STOCHASTIC CALCULUS AND STOCHASTIC DIFFERENTIAL EQUATIONS

SHUHONG LIU

Abstract. This paper introduces stochastic calculus and stochastic differential equations. We start with basic stochastic processes such as martingale and Brownian motion. We then formally define the Itô integral and establish Itô's formula, the fundamental theorem of stochastic calculus. Finally, we prove the Existence and Uniqueness Theorem of stochastic differential equations and present the techniques to solve linear stochastic differential equations.

Contents

1. Introduction

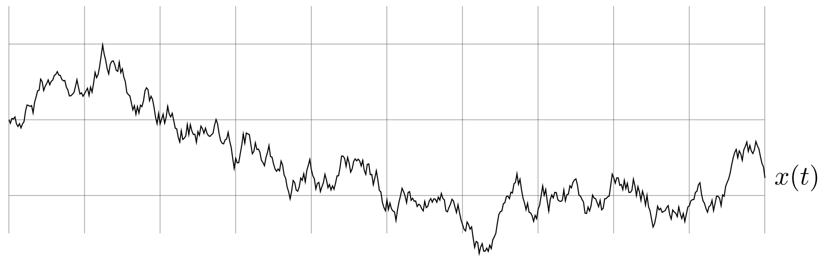
Often times, we understand the change of a system better than the system itself. For example, consider the ordinary differential equation (ODE)

$$\begin{cases} \dot{x}(t) = f(t, x(t)) \\ x(0) = x_0 \end{cases} \tag{1.1}$$

where $x_0 \in \mathbb{R}$ is a fixed point and $f : [0, \infty) \times \mathbb{R} \to \mathbb{R}$ is a smooth function. The goal is to find the trajectory $x(t)$ satisfying the initial value problem. However,

Date: August 2019.

in many applications, the experimentally measured trajectory does not behave as deterministic as predicted. In some cases, the supposedly smooth trajectory $x(t)$ is not even differentiable in t. For example, the trajectory may look like



Therefore, we would like to include some random noise in the system to explain the disturbance. The stochastic representation of the system (1.1) is

$$\begin{cases} \dot{x}(t) = f(t, x(t)) + g(t, x(t))\xi(t) \\ x(0) = x_0 \end{cases} \tag{1.2}$$

where $\xi(t)$ is the white noise. The system (1.2) is called a **stochastic differential equation (SDE)**.

This approach, however, leaves us with some problems to solve:

- define the white noice $\xi(t)$ rigorously;
- define the solution concept of an SDE;
- establish conditions on f, g, and x_0 upon which an SDE has solutions;
- discuss the uniqueness of the solution.

As it turns out, the white noise $\xi(t)$ is related to Brownian motion. More precisely, $\xi(t)$ is the infinitesimal increment along the Brownian path. Therefore, we often write the SDE (1.2) as

$$\begin{cases} dx(t) = f(t, x(t))dt + g(t, x(t))dW_t \\ x(0) = x_0 \end{cases} \tag{1.3}$$

where W_t denotes the standard Brownian motion. We introduce Brownian motion in Section 2, along with two other important stochastic processes, simple random walk and martingale. In particular, martingale and Brownian motion play a huge role in studying stochastic calculus and stochastic differential equations.

In Section 3, we construct the Itô integral. In other words, we will define what it means to integrate against a Brownian motion. We will then work out a simple example to show the difference between the Itô integral and the Riemann integral. Of course, it is impractical to do every calculation by definition. For the Riemann calculus, we have the Fundamental Theorem of Calculus, which gives us the relationship between a differentiable function and its derivative. For stochastic calculus, we have a similar relationship given by Itô's formula. In the latter half of Section 3, we prove Itô's formula and show, through a few examples, how Itô's formula simplifies the calculation of stochastic integral.

In Section 4, we finally discuss stochastic differential equations. Similar to ODE, we will prove the Existence and Uniqueness Theorem for SDE. Knowing there exists a unique solution, we then present the solution methods for linear SDEs and give the general formulas.

Once we solve these problems, we have a powerful mathematical tool to study systems that evolve stochastically. SDEs are heavily used in many different areas such as quantitative finance, partial differential equations, and statistical physics. Since each one of these applications is complicated enough for at least another paper, we will not dive into the details. Instead, we give a brief description for each topic at the end of Section 4 and point to introductory materials for interested readers.

2. Stochastic Processes

In this section, we introduce three basic stochastic processes. We start with simple random walk, the most fundamental stochastic process, to give a sense of the interesting consequences of randomness. We then move on to martingale, the model for "fair games". Martingale has properties, namely the Optional Sampling Theorem and the Martingale Inequality, that are crucial for the study of other stochastic processes as well as stochastic calculus. Finally, we introduce Brownian motion, which is a continuous-time martingale and a scaling limit (in a certain sense) of simple random walk at the same time.

Many results about these stochastic processes are important in studying the Itô integral and stochastic differential equation. However, since these are well-known mathematical objects, most of the proofs will be omitted.

2.1. Simple Random Walk on $\mathbb{Z}$.

Definition 2.1. A **random walk** on the integers $\mathbb{Z}$ with step function F and initial state $x \in \mathbb{Z}$ is a sequence of random variables,

$$S_n = x + \sum_{i=1}^{n} \xi_i$$

where $\xi_1, \ldots, \xi_n$ are i.i.d. random variables with common distribution F.

In particular, a **simple random walk** has Rademacher-$\frac{1}{2}$ increments, i.e.,

$$\mathbb{P}(\xi_i = 1) = \mathbb{P}(\xi_i = -1) = \frac{1}{2}$$

Therefore, a simple random walk is determined by a sequence of fair coin tosses: for each Head, jump to the right; for each Tail, jump to the left.

One (of many) discrete random process modeled by simple random walk is the evolution of the wealth of a gambler whose investment jumps by either ± 1 with equal probability in each period. Naturally, we want to study the following problems:

Problem 2.2. (Gambler's Ruin) Suppose the gambler starts with x dollars. What is the probability that his wealth grows to A dollars before he goes broke? More precisely, define

$$T := \min\{n : S_n = 0 \text{ or } A\}. \tag{2.3}$$

What is $\mathbb{P}^x(S_T = A)$?

Remark 2.4. Before we start, we notice that $\mathbb{P}(T < \infty) = 1$: the game will end as long as there are A consecutive Heads, but if the gambler tosses a fair coin forever, the probability that he does not see any A consecutive Heads is 0.

To solve the problem, we define $u(x) := \mathbb{P}^x(S_T = A)$. Clearly $u(A) = 1$ and $u(0) = 0$. For $0 < x < A$, since after the first jump it is like a new simple random walk starting at either $x \pm 1$, we have a **difference equation**

$$u(x) = \frac{1}{2}u(x-1) + \frac{1}{2}u(x+1)$$

Let $d(x) := u(x) - u(x-1)$. Then from above, $d(x) = d(x+1) = d$ for all x. Since

$$u(x) = u(0) + \sum_{i=1}^{x} d(i) = xd$$

the boundary condition $u(A) = 1$ gives $d = 1/A$, so $u(x) = x/A$.

Proposition 2.5. $\mathbb{P}^x(S_T = A) = x/A$.

Difference equations are heavily used in the study of combinatorics and discrete stochastic processes such as Markov Chains. Of course, most difference equations are much more complicated and interested readers can refer to [1], Section 0.3 for solution methods for more general difference equations. In particular, for linear difference equations with constant term, we need to formulate them as matrix equations, which can then be solved using matrix multiplication. The second problem of the gambler's game is a simple example of this technique.

Problem 2.6. (Expected Duration) How long, on average, will the game last? More precisely, what is $\mathbb{E}T$?

Similar to the first problem, we define $v(x) := \mathbb{E}^x T$. Clearly $v(0) = v(A) = 0$. For $0 < x < A$, since it takes one step to jump to $x \pm 1$, we have

$$v(x) = 1 + \frac{1}{2}v(x-1) + \frac{1}{2}v(x+1)$$

Let $d(x) := v(x) - v(x-1)$. Then

$$\begin{bmatrix} d(x+1) \\ -2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} d(x) \\ -2 \end{bmatrix} \quad \Longrightarrow \quad \begin{bmatrix} d(x) \\ -2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}^x \begin{bmatrix} d(1) \\ -2 \end{bmatrix}$$

where $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}^x = \begin{bmatrix} 1 & x \\ 0 & 1 \end{bmatrix}$ can be shown by induction or eigenvalue decomposition. Hence $d(x) = d(1) - 2(x-1) = v(1) - 2(x-1)$ as $v(0) = 0$. Since

$$v(x) = \sum_{i=1}^{x} d(i) = xv(1) - 2\sum_{i=1}^{x-1} i = xv(1) - x(x-1)$$

the boundary condition $v(A) = 0$ gives $v(1) = A - 1$, so $v(x) = x(A - x)$.

Proposition 2.7. $\mathbb{E}^x T = x(A - x)$.

Remark 2.8. We make a small generalization. Let $T := \min\{n : S_n = A \text{ or } -B\}$ where $A, B \in \mathbb{N}_+$. Then

$$\mathbb{P}^0(S_T = A) = \frac{B}{A+B} \quad \text{and} \quad \mathbb{E}^0 T = AB \tag{2.9}$$

Intuitively, this generalization corresponds to the game where two gamblers with initial wealth A and B dollars, respectively, consecutively bet 1 dollar on the outcome of fair coin tosses until one of them goes broke. We will derive the same results later using martingale theory.

In discrete stochastic processes, there are many random times similar to (2.3). They are non-anticipating, i.e., at any time n, we can determine whether the criterion for such a random time is met or not solely by the "history" up to time n. Strictly speaking, we give the following definitions.

Definition 2.10. A discrete **filtration** of a set Ω is a collection $\{\mathcal{F}_n\}$ of σ-algebras of subsets of Ω such that $\mathcal{F}_n \subset \mathcal{F}_{n+1}$ for all $n \in \mathbb{N}$.

In particular, for a discrete stochastic process $\{X_n\}$, the **natural filtration** $\{\mathcal{F}_n\}$ is such that each $\mathcal{F}_n$ is the σ-algebra generated by $X_1, \ldots, X_n$. We can interpret $\mathcal{F}_n$ as all the information contained in $X_1, \ldots, X_n$.

Definition 2.11. An integer-valued random variable τ is a **stopping time** relative to a filtration $\{\mathcal{F}_n\}$ if for each $n \in \mathbb{N}$, the event $\{\tau = n\}$ is $\mathcal{F}_n$-measurable.

Examples 2.12. First-visit times are usually stopping times. For example, for every $i \in \mathbb{N}$, $\tau_i := \min\{n : S_n = i\}$ is a stopping time. In contrary, last-visit times are usually not stopping times.

Stopping times are important for the study of simple random walk because of the following property.

Proposition 2.13. *(Strong Markov Property) If τ is a stopping time for a random walk $\{S_n\}$, then the post-τ sequence $\{S_{\tau+n}\}$ is also a random walk, with the same step function, starting at S_τ, and independent of $\mathcal{F}_\tau$.*

Proposition 2.5, combined with the Strong Markov Property, renders interesting consequences. For a simple random walk $\{S_n\}$ starting at x, the probability that it reaches A before 0 is x/A, so the probability that it reaches 0 is at least $1 - x/A$. Since this is true for every A, setting $A \to \infty$ gives

$$\mathbb{P}^x\{S_n = 0 \text{ eventually}\} = 1 \tag{2.14}$$

Notice that (2.14) is transition invariant, since for any $i \in \mathbb{N}$, $\{S_n + i\}$ is a simple random walk starting at $x + i$. (2.14) is also reflection invariant, since changing x to $-x$ just reverses the roles of Head and Tail. Therefore, for any $i \in \mathbb{Z}$,

$$\mathbb{P}^x\{S_n = i \text{ eventually}\} = 1 \tag{2.15}$$

Now, consider the stopping times τ_i in Examples 2.12. The Strong Markov Property suggests the post-τ_i sequence is a new simple random walk independent of $\mathcal{F}_{\tau_i}$. Reversing the roles of x and i in (2.15), we know that with probability 1, $\{S_n\}$ will return to x after it visits i. Applying the Strong Markov Property again, we have a new simple random walk starting at x, which will eventually visit i and return to x and so on. Inductively, we can conclude that:

Theorem 2.16. *With probability one, simple random walk visits every state $i \in \mathbb{Z}$ infinitely often.*

Given this property, we say simple random walk is **recurrent**.

2.2. **Martingale.**

A martingale is the model for "fair games" in which the expected future payoff, conditioned on the current payoff, is the same as the current payoff. Although conditional expectation is easy to understand intuitively, the formal definition needs measure theory. Therefore, the definition and properties of conditional expectation are moved to the appendix.

From now on, we will be working in the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with discrete filtration $\{\mathcal{F}_n\}$ or continuous filtration $\{\mathcal{F}_t\}$. Continuous filtration is defined similarly.

Definition 2.17. A continuous **filtration** of a set Ω is a collection $\{\mathcal{F}_t\}$ of σ-algebras of subsets of Ω such that $\mathcal{F}_s \subset \mathcal{F}_t$ for all $s < t$.

When we construct the model for fair games, a minimum requirement is the non-anticipating property.

Definition 2.18. A sequence $\{X_t\}$ of random variables is an **adaptive process** relative to $\{\mathcal{F}_t\}$, if the random variable X_t is $\mathcal{F}_t$-measurable for each t.

On top of being adaptive, we also need fair games to have expected future payoff same as the current payoff. To illustrate the power of martingale theory, we start with the discrete case, so that we can apply the theory to simple random walk.

Definition 2.19. A discrete-time adapted process $\{X_n\}$ of integrable random variables, that is, $\mathbb{E}|X_n| < \infty$ for all $n \in \mathbb{N}$, is a **martingale** relative to $\{\mathcal{F}_n\}$ if $\mathbb{E}(X_{n+1}|\mathcal{F}_n) = X_n$ *a.s.* for all $n \in \mathbb{N}$.

Examples 2.20. In all examples, let $\{\mathcal{F}_n\}$ be the natural filtration.

(1) Let $\{X_n\}$ be a sequence of i.i.d. random variables with $\mathbb{E}X_n = 0$. Then the sequence of partial sums

$$S_n = \sum_{i=1}^{n} X_i \tag{2.21}$$

is a martingale.

(2) Let $\{X_n\}$ be a sequence of i.i.d. random variables with $\mathbb{E}X_n = 0$ and $\mathrm{Var}(X_n) = \sigma^2 < \infty$. Let S_n be the same as (2.21). Then the sequence

$$S_n^2 - n\sigma^2 \tag{2.22}$$

is a martingale.

(3) Let $\{X_n\}$ be a sequence of i.i.d. random variables with finite moment generating function $\varphi(\theta) = \mathbb{E}e^{\theta X_n}$. Let S_n be the same as (2.21). Then the sequence

$$Z_n = \frac{e^{\theta S_n}}{\varphi(\theta)^n}$$

is a positive martingale.

One of the most important theorems in martingale theory is Doob's Optional Sampling Theorem. It states that, under certain conditions, the expected payoff at any stopping time is the same as the initial payoff.

Theorem 2.23. *(Optional Sampling Theorem) Let $\{X_n\}$ be a martingale relative to $\{\mathcal{F}_n\}$. Let τ be a stopping time and let $\tau \wedge n$ denote* $\min\{\tau, n\}$. *Then $X_{\tau\wedge n}$ is a martingale. In particular,*

$$\mathbb{E}X_\tau = \mathbb{E}X_0$$

if any of the following three conditions holds a.s.:

- *τ is bounded.*
- *$|X_{\tau\wedge n}|$ is bounded.*
- *$\mathbb{E}|X_{\tau\wedge n}|^2$ is bounded.*

Proof. See [2], Chapter 1, Section 3. □

Using Optional Sampling Theorem, we can easily solve many problems related to stopping times. For example, we revisit the Gambler's Ruin Problem 2.2 and the Expected Duration Problem 2.6.

Example 2.24. As shown in (2.21), simple random walk $\{S_n\}$ is a martingale. It is easy to check that T as in (2.3) is a stopping time and $|S_{T\wedge n}| \leq \max\{A, B\}$. Hence by Optional Sampling Theorem,

$$0 = \mathbb{E}S_0 = \mathbb{E}S_T = A\mathbb{P}(S_T = A) - B\mathbb{P}(S_T = -B)$$

where $\mathbb{P}(S_T = A) + \mathbb{P}(S_T = B) = 1$. Hence

$$\mathbb{P}(S_T = A) = \frac{B}{A+B}$$

For the second problem, since Rademacher-$\frac{1}{2}$ random variable has variance 1, $S_n^2 - n$ is a martingale, as shown by (2.22). Since $|S_{T\wedge n}^2 - T\wedge n| \leq \max\{A, B\}^2 + T$, by Optional Sampling Theorem,

$$0 = \mathbb{E}(S_0^2 - 0) = \mathbb{E}(S_T^2 - T) = \mathbb{E}(S_T^2) - \mathbb{E}T$$

Hence

$$\mathbb{E}T = \mathbb{E}(S_T^2) = A^2\mathbb{P}(S_T = A) + B^2\mathbb{P}(S_T = -B) = AB$$

We see the results correspond with (2.9), but the derivation is much simpler than solving difference equations.

Now we can discuss continuous-time martingales.

Definition 2.25. A continuous adapted process $\{X_t\}$ of integrable random variables is a **martingale** relative to $\{\mathcal{F}_t\}$ if $\mathbb{E}(X_t|\mathcal{F}_s) = X_s$ *a.s.* for all $s < t$.

Another important property of martingale is the Martingale Inequality, which gives a good estimate on the maximum value attained in a time period. In later sections, the Martingale Inequality will be applied to continuous-time martingales such as Brownian motion and some Itô processes.

Theorem 2.26. *(Martingale Inequality) If $\{X_t\}$ is a martingale and $1 < p < \infty$, then for all t,*

$$\mathbb{E}\left(\max_{0\leq s\leq t} |X_s|^p\right) \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}|X_t|^p$$

Proof. See [4], Chapter 2, Section I and Appendix B. □

2.3. **Brownian Motion.**

Definition 2.27. A standard **Brownian motion** (or **Wiener process**) is a continuous-time stochastic process $\{W_t\}$ satisfying

- $W_0 = 0$;
- for all $t, s \geq 0$, $W_{t+s} - W_s$ has normal $N(0, t)$ distribution;
- for all $s < t$, the random variable $W_t - W_s$ is independent of W_r for all $r \leq s$;
- with probability one, the path $t \to W_t$ is continuous.

This definition has many subtle details.

Remark 2.28. It is not *a priori* clear that such a stochastic process exists, since it is possible that the second and third conditions make the path discontinuous. It was first proved by Norbert Wiener that Brownian motion does exist. Interested readers can refer to [2], Chapter 2, Section 5 for the proof of existence outlining Levy's construction of Brownian motion.

Remark 2.29. Although we mentioned in Section 1 that the white noise $\xi(t)$ is dW_t, Brownian motion is actually nowhere differentiable. The term dW_t should be interpreted as the infinitesimal increment along the Brownian path.

Remark 2.30. Brownian motion can be viewed as the scaling limit of simple random walk. Let $\{\xi_n\}$ be a sequence of i.i.d. random variables with mean 0 and variance 1. For each $n \geq 1$, we define

$$W_t^n = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} \xi_i$$

This is a random step function of size $\xi_i/\sqrt{n}$ at times i/n for i between 1 and nt. Since all ξ_is are independent, W_t^n has independent increments. By the central limit theorem, the distribution of $W_{t+s}^n - W_s^n$ is approximately $N(0,t)$.

So far, it seems W_t^n will converge nicely to a Brownian motion, but in fact, one has to be very careful when taking this limit. Donsker's Theorem proves that as $n \to \infty$, W_t^n converges (in a certain sense) to a standard Brownian motion W_t. The proof uses binary splitting martingales and Skorokhod representation. Interested readers can refer to [3].

This relation between Brownian motion and simple random walk is important. First, it helps explain why Brownian motion is so ubiquitous in nature. Many stochastic processes behave like random walks with small but frequent jumps, especially for long time periods. Second, it suggests that many statistics and properties of simple random walk will have correspondence in Brownian motion. For example, we know that simple random walk is translation and reflection invariant, so naturally, we have the following property about Brownian motion.

Proposition 2.31. *(Symmetry and Scaling Laws) Let $\{W_t\}$ be a standard Brownian motion. Then each of the following is also a standard Brownian motion:*

$$\{-W_t\} \quad \{W_{t+s} - W_t\} \quad \{aW(t/a^2)\} \quad \{tW(1/t)\}$$

Another important property of simple random walk is the Strong Markov Property, which also holds for Brownian motion. We first define stopping times for continuous-time stochastic processes.

Definition 2.32. A nonnegative random variable τ is a stopping time relative to a filtration $\{\mathcal{F}_t\}$ if for each $t \geq 0$, the event $\{\tau \leq t\}$ is $\mathcal{F}_t$-measurable.

Proposition 2.33. *(Strong Markov Property) Let $\{W_t\}$ be a standard Brownian motion with filtration $\{\mathcal{F}_t\}$. Let τ be a stopping time. For all $t \geq 0$, define*

$$W_t^* = W_{t+\tau} - W_\tau$$

and let $\{\mathcal{F}_t^*\}$ be its filtration. Then

- *$\{W_t^*\}$ is also a standard Brownian motion;*
- *for all $t \geq 0$, $\mathcal{F}_t^*$ is independent of $\mathcal{F}_\tau$.*

These two propositions can render useful consequences. We are often interested in the maximal and minimal values attained by Brownian motion in a time period. Formally, define

$$M(t) := \max\{W_s : 0 \le s \le t\} \quad \text{and} \quad m(t) := \min\{W_s : 0 \le s \le t\}$$

We are interested in the events $\{M(t) \ge a\}$ and $\{m(t) \le b\}$ for some a and b. For simplicity, we only deal with the maximum here. Define the **first-passage time**

$$\tau_a := \min\{t : W_t = a\}$$

which is a stopping time. Then the events $\{\tau_a \le t\}$ and $\{M(t) \ge a\}$ are identical. Furthermore, since the Brownian path is continuous, the events $\{M(t) \ge a\}$ and $\{\tau_a < t\}$ are also identical. We claim that if $\tau_a < t$, then W_t is as likely to be above the level a as to be below. Proof of the claim uses the Strong Markov Property and symmetry of the normal distribution. With details omitted, we arrive at one of the most important formulas for Brownian motion.

Proposition 2.34. *(Reflection Principle)*

$$\mathbb{P}(M(t) \ge a) = \mathbb{P}(\tau_a < t) = 2\mathbb{P}(W_t > a) = 2 - 2\Phi(a/\sqrt{t})$$

where Φ is the cumulative density function for the standard normal distribution.

Using the Reflection Principle, we can derive the distribution of τ_a.

Proposition 2.35. *For all a, the first passage time τ_a is finite a.s. and has probability density function*

$$f(t) = \frac{ae^{-a^2/2t}}{\sqrt{2\pi t^3}}$$

Besides the maximal and minimal values, we are also interested in how drifted away the Brownian path is. One measure for this is the quadratic variation.

Definition 2.36. Let $\{X_t\}$ be a continuous-time stochastic process. Fix $T > 0$ and let $\Pi = \{t_0, t_1, \dots, t_n\}$ be a partition of the interval $[0, T]$. We define the **quadratic variation** of $\{X_t\}$ with respect to Π by

$$QV(X_t; \Pi) = \sum_{i=1}^{n} (X_{t_i} - X_{t_{i-1}})^2.$$

Usually, we choose Π to be the nth dyadic partition $\mathcal{D}_n[0, T]$ because we want the partition to be dense and the variation to be summable. Using the Weak Law of Large Numbers, the Chebyshev Inequality, and the Borel-Cantelli Lemma, we can show that as the dyadic partition becomes finer, the quadratic variation of a Brownian path is well-behaved. Again, we omit the details.

Theorem 2.37. *For a standard Brownian motion $\{W_t\}$ and any $T > 0$,*

$$\lim_{n\to\infty} QV(W_t; \mathcal{D}_n[0, T]) = T$$

with probability one.

Proof. See [5], Section 10. □

We will use this result repeatedly later when we construct the Itô integral.

3. Stochastic Calculus and Itô's Formula

Now we have introduced Brownian motion and the white noise dW_t. However, not all random processes can be modeled by Brownian motion, since a Brownian motion starting at $x > 0$ will eventually become negative. Therefore, for positive random processes, like the evolution of stock prices, we need to come up with a better model. Black and Scholes suggested the following model:

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t \tag{3.1}$$

where S_t is the stock price, μ is the instantaneous rate of return, and σ is the **volatility**. The stochastic process $\{S_t\}$ is called a geometric Brownian motion. However, we cannot interpret (3.1) as an ordinary differential equation because the Brownian path is nowhere differentiable. As a result, we introduce the idea of stochastic calculus developed by Itô.

3.1. Construction of Stochastic Integral.

Consider the Riemann integral. We first define the integral of a step function and then extend the definition to a larger class of functions, the Riemann-integrable functions, by approximation.

The Itô integral is defined similarly. We first define the integral of simple processes, which are just random step functions. We then extend the integral to other processes by approximation. Since our purpose is to understand the Itô integral and use Itô's formula to solve stochastic differential equations, the details of the approximation procedure will mostly be omitted.

Definition 3.2. A stochastic process $\{\theta_t\}$ is a **simple process** if for some finite sequence of time $0 < t_1 < t_2 < \cdots < t_m < \infty$ and random variables $\xi_0, \xi_1, \ldots, \xi_{m-1}$ such that each ξ_i is $\mathcal{F}_{t_i}$-measurable, we have

$$\theta_t = \sum_{i=0}^{m-1} \xi_i \cdot I_{(t_i,t_{i+1}]}(t). \tag{3.3}$$

Definition 3.4. For a simple process $\{\theta_t\}$ satisfying (3.3), define the Ito integral of $\{\theta_t\}$ as

$$\int \theta_s dW_s = \sum_{i=0}^{m-1} \xi_i (W_{t_{i+1}} - W_{t_i})$$

Why does this definition make sense? As mentioned before, we should interpret dW_t as the infinitesimal increments along the Brownian path. Therefore, when we add up all the increments in one time period $(t_i, t_{i+1}]$, we should get ξ_i times the total increment over the time period, which is exactly $W_{t_{i+1}} - W_{t_i}$.

Such a definition renders the following properties.

Proposition 3.5. *The mean and variance of the Ito integral in Definition 3.4 are*

$$\mathbb{E}\left(\int \theta_s dW_s\right) = 0$$

$$\mathbb{E}\left(\int \theta_s dW_s\right)^2 = \int \mathbb{E}(\theta_s^2) ds$$

Proposition 3.6. *(Linearity) If $\{\theta_t\}$ and $\{\eta_t\}$ are two simple processes and $a, b \in \mathbb{R}$ are two scalars, then*

$$\int (a\theta_s + b\eta_s) dW_s = a \int \theta_s dW_s + b \int \eta_s dW_s$$

Now we can extend the definition to other processes by approximation. The most important class of stochastic processes is the $\mathcal{H}^2$ processes.

Definition 3.7. An adapted process $\{\theta_t\}$ is of the class $\mathcal{H}^p$ for $1 \leq p < \infty$, if

$$\int \mathbb{E}|\theta_s|^p ds < \infty$$

Theorem 3.8. *Any $\mathcal{H}^2$ process $\{\theta_t\}$ can be approximated arbitrarily closely by simple processes, i.e., there is a sequence of simple processes $\{\theta_t^{(n)}\}$ such that*

$$\lim_{n\to\infty} \int \mathbb{E}(\theta_s - \theta_s^{(n)})^2 ds = 0 \tag{3.9}$$

Proof. See [6], Section 3.2. □

The definition of the integral of $\mathcal{H}^2$ processes follows naturally.

Definition 3.10. Let $\{\theta_t\} \in \mathcal{H}^2$ and let $\{\theta_t^{(n)}\}$ be any sequence of simple processes such that (3.9) holds. Define

$$\int \theta_s dW_s = \lim_{n\to\infty} \int \theta_s^{(n)} dW_s \tag{3.11}$$

Theorem 3.12. *The limit in (3.11) exists in the $\mathcal{L}^2$ sense and is independent of the sequence of approximating simple processes.*

Remark 3.13. $\mathcal{L}^2$ is the space of random variables with finite second moments. The Riesz-Fisher Theorem shows that $\mathcal{L}^2$ is complete.

Proof. Here we only prove existence since uniqueness can be shown similarly.

Suppose $\{\theta_t^{(n)}\}$ is a sequence of simple processes satisfying (3.10). Then $\forall \varepsilon > 0$, $\exists N_\varepsilon$ large enough such that $\forall n \geq N_\varepsilon$,

$$\int \mathbb{E}(\theta_s^{(n)} - \theta_s)^2 ds < \varepsilon$$

By the triangle inequality of the $\mathcal{L}^2$ metric, for $n, m \geq N_\varepsilon$,

$$\begin{aligned}\left(\int \mathbb{E}(\theta_s^{(n)} - \theta_s^{(m)})^2 ds\right)^{1/2} &\leq \left(\int \mathbb{E}(\theta_s^{(n)} - \theta_s)^2 ds\right)^{1/2} + \left(\int \mathbb{E}(\theta_s^{(m)} - \theta_s)^2 ds\right)^{1/2} \\ &\leq 2\sqrt{\varepsilon}\end{aligned}$$

Then, by Proposition 3.5 and 3.6,

$$\mathbb{E}\left(\int \theta_s^{(n)} dW_s - \int \theta_s^{(m)} dW_s\right)^2 = \int \mathbb{E}(\theta_s^{(n)} - \theta_s^{(m)})^2 ds < 4\varepsilon$$

Hence the sequence $\int \theta_s^{(n)} dW_s$ is a Cauchy sequence. Since $\mathcal{L}^2$ is complete,

$$\lim_{n\to\infty} \int \theta_s^{(n)} dW_s$$

exists. □

Of course, we also need to define the integral over time periods other than $(0, \infty)$.

Definition 3.14. Define

$$\int_a^b \theta_s dW_s = \int \theta_s^{(a,b]} dW_s$$

where $\{\theta_t^{(a,b]}\}$ is the adapted $\mathcal{H}^2$ process defined by

$$\theta_t^{(a,b]} = \begin{cases} \theta_t & a < t \leq b \\ 0 & \text{otherwise} \end{cases}$$

Although it is necessary to define the Itô integral of $\mathcal{H}^2$ processes, the most important definition is still the Itô integral for simple processes. Almost all basic properties are derived by considering simple processes first and extending to $\mathcal{H}^2$ processes by approximation. For example, Proposition (3.5) and (3.6) also hold for $\mathcal{H}^2$ processes.

Another important property of the Itô integral is that it is a martingale. It is easy to see from Definition 3.4, using the properties of conditional expectation and the independent increments property of Brownian motion, that the Itô integral of a simple process is a martingale. Then a routine limiting argument extends the property to $\mathcal{H}^2$ processes. The complete proof is left as an exercise for interested readers.

Proposition 3.15. *For an adapted $\mathcal{H}^2$ process $\{\theta_t\}$, let $I_t = \int_0^t \theta_s dW_s$. Then the stochastic process $\{I_t\}$ is a martingale, and*

$$I_t = \mathbb{E}\left(\int \theta_s dW_s \middle| \mathcal{F}_t \right)$$

Finally, we can do some calculations. We will work through a simple example using the definition. After this example, one should understand how the Itô integral differs from the Riemann integral, and how painstaking it is to do calculation by definition (without Itô's formula).

Example 3.16. $\displaystyle\int_0^1 W_s dW_s$

Step 1: consider the process

$$\theta_s = W_s I_{[0,1]}(s).$$

This process is adapted to the natural filtration and is in $\mathcal{H}^2$, since

$$\int_0^\infty \mathbb{E}(\theta_s^2) ds = \int_0^1 \mathbb{E}(W_s^2) ds = \int_0^1 s ds = \frac{1}{2} < \infty$$

Step 2: consider the approximation

$$\theta_s^{(n)} = \sum_{k=0}^{2^n} \theta_{k/2^n} I_{[k/2^n,(k+1)/2^n)}(s).$$

We claim that

$$\lim_{n \to \infty} \int_0^\infty \mathbb{E}(\theta_s - \theta_s^{(n)})^2 ds = 0$$

since

$$\begin{aligned}\int_0^\infty \mathbb{E}(\theta_s - \theta_s^{(n)})^2 ds &= \sum_{k=0}^{2^n-1} \int_{k/2^n}^{(k+1)/2^n} \mathbb{E}(\theta_s - \theta_s^{(n)})^2 ds \\ &= \sum_{k=0}^{2^n-1} \int_{k/2^n}^{(k+1)/2^n} \mathbb{E}(W_s - W_{k/2^n})^2 ds \\ &= \sum_{k=0}^{2^n-1} \int_{k/2^n}^{(k+1)/2^n} (s - k/2^n) ds \\ &= \sum_{k=0}^{2^n-1} \left(\frac{s^2}{2} - \frac{ks}{2^n}\right)\Bigg|_{s=k/2^n}^{s=(k+1)/2^n} \\ &= \sum_{k=0}^{2^n-1} 2^{-(2n+1)} \\ &= \frac{2^n}{2^{2n+1}} \longrightarrow 0\end{aligned}$$

as $n \to \infty$.

Step 3: by Definition 3.4,

$$\int \theta_s^{(n)} dW_s = \sum_{k=0}^{2^n-1} W_{k/2^n}(W_{(k+1)/2^n} - W_{k/2^n})$$

To evaluate this sum, we use the technique of "summation by parts". Notice

$$\begin{aligned}W_1^2 &= \sum_{k=0}^{2^n-1} (W_{(k+1)/2^n}^2 - W_{k/2^n}^2) \\ &= \sum_{k=0}^{2^n-1} (W_{(k+1)/2^n} - W_{k/2^n})(W_{(k+1)/2^n} + W_{k/2^n}) \\ &= \sum_{k=0}^{2^n-1} (W_{(k+1)/2^n} - W_{k/2^n})(W_{(k+1)/2^n} - W_{k/2^n} + 2W_{k/2^n}) \\ &= \sum_{k=0}^{2^n-1} (W_{(k+1)/2^n} - W_{k/2^n})^2 + 2\sum_{k=0}^{2^n-1} W_{k/2^n}(W_{(k+1)/2^n} - W_{k/2^n})\end{aligned}$$

where the first sum is the Quadratic Variation $QV(W_s; \mathcal{D}_n[0,1])$ and the second sum is the Itô integral $\int \theta_s^{(n)} dW_s$. By Theorem 2.37, as $n \to \infty$, we have

$$\begin{aligned}W_1^2 &= \lim_{n\to\infty} QV(W_s; \mathcal{D}_n[0,1]) + 2\lim_{n\to\infty} \int \theta_s^{(n)} dW_s \\ &= 1 + 2\int_0^1 W_s dW_s\end{aligned}$$

Hence

$$\int_0^1 W_s dW_s = \frac{1}{2}(W_1^2 - 1) \tag{3.17}$$

Remark 3.18. We can see that the Itô integral is fundamentally different from the Riemann integral, since the latter gives

$$\int_0^1 W_s dW_s = \int_0^1 W_s W_s' ds = \frac{W_s^2}{2}\bigg|_0^1 = \frac{W_1^2}{2}$$

3.2. **Itô's Formula.**

From Example 3.16, we should see that it is impractical to use the definition to calculate every integral. Fortunately, we have Itô's formula, the fundamental theorem of stochastic calculus. Itô's formula takes many different forms and holds at several levels of generality. We start with the simplest version.

Theorem 3.19. *Let $u(x,t)$ be a function of $x \in \mathbb{R}$ and $t \geq 0$ that is twice differentiable in x and C^1 in t. Let W_t be a standard Brownian motion. Then*

$$u(W_t,t) - u(0,0) = \int_0^t u_x(W_s,s)dW_s + \int_0^t u_t(W_s,s)ds + \frac{1}{2}\int_0^t u_{xx}(W_s,s)ds$$

For simplicity, let $t = 1$. Here we assume that u_t and u_{xx} are uniformly bounded and uniformly continuous. Although the theorem is true without these assumptions, the general case requires further tedious approximation arguments. For each $n \in \mathbb{N}$, let $\mathcal{D}_n$ be the nth dyadic partition of the unit interval $[0,1]$. Let $t_i = i/2^n$ denote the dyadic rationals.

To prove the theorem, we need the following lemma.

Lemma 3.20. *For any uniformly bounded, uniformly continuous, nonnegative function $f(x,t)$,*

$$\lim_{n\to\infty} \sum_{i=0}^{2^n-1} f(W_{t_i},t_i)(W_{t_{i+1}} - W_{t_i})^2 = \int_0^1 f(W_t,t)dt \tag{3.21}$$

and

$$\lim_{n\to\infty} \sum_{i=0}^{2^n-1} o[(W_{t_{i+1}} - W_{t_i})^2] = 0 \tag{3.22}$$

where $o(y)$ means that $o(y)/y \to 0$ as $n \to \infty$.

Proof. Notice that if f is constant, then the lemma will follow directly from the Quadratic Variation Formula. For f not constant, since it is continuous, it is nearly constant over short period of time. More precisely, for any $\varepsilon > 0$, there exists $\delta > 0$ such that

$$|s-t| < \delta \quad\Longrightarrow\quad |f(W_s,s) - f(W_t,t)| < \varepsilon$$

with probability at least $1-\varepsilon$. We can take $\delta = 1/2^m$ for some integer m. Then over each time interval of length 2^{-m}, we can approximate $f(W_t,t)$ by its value at the beginning of the interval. That is, for $n > m$,

$$\begin{aligned}&\sum_{i=0}^{2^n-1} f(W_{t_i},t_i)(W_{t_{i+1}} - W_{t_i})^2 \\ &= \sum_{i=0}^{2^m-1} f(W_{i/2^m}, i/2^m)\left[\sum_{j=0}^{2^{n-m}-1}(W_{(i/2^m)+(j/2^n)} - W_{(i/2^m)+[(j-1)/2^n]})^2\right] + \text{error}\end{aligned}$$

where, as $n \to \infty$,

$$\text{error} \le \varepsilon \sum_{i=0}^{2^n-1} (W_{t_{i+1}} - W_{t_i})^2 \approx \varepsilon$$

by the Quadratic Variation Formula. Notice that each inner sum above is also of the Quadratic Variation type, so as $n \to \infty$, each inner sum converges to $1/2^m$. Hence for each m,

$$\limsup_{n\to\infty} \left| \sum_{i=0}^{2^n-1} f(W_{t_i}, t_i)(W_{t_{i+1}} - W_{t_i})^2 - 2^{-m} \sum_{i=0}^{2^m-1} f(W_{i/2^m}, i/2^m) \right| \le \varepsilon$$

Notice that

$$\lim_{m\to\infty} 2^{-m} \sum_{i=0}^{2^m-1} f(W_{i/2^m}, i/2^m) = \int_0^1 f(W_t, t)dt$$

by the definition of the Riemann integral. Hence

$$\limsup_{n\to\infty} \left| \sum_{i=0}^{2^n-1} f(W_{t_i}, t_i)(W_{t_{i+1}} - W_{t_i})^2 - \int_0^1 f(W_t, t)dt \right| \le \varepsilon$$

which proves (3.21). (3.22) can be shown using similar arguments. $\square$

Proof of Theorem 3.19. We first write $u(W_1, 1) - u(0,0)$ as a telescoping sum:

$$\begin{aligned} u(W_1,1) - u(0,0) &= \sum_{i=1}^{2^n} [u(W_{t_i}, t_i) - u(W_{t_{i-1}}, t_{i-1})] \\ &= \sum_{i=1}^{2^n} [u(W_{t_i}, t_i) - u(W_{t_i}, t_{i-1})] + \sum_{i=1}^{2^n} [u(W_{t_i}, t_{i-1}) - u(W_{t_{i-1}}, t_{i-1})] \\ &:= S_1^{(n)} + S_2^{(n)} \end{aligned}$$

For the first sum $S_1^{(n)}$, the first order Taylor expansion of $u(x,t)$ in t gives

$$\begin{aligned} u(W_{t_i}, t_i) - u(W_{t_i}, t_{i-1}) &= u_t(W_{t_i}, t_i)(t_i - t_{i-1}) + o(t_i - t_{i-1}) \\ &= u_t(W_{t_i}, t_i)2^{-n} + o(2^{-n}) \end{aligned}$$

Since u_t is uniformly bounded, the error term $o(2^{-n})$ is uniformly small. Therefore, when 2^n these errors are summed over all i, the result is $o(1)$. Then

$$S_1^{(n)} = 2^{-n} \sum_{i=1}^{2^n} u_t(W_{t_i}, t_i) + o(1) \longrightarrow \int_0^1 u_t(W_t, t)dt$$

For the second sum $S_2^{(n)}$, the second order Taylor expansion of $u(x,t)$ in x gives

$$\begin{aligned} u(W_{t_i}, t_{i-1}) - u(W_{t_{i-1}}, t_{i-1}) &= u_x(W_{t_{i-1}}, t_{i-1})(W_{t_i} - W_{t_{i-1}}) \\ &\quad + \frac{1}{2} u_{xx}(W_{t_{i-1}}, t_{i-1})(W_{t_i} - W_{t_{i-1}})^2 + o[(W_{t_i} - W_{t_{i-1}})^2] \end{aligned}$$

By Lemma 3.20 and the definition of the Itô integral, as $n \to \infty$,

$$\sum_{i=1}^{2^n} u_x(W_{t_{i-1}}, t_{i-1})(W_{t_i} - W_{t_{i-1}}) \longrightarrow \int_0^1 u_x(W_s, s)dW_s$$

$$\sum_{i=1}^{2^n} u_{xx}(W_{t_{i-1}}, t_{i-1})(W_{t_i} - W_{t_{i-1}})^2 \longrightarrow \int_0^1 u_{xx}(W_s, s)ds$$

$$\sum_{i=1}^{2^n} o[(W_{t_i} - W_{t_{i-1}})^2] \longrightarrow 0$$

Hence

$$\begin{aligned} u(W_1, 1) - u(0, 0) &= \lim_{n\to\infty} S_1^{(n)} + S_2^{(n)} \\ &= \int_0^1 u_t(W_s, s)ds + \int_0^1 u_x(W_s, s)dW_s + \frac{1}{2}\int_0^1 u_{xx}(W_s, s)ds \end{aligned}$$

□

Before we present the more general form of Itô's Formula, we first show the power of this result. We start by revisiting Example 3.16.

Example 3.23. Let $u(x, t) = x^2 - t$. Then $u_x(x, t) = 2x$, $u_t(x, t) = -1$, $u_{xx}(x, t) = 2$. By Ito's formula,

$$W_t^2 - t = \int_0^t 2W_s dW_s + \int_0^t -1ds + \frac{1}{2}\int_0^t 2ds = 2\int_0^t W_s dW_s$$

Hence $\int_0^t W_s dW_s = \frac{1}{2}(W_t^2 - t)$. In particular, $\int_0^1 W_s dW_s = \frac{1}{2}(W_1^2 - 1)$.

We get the same answer but with so much less trouble. The next example gives the geometric Brownian motion satisfying (3.1).

Example 3.24. Let $u(x, t) = e^{\alpha x + \beta t}$. Then $u_x(x, t) = \alpha u(x, t)$, $u_t(x, t) = \beta u(x, t)$, $u_{xx}(x, t) = \alpha^2 u(x, t)$. By Ito's formula,

$$u(W_t, t) = 1 + \alpha\int_0^t u(x, s)dW_s + \left(\beta + \frac{1}{2}\alpha^2\right)\int_0^t u(x, s)ds$$

Let $u(W_t, t) = S_t$. Then we can write the above equation in differential form:

$$\frac{dS_t}{S_t} = \alpha dW_t + \left(\beta + \frac{1}{2}\alpha^2\right)dt$$

In particular, if $\alpha = \sigma$ and $\beta = \mu - \sigma^2/2$, then

$$S_t = \exp\left\{\sigma W_t + \left(\mu - \frac{\sigma^2}{2}\right)t\right\} \tag{3.25}$$

is the geometric Brownian motion satisfying (3.1).

Often times, we want to integrate against stochastic processes that are more complex than Brownian motion. Such processes are usually given by the stochastic differential equation

$$dZ_t = A_t dt + B_t dW_t \tag{3.26}$$

If $\{Z_t\}$ solves the SDE, then it is called an Itô process.

Definition 3.27. An Ito process is a stochastic process of the form

$$Z_t = Z_0 + \int_0^t A_s ds + \int_0^t B_s dW_s$$

where for each $t < \infty$, the process $\{A_s I_{(0,t]}(s)\}$ is in $\mathcal{H}^1$ and the process $\{B_s I_{(0,t]}(s)\}$ is in $\mathcal{H}^2$.

To compute integrals against Itô processes, we make the following generalization of Itô's formula. The proof is similar to the proof of Theorem 3.19 and is thus omitted.

Theorem 3.28. *Let $\{Z_t\}$ be an Ito process satisfying the SDE (3.26). Let $u(x,t)$ be twice differentiable in x and C^1 in t. Define $U_t = u(Z_t, t)$. Then $\{U_t\}$ is an Ito process that satisfies the SDE*

$$\begin{aligned} dU_t &= u_x(Z_t,t)dZ_t + u_t(Z_t,t)dt + \frac{1}{2}u_{xx}(Z_t,t)B_t^2 dt \\ &= u_x(Z_t,t)B_t dW_t + \left(u_x(Z_t,t)A_t + \frac{1}{2}u_{xx}(Z_t,t)B_t^2 + u_t(Z_t,t)\right)dt \end{aligned}$$

We now give another useful formula, Itô's Product Rule, again without proof.

Theorem 3.29. *(Itô's Product Rule) Suppose*

$$\begin{cases} dX_t = A_t dt + B_t dW_t \\ dY_t = C_t dt + D_t dW_t \end{cases}$$

Then

$$d(X_t Y_t) = Y_t dX_t + X_t dY_t + B_t D_t dt$$

With these tools in hand, we can finally start to study stochastic differential equations.

4. Stochastic Differential Equations

So far, we have already seen several stochastic differential equations, such as the geometric Brownian motion (3.1) and Itô processes (3.26). As we mentioned in the introduction, SDE is an extension of ODE allowing randomness in the system. Therefore, we should expect the two kinds of differential equations to have similarities to a certain degree. In this section, we first give the definition of the solution concept of SDE and work through a few examples. Then we prove the Existence and Uniqueness Theorem. Finally, we present techniques of solving linear SDEs and give the general formulas.

4.1. **Definition and Examples.**

Definition 4.1. We say a stochastic process X_t is a solution of the Itô stochastic differential equation

$$\begin{cases} dX_t = A(X_t,t)dt + B(X_t,t)dW_t \\ X_0 = x \end{cases} \tag{4.2}$$

if for all $0 \le t \le T$,

- X_t is $\mathcal{F}_t$-measurable,
- $A(X_t,t) \in \mathbb{L}^1[0,T]$ and $B(X_t,t) \in \mathbb{L}^2[0,T]$,

- $X_t = x + \int_0^t A(X_s, s)ds + \int_0^t B(X_s, s)dW_s$ *a.s.*.

Remark 4.3. (1) $\mathbb{L}^1[0,T]$ denotes the space of all real-valued, adaptive processes $\{X_t\}$ such that

$$\mathbb{E}\left(\int_0^T |X_t|dt\right) < \infty$$

(2) $\mathbb{L}^2[0,T]$ denotes the space of all real-valued, adaptive processes $\{Y_t\}$ such that

$$\mathbb{E}\left(\int_0^T |Y_t|^2 dt\right) < \infty$$

Here are a few simple examples of stochastic differential equations. For now we will only show existence. The uniqueness is proved later by the Existence and Uniqueness Theorem.

Example 4.4. Let f be a continuous function. The unique solution of the SDE

$$\begin{cases} dX_t = f(t)X_t dW_t \\ X_0 = 1 \end{cases}$$

is

$$X_t = \exp\left\{-\frac{1}{2}\int_0^t f(s)^2 ds + \int_0^t f(s)dW_s\right\}$$

Proof. To verify this, let

$$Y_t = -\frac{1}{2}\int_0^t f(s)^2 ds + \int_0^t f(s)dW_s$$

Then

$$dY_t = -\frac{1}{2}f(s)^2 dt + f(s)dW_t$$

Applying Itô's formula to $u(y,t) = e^y$, we have

$$\begin{aligned} dX_t &= u_y(Y_t,t)dY_t + \frac{1}{2}u_{yy}(Y_t,t)g(t)^2 dt \\ &= e^{Y_t}\left(-\frac{1}{2}f(t)^2 dt + f(t)dW_t\right) + \frac{1}{2}e^{Y_t}f(t)^2 dt \\ &= f(t)X_t dW_t \end{aligned}$$

□

Example 4.5. Let f and g be continuous functions. The unique solution of the SDE

$$\begin{cases} dX_t = f(t)X_t dt + g(t)X_t dW_t \\ X_0 = 1 \end{cases}$$

is

$$X_t = \exp\left\{\int_0^t \left(f(s) - \frac{1}{2}g(s)^2\right) ds + \int_0^t g(s)dW_s\right\}$$

Proof. To verify this, notice that the SDE implies

$$\frac{dX_t}{X_t} = f(t)dt + g(t)dW_t$$

Applying Itô's formula to $u(x,t) = \ln x$, we have

$$\begin{aligned} d\ln X_t &= \frac{dX_t}{X_t} - \frac{1}{2}\frac{g(t)^2X_t^2}{X_t^2}dt \\ &= \left(f(t) - \frac{1}{2}g(t)^2\right)dt + g(t)dW_t \end{aligned}$$

i.e.,

$$\ln X_t = \int_0^t \left(f(s) - \frac{1}{2}g(s)^2\right)ds + \int_0^t g(s)dW_s$$

□

Notice that this example also implies the stock price formula we derived in Example 3.24.

Example 4.6. (Stock Prices) Let S_t be the price of stock at time t. We model the price change over time by the SDE

$$\begin{cases} \dfrac{dS_t}{S_t} = \mu dt + \sigma dW_t \\ S_0 = s_0 \end{cases}$$

where μ is the drift and σ is the volatility of the stock. Then by the last example,

$$d\ln S_t = \left(\mu - \frac{1}{2}\sigma^2\right)dt + \sigma dW_t$$

Hence

$$\begin{aligned} S_t &= s_0 \exp\left\{\int_0^t \left(\mu - \frac{1}{2}\sigma^2\right)ds + \int_0^t \sigma dW_s\right\} \\ &= s_0 \exp\left\{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t\right\} \end{aligned}$$

which agrees with (3.25). Notice that the stock price is always positive as long as $s_0 > 0$.

Remark 4.7. The SDE in Example 4.6 implies

$$S_t = s_0 + \int_0^t \mu S_s dt + \int_0^t \sigma S_s dW_s$$

Since $\mathbb{E}\left(\int_0^t \sigma S_s dW_s\right) = 0$, we have

$$\mathbb{E}S_t = s_0 + \int_0^t \mu \mathbb{E}S_s ds$$

Hence the expected stock price is

$$\mathbb{E}S_t = s_0 e^{\mu t}$$

which agrees with the deterministic solution corresponding to $\sigma = 0$.

4.2. **Existence and Uniqueness.**

The Existence and Uniqueness Theorem is the most important theorem in this paper. Similar to Picard's Existence Theorem for ODE, this theorem also builds iterative solutions and applies the fixed point theorem. We start with a simple case.

Problem 4.8. Let $b : \mathbb{R} \to \mathbb{R}$ be a C^1 function with $|b'| \le L$ for some constant L. We want to solve the SDE

$$\begin{cases} dX_t = b(X_t)dt + dW_t \\ X_0 = x \end{cases} \tag{4.9}$$

for some $x \in \mathbb{R}$.

Notice this SDE means that for all t,

$$X_t = x + \int_0^t b(X_s)ds + W_t$$

Similar to building solutions for ODE, we use a successive approximation method. Define $X_t^0 \equiv x$ and

$$X_t^{n+1} := x + \int_0^t b(X_s^n)ds + W_t$$

for $n \in \mathbb{N}$. Define

$$D^n(t) := \max_{0 \le s \le t} |X_s^{n+1} - X_s^n|$$

For a given sample Brownian path, we have

$$D^0(t) = \max_{0 \le s \le t} \left| \int_0^s b(x)dr + W_s \right| \le C$$

for some constant C. We claim that

$$D^n(t) \le C \frac{L^n}{n!} t^n$$

because

$$\begin{aligned} D^n(t) &= \max_{0 \le s \le t} \left| \int_0^s [b(X_r^n) - b(X_r^{n-1})]dr \right| \\ &\le L \int_0^t D^{n-1}(s)ds \\ &\le L \int_0^t C \frac{L^{n-1}s^{n-1}}{(n-1)!} ds \\ &= C \frac{L^n t^n}{n!} \end{aligned}$$

Given the claim, we have, for $m \ge n$,

$$\max_{0 \le s \le t} |X_t^m - X_t^n| \le C \sum_{k=n}^{\infty} \frac{L^k T^k}{k!} \longrightarrow 0$$

as $n \to \infty$. Hence X_t^n converges uniformly to a limit process X_t *a.s.*, where it is easy to check that X_t solves the SDE (4.9).

We now proceed to prove the general Existence and Uniqueness Theorem.

Theorem 4.10. *(Existence and Uniqueness Theorem) Let $A, B : \mathbb{R} \times [0,T] \to \mathbb{R}$ be continuous functions satisfying*

$$\begin{cases} |A(x,t) - A(y,t)| \le \delta |x-y| \\ |B(x,t) - B(y,t)| \le \delta |x-y| \end{cases} \qquad \forall x, y \in \mathbb{R}, t \in [0,T]$$

$$\begin{cases} |A(x,t)| \le \delta(1+|x|) \\ |B(x,t)| \le \delta(1+|x|) \end{cases} \qquad \forall x \in \mathbb{R}, t \in [0,T]$$

for some constant δ. Let X be a random variable independent of the filtration $\mathcal{F}_0$ relative to a standard Brownian motion $\{W_t\}$, such that

$$\mathbb{E}|X|^2 < \infty$$

Then there exists a unique solution $X_t \in \mathbb{L}^2(0,T)$ of the SDE

$$\begin{cases} dX_t = A(X_t,t)dt + B(X_t,t)dW_t \\ X_0 = X \end{cases}$$

Remark 4.11. Here, "unique" means that if X_t and Y_t are both solutions of the SDE, then for all $0 \le t \le T$,

$$\mathbb{P}(X_t = Y_t) = 1$$

To prove the Existence and Uniqueness Theorem, we first need Gronwall's Lemma, which builds a function which satisfies an integral inequality by the solution of the corresponding integral equation.

Lemma 4.12. *(Gronwall's Lemma) Let ϕ and f be nonnegative, continuous functions defined for $0 \le t \le T$. If*

$$\phi(t) \le C + \int_0^t f(s)\phi(s)ds$$

for some constant C and all $0 \le t \le T$, then

$$\phi(t) \le C \exp\left\{ \int_0^t f(s)ds \right\}$$

for all $0 \le t \le T$.

Proof. Let $\Phi(t) = C + \displaystyle\int_0^t f(s)\phi(s)ds$. Then $\Phi(0) = C$, $\phi(t) \le \Phi(t)$, and

$$\Phi'(t) = f(t)\phi(t) \le f(t)\Phi(t)$$

for all $0 \le t \le T$. Let $F(t) := \Phi(t) \exp\left\{ -\displaystyle\int_0^t f(s)ds \right\}$. Notice

$$\begin{aligned} F'(t) &= [\Phi'(t) - f(t)\Phi(t)] \exp\left\{ -\int_0^t f(s)ds \right\} \\ &\le [f(t)\Phi(t) - f(t)\Phi(t)] \exp\left\{ -\int_0^t f(s)ds \right\} \\ &= 0 \end{aligned}$$

Then F is decreasing in t. In particular, $F(t) \le F(0)$, i.e.,

$$\Phi(t) \exp\left\{-\int_0^t f(s)ds\right\} \le C$$

Hence

$$\phi(t) \le \Phi(t) \le C \exp\left\{\int_0^t f(s)ds\right\}$$

□

The proof of the Existence and Uniqueness Theorem goes like this: For uniqueness, we prove by contradiction using Gronwall's Lemma. For existence, similar to the procedure in solving Problem 4.8, we build iterative solutions and use the Martingale Inequality to give estimates on the difference between successive iterations. Then we use the Borel-Cantelli Lemma to show that the iterations converge to a process X_t that solves the SDE. Finally, we prove the process X_t is in $\mathbb{L}^2$.

Proof of Theorem 4.10. We first show uniqueness. Suppose X_t and Y_t are two solutions. Then

$$X_t - Y_t = \int_0^t [A(X_s, s) - A(Y_s, s)]ds + \int_0^t [B(X_s, s) - B(Y_s, s)]dW_s$$

for all $0 \le t \le T$. Since $(a+b)^2 = a^2 + b^2 + 2ab \le 2a^2 + 2b^2$, we have

$$\mathbb{E}|X_t - Y_t|^2 \le 2\mathbb{E}\left|\int_0^t [A(X_s, s) - A(Y_s, s)]ds\right|^2 + 2\mathbb{E}\left|\int_0^t [B(X_s, s) - B(Y_s, s)]dW_s\right|^2$$

The Cauchy-Schwarz inequality gives that for any $t > 0$ and $f : [0, t] \to \mathbb{R}$,

$$\left|\int_0^t f(s)ds\right|^2 \le t\int_0^t |f(s)|^2 ds$$

Hence for the first expectation,

$$\begin{aligned}\mathbb{E}\left|\int_0^t [A(X_s, s) - A(Y_s, s)]ds\right|^2 &\le T\mathbb{E}\left(\int_0^t |A(X_s, s) - A(Y_s, s)|^2 ds\right)\\ &\le T\delta^2 \int_0^t \mathbb{E}|X_s - Y_s|^2 ds\end{aligned}$$

and for the second expectation,

$$\begin{aligned}\mathbb{E}\left|\int_0^t [B(X_s, s) - B(Y_s, s)]dW_s\right|^2 &= \mathbb{E}\left(\int_0^t |B(X_s, s) - B(Y_s, s)|^2 ds\right)\\ &\le \delta^2 \int_0^t \mathbb{E}|X_s - Y_s|^2 ds\end{aligned}$$

Then we can find a constant C such that

$$\mathbb{E}|X_t - Y_t|^2 \le C\int_0^t \mathbb{E}|X_s - Y_s|^2 ds$$

Let $\phi(t) = \mathbb{E}|X_t - Y_t|^2$. Then we have

$$\phi(t) \le C\int_0^t \phi(s)ds$$

By Gronwall's Lemma, $\phi \equiv 0$. That is, $X_t = Y_t$ *a.s.* for all $0 \leq t \leq T$. Hence the solution of the SDE is unique.

Next we show existence. Define

$$\begin{cases} X_t^0 = X \\ X_t^{n+1} = X + \int_0^t A(X_s^n, s)ds + \int_0^t B(X_s^n, s)dW_s \end{cases}$$

for $n \in \mathbb{N}$ and $0 \leq t \leq T$. Let

$$D^n(t) = \mathbb{E}|X_t^{n+1} - X_t^n|^2$$

We claim that for all $n \in \mathbb{N}$,

$$D^n(t) \leq \frac{(Mt)^{n+1}}{(n+1)!}$$

for some constant M depending on δ, T, and X. For the base case $n = 0$,

$$\begin{aligned} D^0(t) &= \mathbb{E}|X_t^1 - X_t^0|^2 \\ &= \mathbb{E}\left|\int_0^t A(X,s)ds + \int_0^t B(X,s)dW_s\right|^2 \\ &\leq 2\mathbb{E}\left|\int_0^t A(X,s)ds\right|^2 + 2\mathbb{E}\left|\int_0^t B(X,s)dW_s\right|^2 \\ &\leq 2\mathbb{E}\left|\int_0^t \delta(1+|X|)ds\right|^2 + 2\int_0^t \delta^2\mathbb{E}(1+|X|)^2 ds \\ &\leq tM \end{aligned}$$

for some M large enough. Suppose the claim holds for $n-1$. Then

$$\begin{aligned} D^n(t) &= \mathbb{E}|X_t^{n+1} - X_t^n|^2 \\ &= \mathbb{E}\left|\int_0^t [A(X_s^n, s) - A(X_s^{n-1}, s)]ds + \int_0^t [B(X_s^n, s) - B(X_s^{n-1}, s)]dW_s\right|^2 \\ &\leq 2\delta^2(1+T)\mathbb{E}\left(\int_0^t |X_s^n - X_s^{n-1}|^2 ds\right) \\ &\leq 2\delta^2(1+T)\int_0^t \frac{M^n s^n}{n!}ds \\ &\leq \frac{(Mt)^{n+1}}{(n+1)!} \end{aligned}$$

provided $M \geq 2\delta^2(1+T)$. This proves the claim.

Notice that

$$\begin{aligned} \max_{0\leq t\leq T} |X_t^{n+1} - X_t^n|^2 \leq 2T\delta^2 &\int_0^T |X_s^n - X_s^{n-1}|^2 ds \\ &+ 2\max_{0\leq t\leq T}\left|\int_0^T [B(X_s^n, s) - B(X_s^{n-1}, s)]dW_s\right|^2 \end{aligned}$$

Hence the Martingale Inequality implies

$$\begin{aligned}\mathbb{E}\left(\max_{0\le t\le T}|X_t^{n+1}-X_t^n|^2\right) &\le 2T\delta^2\int_0^T \mathbb{E}|X_s^n - X_s^{n-1}|^2 ds \\ &\quad + 8\delta^2\int_0^T \mathbb{E}|X_s^n - X_s^{n-1}|^2 ds \\ &\le C\frac{(MT)^n}{n!}\end{aligned}$$

by the claim above. Since

$$\begin{aligned}\mathbb{P}\left(\max_{0\le t\le T}|X_t^{n+1}-X_t^n|^2 > \frac{1}{2^n}\right) &\le 2^{2n}\mathbb{E}\left(\max_{0\le t\le T}|X_t^{n+1}-X_t^n|^2\right) \\ &\le 2^{2n}C\frac{(MT)^n}{n!}\end{aligned}$$

and

$$\sum_{n=1}^{\infty} 2^{2n}C\frac{(MT)^n}{n!} = C(e^{4MT}-1) < \infty$$

the Borel-Cantelli Lemma applies:

$$\mathbb{P}\left(\max_{0\le t\le T}|X_t^{n+1}-X_t^n|^2 > \frac{1}{2^n}\ i.o.\right) = 0$$

Hence

$$X_t^n = X_t^0 + \sum_{i=0}^{n-1}(X_t^{i+1} - X_t^i)$$

converges uniformly to a process X_t almost surely. By definition of X_t^{n+1},

$$X_t = X_0 + \int_0^t A(X_s,s)ds + \int_0^t B(X_s,s)dW_s$$

i.e.,

$$\begin{cases} dX_t = A(X_t,t)dt + B(X_t,t)dW_t \\ X_0 = X \end{cases}$$

Last, we need to show $X_t \in \mathbb{L}^2$. We have

$$\begin{aligned}\mathbb{E}|X_t^{n+1}|^2 &\le C\mathbb{E}|X|^2 + C\mathbb{E}\left|\int_0^t A(X_s^n,s)ds\right|^2 + C\mathbb{E}\left|\int_0^t B(X_s^n,s)dW_s\right|^2 \\ &\le C(1+\mathbb{E}|X|^2) + C\int_0^t \mathbb{E}|X_s^n|^2 ds\end{aligned}$$

for some constant C large enough to satisfy all three parts. By induction,

$$\begin{aligned}\mathbb{E}|X_t^{n+1}|^2 &\le \left(C + C^2 + \cdots + C^{n+2}\frac{t^{n+1}}{(n+1)!}\right)(1+\mathbb{E}|X|^2) \\ &\le C(1+\mathbb{E}|X|^2)e^{Ct}\end{aligned}$$

Letting $n\to\infty$, we have

$$\mathbb{E}|X_t|^2 \le C(1+\mathbb{E}|X|^2)e^{Ct} \le \infty$$

for all $0\le t\le T$. Hence $X_t \in \mathbb{L}^2[0,T]$, which completes the proof. $\square$

Given the Existence and Uniqueness Theorem, we are safe to claim that in all examples in Section 4.1, the solutions are unique. Therefore, the only problem left for solving SDE is to find the unique solution. In this paper, we only present the methods for finding linear stochastic differential equations.

4.3. **Linear Stochastic Differential Equations.**

Definition 4.13. The SDE

$$dX_t = A(X_t, t)dt + B(X_t, t)dW_t$$

is **linear** if A and B are of the form

$$A(x,t) = C(t) + D(t)x \quad \text{and} \quad B(x,t) = E(t) + F(t)x$$

for $C, D, E : [0,T] \to \mathbb{R}$ and $F : [0,T] \to L(\mathbb{R}, \mathbb{R})$, the space of bounded linear maps from $\mathbb{R}$ to $\mathbb{R}$.

In particular, a linear SDE is called **homogeneous** if $C \equiv E \equiv 0$ and is called **linear in the narrow sense** if $F \equiv 0$.

We first present the solution formulas for linear SDE. Then we introduce the technique of breaking the solution process X_t into a product of two processes $Y_t Z_t$ and solving the two processes separately. Using this technique, we can come back and derive the following solution formulas.

Proposition 4.14. *(Linear SDE in the Narrow Sense) First suppose D is a constant. The solution of*

$$\begin{cases} dX_t = (C(t) + DX_t)dt + E(t)dW_t \\ X_0 = X \end{cases}$$

is

$$X_t = e^{Dt}X + \int_0^t e^{D(t-s)}(C(s)ds + E(s)dW_s)$$

More generally, the solution of

$$\begin{cases} dX_t = (C(t) + D(t)X_t)dt + E(t)dW_t \\ X_0 = X \end{cases}$$

is

$$X_t = \Phi(t)\left(X + \int_0^t \Phi(s)^{-1}(C(s)ds + E(s)dW_s)\right)$$

where $\Phi(t)$ is the solution of the ODE

$$\begin{cases} \Phi'(t) = D(t)\Phi(t) \\ \Phi(0) = 1 \end{cases}$$

Proposition 4.15. *(Solutions of general linear SDE) The solution of*

$$\begin{cases} dX_t = (C(t) + D(t)X_t)dt + (E(t) + F(t)X_t)dW_t \\ X_0 = X \end{cases}$$

is

$$X_t = \Phi(t)\left(X + \int_0^t \Phi^{-1}(s)[C(s) - E(s)F(s)]ds\right) + \int_0^t \Phi^{-1}(s)E(s)dW_s$$

where

$$\Phi(t) := \exp\left\{\int_0^t \left(D(s) - \frac{F(s)^2}{2}\right) ds\right\} + \int_0^t F(s)dW_s$$

Now we introduce the technique mentioned above. Itô's formula and Itô's Produce Rule are extensively used in the following derivation.

Example 4.16. Let us revisit Example 4.5 with notations changed according to Definition 4.17. Consider the linear SDE

$$\begin{cases} dX_t = D(t)X_t dt + F(t)X_t dW_t \\ X_0 = X \end{cases} \tag{4.17}$$

We want to find a solution of the form

$$X_t = Y_t Z_t$$

where Y_t satisfies the SDE

$$\begin{cases} dY_t = F(t)Y_t dW_t \\ Y_0 = X \end{cases} \tag{4.18}$$

and Z_t satisfies the SDE

$$\begin{cases} dZ_t = A(t)dt + B(t)dW_t \\ Z_0 = 1 \end{cases} \tag{4.19}$$

where $A(t)$ and $B(t)$ are to be selected. Then by Itô's Product Rule,

$$\begin{aligned} dX_t &= d(Y_t Z_t) \\ &= Y_t dZ_t + Z_t dY_t + F(t)Y_t B(t)dt \\ &= F(t)X_t dW_t + Y_t[dZ_t + F(t)B(t)dt] \end{aligned}$$

Therefore, we want $A(t)$ and $B(t)$ to satisfy

$$dZ_t + F(t)B(t)dt = D(t)Z_t dt$$

which suggests $B \equiv 0$ and $A(t) = D(t)Z_t$. Then (4.19) becomes

$$\begin{cases} dZ_t = D(t)Z_t dt \\ Z_0 = 1 \end{cases}$$

This system is non-random and the solution is

$$Z_t = \exp\left\{\int_0^t D(s)ds\right\}$$

Hence the solution of (4.18) is

$$Y_t = X\exp\left\{\int_0^t F(s)dW_s - \frac{1}{2}\int_0^t F^2(s)ds\right\}$$

and the solution of (4.17) is

$$X_t = Y_t Z_t = X\exp\left\{\int_0^t F(s)dW_s + \int_0^t \left(D(s) - \frac{1}{2}F^2(s)\right) ds\right\}$$

which agrees with the solution in Example 4.5.

Example 4.20. We can also derive the solution formula for general linear SDE, Proposition 4.15, using the same technique. Consider the SDE

$$\begin{cases} dX_t = (C(t) + D(t)X_t)dt + (E(t) + F(t)X_t)dW_t \\ X_0 = X \end{cases}$$

Similarly, we want to find a solution of the form

$$X_t = Y_t Z_t$$

where Y_t satisfies the SDE

$$\begin{cases} dY_t = D(t)Y_t dt + F(t)Y_t dW_t \\ Y_0 = 1 \end{cases}$$

and Z_t satisfies the SDE

$$\begin{cases} dZ_t = A(t)dt + B(t)dW_t \\ Z_0 = X \end{cases}$$

for $A(t)$ and $B(t)$ to be selected. Then

$$\begin{aligned} dX_t &= Z_t dY_t + Y_t dZ_t + F(t)Y_t B(t)dt \\ &= D(t)X_t dt + F(t)X_t dW_t + Y_t[A(t)dt + B(t)dW_t] + F(t)Y_t B(t)dt \end{aligned}$$

Hence we want $A(t)$ and $B(t)$ to satisfy

$$Y_t[A(t)dt + B(t)dW_t] + F(t)Y_t B(t)dt = C(t)dt + E(t)dW_t$$

which suggests

$$\begin{cases} A(t) := [C(t) - F(t)E(t)]Y_t^{-1} \\ B(t) := E(t)Y_t^{-1} \end{cases}$$

By last example, we know $Y_t = \exp\left\{\int_0^t F(s)dW_s + \int_0^t \left(D(s) - \frac{1}{2}F^2(s)\right) ds\right\}$. Hence $Y_t > 0$ $a.s.$. Then

$$Z_t = X + \int_0^t [C(s) - F(s)E(s)]Y_s^{-1} ds + \int_0^t E(s)Y_s^{-1} dW_s$$

Denote Y_t as $\Phi(t)$ as in Proposition 4.15. Then we have

$$X_t = Y_t Z_t = \Phi(t)\left(X + \int_0^t \Phi^{-1}(s)[C(s) - E(s)F(s)]ds\right) + \int_0^t \Phi^{-1}(s)E(s)dW_s$$

which is exactly the formula in Proposition 4.15.

We have proved the Existence and Uniqueness Theorem for stochastic differential equations and presented some techniques for finding the unique solution of linear stochastic differential equations. The natural step forward is to apply the theory of stochastic calculus and stochastic differential equation to real-world problems. As mentioned in the introduction, SDEs are heavily used in quantitative finance, partial differential equations, and statistical physics.

In quantitative finance, SDEs are primarily used for security pricing. For example, the famous **Black-Scholes Formula** for European options pricing is essentially an SDE. However, in order to solve the SDE for the Black-Scholes Formula, we need the Radon-Nikodym Theorem to change the probability measure and the Girsanov

Theorem to drift the stochastic process. Both theorems rely on measure theory and more advanced martingale theory. Interested readers can refer to [2], Chapter 5.

In PDE theory, one interesting problem is the Dirichlet Problem for harmonic functions. That is, given a smooth and bounded domain $U \in \mathbb{R}^n$ and a continuous function $g : \partial U \to \mathbb{R}$, we want to find $u \in C^2(U) \cap C(\overline{U})$ satisfying the PDE

$$\begin{cases} \Delta u = 0 & \text{in } U \\ u = g & \text{on } \partial U \end{cases} \tag{4.21}$$

We can give a probabilistic representation for the solution of the PDE (4.21) using stopping times and Brownian motion. With SDE in hand, we can solve more complex systems like

$$\begin{cases} -\frac{1}{2}\Delta u + cu = f & \text{in } U \\ u = 0 & \text{on } \partial U \end{cases} \tag{4.22}$$

where c and f are smooth functions with $c \geq 0$ in U. The solution of the PDE (4.22) is called the **Feynman-Kac Formula**. Interested readers can refer to [7] for stochastic PDEs.

In statistical physics, SDEs rise naturally in the study of **Schramm-Loewner Evolution (SLE)**, a family of random curves with diffusivity parameter $\kappa \in [0, \infty)$ that arise as a scaling limit candidate for several different lattice models. These models were thought to have no connections, until Oded Schramm discovered that the uniform spanning tree model and the Loop Erased Random Walk model both have this scaling limit with different driving parameters κ. Interested readers can refer to [8].

5. Appendix

5.1. Conditional Expectation.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{G} \subset \mathcal{F}$ be a σ-algebra.

Theorem 5.1. *Let X be a random variable with $\mathbb{E}|X| < \infty$. Then there exists a unique $\mathcal{G}$-measurable random variable $\mathbb{E}(X|\mathcal{G})$ such that, for every bounded $\mathcal{G}$-measurable random variable Y,*

$$\mathbb{E}(XY) = \mathbb{E}(\mathbb{E}(X|\mathcal{G})Y)$$

The unique random variable $\mathbb{E}(X|\mathcal{G})$ is called the **conditional expectation** of X given $\mathcal{G}$. Proof of Theorem 5.1 uses the Radon-Nikodym Theorem. Interested readers can refer to [9], Section 2.

Remark 5.2. In order to verify a random variable $Z = \mathbb{E}(X|\mathcal{G})$, we only need to show that Z is $\mathcal{G}$-measurable and for all $A \in \mathcal{G}$,

$$\mathbb{E}(XI_A) = \mathbb{E}(ZI_A)$$

Using this definition, we can derive the following basic properties of conditional expectation quite easily.

Proposition 5.3. *(Properties of Conditional Expectation)*

- **Linearity**: *$\mathbb{E}(aX + bY|\mathcal{G}) = a\mathbb{E}(X|\mathcal{G}) + b\mathbb{E}(Y|\mathcal{G})$ for all $a, b \in \mathbb{R}$.*
- **Positivity**: *If $X \geq 0$, then $\mathbb{E}(X|\mathcal{G}) \geq 0$.*
- **Stability**: *If Y is $\mathcal{G}$-measurable, then $\mathbb{E}(XY|\mathcal{G}) = Y\mathbb{E}(X|\mathcal{G})$.*
- **Independence Law**: *If X is independent of $\mathcal{G}$, then $\mathbb{E}(X|\mathcal{G}) = \mathbb{E}X$.*

- **Tower Law**: *If $\mathcal{H} \subset \mathcal{G}$, then* $\mathbb{E}(X|\mathcal{H}) = \mathbb{E}(\mathbb{E}(X|\mathcal{G})|\mathcal{H})$.
- **Projection Law**: $\mathbb{E}(\mathbb{E}(X|\mathcal{G})) = X$.
- **Jensen Inequality**: *If $\varphi : \mathbb{R} \to \mathbb{R}$ is convex and $\mathbb{E}|X| < \infty$, then*

$$\mathbb{E}(\varphi(X)) \geq \varphi(\mathbb{E}X) \quad \textit{and} \quad \mathbb{E}(\varphi(X)|\mathcal{G}) \geq \varphi(\mathbb{E}(X|G))$$

5.2. **Borel-Cantelli Lemma.**

Definition 5.4. Let $\{A_n\}$ be a sequence of events in the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then the event

$$\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m = \{\omega \in \Omega \,|\, \omega \text{ belongs to infinitely many of the events } A_n\}$$

is called "A_n infinitely often", or "A_n *i.o.*" for short.

Lemma 5.5. *(Borel-Cantelli Lemma) If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, then $\mathbb{P}\{A_n \ i.o\} = 0$.*

Proof. By definition, for each n,

$$\mathbb{P}\{A_n \ i.o\} \leq \mathbb{P}\left(\bigcup_{m=n}^{\infty} A_m\right) \leq \sum_{m=n}^{\infty} \mathbb{P}(A_m) \to 0$$

as $n \to \infty$ since $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$. $\square$

Acknowledgement

I would like to give special thanks to my mentor, Stephen Yearwood, for his guide through the readings and his help on writing this paper. I also want to thank Professor Peter May for organizing the Math REU. This has been an amazing experience for me.

References

[1] Gregory F. Lawler. Introduction to Stochastic Processes. Chapman & Hall/CRC. 2006.
[2] Gregory F. Lawler. Stochastic Calculus: An Introduction with Applications. http://www.math.uchicago.edu/ lawler/finbook2.pdf
[3] Patrick Billingsley. Convergence of Probability Measures. Wiley-Interscience. 1999.
[4] Lawrence C. Evans. An Introduction to Stochastic Differential Equations. American Mathematical Society. 2014.
[5] Steven P. Lalley. Lecture Notes for Stochastic Calculus and Finance I, Brownian Motion Lecture. 2001. http://galton.uchicago.edu/ lalley/Courses/390/index.html
[6] Ioannis Karatzas & Steven Shreve. Brownian Motion and Stochastic Calculus. Graduate Texts in Mathematics. 1991.
[7] Martin Hairer. An Introduction to Stochastic PDEs. http://www.hairer.org/notes/SPDEs.pdf
[8] Nathanael Berestycki & James R. Norris. Lectures on SchrammLoewner Evolution. http://www.statslab.cam.ac.uk/ james/Lectures/sle.pdf
[9] Steven. P. Lalley. Lecture Notes for Measure-Theoretic Probability II, Conditional Expectation Lecture. 2019. http://galton.uchicago.edu/ lalley/Courses/383/index.html