Stochastic Calculus, Courant Institute, Fall 2022
http://www.math.nyu.edu/faculty/goodman/teaching/StochCalc2022/index.html

Stochastic Calculus

Stochastic integral, Ito integral

Jonathan Goodman, Fall, 2022

1 Introduction

The “calculus” part of Stochastic Calculus involves a new kind of integral, the *Ito integral*, and a new kind of chain rule, *Ito’s lemma*. These go together because the Ito integral is necessary to define the terms that appear in Ito’s lemma. Ito’s lemma requires second derivatives with respect to x, as did the generator calculation from the previous section. The *quadratic variation* is a measure of the “noisiness” of a diffusion X_t. The new “Ito term” that appears in Ito’s lemma involves second derivatives and quadratic variation.

The *Ito calculus* is for functions of time. It is expressed using the notation of differentials, dt, dX_t, and so on. You can think of these informally as Δt and ΔX when Δt is small. The formal interpretation of a differential expression is that it leads to a true formula when integrated. For example, there is the informal expression of an SDE

$$dX_t = a(X_t)dt + b(X_t)dW_t \; . \tag{1}$$

The formal expression involves integrating both sides between two time limits

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

The integral on the left is supposed to “add up” all the small changes dX_t that happen in the time interval $[t_1, t_2]$. The sum of the small changes should be the total change. This leads to the integral formulation of the SDE, which is

$$X_{t_2} - X_{t_1} = \int_{t_1}^{t_2} a(X_t)\, dt + \int_{t_1}^{t_2} b(X_t)\, dW_t \; . \tag{2}$$

Informally, we think of the integral form (2) coming from the differential form (1) by integration. The formal definition is the reverse. The dW integral on the right is an *Ito integral*, and the integral relation (2) is the definition of the SDE (1).

The Ito calculus, like other calculus, involves formulas that are true in the limit $\Delta t \to 0$. The finite Δt approximation may have terms involving various powers of Δt. In ordinary calculus, the terms you ignore in the limit have higher powers of Δt, so that even the sum of many such terms is small when Δt is small. The sums are small even if you replace these terms by their absolute values. In stochastic calculus, there are sums that are small for small Δt not because the terms are very small, but because the terms are somewhat small, and have mean zero, so that the positive terms approximately cancel the negative ones. This cancellation is behind the "Ito rule" (so called, and not by Ito) $dW^2 = dt$. It's true that $\mathrm{E}\Big[(\Delta W)^2\Big] = \Delta t$. But this is not true without the expectation. For example, there is about a 52% chance that $(\Delta W)^2 < .5\Delta t$. As explained just above, the differential formula $dW^2 = dt$ should be interpreted as an integral relation (see Section 3). This integral formula holds because the differences $\Delta W^2 - \Delta t$ have mean zero and cancel in the $\Delta t \to 0$ limit. If you sum the absolute values $\left|\Delta W^2 - \Delta t\right|$, the sum does not vanish in the limit.

We said in the Section 1 notes that you do not have to think of a general diffusion process X_t as being "driven" by Brownian motion. We said then that $b\Delta W$ is just a way of saying "noise with mean zero, variance $b^2\Delta t$". The Ito calculus gives a different meaning to bdW_t through the Ito integral on the right of the integral form (2). If you set $t_1 = 0$ (say), then the right side defines X_{t_2} as a function of t_2 and $W_{[0,t_2]}$ for $t > 0$. The integral on the right involves the Brownian motion path W_t only in the range from $t = 0$ to $t = t_2$. The Ito calculus defines integrals of the form

$$Y_t = \int_0^t b_s dW_s \; . \tag{3}$$

The integrand b_s can depend on the Brownian motion path W. If it does depend on W, then it will be random. But the integrand cannot depend on W in an arbitrary way. The integrand must be *non-anticipating*, which means that future values W_s for $s > t$ are independent of the integrand b_s for $s \leq t$. Said a different way, the integrand value b_t can be a function of t and $W_{[0,t]}$. The "argument" of the function a cannot involve W_s for $s > t$. Functions $b_t(W)$ are also called *adapted* or *progressively measurable*. The integrand $b(X_t)$ in (2) is non-anticipating because X_t is a function of W_s only for $s \leq t$.

A *process* is a function of time. The Ito integral (3) defines the process Y_t. An *Ito process* is any process that can be defined as the sum of an Ito integral and a regular integral

$$Y_t = \int_0^t b_s \, dW_s + \int_0^t a_s \, ds \; . \tag{4}$$

The regular calculus integral (the "Riemann integral") is defined for any continuous integrand a_s even if it is anticipating (not non-anticipating). The limit that defines the regular integral does not rely on cancellation as the Ito integral definition does. The integral version of the SDE formula (2) shows that a diffusion process is an Ito process. It does not go the other way. An Ito process does not have to be a diffusion process because an Ito process does not have to have the Markov property. The differential form of the Ito process formula (4) is

$$dY_t = b_t dW_t + a_t dt \; . \tag{5}$$

As we already said, the Ito integral gives meaning to differential expressions like this. The differential expression (5) is an SDE and defines a diffusion process if a_t and b_t are functions of Y_t.

Most of the derivation in this Section is for one component processes. The extension to more than one component is not hard, if you have understood the one component version.

2 Progressively measurable, filtration

The mathematical theory of diffusion processes uses the concepts of *probability measure*, *sigma algebra*, *filtration*, *conditional expectation*, and *progressively measurable*. This course treats these concepts intuitively and incompletely. The goal is not that you will be able to follow a mathematical proof, but that you will understand what is meant by statements like $Z_t = \mathrm{E}[\,Z|\mathcal{F}_t]$. Here, "understand" means something like "know what it means to simulate Z_t". More precisely: be able to say what is being simulated and how it might be done.

Warnings: (1) This Section is not "linear", which means that things may be defined after they are first used. You may have to read a few times to get it straight. I hope non-linear writing makes the material less mysterious. (2) People use different notation for the same things and the same notation for different things. For example, some people say $Z \in \mathcal{F}$ makes no sense if Z is a random variable and $\mathcal{F}$ is a sigma algebra. Many people write this to express the idea that Z is measurable with respect to $\mathcal{F}$. (3) Concepts may be defined "informally" in terms of imprecise ideas like "information". This may be followed by criteria for computation that still are not mathematical definitions. (4) Some concepts use the same words used in elementary probability but mean something slightly different. *Conditional expectation* is an example. (5) Some definitions here are not the usual ones at the beginnings of probability books. For example, a $\sigma-$algebra is usually defined as a collection of sets, not a collection of functions. The "collection of functions" definition is equivalent, and more often (in my unscientific estimation) useful than the collection of sets version.

A *sigma algebra* (or $\sigma-$algebra) represents a certain state of information. If Z is a random variable, we write $Z \in \mathcal{F}$ to mean that the information in the $\sigma-$algebra $\mathcal{F}$ precisely determines the value of Z. We write $\mathcal{F} \subset \mathcal{G}$ if $\mathcal{G}$ contains all the information of $\mathcal{F}$, which is the same as saying $Z \in \mathcal{F}$ implies that $Z \in \mathcal{G}$. If t represents time, then $\mathcal{F}_t$ represents the information available at time t". Thus, $Z \in \mathcal{F}_t$ means that the information in $\mathcal{F}_t$ determines the value of Z precisely. The family of $\sigma-$algebras $\mathcal{F}_t$ forms a *filtration* if there is no forgetting, $\mathcal{F}_t \subset \mathcal{F}_{t+s}$ if $s \geq 0$.

The σ of $\sigma-$algebra is for "summation", more precisely, infinite sums (like Taylor series). It has to do not with sums in particular, but with *completeness*. The set of real numbers is complete in the sense that if q_n is a sequence of rational numbers that "should have a limit", then there is a real number r with $q_n \to r$ as $n \to \infty$. For example, comes from the Taylor series formula for e^x

when $x = 1$. Define the rational numbers

$$q_n = \sum_{k=0}^{n} \frac{1}{k!} \ .$$

Then

$$\lim_{n\to\infty} q_n = \sum_{k=0}^{\infty} \frac{1}{k!} = e \ .$$

Everything in calculus is defined using limits.

Stochastic calculus needs limits of functions and/or random variables or random paths. For $\mathcal{F}$ to be a $\sigma-$ algebra, it first has to be an algebra in the sense that if $Z_1 \in \mathcal{F}$ and $Z_2 \in \mathcal{F}$ then $Z_1 + Z_2 \in \mathcal{F}$ and $Z_1 Z_2 \in \mathcal{F}$. The σ in $\sigma-$algebra means that if Z_n is a family of functions that "should" have a limit (technical definitions omitted) than there is a $Z \in \mathcal{F}$ so that $Z_n \to Z$ as $n \to \infty$. As an example, for any $\Delta t = T/n$, we can define the Euler Maruyama approximate solution to the SDE as

$$X_{t_{k+1}}^{(\Delta t)} = X_{t_k}^{(\Delta t)} + a_k \Delta t + b_k \Delta W_k \ .$$

We use the abbreviation $a(X_{t_k}^{(\Delta t)}) = a_k$, and similarly for b_k. The actual diffusion path X_t is defined as the limit $\Delta t \to 0$ or $n \to \infty$ of these approximate paths. The approximate paths $X^{(\Delta t)}$ are measurable because they involve only addition and multiplication of quantities derived from ΔW. The limit is measurable because it should exist and $\mathcal{F}$ is a $\sigma-$algebra.

The increasing family $\mathcal{F}_t$ form a *filtration*. This represents states of partial information. At time t you know $\mathcal{F}_t$ completely but there are things in $\mathcal{F}_{t+s}$ that you don't know, if $s > t$. Two component random variables (X, Y) with probability densities $p(x, y)$ illustrate partial information. There is a $\sigma-$algebra $\mathcal{F}_{XY}$ in which you know both X and Y. This is the one we normally use without thinking about it in basic probability. There also is a $\sigma-$algebra $\mathcal{F}_X$ that "knows" the value of X but not the value of Y. Suppose $Z = V(X, Y)$ is some function. Then you know Z if you know X and Y, so $Z \in \mathcal{F}_{XY}$. But if you only know X, and if Z depends on Y in some way, then you do not know Z completely. That means $Z \notin \mathcal{F}_X$. We say that $\mathcal{F}_X$ is the $\sigma-$algebra *generated* by X and $\mathcal{F}_{XY}$ is generated by X and Y. A random variable is in a $\sigma-$algebra if it is a function of the variables used to define the algebra. Thus, $Z = V(X)$ is in $\mathcal{F}_X$ and in $\mathcal{F}_{XY}$. Also $Z = X^2 + Y^2$ (for example) has $Z \in \mathcal{F}_{XY}$ but $Z \notin \mathcal{F}_X$.

Conditional expectation of a random variable with respect to $\mathcal{F}$ is the expected value of Z, conditioned on the information in $\mathcal{F}$. As an example, let (X, Y) be jointly normal with mean zero, each variance 1 and covariance $\mathrm{cov}(X, Y) = \mathrm{E}[XY] = \frac{1}{2}$. Then $Y \notin \mathcal{F}_X$, but if you know X, then the conditional expectation of Y is $\frac{1}{2}X$. This is written

$$\mathrm{E}[\, Y \mid \mathcal{F}_X \,] = \frac{1}{2} X \ . \tag{6}$$

In undergraduate probability, this may be written

$$\mathrm{E}[\, Y \mid X = x \,] = \frac{1}{2} x \; . \tag{7}$$

In undergrad probability, conditional expectation is the ordinary expectation, conditional on some event, such as $X = x$. In stochastic calculus, conditional expectation is is a random variable because the information you are conditioning on is random.

Here's a definition. If Z is a random variable, the conditional expectation

$$U = \mathrm{E}[\, Z \mid \mathcal{F} \,]$$

is another random variable with $U \in \mathcal{F}$ and so that U minimizes

$$\mathrm{E}\Big[\, (V - Z)^2 \Big] \; . \tag{8}$$

over all random variables $V \in \mathcal{F}$. The conditional expectation is the best approximation, in the least squares sense, to a function that is determined by the information in $\mathcal{F}$. This is consistent with ordinary expected value. If X is a random variable, the expected value minimizes the least squares approximation error of X by a number

$$\mu = \arg \, \min_a \mathrm{E}\big[\, (X - a)^2 \big] \; .$$

The undergrad conditional expectation (7) also is found in this way. The stochastic calculus version of this (6) follows from the general minimization definition because any $V \in \mathcal{F}_X$ may be written as a function $g(X)$, and

$$\mathrm{E}\Big[\, (g(X) - Y)^2 \Big] \geq \mathrm{E}\left[\left(\frac{1}{2} X - Y \right)^2 \right] \; .$$

In stochastic calculus, we most often find conditional expectation indirectly or informally without using the definition involving (8) directly.

There is a *standard filtration* or maybe *natural* filtration associated to an SDE (1). In this filtration, $\mathcal{F}_t$ knows the Brownian motion path from the beginning up to time t, which we write $W_{[0,t]}$. A random function Y_t is *adapted* to this filtration if $Y_t \in \mathcal{F}_t$ for all t, which means that Y_t is determined by $W_{[0,t]}$. As explained before, the Euler Maruyama approximation determines X_t using values in $W_{[0,t]}$ and the $\Delta t \to 0$ limit. This shows that X_t is adapted. Therefore $a(X_t)$ and $b(X_t)$ also are adapted.

For proving things in stochastic calculus, we often use the independent increments property of Brownian motion in the form that says any increment in the future of t is independent of $\mathcal{F}_t$, which means independent of any random variable $Y_t \in \mathcal{F}_t$. This applies in particular to approximations to the Ito integral and to solutions of stochastic differential equations. We take always take $\Delta W_k = W_{t_{k+1}} - W_{t_k}$ so that ΔW_k is independent of $\mathcal{F}_{t_k}$. In particular, this means that ΔW_k is independent of $X_{[0,t_k]}$.

Conditional expectation as we have just defined it gives a slightly different way to think of value functions.

$$f(X_t, t) = \mathrm{E}[\, V(X_T) \mid \mathcal{F}_t \,] \ . \tag{9}$$

The quantity on the left side is random, because X_t is random, but $X_t \in \mathcal{F}_t$, so it is known at time t.

The conditional expectation defined here has the *tower property*. If $\mathcal{G} \subset \mathcal{F}$, then we can take the conditional expectation with respect to the information in $\mathcal{F}$ and then take the conditional expectation of that with respect to the smaller amount of information in $\mathcal{G}$. The result is the same as taking conditional expectation with respect to $\mathcal{G}$ in one step;

$$\mathrm{E}[\, \mathrm{E}[\, Z \mid \mathcal{F} \,] \mid \mathcal{G} \,] = \mathrm{E}[\, Z \mid \mathcal{G} \,] \ .$$

One case is when $\mathcal{G}$ is the "trivial" $\sigma-$algebra that has no information. The conditional expectation with respect to no information is just the expectation. Therefore

$$\mathrm{E}[\, \mathrm{E}[\, Z \mid \mathcal{F} \,] \,] = \mathrm{E}[Z] \ .$$

Another case is $\mathcal{F}_t \subset \mathcal{F}_{t+s}$, which gives

$$f(X_t, t) = \mathrm{E}[\, f(X_{t+s}, t+s) \mid \mathcal{F}_t \,] \ .$$

We used this (in a different form, but the same idea) to derive the backward equation.

3 Quadratic variation

The *quadratic variation* is a first example of a $\Delta t \to 0$ limit formula that is true because of cancellation in the error terms in the finite Δt approximation. With $\Delta t > 0$ there are discrete time values $t_k = k\Delta t$. If Y_t is an Ito process (one component for now), the quadratic variation is the limit (I dislike the $[Y]_t$ notation, but people use it)

$$[Y]_t = \lim_{\Delta t \to 0} \sum_{t_k < t} \left(Y_{t_{k+1}} - Y_{t_k} \right)^2 \ . \tag{10}$$

The number of terms on the right is $n_t \approx t/\Delta t$. This has $n_t \to \infty$ as $\Delta t \to 0$ with t fixed. The quadratic variation formula is

$$[Y]_t = \int_0^t b_s^2 \, ds \ . \tag{11}$$

This formula seems natural, as we will see, but the derivation relies on a technical calculation and the hypothesis that a is non-anticipating.

The mathematically correct derivations of Ito calculus formulas like (11) are too technical and long for this quick Stochastic Calculus course. Instead, here is something with the essence but not the details. We replace the integrals (4) with their Δt approximations. We denote increments of Brownian motion by $\Delta W_j = W_{t_{j+1}} - W_{t_j}$.

$$Y_k = \sum_{t_j < t_k} b_{t_j} \Delta W_j + \sum_{t_j < t_k} a_{t_j} \Delta t \; . \tag{12}$$

A crucial point here, and in all of Ito calculus, is that ΔW_j is "in the future of" t_j, and therefore is independent of b_{t_j}. The limits in Ito calculus would either be different or would be wrong if you violate this. For example, the limit defining the Ito integral is different if you use "backward" differences $\Delta W_j = W_{t_j} - W_{t_{j-1}}$. Differences like this do not change the limits defining "regular" calculus, but they do in Ito calculus.

If you make the approximation (12) in the quadratic variation definition (10), the result is

$$[Y]_t^{(\Delta t)} \approx \sum_{t_k < t} \left(b_{t_k} \Delta W_k + a_{t_k} \Delta t \right)^2 \; .$$

The expected value of a typical term in this sum is

$$\mathrm{E}\left[\, b_{t_k}^2 \Delta W_k^2 \right] \; .$$

The following calculation uses the tower property and the independent increments property. It also uses the "obvious" fact that if $U \in \mathcal{F}$, then

$$\mathrm{E}[\, UV \mid \mathcal{F}\,] = U \mathrm{E}[\, V \mid \mathcal{F}\,] \; .$$

This is applied with $U = a_{t_k}$.

$$\begin{aligned} \mathrm{E}\left[\, a_{t_k}^2 \Delta W_k^2 \right] &= \mathrm{E}\left[\, \mathrm{E}\left[\, a_{t_k}^2 \Delta W_k^2 \mid \mathcal{F}_{t_k} \,\right] \right] \\ &= \mathrm{E}\left[\, a_{t_k}^2 \mathrm{E}\left[\, \Delta W_k^2 \mid \mathcal{F}_{t_k} \,\right] \right] \\ &= \mathrm{E}\left[\, a_{t_k}^2 \,\right] \Delta t \; . \end{aligned}$$

4 Ito integral

The Ito integral with respect to a process X_t with integrand b_t is a limit similar to the Riemann integral limit (usual notation and assumptions, $t > 0$, $\Delta t > 0$, $t_k = k\Delta t$)

$$Y_t = \int_0^t b_s \, dX_s = \lim_{\Delta t \to 0} \sum_{t_k < t} b_{t_k} \left(X_{t_{k+1}} - X_{t_k} \right) \; . \tag{13}$$

We can allow different kinds of processes X_t and integrands b_t. The technical details depends on those choices. But we always require that $b_t \in \mathcal{F}_t$ and that we use the forward in time difference $X_{t_{k+1}} - X_{t_k}$. The Riemann integral is less rigid in that allows, for example, backward difference $X_{t_k} - X_{t_{k-1}}$. The following example explains why the Ito integral definition is more rigid

Example, $\int_0^t W_s dW_s$

The ordinary calculus lessons on integration may start with an example where the limit (13) may be computed explicitly. This illustrates "what's going on" but is not the way most integrals are calculated. In ordinary calculus, most integrals are done by "anti-differentiation". The fundamental theorem of calculus and the rules of differentiation are used to guess the anti-derivative. The analogous calculations in Stochastic Calculus are done using *Ito's lemma*.

The example has $b_t = W_t$ and $X_t = W_t$. The integrand is Brownian motion and the process is Brownian motion.

$$Y_t = \int_0^t W_s \, dW_s \ .$$

The sum that approximates the Ito integral is

$$Y_t^{(\Delta t)} = \sum_{t_k < t} W_k (W_{k+1} - W_k) \ . \tag{14}$$

We write W_k instead of W_{t_k} for simplicity. The trick for this specific example is

$$W_k = \frac{1}{2} (W_{k+1} + W_k) - \frac{1}{2} (W_{k+1} - W_k) \ . \tag{15}$$

This makes $Y^{(\Delta t)} = C^{(\Delta t)} - D^{(\Delta t)}$ where we use $W_{k+1} + W_k$ in C and $W_{k+1} - W_k$ in D.

The calculations are simple and different. First

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

What makes it possible to compute this is the formula $(a+b)(a-b) = a^2 - b^2$. Therefore (defining t_n to be the largest t_k with $t_k < t$)

$$C_t^{(\Delta t)} = \frac{1}{2} \left[\left(W_1^2 - W_0^2 \right) + \left(W_2^2 - W_1^2 \right) + \cdots + \left(W_{n+1}^2 - W_n^2 \right) \right] \ .$$

Note that W_1^2 appears with opposite sign in the first and second terms on the right, so W_1^2 cancels in the sum. Similarly, every W_k^2 cancels except W_0 and W_{n+1}. A sum like this is called *telescoping*[1]. It collapses to just

$$C_t^{(\Delta t)} = \frac{1}{2} \left(W_{n+1}^2 - W_0^2 \right) \ .$$

In the limit $\Delta t \to 0$, we have $t_{n+1} \to t$. Also, $W_0 = 0$ (a convention for ordinary Brownian motion), so we are left with just

$$C_t = \lim_{\Delta t \to 0} C_t^{(\Delta t)} = \frac{1}{2} W_t^2 \ .$$

[1] Some small telescopes collapse down to be carried and expanded to be used. Here are some examples. The terms in the sum are like sections of the telescope. The sum collapses, leaving just the beginning of the first term and the end of the last terms.

Notice that the formula for the limit is simple. This is a feature ordinary calculus and Stochastic calculus have in common. The integral has a deep mathematical definition but the formula for it is simple. By contrast, the discrete sums are easy to define but the formulas are complicated.

The D sum is

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

This is the discrete approximation to the total variation, so

$$D_t = \lim_{\Delta t \to 0} D_t^{(\Delta t)} = \frac{1}{2} t \ .$$

Altogether, we get

$$Y_t = \int_0^t W_s dW_s = C_t - D_t = \frac{1}{2} W_t^2 - \frac{1}{2} t \ . \tag{16}$$

We will come back to this formula once to get it from Ito's lemma and again to observe that it is a martingale.

Returning to an earlier point, you get conditional independence (or approximate conditional independence in fancier situations) if you put dX_t in the future of b_t in the Ito integral. That is the reason for using $b_{t_k}(X_{t_{k+1}} - X_{t_k})$ in the limit that defines the Ito integral. Using $b_{t_{k+1}}$ instead of b_{t_k} would mean that the b value used "knows" the increment $(X_{t_{k+1}} - X_{t_k})$, which can change the value of the integral. The present example is an opportunity to illustrate that with a concrete calculation. Suppose we replace $W_k(W_{k+1} - W_k)$ with $W_{k+1}(W_{k+1} - W_k)$ in approximate Ito formula (14). If this were a Riemann integral then the limit would be the same. But here the limit is different. The trick (15) still applies, but in the form

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

The bad sum $B_t^{(\Delta t)}$ that is not an Ito integral approximation can be calculated using the ideas we used for the good one. The result is

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

The limit is

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

This differs from the Ito formula (16) in that the "Ito correction" is positive rather than negative. The bad version B_t is not a martingale.

By now it may be clear that the "Ito answer" (calculation of an integral) is different from what you get from standard calculus. To see that explicitly, here is a calculation as it could be done in basic calculus. If W_t were a differentiable function of t (it is not), then we could write

$$dW_s = \frac{dW_s}{ds} ds \; .$$

The integral would be calculated using the regular calculus chain rule and the fundamental theorem of (regular) calculus:

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

This differs from the Ito answer (16) in that is lacks the correction $-\frac{1}{2}t$.

5 Ito's lemma

Ito's lemma is the chain rule for differentiating (with respect to time) a function of an Ito process. For a function of Brownian motion, the formula is

$$df(W_t, t) = \partial_w f(W_t, t) dW_t + \partial_t f(W_t, t) dt + \frac{1}{2} \partial_w^2 f(W_t, t) dt \; . \tag{17}$$

The "d" on the left side is taken to mean the change in $f(W_t, t)$ in a small increment of time dt, but that interpretation is suspect/false, as we will soon see. The first two terms on the left correspond to the ordinary calculus chain rule. If W_t were a differentiable function of t, in textbook calculus derivative form or using less formal differential form that you get by multiplying both sides by dt:

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

The Ito's lemma formula (17) has the extra "Ito term" $\frac{1}{2}\partial_w^2 f(W_t, t) dt$. The Ito term is possible because W_t is not a differentiable function of t.

The differential expression (17) can be understood as an informal way of writing a formula involving integrals. The integral form makes pure mathematicians happy because the differentials in (17) do not have mathematical definitions. The integral form also explains why the "Ito rule" is appropriate.

It's because the integral involves many (infinitely many) independent $(dW)^2$ contributions so the Ito rule is an expression of the law of large numbers saying that a sum of many independent terms may be replaced by the mean (sort of, see below). The proof of Ito's lemma is a verification of the integral form.

The main postulate of integration is that when you add up (integrate) all the small changes in some quantity, the result is the total change. In this case, that principle is

$$\int_0^t df(W_t,t) = f(W_t,t) - f(W_0,0) \ . \tag{18}$$

The integral of the differential change is the total change. If we apply this to the differential expression (17), the result is

$$f(W_t,t)-f(W_0,0) = \int_0^t \left(\partial_t f(W_s,s) + \frac{1}{2}\partial_w^2 f(W_s,s) \right) ds + \int_0^t \partial_w f(W_s,s) dW_s \ . \tag{19}$$

To remember about it:

- You need Taylor expansion that include all terms