Stochastic Calculus, Courant Institute, Fall 2019
http://www.math.nyu.edu/faculty/goodman/teaching/StochCalc2019/StochCalc.html
Jonathan Goodman, September, 2019

Class 11, Ito

1 Introduction

This class discusses the theory behind much of the stochastic calculus we have done so far. First is a formal definition of the Ito integral and a explanation that the $\Delta t \to 0$ limit that defines the integral exists. Ito's lemma plays the role of the fundamental theorem of calculus, when you take the derivative in time of a function of Brownian motion W_t. Suppose $f(W_t, t)$ is such a function. What is the differential df corresponding to an increment of time dt? If you integrate the changes of f between t_1 and t_2, you get the total change $f(W_{t_1}, t_2) - f(W_{t_1}, t_1)$. That is

$$f(W_{t_1}, t_2) = f(W_{t_1}, t_1) + \int_{t_1}^{t_2} df(W_s, s) \; . \qquad (1)$$ Ili

Ito's lemma is a formula for df that works here.

In the ordinary chain rule, df would involve $(\partial_t f)dt$ and $(\partial_w f)dW$. But if you just use these two terms, then (1) is not satisfied. We will see that (1) works out only if you include the "Ito term" $\frac{1}{2}(\partial_w^2 f)dt$. The Ito term is needed because dW is too big. More specifically, dW is on the order of $\sqrt{dt}$. The differential df has to be accurate to order dt, not just $\sqrt{dt}$. Therefore, we need the second derivative term $\frac{1}{2}(\partial_w^2 f)(dW)^2$. The term $(dW)^2$ has the order of dt but is not equal to dt. We will explain why you can substitute dt for $(dW)^2$ and have (1) still work.

There is a more basic problem. Even without the Ito term, it seems that you have to integrate a function dW_s instead of ds. This is the *Ito integral*. Two big goals today are to define the Ito integral and to use it to prove Ito's lemma.

The indefinite Ito integral is

$$X_t = \int_0^t g_s dW_s \; . \qquad (2)$$ iIi

The integrand is a stochastic process g_t that is *non-anticipating*. This means that g_t may depend on the Brownian motion path $W_{[0,t]}$, but it may not depend on W_s for $s > t$. In fancier language, let $\mathcal{F}_t$ be the filtration of $\sigma-$algebras generated by $W_{[0,t]}$. To be allowed, the integrand must satisfy the equivalent conditions

$$g_t \text{ measurable with respect to } \mathcal{F}_t \;, \quad g_t = \mathrm{E}[\, g_t \mid \mathcal{F}_t] \; .$$

These integrands are allowed, except the last one

g_t	$=$	$1+t^2$	a deterministic function does not depend on W_t at all.
g_t	$=$	W_t	it is non-anticipating to depend on the present.
g_t	$=$	$\begin{cases} 1 & \text{if } \tau > t \\ 0 & \text{if } \tau < t \end{cases}$	integrate up to a hitting time τ.
g_t	$=$	$\max\{W_s \mid 0 \leq s \leq t\}$	may depend on the whole path $W_{[0,t]}$.
g_t	$=$	$\tau - t$	if $\tau > t$ then τ is not known at time t.

Doob's theorem (there are many theorems with this name) states that the process defined by (2) is a *martingale*. This means that if $s > t$, then

$$\mathrm{E}[\, X_s \mid \mathcal{F}_t] = X_t \ .$$

This is because the integral expression (2) has dW_s in the future of $\mathcal{F}_s$. Therefore, at time s, the f_s is known and the value of dW_s is unknown.

The Ito integral allows us to say what it means to be the solution of the stochastic differential equation,

$$dX_t = a(X_t)dt + b(X_t)dW_t \ , \tag{3}$$

sde

First, X_t must be non-anticipating. This is natural in applications, since values of dW in the future of t should not influence X_t. Since X_t is non-anticipating, we can ask it to satisfy the SDE in integral form

$$X_t = X_0 + \int_0^t a(X_s)ds + \int_0^t b(X_s)dW_s \ . \tag{4}$$

sdes

The second integral on the right is an Ito integral with integrand $g_t = b(X_t)$.

Although $\mathrm{E}[X_t] = 0$, there are quadratic functions of X_t whose expectations are non-zero and interesting. The first is the *Ito isometry formula*

$$\mathrm{E}\big[\, X_t^2\big] = \int_0^t \mathrm{E}\big[\, g_s^2\big] \ ds \ . \tag{5}$$

Iif

On the left is the Ito integral (2). On the right is an ordinary (Riemann) integral involving the mean square of the integrand. This formula expresses the fact that dW_s and $dW_{s'}$ (two separate increments of Brownian motion) are independent if their time intervals don't overlap. The left side is the expected square of the sum (the integral) and the right side is the sum of the individual expected values

$$\mathrm{E}\Big[\, (g_s dW_s)^2 \mid \mathcal{F}_s\Big] = g_s^2 ds \ .$$

You get the isometry formula (5) by taking the overall expectation of both sides and integrating.

Quadratic variation is the other interesting quadratic quantity related to X_t. The quadratic variation of X_t up to time t is written $[X]_t$ and may be defined informally as

$$[X]_t = \int_0^t (dX_s)^2 \ .$$

There is a more careful definition of the right side later in this class. Informally, the Ito calculus uses the formula

$$\left(dX_s\right)^2 = g_s^2 dt \; . \tag{6}$$

If you insert this into the informal definition of quadratic variation, you find

$$[X]_t = \int_0^t g_s^2 ds \; . \tag{7}$$

This formula is true, which we will see when we have the $\Delta t \to 0$ definition of quadratic variation on the left side.

For practical calculations with stochastic calculus, the Ito isometry formula (5) allows you to do an ordinary integral to calculate the variance of an Ito process X_t. Suppose you are evaluating the variance by simulation. You might think that the right side estimates the variance more accurately (with less variance), but we will see that the improvement goes to zero as $\Delta t \to 0$. The quadratic variation formula (7) is useful in evaluating financial strategies involving dynamic hedging. There may be a cost proportional to $(\Delta X)^2$. In ordinary calculus, this disappears (go to zero) as $\Delta t \to 0$. But in stochastic calculus, it depends on the quadratic variation.

The isometry formula (5) has important uses in theory. For one thing, it is a way of demonstrating that X_t is a continuous function of t with $|\Delta X|$ is on the order of $(\Delta t)^{\frac{1}{2}}$. For another thing, it provides a way to define the Ito integral when the integrand g_t is not a continuous function of t. In the list of integrands above, the third one is discontinuous at the hitting time $t = \tau$. Integrands like this are used in practice. You will find this approach in most explanations of the Ito integral, but doing it correctly is too technical for this class. [It's not too hard, it's just too long.] Finally, the Ito isometry formula may be used to prove that the SDE (3) has a solution. The existence proof in your ODE class (which you may not have taken or may not remember) probably used *Picard iteration.* A similar proof applies for the SDE, with the Ito isometry formula being used to control the Ito integral term on the right side of (4).

The quadratic variation formula (7) is used to prove half of *Girsanov's theorem* (discussed in a later class). The quadratic variation on the left side is a function of the path $X_{[0,t]}$. If X_t satisfies the SDE

2 The Borel Cantelli lemma

This section has the formal outline but not the "main point" in our definition of the Ito integral. The main point is a definition and a calculation in the next section.

The Riemann integral is defined as a limit as $\Delta t \to 0$. The Ito integral has a similar definition. We will define approximations $X_t^{(\Delta t)}$ and show that there is a limit as $\Delta t \to 0$. Actually, we will take a sequence $\Delta t_m = 2^{-m}$ and take the limit $m \to \infty$. We will write $X_t^{(m)}$ instead of the more correct expression $X_t^{(\Delta t_m)}$.

There are various forms of convergence for random variables. In the central limit theorem, we ask that the probability distribution converges to a Gaussian distribution. This is convergence *in distribution.* Here, we aim for *almost sure pointwise convergence. Pointwise* means that the numbers $X_t^{(m)}$ have a limit in the usual sense (given below). The "point" in "pointwise" is the random Brownian motion path $W_{[0,t]}$. All the approximations $X_t^{(m)}$ are defined using the same path. Pointwise convergence means that for that path

$$\lim_{m\to\infty} X_t^{(m)}(W_{[0,t]}) = X_t(W_{[0,t]}) \;\text{ exists.} \tag{8}$$ $\boxed{\texttt{pc}}$

And it exists for the "point" $W_{[0,t]}$.

We will see that the limit (8) probably does not exist for every continuous path $W_{[0,t]}$. But if the path is a standard Brownian motion then the probability that the limit exists is 1. That is, the limit exists *almost surely.*

The *Borel Cantelli* lemma is a way to prove almost sure convergence by calculating expected values. We use a version of this idea that is a little different (though the same in spirit) from the version in Wikipedia. Let $X^{(m)}$ be a family of random variables. If they converge as $m \to \infty$, then $D_m = \left|X^{(m+1)} - X^{(m)}\right| \to 0$. We show $D_m \to 0$ (almost surely) by showing that

$$S = \sum_{m=1}^{\infty} D_m < \infty \;, \quad \text{(almost surely).}$$

The sum S is non-negative because the numbers D_m are non-negative. We show that $S < \infty$ by showing

$$\mathrm{E}[S] < \infty \;.$$

If the expectation is finite, then S is finite almost surely.

For example, suppose $U \in [0,1]$ is uniformly distributed and $V = U^{-\frac{1}{2}}$. Then

$$\mathrm{E}[V] = \int_0^1 u^{-\frac{1}{2}}\,du = 2u^{\frac{1}{2}}\Big|_0^1 = 2 < \infty \;.$$

This shows that $V < \infty$ almost surely. But V is not always finite. If $U = 0$ then $V = \infty$. This V is finite almost surely without always being finite. Moreover, you prove that $V < \infty$ almost surely not by finding an upper bound $V < a$ almost surely. There is no such upper bound. If $1 < a < \infty$ is a fixed number, then (doing a calculation) $\Pr(V > a) = a^{-2} > 0$. This shows that V can be as large as you want, but the probability of being large is small.

We will define random variables

$$D_m = |X_t^{(m+1)} - X_t^{(m)}| \;. \tag{9}$$ $\boxed{\texttt{Dm}}$

The hard part of the convergence proof is an inequality for the expectations in terms of a geometric series

$$\mathrm{E}[\,D_m] \le d_m = Az^m \;. \tag{10}$$ $\boxed{\texttt{Dmb}}$

Of course $A > 0$ and $z > 0$. The crucial part will be $z < 1$. This implies that

$$\mathrm{E}\left[\sum_{m=1}^{\infty} D_m\right] < \infty \ . \tag{11}$$

`Dsb`

Therefore

$$\sum_{m=1}^{\infty} D_m < \infty \ \text{ almost surely.} \tag{12}$$

`Dsf`

This argument, calculating an expected value to show that a sum is finite almost surely, is what we (but not Wikipedia) call the Borel Cantelli lemma.

Finally, consider the sum

$$\sum_{m=1}^{\infty}\left(X_t^{(m+1)} - X_t^{(m)}\right) \ .$$

If this infinite sum converges, then it defines the limit (8). If (12) is true, then the infinite sum converges absolutely. Therefore, when we prove the basic comparison inequality (10), we will show that the sum converges absolutely, almost surely. This shows that the limit (8) exists almost surely.

It is fitting to end this technical section with a final technical point that we will use later We estimate $\mathrm{E}[D_m]$ indirectly by showing

$$\mathrm{E}\left[\, D_m^2\right] \leq B\, 2^{-m} \ . \tag{13}$$

`Dmsq`

This is possible (see later) because $D_m^2 = \left(X_t^{(m+1)} - X_t^{(m)}\right)^2$ is an algebraic expression that can be expanded in a way that D_m itself cannot be. A bound for the square implies a bound for the quantity itself. If $Y \geq 0$ is any non-negative random variable (such as D_m), the *Cauchy Schwarz* inequality implies that

$$\mathrm{E}[Y] \leq \sqrt{\mathrm{E}[\, Y^2]} \ . \tag{14}$$

`sql`

If you accept these two inequalities, then

$$\mathrm{E}[D_m] \leq \sqrt{B}\left(2^{-\frac{1}{2}}\right)^m \ .$$

This is the desired (10), with $A = \sqrt{B}$ and $z = 1/\sqrt{2}$.

The *Cauchy Schwarz* inequality is

$$\mathrm{E}[YZ] \leq \left(\mathrm{E}\left[\, Y^2\right] \mathrm{E}\left[\, Z^2\right]\right)^{-\frac{1}{2}} \ . \tag{15}$$

`CSi`

You can find the proof in Wikipedia. One interpretation applies with $\mathrm{E}[Y] = 0$ and $\mathrm{E}[Z] = 0$. Then the left side of (refCSi) is the covariance and the right side involves variances. The inequality says that the correlation coefficient satisfies

$$\rho_{YZ} = \frac{\mathrm{cov}(Y,Z)}{\sqrt{\mathrm{var}(Y)\mathrm{var}(Z)}} \leq 1 \ .$$

You can replace Y by $-Y$ and see that $\rho_{YZ} \geq -1$. Altogether,

$$|\rho_{YZ}| \leq 1 \; .$$

A "well known trick" is to apply the Cauchy Schwarz inequality with $|Y|$ and $Z = 1$ to get an inequality about $|Y|$. The inequality is

$$\mathrm{E}[\,|Y|] \leq \sqrt{\mathrm{E}[\,Y^2]}\sqrt{\mathrm{E}[\,1^2]} \; .$$

This is the desired inequality (14). To summarize: we will prove convergence of the Ito integral approximations (given soon) converge almost surely using a calculation that verifies the square inequality (13). This is *almost sure pathwise* convergence, which means that for if $W_{[0,t]}$ is a Brownian motion path, then the approximations converge for this path almost surely.

3 The Ito integral definition, an example

Suppose we have a Δt. The discrete times are $t_k = k\Delta t$. The approximation to the Ito integral is

$$X_t^{(\Delta t)} = \sum_{t_k < t} g_{t_k} \left(W_{t_{k+1}} - W_{t_k} \right) \; . \tag{16}$$

We want this to converge to the limit (2) as $\Delta t \to 0$.

The approximation (16) respects "causality" in putting $\Delta W_k = W_{t_{k+1}} - W_{t_k}$ in the future of t_k where g_{t_k} is defined. This causality is the basic fact about the Ito integral. Informally, it is that dW_s is in the future of g_s, so that at time s, we know g_s but not dW_s. More formally, we have $g_{t_k}\left(W_{t_{k+1}} - W_{t_k}\right)$. We condition on $\mathcal{F}_{t_k}$. Since g_{t_k} is know, this number comes out of the expectation. Therefore

$$\mathrm{E}\big[\, g_{t_k} \left(W_{t_{k+1}} - W_{t_k} \right) \mid \mathcal{F}_{t_k} \big] = g_{t_k} \mathrm{E}\big[\, \left(W_{t_{k+1}} - W_{t_k} \right) \mid \mathcal{F}_{t_k} \big]$$

But the increment $W_{t_{k+1}} - W_{t_k}$ is in the future of $\mathcal{F}_{t_k}$. The independent increments property of Brownian motion implies that the distribution of $W_{t_{k+1}} - W_{t_k}$ is independent of anything in $\mathcal{F}_{t_k}$. In particular

$$\mathrm{E}\big[\, \left(W_{t_{k+1}} - W_{t_k} \right) \mid \mathcal{F}_{t_k} \big] = 0 \; .$$

This implies that

$$\mathrm{E}\left[\, X_t^{(\Delta t)} \right] = 0 \; .$$

This is how causality makes the Ito process X_t into a martingale.

An example is

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

The approximations are

$$X_t^{(\Delta t)} = \sum_{t_k < t} W_{t_k} \left(W_{t_{k+1}} - W_{t_k} \right) \; . \tag{17}$$

This is evaluated by a clever trick, which is the formula

$$W_{t_k} = \frac{1}{2}\left(W_{t_{k+1}} + W_{t_k}\right) - \frac{1}{2}\left(W_{t_{k+1}} - W_{t_k}\right) \; .$$

This allows the sum (17) to be written in terms the two sums

$$S_1 = \frac{1}{2}\sum_{t_k<t}\left(W_{t_{k+1}} + W_{t_k}\right)\left(W_{t_{k+1}} - W_{t_k}\right)$$

$$S_2 = \frac{1}{2}\sum_{t_k<t}\left(W_{t_{k+1}} - W_{t_k}\right)\left(W_{t_{k+1}} - W_{t_k}\right)$$

For S_1, we use the trick

$$\left(W_{t_{k+1}} + W_{t_k}\right)\left(W_{t_{k+1}} - W_{t_k}\right) = W^2_{t_{k+1}} - W^2_{t_k} \; .$$

Suppose the largest time in the sum (17) is $t_n = \max\left\{t_k | t_k < t\right\}$. Then

$$S_1 = \frac{1}{2}\left[\left(W^2_{t_{n+1}} - W^2_{t_n}\right) + \left(W^2_{t_n} - W^2_{t_{n-1}}\right) + \cdots + \left(W^2_{t_1} - W^2_0\right)\right] \; .$$

A sum like this is called "telescoping", to remind us of old hand held telescopes that open and close. All the intermediate values $W^2_{t_k}$ cancel, so that answer involves only the first and last terms. But the last term is $W^2_0 = 0$, so

$$S_1 = \frac{1}{2}W^2_{t_{n+1}} \; .$$

It is clear that $S_1 \to \frac{1}{2}W_t^2$ as $\Delta t \to 0$, because $t_{n+1} \to t$ and W_t is continuous.

We understand the S_2 sum using the law of large numbers. The terms in the sum are independent (the independent increments property) and the expected value is

$$\mathrm{E}\Big[\left(W_{t_{k+1}} - W_{t_k}\right)^2\Big] = \Delta t \; .$$

This implies that

$$\mathrm{E}[\,S_2] = \frac{1}{2}\sum_{t_k<t}\Delta t = \frac{1}{2}t_n \; .$$

Another calculation (see below) shows that $\mathrm{var}(S_2) = O(\Delta t)$. Therefore (you will be able to say this with mathematical confidence after reading the next section) $S_2 \to \frac{1}{2}t$ as $\Delta t \to 0$.

The final formula is

$$\int_0^t W_s dW_s = \frac{1}{2}W_t^2 - \frac{1}{2}t \; . \qquad (18)$$

The "causal ordering" puts dW_t in the future of t. That way the expected value of $g_t dW_t$ is zero. For this reason, the Ito integral is used to model stochastic control and dynamic trading strategies. "Stochastic control" is an engineering term that refers to controlling a stochastic process. At time t you know $\mathcal{F}_t$

but not dW_t. A dynamic trading strategy is similar. You do a trade or "take a position" at time t without knowing what the market will do next.

We argue that the Ito integral captures this idea. We do this by showing that the answer can change if you do not use a causal ordering. We calculate what happens if you replace the causally ordered approximation (17) by a non-causally ordered approximation

$$B_t^{(\Delta t)} = \sum_{t_k < t} W_{t_{k+1}} \left(W_{t_{k+1}} - W_{t_k} \right) \ . \tag{19}$$

bad

We evaluate the limit of $B_t^{(\Delta t)}$ with a variant of the trick we just used.

$$W_{t_{k+1}} \left(W_{t_{k+1}} - W_{t_k} \right) = \tfrac{1}{2} \left(W_{t_{k+1}}^2 - W_{t_k}^2 \right) + \tfrac{1}{2} \left(W_{t_{k+1}} - W_{t_k} \right)^2 \ .$$

We used $W_{t_{k+1}}$ instead of W_{t_k}, and got both terms with $+$ rather than one with $-$. With the telescoping sum, we get

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

When $\Delta t \to 0$, this becomes

$$B_t = \lim_{\Delta t \to 0} B_t^{(\Delta t)} = \tfrac{1}{2} W_t^2 + \tfrac{1}{2} t \ .$$

This is not the answer (18). For the Riemann integral, both approximations (17) and (19) would converge to the same answer. For Ito, they do not.

We get yet a third answer, and a second wrong answer, if we calculate as if W_t is a differentiable function of t. This (wrong) calculation would use the derivative and chain rule relations

$$dW_s = \frac{dW_s}{ds} ds \ , \quad W_s \frac{dW_s}{ds} = \frac{1}{2} \frac{dW_s^2}{ds} \ .$$

With these, you could calculate

$$\int_0^t W_s dW_s = \int_0^t W_s \frac{dW_s}{ds} ds = \frac{1}{2} \int_0^t \frac{dW_s^2}{dw} ds = \frac{1}{2} W_t^2 \ . \tag{20}$$

worse

This is another wrong answer.

We will see that the indefinite Ito integral with respect to Brownian motion (2) is a *martingale*. This means that if $t_2 > t_1$, then

$$\mathrm{E}[\, X_{t_2} \mid \mathcal{F}_{t_1}] = X_{t_1} \ . \tag{21}$$

m

To understand this, first note that

$$X_{t_2} = \int_0^{t_2} g_s dW_s = \int_0^{t_1} g_s dW_s + \int_{t_1}^{t_2} g_s dW_s = X_{t_1} + \int_{t_1}^{t_2} g_s dW_s \ .$$

Everything defining X_{t_1} is known at time t_1, meaning that X_{t_1} is known at time t_1. Therefore, conditioning on $\mathcal{F}_{t_1}$ does not change X_{t_1}

$$\mathrm{E}[\, X_{t_1} \mid \mathcal{F}_{t_1}] = X_{t_1} \; .$$

The martingale property (21) of the indefinite integral is a consequence of the conditional mean value zero property of an Ito integral in the future:

$$\mathrm{E}\left[\int_{t_1}^{t_2} g_s dW_s \mid \mathcal{F}_{t_1} \right] = 0 \; .$$

This will be clear from the definition we are about to get to.

The martingale property helps us tell the right answer (18) from the two wrong answers (19) and (20). Neither of the wrong answers is a martingale. In fact, it is easy to see that

$$\mathrm{E}\big[\, W_{t_2}^2 \mid \mathcal{F}_{t_1} \big] = W_{t_1}^2 + (t_2 - t_1) \; .$$

You find this by squaring the relation $W_{t_2} = W_{t_1} + (W_{t_2} - W_{t_1})$ and using the independent increments property. We see that the right answer is a martingale and neither of the wrong answers is a martingale. If you violate causality, even by the amount dt, you can change the answer from right to wrong.

4 The Ito integral definition

We want the approximations (16) to converge to the limit (2) as $\Delta t \to 0$. Instead of this, we will take $\Delta t^{(m)} = 2^{-m}$ and take the limit $m \to \infty$. You will see just below the advantage of shrinking Δt by factors of 2. But there is a bigger reason for taking a sequence $\Delta t^{(m)} \to 0$ rapidly. We will get a bound for $\mathrm{E}[\,|D_m|]$ that depends on $\Delta t^{(m)}$. If $\Delta t^{(m)} \to 0$ rapidly, then $\mathrm{E}[\,|D_m|] \to 0$ rapidly. This makes it easy to show that $\sum_m \mathrm{E}[\,|D_m|] < \infty$. The summation trick is irrelevant for the ordinary Riemann integral from calculus, because it is not random. But even there it might be convenient to compare the Δt approximation to the $\frac{1}{2}\Delta t$ approximation.

We need a hypothesis on the integrand g_s in our way of proving convergence of the approximations. We will use the hypothesis

$$\mathrm{E}\Big[(g_t - g_s)^2 \Big] \leq C \, |t - s| \; . \tag{22}$$

This is consistent with $|g_t - g_s|$ normally being on the order of $\sqrt{t-s}$, which is the continuity of Brownian motion, on average. The integrands we have in mind that come from stochastic differential equations (4) have this degree of continuity. The convergence proof below uses "hard analysis" (specific inequalities) rather than "soft analysis" (arguments using qualitative properties such as continuity). A proof using hard analysis needs a "hard" hypothesis, such as the specific inequality (22). A soft analysis proof might just ask g_t to be continuous, as for the Riemann integral. A soft analysis proof might be possible for the Ito integral, but it would take longer and be more abstract.

The continuity hypothesis (22) is simpler than it might be. It does not imply that g_t is a continuous function of t. For example, the Poisson arrival process counting function N_t satisfies (22).

If $\Delta t^{(m)} = 2^{-m}$, then $\Delta t^{(m+1)} = \frac{1}{2}\Delta t^{(m)}$. We will write the $\Delta t^{(m)}$ discrete times as

$$t_k^{(m)} = k\,\Delta t^{(m)} \; .$$

The relation $\Delta t^{(m+1)} = \frac{1}{2}\Delta t^{(m)}$ implies that the $\Delta t^{(m+1)}$ even numbered discrete times are

$$t_{2k}^{(m+1)} = (2k)\frac{1}{2}\Delta t^{(m)} = t_k^{(m)} \; .$$

The odd numbered ones are

$$t_{2k+1}^{(m+1)} = (2k+1)\frac{1}{2}\Delta t^{(m)} = t_k^{(m)} + \frac{1}{2}\Delta t^{(m)} \; .$$

The last expression is what you would get if you put $k+\frac{1}{2}$ in the t_k formula, so we use the notation

$$t_{2k+1}^{(m+1)} = t_{k+\frac{1}{2}}^{(m)} \; .$$

For every $\Delta t^{(m)}$ interval $[t_k^{(m)}, t_{k+1}^{(m)}]$, there are two intervals at the $m+1$ level. These are

$$[t_{2k}^{(m+1)}, t_{2k+1}^{(m+1)}] = [t_k^{(m)}, t_{2k+\frac{1}{2}}^{(m)}] \; , \quad \text{and} \quad [t_{2k+1}^{(m+1)}, t_{2k+2}^{(m+1)}] = [t_{k+\frac{1}{2}}^{(m)}, t_{2k+1}^{(m)}] \; .$$

This simple relationship is possible because $\Delta t^{(m+1)} = \frac{1}{2}\Delta t^{(m)}$.

This relationship between the intervals allows us to compare the approximation (16) $X_t^{(m)}$ to $X_t^{(m+1)}$ one term by term. Corresponding to the interval $[t_k^{(m)}, t_{k+1}^{(m)}]$ there is the term

$$g_{t_k^{(m)}}\left(W_{t_{k+1}^{(m)}} - W_{t_k^{(m)}}\right) \; .$$

On the $m+1$ level, this same interval has two terms, which are

$$g_{t_{2k}^{(m+1)}}\left(W_{t_{2k+1}^{(m+1)}} - W_{t_{2k}^{(m+1)}}\right) \; , \quad \text{and} \quad g_{t_{2k+1}^{(m+1)}}\left(W_{t_{2k+2}^{(m+1)}} - W_{t_{2k+1}^{(m+1)}}\right) \; .$$

I now simplify the notation to make the expressions less complicated. I write g_k for $g_{t_k^{(m)}}$, $g_{k+\frac{1}{2}}$ for $g_{t_{2k+1}^{(m+1)}}$, and so on. In this simpler notation we can subtract the level m term from the two corresponding level $m+1$ terms, which gives

$$R_k = g_k\left(W_{k+\frac{1}{2}} - W_k\right) + g_k\left(W_{k+1} - W_{k+\frac{1}{2}}\right) - g_k\left(W_{k+1} - W_k\right) \; .$$

Some algebra shows that

$$R_k = \left(g_{k+\frac{1}{2}} - g_k\right)\left(W_{k+1} - W_{k+\frac{1}{2}}\right) \; . \tag{23}$$

Therefore

$$X_t^{(m+1)} - X_t^{(m)} = \sum_{t_k^{(m)} < t} R_k \; .$$

We prove the almost sure convergence that defines the Ito integral when we verify the inequality (13). More specifically, we have to show that

$$\mathrm{E}\left[\left(\sum_{t_k < t} R_k \right)^2 \right] \leq B \Delta t \; . \tag{24}$$

This inequality involves *cancellation*. The terms R_k are not individually small enough. Write (23) informally as $\Delta g \Delta W$. We argued, and will say more completely soon, that Δg should be roughly the size of ΔW, which would make it on the order of $\sqrt{\Delta t}$. This would make R_k roughly on the order of Δt, as both Δg and ΔW are on the order of $\sqrt{\Delta t}$. If you add up n numbers on the order of Δt, you might think you get something on the order of $n\Delta t = t_n$. This does not go to zero as $\Delta t \to 0$. *Cancellation* happens in a sum when the positive and negative terms roughly cancel, so the sum of the terms, with signs, is smaller than the sum of the absolute values. You might try to get a bound on D_m directly as

$$D_m = \left| \sum_{t_k < t} R_k \right| \leq \sum_{t_k < t} |R_k| \; .$$

This is correct (the inequality is true), but it is not useful because it is too much of an overestimate. The actual D_m is much smaller than the right side suggests it might be.

Instead, we compute the expected square of the sum. As we have seen before, the square of a sum may be written as a double sum.

$$\mathrm{E}\left[\left(\sum_{t_k < t} R_k \right)^2 \right] = \sum_{t_j < t} \sum_{t_k < t} \mathrm{E}[\, R_k R_j]$$

This sum has off-diagonal terms, the ones with $j \neq k$, and diagonal terms that involve R_k^2. The off-diagonal terms have zero expected value. The diagonal terms have order Δt^2, so they add up to something of order $n \Delta t^2$, which is of order Δt.

The cancellation comes from two things. One is that $\mathrm{E}[R_k] = 0$. The other is that R_j is uncorrelated to R_k if $j \neq k$. The cancellation requires some level of decorrelation. It doesn't have to be perfect as it is here (nobody correlated with anyone else), but if all the R_k were equal (perfect correlation), there would be no cancellation and the right side of (24) would be just B, not $B\Delta t$. With mean zero and a large number of uncorrelated terms, any given sample (the question for pathwise almost sure convergence) is likely to have cancellation.

We evaluate the terms using the tower property. If $\mathcal{F}_s$ corresponds to the information in $W_{[0,s]}$, and Q is any random variable, then

$$\mathrm{E}[\, \mathrm{E}[\, Q \mid \mathcal{F}_s] \,] = \mathrm{E}[Q] \; .$$

We start with the off-diagonal terms. Without loss of generality we may assume $k > j$. We condition on knowing everything up to time $t_{k+\frac{1}{2}}$. If you know all that, then you know $g_{k+\frac{1}{2}}$, and g_k, and the four quantities that go into R_j. However, the independent increments property implies that even knowing $W_{[0,t_{k+\frac{1}{2}}]}$, the increment $W_{k+1} - W_{k+\frac{1}{2}}$ is still in the future and has conditional expected value zero. Here is the calculation:

$$
\begin{aligned}
\mathrm{E}[\, R_k R_j] &= \mathrm{E}\Big[\, \mathrm{E}\Big[\, R_k R_j \mid \mathcal{F}_{k+\frac{1}{2}} \Big]\Big] \\
&= \mathrm{E}\Big[\, \mathrm{E}\Big[\Big(W_{k+1} - W_{k+\frac{1}{2}}\Big) \Big(g_{k+\frac{1}{2}} - g_k\Big) \Big(W_{j+1} - W_{j+\frac{1}{2}}\Big) \Big(g_{j+\frac{1}{2}} - g_j\Big) \mid \mathcal{F}_{k+\frac{1}{2}} \Big]\Big] \\
&= \mathrm{E}\Big[\Big(g_{k+\frac{1}{2}} - g_k\Big) \Big(W_{j+1} - W_{j+\frac{1}{2}}\Big) \Big(g_{j+\frac{1}{2}} - g_j\Big) \mathrm{E}\Big[\Big(W_{k+1} - W_{k+\frac{1}{2}}\Big) \mid \mathcal{F}_{k+\frac{1}{2}} \Big]\Big] \\
&= \mathrm{E}\Big[\Big(g_{k+\frac{1}{2}} - g_k\Big) \Big(W_{j+1} - W_{j+\frac{1}{2}}\Big) \Big(g_{j+\frac{1}{2}} - g_j\Big) \cdot 0 \,\Big] \\
&= 0 \; .
\end{aligned}
$$

The main point is $\mathrm{E}\Big[\Big(W_{k+1} - W_{k+\frac{1}{2}}\Big) \mid \mathcal{F}_{k+\frac{1}{2}} \Big] = 0$, and this follows from the independent increments property, because $W_{k+1} - W_{k+\frac{1}{2}}$ is in the future of $t_{k+\frac{1}{2}}$. If we use a non-causal approximation to the Ito integral, this is not true.

The diagonal terms also can be estimated by conditioning. The independent increments property implies that

$$
\mathrm{E}\Big[\Big(W_{k+1} - W_{k+\frac{1}{2}}\Big)^2 \mid \mathcal{F}_{k+\frac{1}{2}} \Big] = \frac{1}{2}\Delta t \; .
$$

This is because it is the variance of the increment over a time interval of length $\frac{1}{2}\Delta t$. The calculation for a diagonal term is

$$
\begin{aligned}
\mathrm{E}\big[\, R_k^2 \big] &= \mathrm{E}\Big[\, \mathrm{E}\Big[\, R_k^2 \mid \mathcal{F}_{k+\frac{1}{2}} \Big]\Big] \\
&= \mathrm{E}\Big[\, \mathrm{E}\Big[\Big(W_{k+1} - W_{k+\frac{1}{2}}\Big)^2 \Big(g_{k+\frac{1}{2}} - g_k\Big)^2 \mid \mathcal{F}_{k+\frac{1}{2}} \Big]\Big] \\
&= \mathrm{E}\Big[\Big(g_{k+\frac{1}{2}} - g_k\Big)^2 \mathrm{E}\Big[\Big(W_{k+1} - W_{k+\frac{1}{2}}\Big)^2 \mid \mathcal{F}_{k+\frac{1}{2}} \Big]\Big] \\
&= \mathrm{E}\Big[\Big(g_{k+\frac{1}{2}} - g_k\Big)^2 \frac{1}{2}\Delta t \Big] \\
&= \frac{1}{2}\Delta t \, \mathrm{E}\Big[\Big(g_{k+\frac{1}{2}} - g_k\Big)^2 \Big] \; .
\end{aligned}
$$

This is where we use the continuity hypothesis (22) on g_s. The expectation on the last line is of Δg over an interval of length $\frac{1}{2}\Delta t$, so the expected square is bounded by $\frac{C\Delta t}{2}$. The final result is

$$
\mathrm{E}\big[\, R_k^2 \big] \leq \frac{1}{4} C \Delta t^2 \; .
$$

The $\frac{1}{4}$ is not important since the point of this calculation is to show that everything goes to zero as $\Delta t \to 0$.

This is the hard part. Now we verify (24) by combing the calculations we have done. In the end I throw in a little trick.

$$
\begin{aligned}
\mathrm{E}\left[\left(\sum_{t_k<t} R_k\right)^2\right] &= \sum_{t_k<t} \mathrm{E}\left[\, R_k^2 \right] \\
&\leq \sum_{t_k<t} \frac{1}{4} C \Delta t^2 \\
&\leq \frac{1}{4} C \Delta t \sum_{t_k<t} \Delta t \\
&= \frac{1}{4} C t \Delta t \ .
\end{aligned}
$$

This verifies (24) with the constant $B = \frac{1}{4}Ct$. With this we have proven the convergence, almost surely, of the approximations. The limit is the Ito integral. This is the definition of the Ito integral.