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

Class 4, Ito integral for Brownian motion

1 Introduction

[In this Class 4 (this lecture), W_t will be standard Brownian motion (no drift), X_t will be a process defined from W_t using an indefinite Ito integral, Y_t will be a process or function defined as an "ordinary" indefinite integral, and $Z_t = X_t + Y_t$ will be a processes defined using both kinds of integral.]

Imagine betting on a Brownian motion path. Let time be broken into small time steps of size Δt, with $t_k = k\Delta t$. At time t_k, you can "buy" f_{t_k} "shares" of the Brownian motion. Then you watch until time t_{k+1}. You get $f_{t_k}\Delta W_{t_k} = f_{t_k}(W_{t_{k+1}} - W_{t_k})$. You started betting at time $t = 0$ and X_{t_k} is the amount you have at time t_k. That is

$$X_{t_k} = \sum_{j=0}^{k-1} f_{t_j}\left(W_{t_{j+1}} - W_{t_j}\right) \; . \qquad (1)$$ dIi

The gain (or loss) in the next time period is

$$X_{t_{k+1}} - X_{t_k} = f_{t_k}\left(W_{t_{k+1}} - W_{t_k}\right) \; . \qquad (2)$$ dId

The betting amounts f_{t_k} can be random (independent of the Brownian motion path), or they can be random in that they depend on the Brownian motion path. But f_{t_k} cannot depend on the *future* of the Brownian motion path. The Brownian motion path up to time t is $W_{[0,t]}$. By "not knowing the future" we mean that there is a function $F(w_{[0,t]}, t)$, which is the *strategy* for betting at time t, and the bet is given by the strategy: $f_{t_k} = F(W_{[0,t_k]})$.

The *Ito integral with respect to Brownian motion* is the limit of a sum like (dIi 1) as $\Delta t \to 0$. This is written

$$X_t = \int_0^t f_s \, dW_s \; . \qquad (3)$$ Ii

The informal *Ito differential* is the limit as $\Delta t \to 0$ of the difference expression, which is the gain/loss over one period

$$dX_t = f_t \, dW_t \; . \qquad (4)$$ Id

The Ito differential is a convenient but informal way of expressing the integral relation

$$X_{t_2} - X_{t_1} = \int_{t_1}^{t_2} f_s \, dW_s \; .$$

The expression (3) the special case of this with $X_0 = 0$. The Ito integral expression (3) is an indefinite integral, but we could fix the endpoints and think of a definite Ito integral.

Suppose g_t is another function and we consider the ordinary (Riemann) indefinite integral

$$Y_t = \int_0^t g_s \, ds \; . \tag{5}$$

iRi

This can be put in informal differential form as

$$dY_t = g_t \, dt \; . \tag{6}$$

Rd

For an applied mathematician creating a mathematical model of something, the differential expression (6) is not just an informal expression of the integral relation (5). The differential expression means that for small Δt,

$$\Delta Y = Y_{t+\Delta t} - Y_t \approx g_t \Delta t \; . \tag{7}$$

delY

The $\approx$ does not mean that the difference between ΔY and $g_t \Delta t$ is small. Both ΔY and $g_t \Delta t$ are small already. It means that the difference between ΔY and $g_t \Delta t$ is *tiny*, which means that even when you add it up, the result is small. When you add up the small contributions

$$\sum_{t_j < t} g_{t_j} \Delta t$$

you get approximately $Y_t - Y_0$, which is not "small" (does not go to zero when $\Delta t \to 0$). The the difference $\Delta Y - g_t \Delta t$ is tiny in that even when you add it up, the result still is small (goes to zero as $\Delta t \to 0$. The mathematical theorem is

$$\lim_{\Delta t \to 0} \left[Y_t - Y_0 - \left(\sum_{t_j < t} g_{t_j} \Delta t \right) \right] = 0 \; .$$

The differential form of *tiny* is

$$\frac{Y_{t+\Delta t} - Y_t - g_t \Delta t}{\Delta t} \to 0 \quad \text{as} \quad \Delta t \to 0 \; .$$

This is also written in "little oh" notation as $Y_{t+\Delta t} - Y_t = g_t \Delta t + o(\Delta t)$. In other words, $O(\Delta t)$ (big Oh) is small and $o(\Delta t)$ (little oh) is tiny, at least when talking about the indefinite Riemann integral (5).

The difference between small and tiny is: "What does it add up to?" This is big Oh versus little oh for the Riemann differential (6). It is more subtle for the Ito differential (4). Even a term that is $O(\Delta t)$ can be tiny if its expected value is zero. Use the notation $\Delta W = W_{t+\Delta t} - W_t$. The small/tiny rules are

$$\begin{aligned} \Delta W = & \text{ small} \\ \Delta t = & \text{ small} \\ \Delta t^2 = & \text{ tiny} \\ (\Delta W)^2 - \Delta t = & \text{ tiny} \; . \end{aligned}$$

Much of ordinary calculus is ignoring the tiny $O(\Delta t^2)$. But $R = (\Delta W)^2 - \Delta t$ is not tiny in the sense of being $O(\Delta t^2)$. In fact (Exercise in Assignment 4) the absolute value is (in expectation) order Δt,

$$\mathrm{E}\Big[\Big|(\Delta W)^2 - \Delta t\Big|\Big] = C\Delta t \;.$$

But $\mathrm{E}[R] = 0$. In a sum like (1) terms of order Δt with expected value zero can add up to something small because of cancellation: the positive terms and negative terms approximately cancel. The cancellation is accurate enough that the sum goes to zero in the limit $\Delta t \to 0$.

The Ito differential (4) and the "Riemann differential" [not a standard term, maybe Newton Leibnitz differential would be better?] (6) may be used together to describe a diffusion process. Suppose Z_t is a random process that satisfies

$$dZ_t = g_t dt + f_t dW_t \;. \tag{8}$$

dZ

Then (we will see this class) the infinitesimal mean of Z is g_t and the infinitesimal variance is f_t^2. As before, the formal expression (8) is equivalent to the integral expression

$$Z_t = Z_0 + \int_0^t g_s\, ds + \int_0^t f_s\, dW_s \;. \tag{9}$$

dXi

The combination of Riemann and Ito differentials in (8) allows us to model any diffusion process in terms of its infinitesimal mean and (square root of its) infinitesimal variance.

Why the infinitesimal mean is g: As we go further into the course we will define infinitesimal mean and variance more precisely, but maybe not entirely precisely. The infinitesimal mean at time t refers to the expected value for a short period in the future, conditioned on what is known up to time t. "What is known up to time t" is the path up to that time, $W_{[0,t]}$. The conditional expectation is $\mathrm{E}[dZ|W_{[0,t]}]$. To be slightly more precise choose a small Δt that will go to zero and (the "small oh" is for a "tiny" error term)

$$\mathrm{E}\big[\,\Delta Z \mid W_{[0,t]}\big] = g_t\,\Delta t \,+\, o(\Delta t) \;. \tag{10}$$

EdZ

This is because the Ito part (X_t) makes no contribution. If X_t is the Ito integral part (3), then (see (2)), we have

$$\mathrm{E}\big[\,\Delta X \mid W_{[0,t]}\big] \approx \mathrm{E}\big[\,(W_{t+\Delta t} - W_t)f_t \mid W_{[0,t]}\big]\big) \;.$$

[The next two points represent one of the most important ideas of Stochastic Calculus!] (*i*) Since f_t is a function of $W_{[0,t]}$, when the path up to time t is known, then f_t also is known. For that reason

$$\mathrm{E}\big[\,(W_{t+\Delta t} - W_t)f_t \mid W_{[0,t]}\big]\big) = f_t\,\mathrm{E}\big[\,(W_{t+\Delta t} - W_t) \mid W_{[0,t]}\big]\big) \;.$$

(*ii*) The independent increments property implies that the increment $(W_{t+\Delta t} - W_t)$ is independent of everything up to time t. In particular, its conditional expected value is still zero. If you condition a random variable on an independent random variable, you don't change its expected value. Therefore

$$\mathrm{E}\big[\,(W_{t+\Delta t} - W_t) \mid W_{[0,t]}\big]) = 0 \; .$$

Altogether (the "little oh" accounting for tiny errors in this argument, "tiny" in the technical sense)

$$\mathrm{E}\big[\,\Delta X \mid W_{[0,t]}\big] = o(\Delta t) \; .$$

The Riemann integral part has $dY_t = g_t\, dt$, which may be expanded to

$$\mathrm{E}\big[\,\Delta Y \mid W_{[0,t]}\big] = g_t\, \Delta t + o(\Delta t) \; .$$

This explains the overall infinitesimal mean formula (EdZ 10).

Why the infinitesimal variance is f_t^2. The reasoning here will be even less formal than for the infinitesimal mean. We will build tools over the next few classes to understand this better. The infinitesimal variance is (we have already seen) equivalent to the expected square

$$\mathrm{var}\big(\,\Delta Z \mid W_{[0,t]}\big) = \mathrm{E}\Big[\,(\Delta Z)^2 \mid W_{[0,t]}\Big] + o(\Delta t) \; .$$

Also, $\Delta Y = O(\Delta t)$, being a regular integral. On the other hand $\Delta X \approx \Delta W f_t$. If f_t is not zero, then ΔX is on the order of $\sqrt{\Delta t}$ because ΔW is on the order of $\sqrt{\Delta t}$. As we said when talking about the infinitesimal mean,

$$\begin{aligned}
\mathrm{E}\Big[\,(\Delta X)^2 \mid W_{[0,t]}\Big] &\approx \mathrm{E}\Big[\,(\Delta W)^2\, f_t^2 \mid W_{[0,t]}\Big] \\
&\approx f_t^2 \mathrm{E}\Big[\,(\Delta W)^2 \mid W_{[0,t]}\Big] \\
&\approx f_t^2 \Delta t \; .
\end{aligned}$$

This leads to

$$\mathrm{E}\Big[\,(\Delta Z)^2 \mid W_{[0,t]}\Big] = f_t^2 \Delta t + o(\Delta t) \; . \tag{11}$$

EdZ2

This shows (suggests, if you're not convinced yet) that *integral expressions like (dXi 9) are able to represent any diffusion process.*

A process that may be represented in the form (dXi 9) is an *Ito processes*. A *diffusion process* is an Ito process that also has the Markov property. Markov means that the distribution of the future depends only on the present, not the past. More specifically, the distribution ΔZ depends on Z_t only, not on Z_s or W_s for $s < t$. This means that the infinitesimal mean and variance in (dZ 8) depend on Z_t only. Tradition tells us to call them $a(z)$ and $b(z)$. Thus, an Ito process is a diffusion if it satisfies a differential relation of the form

$$dZ = a(Z)dt + b(Z)dW_t \; . \tag{12}$$

sde

This is a *stochastic differential equation* (usually called *SDE*). [We will write X for Z most of the time after this class.]

$$dZ = a(Z)dt + b(Z)dW_t \; . \tag{13}$$

sde

A process Z_t is a solution if

$$Z_t = \int_0^t a(Z_s)\, ds + \int_0^t b(Z_s)\, dW_s \ .$$

You create an SDE model of a stochastic process by deciding what the infinitesimal mean $a(z)$ and infinitesimal variance $\mu(z) = b^2(z)$ should be. An important point is that only the infinitesimal variance $\mu(z)$ is relevant to modeling. You have to take the square root $b(z) = \sqrt{\mu(z)}$ to write the SDE, but you get the "same process" if you use $-\sqrt{\mu(z)}$ instead. We will come back to this point in future classes.

Ito's lemma is a stochastic calculus version of the chain rule from ordinary calculus. It answers the question: if X_t depends on t in some stochastic way, and if $u(x)$ depends on z in some differentiable way, then how does $u(X_t)$ depend on t? Ito's lemma for Brownian motion is about processes of the form $Z_t = u(W_t, t)$, with a smooth function $u(w, t)$. Ito's lemma for Brownian motion is a formula for dZ_t:

$$du(W_t, t) = \partial_w u(W_t, t) dW_t + \partial_t u(W_t, t) dt + \frac{1}{2} \partial_w^2 u(W_t, t) dt \ . \qquad (14)$$ IlBm

When we prove Ito's lemma, we will prove it in the integral version

$$u(W_t, t) - u(W_0, 0) = \int_0^t \partial_w u(W_s, s)\, dW_s \qquad (15)$$ IlidW

$$+ \int_0^t \left[\partial_t u(W_s, s) + \frac{1}{2} \partial_w^2 u(W_s, s) \right] ds \qquad (16)$$ Ilidt

This is an Ito process representation (9) [dXi] of $Z_t = u(W_t, t)$ with

$$f_t = \partial_w u(W_t, t)$$
$$g_t = \partial_t u(W_t, t) + \frac{1}{2} \partial_w^2 u(W_t, t) \ .$$

We will use these formulas constantly for the rest of the course.

One can derive (maybe "motivate" is more accurate) Ito's lemma by choosing a small Δt, expanding Δu in Taylor series to include all terms that formally are of size Δt or bigger, and then replacing $(\Delta W)^2$ with Δt. This is justified by the claim (look for support for this claim in the next week or two) that $(\Delta W)^2 - \Delta t$ is tiny. It amounts to replacing $(\Delta W)^2$ with its expected value. We write $\Delta u = u(W_t + \Delta W_t, t + \Delta t) - u(W_t, t)$. We leave out arguments W_t and t, so we write just u for $u(W_t, t)$, etc. We take ΔW to be on the order of $\sqrt{\Delta t}$, so $|\Delta W|^3$ is on the order of $\Delta t^{\frac{3}{2}}$. This makes $|\Delta W|^3$ a tiny $o(\Delta t)$ term. The same reasoning suggests that $|\Delta W| \Delta t = o(\Delta t)$ and, for a simpler reason, $\Delta t^2 = o(\Delta t)$. We expand in a two variable Taylor series and write error terms in big Oh notation. We use absolute values for ΔW because it can be negative.

$$\begin{aligned} u(W_t + \Delta W_t, t + \Delta t) - u &= \partial_w u \Delta W + \frac{1}{2}\partial_w^2 u (\Delta W)^2 + \partial_t u \Delta t \\ &\quad + O(|\Delta W|^3) + O(|\Delta W|\Delta t) + O(\Delta t)^2 \\ &= \partial_w u \Delta W + \frac{1}{2}\partial_w^2 u \Delta t + \partial_t u \Delta t \\ &\quad + O(|\Delta W|^3) + O(|\Delta W|\Delta t) + O(\Delta t)^2 \\ &\quad + \frac{1}{2}\partial_w^2 u \left((\Delta W)^2 - \Delta t\right) \\ &= \partial_w u \Delta W + \left(\frac{1}{2}\partial_w^2 u + \partial_t u\right)\Delta t + o(\Delta t)\ . \end{aligned}$$

The "tiny" $o(\Delta t)$ term on the last line is the sum of the four term above.

2 Geometric Brownian motion

The theory of option pricing in quantitative finance often uses a model for S_t, which is the price of a share of stock at time t. The model is *geometric Brownian motion*

$$dS_t = rS_t dt + \sigma S_t dW_t\ . \tag{17}$$

gBm

This model is built on the natural hypothesis that the expected "return" (the $rS_t dt$ term) and the "risk" (the $\sigma S_t dW_t$) should be proportional to S_t. That way it doesn't matter if you replace each share of stock with price S_t with, say, two shares of stock with price $\frac{1}{2}S_t$. In finance, the parameter r is called the *risk free rate*. I will probably call $rS_t dt$ the "expected return", but a finance person knows the story is more complicated. The parameter σ is the *volatility*, or just *vol*. The geometric Brownian motion SDE expresses the intention that $\mathrm{E}[dS_t|\cdot] = rS_t dt$ and $\mathrm{E}[dS_t|\cdot] = \sigma^2 S_t^2 dt$. We write $\mathrm{E}[\cdots|\cdot]$ to mean the expected value conditional on knowing $S_{[0,t]}$, or, equivalently (we will see), to knowing $W_{[0,t]}$.

If W_t were a "nice" function of t, so $\frac{dW}{dt}$ were well defined, then we could rewrite the SDE (17) as a differential equations class would write it

$$\textbf{(wrong)} \qquad \frac{S_t}{dt} = rS_t + \sigma S_t \frac{dW_t}{dt}\ . \qquad \textbf{(wrong)}$$

The solution would be (check this)

$$\textbf{(wrong)} \qquad S_t = S_0 e^{rt + \sigma W_t}\ . \qquad \textbf{(wrong)}$$

This formula does not satisfy the geometric Brownian motion SDE (17). We see this using Ito's lemma on the function $u(w,t) = S_0 e^{rt+\sigma w}$. The derivatives are

$$
\begin{aligned}
u(w,t) &\xrightarrow{\partial_w} \sigma S_0 e^{rt+\sigma w} \\
&\xrightarrow{\partial_w} \sigma^2 S_0 e^{rt+\sigma w} \\
u(w,t) &\xrightarrow{\partial_t} r S_0 e^{rt+\sigma w}
\end{aligned}
$$

We use $S_t = u(W_t, t)$ and plug into Ito's lemma (14). The result is

$$
\textbf{(wrong)} \qquad dS_t = \sigma S_t dW_t + \frac{1}{2}\sigma^2 S_t dt + r S_t dt \; . \qquad \textbf{(wrong)}
$$

The right side here differs from the right side of the geometric Brownian motion SDE (17) by an extra term $\frac{1}{2}\sigma^2 t S_t dt$.

You can fiddle around trying to see how to fix the first try $S_0 e^{rt+\sigma W_t}$. Our Ito calculation showed that dS of this is too big, so this is too big. Eventually we realize that you can cancel the $\frac{1}{2}\sigma^2 S_t dt$ term by subtracting it from the exponential. [We will have a better method than trial-and-error when we do the more general Ito lemma.] So we try

$$
S_t = S_0 e^{rt - \frac{1}{2}\sigma^2 t + \sigma W_t} \; . \tag{18}
$$

Stf

We check this applying Ito's lemma to the function $u(w,t) = e^{(r-\frac{1}{2}\sigma^2)t+\sigma w}$. It works! The solution to the geometric Brownian motion SDE (17) is the corrected exponential formula (18).

Quant finance people use the solution representation formula (18) for *option pricing*. The Derivative Securities class will have details, but we want to calculate expectations of final time payouts

$$
\mathrm{E}[\, v(S_T)] \; .
$$

The random variable W_t is Gaussian with mean zero and variance t. We can evaluate expectations by putting in this PDF and integrating. The backward equation for this expectation is (except for one term) the *Black Scholes equation* of quant finance. There is an example in Assignment 4.

The solution formula (18) for geometric Brownian motion has some consequences that may seem surprising. Consider the special case $r = 0$. The SDE $dS = \sigma S dW$ may be considered a model of a the long time inter-generational behavior of a "fair" economy. Suppose Δt represents one generation. You are born with wealth S and you leave your one child wealth $S(1+\sigma\Delta W)$. [Every generation consists of a parent and a child who gets the parent's wealth.] The system is fair in the sense that the expected change is zero. But the solution formula is $S_t = S_0 e^{-\frac{1}{2}\sigma^2 t + \sigma W_t}$. For large t, W_t is on the order of $\sqrt{t}$, which is a smaller order than t. It seems natural (and we will show) that [*Almost surely* means with probability 1.]

$$
-\frac{1}{2}\sigma^2 t + \sigma W_t \to -\infty \;\text{ as }\; t \to \infty \text{ almost surely } \; .
$$

Therefore $S_t \to 0$ as $t \to \infty$ almost surely. In this supposedly fair economy almost every family has wealth going to zero. Since the total wealth doesn't change (because the expected change is zero), this can only be because the wealth concentrates in the hands of a small number of families. This point is made by economists in more complicated ways.

Suppose you try to evaluate $\mathrm{E}[S_t]$ by simulation and Monte Carlo. The expectation is $\mathrm{E}[S_t] = m_t = S_0 e^{rt}$, which you can see be evaluating the infinitesimal mean and variance in the geometric Brownian motion SDE (17). If you generate sample paths, the probability that a sample path gives a value as large as the mean is

$$\Pr(S_t > m_t) = \Pr(W_t > \frac{1}{2}\sigma^2 t) \; .$$

We will see that this is "exponentially" unlikely. The probability goes to zero exponentially as $t \to \infty$. The expected value is determined by paths that are exponentially rare. Find the mean by simulation under these conditions is called *rare event simulation*. The best way to do this is not by direct simulation.