = t_k$ (in the habitual notation) and sum over k, giving:

$$\begin{aligned}
V(T) - V(0) \quad = \quad & \int_0^T \partial_W V(W_t,t)dW_t \\
& + \int_0^T \left(\frac{1}{2}\partial_W^2 V(W_t,t) + \partial_t V(W_t,t)\right) dt \\
& + \sum_{k=0}^{n-1} \partial_W^2 V(W_k,t_k)\big(\Delta W_k^2 - \Delta t\big) + O(T\sqrt{\Delta t}) \; .
\end{aligned}$$

2.5. The non Newtonian step: The final step in deriving Ito's lemma has no anologue in the proof of Newton's Fundamental Theorem of calculus. We study the term

$$A = \sum_{k=0}^{n-1} \partial_W^2 V(W_k,t_k)\big(\Delta W_k^2 - \Delta t\big)$$

and show that $A \to 0$ as $\Delta t \to 0$ (actually, as $L \to \infty$ with $\Delta t = T/2^L$) almost surely. Previous experience might lead us to calculate $E[A_L^2]$. This follows a well worn path. We have the double sum expression:

$$E[A_L^2] = \frac{1}{4}\sum_{j,k} E\left[\big(\cdot\big)_j\big(\cdot\big)_k\right] \; .$$

The $j \neq k$ terms have expected value zero because (if $k > j$) $E[\Delta W_k^2 - \Delta t \mid \mathcal{F}_{t_k}] = 0$. We get a bound for the $j = k$ terms using $E[(\Delta W_k^2 - \Delta t)^2 \mid \mathcal{F}_{t_k}] = 2\Delta t^2$:

$$E\left[\partial_W^2 V(W_k, .t_k)^2 \big(\Delta W_k^2 - \Delta t\big)^2 \mid \mathcal{F}_{t_k}\right] \leq C \cdot \Delta t^2 \; .$$

Altogether, we get $E[A_L^2] \leq C\Delta t = C2^{-L}$, which implies that $A_L \to 0$ as $L \to \infty$, almost surely (see the next paragraph). This completes our proof of the first form of Ito's lemma, (6).

2.6. A Technical Detail: Here is a proof that uses the inequalities $E[A_L^2] \leq Ce^{-\beta L}$ for some $\beta > 0$ and proves that $A_L \to 0$ as $L \to \infty$ almost surely. The proof is an easier version of an argument used in the previous lecture. As in that lecture, we start with a observation, this time that $|A_L| \to 0$ as $L \to \infty$ if $\sum_{L=1}^{\infty} |A_L| < \infty$. Also, the sum is finite almost surely if it's expected value is finite. That is, if $\sum_{L=1}^{\infty} E[|A_L|] < \infty$. Finally, the Cauchy Schwartz inequality gives $E[|A_l|] \leq Ce^{-\beta L/2}$. Since this has a finite sum (over L), we get almost sure convergence $A_L \to 0$ as $L \to \infty$.

2.7. Integration by parts: In ordinary Newtonian (and Leibnitzian) calculus, the integration by parts identity is a consequence of the Fundamental Theorem and facts about differentiation (the Leibnitz rule). So let it be for Ito. For instance, integration by parts might lead to

$$\int_0^T t dW_t = TW_T - \int_0^T W_t dt \; . \tag{7}$$

We can check whether this actually is true by taking the Ito differential of tW_t:

$$\begin{aligned} d(tW_t) &= \partial_W (tW_t) dW_t + \frac{1}{2} \partial_W^2 (tW_t) dt + \partial_t (tW_t) dt \\ &= t dW_t + W_t dt \; . \end{aligned}$$

This implies that

$$TW_T = \int_0^T t dW_t + \int_0^T W_t dt \; ,$$

which is a confirmation of (7). We can get a more general version of the same thing if we apply the Ito differential to $f(t)g(W_t)$ (reader: do this).

2.8. Solutions of Ito SDEs: In view of the definition (2) the solution , the Ito SDE (1) is a relation among Ito differentials. We can also compute $dV(X_t)$ (or even $dV(X_t, t)$, which is more complicated but not harder) using the reasoning in paragraphs 2.4 and 2.5 above. I will breeze through the argument, commenting only on the differences. Some of the details are left to assignment 8. We can calculate

$$\Delta V(X) = \partial_X V(x_t) \Delta X_t + \frac{1}{2} \partial_X^2 V(X) \Delta X^2 + O(\Delta t^{3/2}) \; .$$

Also

$$\int_t^{t+\Delta t} \partial_X V(X_s) dX_s = \partial_X V(X_t)\Delta X_t + \frac{1}{2}\partial_X^2 V(X_t)\big(\Delta X^2 - \sigma(X_t)^2\Delta t\big) + O(\Delta t^{3/2}) \; .$$

The new feature is that $E[\Delta X^2] = \sigma(X_t)^2\Delta t + O(\Delta t^{3/2})$. After this, the derivation proceeds as before, eventually giving

$$dV(X_t) = \partial_X V(X_t) dX_t + \frac{1}{2}\partial_X^2 V(X_t)\sigma(X_t)^2 dt \; . \qquad (8)$$

2.9. The "Ito rule" $dW^2 = dt$: The first version of Ito's lemma can be summarized as using Taylor series calculations and neglecting all terms of higher than first order except for dW_t^2, which we replace by dt. You might think this is based on the approximation $\Delta W^2 \approx \Delta t$ for small Δt. The real story is a little more involved. The relative accuracy of the approximation $\Delta W^2 \approx \Delta t$ does not improve as $\Delta t \to 0$. Both sides go to zero, and at the same rate, but they do not get closer to each other in relative terms. In fact, the expected error, $E[|\Delta W^2 - \Delta t|]$, is also of order Δt. If $\Delta t = .1$ then ΔW^2 is just as likely to be .2 as .1, not really a useful approximation. The origin of Ito's rule is that ΔW^2 and Δt have the same expected value. For that reason, if we add up m ΔW^2 values, we are likely to get a number close to $m\Delta t$ if m is large. We might say $\int_a^b dW^2 = \int_a^b dt$, thinking that each side is make up of a large number (an infinite number) of tiny ΔW^2 or Δt values. Remember that for any Q, the Ito dQ is what you have to integrate to get Q. Integrating dW^2 gives the same result as integrating dt.

2.10. Quadratic variation: The informal ideas of the preceding paragraph may be fleshed out using the "quadratic variation" of a process. We already discussed the quadratic variation of Brownian motion. For a general stochastic process, X_t, the quadratic variation is

$$< X >_t = \lim_{\Delta t \to 0} \sum_{k=0}^{n} \big(X_{k+1} - X_k\big)^2 \; . \qquad (9)$$

3 Back to SDE

3.1. Solutions and approximate solutions of SDE's: It is possible to construct solutions to Ito differential equations using the integral formula (2) using "Picard iteration". You might remember a proof that solutions to ordinary differential equations (ODE) exist using this technique. Instead, we will give a proof based on the ODE proof given by the earlier French mathematician Cauchy. We write the forward Euler approximation and show that it converges as $\Delta t \to 0$. This has the advantage of being more concrete and leading to a practical computational algorithm.

Choose a small time step, Δt. Define $t_k = k\Delta t$. We will construct approximations $\overline{X}_k$ which are supposed to approximate X_{t_k}. The notation $\overline{X}$ here is not as used in statistics; it does not imply any averaging. It is true that $\overline{X}_k = X_{t_k}$, though we sill have $W_k = W_{t_k}$. We define $\Delta \overline{X}_k$ and ΔW_k to be the forward differences $\Delta \overline{X}_k = \overline{X}_{k+1} - \overline{X}_k$ and $\Delta W_k = W_{k+1} - W_k$. The forward Euler approximation to (1) is

$$\Delta \overline{X}_k = a(\overline{X}_k, t_k)\Delta t + \sigma(\overline{X}_k, t_k)\Delta W_k \ . \tag{10}$$

Our convergence theorem as $\Delta t \to 0$ is a more elaborate version of the convergence theorem for approximations to Ito integrals. To make the statement of the theorem simpler, we define $\overline{X}_t^{(\Delta t)}$ by linear interpolation from the discrete time values $\overline{X}_k$.

Theorem: Suppose the drift and volatility coefficients are Lipshitz continuous functions of x:

$$|a(x', t) - a(x, t)| \leq C\,|x' - x| \ ,$$

$$|\sigma(x', t) - \sigma(x, t)| \leq C\,|x' - x| \ ,$$

Then (with more thechnical hypotheses on how a and sigma depend on t), for any t,

$$\lim_{\Delta t \to 0} \overline{X}_t^{(\Delta t)} \to X_t$$

exists, is a continuous adapted function of t, and satisfies (2).

Remarks: 1. We should expect that different approximations to (1) will lead to different results, even in the limit $\Delta t \to 0$. There is an example of this below. 2. If there is no drift then the approximations are clearly discrete time martingales (warning: $\overline{X}_t$ is not a continuous time martingale). The limit will also be a martingale in that case.

3.2. Practical paths: Generally, we find solutions of SDE's numerically rather than analytically. The forward Euler method (10) leads to a practical computational procedure. The only question is how we get the Brownian motion increments. What we need are independent increments ΔW_k that are mean zero normals with variance Δt. The standard way to do this is to use a random number generator to make Z_k, iid standard normal random variables, then take $\Delta W_k = \sqrt{\Delta t} Z_k$. These are statistically indistinguishable from Brownian motion increments. The computational version, then, is

$$\overline{X}_{k+1} = \overline{X}_k + a(\overline{X}_k, t_k)\Delta t + \sigma(\overline{X}_k, t_k)\sqrt{\Delta t} Z_k \ . \tag{11}$$

If $m > 1$, the m components of Z_k will be iid standard normals, each of which is multiplied by $\sqrt{\Delta t}$ in the forward Euler approximation.

3.3. Basic Monte Carlo: We might want to make sample paths just to see what they look like. In this case (11) is just what we need. More likely, we are trying to compute $E[F(X)]$, where F is some "functional" of the diffusion path, X_t, for $0 \leq t \leq T$. Simple examples are $F(X) = V(X_T)$ (final payout),

$F(X) = \int_0^T V(X_t)dt$ (continuous time payout), $F(X) = \mathbf{1}_A$, where A is the event $X_t \geq r$ for some $t \leq T$ (hitting probability), etc. A Monte Carlo estimate of $\mu = E[F(X)]$ is the average of many independent samples

$$\widehat{\mu} = \frac{1}{N} \sum_{j=1}^{N} F(\overline{X}(j)) \ , \tag{12}$$

where each of the approximate paths $\overline{X}(j)$ is generated using (11). If there are n steps in each path $T = n\Delta t$, and N paths, then the total work is proportional to nN. There are two sources of error. Bias is the fact that $E[F(X)] \neq E[F(\overline{X})]$ either because F is not computed exactly or because $\overline{X}$ is not an exact sample path. We reduce bias by taking a larger n. Statistical error is the fact that $\widehat{\mu} \neq E[F(\overline{X})]$. We reduce statistical error by taking a larger N. Note that if we take $n \to \infty$ with fixed N, or $N \to \infty$ with fixed n, we do not converge to the exact answer $E[F(X)]$.

3.4. Advanced Monte Carlo: As long as we are trying to evaluate $\mu = E[F(X)]$, the basic Monte Carlo algorithm of the previous paragraph is just the opening bid. Since the number we want, μ, is itself not random, there is in principle no reason to use random numbers to estimate it. For example, we may be able to evaluate μ analytically. There may be other random variables, Y, with $E[Y] = \mu$ but smaller variance than $F(X)$. A simple illustration of this is the method of "antithetic variates" that is effective for short time simulations. In this method, we use the Brownian motion path W_t (or the numbers Z_k), to generate a path, $\overline{X}$, then we use the exact opposite path $-W_t$ (or the numbers $-Z_k$) to generate $\overline{X}'$. Clearly

$$E[F(\overline{X})] = E[F(\overline{X}')] = \frac{1}{2}(E[F(\overline{X})] + E[F(\overline{X}')] \ .$$

If the maps $W \mapsto X \mapsto F$ were linear, the average above would always be exactly zero, resulting in a zero variance (i.e. exact) estimator.

3.5. Avoiding Monte Carlo by PDE: In many cases there is a PDE whose solution determines $E[F]$. Often a numerical solution of the PDE will be faster and more accurate than Monte Carlo, even with fancy tricks.

3.6. Convergence of the forward Euler method: Now, for the last time, we want to compare the Δt approximation to the $\Delta t/2$ approximation and show that

$$E\left[\left|\overline{X}_n^{(\Delta t)} - \overline{X}_{2n}^{(\Delta t/2)}\right|\right] \leq Const \cdot \Delta t^{\alpha} \ , \tag{13}$$

which will show that $n_L = 2^L$ and $n_L \Delta t_L = T$ as $L \to \infty$ the limit of the forward Euler approximations exists. This is more complicated than proving convergence of the Ito integral because now the "stability" of the forward Euler approximation comes in, not just the "consistency" of the finite Δt approximation.

3.7. Consistency of the forward Euler approximation: We want to show that the $\overline{X}_{2k}^{(\Delta t/2)}$ are close to the $\overline{X}_k^{(\Delta t)}$ for $k \leq n$. The technique is to show that the $\overline{X}_{2k}^{(\Delta t/2)}$ almost satisfy the recurrence relations (10) that define the $\overline{X}_k^{(\Delta t)}$. This is the consistency step. The stability step below is to show that if the $\overline{X}_{2k}^{(\Delta t/2)}$ almost satisfy the $\overline{X}_k^{(\Delta t)}$, then the $\overline{X}_{2k}^{(\Delta t/2)}$ are almost the $\overline{X}_k^{(\Delta t)}$.

For this argument we simplify the notation. We write X_k for $\overline{X}_k^{(\Delta t)}$, Y_k for $\overline{X}_{2k}^{(\Delta t/2)}$, and $\tilde{Y}_k$ for $\overline{X}_{2k+1}^{(\Delta t/2)}$. We also invoke the Brownian bridge construction, defining Z_k by $\Delta W_k = \Delta t Z_k$ and $\tilde{Z}_k$ by

$$\begin{aligned} W_{t_{k+1/2}} - W_{t_k} &= \frac{1}{2}\Delta W_k + \frac{1}{2}\sqrt{\Delta t}\tilde{Z}_k \\ &= \frac{1}{2}\sqrt{\Delta t}Z_k + \frac{1}{2}\sqrt{\Delta t}\tilde{Z}_k \end{aligned}$$

The Brownian bridge construction (assignment 6, question 1) shows that Z_k and $\tilde{Z}_k$ are iid standard normals, and, of course, independent for different k values.

We calculate the results of two forward Euler time steps of size $\Delta t/2$ using the simplified notation. First we have

$$\tilde{Y}_k = Y_k + a(Y_k, t_k)\Delta t/2 + \frac{1}{2}\sigma(Y_k, t_k)\sqrt{\Delta t}Z_k + \frac{1}{2}\sigma(Y_k, t_k)\sqrt{\Delta t}\tilde{Z}_k \ .$$

Then

$$Y_{k+1} = \tilde{Y}_k + a(\tilde{Y}_k, t_{k+1/2})\Delta t/2 + \frac{1}{2}\sigma(\tilde{Y}_k, t_{k+1/2})\sqrt{\Delta t}Z_k \ .$$

We want to see what recurrence relation the Y_k satisfy. This means eliminating the $\tilde{Y}_k$. The drift term does not take much work. We write

$$a(\tilde{Y}_k, t_{k+1/2}) = a(Y_k, t_k) + R_k \ ,$$

where

$$|R_k| \leq C\left(\left|\tilde{Y}_k\right| + \Delta t/2\right) \ ,$$

Because we assumed a is Lipschitz continuous. We have to use more information about the diffusion term (we will see why):

3.8. An example, $dX = XdW$. In this example we see the generation of the Ito correction and what happens if we use a discretization inconsistent with the Ito rule that $E[dX_t \mid \mathcal{F}_t] = 0$. The forward Euler approximation here is

$$\overline{X}_{k+1} = \overline{X}_k + \overline{X}_k\Delta W_k = \big(1 + \Delta W_k\big)\overline{X}_k \ .$$

Using the initial condition $X_0 = 1$, this becomes

$$\overline{X}_n = \prod_{k=0}^{n-1}\big(1 + \Delta W_k\big) \ .$$

There is a standard trick for estimating products such as this, convert it into a sum in the exponent. For this purpose, we use the Taylor series calculation $1+\epsilon = e^{\epsilon - \frac{1}{2}\epsilon^2 + O(\epsilon^3)}$. This gives us $1+\Delta W_k = \exp(W_k - \frac{1}{2}W_k^2 + O(\Delta t^{3/2})$. Substitute this into the product, use $e^a e^b = e^{a+b}$, and you will get

$$\overline{X}_n = \exp\left(\sum_{k=0}^{n-1} \Delta W_k - \frac{1}{2}\sum_{k=0}^{n-1} \Delta W_k^2 + O(T\Delta t^{1/2})\right) .$$

By now we recognize that $\sum_{k=0}^{n-1} \Delta W_k = W_n = W_T$, and $\frac{1}{2}\sum_{k=0}^{n-1} \Delta W_k^2 \to T/2$ as $n \to \infty$. This shows that $\overline{X}_n \to e^{W_T - T/2}$ as $\Delta t \to 0$. This is the Ito solution, as it should be since we are using the forward Euler method.

The backward Euler method would be

$$\overline{X}_{k+1} = \overline{X}_k + X_{k+1}\Delta W_k .$$

The difference is that the multiplier of ΔW_k is not the $\mathcal{F}_{t_k}$ measurable value, $\overline{X}_k$, but a number, $\overline{X}_{k_1}$, that depends on ΔW_k. In fact, in order to determine $\overline{X}_{k+1}$ from the backward Euler equation, we have to solve for $\overline{X}_{k+1}$:

$$\overline{X}_{k+1} = \frac{1}{1-\Delta W_k}\overline{X}_k .$$

In order for $\overline{X}_{k+1}$ not to be defined, ΔW_k has to be so large that the probability of this happening is of the order of $e^{-1/\Delta t}$, a forgettable probability. Expanding in exponentials now gives

$$\overline{X}_n = \exp\left(\sum_{k=0}^{n-1} \Delta W_k + \frac{1}{2}\sum_{k=0}^{n-1} \Delta W_k^2 + O(T\Delta t^{1/2})\right) .$$

This converges to $e^{W_T + T/2}$, which is the wrong answer.