Stochastic Calculus, Courant Institute, Fall 2020
http://www.math.nyu.edu/faculty/goodman/teaching/StochCalc2020/index.html

Week 2

Jonathan Goodman, October, 2020

1 Introduction to Ito calculus

This class is an introduction to the Ito calculus. The technical highlights are the Ito integral and Ito's lemma. Ordinary differential equations, ODEs, are dynamical models of systems. The tools for ODEs are differentiation and the Riemann (ordinary) integral. Stochastic differential equations are models for systems with noise. Ito's lemma is for finding the differential of a noisy function. The Ito integral is for accumulating the effect of noise.

For this class, Brownian motion will be denoted W_t, for *Wiener process*, after Norbert Wiener. We usually use a *standard* Brownian motion, which has $W_0 = 0$, and $\mathrm{E}[W_t] = 0$, and $\mathrm{var}(W_t) = t$. The *Ito integral* is a stochastic process, written

$$X_t = \int_0^t a_s dW_s \ . \tag{1}$$

This is analogous to the indefinite integral in ordinary calculus. The *integrand* is another random process a_s, which must be *non-anticipating*, as defined in Section 3.

Section 2 motivates the Ito integral by describing what it would be in discrete time. This has some of the main ideas of stochastic calculus but without the mathematical technicalities. Discrete time is easier to explain and sometimes compute with, but the *continuous time limit* has simpler formulas and is easier to reason with, once you become comfortable with it.

The rest of this class is about the continuous time limits, which are limits as the time increment goes to zero, $\Delta t \to 0$. We have to distinguishing between *small terms* and *tiny terms*. Tiny terms are small enough to ignore in the continuous time limit. Small terms are too big to ignore.

2 Discrete time strategies

We create stochastic models to design and optimize strategies. Investors (investment advisors, actually) and traders use stochastic models of unknown future prices to create dynamic investment and trading strategies. Random walk models illustrate some conceptual ideas without the technicality of the continuous time limit. Let S_n be the price of an asset (something you can buy or sell) at time n, for $n = 0, 1, \cdots$. We put ourselves at time $n = 0$ and imagine that S_0 is known to us, but future prices are not known exactly. We model the price change from time n to $n+1$ as a random variable Y_n. The price process is the random walk

$$S_n = \sum_{k=0}^{n-1} Y_k \ .$$

This may be written as

$$Y_n = S_{n+1} - S_n \ .$$

Consider a simple *trading strategy* that consists of owning a_n amount of the asset at time n. The price changes by amount Y_n in going to time $n+1$, so your gain or loss is $a_n Y_n$. The total gain up to time n is

$$X_n = \sum_{k=0}^{n-1} a_k Y_k \ . \tag{2}$$

It will be useful later to write this in the form

$$X_n = \sum_{k=0}^{n-1} a_k (S_{k+1} - S_k) \ . \tag{3}$$

A strategy is *non-anticipating* (also called *adapted* or *progressively measurable*) if a_k is determined by $(Y_0, \ldots, Y_{k-1})$ only. This way, you put a "bet" on Y_k without knowing what Y_k will be. A strategy starts with an initial bet a_0 that does not depend on the "returns" Y_k at all. After that, the bets are $a_1(Y_0)$, $a_2(Y_0, Y_1)$, and so on up to $a_k(Y_0, \ldots, Y_{k-1})$. This is "decision making under uncertainty". You decide using a probability model of the future, but not a forecast of the future. You have a probability distribution of Y_k, but you do not predict the value of Y_k. A non-anticipating trading strategy a_k produces a sequence of random variables X_n, The questions are: (1) What random variables can be achieved in this way? (2) Which of these do we like the best? (3) What strategy give the X_n we like the best?

A particularly important case is when the price process S_n is a simple random walk. This means that the steps Y_k satisfy these three properties:

$$\mathrm{E}[\,Y_k] = 0 \ , \ \ \mathrm{var}(\,Y_k) = \sigma^2 \ , \ \ Y_k \text{ are i.i.d.} \tag{4}$$

In that case, the total gain process (2) has two properties that carry over to the continuous time versions in Section 3. These are the *Doob martingale theorem* and the *Ito isometry formula*.

The Doob martingale theorem is named for mathematician Joseph Doob. We call a sequence $(X_0, X_1, \ldots, X_n)$ a *discrete path*. What's discrete is time, not necessarily X_k. A random process X_n is a *martingale* if its increments have zero conditional expectation. More precisely, for any discrete path $(x_0, \ldots, x_n)$,

$$\mathrm{E}[\,X_{n+1} - X_n \mid (X_0, \ldots, X_n) = (x_0, \ldots, x_n)] = 0 \ . \tag{5}$$

The distribution of the *increment* $S_{n+1} - S_n$ can depend on the path in other ways. For example you could have (if you wanted to)

$$\begin{aligned}\text{var}(\, X_{n+1} - X_n \mid (x_0, \ldots, x_n)) &= \text{E}\big[\, (X_{n+1} - x_n)^2 \mid (X_0, \ldots, X_n) = (x_0, \ldots, x_n)\big] \\ &= x_n^2 \ .\end{aligned}$$

If the steps Y_k satisfy (4) then the non-anticipating strategy $a_k = \frac{1}{\sigma} X_k$ in 2 would do it.

The martingale process has zero mean, but a zero mean process might not be a martingle. You can average the zero conditional mean (5) over all possible paths $(x_0, \ldots, x_n)$. The result is

$$\text{E}[\, X_{n+1} - X_n] = 0 \ .$$

This implies that

$$\text{E}[X_n] = 0 \ , \tag{6}$$

because $X_0 = 0$ (you have gain or loss before you start investing) and the expected changes are all zero. It takes more than this to be a martingale. For that, the expected increment is zero no matter what path you took up to time n. Exercise 1 has an example to illustrate the difference.

The "Doob martingale theorem" states that the gain process (2) is a martingale if the one-period returns Y_k satisfy (4). [There are many facts called "Doob martingale theorem", including this one.] The proof is simple. The gain in one period is $a_n Y_n$, with Y_n independent of anything that came before. The "weight" a_n can depend on $X_0, \ldots, X_n$. This allows Y_k to influence a_n if $n > k$, but it still leaves Y_n independent. Therefore

$$\begin{aligned}\text{E}[\, X_{n+1} - X_n \mid (X_0, \ldots, X_n) = (x_0, \ldots, x_n)] &= \text{E}[\, Y_{n+1}\, a_n(x_0, \ldots, x_n)] \\ &= 0 \ .\end{aligned}$$

The point here is that at time n you don't know Y_{n+1}.

Doob's martingale theorem has been interpreted informally as saying that there is no benefit in fancy trading strategies. I disagree with that view. Active strategies cannot turn a martingale into an expected profit, but they can give lower risk for the same expected return. This is discussed more in the class of Week 5.

The *Ito isometry formula* for this situation (fancier ones are coming) is

$$\text{E}\big[\, X_n^2\big] = \sigma^2 \sum_{k-0}^{n-1} \text{E}\big[\, a_k^2\big] \ . \tag{7}$$

In math, an *isometry* is a mapping that doesn't change the sizes of things or the distances between them. In this case, the mapping is the sum (2) that takes the trading strategy weights a_n and produces the total return X_n. The size on the left of the Ito isometry formula is the variance of X_n. The size on the right is the expected squares of the weights. Actually, the "size" is usually taken to be the square root of a sum of squares. You could take the square roots of both sides of the isometry formula to get an equation about sizes rather than squares of sizes.

Here is a more general "Doob martingale theorem" that has the same proof we just gave. The Y_k don't have to be independent or identically distributed. They just need to be *martingale differences*, or martingale *increments*. To explain this, we change our hypothesis about S_n and just assume S_n is a martingale. Then the numbers $Y_n = S_{n+1} - S_n$ are *martingale differences*. The sums (2) or (3) may be interpreted as the total gain or loss when the asset price is any martingale. The theorem is that if S_n is a martingale and if the weights a_n are adapted, then X_n is a martingale too. The proof is the same – look and see.

3 Strategies in continuous time

Last week we saw that Brownian motion is continuous time limit of random walk, given the proper scalings. The Ito integral (1) is the continuous time limit of the "strategy" sum (3) in the same scaling. It is an idealization of a situation where the price changes happen quickly and the strategy changes just as fast. Suppose that W_t is standard Brownian motion and that it represents the price of an asset at time t. Choose a small Δt that represents the time scale on which the price can be reported and the strategy can change. This divides time into time periods labelled by starting times $t_k = k\Delta t$. At time t_k, the agent places a "bet" of size a_{t_k} on the asset. The price change over the next period is $W_{t_{k+1}} - W_{t_k}$. The profit or loss is

$$a_{t_k}\left(W_{t_{k+1}} - W_{t_k}\right) .$$

The total profit up to time t is

$$X_t^{\Delta t} = \sum_{t_k < t} a_{t_k}\left(W_{t_{k+1}} - W_{t_k}\right) . \tag{8}$$

The *Ito integral* with respect to Brownian motion is the limit as $\Delta t \to 0$

$$X_t = \int_0^t a_s\, dW_s = \lim_{\Delta t \to 0} \sum_{t_k < t} a_{t_k}\left(W_{t_{k+1}} - W_{t_k}\right) . \tag{9}$$

The continuous time limit $\Delta t \to 0$ is like the continuous time limit of the Riemann (ordinary) integral

$$\int_0^t b_s\, ds = \lim_{\Delta t \to 0} \sum_{t_k < t} b_{t_k} \Delta t . \tag{10}$$

The Riemann integral limit exists as long as b_t is a continuous function of t.

The Ito integral limit (9) exists if a_t is continuous and adapted. *Adapted* means that a_t can be random, but the randomness is a function of the path up to time t. This means that there is a function $A(w_{[0,t]}, t)$ so that

$$a_t = A(W_{[0,t]}, t) \; .$$

For example, you could ignore the random part and take $a_t = t^2$. The Ito integral would be the random variable

$$X_t = \int_0^t s^2 \, dW_s \; .$$

Another example is $a_t = W_t$. This example is worked out in Section 4. You could take $a_t = X_t$. This choice is non-anticipating (adapted) because $1 + X_t$ depends on W_s for $s \leq t$. The formula is

$$X_t = \int_0^t (1 + X_s) \, dW_s \; .$$

Integrals like this come up in solutions of stochastic differential equations we will see in Week 4.

The limit (9) is a sequence of random variables $X_t^{\Delta t}$ converging to another random variable X_t. There are several kinds of convergence for random variables like this, including convergence *in distribution*, convergence *in probability*, and convergence *almost surely*. Convergence in distribution means that the probability distribution of the random paths $X_t^{\Delta t}$ converges to the probability distribution of the path X_t. This says that simulating the process $X_t^{\Delta t}$ gives results similar to the theoretical process X_t that we cannot simulate exactly. Convergence in probability is the statement that for every $\epsilon > 0$

$$\Pr\big(\big| X_t^{\Delta t} - X_t \big| > \epsilon \big) \;\to\; 0 \;\text{ as }\; \Delta t \to 0 \; .$$

This would be natural if you were trying to estimate X_t using $X_t^{\Delta t}$. It might be less natural here because there usually is no independent way to estimate X_t. Almost sure convergence says that "with probability 1" the limit formula (9) is true. The Week 6 class has more to say about the phrase "with probability 1", which is what probability people mean by "almost surely". Almost sure convergence is both the hardest to prove theoretically and the hardest to check in computations. You would have to do a sequence of simulations with decreasing Δt. For convergence in probability, you do one simulation with a single Δt. The smaller Δt is, the less likely you are to be off by more than ϵ. For almost sure convergence, you have to be unlikely ever to be wrong by more than ϵ as $\Delta t \to 0$.

There is a "Doob martingale theorem" for the Ito integral (9). The theorem is that if a_t is non-anticipating then X_t is a martingale. The definition of martingale in continuous time is X_t is a martingale with respect to the Brownian motion path W if

$$\mathrm{E}\big[\, X_{t+s} \mid W_{[0,t]} \big] = X_t \; . \tag{11}$$

In this statement, $W_{[0,t]}$ is the Brownian motion path on the time interval $[0,t]$. Part of the statement in (11) is that X_t is a function of $W_{[0,t]}$. This is true if X_t is an Ito integral with respect to W. This definition is a little different from the definition of martingale in discrete time. In discrete time, the definition involved X_n and X_{n+1}, This is the discrete time process X at time n and the next time $n+1$. In continuous time, there is no "next time". That's why the definition (11) refers to all times $s > t$.

The proof of this version of Doob's theorem is just from definitions and limits. The Ito integral is the continuous time limit ($\Delta t \to 0$) of a discrete time sum (8). Therefore, $X_t^{\Delta t}$ is a martingale, as was explained in Section 2. The full mathematical proof is a little involved, because $X_t^{\Delta t}$ is a martingale only at the discrete times t_k. These get more finely spaced in the continuous time limit. It is a basic technical theorem in probability that if a discrete time martingale converges in this way, the limit is a continuous time martingale.

The continuous time Ito isometry formula for this kind of Ito integral is

$$\mathrm{E}\left[\, X_t^2 \right] = \int_0^t \mathrm{E}\left[\, a_s^2 \right] \, ds \; . \tag{12}$$

You prove this by using the discrete time Ito isometry theorem (7) and taking the continuous time limit. The parameter σ^2 in the discrete time formula is the variance of the martingale difference. The martingale difference for $X_t^{\Delta t}$ is $\Delta W_k = W_{t_{k+1}} - W_{t_k}$. The variance of that is

$$\sigma^2 = \mathrm{var}(\Delta W_k) = \mathrm{E}\Big[\, (W_{t_k+\Delta t} - W_{t_k})^2 \Big] = \Delta t \; .$$

Therefore the discrete time formula (7) implies that

$$\mathrm{E}\Big[\, \big(X_{t_n}^{\Delta t}\big)^2 \Big] = \Delta t \sum_{k=0}^{n-t} \mathrm{E}\left[\, a_{t_k}^2 \right] \; .$$

In the continuous time limit, the right side in the sum converges to the Riemann integral on the right side of the continuous time isometry formula (12).

It is important to keep in mind that the approximation formula (8) requires ΔW_k to be in the future of t_k.

4 An example

The example $a_t = W_t$ illustrates many features of the Ito integral. The calculation that leads to (17) is one of the important ideas in Ito's lemma in Section 5. The calculations show what can go wrong if you make an approximation to the Ito integral. Suppose a_t is a non-anticipating strategy. It is possible that $a_{t_{k+1}}$ "knows" the price change over the period $W_{t_{k+1}} - W_{t_k}$. A trading "strategy" (not a strategy you can use because it knows the future) using the future a value might be

$$B_t^{\Delta t} = \sum_{t_k < t} a_{t_{k+1}} (W_{t_{k+1}} - W_{t_k}) \; . \tag{13}$$

You might think using $a_{t_{k+1}}$ instead of a_{t_k} is not serious because it only looks Δt into the future. At the end of this section, we calculate an example that shows that even this small change can give a different continuous time limit.

We find an explicit formula for

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

This a_t is adapted because it depends on the path W_s for $s \leq t$. We will see that the limit (8) exists. With the formula for X_t, we will be able to verify that X_t is a martingale (Doob's theorem) and the Ito isometry formula. We will see that other ways to define the integral, ways that give the same answer for the Riemann integral (the ordinary integral from calculus) are not equivalent for the Ito integral and give different limits. It is crucial to put $\Delta W = W_{t_{k+1}} - W_{t_k}$ in the future of a_{t_k}. From the strategy point of view, it is important that at you make the bet at time t_k when you do not know whether ΔW is positive or negative.

We simplify the calculations by writing W_k instead of W_{t_k}. In this notation, the approximation (8) is

$$X_t^{\Delta t} = \sum_{t_k < t} W_k \left(W_{k+1} - W_k\right) \ .$$

The trick for evaluating this (which is in all the books) is the clever formula

$$W_k = \frac{1}{2}\left(W_{k+1} + W_k\right) - \frac{1}{2}\left(W_{k+1} - W_k\right) \ . \tag{14}$$

Using this, we get

$$X_t^{\Delta t} = \frac{1}{2}\sum_{t_k<t}\left(W_{k+1} + W_k\right)\left(W_{k+1} - W_k\right) - \frac{1}{2}\sum_{t_k<t}\left(W_{k+1} - W_k\right)^2 \ . \tag{15}$$

For the first sum on the right, note that $\left(W_{k+1} + W_k\right)\left(W_{k+1} - W_k\right) = W_{k+1}^2 - W_k^2$. The largest k value in the sum is $n_t = \max\left\{k | t_k < t\right\}$. The first sum is

$$\begin{aligned}\sum_{t_k<t}\left(W_{k+1} + W_k\right)\left(W_{k+1} - W_k\right) &= \sum_{t_k<t}\left(W_{k+1}^2 - W_k^2\right) \\ &= \frac{1}{2}W_{n_t}^2 \ .\end{aligned}$$

This converges to $\frac{1}{2}W_t^2$, because $|t - t_{n_t}| \leq \Delta t$.

The other term involves the *quadratic variation*

$$Q_t^{\Delta t} = \sum_{t_k<t}\left(W_{k+1} - W_k\right)^2 \ .$$

We evaluate the $\Delta t \to 0$ limit by calculating the mean and variance of $Q_t^{\Delta t}$. The mean has a clear limit and the variance goes to zero as $\Delta t \to 0$. This gives the limit *in probability* of Q. [Limit in probability was defined in Week 1.] Since ΔW is an increment of Brownian motion over a time interval of length Δt, the variance is equal to Δt. Therefore:

$$\begin{aligned}
\mathrm{E}\left[\, Q_t^{\Delta t} \right] &= \sum_{t_k<t} \mathrm{E}\left[\, (W_{k+1} - W_k)^2 \right] \\
&= \sum_{t_k<t} \Delta t \\
&= t_{n_t} \ .
\end{aligned}$$

This shows that $\mathrm{E}\left[\, Q_t^{\Delta t} \right] \to t$ as $\Delta t \to 0$.

For the variance, we subtract the mean of Q, which is $t_{n_1} = \sum \Delta t$, which leads to

$$\begin{aligned}
\mathrm{var}\left(\, Q_t^{\Delta t} \right) &= \mathrm{E}\left[\, \left(Q_t^{\Delta t} - t_{n_t}\right)^2 \right] \\
&= \mathrm{E}\left[\left(\sum_{t_k<t} \left[(W_{k+1} - W_k)^2 - \Delta t \right] \right)^2 \right] \\
\mathrm{var}\left(\, Q_t^{\Delta t} \right) &= \mathrm{E}\left[\left(\sum_{t_k<t} R_k \right)^2 \right] \qquad (16)
\end{aligned}$$

with

$$R_k = (W_{k+1} - W_k)^2 - \Delta t \ .$$

The right side of (16) is the square of a sum. You can write it as a double sum to put the expectation on each term:

$$\left(\sum_{t_k<t} R_k \right)^2 = \sum_{t_k<t} \sum_{t_j<t} R_k R_j \ .$$

For example, if there are two terms in the square, the expansion has four terms:

$$(R_1 + R_2)^2 = R_1^2 + R_1 R_2 + R_2 R_1 + R_2^2 \ .$$

You could combine the middle terms, $R_1 R_2 + R_2 R_1 = 2 R_1 R_2$. But it is not convenient to do so. It is simpler not to use the fact that $R_k R_j = R_j R_k$. With this, (16) becomes

$$\mathrm{var}\left(\, Q_t^{\Delta t} \right) = \sum_{t_k<t} \sum_{t_j<t} \mathrm{E}[\, R_k R_j] \ .$$

We use the *independent increments* property from Week 1 to see that R_k and R_j are independent of $j \neq k$. This is because they involve the increments of Brownian motion over disjoint intervals of time. Furthermore,

$$\mathrm{E}[\, R_k] = \mathrm{E}\left[\, (W_{k+1} - W_k)^2 \right] - \Delta t = 0 \ .$$

Therefore, if $k \neq j$, we have $\mathrm{E}[\, R_k R_j] = 0$. We call the terms $j \neq k$ the *off diagonal* terms, because $j \neq k$ corresponds to an off-diagonal entry in a matrix. Only the diagonal terms ($k = j$) are different from zero, and they are all equal.

$$\mathrm{var}\big(\, Q_t^{\Delta t}\big) = \sum_{t_k < t} \mathrm{E}\big[\, R_k^2\big] = n_t \mathrm{E}\big[\, R_0^2\big] \; .$$

The final part of this calculation is

$$\begin{aligned}
\mathrm{E}\big[\, R_0^2\big] &= \mathrm{E}\big[\, R_0^2\big] \\
&= \mathrm{E}\left[\left((W_1 - W_0)^2 - \Delta t\right)^2\right] \\
&= \mathrm{E}\Big[(W_1 - W_0)^4\Big] - 2\mathrm{E}\Big[(W_1 - W_0)^2\Big]\, \Delta t + \Delta t^2 \\
&= 2\Delta t^2 \; .
\end{aligned}$$

The last line uses the formula for the fourth moment of a mean zero Gaussian. If $Y \sim \mathcal{N}(0, \sigma^2)$, then $\mathrm{E}\big[\, Y^4\big] = 3\sigma^4$. In our case $W_1 - W_0 = W_1$ is Gaussian with variance $\sigma^2 = \Delta t$, so $\mathrm{E}\big[\, W_1^4\big] = 3\Delta t^2$.

As a final trick, use the formula we have used already:

$$\sum_{t_k < t} \Delta t = t_{n_t} < t \; .$$

Therefore

$$\begin{aligned}
\mathrm{var}\big(\, Q_t^{\Delta t}\big) &= \sum_{t_k < t} \Delta t^2 \\
&= \left(\sum_{t_k < t} \Delta t\right) \Delta t \\
&= t_{n_t} \Delta t \\
&\leq t \Delta t \; .
\end{aligned}$$

This shows the expected value of $Q_t^{\Delta t}$ converges to t and the variance converges to zero. Therefore

$$Q_t^{\Delta t} \to t \; , \;\; \text{in probabilty, as } \Delta t \to 0 \; . \tag{17}$$

With all this, we can see that

$$\int_0^t W_s \, dW_s = \frac{1}{2} W_t^2 - \frac{1}{2} t \; . \tag{18}$$

We will return to this formula many times, and to the calculations that led to it.

The anticipating (not non-anticipating) approximation (13) for this example gives

$$B_t^{\Delta t} = \sum_{t_k < t} W_{k+1} \left(W_{k+1} - W_k\right) \; .$$

The trick (14) gives something that differs from (15) by a minus sign

$$B_t^{\Delta t} = \frac{1}{2}\sum_{t_k<t} \left(W_{k+1} + W_k\right)\left(W_{k+1} - W_k\right) + \frac{1}{2}\sum_{t_k<t} \left(W_{k+1} - W_k\right)^2 \ .$$

The limit of this is calculated in the same way:

$$B_t = \frac{1}{2}W_t^2 + \frac{t}{2} \ .$$

Unlike the Ito answer (18), this bad answer is not a martingale.

5 Ito's lemma for Brownian motion

Ito's lemma is an expression for the small change in a function of Brownian motion in a small increment of time. Consider a process that is a function of W_t and t,

$$X_t = f(W_t, t) \ .$$

This process involves the function $f(w,t)$ with partial derivatives $\partial_w f$, etc. Ito's lemma is the formula

$$dX_t = \partial_w f(W_t,t)\, dW_t + \left(\partial_t f(W_t,t) + \frac{1}{2}\partial_w^2 f(W_t,t) \right) dt \ . \tag{19}$$

This is a convenient but informal expression.

The more formal mathematical expression is the integral version. The idea is that dX_t is the change of X in time dt. If you add up, or integrate, all these small changes, you get the total change:

$$\int_{T_1}^{T_2} dX_t = X_{T_2} - X_{T_1} \ . \tag{20}$$

Ito's lemma is a formula for the left side, which is

$$\int_{T_1}^{T_2} \partial_w f(W_t,t)\, dW_t \ + \int_{T_1}^{T_2} \left(\partial_t f(W_t,t) + \frac{1}{2}\partial_w^2 f(W_t,t) \right) dt = X_{T_2} - X_{T_1} \ . \tag{21}$$

Ito's lemma is a version of the chain rule appropriate for Brownian motion. To see this, suppose U_t is a differentiable function of t. Then the ordinary chain rule in the language of college calculus is

$$\frac{d}{dt} f(U_t,t) = \partial_u f(U_t,t)\frac{dU_t}{dt} + \partial_t f(U_t,t) \ .$$

In the informal language of differentials, this would be

$$df(U_t,t) = \partial_u f(U_t,t)\, dU_t + \partial_t f(U_t,t)\, dt \ .$$

The Ito expression (19) includes the second derivative term involving $\partial_w^2 f$. This is related to the informal "derivation" of Ito's lemma in which you expand f to second order in dW and then use the *Ito rule*

$$(dW_t)^2 = dt \; . \tag{22}$$

The expansion and "Ito rule" calculation is like this. We simplify by leaving out arguments when they are at time t. For example, we write f for $f(W_t, t)$ and $\partial_w f$ for $\partial_w f(W_t, t)$.

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

The derivation of Ito's lemma (21) has to explain why these Taylor series terms are exactly the ones you need (all of these and no more), and the "Ito rule" 22.

Example. Here is the Ito's lemma approach to the example of Section 4. Consider the function $f(w, t) = \frac{1}{2}w^2$. This has derivatives

$$\begin{aligned} \partial_w f &= w \\ \partial_w^2 f &= 1 \\ \partial_t f &= 0 \; . \end{aligned}$$

From Ito's lemma in differential form (19) we calculate

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

We can integrate this, and we get

$$\frac{1}{2}W_t^2 = \int_0^t W_s dW_s + \frac{1}{2}t \; .$$

This is the formula (18) we had in Section 4. Another way to do this is to take $f(w, t) = \frac{1}{2}w^2 - \frac{1}{2}t$. In this case, the partial derivatives are

$$\begin{aligned} \partial_w f &= w \\ \partial_w^2 f &= 1 \\ \partial_t f &= -\frac{1}{2} \; . \end{aligned}$$

The answer (18) comes from integrating this.

Example. This calculation allows us to understand some surprising features of geometric Brownian motion. If $f(w,t) = e^w$, then the partial derivatives are

$$\begin{aligned}\partial_w f &= e^w \\ \partial_w^2 f &= e^w \\ \partial_t f &= 0 \ .\end{aligned}$$

Therefore,

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

A more general $f = e^{aw+bt}$ has partial derivatives

$$\begin{aligned}\partial_w f &= af \\ \partial_w^2 f &= a^2 f \\ \partial_t f &= bf \ .\end{aligned}$$

Therefore, if

$$X_t = f(W_t, t) = f = e^{aW_t + bt} \ ,$$

then

$$dX_t = aX_t\, dW_t + \left(b + \frac{1}{2}a^2\right) X_t\, dt \ .$$

The dt term disappears if we choose $b = -\frac{1}{2}a^2$. This gives the process

$$X_t = e^{aX_t - \frac{1}{2}a^2 t} \ ,$$

that satisfies $dX_t = aX_t dW_t$, or, in integral form,

$$X_T - X_0 = a \int_0^T X_t dW_t \ .$$

Exercise 4 explores this process. The Week 5 class has much more about geometric Brownian motion.

The proof of Ito's lemma starts with a discrete version of (20):

$$X_{T_2} - X_{T_1} \approx \sum_{T_1 \le t_k < T_2} \Delta X_k \ .$$

This uses the same notation system, with $\Delta X_k = X_{t_{k+1}} - X_{t_k}$. Some Taylor approximations give the local approximation that is something like the differential form of Ito's lemma (19)

$$\Delta X_k = a_k \Delta W_k + b_k \Delta t + R_k \ . \tag{23}$$

We add these up:

$$\begin{aligned}\sum_{T_1 \le t_k < T_2} \Delta X_k &= \sum_{T_1 \le t_k < T_2} a_k \Delta W_k + \sum_{T_1 \le t_k < T_2} b_k \Delta t + \sum_{T_1 \le t_k < T_2} R_k \\ &= \qquad S_1^{\Delta t} \qquad + \qquad S_2^{\Delta t} \qquad + \qquad S_3^{\Delta t} \ .\end{aligned}$$

This formula uses the informal notation $a_k = a(t_k)$, $b_k = b(t_k)$, and $\Delta W_k = W_{t_{k+1}} - W_{t_k}$. In the limit $\Delta t \to 0$, the left side sum converges to $X_{T_2} - X_{T_1}$. The right side sum $S_1^{\Delta t}$ converges to the Ito integral

$$S_1^{\Delta t} = \sum_{T_1 \leq t_k < T_2} a_k \Delta W_k \quad \to \quad \int_{T_1}^{T_2} a_t dW_t \ . \tag{24}$$

The second sum converges to the Riemann integral (the "ordinary" integral)

$$S_2^{\Delta t} = \sum_{T_1 \leq t_k < T_2} b_k \Delta t \quad \to \quad \int_{T_1}^{T_2} b_t dt \ . \tag{25}$$

The last sum vanishes in the limit:

$$S_3^{\Delta t} \to 0 \ , \quad \text{as } \Delta t \to 0 \ .$$

The terms with non-zero limits (24) and eqrefS2 converge to the integrals on the right side of (21).

There is some informal language for this. When Δt is small, all the terms on the right of the expression (23) for ΔX_k are small. The term $a_k \Delta W_k$ is small, but when you add them together in the sum (24) the answer is not small. As $\Delta t \to 0$ the number of terms in the sum goes to infinity in a way that the limit is finite. The terms $b_k \Delta t$ are small in the same sense. There are enough small terms $b_k \Delta t$ in the sum (25) to get a non-zero answer as $\Delta t \to 0$. The terms R_k are smaller than small, so we call them *tiny*. Being tiny means that they are so small that even when you add them up the result goes to zero. The art in Ito's lemma is being able to tell which terms are small and which are tiny.

When $X_t = f(W_t, t)$, we can use Taylor expansion to get

$$\Delta X_k = \partial_w f \Delta W_k + \frac{1}{2} \partial_w^2 f \Delta W_k^2 + \partial_t f \Delta t + (*) \Delta W_k^3 + (**) \Delta W_k \Delta t \ .$$

The terms $(*)$ and $(**)$ are remainder terms in the Taylor series. The terms are "tiny" because, for example

$$\mathrm{E}\Big[\sum \left|\Delta W_k\right|^3 \Big] \sim (T_2 - T_1) \Delta t \ .$$

The Ito rule (22) comes from the fact that this term also is tiny:

$$\frac{1}{2} \partial_w^2 f \left(\Delta W_k^2 - \Delta t \right) \ .$$

The calculation to show this is like the calculation in Section 4.

$$\mathrm{E}\left[\left\{ \sum_{T_1 \leq t_k < T_2} \partial_w^2 f_k \left(\Delta W_k^2 - \Delta t \right) \right\}^2 \right] \sim (T_2 - T_1) \Delta t \ .$$

This shows that the error term goes to zero in distribution as $\Delta t \to 0$. As a result, you can do the replacement

$$\sum_{T_1 \leq t_k < T_2} \partial_w^2 f_k \Delta W_k^2 \implies \sum_{T_1 \leq t_k < T_2} \partial_w^2 f_k \Delta t \to \int_{T_1}^{T_2} \partial_w^2 f(W_t, t) \, dt \ .$$

6 Exercises

Notes on computational exercises. Each computational exercise is supposed to be a small research project. It will take some time to get a code working with well formatted output and graphs, but once you've done that, you should spend some time "playing" with the code using your natural curiosity. See what you can get it to do and think about why it might do that. The grader will reward creative computational work.

1. Suppose $0 < r < 1$, $X_0 = 0$, and, for $n > 0$,

$$X_{n+1} = rX_n + Y_n \ .$$

   Show that if the Y_k satisfy (4), then X_n has the zero mean property (6) but is not a martingale. A process like this is called *mean reverting.* In this case it reverts to the mean value zero.

2. The *gambler's ruin paradox* is an apparent paradox about martingales. The gambler makes a sequence of bets on i.i.d. random variables Y_n with $\Pr(Y_n = 1) = \Pr(Y_n = -1) = \frac{1}{2}$. The first bet is $a_0 = 1$. After that, the betting strategy is

$$a_k = \begin{cases} 0 & \text{if } X_k = 1 \\ 2^k & \text{otherwise.} \end{cases}$$

   Here, X_k is the win/loss (2). The strategy is called "double or nothing". Each time the gambler loses, he doubles his bet. He stops the first time he wins.

   (a) Find the possible values of X_n, calculate the probabilities of these values and show explicitly that $\mathrm{E}[X_n] = 0$.

   (b) Show that "the gambler eventually wins" by showing that Pr(never wins) = 0. The event "never wins" is the same as "lose at step 0, then lose at step 1, then $\cdots$". This seems to be a violation of the Doob martingale theorem because it a a guaranteed (probability 1) gain of \$1 betting on a martingale. Note that Doob's theorem, as given above, applies only at finite time n. Part (a) shows the conclusion is valid for any finite time.

   (c) Let $p_w = \Pr(Y_n = 1)$ and suppose $p_w < \frac{1}{2}$. Show that any strategy with $a_k \geq 0$, other than $a_k = 0$ for all k, has an expected loss at any finite time, but the gambler still wins \$1 with probability 1 if allowed to play arbitrarily long.

   (d) Suppose the gambler has a "capital requirement", that he is not allowed to bet if $X_n < -R$. Let p_B = Pr(hit the capital limit before winning \$1). This is the probability that the gambler "blows up" (a term for traders who are forced to stop trading because they lose too much money). Show that $p_b \to 0$ as $R \to \infty$.

The book *Fooled by Randomness* by Nassim Taleb makes the same point this exercise is making. A trader has ways to seem good by "beating the market" most of the time. These strategies don't have positive expected returns and are dangerous.

3. Verify the Ito Isometry formula for the example (18). To do this, first evaluate three terms on the right

$$Q_t = \mathrm{E}\left[\left(\frac{1}{2} W_t^2 - \frac{1}{2} t \right)^2 \right] = \frac{1}{4} \mathrm{E}\big[\, W_t^4 \big] - \frac{t}{2} \mathrm{E}\big[\, W_t^2 \big] + \frac{t^2}{4} \; .$$

   Then evaluate

$$q_s = \mathrm{E}\big[\, W_s^2 \big] \; .$$

   Then show that

$$\int_0^t q_s \, ds$$

   has the value the Ito isometry formula says it should have.

4. Download and run the code `ProportionalStrategy.py`. It should make a plot file that matches the posted `ProportionalStrategy.pdf`. This computation demonstrates the convergence in distribution of the Ito integral in the continuous time limit. The code computes the Ito integral corresponding to the following trading strategy. At time t, your wealth is X_t. You invest rX_t by betting on the Brownian motion outcome dW_t. This makes your gain or loss $rX_t dW_t$. The Ito integral adds this up, to give the wealth process

$$X_t = 1 + \int_0^t r X_s dW_s \; .$$

   The integrand is $a_s = rX_s$. It is adapted because it depends on the Brownian motion increments only up to time s. The code approximates the integral with a sum of the type (8). Please "play with" the code by seeing what happens when you change parameters. Here are some suggestions, but you don't have to do just this.

   (a) Increase the value of r and see that the large Δt curve is lower than the others. Why?

   (b) The Ito isometry formula in this case is

$$\mathrm{var}(X_T) = \int_0^T \mathrm{E}\big[\, a_t^2 \big] \, dt \; .$$

   Explain how to get this from (12). Warning: what is called X_t there is not what is called X_t here. Check this conclusion numerically by getting the left side from the simulated X_T values and the right side from the the third return value of the function `sim(...)`. The results should be presented in a small well formatted table. You can borrow from the Week 1 code to make a good table.

(c) (*extra credit, do only if time permits*) Example 2 of Section 5 explains how to derive a relation between X_T and W_T. Check whether this relation is satisfied as $\Delta t \to 0$. Present results in the form of a comparison of two or more cdf functions.

5. Merton's theory of default is a proposal for evaluating the default risk of a corporate bond. In a simplified version, a company agrees to pay a *coupon* which is $c\,dt$ in time dt. This is a continuous time idealization of real bonds that pay coupons are regular and frequent but discrete times. Assume there is a *discount rate* ρ and that the present value of a payment at time t is discounted by a factor $e^{-\rho t}$. Suppose the payments are supposed to continue to time T, but the payment may end at time $\tau < T$ if the company *defaults* at time τ. The "wedge" $\wedge$ mean "minimum", so the payments end at time $T \wedge \tau$. The total expected present value of the coupons is

$$Y = \mathrm{E}\left[\int_0^{T\wedge\tau} e^{-\rho t}\, dt \right] \; .$$

   There is a simple formula for the integral $[\cdots]$.

   Merton's "theory" is a model of the random variable τ. In this theory, the company (the *firm*) has a wealth X_t that is given by a process like the one of Exercise 4. There is a "drawdown level" $x_d < 1$ so that the company defaults at the first time when X_t hits x_d.

$$\tau = \min\left\{ t \mid X_t = x_d \right\} \; .$$

   Modify the code `ProportionalStrategy.py` to make a cdf of Y and estimate its expected value. It should use a sequence (not a long sequence) of Δt values and observe convergence. Choose values of r and T and x_d so that default is reasonably likely and non-default (by time T) is also likely.

Notes on the exercises

1. If you are trying to decide whether some mathematical statement is true or not, try it on the simplest examples you can think of. The linear autoregressive process should be on your list of examples.

2. If X_n is a sequence of random variables and $X_n \to X$ as $n \to \infty$, it might be that $\mathrm{E}[X_n]$ does not converge to $\mathrm{E}[X]$. For example, if $X_n = 2^n$ with probability 2^{-n} and $X_n = 0$ otherwise. This has $[X_n] = 1$. To go from n to $n+1$, "toss a coin" and take $X_{n+1} = 2X_n$ with probability $\frac{1}{2}$ and $X_[n+1] = 0$ otherwise. This is "double or nothing". These simple examples have relevance in finance.

3. The Ito isometry formula says two things are equal. Both are expected values of squares, so it might not be clear just how different they are. This exercise exercise emphasizes the difference.

4. You can learn the relation between probability distributions for two one component random variables by comparing the CDF. The CDF turns out not to be a strong differentiator between distributions.

   (a) Until you do this, it looks like you get accurate answers with a rather large Δt. This disappears when you try a harder problem. See why this goes wrong and you'll have a feel for geometric Brownian motion.

   (b) Part of this problem is taking the time to make good output. You might be curious why the results are so good.

   (c) This is another demonstration of the convergence of the Ito integral and the meaning of Ito's lemma.

5. Default modeling is one of the hard and important issues in finance. Merton's model is one of the classics that people build from. From a stochastic calculus point of view, this exercise illustrates the convergence of the distribution of the path $X_{[0,T]}$, not just the distribution of X_T. The hitting time τ is a function of the whole path, not just one value.