Stochastic Calculus, Courant Institute, Fall 2018
http://www.math.nyu.edu/faculty/goodman/teaching/StochCalc2018/index.html
Jonathan Goodman, October, 2018

Lesson 4, Ito's lemma

1 Introduction

Ito's lemma is the chain rule for stochastic calculus. If X_t is a diffusion process with infinitesimal mean $a(x,t)$ and infinitesimal variance $v(x,t)$, and if $u(x,t)$ is a function with enough derivatives, then $Y_t = u(X_t,t)$ is another stochastic process. This satisfies

$$du(X_t,t) = \partial_t u(X_t,t)\,dt + \partial_x u(x,t) dX_t + \frac{1}{2}\partial_x^2 u(X_t,t)\,v(X_t)dt\;. \qquad (1)$$

The first two terms on the right are from the ordinary chain rule that would apply if X_t were a differentiable function of t. The last term is new to diffusion processes. It arises from the fact that dX^2 is of the order of dt. The chain rule is a relation that holds to order dt, so you have to keep all terms of that order.

The formal Ito's lemma relation (1) is formal. The terms dX and dt do not have an independent mathematical meaning. The scientist's understanding of (1), which is usually a simple and reliable way to think about differentials, is actually wrong here. Suppose $\Delta t > 0$ is a small time step and $\Delta u = u(X_{t+\Delta t}, t+\Delta t) - u(X_t,t)$. It is not true that

$$\Delta u = \partial_t u(X_t,t)\Delta t + \partial_x u(X_t,t)\Delta X + \frac{1}{2}\partial_x^2 u(X_t,t) v(X_t,t)\Delta t + O(\Delta t^2)\;.$$

This is because the difference

$$r = (\Delta X)^2 - v(X_t)\Delta t$$

is actually on the order of Δt. We can ignore r (as we will see) because it is of order Δt and it has mean zero. You cannot replace $(\Delta X)^2$ with $v(X_t)\Delta t$ "pointwise" at any specific time. But you can use $v(X_t)\Delta t$ as a substitute for $(\Delta X)^2$ in an average sense. This is the main technical issue of this lesson.

The formal expression (1) is meant to be a simply way to express the integral relations. Integrate both sides over the time integral $[T_1, T_2]$. From the left side of (1) you get

$$\int_{T_1}^{T_2} du(X_t,t) = u(X_{T_2}, T_2) - u(X_{T_1}, T_1)\;.$$

We have not given a mathematical definition of $\int du$, so we can take this as the definition. If we then integrate the terms on the right of (1), the result seems to be

$$u(X_{T_2},T_2) - u(X_{T_1},T_1) = \int_{T_1}^{T_2} \left(\partial_t u(X_t,t) + \frac{1}{2}\partial_x^2 u(X_t,t) v(X_t) \right) dt \tag{2}$$

$$+ \int_{T_1}^{T_2} \partial_x u(X_t,t)\, dX_t \; . \tag{3}$$

The integral on the right side on the first line is the ordinary Riemann integral of the continuous integrand $(\cdots)$. The integral on the second line is the Ito integral with respect to the diffusion dX_t defined in Lesson 3. We prove Ito's lemma by proving the integral version (2)(3).

Ito's lemma also serves as the stochastic version of the fundamental theorem of calculus. Without it, we would struggle to evaluate Ito integrals from the definition, as on Assignment 3 with

$$\int_0^T W_t dW_t = \frac{1}{2}W_t^2 - \frac{1}{2}T \; . \tag{4}$$

In an ordinary calculus class, there may be some examples where Riemann integrals are done directly from the definition, such as

$$\int_0^a x\, dx = \frac{1}{2}a^2 \; .$$

This may be done directly from the definition using the identity

$$\sum_{k=1}^{n} k = \frac{1}{2}n^2 - \frac{1}{2}n \; .$$

But the easier way is to note that $\frac{d}{dx}\left(\frac{1}{2}x^2\right) = x$ and then use the fundamental theorem of calculus. The integral (4) can be done in the same way. Apply Ito's lemma (1) to the function $u(w,t) = \frac{1}{2}w^2 - \frac{1}{2}t$. The necessary derivatives are $\partial_w u = w$, $\partial_w^2 u = 1$, and $\partial_t u = \frac{1}{2}$. Therefore

$$d\left(\frac{1}{2}W_t^2 - \frac{1}{2}t\right) = W_t dW_t + \frac{1}{2}dt - \frac{1}{2}dt = W_t dW_t \; .$$

The integral relations (2) and (3), together with this calculation, imply (4). It is rare to find "indefinite integral" in this way using Ito's lemma. It happens only is special examples. Even for ordinary calculus, most integrands do not have an indefinite integral in closed form.

2 Proof of Ito's lemma

The proof of Ito's lemma has two steps. First, we do a Taylor expansion of Δu and identify the terms of order Δt or higher. Then we show that adding up the terms $(\Delta X)^2$ and adding up the terms $v(X_t)\Delta t$ have the same limit as $\Delta t \to 0$. Both of these arguments use ideas from Lesson 3 and Assignment 3. There also is an application of Borel Cantelli to show that the arguments are correct almost surely. For simplicity, we take the lower limit T_1 to be zero. We write the upper limit T_2 as T.

Use the notation of Lesson 3. Take $h_n = 2^{-n}$, $t_k = kh_n$, and write X_k for X_{t_k}, etc. Then[1] It is easy to give a more correct argument, but it takes longer and isn't more interesting.

$$u(X_T, T) - u(X_0, 0) = \sum_{t_k < T} \Delta u_k \ ,$$

where

$$\Delta u_k = u(X_k + \Delta X_k, t_k + \Delta t) - u(X_k, t_k) \ , \ \ \Delta X_k = X_{k+1} - X_k \ .$$

The Taylor expansion is

$$\begin{aligned}\Delta u_k = \partial_x u_k \Delta X_k + \frac{1}{2}\partial_x^2 u_k \left(\Delta X_k\right)^2 + \partial_t u_k \Delta t \\ + O(|\Delta X_k|^3) + O(\Delta t\,|\Delta X_k|) + O(\Delta t^2) \ .\end{aligned}$$

We sum over k. On the left side we get $u(X_T, t) - u(X_0, 0)$. There are six sums on the right to consider.

1. The first term on the right leads to the sum

$$\sum_{t_k < T} u(X_k, t_k)\Delta X_k \ .$$

We showed in Lesson 3 that this converges almost surely as $n \to \infty$ to the Ito integral

$$\int_0^T u(X_t, t) dX_t \ .$$

2. The second term is the most interesting one. Subsection 2.1 is devoted to it.

3. The third term is the Riemann sum approximation to

$$\int_0^T \partial_t u(X_t, t)\, dt \ .$$

4. The fourth term involves

$$\mathrm{E}\Big[\, |\Delta X_k|^3 \Big] \ .$$

[1]This identity is a little wrong because there is a tiny piece of time between the largest $t_k < T$ and T. This was ignored in Lesson 3 and will be ignored here.

This may be bounded using the Cauchy Schwarz inequality, and $|\Delta X_k|^3 = |\Delta X_k|\,(\Delta X_k)^2$. Therefore

$$\begin{aligned}
\mathrm{E}\Big[\,|\Delta X_k|^3\Big] &= \mathrm{E}\Big[\,|\Delta X_k|\,(\Delta X_k)^2\Big] \\
&\leq \Big\{\mathrm{E}\Big[\,(\Delta X_k)^2\Big]\,\mathrm{E}\Big[\,(\Delta X_k)^4\Big]\Big\}^{\frac{1}{2}} \\
&\leq \left\{C\Delta t \cdot \Delta t^2\right\} \\
&\leq C\Delta t^{\frac{3}{2}} \; .
\end{aligned}$$

The philosophy for this term is that higher moments control lower moments. In this case, the fourth moment $\mathrm{E}\Big[\,(\Delta X_k)^4\Big]$ controls the third moment $\mathrm{E}\Big[\,|\Delta X_k|^3\Big]$. If you know the fourth moment is bounded then you know that the third moment is bounded. You calculate the bound using the Cauchy Schwarz inequality.

2.1 The second sum

It is similar (or the same as) the quadratic variation problems on Assignment 3. It starts with a technical trick to avoid a complicated mess. Instead of the "pointwise" infinitesimal variance, we use a related quantity that is exact for time Δt

$$w(x, \Delta t) = \mathrm{E}_{x,t}\Big[\,(X_{t+\Delta t} - x)^2\Big] \; .$$

As a reminder, the subscript $\mathrm{E}_{x,t}$ means to take the expectation with the condition that $X_t = x$. An equivalent definition is

$$w(X_t, \Delta t) = \mathrm{E}\Big[\,(X_{t+\Delta t} - X_t)^2 \mid \mathcal{F}_t\Big] \; . \tag{5}$$

If you know the path up to time t, which is the information in $\mathcal{F}_t$, then you know X_t. It is natural that the conditional expectation is a function of X_t. This is almost the same as $v(x)\Delta t$ but not quite. Our definition of a diffusion process included the hypothesis that

$$w(x, \Delta t) = v(x)\Delta t + O(\Delta t^2) \; . \tag{6}$$

The proof that follows is "simplified" (avoiding a big mess) by using w instead of v so as to not have an $O(\Delta t^2)$ "error term" someplace.

The second sum is

$$S_n = \frac{1}{2}\sum_{t_k < T} \partial_x^2 u(X_k, t_k)\Delta X_k^2 \; .$$

This can be written as the sum of a "mean" and a fluctuating part:

$$S_n = \frac{1}{2}\sum_{t_k<T} \partial_x^2 u(X_k, t_k) w(X_k, h_n) + \frac{1}{2}\sum_{t_k<T} \partial_x^2 u(X_k, t_k)\left(\Delta X_k^2 - w(X_k, h_n)\right) \; .$$

The first sum on the right, the "mean" part, is more or less the Riemann sum approximation to the dt integral.

$$\sum_{t_k<T} \partial_x^2 u(X_k,t_k) w(X_k,h_n) = \sum_{t_k<T} \partial_x^2 u(X_k,t_k) v(X_k)\Delta t + \sum_{t_k<T} O(\Delta t^2) \; .$$

The first sum is the actual Riemann sum, which converges to the integral as $\Delta t \to 0$ (which is the same as $n \to \infty$):

$$\int_0^T \partial_x^2 u(X_t,t) v(X_t)\, dt \; .$$

The second sum is $O(\Delta t)$ for a reason we saw in Lesson 3:

$$\begin{aligned}
\left| \sum_{t_k \leq T} O(\Delta t^2) \right| &\leq C \sum_{t_k\leq T} \Delta t^2 \\
&\leq C\Delta t \sum_{t_k \leq T} \Delta t \\
&\leq CT\Delta t = O(\Delta t) \; .
\end{aligned}$$

The "fluctuation sum" is the part that has mean zero and turns out to go to zero almost surely as $n \to \infty$. It is

$$R_n = \sum_{t_k<T} \partial_x^2 u(X_k,t_k) \left(\Delta X_k^2 - w(X_k,\Delta t)\right) \; .$$

As in Lesson 3, we show $R_n \to 0$ almost surely as $n \to \infty$ by calculating $\mathrm{E}\big[R_n^2\big]$. Define

$$V_k = \partial_x^2 u(X_k,t_k) \left(\Delta X_k^2 - w(X_k,\Delta t)\right) \; .$$

Then

$$\mathrm{E}\big[R_n^2\big] = \sum_{t_k<T} \sum_{t_j<T} \mathrm{E}[V_k V_j] \; .$$

There are diagonal terms ($j = k$) and off-diagonal terms ($j < k$ or $j > k$). All of the off-diagonal expectations are zero. To see this, suppose $k > j$ and condition on $\mathcal{F}_{t_k}$. The values of V_j and X_k are known at time t_k so V_j and X_k come out of the conditional expectation. The conditional expectation of V_k is zero because of the definition (5) of w. Therefore

$$\begin{aligned}
\mathrm{E}[\,V_j V_k] &= \mathrm{E}[\,\mathrm{E}[\,V_j V_k \mid \mathcal{F}_{t_k}]] \\
&= \mathrm{E}[\,V_j \,\mathrm{E}[\,V_k \mid \mathcal{F}_{t_k}]] \\
&= \mathrm{E}\big[\,V_j\, \partial_x^2 u(X_k,t_k)\, \mathrm{E}\big[\,\big(\Delta X_k^2 - w(X_k,\Delta t)\big) \mid \mathcal{F}_{t_k}\big]\big] \\
&= \mathrm{E}\big[\,V_j\, \partial_x^2 u(X_k,t_k)\, \mathrm{E}\big[\,\Delta X_k^2 \mid \mathcal{F}_{t_k}\big]\big] \;-\; \mathrm{E}\big[\,V_j\, \partial_x^2 u(X_k,t_k)\, w(X_k,\Delta t\big] \; .
\end{aligned}$$

This is zero because of (5), as $\mathrm{E}\big[\,\Delta X_k^2 \mid \mathcal{F}_{t_k}\big] = w(X_k,\Delta t)$.

The diagonal terms have the form

$$\mathrm{E}\big[\,V_k^2\big] = \mathrm{E}\Big[\,\big(\partial_x^2 u(X_k,t_k)\big)^2\,\big(\Delta X_k^2 - w(X_k,\Delta t)\big)^2\Big]\ .$$

We will see that $\mathrm{E}\big[\,V_k^2\big] = O(\Delta t^2)$. This implies that the sum of the diagonal terms is $O(\Delta t) \to 0$ as $n \to \infty$. The factor involving $\partial_x^2 u$ is bounded by assumption (u has enough bounded derivatives).

Here's a "back of the envelope" summary of the argument. Suppose Y_k is a family of random variables with mean μ and variance σ^2. To keep it simple, suppose the Y_k are independent and gaussian. Define

$$S_n = \sum_{t_k<T} Y_k\ .$$

Suppose $m_n = \mathrm{E}[\,S_n]$ has a limit as $\Delta t \to 0$. Let t_N be the largest t_k. That is, let N_n be so that $t_N = \max\{t_k \mid t_k < T\}$. Then t_N is within Δt of T. The number of terms is $N \approx T/\Delta t$. The sum S_n is the sum of N independent terms. Therefore $m_n = N_n\mu \approx T\mu/\Delta t$. The mean of S_n will have a finite limit m as $\Delta t \to 0$ if $\mu \approx m\Delta t/T$. That is, if $\mu = O(\Delta t)$. In the calculations that went into our Ito integral/Ito Lemma proofs, ΔX_k and $\partial_x^2 u_k \Delta X_k^2$ have expected value on the order of Δt.

The variance calculation is similar. We have

$$\mathrm{var}(S_n) = N\sigma^2 \approx T\sigma^2/\Delta t\ .$$

This has a finite limit as $\Delta t \to 0$ if $\sigma^2 = O(\Delta t)$. The Ito integral contributions $f_{t_k}\Delta X_k$ have variance of order Δt. That's why the sum that approximates the Ito integral has a finite variance. But ΔX_n^2 has variance $O(\Delta t^2)$ (an exercise). If $\sigma^2 = O(\Delta t^2)$ then $\mathrm{var}(S_n) = O(\Delta t)$. This goes to zero as $\Delta t \to 0$.

3 Applications

The easy applications are the reward for working through all that theory. Stochastic calculus will now start looking more like applied math and less like theorem/proof pure math.

3.1 Geometric Brownian motion

Geometric Brownian motion is the solution to the SDE

$$dS = \mu S_t dt + \sigma S_t dW_t\ . \tag{7}$$

The trick is to find the SDE satisfied by

$$X_t = \log(S_t)\ .$$

The Ito calculus applies. Take $u(s) = \log(S)$, with derivatives $\partial_s u(x) = \frac{1}{s}$ and $\partial_s^2 u(s) = -\frac{1}{s^2}$. The infinitesimal variance is $\sigma^2 S_t^2$. Informally we write

$$(\sigma S_t dW_t)^2 = \sigma^2 S_t^2 \left(dW_t\right)^2 = \sigma^2 S_t^2 dt \ ,$$

But this isn't really true on the differential level, only on the integral level. Ito's lemma (1) and the SDE (**??**) are used in the following calculation:

$$\begin{aligned}
dX_t &= d\log(S_t) \\
&= \partial_s u(S_t) dS_t + \frac{1}{2}\partial_s^2 u(S_t) v(S_t) dt \\
&= \frac{1}{S_t}\left(\mu S_t dt + \sigma S_t dW_t\right) - \frac{1}{2}\frac{1}{S_t^2}\sigma^2 S_t^2 dt \\
&= \left(\mu - \frac{1}{2}\sigma^2\right) dt + \sigma dW_t \ .
\end{aligned}$$

You can integrate the two terms, and use $W_0 = 0$, to get

$$X_T = X_0 + \left(\mu - \frac{1}{2}\sigma^2\right) T + \sigma W_T \ .$$

Therefore (using $e^{X_0} = S_0$)

$$\begin{aligned}
S_T &= e^{X_T} \\
&= e^{X_0} e^{\left(\mu - \frac{1}{2}\sigma^2\right)T + \sigma W_T} \\
S_T &= S_0 e^{\left(\mu - \frac{1}{2}\sigma^2\right)T + \sigma W_T} \ .
\end{aligned} \tag{8}$$

We see that S_T is a function of W_T only. In general, the solution to an SDE depends on the whole path $W_{[0,T]}$. There is an example of this in Assignment 4, the Ornstein Uhlenbeck process.

It is interesting to examine the solution in the special case of zero mean growth, $\mu = 0$. In this case, S_t is a martingale:

$$\mathrm{E}[\, dS_t \mid \mathcal{F}_t] = 0 \ .$$

This implies that

$$\mathrm{E}[\, S_t] = S_0 \ .$$

Some economists and political philosophers take S_t as a model of long term wealth accumulation in a "fair" society. Suppose dt is the time period of a generation. Then you are born with S_t wealth and you leave to your child (assuming one parent $\Longrightarrow$ one child in this simple model) $S_t + \sigma dW_t$. That is, you might die richer or poorer, but the expected value is zero. The wealth moves from family to family in one generation without being created or destroyed. Everyone has the same independent random process – the same chance to "get ahead". Your expected wealth after t generations is still S_0.

But the solution formula (8) implies that $S_t \to 0$ as $t \to \infty$ almost surely. This is because W_t is on the order of $\sqrt{t}$, so the Brownian motion part is dominated by the deterministic part $-\frac{1}{2}\sigma^2 t$. There is a proof in assignment 3. If we think of a society as made of many independent "copies" of the process S_t, then most of them have $S_t \to 0$ as $t \to \infty$. Only a few have $S_t \gg S_0$, so that the average is still S_0. The total wealth becomes concentrated in fewer richer families as $t \to \infty$.

3.2 Ornstein Uhlenbeck, linear processes

The Ornstein Uhlenbeck process is governed by the SDE

$$dX_t = -\gamma X_t dt + \sigma dW_t \; . \tag{9}$$

Models like this are used to model small fluctuations about a steady state. The resting state is $X = 0$. The noise parameter σ represents the strength of the noise that drives X away from equilibrium. The *damping* parameter γ represents the strength of the force (or some other tendency) that returns X to its resting state. Assignment 4 examines the OU process in some technical detail.

The OU process X_t settles down to a statistical equilibrium. The value of X_t never stops changing, but the PDF $f(x,t)$ converges to a limit as $t \to \infty$. The limiting PDF, which is the *equilibrium density* is $f_\infty(x)$. Assignment 4 gives a simple formula for the equilibrium density.

We can learn about the equilibrium density using Ito calculations inside expectations. Suppose $u(x)$ is "any" function in the sense that the precise form does not matter but the calculations may not apply. Let X_t be a diffusion with infinitesimal mean $a(x)$ and infinitesimal variance $v(x)$. The time derivative of the expectation value may be calculated as follows

$$\begin{aligned} d\,\mathrm{E}[\,u(X_t)] &= \mathrm{E}[\,d\,u(X_t)] \\ &= \mathrm{E}\left[\,\partial_x u(X_t) dX_t + \frac{1}{2}\partial_x^2 u(X_t)\, v(X_t) dt\right] \\ &= \mathrm{E}\left[\,\partial_x u(X_t) a(X_t) dt + \frac{1}{2}\partial_x^2 u(X_t)\, v(X_t) dt\right] \\ &= \mathrm{E}\left[\,\partial_x u(X_t) a(X_t) + \frac{1}{2}\partial_x^2 u(X_t)\, v(X_t)\right] \, dt \; . \end{aligned}$$

We were able to replace dX_t with its expectation value $a(X_t)dt$ because we are taking expectations. If you think about this point more deeply, you may conclude that the tower property is involved. In more common notation, this may be written

$$\frac{d}{dt}\mathrm{E}[\,u(X_t)] = \mathrm{E}\left[\,\partial_x u(X_t) a(X_t) + \frac{1}{2}\partial_x^2 u(X_t)\, v(X_t)\right] \; . \tag{10}$$

For example, for the OU process we calculate

$$\frac{d}{dt}\mathrm{E}[\,X_t] = -\gamma \mathrm{E}[\,X_t] \; .$$

This implies that

$$\mathrm{E}[\,X_t] = e^{-\gamma t}\mathrm{E}[\,X_0] \ .$$

The expected value converges to the resting value exponentially with rate γ as $t \to \infty$.

But X_t does not converge to the resting value X_0 as $t \to \infty$. We can understand this to some extent by assuming mean zero ($\mathrm{E}[\,X_0] = 0$) and computing the time dependence of the variance

$$\begin{aligned}\frac{d}{dt}\mathrm{E}\left[\,X_t^2\right] &= 2\mathrm{E}[\,X_t(-\gamma X_t)] + \mathrm{E}\left[\,\sigma^2\right] \\ &= -2\gamma\mathrm{E}\left[\,X_t^2\right] + \sigma^2 \ .\end{aligned}$$

We write $S_t = \mathrm{E}\left[\,X_t^2\right]$. In the equilibrium probability density, $\frac{d}{dt}S_t = 0$. Here, that leads to

$$\begin{aligned}-2\gamma S_\infty + \sigma^2 &= 0 \\ S_\infty &= \frac{\sigma^2}{2\gamma} \ .\end{aligned}$$

The differential equation

$$\frac{d}{dt}S_t = -2\gamma S_t + \sigma^2$$

has solution

$$S_t = \frac{\sigma^2}{2\gamma} + e^{-2\gamma t}\left(S_0 - \frac{\sigma^2}{2\gamma}\right) \ .$$

The variance at time t converges exponentially to the steady state variance, which is not zero.

Physicists refer to these calculations as the *fluctuation dissipation theorem.* It is possible to determine S_∞ from the principles of equilibrium statistical mechanics. It is possible to determine γ from simple dynamical models. Finding σ is harder, as it requires modeling the noise process. The fluctuation dissipation theorem tells you that you can determine σ from S_∞ and γ. This argument was first used by Einstein in his theory of Brownian motion.

4 The Ito isometry formula

The formula applies to a stochastic integral with respect to a martingale X_t with infinitesimal variance $v(x)$. It is important that X_t is a martingale – the formula is not true if the infinitesimal mean is not zero.

$$\mathrm{E}\left[\int_0^T f_t dX_t\right]^2 = \int_0^T \mathrm{E}\left[\,f_t^2 v(X_t)\right] dt \ . \tag{11}$$

This formula is a fancy version of the fact that when you add independent random variables the variance of the sum is the sum of the variances. This fact applies not only to independent random variables, but also to martingales. The calculation is the same. The left side of (11) is the variance of the sum (the integral). The right side is the sum (the integral) of the variances. As an example, apply this to $\int W dW$. The conclusion should be

$$\frac{1}{4}\mathrm{E}\left[\, W_T^4 \right] - \frac{t}{2}\mathrm{E}\left[\, W_T^2 \right] + \frac{1}{4}T^2 = \int_0^T t\, dt \; .$$

There is more than one way to derive the Ito isometry formula. One uses the Ito calculus. Another starts with the approximations to the Ito integral, does a calculation on the sum that involves figuring out why the off diagonal terms in the square have expected value zero, then taking the limit $n \to \infty$.