Week 5
Integrals with respect to Brownian motion

Jonathan Goodman

October 7, 2012

1 Introduction to the material for the week

This week starts the other 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 W_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 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.

I.i.d. noise processes cannot have general distributions in continuous time. A continuous time i.i.d. noise processes, *white noise*, is Gaussian. The continuous time scaling limit for Brownian motion is

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

The CLT implies that W_t is Gaussian regardless of the distribution of Z_n. White noise 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 relation between X and W may be expressed informally in the Ito differential form

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

The integrand, f, must be 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

$$\mathrm{E}[\, dX_t \mid \mathcal{F}_t] = f_t \mathrm{E}[\, dW_t \mid \mathcal{F}_t] = 0 \; , \tag{7}$$

and

$$\mathrm{E}\left[\, dX_t^2 \mid \mathcal{F}_t\right] = f_t^2 \mathrm{E}\left[\, dW_t^2 \mid \mathcal{F}_t\right] = f_t^2 \, dt \; . \tag{8}$$

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 (7). More or less any such martingale can be represented as an Ito integral (27). 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 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 to the Ito integral, X_t^m. This section is a description of some technical tools that can show that the X_t^m converge to a limit as $m \to \infty$. What we describe is related to the standard *Borel Cantelli lemma* but it is not the same. 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

$$s = \sum_{m=1}^{\infty} a_m < \infty \ . \tag{9}$$

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 with $|b_m| \leq a_m$. Consider the sum

$$x = \sum_{m=1}^{\infty} b_m \ . \tag{10}$$

The sum converges *absolutely* if the a_m have a finite sum. Therefore (9) implies that x is well defined. The partial sums for (10) are

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

These satisfy

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

as $n \to \infty$. If x_m is a sequence of numbers with $b_m = x_{m+1} - x_m$, then the limit

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

is well defined. Moreover,

$$|x - x_n| < r_n \to 0 \ , \quad \text{as } n \to \infty \ . \tag{11}$$

Suppose A_m is a sequence of non-negative random numbers. 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).

Many people refuse to distinguish between outcomes that are impossible, which would be $\omega \notin \Omega$, and events that have probability zero. We will be sloppy with the distinction in this class, and ignore it much of the time.

Our strategy will be to show that $S < \infty$ a.s. by showing that $\mathrm{E}[\,S] < \infty$. That is

$$\sum_{j=m}^{\infty} \mathrm{E}[\,A_m] < \infty \quad \Longrightarrow \quad \sum_{m=1}^{\infty} A_m < \infty \;\text{ a.s.}$$

In particular, let X_t^m be a sequence of random paths. Suppose you can show that

$$\mathrm{E}\left[\,\left|X_t^{m+1} - X_t^m\right|\right] \leq a_m \;\;, \qquad \text{with} \quad \sum_{m=1}^{\infty} a_m < \infty \;, \tag{12}$$

for all $t \leq T$. Then you know that the following limit exists almost surely

$$X_t = \lim_{j \to \infty} X_t^m \;. \tag{13}$$

This is our version of the *Borel Cantelli* lemma. We calculate expected values to verify the hypothesis (12), then we conclude that the limit exists pathwise almost surely.

3 Riemann sums for the Ito integral

We use the following Riemann sum approximation for the Ito integral (27):

$$X_t^m = \sum_{t_j < t} f_{t_j} \Delta W_j \;. \tag{14}$$

The notation is

$$\Delta t = 2^{-m} \;, \tag{15}$$

$$t_j = j \Delta t \;, \tag{16}$$

W_t is a standard Brownian motion, and

$$\Delta W_j = W_{t_{j+1}} - W_{t_j} \;, \tag{17}$$

The pathwise convergence will be that for almost every Brownian motion path, the approximations (14) converge to a limit. This limit will be measurable in $\mathcal{F}_t$ because X_t is a function of $W_{[0,t]}$.

The Riemann sum approximation (14) needs lots of explanation. The Brownian motion increment used at time t_j (17) is in the future of t_j. We assume that f_{t_j} is measurable in $\mathcal{F}_{t_j}$, so this makes ΔW_j independent of f_{t_j}. In particular,

$$\mathrm{E}\left[\, f_{t_j}\Delta W_j \mid \mathcal{F}_{t_j}\right] = 0 \; , \tag{18}$$

and

$$\mathrm{E}\left[\, \left(f_{t_j}\Delta W_j\right)^2 \mid \mathcal{F}_{t_j}\right] = f_{t_j}^2 \Delta t \; . \tag{19}$$

The Riemann sum definition (14) 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 use the limit $m \to \infty$ rather than $\Delta t \to 0$. It is easy to compare the $\Delta t_m = 2^{-m}$ approximation to the one with $\Delta t_{m+1} = \frac{1}{2}\Delta t_m$, as we will see. Moreover, taking $\Delta t \to 0$ rapidly makes it easier for the sum (12) to converge.

We assume that the integrand f_t is continuous in some way. Specifically, we assume that if $s > 0$, then

$$\mathrm{E}\left[\, (f_{t+s} - f_t)^2 \mid \mathcal{F}_t\right] \le Cs \; . \tag{20}$$

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.

The main step in the proof is the estimation of the terms in (12). The move from m to $m+1$ replaces Δt_m by $\Delta t_{m+1} = \frac{1}{2}\Delta t_m$. We can write X_t^{m+1} in terms of the extended m definition $t_{j+\frac{1}{2}} = (j+\frac{1}{2})\Delta t$. For 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.

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

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

$$X_t^{m+1} - X_t^m = \sum_{t_j < t} \left(f_{j+\frac{1}{2}} - f_j \right) \left(W_{j+1} - W_{j+\frac{1}{2}} \right) \;\; + Q \; . \tag{21}$$

The terms on the right side of (21) have mean zero. This implies that the sum has cancellations that may be hard to see if we take absolute values too soon. We find the cancellations by calculating the square and using the *Cauchy*

Schwarz inequality. In probability, a form of the Cauchy Schwarz inequality is that if U and V are two random variables, then (proof in the next paragraph)

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

For $V = 1$, this is just

$$\mathrm{E}[U] \leq \sqrt{\mathrm{E}[U^2]}\ .$$

Computing the square of (21) gives

$$\mathrm{E}\left[\,\left|X_t^{m+1} - X_t^m\right|\right] \leq a_m\ ,$$

where

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

This is something we can caluclate.

(Here is a proof of the Cauchy Schwarz inequality in the form we need. The following quantity is non-negative for any α

$$0 \leq \mathrm{E}\left[\left(U - \alpha V\right)^2\right] = \mathrm{E}\left[U^2\right] - 2\alpha\mathrm{E}[UV] + \alpha^2\mathrm{E}\left[V^2\right]\ .$$

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

$$0 \leq \mathrm{E}\left[U^2\right] - \frac{\mathrm{E}[UV]^2}{\mathrm{E}[V^2]}\ .$$

Multiply through by $\mathrm{E}\left[V^2\right]$ and you get Cauchy Schwarz.)

Denote a typical term in the sum on the right of (21) 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

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

It follows from the tower property that $\mathrm{E}[Y_j \mid \mathcal{F}_j] = 0$ If $k < j$, then Y_k is known in $\mathcal{F}_j$, so

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

In the expected value of $(\sum Y_j^2) = \sum(Y_jY_k)$ there are two kinds of terms. We just saw that *off diagonal* terms, those with $j \neq k$ have expected value equal to zero. A typical diagonal term has

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

The next expectation, and (20) gives the desired inequality

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

Finally,

$$a_m^2 \leq C \sum_{t_j < t} \Delta t^2 = C \Delta t \sum_{t_j < t} \Delta t \leq C t \, \Delta t_m \; .$$

You can check that adding Q to this calculation does not change the conclusion.

The last inequality may be written

$$a_m \leq C \sqrt{t} \sqrt{\Delta t_m} \leq C \sqrt{t} \alpha^m \; ,$$

where $\alpha = 2^{-1/2} < 1$. The sum in (12) becomes a convergent geometric series. This completes the proof that the approximations (14) converge to something.

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 (12) a convergent geometric series. In another week (which we will not do in this course), we could show that the restriction to powers of 2 for Δt is unnecessary. You can see how to relax our assumption (20). 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 $\sqrt{s}$. This would just lead to a different $\alpha < 1$ in the final geometric series.

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 way to approach actual calculations. This is as in ordinary calculus. Riemann sums are a good way to define the Riemann integral, but the fundamental theorem of calculus is an easier way to compute specific examples.

The first example is

$$X_t = \int_0^t W_s dW_s \; . \tag{22}$$

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 $t_n = \max\left\{t_j \mid t_j < t\right\}$, 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

$$S = \sum_{t_j<t}\Delta W_j^2\ . \tag{23}$$

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

$$\mathrm{var}(S) = 2\Delta t\left(\sum_{t_j<t}\Delta t\right) = 2\Delta t\, t_n \le 2t2^{-m}\ .$$

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

$$\int_0^t W_s dW_s = \frac{1}{2}\left(W_t^2 - t\right)\ . \tag{24}$$

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

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

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

The Ito result (24) 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 (23) converges to zero as $\Delta t \to 0$. That is because ΔW_j^2 scales like $\Delta t t^2$ 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 answer differentiable calculus 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

$$\mathrm{E}[\, X_t \mid \mathcal{F}_s] = X_s \; . \tag{25}$$

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 (24) is a martingale. The "correction" $W_t^2 \to W_t^2 - t$ accomplishes this.

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

$$\int_a^t f_s dW_s = \lim_{m\to\infty} \sum_{a \le t_j < t} f_{t_j}\left(W_{t_{j+1}} - W_{t_j}\right) \; . \tag{26}$$

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 (26) 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

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

This is the martingale property for X_t.

The Ito isometry formula is

$$\mathrm{E}\left[\left(\int_0^t f_s dW_s \right)^2 \right] = \int_0^t \mathrm{E}\left[f_s^2 \right] ds \ . \tag{27}$$

The variance of the Ito integral is equal the the ordinary integral of the expected square of the integrand. The ideas we have been using make the proof of this formula routine. Informally, we write

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

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 $\mathrm{E}\left[(dW_s)^2 \mid \mathcal{F}_s \right] = dt$. Then

$$\left(\int_0^t f_s dW_s \right)^2 = \int_0^t f_s dW_s \cdot \int_0^t f_s' dW_s' = \int_0^t \int_0^t f_s df_{s'} dW_s W_{s'} \ .$$

Taking expectations,

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

A more formal, but not completely rigorous, version of this argument is little different from this. We merely switch to the Riemann sum approximation and take the limit at the end:

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

The last line is the Riemann sum approximation to the right side of (27).

Let us check the Ito isometry formula on the example (24). For the Ito integral part we have (recall that $X \sim \mathcal{N}(0, \sigma^2)$ implies $\mathrm{var}(X^2) = 2\sigma^4$)

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

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 (27) 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 ($f_s^2 = \mathrm{E}\left[\, f_s^2 \right] = s^4$)

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

This may be easier than the method used in question (3) of Assignment 3.