STOCHASTIC CALCULUS AND NUMERICAL METHODS FOR SOLVING STOCHASTIC DIFFERENTIAL EQUATIONS

BRADLEY YU

Abstract. Stochastic calculus is concerned with finding the solutions to stochastic differential equations. However, this is difficult since the usual tools of integral and differential calculus are not well-defined on random continuous motion. Thus we introduce a new form of calculus to assist us in solving stochastic differential equations. In particular, we introduce the construction of the stochastic integral and derive Ito's Formula as a consequence. We can use Ito's Formula to solve for the solutions to certain SDEs, however not all SDEs are solvable analytically. Finally, we will introduce numerical methods as an alternative scheme to assist in solving SDEs and discuss their orders of convergences to the solution. This paper assumes basic knowledge of probability theory, but we will introduce all concepts in this paper as needed. No prior knowledge of differential equations is needed.

Contents

1. Introduction 1
2. Probability Theory 2
3. Martingales 4
4. Brownian Motion 7
5. Quadratic Variation 9
6. Stochastic Calculus 12
7. Itô's Formula 16
8. Numerical Methods and Simulations 20
Acknowledgments 28
References 28
Appendix A. Code for Graphs and Numerical Simulations 28

1. Introduction

Stochastic calculus is about extending calculus to random variables, specifically random continuous motion, or Brownian motion. For example, you might want to calculate the expected winnings of a certain strategy in the stock market. As it turns out, we can model asset prices as a random continuous process. One of the most famous applications of stochastic calculus is the Black-Scholes formula, which gives the theoretical price of a European Option at some future time. Aside from financial mathematics, other applications of stochastic calculus include control theory and mathematical biology.

Date: DEADLINES: Draft AUGUST 14 and Final version AUGUST 28, 2021.

In Section 2, we review probability theory. In Sections 3 and 4, we go over martingales and the construction of Brownian motion through martingales. Section 5 introduces the concept of quadratic variation, which is used in Section 6 to develop the stochastic integral. We derive Ito's Formula in Section 7. In Section 8, we go over weak and strong convergence of Euler-Maruyama and Milstein schemes for stochastic differential equations.

2. Probability Theory

While the mathematical foundations of probability theory are rooted in measure theory, most of what we need from probability theory does not require the full machinery of measure theory.

Definition 2.1. A *sample space* Ω is a set. Elements of the sample space $\omega \in \Omega$ are called outcomes.

Intuitively, a sample space is the range of possible outcomes that a random variable can take. For example, suppose that our random variable is the number on a dice roll. Then our sample space would be $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Definition 2.2. The set of events $\mathcal{F}$ is a *σ-algebra* of subsets of Ω that satisfy the following properties:

- $\emptyset \in \mathcal{F}$
- If $A \in \mathcal{F}$, then $\Omega \setminus A \in \mathcal{F}$
- If $A_1, A_2, \ldots \in \mathcal{F}$, then $\bigcup_{k=1}^{\infty} A_k \in \mathcal{F}$

We can see that if $\emptyset \in \mathcal{F}$, then it follows by the second bullet point that $\Omega \in \mathcal{F}$. Each element $E \in \mathcal{F}$ is called an *event*.

Example 2.3. The power set $\mathcal{F} = 2^{\Omega}$ is a σ-algebra on Ω.

Example 2.4. If Ω is a topological space, then $\sigma\{A \subset \Omega : A \text{ is open}\}$ is called the *Borel σ-algebra* on Ω and is denoted by $\mathcal{B}(A)$.

When working with continuous spaces, such as $\Omega = \mathbb{R}$, often times we will let $\mathcal{F} = \mathcal{B}(\Omega)$.

Definition 2.5. A *probability measure* is a function $\mathbb{P} : \mathcal{F} \to [0, 1]$ such that

- $\mathbb{P}(\emptyset) = 0$ and $\mathbb{P}(\Omega) = 1$
- If $E_1, E_2, \ldots \in \mathcal{F}$ are pairwise disjoint, then $\mathbb{P}(\bigcup_{k\in\mathbb{N}} E_k) = \sum_{k\in\mathbb{N}} \mathbb{P}(E_k)$

Definition 2.6. A probability space is a measure space with total measure one. It is denoted by $(\Omega, \mathcal{F}, \mathbb{P})$.

The simplest probability space is a discrete probability space where Ω is finite or countable. Often times, we will take Ω to be $\mathbb{R}$.

Definition 2.7. A *random variable* (or equivalently, $\mathcal{F}$-measurable function), is a function $f : \Omega \to \mathbb{R}$ from $(\Omega, \mathcal{F})$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that for every $A \in \mathcal{B}(\mathbb{R})$, the preimage

$$f^{-1}(A) \equiv \{\omega \in \Omega \colon f(\omega) \in A\} \in \mathcal{F}$$

.

For example, suppose we have some process $X : \Omega \to \mathbb{R}$ that returns a real number. We want to measure the probability that the outcome of this process occurs in some

set $A \in \mathcal{B}(\mathbb{R})$. Ideally, this would be $\mathbb{P}[X^{-1}(A)]$, which is only well-defined if X is a measurable function.

Definition 2.8. A *simple random variable* $X : \Omega \to \mathbb{R}$ is a random variable that can only take finitely many values $X(\Omega) = \{x_1, x_2, \ldots, x_n\}$. We define the *expectation* of X to be

$$\mathbb{E}[X] = \sum_{i=1}^{n} x_i \mathbb{P}[X = x_i]$$

An example of a simple random variable would be the result of a dice roll $X(\Omega) = \{1, 2, 3, 4, 5, 6\}$. The expectation of a random variable can be interpreted as its weighted average of all the possible outcomes.

The natural continuation of expectation to apply to continuous random variables would be to integrate over all the possible outcomes.

Definition 2.9. The expectation of a random variable is given by

$$\mathbb{E}[X] = \int_{\omega \in \Omega} X(\omega) \mathbb{P}[d\omega] = \int_{\Omega} X \, d\mathbb{P}$$

where $\int$ is the Lebesgue integral. (The Lebesgue integral generalizes the Riemann integral to measurable functions, while the Riemann integral can only integrate continuous functions)

Definition 2.10. The variance of a random variable X is given by

$$\mathbb{V}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

A specific type of random variable is called the *indicator function* of an event.

Definition 2.11. An indicator function for an event E is a random variable $1_E : \Omega \to \mathbb{R}$ defined by

$$1_E(\omega) = \begin{cases} 1 & \omega \in E \\ 0 & \omega \notin E \end{cases}$$

Definition 2.12. The conditional expectation of a random variable X given a random variable Y is a unique random variable denoted as $\mathbb{E}[X \mid Y]$. The conditional expectation satisfies:

- $\mathbb{E}[X \mid Y]$ is $\sigma(Y)$-measurable.
- For every Y-measurable event A, $\mathbb{E}[\mathbb{E}[X \mid Y]1_A] = \mathbb{E}[X1_A]$

Intuitively, the conditional expectation of a random variable X given Y is the "best guess" of X given the information Y, where Y is a random variable. If X is independent of Y, then $\mathbb{E}[X \mid Y] = \mathbb{E}[X]$. If X is Y-measurable, then $\mathbb{E}[X \mid Y] = X$. By taking $A = \Omega$, then we can see that $\mathbb{E}[\mathbb{E}[X \mid Y]] = \mathbb{E}[X]$, i.e. our expected value of a random variable does not change based on what information we have.

Proposition 2.13. *The conditional expectation $\mathbb{E}[X \mid Y]$ satisfies the following properties:*

- ***Linearity**: If a, b are constants and X, Z are random variables, then*

$$\mathbb{E}[aX + bZ \mid Y] = a\mathbb{E}[X \mid Y] + b\mathbb{E}[Z \mid Y]$$

- ***Tower Property**: If Z is Y-measurable, then*

$$\mathbb{E}[\mathbb{E}[X \mid Y] \mid Z] = \mathbb{E}[X \mid Z]$$

- *If Z is a Y-measurable random variable, then when conditioning with respect to Y, Z acts like a constant.*

$$\mathbb{E}[XZ \mid Y] = Z\mathbb{E}[X \mid Y]$$

Proof. By the definition of conditional expectation, $a\mathbb{E}[X \mid Y] + b\mathbb{E}[Z \mid Y]$ is a Y-measurable random variable. Then for any Y-measurable event A,

$$\begin{aligned}\mathbb{E}[1_A(a\mathbb{E}[X \mid Y] + b\mathbb{E}[Z \mid Y])] &= a\mathbb{E}[1_A\mathbb{E}[X \mid Y]] + b\mathbb{E}[1_A\mathbb{E}[Z \mid Y]] \\ &= a\mathbb{E}[1_A X] + b\mathbb{E}[1_A Z] \\ &= \mathbb{E}[1_A(aX + bZ)]\end{aligned}$$

By the uniqueness of conditional expectation, then

$$\mathbb{E}[aX + bZ \mid Y] = a\mathbb{E}[X \mid Y] + b\mathbb{E}[Z \mid Y]$$

Let A be a Z-measurable event. Since Z is Y-measurable, then A is also a Y-measurable event. It follows by the definition of conditional expection that

$$\begin{aligned}\mathbb{E}[\mathbb{E}[\mathbb{E}[X \mid Y] \mid Z]1_A] &= \mathbb{E}[\mathbb{E}[X \mid Y]1_A] \\ &= \mathbb{E}[X1_A]\end{aligned}$$

By the uniqueness of conditional expectation, we have

$$\mathbb{E}[\mathbb{E}[X \mid Y] \mid Z] = \mathbb{E}[X \mid Z]$$

□

3. Martingales

First we will introduce two new definitions.

Definition 3.1. A *filtration* $\{\mathcal{F}_t\}$ is a collection of non-decreasing σ-algebras of $\mathcal{F}$ such that if $s < t$, then $\mathcal{F}_s \subseteq \mathcal{F}_t$.

Intuitively, a filtration $\{\mathcal{F}_t\}$ represents a collection of indexed historical information. For any time t, the filtration $\mathcal{F}_t$ represents all historical information and properties we know of our process up to t.

Definition 3.2. A *stochastic process* is a collection of random variables $\{X_t\}$ that are indexed by time t. We can index in both continuous time ($t \in [0, \infty)$) as well as discrete time ($t \in \{0, 1, 2, \ldots\}$).

A martingale is the representation of a zero-sum game. First, we will give a discrete example of a martingale. Consider the following game: you flip a fair coin and if it lands on heads, you gain a dollar, otherwise you lose a dollar. Let the random variables $X_1, X_2, \ldots$ be a stochastic process representing the results of the coin flip at times $t = 1, 2, \ldots$. We can represent the winnings of the game at time t as the random variable

$$W_t = X_1 + X_2 + \ldots + X_t$$

We can see that $\mathbb{P}[X_n = 1] = \mathbb{P}[X_n = -1] = \frac{1}{2}$. It follows that $\mathbb{E}[X_n] = 0$ for any n. By the linearity property of expected value, we have that

$$\mathbb{E}[W_t] = \mathbb{E}[X_1] + \ldots + \mathbb{E}[X_t] = 0$$

in other words, our expected winnings for this game at any finite time step t is 0.

To find the variance, we want to find $\mathbb{E}[W_t^2]$. Then

$$\begin{aligned}\mathbb{E}[W_t^2] &= \mathbb{E}[(X_1 + X_2 + \ldots + X_t)^2] \\ &= \mathbb{E}\left[\sum_{i=1}^{t} X_i^2 + \sum_{i<j} 2X_iX_j\right] \\ &= \mathbb{E}\left[\sum_{i=1}^{t} X_i^2\right] + \mathbb{E}\left[\sum_{i<j} 2X_iX_j\right]\end{aligned}$$

Since for any $i \neq j$, the results of X_i and X_j are independent, then

$$\mathbb{E}[2X_iX_j] = 2\mathbb{E}[X_i]\mathbb{E}[X_j] = 0$$

for $i < j$. It follows immediately that

$$\mathbb{E}\left[\sum_{i<j} 2X_iX_j\right] = 0$$

Since $X_i^2 = 1$ for all i, then $\mathbb{E}[X_i^2] = 1$. Thus,

$$\begin{aligned}\mathbb{E}\left[\sum_{i=1}^{t} X_i^2\right] &= \sum_{i=1}^{t} \mathbb{E}[X_i^2] \\ &= t\end{aligned}$$

Therefore

$$Var(W_t) = t$$

The winnings of the game $W_1, W_2, \ldots$ is one example of a *simple random walk*. This brings us to our definition of a martingale:

Definition 3.3. A sequence of random variables $M_0, M_1, \ldots$ is a *martingale* with respect to filtration $\{\mathcal{F}_n\}$ if it satisfies the following:

- For all n, M_n is $\mathcal{F}_n$-measurable and $\mathbb{E}[M_n] < \infty$.
- If $m < n$, then $\mathbb{E}[M_n \mid \mathcal{F}_m] = M_m$

We can see that our winnings $W_1, W_2, \ldots$ from the coin toss game is a martingale.

Proof. At time $t = 0$, $W_0 = 0$. It suffices to show by induction and the tower property that for any time t, $\mathbb{E}[W_{t+1} \mid \mathcal{F}_t] = W_t$. Let X_{t+1} be the result of the coin flip at time $t + 1$. Then

$$\begin{aligned}\mathbb{E}[W_{t+1} \mid \mathcal{F}_t] &= \mathbb{E}[W_t + X_{t+1} \mid \mathcal{F}_t] \\ &= \mathbb{E}[W_t \mid \mathcal{F}_t] + \mathbb{E}[X_{t+1} \mid \mathcal{F}_t]\end{aligned}$$

Since W_t is $\mathcal{F}_t$-measurable, then $\mathbb{E}[W_t \mid \mathcal{F}_t] = W_t$. Since X_{t+1} is independent of $\mathcal{F}_t$, then $\mathbb{E}[X_{t+1} \mid \mathcal{F}_t] = \mathbb{E}[X_{t+1}] = 0$. Thus $\mathbb{E}[W_{t+1} \mid \mathcal{F}_t] = W_t$ □

Now, instead of only betting one dollar at each coin toss, we can bet any finite amount of money. Suppose at time t, we can make a bet B_t. If X_t lands heads, then we gain B_t dollars. If X_t lands tails, then we lose B_t dollars. Let $X_1, X_2, \ldots$ represent the result of the coin flip. Let $M_1, M_2, \ldots$ represent the simple random

walk (martingale) given by the results of the coin toss. It follows that we can view our winnings at time t as

$$\begin{aligned} W_t &= \sum_{i=1}^{t} B_i X_i \\ &= \sum_{i=1}^{t} B_i (M_i - M_{i-1}) \end{aligned}$$

If we let $\Delta M_i = M_i - M_{i-1}$, then we can write

$$W_t = \sum_{i=1}^{t} B_i \Delta M_i$$

W_t is an example of a *discrete stochastic integral*.

Proposition 3.4. *W_t is a martingale.*

Proof. Since our bet B_i is finite, for all i, there exists some $K_i < \infty$ such that $|B_i| \leq K_i$. It follows that

$$\begin{aligned} \mathbb{E}[W_t] &= \sum_{i=1}^{t} \mathbb{E}[B_i(M_i - M_{i-1})] \\ &\leq \sum_{i=1}^{t} \mathbb{E}[|B_i||M_i - M_{i-1}|] \\ &\leq \sum_{i=1}^{t} K_i \mathbb{E}[|M_i - M_{i-1}] \\ &< \infty \end{aligned}$$

This satisfies the first condition for martingales.

By the nature of our bet, we must place our bet at time t, B_t, before the result of X_t, i.e. B_t is $\mathcal{F}_{t-1}$-measurable. Also, we can see that W_t is $\mathcal{F}_t$-measurable. Then,

$$\begin{aligned} \mathbb{E}[W_{t+1} \mid \mathcal{F}_t] &= \mathbb{E}[W_t + B_{t+1}(M_{t+1} - M_t) \mid \mathcal{F}_t] \\ &= \mathbb{E}[W_t \mid \mathcal{F}_t] + \mathbb{E}[B_{t+1}(M_{t+1} - M_t) \mid \mathcal{F}_t] \end{aligned}$$

Since W_t is F_t-measurable, then $\mathbb{E}[W_t \mid \mathcal{F}_t] = W_t$. Also, since B_{t+1} is F_t-measurable and M_t is a martingale, then

$$\mathbb{E}[B_{t+1}(M_{t+1} - M_t) \mid \mathcal{F}_t] = B_{t+1}\mathbb{E}[M_{t+1} - M_t \mid \mathcal{F}_t] = 0$$

Thus $\mathbb{E}[W_{t+1} \mid \mathcal{F}_t] = W_t$. □

What is interesting is that no matter what betting strategy $\{B_t\}$ that we pick, our expected winnings at a finite time t in the future will always be zero, i.e. we cannot beat the game using any strategy $\{B_t\}$ in a finite amount of time.

Interestingly enough, this does not necessarily hold true for infinite time. Consider the following betting strategy: We start by betting \$1. If the coin flips heads, we bet \$0 for the rest of the time. Otherwise, we bet \$2, then \$4, and so on continuously doubling our bet. If we win one bet, then our total winnings would be \$1. With probability one, we will win at least once.

To represent this mathematically, let $B_1 = 1$. If $X_1 = X_2 = \ldots = X_n = -1$, then let $B_{n+1} = 2^n$. Otherwise, let $B_{n+1} = 0$. By our previous claim, we know that any betting strategy cannot win. We lose the game only if $X_1 = \ldots = X_t = -1$. This happens with probability 2^{-t} and our winnings will be

$$W_t = \sum_{i=1}^{t} -(2^{i-1}) = -(2^t - 1)$$

We win the game if we do not lose so this happens with probability $1 - 2^{-t}$, where our winnings will always be $W_t = 1$. Then we can compute the expected winnings at time t to be:

$$\mathbb{E}[W_t] = 1 \cdot (1 - 2^{-t}) - (2^t - 1) \cdot 2^{-t} = 0$$

However, we will eventually win with probability one so

$$\mathbb{E}[W_\infty] = \lim_{t \to \infty} \mathbb{E}[W_t] = 1$$

which means that we can beat the game given an infinite amount of time.

We have introduced the concept of a discrete stochastic integral with our coin tossing game. We can now generalize this concept.

Let $X_1, X_2, \ldots$ be identically distributed independent random variables with mean 0 and variance σ^2. Let $S_n = X_1 + \ldots + X_n$. Let $\{\mathcal{F}_n\}$ represent the filtration generated by $X_1, \ldots, X_n$. Let $\{B_n\}$ be a sequence of random variables such that for each n, B_n is F_{n-1}-measurable. (Note that in our previous example, this was our bet on each result of the coin-flip.) Then the integral of B_n with respect to S_n is defined by

$$Z_n = \sum_{i=1}^{n} B_i(S_i - S_{i-1}) = \sum_{i=1}^{n} B_i \Delta S_i$$

Proposition 3.5. *The integral Z_n satisfies the following properties:*

- ***Linearity:*** *If J_n, K_n are predictable sequences and a, b are constants, then $aJ_n + bK_n$ is a predictable sequence and*

$$\sum_{i=1}^{n} (aJ_i + bK_i)\Delta S_i = a \sum_{i=1}^{n} J_i \Delta S_i + b \sum_{i=1}^{n} K_i \Delta S_i$$

- ***Martingale property:*** *The integral Z_n is a martingale with respect to $\{\mathcal{F}_n\}$. This was shown in Proposition 2.4.*
- ***Variance rule:***

$$Var\left[\sum_{i=1}^{n} B_i \Delta S_i\right] = \mathbb{E}\left[\left(\sum_{i=1}^{n} B_i \Delta S_i\right)^2\right] = \sigma^2 \sum_{i=1}^{n} \mathbb{E}[B_i^2]$$

Proof. Proof is on page 27 of [3] □

4. Brownian Motion

Now that we have shown properties of discrete random walks, a natural continuation is to ask: What properties would a continuous random walk have? We can sort of construct a continuous random walk by taking the limit of a discrete random walk time step and step interval to zero which we show below.

Let our time step be denoted as $\Delta t = \frac{1}{N}$ where N is large. In a simple random walk, we let $\Delta t = 1$. Let our step increment be denoted as Δx, such that $\mathbb{P}\{X_n =$

$-\Delta x\} = \mathbb{P}\{X_n = \Delta x\} = \frac{1}{2}$ for each n. Our martingale sequence will be $\{B_n\}$ where $B_n = X_1 + \ldots + X_n$. We can view the sequence at times $t = \Delta t, 2\Delta t, 3\Delta t \ldots$. If we want to view our process at time $t = 1$, then $B_1^{(N)} = X_1 + \ldots + X_N$. Suppose we want to pick Δx such that $Var(B_1^{(N)}) = 1$.

To do so, let $\{Y_n\}$ be a stochastic process such that for each n, $\mathbb{P}\{Y_n = -1\} = \mathbb{P}\{Y_n = 1\} = \frac{1}{2}$. Note this is the stochastic process generated by our coin-tossing results. We can see that $X_n = Y_n \Delta x$ for each n. Then

$$\begin{aligned} Var(B_1^{(N)}) &= Var(\Delta x(Y_1 + \ldots + Y_N)) \\ &= (\Delta x)^2 Var(Y_1 + \ldots + Y_N) \\ &= (\Delta x)^2 N \end{aligned}$$

Since we want $(\Delta x)^2 N = 1$, we should pick

$$\Delta x = \sqrt{\frac{1}{N}} = \sqrt{\Delta t}$$

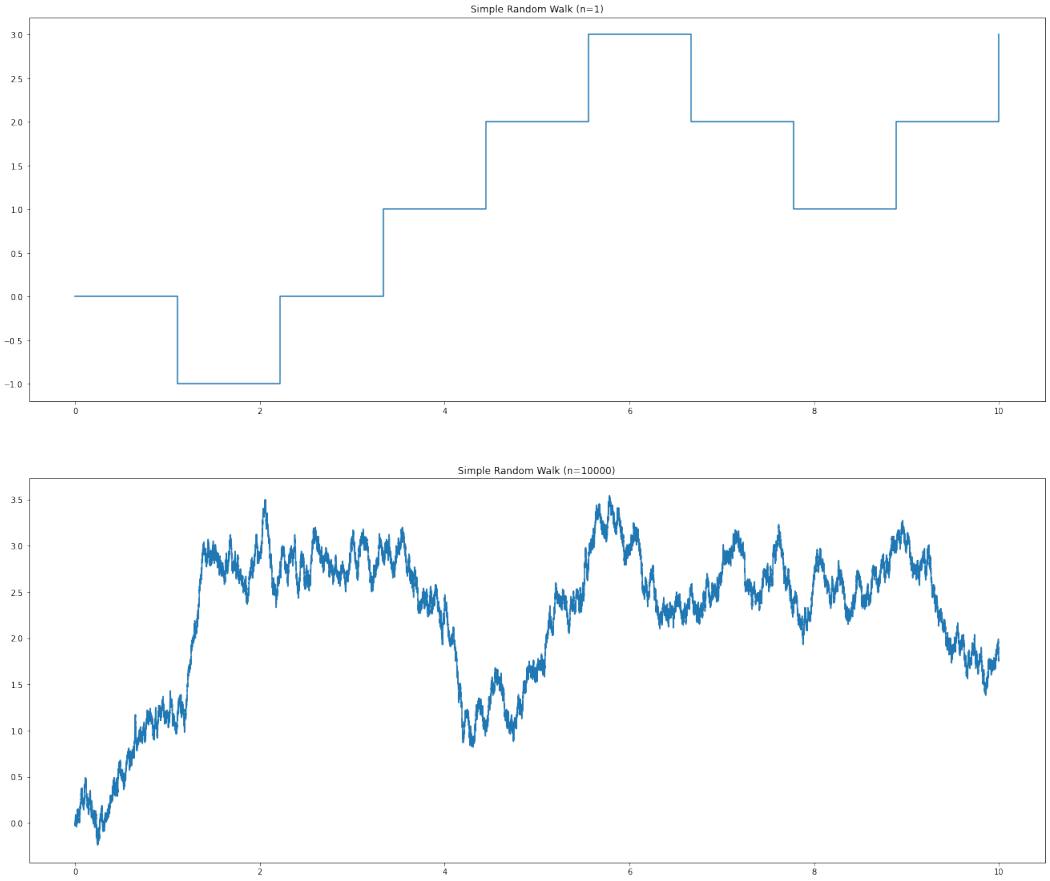

FIGURE (1) Constructing sample paths of a simple random walk $X_1 \sim N(0, 1)$ for $N = 1$ (top), and $N = 10000$ (bottom). Large values of N can be used to simulate a standard Brownian motion.

This is just one way in which we can construct Brownian motion using simple random processes. We need to add additional assumptions in what it means to have "continuous random motion."

Let B_t denote the value of a process at time t. If B_t is a random variable, then we have a stochastic process denoted by a random function $t \longmapsto B_t$. For $\{B_t\}$ to be a Brownian Motion, it has to satisfy these assumptions:

- If $s < t$, the distribution of $B_t - B_s$ equals the distribution of $B_{t-s} - B_0$. (Stationary Increments)
- If $s < t$, then $B_t - B_s$ is independent of $B_r - B_0$ for $r \leq s$. (Independent increments)
- The function $t \longmapsto B_t$ is continuous.

Let μ, σ^2 denote the mean and variance of B_1. For $s < t$, since B_t has stationary and independent increments, then

$$\mathbb{E}[B_t] = \mathbb{E}[B_s + B_t - B_s] = \mathbb{E}[B_s] + \mathbb{E}[B_{t-s}]$$

and

$$Var(B_t) = Var(B_s) + Var(B_{t-s})$$

By repeatedly taking $s = 1$, then it follows that $\mathbb{E}[B_t] = t\mu$ and $Var(B_t) = t\sigma^2$.

Definition 4.1. A stochastic process B_t is a Brownian motion with drift μ and variance σ^2 if

- $B_0 = 0$
- If $s < t$, then the distribution of $B_t - B_s$ is normal with mean $\mu(t-s)$ and variance $\sigma^2(t-s)$
- B_t has independent increments
- The function $t \longmapsto B_t$ is continuous

If B_t has parameters $\mu = 0$ and $\sigma^2 = 1$, then B_t is called a *standard Brownian motion*. Note that if B_t is a standard Brownian motion, and X_t is a Brownian motion with parameters μ, σ^2, then $X_t = \mu t + \sigma B_t$.

Notation 4.2. For the rest of the paper, we will typically denote standard Brownian motion using B_t or W_t.

Theorem 4.3. *Brownian motion is nowhere differentiable with probability one.*

Proof. The curious reader can refer to page 53 of [3] for the full proof. □

Intuitively, this should make sense given how Brownian motion is defined. At any point t, we know that Brownian motion is continuous, however we do not know what "direction" it may take since it is independent of any time $s < t$.

5. Quadratic Variation

In the following section, we will show the existence of quadratic variation for Brownian motion. With Riemann-Stieltjes integration, we assume the integrator to be monotone or have bounded variation. Since Brownian motion is nowhere differentiable, then it would turn out that it may not have finite variation: at every interval, it will make infinitely many small jumps. This poses a problem in defining an integral with respect to Brownian motion. However, this result on the quadratic variation of Brownian motion will allow us to show that a stochastic integral is well-defined.

Definition 5.1. If X_t is a process, then the quadratic variation at time t is defined as

$$\langle X \rangle_t = \lim_{n \to \infty} \sum_{j \leq tn} \left[X\left(\frac{j}{n}\right) - X\left(\frac{j-1}{n}\right) \right]^2$$

The following is a useful property of the quadratic variation of a standard Brownian motion. First we will prove the following lemma.

Lemma 5.2. *Let B_t be a standard Brownian motion. Then the distribution of the random variable*

$$X_{j,n} = \sqrt{n}\left(B\left(\frac{j}{n}\right) - B\left(\frac{j-1}{n}\right)\right)$$

is $X_{j,n} \sim N(0,1)$.

Proof. By the properties of Brownian motion, the distribution of

$$B\left(\frac{j}{n}\right) - B\left(\frac{j-1}{n}\right) \sim B\left(\frac{1}{n}\right) \sim N(0, \frac{1}{n})$$

Then $X_{j,n} \sim \sqrt{n} \cdot N(0, \frac{1}{n}) \sim N(0,1)$ □

Proposition 5.3. *Let B_t be a standard Brownian motion. Then*

$$\langle B \rangle_t = t$$

Proof. Let t be fixed. Let

$$Q_n = \sum_{j \leq tn} \left[B\left(\frac{j}{n}\right) - B\left(\frac{j-1}{n}\right) \right]^2$$

We can rewrite this as

$$\begin{aligned} \sum_{j \leq tn} \left[B\left(\frac{j}{n}\right) - B\left(\frac{j-1}{n}\right) \right]^2 &= \sum_{j \leq tn} \frac{1}{n} \left[\sqrt{n}\left(B\left(\frac{j}{n}\right) - B\left(\frac{j-1}{n}\right)\right) \right]^2 \\ &= \frac{1}{n} \sum_{j \leq tn} \left[\sqrt{n}\left(B\left(\frac{j}{n}\right) - B\left(\frac{j-1}{n}\right)\right) \right]^2 \end{aligned}$$

Let

$$Y_j = \left[\sqrt{n}\left(B\left(\frac{j}{n}\right) - B\left(\frac{j-1}{n}\right)\right) \right]^2$$

so that by substitution

$$Q_n = \frac{1}{n} \sum_{j \leq tn} Y_j$$

By Lemma 5.2, we know that $Y_j = Z^2$, where $Z \sim N(0,1)$ so

$$\mathbb{E}[Y_j] = \mathbb{E}[Z^2] = 1$$

Using integration by parts, we can find that $\mathbb{E}[Z^4] = 3$. Then

$$Var(Y_j) = Var(Z^2) = \mathbb{E}[Z^4] - 1 = 2$$

By the linearity of expectation,

$$\mathbb{E}[Q_n] = \frac{1}{n} \sum_{j \leq tn} \mathbb{E}[Y_j] = \frac{1}{n} \cdot \lfloor tn \rfloor$$

and

$$Var(Q_n) = \frac{1}{n^2} \sum_{j \leq tn} Var(Y_j) = \frac{1}{n^2} \cdot 2\lfloor tn \rfloor$$

Note that as $n \to \infty$, then $\mathbb{E}[Q_n] \to t$ and $Var(Q_n) \to \frac{2t}{n} \to 0$, so Q_t approaches a constant random variable with value t and variance zero. Thus

$$\langle B \rangle_t = \lim_{n \to \infty} Q_n = t$$

□

Proposition 5.4. *Let W_t be a Brownian motion with parameters μ, σ, i.e. $W_t = \mu t + \sigma B_t$ where B_t is a standard Brownian motion. Then*

$$\langle W \rangle_t = \sigma^2 t$$

Proof. Let t be fixed and let

$$Q_n = \sum_{j \leq tn} \left[W\left(\frac{j}{n}\right) - W\left(\frac{j-1}{n}\right) \right]^2$$

Using stationary and independent increments of Brownian motion, then

$$\begin{aligned}
Q_n &= \sum_{j \leq tn} \left[W\left(\frac{1}{n}\right) \right]^2 \\
&= \sum_{j \leq tn} \left[\frac{\mu}{n} + \sigma B\left(\frac{1}{n}\right) \right]^2 \\
&= \sum_{j \leq tn} \frac{\mu^2}{n^2} + 2 \sum_{j \leq tn} \sigma B\left(\frac{1}{n}\right) \frac{\mu}{n} + \sum_{j \leq tn} \sigma^2 B\left(\frac{1}{n}\right)^2 \\
&= \frac{\mu^2 \lfloor tn \rfloor}{n^2} + \frac{2\sigma\mu}{n} \sum_{j \leq tn} \left[B\left(\frac{j}{n}\right) - B\left(\frac{j-1}{n}\right) \right] + \sigma^2 \sum_{j \leq tn} \left[B\left(\frac{j}{n}\right) - B\left(\frac{j-1}{n}\right) \right]^2
\end{aligned}$$

As $n \to \infty$, then

$$\frac{\mu^2 \lfloor tn \rfloor}{n^2} \to 0$$

$$\frac{2\sigma\mu}{n} \sum_{j \leq tn} \left[B\left(\frac{j}{n}\right) - B\left(\frac{j-1}{n}\right) \right] \to \frac{2\sigma\mu}{n} [B_t] \to 0$$

and using Proposition 5.3,

$$\sigma^2 \sum_{j \leq tn} \left[B\left(\frac{j}{n}\right) - B\left(\frac{j-1}{n}\right) \right]^2 \to \sigma^2 \langle B \rangle_t \to \sigma^2 t$$

Therefore

$$\langle W \rangle_t = \lim_{n \to \infty} Q_n = \sigma^2 t$$

□

6. Stochastic Calculus

The goal of stochastic calculus is to make sense of the expression

$$dX_t = m(t, X_t)dt + \sigma(t, X_t)dB_t \tag{6.1}$$

where B_t is a standard Brownian motion. In differential and integral calculus, you would just take the integral of each side and solve it analytically or numerically. However, since Brownian motion is nowhere differentiable, we cannot define a derivative easily. Instead we focus on the integral. A solution X_t to the differential equation (6.1) must satisfy

$$X_t = X_0 + \int_0^t m(s, X_s)\ ds + \int_0^t \sigma(s, X_s)\ dB_s$$

The stochastic integral (also called the Itô integral) gives meaning to the second expression

$$\int_0^t \sigma(s, X_s)\ dB_s$$

or more generally the expression

$$Z_t = \int_0^t A_s\ dB_s$$

Analogous to the discrete stochastic integral, we can think of A_s as the "bet" on whether B_s goes up or down. Thus we want to restrict our betting strategy such that it cannot see into the future.

Similar to in Riemann integration, we first define the Itô integral on simple discrete processes, then take the limit to extend the integral to continuous processes.

Definition 6.2. A process A_t is a *simple process* if there exist times

$$0 = t_0 < t_1 < \ldots < t_n < \infty$$

and random variables $Y_j, j = 0, 1, \ldots, n$ that are $\mathcal{F}_{t_j}$-measurable such that

$$A_t = Y_j, \quad t_j \leq t < t_{j+1}$$

This is the stochastic analogy to a step function. It can be thought of as a discrete betting strategy in which one can only change their bet A_t at finite time intervals $t_0, t_1, \ldots$.

We now define the stochastic integral for simple processes

Definition 6.3. If A_t is a simple process, then we define

$$Z_t = \int_0^t A_s\ dB_s$$

by

$$Z_{t_j} = \sum_{i=0}^{j-1} Y_i[B_{t_{i+1}} - B_{t_i}]$$

and extending the integral to all t,

$$Z_t = Z_{t_j} + Y_j[B_t - B_{t_j}], \quad t_j \leq t < t_{j+1}$$

We also define

$$\int_s^r A_t\ dB_t = Z_r - Z_s$$

Proposition 6.4. *If B_t is a standard Brownian motion with respect to filtration $\{\mathcal{F}_t\}$ and A_t, C_t are simple processes, then the following properties hold:*

- ***Linearity:*** *If a, b are constants, then $aA_t + bC_t$ is a simple process and*

$$\int_0^t (aA_s + bC_s)\ dB_s = a\int_0^t A_s\ dB_s + b\int_0^t C_s\ dB_s$$

- ***Martingale property:*** *The process*

$$Z_t = \int_0^t A_s\ dB_s$$

  is a martingale with respect to $\{\mathcal{F}_t\}$.
- ***Variance rule:*** *Z_t is square integrable and*

$$Var(Z_t) = \mathbb{E}\left[\left(\int_0^t A_s\ dB_s\right)^2\right] = \int_0^t \mathbb{E}[A_s^2]\ ds$$

- *With probability one, the function $t \mapsto Z_t$ is continuous.*

Proof. Linearity follows from the definition of the integral and summation.

To show Z_t is a martingale, we want to show

$$\mathbb{E}[Z_t \mid \mathcal{F}_s] = Z_s, \quad \text{when } t \geq s$$

Suppose we have j, k such that $t_j \leq s < t_{j+1}$ and $t_k \leq t < t_{k+1}$ and $j \leq k$. By the definition of the stochastic integral, we have

$$Z_s = \sum_{i=0}^{j-1} Y_i[B_{i+1} - B_i] + Y_j[B_s - B_{t_j}]$$

$$Z_t = Z_s Y_j[B_{t_{j+1}} - B_s] + \sum_{i=j+1}^{k-1} Y_i[B_{i+1} - B_i] + Y_k[B_t - B_{t_k}]$$

By linearity of conditional expectation, then

$$\begin{aligned}
\mathbb{E}[Z_t \mid \mathcal{F}_s] = {} & \mathbb{E}[Z_s \mid \mathcal{F}_s] && (6.5)\\
& + \mathbb{E}[Y_j[B_{t_{j+1}} - B_s] \mid \mathcal{F}_s] && (6.6)\\
& + \sum_{i=j+1}^{k-1} \mathbb{E}[Y_i[B_{i+1} - B_i] \mid \mathcal{F}_s] && (6.7)\\
& + \mathbb{E}[Y_k[B_t - B_{t_k}] \mid \mathcal{F}_s] && (6.8)
\end{aligned}$$

For (6.5), since Z_s is $\mathcal{F}_s-$measurable, then

$$\mathbb{E}[Z_s \mid \mathcal{F}_s] = Z_s$$

For (6.6), (6.7), and (6.8), note that each term is of the form $\mathbb{E}[Y_a[B_b - B_{t_a}] \mid \mathcal{F}_s]$ where $s \leq t_a \leq b$. It follows that $B_b - B_{t_a}$ is independent of $\mathcal{F}_{t_a}$ so

$$\mathbb{E}[B_b - B_{t_a} \mid \mathcal{F}_{t_a}] = \mathbb{E}[B_b - B_{t_a}] = 0$$

Also Y_a is $\mathcal{F}_{t_a}$-measurable so it acts as a constant. Then by the tower property,

$$\begin{aligned}\mathbb{E}[Y_a(B_b - B_{t_a}) \mid \mathcal{F}_s] &= \mathbb{E}[\mathbb{E}[Y_a(B_b - B_{t_a}) \mid \mathcal{F}_{t_a}] \mid \mathcal{F}_s] \\ &= \mathbb{E}[Y_a \mathbb{E}[B_b - B_{t_a} \mid \mathcal{F}_{t_a}] \mid \mathcal{F}_s] \\ &= \mathbb{E}[Y_a \cdot 0 \mid \mathcal{F}_s] \\ &= 0\end{aligned}$$

Each of the terms (6.6), (6.7), and (6.8) equals zero so then

$$\mathbb{E}[Z_t \mid \mathcal{F}_s] = Z_s$$

We will now prove the variance rule. It suffices to show it for $t = t_n$. For simplicity's sake, let $\Delta B_i = B_{t_i} - B_{t_{i+1}}$. We want to show

$$\mathbb{E}\left[\left(\int_0^t A_s \, dB_s\right)^2\right] = \int_0^t \mathbb{E}[A_s^2] \, ds$$

By the definition of the stochastic integral,

$$\begin{aligned}Z_t^2 &= \left(\int_0^t A_s \, dB_s\right)^2 \\ &= \left(\sum_{i=0}^{n-1} Y_i \Delta B_i\right)\left(\sum_{j=0}^{n-1} Y_i \Delta B_j\right) \\ &= \sum_{i=0}^{n-1} Y_i^2 (\Delta B_i)^2 + \sum_{i<j\colon n-1} Y_i Y_j \Delta B_i \Delta B_j\end{aligned}$$

Taking the expected value, we get

$$\mathbb{E}[Z_t^2] = \sum_{i=0}^{n-1} \mathbb{E}[Y_i^2 (\Delta B_i)^2] + 2 \sum_{i<j\colon n-1} \mathbb{E}[Y_i Y_j \Delta B_i \Delta B_j]$$

For each term $\mathbb{E}[Y_i Y_j \Delta B_i \Delta B_j]$ where $i < j$, we know that Y_i, Y_j, and ΔB_i are $\mathcal{F}_{t_j}$-measurable. We also know that ΔB_j is independent of $\mathcal{F}_{t_j}$ and $\mathbb{E}[\Delta B_j] = 0$. Then by the properties of conditional expectation

$$\begin{aligned}\mathbb{E}[Y_i Y_j \Delta B_i \Delta B_j] &= \mathbb{E}[Y_i Y_j \Delta B_i \Delta B_j \mid \mathcal{F}_{t_j}] \\ &= Y_i Y_j \Delta B_i \mathbb{E}[\Delta B_j \mid \mathcal{F}_{t_j}] \\ &= Y_i Y_j \Delta B_i \mathbb{E}[\Delta B_j] \\ &= Y_i Y_j \Delta B_i \cdot 0 \\ &= 0\end{aligned}$$

Now we are left with the term

$$\sum_{i=0}^{n-1} \mathbb{E}[Y_i^2 (\Delta B_i)^2]$$

For each term $\mathbb{E}[Y_i^2(\Delta B_i)^2]$, Y_i is $\mathcal{F}_{t_i}$-measurable and $(\Delta B_i)^2$ is independent of $\mathcal{F}_{t_i}$. Then

$$\begin{aligned}\mathbb{E}[Y_i^2(\Delta B_i)^2] &= \mathbb{E}[\mathbb{E}[Y_i^2(\Delta B_i)^2 \mid \mathcal{F}_{t_i}]] \\ &= \mathbb{E}[Y_i^2\mathbb{E}[(\Delta B_i)^2 \mid \mathcal{F}_{t_i}] \\ &= \mathbb{E}[Y_i^2\mathbb{E}[(\Delta B_i)^2]]\end{aligned}$$

Note that

$$\mathbb{E}[(\Delta B_i)^2] = Var(B_{t_{i+1}} - B_{t_i}) = t_{i+1} - t_i$$

Then

$$\begin{aligned}\mathbb{E}[Y_i^2(\Delta B_i)^2] &= \mathbb{E}[Y_i^2(t_{i+1} - t_i)] \\ &= \mathbb{E}[Y_i^2](t_{i+1} - t_i)\end{aligned}$$

Going back to the RHS of the statement:

$$\int_0^t \mathbb{E}[A_s^2]\, ds$$

we know that A_t^2 is a process such that

$$A_s^2 = Y_i^2, \quad \text{when } t_i \le s \le t_{i+1}$$

Then $t \mapsto \mathbb{E}[A_s^2]$ is a step function that takes values $\mathbb{E}[Y_i^2]$ when $t_i \le s \le t_{i+1}$. Therefore we have

$$\mathbb{E}[Z_t^2] = \sum_{i=0}^{n-1} \mathbb{E}[Y_i^2](t_{i+1} - t_i) = \int_0^t \mathbb{E}[A_s^2]\, ds$$

□

Our next step is to extend the integral to continuous processes. To do so, we use the following lemma.

Lemma 6.9. *Let A_t be a stochastic process with continuous paths such that A_t is $\mathcal{F}$-adapted (or analogously, $\mathcal{F}_t$-measurable). Suppose there exists a $C < \infty$ such that $|A_t| \le C$ for all t. Then there exists a sequence of simple processes $\{A_t^{(n)}\}$ such that*

$$\lim_{n\to\infty} \int_0^t \mathbb{E}[|A_t - A_t^{(n)}|^2] = 0$$

Also for all n, t, $|A_t^{(n)}| \le C$

Proof. See 3.2.2 in [3] for the proof □

We use this result to define the stochastic integral for bounded, continuous processes.

Definition 6.10. Let A_s be a bounded, continuous processes. To define

$$\int_0^t A_s\, dB_s$$

find a sequence of simple processes $\{A_t^{(n)}\}$ that satisfy Lemma 6.9. Then for each t, there is a square-integrable random variable Z_t such that

$$Z_t = \lim_{n\to\infty} \int_0^t A_s^{(n)}\ dB_s$$

We now define

$$\int_0^t A_s\ dB_s = Z_t$$

Definition 6.11. Let A_s be adapted and continuous, but not necessarily bounded. Then we can define

$$\int_0^t A_s\ dB_s$$

as follows. We want to create increasing bounds for each process in the sequence which allows us to approximate unbounded values. To do so, for each n, let $T_n = \min\{t\colon |A_t| = n\}$ be the minimum stopping time such that $|A_t|$ reaches n. Let $A_s^{(n)} = A_{\min\{s,T_n\}}$. We create a sequence of continuous, processes $A_s^{(n)}$ that are each bounded by n, so each $A_s^{(n)}$ is a bounded, continuous process. We now define

$$\int_0^t A_s\ dB_s = \lim_{n\to\infty} \int_0^t A_s^{(n)}\ dB_s$$

Proposition 6.12. *Let B_t be a standard Brownian motion with respect to filtration $\{\mathcal{F}_t\}$ and let A_t and C_t be adapted continuous processes. Then the following conditions hold:*

- ***Linearity:*** *If a, b are constants then*

$$\int_0^t (aA_s + bC_t)\ dB_s = a\int_0^t A_s\ dB_s + b\int_0^t C_t\ dB_s$$

- ***Variance rule:*** *Let*

$$Z_t = \int_0^t A_s\ dB_s$$

Then

$$Var(Z_t) = \mathbb{E}[Z_t^2] = \int_0^t \mathbb{E}[A_s^2]\ dB_s$$

- ***Continuous paths:*** *With probability one, the function $t \mapsto Z_t$ is continuous.*

7. Itô's Formula

Ito's formula is the fundamental theorem of stochastic calculus. We will provide a derivation of Ito's formula, but leave out some technical details relating to taking the limit of sequences of random variables. For the full derivation, see pg 102 of [3]. Note: a function f is C^k in x if it has k continuous derivatives, i.e. $f_x, f_{xx}, \ldots, f_{x^{(k)}}$ exist and are continuous.

Theorem 7.1. ***Ito's Formula*** *Let $f(t,x)$ be a function that is C^1 in t and C^2 in x. If B_t is a standard Brownian motion, then*

$$f(t,B_t) = f(0,B_0) + \int_0^t \partial_x f(s,B_s)\ dB_s + \int_0^t \left[\partial_t f(s,B_s) + \frac{1}{2}\partial_{xx} f(s,B_s)\right]\ ds$$

In differential form,

$$df(t, B_t) = \partial_x f(t, B_t)\ dB_t + \left[\partial_t f(t, B_t) + \frac{1}{2}\partial_{xx} f(t, B_t)\right]\ dt$$

Proof. We first write out the telescoping series for $f(t, B_t)$

$$f(t, B_t) = f(0, B_0) + \sum_{i=1}^{tn} \left[f(\tfrac{i}{n}, B_{\frac{i}{n}}) - f(\tfrac{i-1}{n}, B_{\frac{i-1}{n}})\right] \tag{7.2}$$

Consider the second order Taylor approximation of f

$$f(t + \Delta t, x + \Delta x) - f(t, x) =$$
$$\partial_t f(t, x)\Delta t + o(\Delta t) + \partial_x f(t, x)\Delta x + \frac{1}{2}\partial_{xx} f(t, x)(\Delta x)^2 + o((\Delta x)^2)$$

where $o(x)/x \to 0$ as $x \to 0$. We apply the Taylor approximation to each term $f\left(\frac{i}{n}, B_{i/n}\right) - f\left(\frac{i-1}{n}, B_{(i-1)/n}\right)$ in (7.2) so we get

$$\begin{aligned} f(\tfrac{i}{n}, B_{\frac{i}{n}}) - f(\tfrac{i-1}{n}, B_{\frac{i-1}{n}}) = {} & \partial_t f(\tfrac{i-1}{n}, B_{\frac{i-1}{n}})(\tfrac{1}{n}) \\ & + \partial_x f(\tfrac{i-1}{n}, B_{\frac{i-1}{n}})\Delta B_{i/n} \\ & + \frac{1}{2}\partial_{xx} f(\tfrac{i-1}{n}, B_{\frac{i-1}{n}})(\Delta B_{i/n})^2 \\ & + o(\tfrac{1}{n}) + o((\Delta B_{i/n})^2) \end{aligned} \tag{7.3}$$

where $\Delta B_{i/n} = B_{\frac{i}{n}} - B_{\frac{i-1}{n}}$.
Subbing (7.3) into (7.2), and taking the limit as $n \to \infty$, we get

$$f(t, B_t) = f(0, B_0) \tag{7.4}$$
$$+ \lim_{n\to\infty} \sum_{i=1}^{tn} \partial_t f(\tfrac{i-1}{n}, B_{\frac{i-1}{n}})(\tfrac{1}{n}) \tag{7.5}$$
$$+ \lim_{n\to\infty} \sum_{i=1}^{tn} \partial_x f(\tfrac{i-1}{n}, B_{\frac{i-1}{n}})\Delta B_{i/n} \tag{7.6}$$
$$+ \frac{1}{2}\lim_{n\to\infty} \sum_{i=1}^{tn} \partial_{xx} f(\tfrac{i-1}{n}, B_{\frac{i-1}{n}})(\Delta B_{i/n})^2 \tag{7.7}$$
$$+ \lim_{n\to\infty} \sum_{i=1}^{tn} o(\tfrac{1}{n}) + o((\Delta B_{i/n})^2) \tag{7.8}$$

The term in (7.5) is the approximation of the Riemann integral for $\partial_t f(t, B_t)$ so

$$\lim_{n\to\infty} \sum_{i=1}^{tn} \partial_t f(\tfrac{i-1}{n}, B_{\frac{i-1}{n}})(\tfrac{1}{n}) = \int_0^t \partial_t f(s, B_s)\ ds$$

The term in (7.6) is the simple process approximation of the stochastic integral for $\partial_x f(t, B_t)$ so

$$\lim_{n\to\infty} \sum_{i=1}^{tn} \partial_x f(\tfrac{i-1}{n}, B_{\frac{i-1}{n}})\Delta B_{i/n} = \int_0^t \partial_x f(s, B_s)\ dB_s$$

For the term in (7.7), we need to deal with the $(\Delta B_{i/n})^2$ somehow. We do this by approximating $\partial_{xx} f(t, B_t)$ using step functions which allows us to treat it as a

constant on each interval. For sake of clarity, let $g(t) = \partial_{xx} f(t, B_t)$. The function g is continuous, so for every $\epsilon > 0$, there exists a step function g_ϵ, such that

$$|g(t) - g_\epsilon(t)| < \epsilon$$

Fix ϵ and consider the step function g_ϵ such that

$$g_\epsilon(s) = g(t_i), \quad \text{when } t_j \leq s < t_{j+1}$$

For any interval $[t_j, t_{j+1}]$, since $g_\epsilon(s)$ is constant, we can use our result on the quadratic variation of B_t in Proposition 4.3 to get the limit of the sum on that interval:

$$\begin{aligned}
\lim_{n\to\infty} \sum_{i=t_j+1}^{t_{j+1}n} g_\epsilon(\tfrac{i-1}{n})(\Delta B_{i/n})^2 &= \lim_{n\to\infty} g(t_j) \sum_{i=t_j+1}^{t_{j+1}n} (B_{i/n} - B_{(i-1)/n})^2 \\
&= g(t_j)(\langle B\rangle_{t_j} - \langle B\rangle_{t_{j-1}}) \\
&= g(t_j)(t_j - t_{j-1})
\end{aligned}$$

We can add each of the intervals and take the limit as $\epsilon \to 0$ to get

$$\begin{aligned}
\lim_{n\to\infty} \sum_{i=1}^{tn} g_\epsilon(\tfrac{i-1}{n})(\Delta B_{i/n})^2 &= \sum_{j=1}^{m} g(t_j)(t_j - t_{j-1}) \\
&= \int_0^t g_\epsilon(s)\, ds \\
&\to \int_0^t g(s)\, ds
\end{aligned}$$

Now we want to show that the two sums converge when $n \to \infty$. It follows that

$$\begin{aligned}
\left| \sum_{i=1}^{tn} g(\tfrac{i-1}{n})(\Delta B_{i/n})^2 - \sum_{i=1}^{tn} g_\epsilon(\tfrac{i-1}{n})(\Delta B_{i/n})^2 \right| &= \left| \sum_{i=1}^{tn} [g(\tfrac{i-1}{n}) - g_\epsilon(\tfrac{i-1}{n})](\Delta B_{i/n})^2 \right| \\
&\leq \epsilon \left| \sum_{i=1}^{tn} (\Delta B_{i/n})^2 \right| \\
&\to \epsilon \langle B\rangle_t \\
&= \epsilon t
\end{aligned}$$

Therefore combining the last two results, we have

$$\begin{aligned}
\frac{1}{2} \lim_{n\to\infty} \sum_{i=1}^{tn} g(\tfrac{i-1}{n})(\Delta B_{i/n})^2 &= \frac{1}{2} \lim_{\epsilon\to 0} \int_0^t g_\epsilon(s)\, ds \\
&= \frac{1}{2} \int_0^t g(s)\, ds \\
&= \frac{1}{2} \int_0^t \partial_{xx} f(s)\, ds
\end{aligned}$$

Lastly for the term in (7.8), as $n \to \infty$,

$$\sum_{i=1}^{tn} o(\tfrac{1}{n}) = o(\tfrac{1}{n}) tn = \frac{o(\frac{1}{n})}{\frac{1}{tn}} \to 0$$

and by normality of Brownian increments, $(\Delta B_{i/n})^2 \sim \frac{1}{n}$, then

$$\sum_{i=1}^{tn} o((\Delta B_{i/n})^2) \to 0$$

Combining all the results together, then

$$f(t, B_t) = f(0, B_0) + \int_0^t \partial_x f(s, B_s)\ dB_s + \int_0^t \left[\partial_t f(s, B_s) + \frac{1}{2}\partial_{xx} f(s, B_s)\right]\ ds$$

□

We will now use Ito's Formula to calculate the solutions to some stochastic differential equations (SDE). Formally, we can use the following convention to assist in our calculations

$$(dt)^2 = 0 \qquad (dB_t)(dt) = 0 \qquad (dB_t)^2 = dt$$

The reason for this is that in doing formal calculations, we will drop terms that have order less that dt.

Example 7.9. Let X_t be a process that satisfies

$$dX_t = \mu_t X_t dt + \sigma_t X_t dB_t \tag{7.10}$$

where B_t is a standard Brownian motion.

We want to solve for X_t. Let $f(t, B_t) = \log(X_t)$. Then

$$\begin{aligned}
\partial_t(t, B_t) &= 0, \\
\partial_x f(t, B_t) &= \frac{dX_t}{X_t} \\
&= \mu_t dt + \sigma_t dB_t, \\
\partial_{xx} f(t, B_t) &= -\frac{(dX_t)^2}{X^2} \\
&= -\frac{1}{X_t^2}\left(\mu_t^2 X_t^2 (dt)^2 + 2\mu_t\sigma_t X_t^2 (dB_t)(dt) + \sigma_t^2 X_t^2 (dB_t)^2\right) \\
&= -\sigma_t^2 dt
\end{aligned}$$

Applying Ito's Formula (Theorem 6.1) to $f(t, B_t)$, then

$$\begin{aligned}
\log(X_t) &= f(0, B_0) + \int_0^t \partial_x f(s, B_s)\ dB_s + \int_0^t \left[\partial_t f(s, B_s) + \frac{1}{2}\partial_{xx} f(s, B_s)\right]\ ds \\
&= \log(X_0) + \int_0^t (\mu_t d_t + \sigma_t dB_t) - \int_0^t \frac{\sigma_t^2}{2} dt
\end{aligned}$$

Rearranging the terms and taking the exponential, we get

$$X_t = X_0 \exp\left\{\int_0^t \left(\mu_s - \frac{\sigma_s^2}{2}\right)\ ds + \int_0^t \sigma_s\ dB_s\right\}$$

If $\mu_t = \mu$ and $\sigma_t = \sigma$ are constant, then we have the following expression

$$X_t = X_0 \exp\left\{\left(\mu - \frac{\sigma^2}{2}\right) t + \sigma B_t\right\} \tag{7.11}$$

If a process X_t satisfies the SDE (7.10), then it is called a *geometric Brownian motion* with parameters (μ, σ). This is often a natural model for stock prices, since it measures the change in percentage of the stock prices rather than the absolute price.

Definition 7.12. A stochastic process X_t is called an *Ito process* if there exist continuous, $\mathcal{F}_t$-adapted processes A_t, C_t such that

$$dX_t = A_t dt + C_t dB_t$$

where B_t is a standard Brownian motion

Theorem 7.13. ***Ito's Formula for Ito Processes*** *Let $f(t,x)$ be a function that is C^1 in t and C^2 in x. Let X_t be an Ito process such that*

$$dX_t = A_t dt + C_t dB_t$$

where B_t is a standard Brownian motion. Let H_t be a $\mathcal{F}_t$-adapted process. We define

$$\int_0^t H_s dX_s := \int_0^t H_s(A_s ds + C_s dB_s) = \int_0^t H_s A_s \, ds + \int_0^t H_s C_s dB_s$$

Then

$$\begin{aligned} df(t, X_t) &= \partial_t f(t, X_t)dt + \partial_x f(t, X_t)dX_t + \frac{1}{2}\partial_{xx} f(t, X_t)(dX_t)^2 \\ &= \left[\partial_t f(t, X_t) + A_t \partial_x f(t, X_t) + \frac{1}{2}C_t^2 \partial_{xx} f(t, X_t)\right] dt + C_t \partial_x f(t, X_t) dB_t \end{aligned}$$

or in integral form,

$$\begin{aligned} f(t, X_t) = f(0, X_0) &+ \int_0^t \left[\partial_t f(s, X_s) + A_s \partial_x f(s, X_s) + \frac{1}{2}C_s^2 \partial_{xx} f(s, X_s)\right] \, ds \\ &+ \int_0^t C_s \partial_x f(s, X_s) \, dB_s \end{aligned}$$

Proof. The proof for the first equality is similar to Ito's Formula. By expanding the dX_t and $(dX_t)^2$ terms, we get the second equality. □

Using Ito's formula, we now have a method to solve stochastic differential equations analytically.

8. Numerical Methods and Simulations

Often times, a stochastic differential equation cannot be solved analytically. In these cases, mathematicians will turn to numerical methods in order to simulate a solution. We present two methods for numerical analysis of SDEs and also introduce the usage of Monte Carlo simulations.

In ordinary calculus, usually one wants to find values of a function given the rate of change. For example, let $f(t)$ denote the position of a particle at time t. If we are given the function for the rate of change:

$$df(t) = C(t, f(t))dt$$

then given an initial condition $f(0) = x_0$, we can approximate the function for some small increment t using *Euler's Method*:

$$f(t + \Delta t) = f(t) + C(t, f(t))\Delta t$$

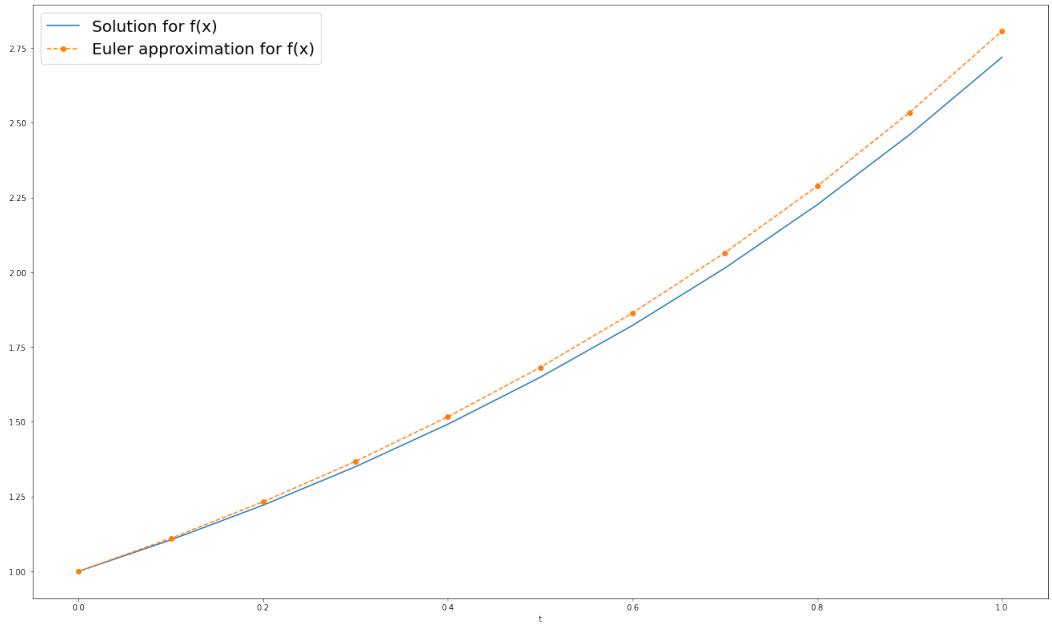

FIGURE (2) Graphical representation of Euler's method (orange) with step size $\Delta t = 0.1$, compared to the exact solution (blue) for the differential equation $df(t) = e^t dt$

Euler's method is often the simplest way to approximate a differential equation. By taking smaller step sizes for Δt, we can reach closer and closer approximations to the exact value of the function.

In a similar vein, the simplest method to approximate a stochastic differential equation is the *stochastic Euler method* otherwise know as the *Euler-Maruyama Method*. Given a stochastic differential equation

$$dX_t = m(t, X_t)dt + \sigma(t, X_t)dB_t$$

where B_t is a standard Brownian motion, and an initial condition, $X(0) = x_0$, we can approximate the $X(t)$ using the formula:

$$X(t + \Delta t) = X(t) + m(t, X(t))\Delta t + \sigma(t, X(t))N_k\sqrt{\Delta t} \tag{8.1}$$

where $N_k \sim N(0, 1)$.

We will use the Euler-Maruyama method to approximate the solution to a geometric SDE. Let X_t be a process that satisfies

$$dX(t) = \mu X(t)dt + \sigma X(t)dB_t, \quad X(0) = x_0 \tag{8.2}$$

and μ, σ are constants. According to (7.1), we have $m(t, X(t)) = \mu X(t)$ and $\sigma(t, X(t)) = \sigma X(t)$. As shown in Example 6.9, the solution to (7.2) is given by

$$X(t) = x_0 \exp\left\{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma B_t\right\}$$

For this specific simulation, we used the parameters $\mu = 2$, $\sigma = 1$, and $X(0) = 1$. For the discretized Brownian motion increments, we used a step size of $dt = 2^{-15}$. To approximate the solution using Euler-Maruyama, we used step sizes of $\Delta t =$

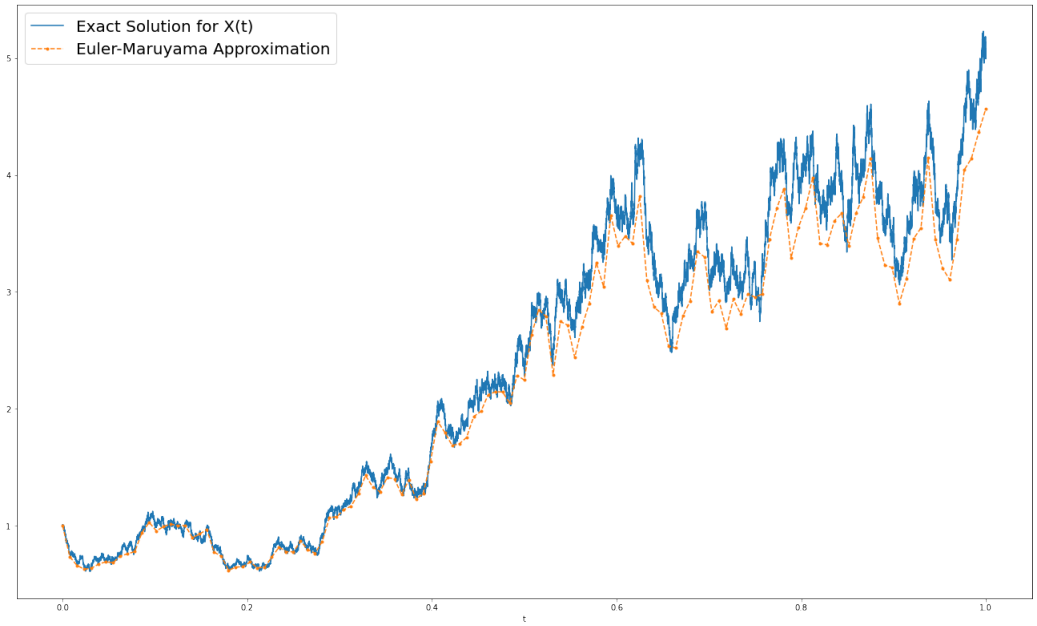

FIGURE (3) Euler-Maruyama approximation with $\Delta t = 2^{-7}$ (orange) and the exact solution (blue) for geometric Brownian motion

2^{-7}, $\Delta t = 2^{-9}$, and $\Delta t = 2^{-11}$ and produced an approximation $X_N(t)$. Our errors $|X_N(t) - X(t)|$ at $t = T = 1$ were computed to be 0.488, 0.186, and 0.075 respectively.

As we can see, when we reduce our step size Δt, the Euler-Maruyama approximation seems to converge to the exact solution. This brings into a new question: how fast does the Euler-Maruyama method converge to its solution? To answer this, we bring in two important new definitions.

Definition 8.3. Let $X(t)$ be a process. Let X_n denote the approximation of the value of $X(T)$ where $T = n\Delta t$. The *weak error* $e_{\Delta t}^{\text{weak}}$ with respect to some function f is defined by

$$e_{\Delta t}^{\text{weak}} = \sup_{0 \le t_n \le T} |\mathbb{E}[f(X_n)] - \mathbb{E}[f(X(t_n))]|$$

We say that X_n *converges weakly* to $X(t)$ with order $\rho > 0$ if there exists a constant $C > 0$ independent of Δt such that

$$e_{\Delta t}^{\text{weak}} \le C\Delta t^{\rho} \tag{8.4}$$

for all sufficiently small Δt.

Definition 8.5. We define *strong error* $e_{\Delta t}^{\text{strong}}$ as

$$e_{\Delta t}^{\text{strong}} = \sup_{0 \le t_n \le T} \mathbb{E}[|X_n - X(t_n)|$$

We say X_n *converges strongly* to $X(t)$ with order $\rho > 0$ if there exists a constant $C > 0$ independent of Δt such that

$$e_{\Delta t}^{\text{strong}} \le C\Delta t^{\rho} \tag{8.6}$$

for all sufficiently small Δt.

We can view weak convergence as the decay in the rate of the "error of the means" while strong convergence is the decay in the rate of the "mean of the errors." Strong convergence implies weak convergence, but the opposite is not necessarily true.

If a numerical scheme converges with order ρ, then by decreasing our step size by a factor of k reduces our error by a factor of k^ρ. For example, an order of $\rho = 1$ means that decreasing our step size by a factor of 100 also decreases our error by a factor of 100. Meanwhile, an order of $\rho = \frac{1}{2}$ means that decreasing our step size by a factor of 100 decreases our error only by a factor of 10.

We claim that the Euler-Maruyama method converges strongly with order $\rho = \frac{1}{2}$. We will test this claim using a numerical experiment for the value of $X(T)$ at $T = 1$. Since (7.6) should hold for all $T \in [0, 1]$ then it will hold for $T = 1$. Applying (7.6) for $\rho = \frac{1}{2}$, then the following inequality should be true:

$$e_{\Delta t}^{\text{strong}} \leq C\Delta t^{\frac{1}{2}} \tag{8.7}$$

We conducted $M = 1000$ sample paths to calculate the mean of the error for values of $\Delta t = 2^{-x}$ for $5 \leq x \leq 14$. If (7.7) holds, then taking the log of both sides, we get

$$\log(e_{\Delta t}^{\text{strong}}) \approx \log(C) + \frac{1}{2}\log(\Delta t)$$

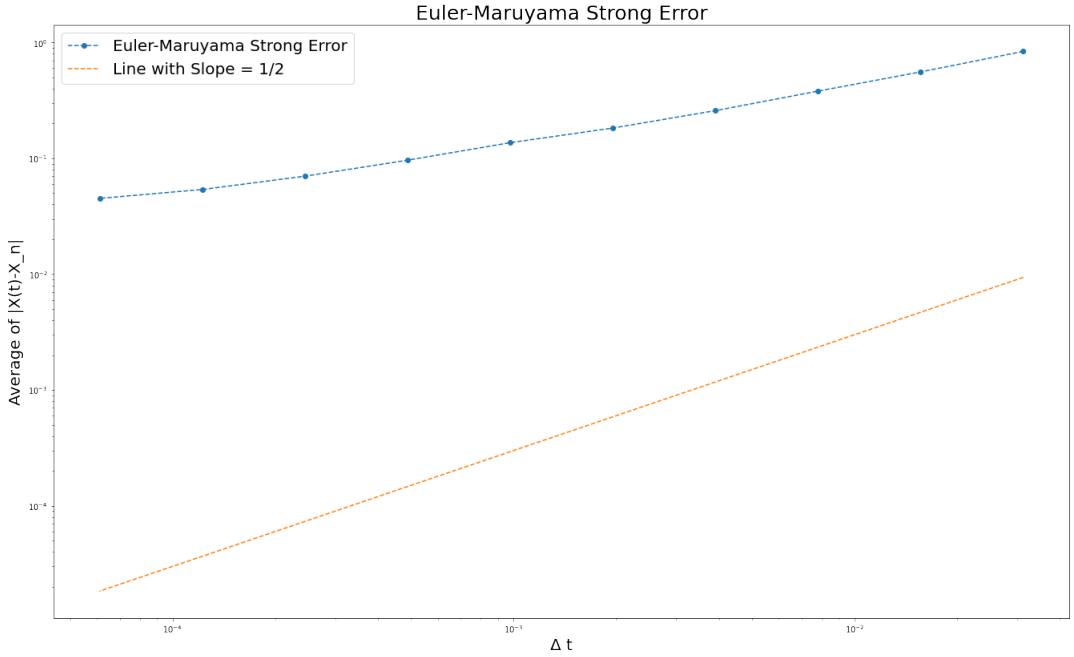

FIGURE (4) Euler-Maruyama strong error (blue) $|X(t) - X_n|$ for values of Δt with $\log(X) = \frac{1}{2}\log(X)$ (orange) for comparison.

We graph the relation between $e_{\Delta t}^{\text{strong}}$ and Δt on a log-log scale (Figure 4). Comparing the strong error line (blue) to the orange line, the slopes seem to match visually. To further confirm this, we compute the least-squares fit for $\log(C)$ and ρ and obtain the value of $\rho = 0.475$ with a least squares residual of 0.0649. Thus, our results are consistent with strong order of convergence $\rho = \frac{1}{2}$.

We also claim that the Euler-Maruyama numerical scheme converges weakly with respect to the function $f(X) = X$ with order $\rho = 1$, i.e.

$$e_{\Delta t}^{\text{weak}} \leq C\Delta t \tag{8.8}$$

To test this hypothesis, first we want to find $\mathbb{E}[X(t)]$. Recall that

$$dX_t = \mu X_t dt + \sigma X_t dB_t$$

represents

$$X_t = X_0 + \mu \int_0^t X_s \, ds + \sigma \int_0^t X_s \, dB_s$$

Taking the expectation of both sides, we get

$$\mathbb{E}[X_t] = \mathbb{E}[X_0] + \mu\mathbb{E}\left[\int_0^t X_s ds\right] + \sigma\mathbb{E}\left[\int_0^t X_s \, dB_s\right] = \mathbb{E}[X_0] + \mu \int_0^t \mathbb{E}[X_s] \, ds + 0$$

This implies

$$d\mathbb{E}[X_t] = \mu\mathbb{E}[X_t]dt, \quad \mathbb{E}[X_0] = x_0$$

which has a known solution

$$\mathbb{E}[X_t] = x_0 e^{\mu t}$$

Now we conduct a numerical experiment over $M = 10^6$ sample paths for values of $\Delta t = 2^{-x}$, $5 \leq x \leq 10$. We calculate the error as

$$|\mathbb{E}[X_n] - \mathbb{E}[X(T)]|, \quad n\Delta t = T$$

for $T = 1$.

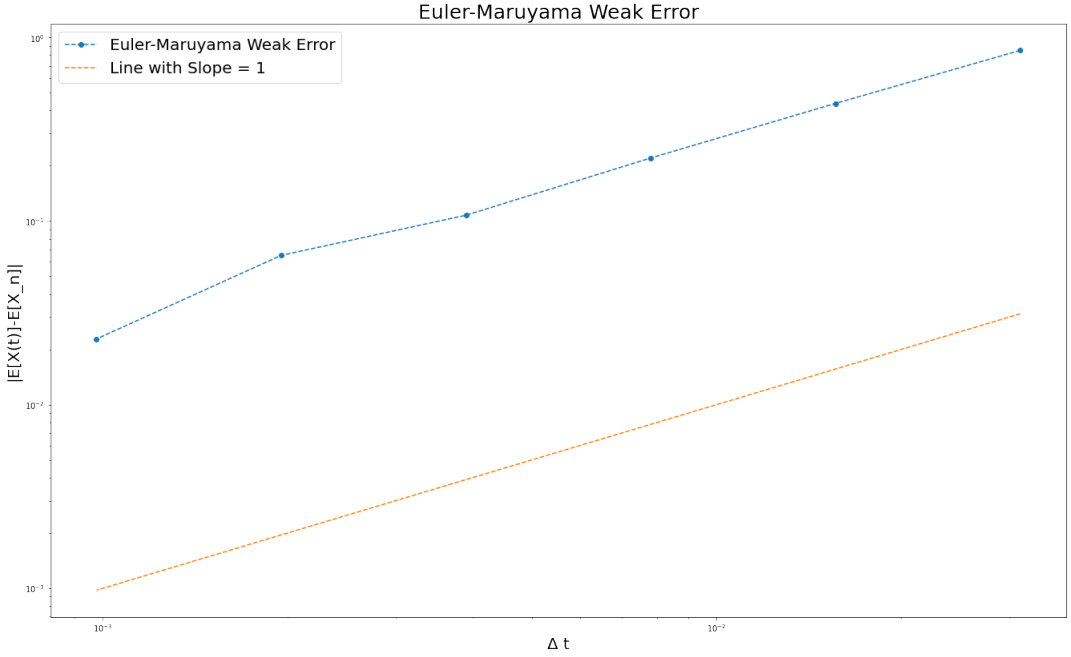

FIGURE (5) Euler-Maruyama weak error (blue) $|\mathbb{E}[X(t)] - \mathbb{E}[X_n]|$ for values of Δt with $\log(X) = \log(X)$ (orange) for comparison.

Visually, the slope of the weak error with respect to the orange line seem similar in Figure 5. We confirm this with a least-squares fit of $\log(e_{\Delta t}^{\text{weak}})$ with $\rho = 1.011$ and a residual of 0.0629. Our findings are consistent with strong order of convergence $\rho = 1$. The order of convergence of the Euler-Maruyama method can also be formally proven. For a formal proof of the strong and weak convergence of the Euler-Maruyama scheme, see page 55 of [2].

Recall that the Euler-Maruyama method had a strong convergence with order $\frac{1}{2}$. We can make an adjustment to the Euler-Maruyama method by adding in a correction term, which will produce the Milstein method:

$$\begin{aligned} X(t+\Delta t) = X(t) + m(t, X(t))\Delta t + \sigma(t, X(t))(B_{t+\Delta t} - B_t) \\ + \frac{1}{2}\sigma(t, X(t))\sigma'(t, X(t))((B_{t+\Delta t} - B_t)^2 - \Delta t) \end{aligned} \tag{8.9}$$

In order to see how to obtain this expression, we perform a *truncated Ito-Taylor expansion*. Consider the SDE

$$X_t = X_{t_0} + \int_{t_0}^{t} m(X_s)\ ds + \int_{t_0}^{t} \sigma(X_s)\ dB_s, \quad X_{t_0} = x_0 \tag{8.10}$$

where $m(X_t), B(X_t)$ are smooth adapted processes with respect to $\{\mathcal{F}_t\}$ and B_t is a standard Brownian motion. Recall Ito's Formula for Ito Processes (Theorem 7.13) for an arbitrary function $f\colon \mathbb{R} \to \mathbb{R}$ that is sufficiently smooth:

$$\begin{aligned} f(X_t) = f(X_{t_0}) + \int_{t_0}^{t} \left[m(X_s)f'(X_s) + \frac{1}{2}(\sigma(X_s))^2 f''(X_s) \right]\ ds \\ + \int_{t_0}^{t} \sigma(X_s)f'(X_s)\ dB_s \end{aligned} \tag{8.11}$$

For simplicity of notation, let the operators $L_1 f(x)$ and $L_2 f(x)$ denote

$$\begin{aligned} L_1 f(x) &= m(x)f'(x) + \frac{1}{2}(\sigma(x))^2 f''(x) \\ L_2 f(x) &= \sigma(x)f'(x) \end{aligned}$$

so we can rewrite (8.11) as

$$f(X_t) = f(X_{t_0}) + \int_{t_0}^{t} L_1 f(X_s)\ ds + \int_{t_0}^{t} L_2 f(X_s)\ dB_s \tag{8.12}$$

We apply (8.12) for functions $f(x) = m(x)$ and $f(x) = \sigma(x)$ to get

$$m(X_t) = m(X_{t_0}) + \int_{t_0}^{t} L_1 m(X_s)\ ds + \int_{t_0}^{t} L_2 m(X_s)\ dB_s \tag{8.13}$$

$$\sigma(X_t) = \sigma(X_{t_0}) + \int_{t_0}^{t} L_1 \sigma(X_s)\ ds + \int_{t_0}^{t} L_2 \sigma(X_s)\ dB_s \tag{8.14}$$

Substituting (8.13) and (8.14) into (8.10), we get the expression

$$\begin{aligned} X_t &= X_{t_0} + \int_{t_0}^{t} \left(m(X_{t_0}) + \int_{t_0}^{s} L_1 m(X_r)\ dr + \int_{t_0}^{s} L_2 m(X_r)\ dB_r \right) ds \\ &\quad + \int_{t_0}^{t} \left(\sigma(X_{t_0}) + \int_{t_0}^{s} L_1 \sigma(X_r)\ dr + \int_{t_0}^{s} L_2 \sigma(X_r)\ dB_r \right) dB_s \\ &= X_{t_0} + m(X_{t_0}) \int_{t_0}^{t} ds + \sigma(X_{t_0}) \int_{t_0}^{t} dB_s \\ &\quad + \int_{t_0}^{t} \int_{t_0}^{s} L_2 \sigma(X_r)\ dB_r dB_s + \mathcal{R} \end{aligned} \tag{8.15}$$

where $\mathcal{R}$ denotes the remaining terms of order greater than $t^{3/2}$:

$$\mathcal{R} = \int_{t_0}^{t} \int_{t_0}^{s} L_1 m(X_r)\ dr\ ds + \int_{t_0}^{t} \int_{t_0}^{s} L_2 m(X_r)\ dB_r\ ds + \int_{t_0}^{t} \int_{t_0}^{s} L_1 \sigma(X_r) dr\ dB_s \tag{8.16}$$

Once more, we use (8.12) for the function $f(x) = L_2\sigma(x)$ to obtain

$$L_2\sigma(X_t) = L_2\sigma(X_{t_0}) + \int_{t_0}^{t} L_1 L_2 \sigma(X_s)\ ds + \int_{t_0}^{t} L_2 L_2 \sigma(X_s)\ dB_s \tag{8.17}$$

Substituting (8.17) into the double integral term of (8.15), we get

$$\begin{aligned} X_t &= X_{t_0} + m(X_{t_0}) \int_{t_0}^{t} ds + \sigma(X_{t_0}) \int_{t_0}^{t} dB_s \\ &\quad + L_2\sigma(X_{t_0}) \int_{t_0}^{t} \int_{t_0}^{s} dB_r dB_s + \tilde{\mathcal{R}} \end{aligned} \tag{8.18}$$

where $\tilde{\mathcal{R}}$ denotes the remainder

$$\begin{aligned} \tilde{\mathcal{R}} &= \mathcal{R} + \int_{t_0}^{t} \int_{t_0}^{s} \int_{t_0}^{r} L_1 L_2 \sigma(X_u)\ du\ dB_r\ ds \\ &\quad + \int_{t_0}^{t} \int_{t_0}^{s} \int_{t_0}^{r} L_2 L_2 \sigma(X_u)\ dB_u\ dB_r\ ds \end{aligned} \tag{8.19}$$

Now we are simply left with dealing with the double integral term in (8.18). Under the standard (Lebesgue) integral, we have

$$\int_{t_0}^{t} \int_{t_0}^{s} dB_r dB_s = \int_{t_0}^{t} (B_s - B_{t_0})\ dB_s = \int_{t_0}^{t} B_s\ dB_s - B_{t_0} \int_{t_0}^{t} dB_s$$

The second term is just

$$B_{t_0} \int_{t_0}^{t} dB_s = B_{t_0}(B_t - B_{t_0}) = B_{t_0} B_t - B_{t_0}^2 \tag{8.20}$$

The first term is a stochastic integral, so applying Ito's formula for $f(x) = \frac{1}{2}x^2$, we get

$$\frac{1}{2} B_t^2 = \frac{1}{2} B_{t_0}^2 + \int_{t_0}^{t} B_s\ dB_s + \frac{1}{2} \int_{t_0}^{t} ds$$

which implies

$$\int_{t_0}^{t} B_s dB_s = \frac{1}{2} B_t^2 - \frac{1}{2} B_{t_0}^2 - \frac{1}{2}(t - t_0) \tag{8.21}$$

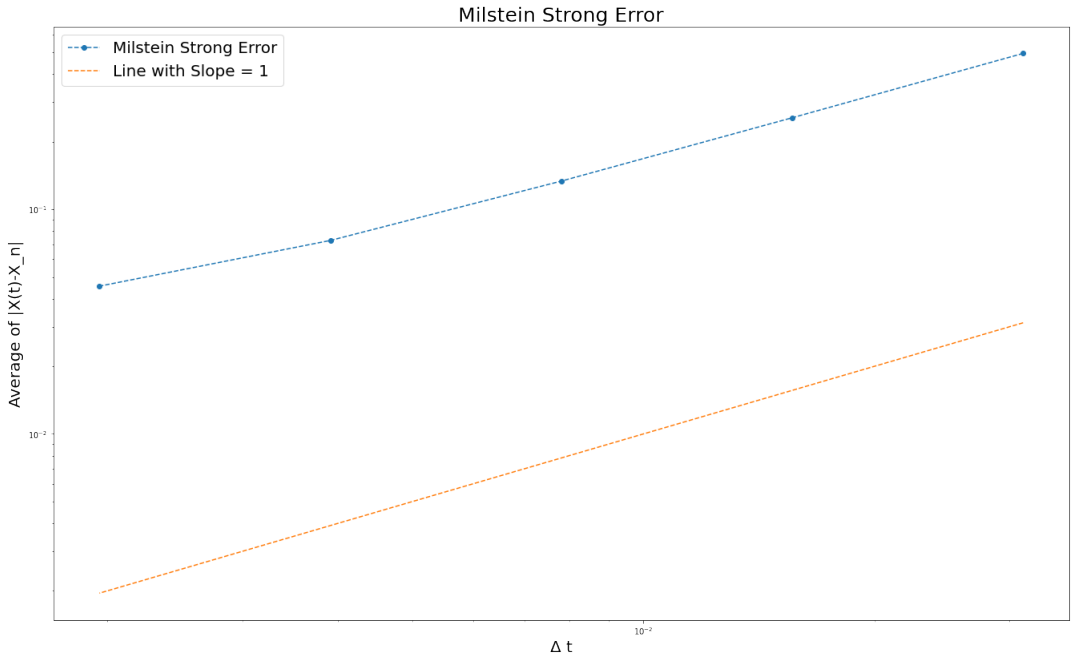

FIGURE (6) Milstein strong error (blue) $|X(t) - X_n|$ for values of Δt with $\log(X) = \log(X)$ (orange) for comparison.

Combining the results of (8.20) and (8.21), then

$$\int_{t_0}^{t}\int_{t_0}^{s} dB_r dB_s = \frac{1}{2}((B_t - B_{t_0})^2 - (t - t_0)) \tag{8.22}$$

Adding (8.22) back into (8.18), replacing $L_2\sigma(x) = \sigma(x)\sigma'(x)$ and solving the integrals, we obtain

$$\begin{aligned} X_t = X_{t_0} + m(X_{t_0})(t - t_0) + \sigma(X_{t_0})(B_t - B_{t_0}) \\ + \frac{1}{2}\sigma(X_{t_0})\sigma'(X_{t_0})((B_t - B_{t_0})^2 - (t - t_0)) + \tilde{\mathcal{R}} \end{aligned} \tag{8.23}$$

Thus we have obtained the Ito-Taylor expansion of our stochastic differential equation. By truncating the remainder, we can derive a numerical stochastic integration method

$$\begin{aligned} X(t + \Delta t) = X(t) + m(t, X(t))\Delta t + \sigma(t, X(t))(B_{t+\Delta t} - B_t) \\ + \frac{1}{2}\sigma(t, X(t))\sigma'(t, X(t))((B_{t+\Delta t} - B_t)^2 - \Delta t) \end{aligned}$$

which turns out to be the Milstein method. Also note that by only leaving first order terms and truncating the rest, we have the Euler-Maruyama method.

We claim the the Milstein method has strong order of convergence $\rho = 1$ which is an improvement to the Euler-Maruyama method. To test this, we compute a numerical experiment of 1000 sample Brownian paths for values of $\Delta t = 2^{-x}$, $5 \leq x \leq 10$. Graphing the error terms with respect to Δt, we get Figure 5. We further confirm our hypothesis with a least-squares fit of 0.870 and a residual of 0.0146.

Acknowledgments

I would like to thank my mentor Stephen Yearwood for guiding me through the material. I would also like to thank Professor Peter May for providing the opportunity to create this project through the REU and also Professor Gregory Lawler for running the probability subprogram.

References

[1] https://web.math.princeton.edu/ rvan/acm217/ACM217.pdf
[2] Tobias Jahnke Numerical methods in mathematical finance https://www.math.kit.edu/ianm3/lehre/numfima2016w/media/num-meth-math-fin.pdf
[3] Gregory Lawler Stochastic Calculus: An Introduction with Applications https://www.math.uchicago.edu/ lawler/finbook.pdf
[4] Desmond J. Higham An Algorithmic Introduction to Numerical Simulation of Stochastic Differential Equations https://epubs.siam.org/doi/pdf/10.1137/S0036144500378302

Appendix A. Code for Graphs and Numerical Simulations

The Python script below shows the discretized versions of turning a simple random walk into Brownian motion.

```
def generate_walk(n):
    index = np.linspace(start = 0, stop = 10, num = 10*n)
    randomlist = np.array(random.choices([-1*np.sqrt(1/n),
                                          1*np.sqrt(1/n)],
                                          k = 10*n))
    randomlist[0] = 0
    return index, randomlist.cumsum()
index_1, walk_1 = generate_walk(1)
index_10000, walk_10000 = generate_walk(10000)

fig = plt.figure(figsize=(50,30))
ax1 = fig.add_subplot(321)
ax1.set_title("Simple Random Walk (n=1)")
ax1.step(index_1, walk_1, where = "post")

ax2 = fig.add_subplot(323)
ax2.set_title("Simple Random Walk (n=10000)")
ax2.step(index_10000, walk_10000, where = "post")
```

The Python script below provides the Euler approximation method for first order differential equations

```
### Euler's Method vs function

df = lambda t: np.exp(t) #differential equation
x0 = 1     #initial value
dt = 0.1     #step size
t = np.arange(0,1+dt,dt) #values from 0-1

dft = df(t) #values of differential equation
```

```
dft[0] = 0

euler = x0 + (dft * dt).cumsum() #euler approximation
f = lambda t: np.exp(t) #exact solution


fig = plt.figure(figsize=(50,30))
ax3 = fig.add_subplot(221)
ax3.plot(t, f(t), label = "Solution for f(x)")
ax3.plot(t, euler, linestyle = 'dashed', marker = 'o',
         label = "Euler approximation for f(x)")
ax3.set_xlabel('t')
ax3.legend(prop={'size': 20})
plt.show()
```

The Python script below provides the Euler-Maruyama approximation compared to a geometric Brownian motion.

```
## Euler–Maruyama Method
# dX(t) = mu*X(t)*dt + sigma*X(t)*dB(t) <- stochastic differential equation
# Parameters: mu = 2, sigma = 1, X(0) = 1
# X(t) = X(0)exp{(mu - (sigma^2/2)t + sigma*B(t))} <- exact solution
#


mu = 2
sigma = 1
x0 = 1

N = 2**15
T = 1
dt = T/N
np.random.seed(14)
dB = math.sqrt(dt) * (np.random.normal(0, 1, size = N+1))
B = dB.cumsum() #discrete Brownian motion simulation



I = 2**7 #number of Euler increments
step = int(N/I)
Dt = step * dt
t_I = np.arange(0,1 + Dt,Dt)
DBt = np.array([sum(dB[k*step:(k+1)*step]) for k in range(I)])
       #Brownian motion increments of size Dt

X_euler = np.zeros(I + 1) #initializing empty array
```

```
X_euler[0] = x0
X_euler[1:] = x0 * ((1 + mu * Dt + sigma * DBt).cumprod())
    #Euler approximation for Geometric Brownian motion


t_N = np.arange(0,1 + dt,dt) #values from 0-1
X_exact = x0 * np.exp((mu-0.5*sigma**2)*t_N + sigma * B) #exact solution

fig = plt.figure(figsize=(50,30))
ax4 = fig.add_subplot(221)
ax4.plot(t_N, X_exact, label = "Exact Solution for X(t)")
ax4.plot(t_I, X_euler, linestyle = 'dashed', marker = '.',
         label = "Euler-Maruyama Approximation")
ax4.set_xlabel('t')
ax4.legend(prop={'size': 20})
plt.show()

error = X_exact[-1] - X_euler[-1]
```

The Python script below tests for the strong convergence of the Euler-Maruyama method.

```
%%time
#test strong convergence E[|X_euler - X_exact|]
# dX(t) = mu*X(t)*dt + sigma*X(t)*dB(t) <- stochastic differential equation
# Parameters: mu = 2, sigma = 1, X(0) = 1
# X(t) = X(0)exp{(mu - (sigma^2/2)t + sigma*B(t))} <- exact solution
# testing M = 10000 sample paths at T = 1 for Dt = 2^-5 -> 2^-14

mu = 2
sigma = 1
x0 = 1

N = 2**15 #number of discrete Brownian increments
T = 1 #end time
dt = T/N


M = 1000 #number of sample paths
errors = np.zeros([M,10])

for m in range(M):
    dB = math.sqrt(dt) * np.random.normal(0, 1, size = N+1)
    B = dB.cumsum() #discrete Brownian motion simulation

    for i in range(5,15):
```

```
        I = 2**i #number of Euler increments
        step = int(N/I)
        Dt = step * dt
        DBt = np.array([sum(dB[k*step:(k+1)*step]) for k in range(I)])
              #Brownian motion increments of size Dt

        X_euler = np.zeros(I + 1) #initializing empty array
        X_euler[0] = x0
        X_euler[1:] = x0 * ((1 + mu * Dt + sigma * DBt).cumprod())
             #Euler approximation for Geometric Brownian motion


        X_exact = x0 * np.exp((mu-0.5*sigma**2)*T + sigma * B[-1])
             #exact solution

        errors[m,i-5] = abs(X_exact - X_euler[-1])

Dts = np.array([2**x for x in range(-5,-15,-1)])
err_means = errors.mean(axis = 0)

fig = plt.figure(figsize=(50,30))
ax5 = fig.add_subplot(221)
ax5.loglog(Dts, err_means, marker = "o", linestyle = "dashed",
           label = "Euler-Maruyama Strong Error")
ax5.loglog(Dts, 0.5*Dts, linestyle = "dashed",
           label = "Line with Slope = 1/2")
ax5.set_xlabel('\u0394 t', size = 20)
ax5.set_ylabel('Average of |X(t)-X_n|', size = 20)
ax5.set_title("Euler-Maruyama Strong Error", size = 25)
ax5.legend(prop={'size': 20})
plt.show()
```

The Python script below tests the weak convergence of the Euler-Maruyama method.

```
#test weak convergence |E[X_euler] - E[X_exact]|
# dX(t) = mu*X(t)*dt + sigma*X(t)*dB(t) <- stochastic differential equation
# Parameters: mu = 2, sigma = 1, X(0) = 1
# X(t) = X(0)exp{(mu - (sigma^2/2)t + sigma*B(t))} <- exact solution
# testing M = 10000 sample paths at T = 1 for Dt = 2^-5 -> 2^-

mu = 2
sigma = 1
x0 = 1
```

```
N = 2**15 #number of discrete Brownian increments
T = 1 #end time
dt = T/N


M = 1000000 #number of sample paths
X_euler = np.zeros([6,1])
for i in range(5,11):
    I = 2**i #number of Euler increments
    step = round(N/I)
    Dt = step * dt
    temp = np.full(M, x0)
    for j in range(I-1):
        DBt = np.sqrt(Dt) * np.random.normal(0,1,M)
        temp = temp + mu * temp * Dt + sigma * temp * DBt

    X_euler[i-5] = temp.mean()

weak_errors = abs(X_euler - np.exp(mu))
weak_errors = weak_errors.T[0]
Dts = np.array([2**x for x in range(-5,-11,-1)])

fig = plt.figure(figsize=(50,30))
ax6 = fig.add_subplot(221)
ax6.loglog(Dts, abs(weak_errors), marker = "o", linestyle = "dashed",
           label = "Euler-Maruyama Weak Error")
ax6.loglog(Dts, Dts, linestyle = "dashed", label = "Line with Slope = 1")
ax6.set_xlabel('\u0394 t', size = 20)
ax6.set_ylabel('|E[X(t)]-E[X_n]|', size = 20)
ax6.set_title("Euler-Maruyama Weak Error", size = 25)
ax6.legend(prop={'size': 20})
plt.show()
```

The Python script below tests for the strong converence of the Milstein method

```
#Milstein Method
#test strong convergence E[|X_mil - X_exact|]
# dX(t) = mu*X(t)*dt + sigma*X(t)*dB(t) <- stochastic differential equation
# Parameters: mu = 2, sigma = 1, X(0) = 1
# X(t) = X(0)exp{(mu - (sigma^2/2)t + sigma*B(t))} <- exact solution
# testing M = 10000 sample paths at T = 1 for Dt = 2^-5 -> 2^-14

mu = 2
sigma = 1
x0 = 1
```

```
N = 2**15 #number of discrete Brownian increments
T = 1 #end time
dt = T/N


M = 1000 #number of sample paths

dB = np.sqrt(dt) * np.random.normal(0, 1, size = [M, N+1])
B = dB.cumsum(axis = 1)

X_exact = x0 * np.exp((mu-0.5*sigma**2)*T + sigma * B[:,-1])
       #exact value at T = 1
errors = np.zeros(5)
for i in range(5,10):
    I = 2**i #number of Euler increments
    step = int(N/I) #2^-i sized
    Dt = step * dt
    X_mil = np.zeros([M,I+1])
         #initialize empty array for Milstein approximation
    X_mil[:,0] = x0
    for j in range(1,I+1):
        DBt = dB[:,(j-1)*step:j*step].sum(axis = 1)
        X_mil[:,j] = X_mil[:,j-1] + mu * X_mil[:,j-1] * Dt +
                          sigma * X_mil[:,j-1] * DBt +
                          0.5 * sigma**2 * X_mil[:,j-1]* (DBt**2 - Dt)

    errors[i-5] = abs(X_mil[:,-1] - X_exact).mean()
Dts = np.array([2**x for x in range(-5,-10,-1)])
err_means = errors.mean(axis = 0)

fig = plt.figure(figsize=(50,30))
ax5 = fig.add_subplot(221)
ax5.loglog(Dts, errors, marker = "o", linestyle = "dashed",
           label = "Milstein Strong Error")
ax5.loglog(Dts, Dts, linestyle = "dashed", label = "Line with Slope = 1")
ax5.set_xlabel('\u0394 t', size = 20)
ax5.set_ylabel('Average of |X(t)-X_n|', size = 20)
ax5.set_title("Milstein Strong Error", size = 25)
ax5.legend(prop={'size': 20})
plt.show()
```

This function was used to compute the least-squares fit and residual for the errors.

```
#compute least squares and residual
```

```
def least_squares_resid(Dts, error):
    A = np.vstack([np.log(Dts), np.ones(len(Dts))]).T
    y = np.log(error)
    x = np.linalg.lstsq(A, y, rcond=None)[0]
    resid = np.linalg.norm(A.dot(x) - y) ** 2
    return x, resid
```