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. These are stochastic calculus tools related to the chain rule and ordinary (Riemann) integral of ordinary calculus. Integration and the chain rule are used to understand and solve ordinary differential equations. These need to be combined with the Ito integral and Ito's lemma for stochastic differential equations.

Here is a review of the ODE world in notation taken from the SDE world. An ODE may be written

$$dX_t = a(X_t)\, dt \; . \tag{1}$$

This is given a rigorous meaning using the ordinary integral:

$$X_t - X_0 = \int_0^t a(X_s)\, ds \; . \tag{2}$$

A related statement involves a function of the solution $f(X_t)$. The chain rule gives

$$df(X_t) = f'(X_t)\, dX_t = f'(X_t)\, a(X_t)\, dt \; .$$

The integral form of this differential relation is

$$f(X_t) = f(X_0) + \int_0^t f'(X_s)\, a(X_s)\, ds \; .$$

Again, the informal expression using differentials and the chain rule is equivalent to a formula involving integrals where all the terms have mathematical definitions. The rigorous integral formulas don't replace the differential formulas. Differential formulas are how people think about dynamics. But integral versions give us confidence that the informal reasoning is not silly.

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

$$X_t = \int_0^t a_s dW_s \; . \tag{3}$$

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.

A *stochastic differential equation* is a dynamical model of the form

$$dX_t = a(X_t)\, dt + b(X_t)\, dW_t \; . \tag{4}$$

This equation is understood in two ways, which we later will call *weak* and *strong*. The weak interpretation is that X_t is a stochastic process with infinitesimal mean and variance as explained in Week 1:

$$\mathrm{E}\big[\, dX_t \mid X_{[0,t]} \big] = a(X_t)\, dt \tag{5}$$

$$\mathrm{var}\big[\, dX_t \mid X_{[0,t]} \big] = b^2(X_t)\, dt \; . \tag{6}$$

These relations follow from the informal SDE (4) using the basic properties of Brownian motion that we "derived" in Week 1 ("motivated" would be more accurate):

$$\mathrm{E}[\, dW_t] = 0 \tag{7}$$

$$\mathrm{var}[\, dW_t] = dt \; . \tag{8}$$

You just take the conditional mean and variance of (4) using these, and you get the infinitesimal mean and variance expressions (5) and (6). The *strong* interpretation is that you can integrate both sides of the SDE (4) from time t_1 to time t_2 and get get

$$X_{t_2} - X_{t_1} = \int_{t_1}^{t_2} a(X_t)\, dt \, + \int_{t_1}^{t_2} b(X_t)\, dW_t \; . \tag{9}$$

The first integral on the right is an ordinary Riemann integral. The second is an Ito integral with respect to Brownian motion, which is one of the main topics for this Week.

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

Stochastic models are used to design and optimize strategies. Investors and traders use stochastic models of unknown future prices to create dynamic investment and trading strategies. Engineers use stochastic models to design control algorithms. Random walk models illustrate some conceptual ideas in stochastic control 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 "shares" 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 $a_n Y_n$. We assume, in this simplest model, that trading (having $a_{n+1} \neq a_n$) is free. We also assume that the trader is a *price taker*, which means that any numbers a_n are allowed (the market has an infinite supply of shares) and that buying or selling does not effect the price. The value of the "position" at the beginning of period k is $a_k S_k$, which is a_k shares, each of which has price S_k. At the end of period k, the price has changed to $S_{k+1} = S_k + Y_k$. The value of the "position" changes by $a_k(S_{k_1} - S_k)$. The total gain up to time n is

$$X_n = \sum_{k=0}^{n-1} a_k (S_{k+1} - S_k) \; . \tag{10}$$

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 use a probability model of the future that may include an uncertain forecast, but at time k, Y_k is to some degree random. 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 you 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:

$$\mathrm{E}[\,Y_k] = 0 \; , \;\; \mathrm{var}(\,Y_k) = \sigma^2 \; , \;\; Y_k \text{ are i.i.d.} \tag{11}$$

In that case, the total gain process (10) 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. There are several versions that apply to different kinds of random processes. They all refer to the conditional expectation of the future value given that you know the process up to a given time. A process is a *martingale* if the conditional expectation of the future value is equal to the known present value, even given the whole past history of the process. The theorem is that any non-anticipating strategy applied to a martingale produces another martingale. You cannot manufacture a positive expected value by strategizing on a zero expected return process.

Here is one theorem in the family of Doom martingale theorems. We call a sequence $(S_0, S_1, \cdots, S_k, \cdots, S_n)$ a *discrete path*. The time variable k is discrete but the value S_k might be continuous or discrete. The process is a *martingale* if the conditional expectation of the future value is the present value:

$$\mathrm{E}[\, S_{n+1} \mid (S_0, \ldots, S_n) = (s_0, \ldots, s_n)] = s_n \; . \tag{12}$$

The martingale property be stated in terms of the *increments* (also called *martingale differences*), given equivalently in increment or difference form:

$$S_{k+1} = S_k + Z_k \; , \quad S_{k+1} - S_k = Z_k \; .$$

It is

$$\mathrm{E}[\, Z_n \mid (S_0, \ldots, S_n) = (s_0, \ldots, s_n)] = 0 \; . \tag{13}$$

You can see this by noting that $Z_n = S_{n+1} - s_n$ (unknown value minus known value), conditional on knowing S_n.

The martingale property and the Markov property seem similar, but they are not the same. A Markov process does not have to be a martingale. In fact, we will see that the infinitesimal mean formula (5) implies that an SDE process is a martingale only if the infinitesimal mean (the drift) is zero. Continuous time martingales are defined in Section 3. Going the other way, the distribution of a martingale increment may depend on the whole path up to step n. In a Markov process, the distribution of the increment would depend on s_n only. For example, if the numbers Y_k satisfy the unbiased random walk properties (11), then the following (nonsensical but allowed) dynamical formula defines a martingale:

$$S_{n+1} = S_n + S_{n-1} Y_n + S_{n-2}(Y_n^2 - \sigma^2) \; .$$

All the last two terms on the right both have expected value zero, even knowing the values $S_{n-1} = s_{n-1}$ and $S_{n-2} = s_{n-2}$. The distribution of the increment may depend on the whole path, but no matter which path up to time n, the increment has mean zero.

A caution: it is possible that the unconditional increments have mean zero even though the process is not a martingale. The unconditional expected value is

$$\mathrm{E}[\, S_{n+1} - S_n] \; .$$

This can be zero even though the conditional expected values (13) are not all zero. On the other hand, if S_N is a martingale, then the unconditional means are zero because of what in stochastic processes is often called the *tower property*: the unconditional mean is the mean of the conditional means. Exercise 1 has an example of this.

Finally, the Doob theorem for this kind of martingale says that if S_n is a martingale, and X_n results from the sequence S_k by a strategy of the form (10), and if the strategy "weights" a_k are non-anticipating, then X_n is a martingale. The proof is simple. A strategy is non-anticipating means that a_n is some "strategic" function of everything that has come before, which means that we may write

$$a_n = a(s_1, \cdots, s_n, n) \; .$$

The function on the right refers to any a that can be determined from the values up to time n. The strategy may depend on n also. The increment of the X process is $Z_n = Y_n a(S_1, \cdots, S_n, n)$. Its conditional expectation, given the S path up to n, is

$$\begin{aligned} & \mathrm{E}[\, Y_n a(S_1, \cdots, S_n, n) \mid S_1 = s_1, \cdots, S_n = s_n] \\ &= \mathrm{E}[\, Y_n \mid S_1 = s_1, \cdots, S_n = s_n] \cdot a(s_1, \cdots, s_n, n) \\ &= 0 \; . \end{aligned}$$

The constant $a(s_1, \cdots, s_n, n)$ comes outside the expectation because it is just a number, not a random variable.

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

$$\mathrm{E}\big[\, X_n^2 \big] = \sigma^2 \sum_{k-0}^{n-1} \mathrm{E}\big[\, a_k^2 \big] \; . \tag{14}$$

In this theorem the sequence X_n is constructed from the sequence S_n by any non-anticipating strategy (10), and Y_k is an i.i.d. mean zero sequence (11). Here is a proof of the isometry formula using ideas that will become familiar in this course. We write X_n as a sum, writing the formula twice with a different summation index

$$X_n = \sum_{k=1}^{n} a_k Y_k = \sum_{j=1}^{n} a_j Y_j \; .$$

The strategy coefficients a_k could be complicated functions of $Y_1, \cdots, Y_{k-1}$. The proof starts with the trick of using both sums to express the square, and then interpreting the product of sums as a sum of products:

$$\begin{aligned} \mathrm{E}\big[\, X_n^2 \big] &= \mathrm{E}\left[\left(\sum_{k=1}^{n} a_k Y_k \right) \left(\sum_{j=1}^{n} a_j Y_j \right) \right] \\ &= \mathrm{E}\left[\sum_{k=1}^{n} \sum_{j=1}^{n} a_k Y_k a_j Y_j \right] \\ &= \sum_{k=1}^{n} \sum_{j=1}^{n} \mathrm{E}[\, a_k a_j Y_k Y_j] \; . \end{aligned}$$

The next step is to look at the individual terms and separate them into three groups: those with $j = k$ (the *diagonal* terms), the ones with $j < k$ and the ones with $j > k$. If $j = k$, the term is

$$\mathrm{E}\big[\, a_k^2 Y_k^2 \big] = \mathrm{E}\big[\, a_k^2 \big]\, \mathrm{E}\big[\, Y_k^2 \big] = \sigma^2 \mathrm{E}\big[\, a_k^2 \big] \ .$$

The first equality is true because a_k is independent of Y_k. That's true because Y_k is independent of all the other Y_j and a_k is a function only of those other Y_j.

Next, look at the *off-diagonal* terms with $k > j$. For those, we have

$$\mathrm{E}[\, a_k a_j Y_k Y_j] = \mathrm{E}[\, Y_k]\, \mathrm{E}[\, a_k a_j Y_j] = 0 \cdot \mathrm{E}[\, a_k a_j Y_j] = 0 \ .$$

This is true because Y_k is independent of everything else. The off-diagonal terms with $j > k$ also vanish, using similar reasoning. Thus, the double sum representing $\mathrm{E}[X_n^2]$ reduces to the sum of its diagonal terms. These are the terms in the isometry formula (14).

The name *isometry formula* comes from the mathematical terminology in which a transformation that does not change sizes is called an *isometry.* In this case, one of the objects is the sequence of random variables a_k. The size of this sequence is given by the sum of the expected squares (a kind of L^2 norm for fancy mathematicians). The other object is the random variable X_n, whose size is measured by $\mathrm{E}[X_n^2]$. You don't have to understand this explanation, but it may help you remember the name of the formula.

3 Strategies in continuous time

Last week we saw that Brownian motion is continuous time limit involving random walk. The Ito integral (3) is the corresponding continuous time limit of the "strategy" sum (10). 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

$$X_t^{\Delta t} = \sum_{t_k < t} a_{t_k}\left(W_{t_{k+1}} - W_{t_k} \right) \ . \qquad (15)$$

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

$$X_t = \int_0^t a_s \, dW_s = \lim_{\Delta t \to 0} \sum_{t_k < t} a_{t_k}\left(W_{t_{k+1}} - W_{t_k} \right) \ . \qquad (16)$$

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

$$\int_0^t b_s \, ds = \lim_{\Delta t \to 0} \sum_{t_k < t} b_{t_k} \Delta t \; . \tag{17}$$

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

The Ito integral limit (16) exists if a_t is continuous and adapted. *Adapted* (the same as *non-anticipating* in Section 2) 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. This choice is non-anticipating (adapted) because $1 + X_t$ depends on W_s for $s \leq t$. An important example involves the integral equation

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

The integrand on the right, which is $a_t = X_t$, is adapted because of the integral formula. The left side, X_t, is defined in terms of W_s for $s < t$. Integrals like this come up in solutions of stochastic differential equations we will see in Week 4.

The limit (16) 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 $\epsilon > 0$

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

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 (16) 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 $\Delta t \to 0$. A longer version of this Stochastic Calculus course had one class devoted to a proof that the limit formula (16) is true in probability.

There is a Doob martingale theorem for the Ito integral (16). 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

$$\mathrm{E}\big[\, X_{t+s} \mid W_{[0,t]} \big] = X_t \; . \tag{18}$$

In this statement, $W_{[0,t]}$ is the Brownian motion path on the time interval $[0, t]$. Part of the statement in (18) 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 (18) 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 (15). 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

$$\mathrm{E}\big[\, X_t^2 \big] = \int_0^t \mathrm{E}\big[\, a_s^2 \big] \; ds \; . \tag{19}$$

You prove this by using the discrete time Ito isometry theorem (14) 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

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

Therefore the discrete time formula (14) implies that

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

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 (19).

It is important to keep in mind that the approximation formula (15) 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 (**??**) 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

$$B_t^{\Delta t} = \sum_{t_k < t} a_{t_{k+1}} (W_{t_{k+1}} - W_{t_k}) \ . \tag{20}$$

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 (15) 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 (15) 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

$$W_k = \frac{1}{2} \left(W_{k+1} + W_k \right) - \frac{1}{2} \left(W_{k+1} - W_k \right) \ . \tag{21}$$

Using this, we get

$$X_t^{\Delta t} = \frac{1}{2} \sum_{t_k < t} \left(W_{k+1} + W_k \right) \left(W_{k+1} - W_k \right) - \frac{1}{2} \sum_{t_k < t} \left(W_{k+1} - W_k \right)^2 \ . \tag{22}$$

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 $n_t = \max \{ k | t_k < t \}$. 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*

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

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:

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

This shows that $\mathrm{E}\big[\, Q_t^{\Delta t}\big] \to t$ 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

$$\begin{aligned} \mathrm{var}\big(\, Q_t^{\Delta t}\big) &= \mathrm{E}\Big[\, \big(Q_t^{\Delta t} - t_{n_t}\big)^2\Big] \\ &= \mathrm{E}\left[\left(\sum_{t_k<t}\Big[(W_{k+1} - W_k)^2 - \Delta t\Big]\right)^2\right] \\ \mathrm{var}\big(\, Q_t^{\Delta t}\big) &= \mathrm{E}\left[\left(\sum_{t_k<t} R_k\right)^2\right] \qquad (23) \\ &\text{with} \\ R_k &= (W_{k+1} - W_k)^2 - \Delta t \ . \end{aligned}$$

We use the square of a sum = sum of products trick we used before. The result is:

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

The diagonal terms have (see Exercise 3) for the last line)

$$\begin{aligned} \mathrm{E}\big[\, R_k^2\big] &= \mathrm{E}\Big[\, (W_{k+1} - W_k)^4\Big] - 2\mathrm{E}\Big[\, (W_{k+1} - W_k)^2\Big]\,\Delta t + \Delta t^2 \\ \mathrm{E}\big[\, R_k^2\big] &= 2\Delta t^2 \ . \qquad (24) \end{aligned}$$

The off-diagonal terms vanish because of the independent increments property. For example, if $k \neq j$, then R_k is independent of R_j and both have expected value zero.

We finish the calculation with another simple trick. The variance (??) is given by the sum of the diagonal terms

$$\begin{aligned}\operatorname{var}\left(Q_t^{\Delta t}\right) &= \sum_{t_k<t} 2\Delta t^2 \\ &= 2\Delta t \cdot \sum_{t_k<t} \Delta t \\ &\approx 2\Delta t \cdot t \;.\end{aligned}$$

The approximation $\approx$ is exact if t is one of the discrete times t_k. Otherwise the error is less than Δt. Either way, $\operatorname{var}\left(Q_t^{\Delta t}\right) \to 0$ as $\Delta t \to 0$. Therefore $Q_t^{\Delta t} \to t$ in probability.

All this gives the result

$$\int_0^t W_s \, dW_s = \frac{1}{2} W_t^2 - \frac{1}{2} t \;. \tag{25}$$

The second term is the "Ito term", which has no analogue in the Riemann integral. If W_t were a differentiable function of t, we would calculate

$$\int_0^t W_s \, dW_s = \int_0^t W_s \frac{dW_s}{ds} \, ds = \frac{1}{2} \int_0^t \frac{d}{ds} \left(W_s^2\right) \, ds = \frac{1}{2} W_t^2 \;.$$

The Ito result is different from this. Without the Ito term, the formula would not satisfy the Doob martingale theorem or the Ito isometry formula (Exercise 4).

You will be happy to learn that we rarely evaluate Ito integrals directly from the definition in this way. Instead Ito's lemma plays the role of the fundamental theorem of calculus for the Riemann integral. We try to find the "anti Ito derivative". That almost never leads to a simple formula for the integral, even for the simpler Riemann integral it doesn't.

The anticipating (not non-anticipating) approximation (20) for this example gives

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

The trick (21) gives something that differs from (22) 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 (25), this bad answer is not a martingale. You can define a Riemann integral, using either endpoint of the interval, or the midpoint. That is not true for the Ito integral. It has to be the beginning of the interval: a_{t_k}, not $a_{t_{k+1}}$. Anticipating, even by just Δt, changes the answer.

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

$$dX_t = \partial_w f(W_t,t)\, dW_t + \left(\partial_t f(W_t,t) + \frac{1}{2}\partial_w^2 f(W_t,t) \right) dt\ . \tag{26}$$

This is a convenient but informal expression.

We think of the differential Ito's lemma formula as a relation between dX_t, dW_t and dt. To prove it, or to understand why it's true, we replace it with a corresponding integral formula. 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:

$$\int_{t_1}^{t_2} dX_t = X_{t_2} - X_{t_1}\ . \tag{27}$$

The integral form of Ito's lemma is the formula you get by integrating both of the terms on the right side of the differential formula (26):

$$X_{t_2} - X_{t_1} = \int_{t_1}^{t_2} \partial_w f(W_t,t)\, dW_t + \int_{t_1}^{t_2} \left(\partial_t f(W_t,t) + \frac{1}{2}\partial_w^2 f(W_t,t) \right) dt\ . \tag{28}$$

This is the stochastic calculus version of the fundamental theorem of calculus.

Ito's lemma is a version of the chain rule appropriate for Brownian motion. In the ordinary chain rule concerns a differentiable function of t, which we write as U_t. The chain rule formula, in language of ordinary calculus language, 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 stochastic calculus 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 (26) includes the second derivative term involving $\partial_w^2 f$, which is not present in the ordinary calculus chain rule.

There is an informal "derivation" of Ito's lemma that tells you how to remember it but not why it's true. You just expand $df(W_t,t)$ to second order and apply "Ito's rule", which is $(dW_t)^2 = dt$. You get

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

A real derivation, one that explains why it's true, would have to explain why you expand to second order in W but not in t, among other things.
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 (26) 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 (25) 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 (25) 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

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

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 5 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 (27):

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

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 (26)

$$\Delta X_k = a_k \Delta W_k + b_k \Delta t + R_k \; . \tag{29}$$

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

$$S_1^{\Delta t} = \sum_{T_1 \le t_k < T_2} a_k \Delta W_k \;\; \to \;\; \int_{T_1}^{T_2} a_t dW_t \; . \tag{30}$$

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

$$S_2^{\Delta t} = \sum_{T_1 \le t_k < T_2} b_k \Delta t \;\; \to \;\; \int_{T_1}^{T_2} b_t dt \; . \tag{31}$$

The last sum vanishes in the limit:

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

The terms with non-zero limits (30) and eqrefS2 converge to the integrals on the right side of (28).

There is some informal language for this. When Δt is small, all the terms on the right of the expression (29) for ΔX_k are small. The term $a_k \Delta W_k$ is small, but when you add them together in the sum (30) 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 (31) 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

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

The Ito rule (**??**) 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.

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

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

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

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 $X_0 = 0$, and, for $n > 0$,

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

Show that if the Y_k satisfy (11), then X_n has the zero mean property $\mathrm{E}[X_n] = 0$ but is not a martingale. This process is called *mean reverting* if $|r| < 1$. It mean reverts to the to zero if $X_0 = 0$.

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 $\Pr(Y_n = 1) = \Pr(Y_n = -1) = \frac{1}{2}$ (we call these "coin tosses" or "fair" coin tosses). The first bet is $a_0 = 1$. The strategy is to double the bet each time you lose (get $Y_k = -1$) and stop betting the first time you win (get $Y_k = 1$). In other words

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

Here, X_k given by (10) with this strategy. The strategy is called "double or nothing".

   (a) Find the possible values of X_n, calculate the probabilities of these values and show explicitly that $\mathrm{E}[X_n] = 0$. How is this related to the martingale property?

   (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 $p_B = \Pr(\text{hit the capital limit before winning \$1})$. 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. Suppose $Y \sim \mathcal{N}(0, \sigma^2)$. Show that $\text{var}\big(Y^2\big) = 2\sigma^4$. For this, you probably need to verify that

$$\int_{-\infty}^{\infty} z^4 e^{-\frac{1}{2}z^2}\, dz = 3 \int_{-\infty}^{\infty} z^2 e^{-\frac{1}{2}z^2}\, dz \;.$$

4. Verify that the example (25) satisfies the continuous time Doob martingale theorem (the right side is a martingale) and the Ito isometry formula. For the isometry formula, it may help to do Exercise 3 first.

5. 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

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

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 (15). 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

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

   Explain how to get this from (19). 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.

6. (*This is for "extra credit" and is not required as part of the assignment for Week 2*. Do this only if you have extra time.)

   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 present value of the coupons is the random variable

$$Y = \int_0^{T\wedge\tau} e^{-\rho t}\, dt \; .$$

   There is a simple formula for the integral.

   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 5. There is a "drawdown level" $x_d < 1$ so that the company defaults at the first time when X_t hits x_d.

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

   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 auto-regressive 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.