Derivation of Itô's lemma

Himalaya Senapati

October 20, 2024

Abstract

We begin by introducing the notion of an ordinary integral and numerically establishing why it does not extend easily to stochastic integrals. We then introduce Itô's lemma also called Itô–Doeblin formula and give a heuristic derivation via Taylor's series. We conclude by deriving the Geometric Brownian Motion solution of the stochastic differential equation followed by the stock prices.

Contents

1 Motivation **2**

2 Riemann Integration **2**

2.1 Numerical methods . . . 4

3 Stochastic Differential Equation **5**

3.1 Brownian Motion . . . 6

3.2 Why Riemann Integral does not work: Bounded Quadratic Variation . . . 6

3.3 Itô's lemma . . . 8

3.4 Geometric Brownian Motion . . . 8

1 Motivation

• Stochastic calculus/Itô calculus deals with stochastic/random processes and allows the modeling of random systems.

• Analyzing historical data, we notice that stock price dynamics can be modeled using a diffusion process called 'Brownian motion' or 'Wiener process' (see Sec. 3.1). We also want to compute the prices of 'derivative products' whose pay off depends on the value of the underlying stocks. It thus becomes imperative to be able to compute the so called 'Itô integrals' that tells us how to integrate with respect to the Brownian Motion measures.

• We would also want to be able to differentiate time-dependent functions of Brownian motions motivating the need for a 'stochastic version' of the chain rule in ordinary differential calculus: *Itô's lemma*.

• Itô calculus/ Itô's lemma is extensively used in mathematical finance:

- Option Pricing Formulas, for example, Black-Scholes.
-

• We begin by recalling Riemann Integration that tells us how to integrate deterministic functions. We will then introduce Brownian Motions and see that some of the intuitions do not carry over to integration of such stochastic functions. Thus, it needed the development of a new theory, 'Itô Calculus'. We will then heuristically derive 'Itô's Lemma'. We will conclude by giving a closed form solution to a SDE model of stock price dynamics.

2 Riemann Integration

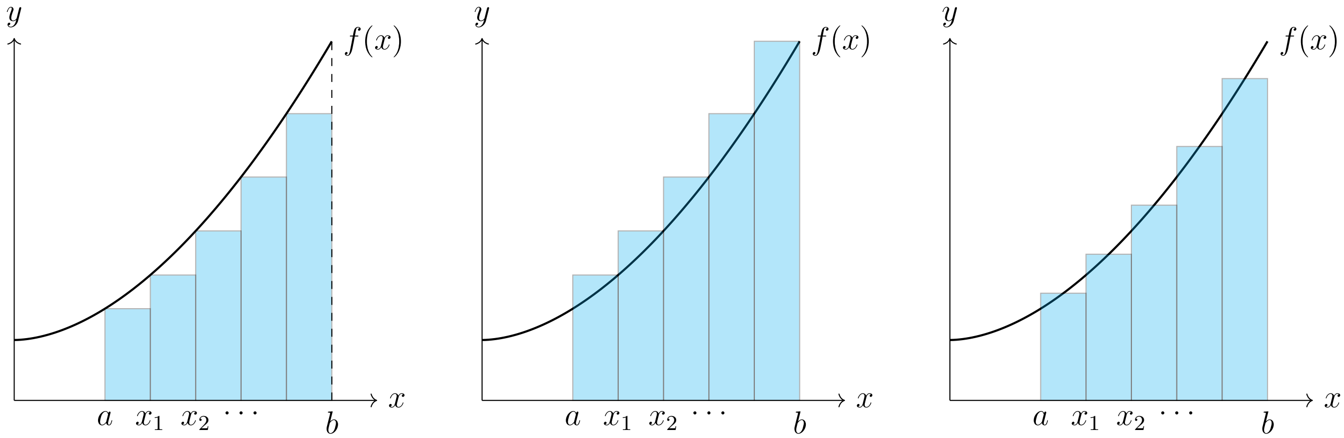
• The integral of a function $f(x)$ on a closed interval $[a, b]$ is defined as the area under the curve of the function on that interval. *Riemann integral* formalizes this notion of integration and gives a rigorous definition. Moreover, it motivates naive numerical approaches as we will see in Sec. 2.1.

• The idea is to approximate the area under the curve of the function by adding up the areas of rectangles as shown in Fig. 1. As the rectangles become narrower (i.e., the partitions get finer), the total area of the rectangles approaches the true area under the curve.

• More formally, to define a Riemann integral, the interval $[a, b]$ is divided into smaller subintervals. A partition of the interval $[a, b]$ is a finite set of points $x_0, x_1, x_2, \ldots, x_n$ such that:

$$a = x_0 < x_1 < x_2 \cdots < x_n = b. \tag{1}$$

We then choose a sample point c_i in each subinterval $[x_{i-1}, x_i]$ for $i = 1, 2, \ldots n$.



(a) Left sum (b) Right sum (c) Center sum

Figure 1: Riemann Integral: first-order (a,b) and second-order (c) numerical methods.

• The Riemann sum is defined as $S_P = \sum_{i=1}^{n} f(c_i)\Delta x_i$ where $\Delta x_i = x_i - x_{i-1}$ is the length of the i^{th} subinterval and P denotes the mesh size, say $\max(\Delta x_i)$.

• The Riemann Integral is now defined as

$$\boxed{\int_a^b f(x)\,dx = \lim_{\max(\Delta x_i)\to 0} \sum_{i=1}^{n} f(c_i) \times \Delta x_i.} \tag{2}$$

• Note that, here we are asserting that the integral is independent of all sampling choices c_i. In particular, $c_i = x_{i-1}$ (left sum), $c_i = x_i$ (right sum) and $c_i = (x_{i-1} + x_i)/2$ (center sum) will all converge to the same limiting value.

• **Why Riemann integrability?** Riemann integral is the first rigorous definition of integration. One would think that defining integral as area under the curve is already rigorous. But, then the question becomes what is 'area'. In fact, as we learnt in the measure theory notes, there are subsets in the plane whose area we cannot define – they are non-measurable sets. Thus, one can also take Riemann integral as the rigorous definition of area of a class of 'special regions'.

• We are also interested in finding the set of all functions, and set of all intervals, for which the Riemann integral exists. These functions would be called 'Riemann integrable' over those intervals. Nicely behaving 'bounded' and 'continuous' functions are always Riemann integrable over all bounded intervals. Intuitively, it is clear that 'bounded' but piece-wise continuous functions that are discontinuous at finitely many points should also be Riemann integrable. In fact, as long as a function is bounded as well as 'continuous almost everywhere', it is Riemann integrable over any bounded intervals.

• **Riemann integrability is not general enough:** Suppose we want to find the integral of the so-called Dirichlet function

$$f(x) = \begin{cases} 1 & \text{if} \quad x \in \mathbb{Q} \\ 0 & \text{if} \quad x \notin \mathbb{Q} \end{cases} \tag{3}$$

over the interval $[0,1]$: $\int_0^1 f(x)\ dx$. Note that $f(x)$ is not Riemann integrable, as it is *NOT continuous anywhere.* This is because, near any irrational number, we can get as close a rational number[1] as we want and vice versa. Thus, at all points on the interval, there are very close points, with $f(x)$ taking differing values of 0 and 1 on them. Though $f(x)$ is not Reimann integrable, recall from measure theory notes that $\mathbb{Q}$ is of measure zero. Hence, we would intuitively expect $\int_0^1 f(x)\ dx = 0$. This suggests that we would want to expand the scope of integrable functions.

what it means?

• **Riemann vs Lebesgue integrals:** Going back to intuition, integration is defined as the area under the curve. Thus, we can expect the integration to be defined for functions where the region under the curve is 'measurable', i.e., we can assign an area to them. This leads to a measure theoretic definition of integrals that generalizes Riemann integrals to Lebesgue integrals. In practice, we do not need to worry about the difference between *Riemann* and *Lebesgue*.

• There exists another generalization to *Improper* Riemann integrals. Note that Riemann integral is only defined for bounded functions on bounded intervals. However, it can be generalized to *improper* Riemann integrals via limits: a function $f(x)$ is improperly Riemann-integrable on the open interval (a,b) if $\int_c^d f(x)\ dx$ exists for all c and d with $a < c < d < b$ and

$$\lim_{\substack{c\to a\\ d\to b}} \int_c^d f(x)\ dx \quad \text{converges.} \tag{4}$$

There can be functions that are *improper Riemann integrable* but **not** *Lebesgue integrable*, eg., $\int_{\mathbb{R}} \frac{\sin(x)}{x}\ dx$.

2.1 Numerical methods

• We are also interested in finding numerical methods to compute integrals of functions where a closed form solution is not available. We take inspiration from the definition of Riemann integral, and can compute either of the three candidates, see Fig. 1 –

- Left sum $= \sum_{i=1}^n f(x_{i-1}) \times \Delta x_i$.
- Right sum $= \sum_{i=1}^n f(x_i) \times \Delta x_i$.
- Trapezoidal sum/rule $= \sum_{i=1}^n \frac{f(x_{i-1})+f(x_i)}{2} \times \Delta x_i$ [kind of similar to center-sum].

• Of course in the limit of $\max(\Delta x_i) \to 0$, the three quantities will converge. However, we are interested in the error for finite mesh size, i.e., size of $\max(\Delta x_i)$. In practice, we take equally spaced intervals, so that $\Delta x_i = \Delta x$.

• We will *heuristically* derive the errors assuming 'nicely behaved functions' so that we can apply Taylor series for approximations:

$$\boxed{f(x_i) = f(x_{i-1} + \Delta x) = f(x_{i-1}) + f'(x)|_{x=x_{i-1}}\Delta x + \frac{1}{2} f''(x)|_{x=x_{i-1}}(\Delta x)^2} \tag{5}$$

[1] Formally, rationals are *dense* in reals.

where $f'(x) = \frac{d}{dx} f(x)$. Roughly speaking, errors for a single sub interval for Left or Right sum would be sums of the areas of the triangular regions below/above the curve in Fig. 1. Similarly, the error for the Trapezoidal sum would involve differences of areas of blue and white triangles above and below the curve similar to the center sum in Fig. 1c. For brevity, we give a rough sketch of the errors in both methods below, as the exact coefficients are not that important.

• Approximating an area of the small rectangles either via the 'Left sum' or the 'Right sum', the error would be of the order *x-length* $= \Delta x$ multiplied by *y-error* $= \mathcal{O}\left(f'(x) \times \Delta x\right)$. The total error is obtained after summing over each rectangle errors:

$$\mathcal{O}\left(\Delta x \times f'(x) \times \Delta x \times n\right) \sim \mathcal{O}\left(\frac{(b-a)^2}{n}\right). \tag{6}$$

• Proceeding similarly for the trapezoidal rule, we note that $\frac{f(x_i)+f(x_{i-1})}{2}$ differs from both $f(x_i)$ and $f(x_{i-1})$ in the order of $f''(x)|_{x=x_{i-1}} \times \mathcal{O}((\Delta x)^2)$. Thus, the error estimate in this method turns out to be

$$f''(x) \times \mathcal{O}(\Delta x^3) \times n \sim \mathcal{O}\left(\frac{(b-a)^3}{n^3}\right) \times n \sim \mathcal{O}\left(\frac{(b-a)^3}{n^2}\right). \tag{7}$$

• It is clear that, for large n, the error for left and right sums go as $1/n$, these are called first-order methods. Similarly, the error for the trapezoidal rule goes like $1/n^2$, this is a second-order method.

3 Stochastic Differential Equation

• Asset prices are often modeled as stochastic processes. In particular, analyzing historical prices, we see that the stock prices follow the following 'stochastic differential equation' (SDE):

$$\boxed{\frac{dS(t)}{S(t)} = \mu\, dt + \sigma\, dB(t)} \tag{8}$$

where B(t) is the canonical Brownian Motion to be defined below. Moreover, if we consider more complex 'derivative' products whose prices depend on the 'underlying' stock prices, they follow more complicated SDEs. Thus, we would like to understand how to represent dS_t/S_t in terms of $dF(S_t)$ and dt. For example, suppose we can write

$$\frac{dS(t)}{S(t)} = \alpha \times dF(S_t) + \beta \times dt. \tag{9}$$

Then, we would have

$$\alpha \times dF(S_t) = (\mu - \beta)\, dt + \sigma\, dB(t) \Rightarrow S_T = F^{-1}\left(F(S_0) + \frac{\mu - \beta}{\alpha} \times T + \frac{\sigma}{\alpha}\, B(T)\right). \tag{10}$$

• It thus becomes essential to be able to represent $dF(S(t), t)$ as a function of $d(S(t))$ and dt giving rise to Ito's lemma.

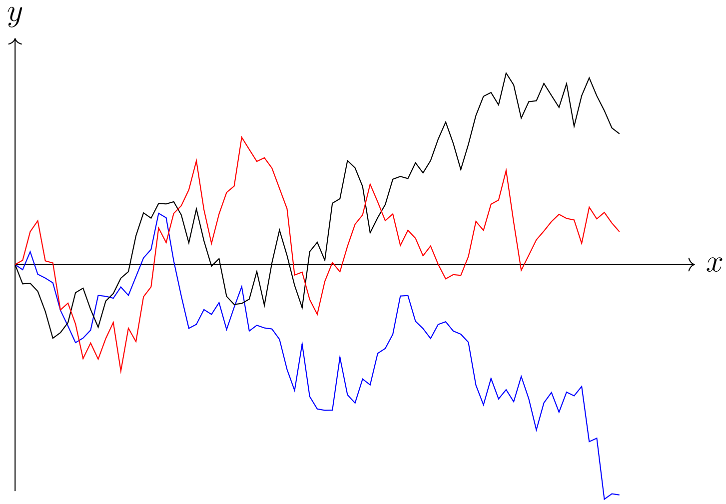


Figure 2: Instances of Brownian motion.

3.1 Brownian Motion

• Formally, the canonical Brownian motion/Wiener process denoted $B(t)$ or $W(t)$ is characterized by four axioms/facts

1. $W(0) = 0$
2. $W(t)$ has independent increments
3. $W(t) - W(s) \sim \mathcal{N}(0, t-s)$ for $0 \leq s \leq t$.
4. $W(t)$ is continuous *almost surely.* revise what is almost surely

The 4^{th} condition is a bit subtle, and at a first glance is *not* required. However, it is possible to construct stochastic processes[2] that obey conditions (1)–(3) but are not continuous almost surely. We will not expand on this. In practice, we can simply work with (1)–(3) and get the relevant results.

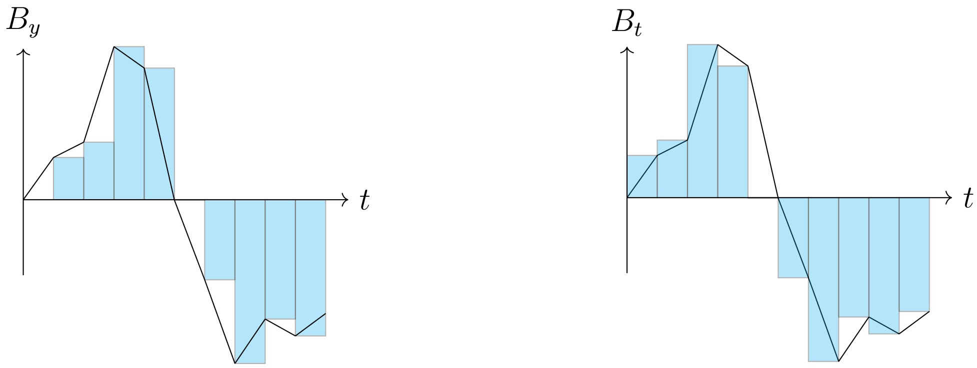
• In practice, we can simulate Brownian motions by partitioning $[0, T]$ into subintervals $0 = t_0 < t_1 < t_2 < \cdots t_n = T$, setting $B(0) = 0$ and incrementing $B(t_i)$ by sampling from $\mathcal{N}(0, t_{i+1} - t_i)$. See Fig. 2 for examples.

3.2 Why Riemann Integral does not work: Bounded Quadratic Variation

do it later

• To clarify, as $B(t)$ is bounded, $\int_0^T B(t)dt$ should be Riemann integrable. In fact, one can show that the left sum as well as the right sum, both converge to the same distribution: $\mathcal{N}(0, T^3/3)$, see Fig. 3 [left as an exercise, :)]. However, the difficulty is in integrating with respect to the Brownian motion measure, i.e., to compute $\int_0^T f(t, B_t)dB_t$.

[2] $V_t = W_t$ except for $s \in [0, 1]$ from uniform distribution with $V_s = 0$, link.



(a) Left sum (b) Right sum

Figure 3: Area under the Brownian motion curve.

• Let us take the integral $\int_0^T B_t \, dB_t$. We notice from the accompanying Jupyter notebook that, numerical simulations illustrate the difference between the Left Sum and the Right Sum. In fact, the difference, also related to quadratic variation,

$$\sum_{i=1}^{n} B(t_{i+1})\Delta B(t_i) - \sum_{i=1}^{n} B(t_i)\Delta B(t_i) = \sum_{i=1}^{n} \Delta B(t_i)^2 \tag{11}$$

does not vanish in the limit to 0. Rather it seems to go like t. This is called the Bounded Quadratic Variation property of the Brownian motions.

• As the left and right sums do not converge, Riemann integral is not well defined here. In other words, the value of the integral depends on the way we sample points from the sub intervals while computing the areas of the rectangles.

• **Heuristically deriving Bounded Quadratic Variation:** We want to compute $\sum_{i=1}^{n} \Delta B(t_i)^2$ in the limit. Notice that, $\Delta B(t_i) \sim \mathcal{N}(0, \Delta t)$. This implies $E(\Delta B(t_i)^2) = \Delta t$ implying that

$$E\left(\sum_{i=1}^{n} \Delta B(t_i)^2\right) = T. \tag{12}$$

Moreover, $Var(\Delta B(t_i)^2) = 2(\Delta t)^2$ implying that the variance

$$Var\left(\sum_{i=1}^{n} \Delta B(t_i)^2\right) = 2T\Delta t \tag{13}$$

vanishes in the limit $\Delta t \to 0$. Thus, it makes sense that the difference between the Right and the Left sums is given by

$$\lim_{\Delta t \to 0} \sum_{i=1}^{n} \Delta B(t_i)^2 = t. \tag{14}$$

• This is the reason we can write $dB \wedge dB = dB^2 \sim dt$.

3.3 Itô's lemma

• Itô's lemma is in a sense a stochastic calculus version of the *chain rule* and is used to express $dF(B(t), t)$ as a function of $dB(t)$ and dt.

• **Heuristic derivation via Taylor's series:**

these vanish in limit

$$\begin{aligned} dF &= \frac{\partial}{\partial t}F \times dt + \boxed{\frac{1}{2}\frac{\partial^2}{\partial t^2}F \times dt^2} + \frac{\partial}{\partial B}F \times dB + \frac{1}{2}\frac{\partial^2}{\partial B^2}F \times dB^2 + \boxed{\frac{\partial^2}{\partial t \partial B}F \times dt\ dB} + \dots \\ &= \left(\frac{\partial}{\partial t}F + \frac{1}{2}\frac{\partial^2}{\partial B^2}F\right) \times dt + \frac{\partial}{\partial B}F \times dB. \end{aligned} \tag{15}$$

• We essentially used the fact that the only terms that do not vanish in the limit are dt, dB and $dB^2 = dt$.

• **Keep track of second order terms:** The *cardinal* rule in stochastic calculus is to be careful with the second order terms, as we are used to taking them to zero in the ordinary calculus. While $dt^2 = dB\ dt = 0$, $dB^2 = dt$.

• **Chain rule:** It is evident that $d(XY) = X\ dY + Y\ dX + dX\ dY$ as we cannot ignore the second order term.

3.4 Geometric Brownian Motion

• Empirically, it is seen that dynamics of stock prices is captured well by the following Stochastic Differential Equation:

$$\frac{dS}{S} = \mu\ dt + \sigma\ dB_t. \tag{16}$$

• We would like to understand the distribution of $S(T)$ given $S(0), \mu$ and σ. To this effect, we would like to express $\frac{dS}{S}$ in terms of $df(S)$ for some function $f(S)$. Our intuition from usual calculus would suggest to try the logarithmic function:

$$d\log(S) = \frac{dS}{S} - \frac{1}{2S^2}dS \wedge dS. \tag{17}$$

use ito's lemma to arrive at 17

As noted before, we always have to be careful about the second order effects, they do not always vanish. In fact, from eq. (16), and using the heuristic $dt^2 = 0$, $dB\ dt = 0$ and $dB^2 = dt$,

$$dS \wedge dS = \sigma^2 \times S^2\ dt. \tag{18}$$

squre the equation 16

This implies

$$d\log(S) = \frac{dS}{S} - \frac{1}{2}\sigma^2\ dt \Rightarrow \frac{dS}{S} = d\log(S) + \frac{1}{2}\sigma^2\ dt. \tag{19}$$

Hence, from eq. (16),

$$d\log(S) = \left(\mu - \frac{1}{2}\sigma^2\right) dt + \sigma\ dB_t. \tag{20}$$

Integrating both sides, we get

$$S(T) = S(0) \times \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)T + \sigma B(T)\right). \tag{21}$$

We say that, $S(t)$ follows a Geometric Brownian Motion.