Week 5
Integrals with respect to Brownian motion

Jonathan Goodman

October 7, 2013

1 Introduction to the material for the week

This week continues the calculus aspect of stochastic calculus, the limit $\Delta t \to 0$ and the Ito integral. This is one of the most technical classes of the course. Look for applications in coming weeks. Brownian motion plays a new role this week, as a source of *white noise* that drives other continuous time random processes. Starting this week, W_t usually denotes standard Brownian motion, so that X_t can denote different random process *driven* by W in some way. The driving white noise is written informally as dW_t.

White noise is a continuous time analogue of a sequence of i.i.d. random variables. Let Z_n be such a sequence, with $\mathrm{E}[\, Z_n] = 0$ and $\mathrm{E}\left[\, Z_n^2\right] = 1$. These generate a *random walk*,

$$V_n = \sum_{k=0}^{n-1} Z_k \; . \tag{1}$$

The V_n can be expressed in a more dynamical way by saying $V_0 = 0$ and $V_{n+1} = V_n + Z_n$. If the sequence V_n is given, then

$$Z_n = V_{n+1} - V_n \; . \tag{2}$$

In the continuous time limit, a properly scaled V_n converges to Brownian motion. The discrete time "independent increments property" is the statement that the Z_n defined by (2) are independent. The discrete time analogue of the fact that Brownian motion is homogeneous in time is the statement that the Z_n are identically distributed. We can also write

$$\mathrm{E}\left[\, (V_n - V_m)^2 \mid \mathcal{F}_m \right] = n - m \; ,$$

Which is the analogue of the corresponding Brownian motion formula.

From their basic definitions, the continuous time white noise and Brownian motion must be Gaussian. (This is part of the *Levy uniqueness theorem.*) Suppose that the random variables Z_n are i.i.d. but not Gaussian. Even then, the scaling limits of V_n are Gaussian Brownian motion, and the scaling limit of the

Z_n process, which is harder to define, is Gaussian white noise. The continuous time scaling limit for Brownian motion is

$$\frac{1}{\sqrt{\Delta t}} V_n \stackrel{\mathcal{D}}{\rightharpoonup} W_t \ , \quad \text{as } \Delta t \to 0 \text{ with } t_n = n\Delta t, \text{ and } t_n \to t. \tag{3}$$

The CLT implies that W_t is Gaussian regardless of the distribution of Z_n. The *white noise* "process" dW_t is Gaussian as well, in whatever way it makes sense.

In continuous time, it is simpler to define white noise from Brownian motion rather than the other way around. The continuous time analogue of (2) is to write dW_t as the source of noise. The continuous time analogue of (1) would be to define a white noise process Z_t somehow, then get Brownian motion as

$$W_t = \int_0^t Z_s \, ds \ . \tag{4}$$

The numbers W_t make sense as random variables and the path W_t is a continuous function of t. The numbers Z_t do not make sense in the same way.

The Ito integral with respect to Brownian motion is written

$$X_t = \int_0^t f_s dW_s \ . \tag{5}$$

The *integrand*, is f_t. It can be random, but there is an important constraints: the value of f_t must be known at time t. The relation between X and W may be expressed informally in the Ito differential form

$$dX_t = f_t dW_t \ . \tag{6}$$

The discrete analogue would be

$$X_n = \sum_{k=0}^{n} f_n (V_{n+1} - V_n) \tag{7}$$

$$= \sum_{k=0}^{n} f_n Z_n \ . \tag{8}$$

The "integrand", f_n is *nonanticipating* if its value is "known at time n". The more formal statement is that the future noise values, Z_k for $k \geq n$, are independent of f_n. In this case,

$$\mathrm{E}[\, X_n - X_{n-1}] = \mathrm{E}[\, f_n Z_n] = 0 \ .$$

The stronger statement (below) is that X_n is a *martingale*.

The discrete version (7) is defined even if f_n is not non-anticipating. But the $\Delta t \to 0$ does not work for the continuous time Ito integral (5) unless f_t is adapted to the filtration generated by W. If $\mathcal{F}_t$ is generated by the path $W_{[0,t]}$, then f_t must be measurable in $\mathcal{F}_t$. The Ito integral is different from other stochastic integrals (e.g. Stratonovich) in that the increment dW_t is taken to be in the future of t and therefore independent of $f_{[0,t]}$. This implies that

$$\mathrm{E}[\,dX_t \mid \mathcal{F}_t] = f_t \mathrm{E}[\,dW_t \mid \mathcal{F}_t] = 0\ , \tag{9}$$

and

$$\mathrm{E}\big[\,dX_t^2 \mid \mathcal{F}_t\big] = f_t^2 \mathrm{E}\big[\,dW_t^2 \mid \mathcal{F}_t\big] = f_t^2\, dt\ . \tag{10}$$

The Ito integral is important because more or less any continuous time continuous path stochastic process X_t can be expressed in terms of it. A *martingale* is a process with the mean zero property (9). More or less any such martingale can be represented as an Ito integral (5). This is in the spirit of the central limit theorem. In the continuous time limit, a process is determined by its mean and variance. If the mean is zero, it is only the variance, which is f_t^2.

The mathematics this week is reasonably precise yet not fully rigorous. You should be able to understand it even if you have not studied "mathematical analysis". This material is not "for culture". You are expected to master it along with the rest of the course. If this were not possible, or not important, the material would not be here.

The approach taken here is not the standard approach using approximation by "simple functions" and the *Ito isometry* formula. You can find the standard approach in the book by Oksendal, for example. The standard approach is simpler but relies more results from measure theory. The approach here will look almost the same as the standard approach if you do it completely rigorously, which we do not.

2 Pathwise convergence and the Borel Cantelli lemma

Section 3 constructs a sequence of approximations, X_t^m, that converges to the Ito integral as $m \to \infty$. This section describes some technical tools that help us prove such limits. The method a version of the standard *Borel Cantelli lemma*. This section is written without the usual motivations. You may need to read it twice to see how things fit together.

Suppose $a_m > 0$ is a sequence of numbers with a finite sum

$$s = \sum_{m=1}^{\infty} a_m < \infty\ . \tag{11}$$

Let r_n be the *tail sum*

$$r_n = \sum_{m>n} a_m\ .$$

Then $r_n \to 0$ as $n \to \infty$. The proof of this is that the *partial sums*

$$s_n = \sum_{m=1}^{n} a_m$$

converge to s, and $s_n + r_n = s$ for any n, so $s - s_n = r_n \to 0$ as $k \to \infty$.

Now suppose b_m is a sequence of numbers, not necessarily positive, and consider the sum

$$x = \sum_{m=1}^{\infty} b_m \ . \tag{12}$$

The sum converges *absolutely* if

$$\sum_{m=1}^{\infty} |b_m| < \infty \ .$$

You can prove that the sum converges absolutely by finding $a_m > 0$ that satisfy $|b_m| \leq a_m$ and (11). For example, suppose $b_m = \frac{1}{m^2}\cos(mt)$, with t being some fixed number. Rather than spending time trying to figure out the sum

$$\sum_{m=1}^{\infty} |b_m| = \sum_{m=1}^{\infty} \frac{1}{m^2} |\cos(mt)| \ ,$$

you can just say $|b_m| \leq a_m = \frac{1}{m^2}$ and know that

$$\sum_{m=1}^{\infty} a_m = \sum_{m=1}^{\infty} \frac{1}{m^2} < \infty \ .$$

The partial sums for (12) are

$$x_n = \sum_{m=1}^{n} b_m \ .$$

By definition, the sum (12) converges, and is equal to x, if $x_n \to x$ as $n \to \infty$. If we have an *upper bound* sequence a_m, then

$$|x - x_n| = \left| \sum_{m>n} b_m \right| \leq \sum_{m>n} |b_m| \leq \sum_{m>n} a_j = r_n \to 0 \ ,$$

as $n \to \infty$.

We apply this idea to proving convergence of a sequence. Suppose x_n is a given sequence of numbers, and we want to show it converges to a limit as $n \to \infty$. We define the sequence of differences $b_m = x_m - x_{m-1}$. We define the first b assuming that $x_0 = 0$, which gives $b_1 = x_1$. Then the x_n limit is the same as the b_m sum:

$$x = \lim_{n\to\infty} x_n = \lim_{n\to\infty} \sum_{m=1}^{n} b_m = \sum_{m=1}^{\infty} b_m$$

If we can find a_m that satisfies the conditions

$$|x_m - x_{m-1}| = |b_m| \leq a_m \ , \qquad \sum_{m=1}^{\infty} a_m < \infty \ ,$$

that proves that the x_n limit exists.

Suppose A_m is a sequence of non-negative random numbers with a random sum

$$S = \sum_{m=1}^{\infty} A_m \ .$$

An example would be

$$A_m = Y_m^2 \ , \quad \text{with} \ \ Y_m \sim \mathcal{N}(0, \frac{1}{m^2}) \ .$$

Typically, the A_m can be arbitrarily large and so it might happen that $S = \sum A_m = \infty$. We hope to show that the probability it will happen is zero. The event $S = \infty$ is a measurable set, which in some sense means it is a possible outcome. But if $\mathrm{P}(S = \infty) = 0$, you will never see that outcome. We say that an event $D \subset \Omega$ happens *almost surely* if $\mathrm{P}(D) = 1$. This is abbreviated as *a.s.,* as in $S < \infty$ almost surely, or $S < \infty$ a.s. Other expressions are *a.e.*, for *almost everywhere*, and *p.p.*, for *presque partout* (almost everywhere, in French). One can distinguish between outcomes that are impossible, which would be $\omega \notin \Omega$, and events that have probability zero. We will ignore this distinction most of the time.

Our strategy is to show that $S < \infty$ a.s. by showing that $\mathrm{E}[\,S] < \infty$. If the expected value is finite:

$$\mathrm{E}[\,S] = \mathrm{E}\left[\sum_{m=1}^{\infty} A_m\right] = \sum_{j=m}^{\infty} \mathrm{E}[\,A_m] < \infty \ ,$$

then the sum is finite, almost surely:

$$S = \sum_{m=1}^{\infty} A_m < \infty \quad , \qquad \text{a.s.}$$

In particular, let X_t^m be a sequence of random paths. Suppose there is a sequence of numbers, not random, so that

$$\mathrm{E}\left[\,\left|X_t^{m+1} - X_t^m\right|\right] \le a_m \quad , \qquad \text{with} \qquad \sum_{m=1}^{\infty} a_m < \infty \ . \tag{13}$$

Then you know that the following limit exists almost surely

$$X_t = \lim_{j\to\infty} X_t^m \ . \tag{14}$$

This is our version of the *Borel Cantelli* lemma. We calculate expected values to verify the hypothesis (13), then we conclude that the limit exists pathwise almost surely.

Although these are the major quantitative arguments, they are not complete mathematical proofs. For example, we did not give a full definition of the probability space or probability measures involved. We did not give the mathematical definition of expectation with respect to a probability measure. We did not prove that $\mathrm{E}[\sum A_m] = \sum \mathrm{E}[A_m]$. You can find details like these in a graduate level course on theoretical probability, such as the Courant Institute course *Probability Limit Theorems*.

3 Riemann sums for the Ito integral

We use the following Riemann sum approximation for the Ito integral (5):

$$X_t^m = \sum_{t_j < t} f_{t_j} \Delta W_j \; . \tag{15}$$

The notation is

$$\Delta t = 2^{-m} \; , \tag{16}$$

$$t_j = j \Delta t \; , \tag{17}$$

W_t is a standard Brownian motion, and

$$\Delta W_j = W_{t_{j+1}} - W_{t_j} \; , \tag{18}$$

We always assume that f_t is measurable with respect to $\mathcal{F}_t$, which is the formal way of saying that "f_t is known at time t". We will show that the sequence of approximations (15) converges as $m \to \infty$ for almost every Brownian motion path. This limit will be measurable in $\mathcal{F}_t$ because X_t is a function of $W_{[0,t]}$.

The Riemann sum approximation (15) needs lots of explanation. The Brownian motion increment used at time t_j (18) is in the future of t_j. The f_{t_j} are measurable in $\mathcal{F}_{t_j}$ (or *progressively measurable*, or *non-anticipating*, or *adapted* to the filtration $\mathcal{F}_t$), so ΔW_j independent of f_{t_j}. In particular,

$$\mathrm{E}\left[\, f_{t_j} \Delta W_j \mid \mathcal{F}_{t_j} \right] = 0 \; , \tag{19}$$

and

$$\mathrm{E}\left[\left(f_{t_j} \Delta W_j \right)^2 \mid \mathcal{F}_{t_j} \right] = f_{t_j}^2 \, \mathrm{E}\left[\Delta W_j^2 \mid \mathcal{F}_{t_j} \right] = f_{t_j}^2 \Delta t \; . \tag{20}$$

The Riemann sum definition (15) definies X_t^m for all t. It gives a path that is discontinuous at the times t_j. Sometimes it is convenient to re-define X_t^m by linear interpolation between t_j and t_{j+1} so that it is continuous. Those subtleties do not matter this week. We mention them because they seem to play a big role in other treatments of the subject.

Taking $m \to \infty$ has the effect of sending $\Delta t_m = 2^{-m}$ to zero. This is not the same as just letting $\Delta t \to 0$, because not all possible small values of Δt are considered. The $m \to \infty$ approach simplifies the technical details in two ways. One is that the time steps Δt_m converge to zero quickly. The other is that it is easy to compare the Δt_m and $\Delta t_{m+1} = \frac{1}{2} \Delta t_m$ approximations. Ultimately, we want to understand the integral (5) rather than the technical approximations used to define it.

We assume that the integrand f_t is continuous in some way. Specifically, we assume that for any T, there is a C_T so that if $t \leq T$ and $s > 0$, then

$$\mathrm{E}\Big[(f_{t+s} - f_t)^2 \mid \mathcal{F}_t \Big] \leq C_T \, s \; . \tag{21}$$

This allows integrands like $f_t = W_t$, or $f_t = tW_t$. Some of the integrands we use later in the course do not satisfy this hypotheses, but most are close. We will re-examine the conditions on f_t below to see what is really necessary.

Here is the strategy for proving that the limit

$$X_t = \lim_{m \to \infty} X_t^m$$

exists. We use the criterion (13), and seek an upper bound a_m so that

$$\mathrm{E}\big[\left| X_t^{m+1} - X_t^m \right| \big] \leq a_m \; . \tag{22}$$

We do this, in turn, by finding a_m^2 so that

$$\mathrm{E}\Big[\left(X_t^{m+1} - X_t^m \right)^2 \Big] \leq a_m^2 \; . \tag{23}$$

The *Cauchy Schwarz inequality* (see below) implies that (22) is a consequence of (23). In fact, if U is any random variable, then the Cauchy Schwarz implies that

$$\mathrm{E}[\, |U| \,] \leq \sqrt{\mathrm{E}[\, U^2 \,]} \; . \tag{24}$$

One can also derive (24) using *Jensen's inequality*, but that takes longer to explain.

The Cauchy Schwartz inequality for random variables is the following theorem. Suppose U and V are any two random variables (correlated or not), then

$$\mathrm{E}[\, UV \,] \leq \sqrt{\mathrm{E}[\, U^2 \,] \, \mathrm{E}[\, V^2 \,]} \; . \tag{25}$$

A small trick gets (24) from this. Define V from U as $V = 1$ if $U \geq 0$, and $V = -1$ if $U < 0$, so that $UV = |U|$ and $\mathrm{E}\big[V^2 \big] = 1$. We prove the Cauchy Schwarz inequality (25) using $(U - \alpha V)^2$, which is non-negative for any α. Therefore

$$0 \leq \mathrm{E}\Big[(U - \alpha V)^2 \Big] = \mathrm{E}\big[U^2 \big] - 2\alpha \mathrm{E}[\, UV \,] + \alpha^2 \mathrm{E}\big[V^2 \big] \; .$$

We minimize the right side by taking $\alpha = \mathrm{E}[\, UV \,] / \mathrm{E}\big[V^2 \big]$. Putting this in the second expression gives

$$0 \leq \mathrm{E}\big[U^2 \big] - \frac{\mathrm{E}[\, UV \,]^2}{\mathrm{E}[\, V^2 \,]} \; .$$

Multiply through by $\mathrm{E}\big[V^2 \big]$ and you get (25), which implies (24). And (24) is the reason our desired (22) follows from the more convenient (23).

Calculating squares, as in (23) rather than (22), is informative because it can reveal *cancellation* in a sum. Consider a generic sequence of random variables Y_k with $\mathrm{E}[\,Y_k] = 0$, and look at the sums

$$S_m = \sum_{k=1}^{m} Y_k \ .$$

We say there is *cancellation* in the sum if

$$|S_m| \ll \sum_{k=1}^{m} |Y_k| \ .$$

The symbol $\ll$ means "is much less than". It is a little vague, as is the rest of this motivational paragraph. A sum has cancellation if the positive terms are nearly balanced by the negative terms. This requires the terms to be different, obviously. for example, suppose $Y_k = Y$ for all k. Then $S_m = mY$, and there is no cancellation. The opposite extreme is the case where $Y_k \sim Y$ but are independent. In that case, we calculate

$$\mathrm{E}\left[\,S_m^2\right] = \mathrm{E}\left[\left(\sum_{k=1}^{m} Y_k\right)^2\right] \ .$$

You can see how to expand the square on the right by writing

$$\begin{aligned}(a+b+c)^2 &= (a+b+c)(a+b+c) \\ &= a^2 + ab + ac + ba + b^2 + bc + ca + cb + c^2 \ .\end{aligned}$$

In the same way

$$\begin{aligned}\left(\sum_{k=1}^{m} Y_k\right)^2 &= \left(\sum_{j=1}^{m} Y_j\right)\left(\sum_{k=1}^{m} Y_k\right) \\ &= \sum_{j=1}^{m}\sum_{k=1}^{m} Y_j Y_k \ .\end{aligned}$$

The *diagonal terms* are the terms on the right with $j = k$, by analogy to the diagonal entries of a matrix. The *off diagonal* terms are the ones with $j \neq k$. The diagonal terms have expected value

$$\mathrm{E}\left[\,Y_k^2\right] = \mathrm{E}\left[\,Y^2\right] \ .$$

The off diagonal terms, if Y_k is independent of Y_j for $k \neq j$, are

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

Therefore, adding the diagonal and off diagonal terms,

$$\mathrm{E}\left[\,S_m^2\right] = \sum_{k=1}^{m} \mathrm{E}\left[\,Y_k^2\right] + \sum_{j\neq k} \mathrm{E}[\,Y_j Y_k] = \sum_{k=1}^{m} \mathrm{E}\left[\,Y^2\right] = m\sigma_Y^2 \ .$$

The Cauchy Schwarz inequality turns this into an estimate for $|S_m|$:

$$\mathrm{E}[\,|S_m|] \leq \sqrt{m}\sqrt{\sigma_Y^2} \ .$$

Thus, although S_m is the sum of m terms, it is only on the order of $\sqrt{m}$ because of cancellation. We found the cancellation by computing the expected square.

With all this motivation, we estimate

$$\mathrm{E}\Big[\left(X_t^{m+1} - X_t^m\right)^2\Big] \ .$$

The time increments in the X_t^m sum are of the form

$$[t_j^m, t_{j+1}^m] = [j\Delta t_m, (j+1)\Delta t_m] \ .$$

This time interval contributes $f_{t_j^m}(W_{t_{j+1}^m} - W_{t_j^m})$ to X_t^m. This level m interval consists of exactly two level $m+1$ intervals:

$$[t_j^m, t_{j+1}^m] = [t_{2j}^{m+1}, t_{2j+1}^{m+1}] \cup [t_{2j+1}^{m+1}, t_{2j+2}^{m+1}] \ .$$

You can verify this starting from the fact that $\Delta t_{m+1} = \frac{1}{2}\Delta t_m$, so $t_{2j}^{m+1} = 2j\,\Delta t_{m+1} = 2j\,\frac{1}{2}\Delta t_m = t_j^m$. The following notation simplifies the discussion. We fix m and leave out the m superscripts and subscripts, writing t_j for t_j^m, etc. We write $t_{j+\frac{1}{2}} = (j+\frac{1}{2})\Delta t$ for the midpoint of the level m interval $[t_j^m, t_{j+1}^m]$. In this notation, we have

$$[t_j, t_{j+1}] = [t_j, t_{j+\frac{1}{2}}] \cup [t_{j+\frac{1}{2}}, t_{j+1}] \ .$$

For even more simplicity, we write skip the t's and write $f_{j+\frac{1}{2}}$ for $f_{t_{j+\frac{1}{2}}}$, and $W_{j+\frac{1}{2}}$ for $W_{t_{j+\frac{1}{2}}}$, etc. In this notation, we have

$$X_t^{m+1} = \sum_{t_j<t} \Big[f_j \left(W_{j+\frac{1}{2}} - W_j\right) + f_{j+\frac{1}{2}} \left(W_{j+1} - W_{j+\frac{1}{2}}\right)\Big] \ + Q \ .$$

The Q on the end is the term that may result from X_t^{m+1} having an odd number of terms in its sum. In that case, Q is the last term. It makes a negligible contribution to the sum. We subtract from X_t^{m+1} the X_t^m sum

$$X_t^m = \sum_{t_j<t} f_j \left(W_{j+1} - W_j\right) \ .$$

The result is

$$X_t^{m+1} - X_t^m = \sum_{t_j<t} \left(f_{j+\frac{1}{2}} - f_j\right)\left(W_{j+1} - W_{j+\frac{1}{2}}\right) \ + Q \ . \qquad (26)$$

Now the calculation starts.

Denote a typical term in the sum on the right of (26) as

$$Y_j = \left(f_{j+\frac{1}{2}} - f_j\right)\left(W_{j+1} - W_{j+\frac{1}{2}}\right) \ .$$

It is clear from the definition that

$$\mathrm{E}\Big[\,Y_j \mid \mathcal{F}_{j+\frac{1}{2}}\Big] = \left(f_{j+\frac{1}{2}} - f_j\right)\mathrm{E}\Big[\,W_{j+1} - W_{j+\frac{1}{2}} \mid \mathcal{F}_{j+\frac{1}{2}}\Big] = 0 \ .$$

It follows from the tower property that $\mathrm{E}[\,Y_j \mid \mathcal{F}_j] = 0$. The off diagonal expected values are zero. To see this, suppose $k < j$. Then Y_k is known in $\mathcal{F}_j$, so

$$\mathrm{E}[\,Y_k Y_j \mid \mathcal{F}_j] = Y_k\,\mathrm{E}[\,Y_j \mid \mathcal{F}_j] = 0 \ .$$

We calculate the diagonal terms in two tower property steps, starting with

$$\begin{aligned}
\mathrm{E}\Big[\,Y_j^2 \mid \mathcal{F}_{j+\frac{1}{2}}\Big] &= \mathrm{E}\left[\left(f_{j+\frac{1}{2}} - f_j\right)^2\left(W_{j+1} - W_{j+\frac{1}{2}}\right)^2 \,\middle|\, \mathcal{F}_{j+\frac{1}{2}}\right] \\
&= \left(f_{j+\frac{1}{2}} - f_j\right)^2 \mathrm{E}\left[\left(W_{j+1} - W_{j+\frac{1}{2}}\right)^2 \,\middle|\, \mathcal{F}_{j+\frac{1}{2}}\right] \\
&= \left(f_{j+\frac{1}{2}} - f_j\right)^2 \frac{\Delta t}{2} \ .
\end{aligned}$$

Now we use (21) and go from $\mathcal{F}_{j+\frac{1}{2}}$ to $\mathcal{F}_j$, using the previous result and the tower property:

$$\mathrm{E}\big[\,Y_j^2 \mid \mathcal{F}_j\big] = \mathrm{E}\left[\left(f_{j+\frac{1}{2}} - f_j\right)^2 \mid \mathcal{F}_j\right]\frac{\Delta t}{2} \le C\Delta t^2 \ .$$

One more application of the tower property gives

$$\mathrm{E}\big[\,Y_j^2\big] = \mathrm{E}\big[\,\mathrm{E}\big[\,Y_j^2 \mid \mathcal{F}_j\big]\big] \le \mathrm{E}\big[\,C\Delta t^2\big] = C\Delta t^2 \ . \tag{27}$$

This is the estimate we need.

The expected square is the sum of the diagonal terms (27):

$$\begin{aligned}
\mathrm{E}\Big[\,\big(X_t^{m+1} - X_t^m\big)^2\Big] &= \sum_{t_j \le t} \mathrm{E}\big[\,Y_j^2\big] \\
&= C \sum_{t_j \le t} \Delta t_m^2 \ .
\end{aligned}$$

Note that $\sum_{t_j^m < t} \Delta t_m \le t$. Therefore

$$\sum_{t_j^m < t} C\Delta t_m^2 \le C\Delta t_m \sum_{t_j^m < t} \Delta t \le C\,t\,\Delta t_m \ .$$

That leads to

$$\mathrm{E}\Big[\,\big(X_t^{m+1} - X_t^m\big)^2\Big] \le C_t \Delta t_m = C_t 2^{-m} \ . \tag{28}$$

The Cauchy Schwarz inequality (24) gives

$$\mathrm{E}\left[\,\left|X_t^{m+1} - X_t^m\right|\right] \leq C_t 2^{-m/2} \;. \tag{29}$$

If you have not seen this before, you might be worried that what is called C_t in (29) is the square root of what is called C_t in (28). Mathematicians use C, to mean "some constant". This constant, the value of C, could be different in different places. Similarly, C_t means "some constant whose value depends on t". It is not called $C_{t,m}$ because its value does not depend on m. The part corresponding to Q, which we have ignored until now, can indeed be ignored (check if you don't believe me).

We can now apply the Borel Cantelli lemma. The estimate (29) implies that the hypotheses (13) are satisfied, with $a_m = C_t 2^{-m/2}$ and therefore $\sum a_m < \infty$. This implies that the Riemann sums (15) have a limit as $m \to \infty$.

We used the powers of two in two ways. First, it made it easy to compare X_t^m to X_t^{m+1}. Second, it made the sum on the right of (13) a convergent geometric series. If we had taken $\Delta t_m = \frac{1}{m}$ and proved the estimate (29), that would not have proven convergence as $m \to \infty$, because $\sum a_m$ would be infinite in that case. It is possible to prove convergence of the approximations (in hopefully clear notation) $X_t^{\Delta t}$ as $\Delta t \to 0$, but we do not have the time for that proof in this course.

It is possible to prove a stronger *uniform convergence* theorem. In fact, we may sometimes assume uniform convergence in coming weeks. We describe uniform convergence using the maximum difference up to some time T:

$$D_T^m = \max_{t \leq T} \left|X_t^{m+1} - X_t^m\right| \;.$$

There is a well known theorem in probability ("Well known to those who know it well" – Mal Kalos) called *Doob's martingale inequality* that uses what we already proved to show that if $\Delta t_m = 2^{-m}$, then

$$\mathrm{E}[\,D_T^m] \leq C_T \Delta t_m \;.$$

The assumption (21) can be relaxed. For example, it suffices to take $\mathrm{E}\left[\,(f_{t+s} - f_t)^2\right] \leq Cs$, rather than the conditional expectation. This allows discontinuous integrands that depend on hitting times. It is possible to substitute a power of s less than 1, such as $\sqrt{s}$.

4 Example

There are a few Ito integrals that can be computed directly from the definition. Ito's lemma, which we will see next week, is a better approach actual calculations. Ito's lemma is the stochastic integral analogue of the fundamental theorem of calculus. Riemann sums define the integral in ordinary calculus. But it is easier to integrate by anti-differentiation than by taking the limit of Riemann sums.

The first example is

$$X_t = \int_0^t W_s dW_s \ . \tag{30}$$

The Riemann sum approximation is

$$X_t^m = \sum_{t_j < t} W_{t_j} \left(W_{t_{j+1}} - W_{t_j} \right) \ .$$

The trick for doing this is

$$W_{t_j} = \frac{1}{2} \left(W_{t_{j+1}} + W_{t_j} \right) - \frac{1}{2} \left(W_{t_{j+1}} - W_{t_j} \right) \ .$$

This leads to

$$X_t^m = \frac{1}{2} \sum_{t_j < t} \left(W_{t_{j+1}} + W_{t_j} \right) \left(W_{t_{j+1}} - W_{t_j} \right) - \frac{1}{2} \sum_{t_j < t} \left(W_{t_{j+1}} - W_{t_j} \right) \left(W_{t_{j+1}} - W_{t_j} \right) \ .$$

A general term in the first sum is

$$\left(W_{t_{j+1}} + W_{t_j} \right) \left(W_{t_{j+1}} - W_{t_j} \right) = W_{t_{j+1}}^2 - W_{t_j}^2 \ .$$

Therefore, the first sum is a *telescoping sum*,[1] which is a sum of the form

$$(a - b) + (b - c) + \cdots + (x - y) + (y - z) = a - z \ .$$

Let $t_n = \max \{ t_j \mid t_j < t \}$, then the first sum is $\frac{1}{2} \left(W_{t_{n+1}}^2 - W_0^2 \right)$. This simplifies more because $W_0 = 0$ to $\frac{1}{2} W_{t_{n+1}}^2$. Clearly, $W_{t_{n+1}} \to W_t$ as $\Delta t \to 0$.

The second sum involves

$$S = \sum_{t_j < t} \Delta W_j^2 \ . \tag{31}$$

The mean and variance describe the answer as precisely as we need. For the mean, we have $\mathrm{E}\left[\Delta W_j^2 \right] = \Delta t$, so

$$\mathrm{E}[\, S] = \sum_{t_j < t} \Delta t = t_n \to t \ \text{ as } \ \Delta t \to 0 \ .$$

For the variance, the terms ΔW_j are independent, and $\mathrm{var}\left(\Delta W_j^2 \right) = 2 \Delta t^2$ (recall: ΔW_j is Gaussian and we know the fourth moments of a Gaussian) Therefore

$$\mathrm{var}(S) = 2 \Delta t \left(\sum_{t_j < t} \Delta t \right) = 2 \Delta t \, t_n \leq 2t 2^{-m} \ .$$

[1] The term comes from a *collapsing telescope*. You can find pictures of these on the web.

These two calculations show that $S \to t$ as $m \to \infty$. Therefore

$$X_t^m \to \frac{1}{2}\left(W_t^2 - t\right) \;\text{ as }\; m \to \infty \;.$$

This gives the famous result

$$\int_0^t W_s dW_s = \frac{1}{2}\left(W_t^2 - t\right) \;. \tag{32}$$

We have much to say about this result, starting with what it is not. It is not the answer you would get if W_t were a differentiable function of t. If W_t is differentiable, then $dW_s = \frac{dW}{ds}ds$, and

$$\int_0^t W_s dW_s = \int_0^t W_s \frac{dW}{ds}\, ds = \frac{1}{2}\int_0^t \frac{d}{ds} W_s^2 \, ds = \frac{1}{2}W_t^2 \;.$$

The Ito result (32) is different. The Ito calculus for rough functions like Brownian motion gives results that are not what you would get using the ordinary calculus. In ordinary calculus, the sum (31) converges to zero as $\Delta t \to 0$. That is because ΔW_j^2 scales like Δt^2 if W_t is a differentiable function of t, so S is like $\Delta t \sum_{t_j < t} \Delta t = \Delta t \, t$. But ΔW scales like Δt for Brownian motion. That is why S makes a positive contribution to the Ito integral.

The wrong answer $\frac{1}{2}W_t^2$ is wrong because it is not a *martingale.* A martingale is a stochastic process so that if $t > s$, then

$$\mathrm{E}[\, X_t \mid \mathcal{F}_s] = X_s \;. \tag{33}$$

The Ito integral is a martingale. But

$$\mathrm{E}\left[\, W_t^2 \mid \mathcal{F}_s\right] = W_s^2 + (t - s) \;,$$

so W_t^2 is not a martingale (see Section 5). The correct formula (32) is a martingale. The "correction" $W_t^2 \to W_t^2 - t$ makes this happen.

This example illustrates the general principle that ΔW_j must be in the future of f_j in the Riemann sum approximation (15). This implies that $\mathrm{E}[\, f_j \Delta W_j] = 0$, and the stronger statement that $\mathrm{E}[\, f_j \Delta W_j \mid \mathcal{F}_j] = f_j\, \mathrm{E}[\, \Delta W_j \mid \mathcal{F}_j] = 0$. Suppose we violate this and propose a "trapezoid rule" approximation

$$(\textit{Wrong}) \qquad \int_{t=t_j}^{t_{j+1}} W_t \, dW_t \approx \frac{W_{t_{j+1}} + W_{t_j}}{2}\left(W_{t_{j+1}} - W_{t_j}\right) \;. \qquad (\textit{Wrong})$$

This leads to the incorrect integral approximation

$$(\textit{Wrong}) \qquad \int_0^t W_s \, dW_s \approx \frac{1}{2}\sum_{t_j < t}\left(W_{t_{j+1}} + W_{t_j}\right)\left(W_{t_{j+1}} - W_{t_j}\right) \;. \qquad (\textit{Wrong})$$

You can check that this is the telescoping sum part of the correct approximation and converges to $\frac{1}{2}W_t^2$, with out the correction that makes it a martingale. The problem here is that ΔW_j is not in the future of the trapezoid rule approximation $\frac{1}{2}\left(f_{t_{j+1}} + f_{t_j}\right)$. For the integrand of this example, $f_t = W_t$, one easily checks that

$$\mathrm{E}\left[\frac{1}{2}\left(W_{t_{j+1}} + W_{t_j}\right)\left(W_{t_{j+1}} - W_{t_j}\right)\right] = \frac{1}{2}\Delta t \neq 0 \ .$$

It would be OK to get a non-zero expectation if it were Δt^2, because the total contribution from these is $O(\Delta t)$ and vanishes as $\Delta t \to 0$. But this $O(\Delta t)$ error makes an $O(1)$ contribution when you add up over j.

5 Properties of the Ito integral

This section discusses two properties of the Ito integral: (1) the martingale property, (2) the *Ito isometry formula*.

Two easy steps verify the martingale property. Step one is to say that we can define the Ito integral with a different start time as

$$\int_a^t f_s dW_s = \lim_{m\to\infty} \sum_{a \le t_j < t} f_{t_j}\left(W_{t_{j+1}} - W_{t_j}\right) \ . \tag{34}$$

This has the additivity property

$$\int_0^a f_s dW_s + \int_a^t f_s dW_s = \int_0^t f_s dW_s \ .$$

Step two is that

$$\mathrm{E}\left[\int_a^t f_s dW_s \,\middle|\, \mathcal{F}_a\right] = 0 \ .$$

This is because the right side of (34) has expected value zero. That is because all the terms on the right are in the future of $\mathcal{F}_a$. That zero expectation is preserved in the limit $\Delta t \to 0$. A general theorem in probability says that if Y_m is a family of random variables and $Y_m \to Y$ as $m \to \infty$, and if another technical condition is satisfied (discussed in Week 8), then $\mathrm{E}[Y_m] \to \mathrm{E}[Y]$ as $m \to \infty$.

When we use these facts together, we conclude that

$$\begin{aligned}\mathrm{E}\left[\int_0^t f_s dW_s \,\middle|\, \mathcal{F}_a\right] &= \mathrm{E}\left[\int_0^a f_s dW_s \,\middle|\, \mathcal{F}_a\right] + \mathrm{E}\left[\int_a^t f_s dW_s \,\middle|\, \mathcal{F}_a\right] \\ &= \mathrm{E}\left[\int_0^a f_s dW_s \,\middle|\, \mathcal{F}_a\right] \\ &= X_a \ .\end{aligned}$$

This is the martingale property for X_t.

The Ito isometry formula is

$$\mathrm{E}\left[\left(\int_0^t f_s dW_s\right)^2\right] = \int_0^t \mathrm{E}\left[f_s^2\right] ds \ . \tag{35}$$

The variance of the Ito integral is equal the the ordinary integral of the expected square of the integrand. We explain the idea first informally, then more formally in the next paragraph. Suppose $[s, s+dt]$ and $[s', s'+dt]$ are two small time intervals of length $dt > 0$. Let $dW_s = W_{s+dt} - W_s$ and $dW_{s'} = W_{s'+dt} - W_{s'}$ be the corresponding Brownian motion increments. Then

$$\mathrm{E}[\, f_s dW_s f_{s'} dW_{s'}] = \begin{cases} 0 & \text{if } s \neq s' \\ \mathrm{E}\left[\, f_s^2\right] ds & \text{if } s = s' \; . \end{cases}$$

The unequal time formula on the top line reflects that either dW_s of $dW_{s'}$ is in the future of everything else in the formula. The equal time formula on the bottom line reflects the informal $\mathrm{E}\Big[\,(dW_s)^2 \mid \mathcal{F}_s\Big] = dt$. Then

$$\begin{aligned} \left(\int_0^t f_s dW_s\right)^2 &= \int_0^t f_s dW_s \cdot \int_0^t f_{s'} dW_{s'} \\ &= \int_0^t \int_0^t f_s f_{s'} dW_s W_{s'} \; . \end{aligned}$$

Taking expectations,

$$\begin{aligned} \mathrm{E}\left[\left(\int_0^t f_s dW_s\right)^2\right] &= \int_0^t \int_0^t \mathrm{E}[\, f_s d f_{s'} dW_s W_{s'}] \\ &= \int_0^t \mathrm{E}\left[\, f_s^2\right] ds \; . \end{aligned}$$

A more formal version of this argument is similar to the informal one. We just use the Riemann sum approximation. Only the diagonal terms in the double sum have non-zero expected value:

$$\begin{aligned} \mathrm{E}\left[\left(\sum_{t_j<t} f_{t_j} \Delta W_{t_j}\right)^2\right] &= \mathrm{E}\left[\sum_{t_j<t}\sum_{t_k<t} f_{t_j} f_{t_k} \Delta W_{t_j} \Delta W_{t_k}\right] \\ &= \sum_{t_j<t}\sum_{t_k<t} \mathrm{E}\left[\, f_{t_j} f_{t_k} \Delta W_{t_j} \Delta W_{t_k}\right] \\ &= \sum_{t_j<t} \mathrm{E}\Big[\, f_{t_j}^2 \, \mathrm{E}\Big[\, \Delta W_{t_j}^2 \mid \mathcal{F}_{t_j}\Big]\Big] \\ &= \sum_{t_j<t} \mathrm{E}\Big[\, f_{t_j}^2\Big] \Delta t \; . \end{aligned}$$

The last line is the Riemann sum approximation to the right side of (35).

Let us check the Ito isometry formula on the example (32). For the Ito integral part we have (recall that $X \sim \mathcal{N}(0, \sigma^2)$ implies $\mathrm{var}\big(X^2\big) = 2\sigma^4$)

$$\mathrm{var}\left(\int_0^t W_s dW_s\right) = \frac{1}{4}\mathrm{var}\big(W_t^2 - t\big) = \frac{1}{4}\mathrm{var}\big(W_t^2\big) = \frac{1}{4}\, 2t^2 = \frac{t^2}{2} \; .$$

For the Riemann integral part, we have

$$\int_0^t \mathrm{E}\left[W_s^2 \right] ds = \int_0^t s \, ds = \frac{t^2}{2} \; .$$

As the Ito isometry formula (35) says, these are equal.

A simpler example is $f_s = s^2$, and

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

This is more typical of general Ito integrals in that X_t is not a function of W_t alone. Since X is a linear function of W, X is Gaussian. Since X is an Ito integral, $\mathrm{E}[\, X_t] = 0$. Therefore, we characterize the distribution of X_t completely by finding its variance. The Ito isometry formula gives ($f_s^2 = \mathrm{E}\left[f_s^2 \right] = s^4$)

$$\mathrm{var}(X_t) = \int_0^t s^4 \, ds = \frac{s^5}{5} \; .$$