MEASURE THEORY, STOCHASTIC CALCULUS, AND THE BLACK-SCHOLES-MERTON MODEL

AGUSTÍN ESTEVA

Abstract. This expository paper develops the theory behind stochastic calculus, placing special emphasis on the measure-theoretic "risk-neutral" derivation of the Black-Scholes-Merton equation. Along the way, this paper will explore the topics of the Radon-Nikodym derivative, Brownian motion, the stochastic integral, the Itô-Doeblin formula, and Girsanov's theorem. Basic measure theory and probability theory knowledge is assumed, along with calculus and some differential equations.

Contents

1. Motivation and Introduction

Kiyosi Itô, despite Japanese paper shortages, having to mimeograph work, and publishing his papers in American journals, was able to introduce calculus to the world of random (stochastic) processes when he published his 1944 paper, *Stochastic Integral* [1]. In it, he introduced the Itô integral,

$$I(t) = \int_0^t \Delta_s dB_s, \tag{1.1}$$

Date: AUGUST 28, 2024.

where B_t is a continuous random motion. Seven years later, he established the foundations of stochastic differential calculus by publishing his change of variables formula in [2]. The formula is given by

(1.2)

$$d\varphi(t, X_t) = \left[\partial_t\varphi(t, X_t) + m_t X_t \partial_x \varphi(t, X_t) + \frac{\sigma_t^2 X_t^2}{2}\partial_{xx}\varphi(t, X_t)\right] dt + \sigma_t X_t \partial_x \varphi(t, X_t) dB_t$$

where X_t is a *geometric Brownian motion* satisfying the stochastic differential equation,

$$\begin{cases} dX_t = m_t X_t dt + \sigma_t X_t dB_t \\ X(0) = 0. \end{cases} \tag{1.3}$$

In 1940, on the other side of the war, a 25-year old Wolfgang Doeblin dies in battle, burning his math notes as a last stand against the Nazis. Paul Lévy compared this Frenchman to the likes of Gaolois and Able, but his name was forgotten to history until his Pli (a mathematical black box used in wartime) was opened in 2000. The Pli revealed that 2 years before Itô published his first papers on stochastic differential equations, Doeblin's notes contained his own change of variables formula,

$$d\varphi(X_t, t) = \left[\partial_t\varphi(X_t, t) + mX_t\partial_x\varphi(X_t, t) + \frac{\sigma^2 X_t^2}{2}\partial_{xx}\varphi(X_t, t)\right] dt + d\delta(\overline{H}_t),$$

where $\delta(u)$ is a Brownian motion and $\overline{H}_t = \int_0^t [\sigma(X_s, s)\partial_x\varphi(X_s, s)]^2 \, ds$.[1] Doeblin, depressed by the war, wrote that he was happy during the hours he spent developing his theory of stochastic analysis, which would go unnoticed for 60 years. [3]

The goal of this paper is to understand the importance of their work by taking a measure-theretic approach to stochastic analysis and deriving the Ito-Doeblin formula along with its most important financial application, the Black-Scholes-Merton equation. First, however, we begin by recalling definitions for stochastic processes, proving some basic probability results, and introducing Martingales and conditional expectation via Radon-Nikodym derivatives.

1.1. **Definitions.** Louis Bachalier's doctoral thesis, *The Theory of Speculation* (1900), was the first attempt to use mathematics to model finances. In it, he made use of Brownian motion, or continuous random motion. The following definitions are introduced in order to understand what Brownian motion is, and more generally, what random processes are, and how they can be used to model the markets.

Definition 1.4. A *probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ is a triple where Ω is an arbitrary set, $\mathcal{F}$ is a $\sigma-$algebra of Ω containing events, and $\mathbb{P}$ is a probability measure on $(\Omega, \mathcal{F})$.

[1]The difference between (1.2) and Deoblin's formula was later explained by Dúbins-Shwarz and Dambis in 1965 with their representation of continuous martingales.

For the rest of the paper, it is understood, but not always stated, that we are always acting within a probability space.

Definition 1.5. A *stochastic process* is a sequence of random variables in a probability space.

Definition 1.6. A *filtration* $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$ of $(\Omega, \mathcal{F})$, where $\mathbb{T}$ is the time index set, is a sequence of $\sigma-$algebras, $\mathcal{F}_t \subseteq \mathcal{F}$ such that if $s < t$, then $\mathcal{F}_s \subseteq \mathcal{F}_t$.

It is convenient to think of a filtration as an increasing sequence of the information contained in the random variables. Thus, $\mathcal{F}$ contains all the information that could possibly be known in the probability space.

Remark 1.7. Let $\{X_t\}_{t\in\mathbb{T}}$ be a stochastic process. We say that the *natural filtration* of X_t is the sigma algebra generated by X_t. That is,

$$\mathcal{F}_t = \sigma(X_1, X_2, \ldots, X_t).$$

Definition 1.8. We say a process $\{X_t\}_{t\in\mathbb{T}}$ is *adapted* to a filtration $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$ if X_t is $\mathcal{F}_t-$measurable for each t.

Definition 1.9. A random variable $T : \Omega \to \mathbb{T} \cup \{\infty\}$ is called a *random time.* A random time is defined as a *stopping time* (with respect to the filtration $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$) if

$$\{T \leq t\} = \{\omega : T(\omega) \leq t\} \in \mathcal{F}_t$$

One can think of a stopping time as a random time T such that the event that T has occurred before time t is known by time t.

The next example gives an instance of these definitions with a simple scenario.

Example 1.10. Suppose we throw a fair coin three times, then the sample space is

$$\Omega = \{HHH, HHT, HTH, HTT, TTT, TTH, THT, THH\}.$$

The first toss divides Ω by the events which start with either a head, $\mathcal{A}_H$, or tail, $\mathcal{A}_T$. Thus,

$$\mathcal{F}_1 = \{\Omega, \emptyset, \mathcal{A}_H, \mathcal{A}_T\}.$$

Note that Ω and $\emptyset$ are included in order for $\mathcal{F}_1$ to be a valid $\sigma-$algebra; both of these sets are redundant information at this point, as a toss (which already excludes half of Ω's options) has already been made.

The second toss divides $\mathcal{A}_H$ into two sets, those whose second toss is tails, $\mathcal{A}_{HT}$, and those who's second toss is heads, $\mathcal{A}_{HH}$. Similarly, we can create $\mathcal{A}_{TT}$ and $\mathcal{A}_{TH}$. However, for this to be a valid $\sigma-$algebra, we must take complements and unions,

$$\begin{aligned}\mathcal{F}_2 = \{&\emptyset, \Omega, \mathcal{A}_H, \mathcal{A}_T, \mathcal{A}_{HH}, \mathcal{A}_{HT}, \mathcal{A}_{TH}, \mathcal{A}_{TT},\ \mathcal{A}_{HH}^c, \mathcal{A}_{HT}^c, \mathcal{A}_{TH}^c, \mathcal{A}_{TT}^c,\\ &\mathcal{A}_{HH} \cup \mathcal{A}_{TH}, \mathcal{A}_{HH} \cup \mathcal{A}_{TT}, \mathcal{A}_{HT} \cup \mathcal{A}_{TT}, \mathcal{A}_{HT} \cup \mathcal{A}_{TH}\}.\end{aligned}$$

Then since all the information will be known for the third coin toss, $\mathcal{F}_3 = \mathcal{F}$, which contains 2^8 elements. Suppose $\omega = \omega_1\omega_2\omega_3$ is the sequence of coin flips of an event in Ω, where ω_n is the outcome of the nth toss. Let

$$X_j = \begin{cases} 1, & \omega_j = H \\ -1, & \omega_j = T \end{cases}$$

Thus, $\mathbb{P}\{X_j = 1\} = \mathbb{P}\{X_j = -1\} = \frac{1}{2}$. Define the stochastic process M_n by

$$M_n := \sum_{j=1}^{n} X_j.$$

Since $\mathcal{F}_n$ contains information up to the sequence of n tosses, we see that M_n is adapted to $\{\mathcal{F}_n\}_{n=1}^{3}$.

The following two famous lemmas are given in order to give a taste for measure theoretic proofs, and also to refer back to for later use.

Lemma 1.11. *(Chebyshev's Inequality) If* $1 \leq p < \infty$, *then for any* $\lambda > 0$, *we have that*

$$\mathbb{P}\{|X| \geq \lambda\} \leq \frac{\mathbb{E}[|X|^p]}{\lambda^p}.$$

Proof. Define $A := \{\omega \in \Omega; |X(\omega)| \geq \lambda\}$, then

$$\mathbb{E}[|X|^p] = \int_\Omega |X|^p d\mathbb{P} \geq \int_A |X|^p d\mathbb{P} \geq \lambda^p \int_A d\mathbb{P} = \lambda^p \mathbb{P}\{A\}.$$

□

For the Borel-Cantelli lemma, we first need a definition.

Definition 1.12. Suppose we have $(\Omega, \mathcal{F}, \mathbb{P})$ as our probability space, $\{A_n\}$ a sequence of events where $A_n \in \mathcal{F}$ for all n. We define (A_n i.o), or A_n *infinitely often*, by

$$(A_n \text{ i.o.}) = \bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i = \{\omega \in \Omega | \omega \text{ belongs to infinitely many } A_n\}.$$

Lemma 1.13. *(Borel-Cantelli) Let* $\{A_n\}$ *be a sequence of events. If* $\sum_{n=1}^{\infty} \mathbb{P}\{A_n\} < \infty$, *then* $\mathbb{P}\{(A_n \text{ i.o.})\} = 0$.

Proof. Since $(A_n \text{ i.o.}) \subseteq \bigcup_{i=n}^{\infty} A_i$, we have that

$$\mathbb{P}\{(A_n \text{ i.o.})\} \leq \lim_{n\to\infty} \mathbb{P}\{\bigcup_{i=n}^{n} A_i\} \leq \lim_{n\to\infty} \sum_{i=n}^{n} \mathbb{P}\{A_i\},$$

where, as $n \to \infty$, we have that $\lim_{n\to\infty} \sum_{i=n}^{n} \mathbb{P}\{A_i\} \to 0$ (a.s.). □

1.2. **The Radon-Nikodym Theorem.** We state and prove the Radon-Nikodym Theorem (Theorem 1.17) in order to be able to define conditional expectation later on, and thus be able to introduce martingales and Brownian motion. First, we must introduce a few definitions.

Definition 1.14. A measure ν is said to be *absolutely continuous* with respect to a measure μ if whenever $\mu(E) = 0$, then $\nu(E) = 0$. We write this as $\nu \ll \mu$.

Definition 1.15. Let $(\Omega, \mathcal{F}, \mu)$, $(\Omega, \mathcal{F}, \nu)$ be two measure spaces. We say that μ and ν are *mutually singular*, denoted by $\mu \perp \nu$, if there exists some $A \in \Omega$ such that $\mu(A) = 0$ and $\nu(A^C) = 0$.

We state a lemma which will allows us to prove the Radon-Nikodym Theorem, the proof of which is simple using the Hahn decomposition theorem.

Lemma 1.16. *Let μ and ν be finite positive measures on a measurable space $(X, \mathcal{F})$. Either $\mu \perp \nu$ or else there exists $\epsilon > 0$ and $G \in \mathcal{F}$ such that $\mu(G) > 0$ and G is a positive set for $\nu - \epsilon\mu$.*

Theorem 1.17. *(Radon-Nikodym) Suppose μ and ν are measures on $(\Omega, \mathcal{F})$ with $\nu \ll \mu$ such that*

$$\Omega = \bigcup_{n=1}^{\infty} A_n, \qquad \textit{with} \qquad \mu(A_n), \nu(A_n) < \infty \quad \forall n.$$

Then there exists a $\mu-$integrable non-negative function f which is measurable with respect to $\mathcal{F}$ such that for every $A \in \mathcal{F}$,

$$\nu(A) = \int_A f \mathrm{d}u.$$

Moreover, f is unique almost everywhere with respect to μ.

The function f is called the *Radon-Nikodym derivative* of ν with respect to μ and is written $\mathrm{d}\nu = f\mathrm{d}\mu$.

Proof. We give the proof in [4] and begin by defining f. Let

$$\mathcal{A} = \{g \text{ measurable} | g \geq 0, \int_A g d\mu \leq \nu(A), A \in \mathcal{F}\}.$$

Since $\mathcal{A}$ is bounded above by some $\nu(A)$ and $0 \in \mathcal{A}$, we let $L := \sup\{\int g d\mu \mid g \in \mathcal{A}.\}$ Let $g_n \in \mathcal{A}$ for all n with $\int g_n d\mu \to L$, then define $h_n = \max\{g_1, \ldots, g_n\}$. We claim that $h_n \in \mathcal{A}$ by induction and prove the $n = 2$ case. Let $B := \{x \mid g_1(x) \geq g_2(x)\}$, so then

$$\begin{aligned} \int_A h_2 d\mu &= \int_{A \cap B} h_2 d\mu + \int_{A \cap B^c} h_2 d\mu \\ &= \int_{A \cap B} g_1 d\mu + \int_{A \cap B^c} g_2 d\mu \\ &\leq \nu(A \cap B) + \nu(A \cap B^c) \\ &= \nu(A). \end{aligned}$$

And thus $h_2 \in \mathcal{A}$. Because $g \geq 0$, then h_n is increasing up to some f, where by the monotone convergence theorem,

$$\int_A g_n d\mu \leq \int_A h_n d\mu \leq \int_A f d\mu \leq \nu(A).$$

Because this holds for all n, then $\int f d\mu = L$. Define a positive measure by

$$\lambda(A) = \nu(A) - \int_A f d\mu.$$

Assume, for the sake of contradiction, that λ is not absolutely singular to μ. By Lemma 1.15, there exists some $\epsilon > 0$ and $G \in \mathcal{F}$ such that $\mu(G) > 0$ and G is a positive set for $\lambda - \epsilon\mu$. Thus,

$$\nu(A) - \int_A f d\mu = \lambda(A) \geq \lambda(A \cap G) \geq \epsilon\mu(A \cap G) = \int_A \epsilon \mathbb{1}_G d\mu,$$

where $\mathbb{1}_G$ is the indicator function for G. Thus, since $\nu(A) \geq \int_A (f + \epsilon \mathbb{1}_G) d\mu$, we have that $(f + \epsilon \mathbb{1}_G) \in \mathcal{A}$. However, since

$$\int_A (f + \epsilon \mathbb{1}_G) d\mu = L + \int_A \epsilon \mathbb{1}_G d\mu \geq L,$$

then $L \neq \sup\{\int g d\mu \mid g \in \mathcal{A}\}$, which is a contradiction.
Thus, there exists some $E \in \mathcal{F}$ such that $\mu(E) = 0$ and $\lambda(E^c) = 0$, where, since $\nu \ll \mu$, $\nu(E) = 0$. Because λ is a positive measure, we must write

$$\lambda(E) = \nu(E) - \int_E f d\mu = 0.$$

Thus, $\nu(A) = \int_A f d\mu$ for all $A \in \mathcal{F}$. The rest of the proof can be found in [4]. □

Example 1.18. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and suppose Q is a probability measure with $Q \ll \mathbb{P}$, then the Radon-Nikodym derivative

$$X = \frac{dQ}{d\mathbb{P}}$$

is a nonnegative random variable with $\mathbb{E}[X] = 1$ satisfying

$$Q(E) = \mathbb{E}_{\mathbb{P}}[\mathbf{1}_E X] \qquad \Leftrightarrow \qquad Q(E) = \int_E X d\mathbb{P}.$$

Definition 1.19. Suppose $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $\mathcal{G} \subset \mathcal{F}$. We define the *conditional expectation* of X given $\mathcal{G}$ as

$$\mathbb{E}[X|\mathcal{G}] = \frac{dQ}{d\mathbb{P}|_{\mathcal{G}}}.$$

Remark 1.20. We provide the Radon-Nikodym derivative definition of conditional expectation in order to be able to show its existence, which is not immediate by its usual definition. It takes little work to show that this definition satisfies the usual definition of conditional expectation: $\mathbb{E}[X|\mathcal{G}]$ is $\mathcal{G}-$measurable random variable; and for all $G \in \mathcal{G}$, $\int_G \mathbb{E}[X|\mathcal{G}] d\mathbb{P} = \int_G X d\mathbb{P}$. Another useful aspect of the Radon-Nikodym conditional expectation is that it is unique.

Proposition 1.21. *Let X, Y be random variables and $\mathcal{G}$ be a $\sigma-$algebra.*

- *If a, b are constants, then $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$.*
- *If Y is $\mathcal{G}$ measurable, then $\mathbb{E}[Y|\mathcal{G}] = Y$.*
- *If Y is independent of $\mathcal{G}$, then $\mathbb{E}[Y|\mathcal{G}] = \mathbb{E}[Y]$.*
- *(Tower Property) If $\mathcal{G} \subset \mathcal{F}$, then $\mathbb{E}[\mathbb{E}[Y|\mathcal{F}]|\mathcal{G}] = \mathbb{E}[Y|\mathcal{G}]$.*

1.3. **Martingales.** Consider a fair gambling game. The expected winnings in future games is \$0, regardless of the games already played. Mathematically, if our winnings are denoted by M_t, then if $m < n$, we can express this as

$$\mathbb{E}[M_n - M_m|\mathcal{F}_m] = 0 \quad \text{or} \quad \mathbb{E}[M_n|\mathcal{F}_m] = M_m.$$

We call fair games like this martingales.

Definition 1.22. Let $\mathbb{T} = \mathbb{N}$. We say that a real-valued stochastic process $\{M_n\}_{n\in\mathbb{T}}$ is a (discrete) *martingale* with respect the filtration $\{\mathcal{F}_n\}_{n\in\mathbb{T}}$ if

1. $\mathbb{E}[|M_n|] < \infty$ for each n;
2. $\{M_n\}_{n\in\mathbb{T}}$ is adapted to $\{\mathcal{F}_n\}_{n\in\mathbb{T}}$;
3. $M_m = \mathbb{E}[M_n|\mathcal{F}_m]$ for all $m \leq n$.

We shall later see that Brownian motion is a martingale, and thus provide these next two theorems to outline some important properties of martingales.

Theorem 1.23. *(Doob's Optional Stopping Theorem) Let $\{\mathcal{F}_n\}_{n\in\mathbb{T}}$ be a filtration on $\mathcal{F}$ and M_n be a martingale with respect to $\{\mathcal{F}_n\}$. If T is a stopping time bounded above by some integer K, then $\mathbb{E}[M_n] = \mathbb{E}[M_k]$.*

The stopping theorem tells us that there is no beating a fair game. Intuitively, it says that on average, a gambler leaves a fair game with the same amount of money as when he started.

Definition 1.24. Suppose $\Delta_1, \Delta_2, \dots$ is an adaptable sequence with respect to $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$, and $\mathbb{E}[\Delta_n^2] < \infty$ for all n. We define the *discrete stochastic integral* as

$$Z_n = \sum_{j=1}^{n} \Delta_j (M_j - M_{j-1}). \tag{1.25}$$

We can think of Δ_j as a betting strategy on the Drunkard's walk, in which we have some bet predicting whether the interval $M_j - M_{j-1}$ will go up or down. Thus, the discrete stochastic integral describes the winnings in our game. We give without proof the following properties.

Proposition 1.26. *The discrete stochastic integral described in (1.25) satisfies the following:*

1. *(Martingale Property) the integral Z_n is a martingale with respect to $\{\mathcal{F}_n\}$;*
2. *(Linearity) suppose Δ_n and Θ_n are adaptable sequences with a, b real numbers, then*

$$\sum_{j=1}^{n}(a\Delta_j + b\Theta_j)(M_j - M_{j-1}) = a\sum_{j=1}^{n}\Delta_j(M_j - M_{j-1}) + b\sum_{j=1}^{n}\Theta_j(M_j - M_{j-1});$$

3. *(Isometry)*

$$Var\left[\sum_{j=1}^{n}\Delta_j(M_j - M_{j-1})\right] = \sum_{j=1}^{n}\mathbb{E}[\Delta_j]^2.$$

This next theorem provides insight into the behaviors of martingales motion in the long term.

Theorem 1.27. *(Martingale Convergence Theorem) Suppose M_n is a martingale with respect to $\{\mathcal{F}_n\}_{n\in\mathbb{T}}$ and there exists a constant $C < \infty$ such that $\mathbb{E}[|M_n|] \leq C$ for all n. Then there exists a random variable M_∞ such that with probability one,*

$$\lim_{n\to\infty} M_n = M_\infty.$$

We give the beautiful proof for discrete martingales by Greg Lawler in [7].

Proof. Let $M_0, M_1, \dots$ be a martingale with respect to its natural filtration $\{\mathcal{F}_n\}_{n\in\mathbb{N}_0}$. We claim that if $a < b$ are real, then M_n cannot infinitely fluctuate above b and below a. Define

$$S_1 = \min\{n \mid M_n \leq a\}, \qquad T_1 = \min\{n > S_1 \mid M_n \geq b\}$$

as stopping times denoting S_1 as the first time the martingale drops below a and T_1 as the first time after S_1 the martingale rises above b. For $j > 1$, define

$$S_j = \min\{n > T_{j-1} \;\; M_n \leq a\}, \qquad T_j = \min\{n > S_j \;\; M_n \geq b\}.$$

Define the discrete stochastic integral as

$$W_n = \sum_{k=0}^{n} \Delta_k [M_k - M_{k-1}],$$

where Δ_n is a "bet" we make such that $\Delta_n = 0$ if $n - 1 < S_1$ (we don't buy if the "price" of the martingale hasn't dropped below a), $\Delta_n = 1$ if $S_j \leq n - 1 < T_j$ (we "buy" if the price just dropped below a) and $\Delta_n = 0$ if $T_j \leq n - 1 < S_{j+1}$ (we "sell" once the price rises above b). Define the number of fluctuations between a and b as

$$U_n = j, \qquad T_j < n \leq T_{j+1}.$$

Thus, since each U_j results in a profit of at least $b - a$, we have that

$$W_n \geq U_n(b-a) + (M_n - a), \tag{1.28}$$

where the last term represents the loss of holding the asset at the present. By Proposition 1.26, W_n is a martingale, and thus by Theorem 1.23, $\mathbb{E}[W_n] = \mathbb{E}[W_0] = 0$. Therefore, taking expectations of (1.28), we have that, for every n,

$$\mathbb{E}[U_n] \leq \frac{\mathbb{E}[a - M_n]}{b-a} \leq \frac{|a| + \mathbb{E}[|M_n|]}{b-a} \leq \frac{|a| + C}{b-a} < \infty.$$

Thus, since $\lim_{n\to\infty} \mathbb{E}[U_n] < \infty$, a result from measure theory[2] shows that $\lim_{n\to\infty} U_n < \infty$, implying a finite number of fluctuations. □

Remark 1.29. The definition of a *continuous-time* martingale is analogous to the discrete-time martingale, with the distinction that $\mathbb{T} = [0, \infty)$. It is not hard to show that the two above theorems hold for continuous-time martingales.

Definition 1.30. Let $\{X_t\}_{t\in\mathbb{T}}$ be a process and suppose $\Pi_N = \{0 = t_0^N < t_1^N < \cdots < t_N^N = t\}$ is a partition of $\mathbb{T}$ by stopping times $\{t_i^N\}_{i,N\in\mathbb{N}}$. Suppose further that $\|\Pi\| \to 0$ as $N \to \infty$. We define the *quadratic variation* of M_t to be

$$\sum_{i=1}^{N} (M_{t_{i+1}^N} - M_{t_i^N})^2 \to \langle M \rangle_t$$

as $N \to \infty$.

The rest of the paper focuses on a specific example of a continuous-time martingale: Brownian motion.

2. Brownian Motion

We first define Brownian motion and then provide intuition for it.

Definition 2.1. A continuous adapted process $\{B_t\}_{t\in[0,\infty)}$ taking values in $\mathbb{R}^d$ is called a *(d-dimensional) Brownian motion* with *drift* m and *variance* σ^2 if, for all $0 = t_0 < t_1 < \cdots < t_n$, we have that

(1) $B_0 = 0$;
(2) the distribution of $\{B_{t_{i+1}} - B_{t_i}\}_{i=0}^{n-1}$ is normal with
$$B_{t_{i+1}} - B_{t_i} \sim N(m(t_{i+1} - t_i), \sigma^2(t_{i+1} - t_i));$$
(3) $B_{t_{i+1}} - B_{t_i}$ is independent of $\mathcal{F}_{t_i}$.

[2] If f is measurable and $\int f d\mu < \infty$, then $f < \infty$ a.s.

We often make use of the *standard Brownian motion*, which is a Brownian motion with drift 0 and variance 1.

2.1. **Drunkard's Walk.** To construct a continuous random motion, it is helpful to first construct the discrete case. A *drunkard's walk*, or *symmetric random walk*, is the stochastic process $\{M_n\}_{n\in\mathbb{T}}$ defined in Example 1.9, with the change that we let $\omega = \omega_1\omega_2\ldots$ be an infinite sequence of coin tosses instead of a finite one. We allow the tosses happen every Δt time increment.

Proposition 2.2. *The drunkard's walk satisfies the following properties:*

1. *if* $0 = k_0 < k_1 < \cdots < k_n$ *are integers, then the random variables* $\{B_{t_{i+1}} - B_{t_i}\}_{i=1}^{n-1}$ *are independent with mean* 0 *and variance* $k_{i+1} - k_i$;
2. *the drunkard's walk is a martingale;*
3. *the quadratic variation of the drunkard's walk is*

$$\langle M\rangle_k = \sum_{j=1}^{k}(M_j - M_{j-1})^2 = k.$$

Proof. We use Propostion 1.21.

1. The first proposition is immediate by construction.
2. Let $m < n$ be natural numbers, then

$$\begin{aligned}\mathbb{E}[M_n|\mathcal{F}_m] &= \mathbb{E}[(M_n - M_m) + M_m|\mathcal{F}_m]\\ &= \mathbb{E}[(M_n - M_m)|\mathcal{F}_m] + \mathbb{E}[M_m|\mathcal{F}_m]\\ &= \mathbb{E}[(M_n - M_m)] + M_m = M_m. \qquad (2.3)\end{aligned}$$

   Where the equalities in (2.3) hold due to independence, M_m being $\mathcal{F}_m$ measurable, and the first property.
3. For any j, $M_j - M_{j-1} = \pm 1$, and thus $\sum_{j=1}^{k}(\pm 1)^2 = k$.

□

Remark 2.4. We can approximate Brownian motion as a limit of the Drunkard's walk where the speed of coin tosses is increased and the step size is decreased, we define a *scaled random walk* by

$$W_{Nt}^{(N)} = \frac{1}{\sqrt{N}}M_{Nt},$$

where N is a fixed integer, Nt is an integer, and $\frac{1}{\sqrt{N}}$ is the step size of the walk.

We prove a few properties of Proposition 2.2, now applied to the scaled random walk, and leave the rest for the reader to check.

1. While the expectation is still obviously 0, we need to see if the variance remains the same. Let $s < t$ such that Ns, Nt are integers, then

$$\begin{aligned}\mathrm{Var}[W_t^{(N)} - W_s^{(N)}] &= \mathrm{Var}[\frac{1}{\sqrt{N}}(M_{Nt} - M_{Ns})]\\ &= \frac{1}{N}(\mathrm{Var}[X_1] + \mathrm{Var}[X_2] + \cdots + \mathrm{Var}[X_{N(t-s)}]) = (t - s).\end{aligned}$$

(2) For $t \geq 0$ such that Nt is an integer,

$$\langle W_t^{(N)} \rangle_t = \sum_{j=1}^{Nt} \left[W^{(N)}\left(\frac{j}{N}\right) - W^{(N)}\left(\frac{j-1}{N}\right) \right]^2 = \sum_{j=1}^{Nt} \left[\frac{1}{\sqrt{N}} X_j \right]^2 = t.$$

Moreover, we will state, but not prove, one final theorem from [5] for the scaled random walk.

Theorem 2.5. *(Central limit) Let $t \geq 0$. As $N \to \infty$, the distribution of the scaled random walk $W^{(n)}(t)$ evaluated at time t converges to the normal distribution with mean zero and variance t.*

Thus, as Figure 1 below shows, we can think of Brownian motion as the limit of scaled random walks.

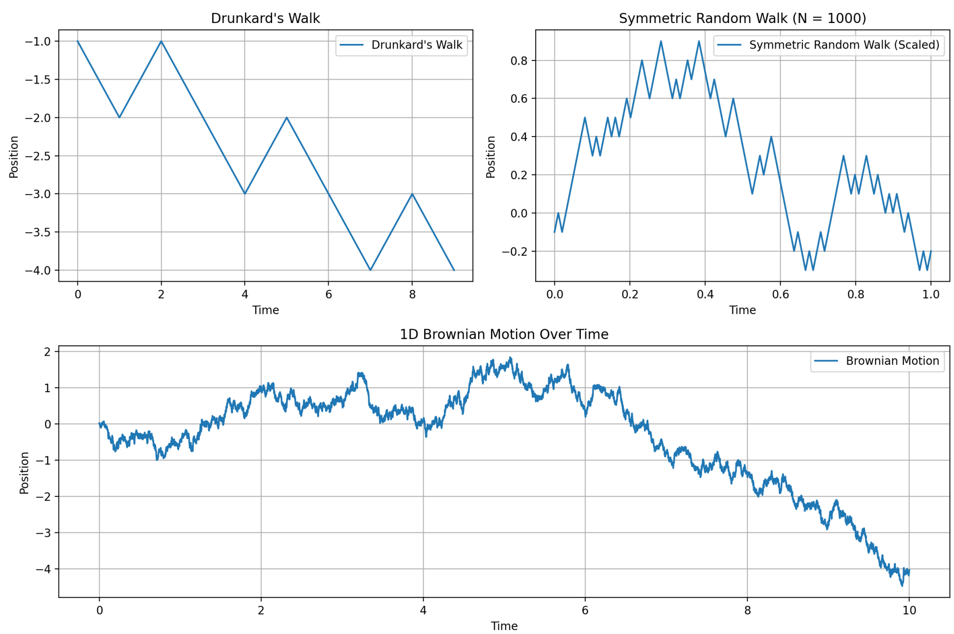

Figure 1. Drunkard's Walk, Symmetric Random Walk, and Brownian Motion

2.2. **Lévy's Construction of Brownian Motion.** To show that there exists a Brownian motion, we will give the Lévy construction from [6]. We first need a preliminary lemma.

Lemma 2.6. *Let $\{X^n\}_{n\in\mathbb{N}}$ be a sequence of a.s. continuous functions which converge uniformly in probability to a process X. That is, for any $\epsilon > 0$,*

$$\lim_{n\to\infty} \mathbb{P}\{ \sup_{s\in[0,t]} ||X_s^n - X_s|| < \epsilon\} = 1$$

for all t. Then X is also continuous.

We leave this lemma without proof, but note that for fixed ω, this is just a classical uniform convergence statement. For fixed t, there exists a sub-sequence in n such that convergence is almost sure. We can now build a Brownian motion.

For $n \in \mathbb{N}_0$, define

$$D_n := \{\frac{k}{2^n} | k \in \mathbb{N}\},$$

$D_0 = \mathbb{N}$. Then we have that $\mathcal{D} = \bigcup_n D_n$ is the set of Dyadic rationals. Let $\{Z_m\}_{m \in \mathcal{D}}$ be a collection of random variables such that $Z_m \sim N(0,1)$, and $Z_0 = 0$. Determine the value of the nth approximation X_t^n on $t \in D_n$ by defining

$$X_t^0 = \sum_{k \in D_0 : k < t} Z_k.$$

For $n > 0$, define $X_t^n = X_t^{n-1}$ for all $t \in D_{n-1}$. For $t \in D_n \setminus D_{n-1}$, let

$$X_t^n = X_t^{n-1} + \frac{Z_t}{2^{n+1}} \tag{2.7}$$

and use linear interpolation to define X_t^n for all value of t. We can now formally interpolate between $\{X_t^n\}_{t \in D_n}$:

$$X_t^n = X_{\lfloor t \rfloor_n}^n + \frac{t - \lfloor t \rfloor_n}{\lceil t \rceil_n - \lfloor t \rfloor_n}(X_{\lceil t \rceil_n}^n - X_{\lfloor t \rfloor_n}^n)$$

where $\lfloor t \rfloor_n$ is the maximum $s \in D_n$ less than t, and $\lceil t \rceil_n$ is defined similarly.

Theorem 2.8. *The processes X^n defined in (2.7) converge, in its natural filtration, a.s. uniformly to a process Brownian motion starting at zero.*

Proof. First we will show convergence. By construction, we have that

$$\sup_{s \in [0,t]} ||X_s^n - X_s^{n+1}|| = \max_{s \in \{D_{n+1} \setminus D_n | s < t\}} ||\frac{Z_s}{2^{n+1}}||.$$

Note that $\{D_{n+1} \setminus D_n | s < t\}$ has $t2^n$ elements. Let $F(x) := \mathbb{P}\{||Z_s||^2 \leq x\}$ be the distribution function of $||Z_s||^2$.[3] Then

$$\begin{aligned}
\mathbb{P}\{\sup_{s \in [0,t]} ||X_s^n - X_s^{n+1}|| > \epsilon\} &= \mathbb{P}\{\max_{\{D_{n+1} \setminus D_n | s < t\}} ||Z_s|| > 2^{n+1}\epsilon\} \\
&\leq \sum_{s \in \{D_{n+1} \setminus D_n | s < t\}} \mathbb{P}\{||Z_s|| > 2^{n+1}\epsilon\} \\
&= t2^n(1 - F(2^{2n+2}\epsilon^2)) \\
&= t2^n \exp\{-2^{2n+1}\epsilon^2\} \\
&\leq te^{-n} < \infty.
\end{aligned}$$

Where the last inequality stands since as $n \to \infty$, we can choose N large enough such that $N(\ln(2) + 1) < 2^{2N+1}\epsilon^2$, and thus the inequality hold for $n > N$. Thus, by Borel-Cantelli (Lemma 1.13), we have that

$$\mathbb{P}\{\sup_{s \in [0,t]} ||X_s^n - X_s^{n+1}|| \geq \epsilon \text{ for infinitely many } n\} = 0.$$

By Lemma 2.9, X^n converges uniformly on $[0,t]$ and so thus X_t is a continuous process.

[3] It is known that $||Z_s||^2$ has a χ^2−distribution with $d = 2$, and thus $F(x) = 1 - e^{\frac{-x}{2}}$.

Now we prove that X_t is a Brownian motion. For s, t such that $\lceil s \rceil_n < t$, $t \in D_n \setminus D_{n+1}$, we know that Z_t is independent of $\mathcal{F}_s = \sigma(X_u | u \leq s)$ since Z_t is not involved in the construction of X_s. Moreover,

$$X_t - X_s = X_t^0 - X_s^0 = \sum_{k \in D_0 : s < k < t} Z_k \sim N(0, (t-s))$$

and $X_t - X_s$ is independent of $\mathcal{F}_s$ by the above logic. If the result holds for any $s, t \in D_n$, then if $u \in D_{n+1} \setminus D_n$,

$$X_u - X_{\lfloor u \rfloor_n} = \frac{X_{\lceil u \rceil_n} + X_{\lfloor u \rfloor_n}}{2} + \frac{Z_u}{2^{n+2}} = \frac{2^{-(n+1)} Z_{\lceil u \rceil_n}}{2} + \frac{Z_u}{2^{n+2}} \sim N(0, \frac{1}{2^{n+1}}).$$

Similarly, we have that $X_{\lceil u \rceil_n} - X_u \sim N(0, \frac{1}{2^{n+1}})$. Both intervals are independent of $\mathcal{F}_{\lfloor u \rfloor_n}$. Thus, for all $s, t \in D_{n+1}$,

$$X_t - X_s = (X_t - X_{\lfloor t \rfloor_n}) + (X_{\lfloor t \rfloor_n} - X_{\lceil s \rceil_n}) + (X_{\lceil s \rceil_n} - X_s) \sim N(0, (t-s)).$$

The first two terms of the sum are independent of $\mathcal{F}_{\lceil s \rceil_n}$, and thus independent of $\mathcal{F}_s$. The last is independent of both $\mathcal{F}_{\lfloor s \rfloor_n}$ and $X_s - X_{\lfloor s \rfloor_n}$. To prove that $X_{\lceil s \rceil_n}$ is independent of $\mathcal{F}_s$, simply write

$$\mathcal{F}_s = \mathcal{F}_{\lfloor s \rfloor_n} \vee \sigma(X_s - X_{\lfloor s \rfloor_n}) \vee \sigma(Z_u | u \in (\lfloor s \rfloor_n), s).$$

Thus, by inducting, we see that for any $D_n \in \mathcal{D}$, if $s, t \in D_{n+1}$, we have that $X_t - X_s$ is normally distributed and independent of $\mathcal{F}_s$.

Finally, if $s < t$, we can find sequences $s_n \downarrow s$ and $t_n \uparrow t$ with $s_n, t_n \in D_n$ and $s_k \leq t_k$ for some $k \geq 0$. Then $X_{t_n} - X_{s_n} \sim N(0, (t_n - s_n))$ and by continuity of X, we have that

$$X_t - X_s = X_{t_k} - X_{s_k} + \sum_{n=k+1}^{\infty} (X_{t_n} - X_{t_{n-1}} - X_{s_n} + X_{s_{n-1}}) \sim N(0, (t-s)).$$

Thus, X is a Brownian motion starting at zero (by construction) in its natural filtration. □

2.3. **Brownian Motion.** Having now showed the existence of Brownian motion, we can talk about some of its properties.

Theorem 2.9. *Brownian motion is a martingale.*

The proof of the theorem is identical to its discrete version in Proposition 2.2. This theorem now allows us to apply Theorems 1.23 and 1.27 to Brownian motion.

From now on, it is implied that the index time of the processes $\mathbb{T}$ is the continuous set $[0, \infty)$.

Theorem 2.10. *(Markov Property) Let B_t be a Brownian motion and T a stopping time with $\mathbb{P}\{T < \infty\} = 1$. Then the process defined by*

$$Y_t = B_{T+t} - B_T$$

is a standard Brownian motion with respect to the filtration $\{\mathcal{F}_T\}$.

Intuitively, the Markov Property states that stopping the Brownian motion at any time T and starting it up again creates a Brownian motion independent of its past. Thus, Brownian motion is "memory-less." We provide a result for Brownian motion which is often used to compute specific probabilities associated with the randomness.

Corollary 2.11. *(Reflection Principle) If B_t is a standard Brownian motion with $B_0 = 0$, then for every $a > 0$,*

$$\mathbb{P}\{\max_{0\leq s\leq t} B_s \geq a\} = 2\mathbb{P}\{B_t > a\}.$$

Proof. Let τ_a be the first time B_t hits a, that is, $B_{\tau_a} = a$. Then we know that

$$\mathbb{P}\{\max_{0\leq s\leq t} B_s \geq a\} = \mathbb{P}\{\tau_a < t\}.$$

Moreover, we have that

$$\begin{aligned}\mathbb{P}\{B_t > a\} &= \mathbb{P}\{\tau_a < t \quad \text{and} \quad B_t > a\}\\ &= \mathbb{P}\{\tau_a < t\}\mathbb{P}\{B_t - B_{\tau_a} > 0 \mid \tau_a < t.\}\end{aligned}$$

By Theorem 2.10, we know that the probability of $Y_t = B_t - B_{\tau_a}$ being positive is equal to it being negative, since it is a Brownian motion. Thus,

$$\mathbb{P}\{B_t - B_{\tau_a} > a \mid \tau_a < t\} = \frac{1}{2}.$$

□

It is easy to show that if f is C^1, that is, f has a continuous derivative, then its quadratic variation is zero. Later on, we will see that quadratic variation is the source for the volatility term in the Black-Scholes-Merton PDE. The next theorem provides us with a useful characterization of Brownian motion that provides much meaning to the Itô-Doeblin formulas.

Theorem 2.12. *Let B be a Brownian motion. Then $\langle B\rangle_t = t$ for all $t \geq 0$ a.s.*

Proof. Suppose T is a stopping time and $0 \leq t \leq T$. Let $\Pi = \{t_0, t_1, \dots, t_n\}$ be a partition of $[0, T]$. It suffices to show that the *sampled quadratic variation,*

$$Q_\Pi = \sum_{j=0}^{n-1} \left(B_{t_{j+1}} - B_{t_j}\right)^2$$

converges to T as $\|\Pi\| \to 0$. To do this, we must conclude that $\mathbb{E}[Q_\Pi] \to T$ and $\mathrm{Var}[Q_\Pi] \to 0$. Note that because the intervals are independent normal variables with mean zero, we have that the variance of the intervals is

$$\mathbb{E}\left[\left(B_{t_{j+1}} - B_{t_j}\right)^2\right] = t_{j+1} - t_j.$$

To compute $\mathbb{E}[Q_\Pi]$, we use the linearity of expectation and bring it inside the telescoping sum, giving $\mathbb{E}[Q_\Pi] = T$.

To see that the variance of Q_Π converges to 0, note that[4]

$$\begin{aligned}\mathrm{Var}\left[\left(B_{t_{j+1}} - B_{t_j}\right)^2\right] &= \mathbb{E}\left[\left[\left(B_{t_{j+1}} - B_{t_j}\right)^2 - (t_{j+1} - t_j)\right]^2\right]\\ &= \mathbb{E}\left[\left(B_{t_{j+1}} - B_{t_j}\right)^4\right]\\ &\quad - 2(t_{j+1} - t_j)\mathbb{E}\left[\left(B_{t_{j+1}} - B_{t_j}\right)^2\right] + (t_{j+1} - t_j)^2\\ &= 3(t_{j+1} - t_j)^2 - 2(t_{j+1} - t_j)^2 + (t_{j+1} - t_j)^2 \qquad (2.13)\\ &= 2(t_{j+1} - t_j)^2.\end{aligned}$$

[4]We use a well known fact, known as *normal kurtosis*, to derive the first term in (2.19). Read more about it in Exercise 3.3 of [5]

Again, using the linearity of variance, we obtain

$$\mathrm{Var}[Q_\Pi] = \sum_{j=0}^{n-1} 2(t_{j+1-t_j})^2 \leq 2\|\Pi\| \sum_{j=0}^{n-1}(t_{j+1} - t_j) = 2\|\Pi\|T.$$

Thus, as $\|\Pi\| \to 0$, we conclude that $\mathrm{Var}[Q_\Pi] \to 0$. □

We conclude the section with a dilemma. A known result in stochastic analysis is a theorem stating that Brownian motion is nowhere differentiable. The proof, which can be found on pages 48-55 of [7], deals with the fact that Brownian motion is $\alpha-$Hölder continuous for $\alpha < \frac{1}{2}$. Intuitively, if one could determine the derivative of a Brownian motion B at time t by looking at B_s, $0 \leq s \leq t$, then the derivative would provide information on $B_{t+\Delta t} - B_t$, contradicting Definition 2.1's independence statement. Thus, how could Doeblin and Itô talk about dX_t in (1.2)? The answer lies in the foundations of stochastic calculus: the stochastic integral.

3. Stochastic Calculus

Now that we have introduced Brownian motion, we can look at some of the tools which will be used in order to utilize Brownian motion in finance in section 5.

3.1. **The Itô-Integral.**

We shall finally make sense of (1.1), and with it, (1.2). To do this, it will be helpful to first recall a construction of the classic Riemann integral.

Definition 3.1. Suppose $f : [a, b] \to \mathbb{R}$ is a continuous function, and Π is a partition of $[a, b]$ with $a = t_0 < t_1 < \cdots < t_n = b$. If f is approximated by a step function, $f_n(x) = f(s_j)$, where $t_{j-1} < t \leq t_j$ and $s_j \in [t_{j-1}, t_j]$, then we define the *Riemann Integral* by

$$R_t = \int_a^b f(x)dx = \lim_{n\to\infty} \int_a^b f_n(x)dx = \lim_{n\to\infty} \sum_{j=1}^{n} f(s_j)(t_j - t_{j-1}).$$

We can now compare this definition with the definition of the stochastic integral, often referred to as an Itô integral.

Definition 3.2. Let B_t be a Brownian motion and suppose $\{X_j\}_{j\in[n]}$ is a stochastic process adapted to a filtration $\{\mathcal{F}_{t_j}\}$, with $0 = t_0 < t_1 < \cdots < t_n = t < \infty$, and $\mathbb{E}[X_j^2] < \infty$. We say a Δ_t is a *simple process* if $\Delta_t = X_j$, where $t_j \leq t < t_{j+1}$. We define the *Itô integral* by

(3.3)

$$I_t = \int_0^t \Delta_s dB_s = \sum_{i=0}^{j-1} \Delta_i[B_{t_{i+1}} - B_{t_i}] + \Delta_j[B_t - B_{t_j}] = \sum_{i}^{j} \Delta_{t_i}[B_{t_{i+1}\wedge t} - B_{t_i\wedge t}].$$

Intuition for the stochastic integral can be gained by regarding B_t as a Brownian motion modeling the price per share of some asset at time t, and thinking of $t_0, t_1, \dots, t_n$ as trading dates for the asset. It is natural to assume a position taken in the asset at each trading date, and thus Δ_t is simple and adapted to the information known at time t, that is, $\mathcal{F}_t$. Thus, $I(t) = I_t$ models the gains from trading at each time t.

Proposition 3.4. *The Itô integral defined in (3.3) satisfies the following properties:*

(1) I_t is a martingale;

(2) (Itô Isometry)

$$\mathbb{E}[I_t^2] = \mathbb{E}\left[\int_0^t \Delta_s^2 ds\right]; \tag{3.5}$$

(3) if a, c are constants and A_t, C_t are simple processes, then $aA_t + cC_t$ are simple processes and

$$\int_0^t (aA_s + cC_s)dB_s = a\int_0^t A_s dB_s + c\int_0^t C_s dB_s; \tag{3.6}$$

(4) (Quadratic Variation)

$$\langle I\rangle_t = \int_0^t \Delta_s^2 ds. \tag{3.7}$$

Proof. (1) Let $t_\ell < t_k$ such that $t_\ell, t_k \in \Pi$, where $s \in [t_\ell, t_{\ell+1})$, and $t \in [t_k, t_{k+1})$. Rewriting (3.3) gives

$$\begin{aligned} I_t = &\sum_{j=0}^{\ell-1} \Delta_{t_j}[B_{t_{j+1}} - B_{t_j}] \\ &+ \Delta_{t_\ell}[B_{t_{\ell+1}} - B_{t_\ell}] \\ &+ \sum_{j=\ell+1}^{k-1} \Delta_{t_j}[B_{t_{j+1}} - B_{t_j}] \\ &+ \Delta_{t_k}[B_{t_{k+1}} - B_{t_k}] \end{aligned} \tag{3.8}$$

We proceed by taking conditional expectations of each term in (3.8). Because $t_\ell < s$, then the terms in the first sum in (3.8) are $\mathcal{F}_s$ measurable. Thus, the first sum remains the same after taking the conditional expectation.

For the second sum in (3.8), since we know that B_t is a martingale by Theorem 2.9, we have that

$$\mathbb{E}[\Delta_{t_\ell}[B_{t_{\ell+1}} - B_{t_\ell}] \mid \mathcal{F}_s] = \Delta_{t_\ell}[\mathbb{E}[B_{t_{\ell+1}} \mid \mathcal{F}_s] - B_{t_\ell}] = \Delta_{t_\ell}[B_{t_s} - B_{t_\ell}].$$

Thus, adding the first two terms yields I_s. The third sum, whose time-terms are greater than s, yields by the tower property of expectation that

$$\begin{aligned} \mathbb{E}\left[\Delta_{t_j}[B_{t_{j+1}} - B_{t_j}] \mid \mathcal{F}_s\right] &= \mathbb{E}\left[\mathbb{E}\left[\Delta_{t_j}[B_{t_{j+1}} - B_{t_j}] \mid \mathcal{F}_{t_j}\right] \mid \mathcal{F}_s\right] \\ &= \mathbb{E}\left[\Delta_{t_j}[\mathbb{E}[B_{t_{j+1}} \mid \mathcal{F}_{t_j}] - B_{t_j}] \mid F_s\right] \\ &= \mathbb{E}\left[\Delta_{t_j}[B(t_j) - B(t_j) \mid F_s]\right] = 0. \end{aligned}$$

By the same logic, the fourth term is also 0. We conclude that $\mathbb{E}[I_t|\mathcal{F}_s] = I_s$.

(2) We must prove that the expectation of the cross terms in I_t^2 is zero. For $i < j$,

$$\begin{aligned} &\mathbb{E}\left[\Delta_{t_i}\Delta_{t_j}(B_{t_{i+1}\wedge t} - B_{t_i\wedge t})(B_{t_{j+1}\wedge t} - B_{t_j\wedge t})\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\Delta_{t_i}\Delta_{t_j}(B_{t_{i+1}\wedge t} - B_{t_i\wedge t})(B_{t_{j+1}\wedge t} - B_{t_j\wedge t})\right] \mid \mathcal{F}_{t_j}\right] \\ &= \mathbb{E}\left[\Delta_{t_i}\Delta_{t_j}(B_{t_{i+1}\wedge t} - B_{t_i\wedge t})\mathbb{E}\left[(B_{t_{j+1}\wedge t} - B_{t_j\wedge t}) \mid \mathcal{F}_{t_j}\right]\right] = 0, \end{aligned}$$

and for the remaining terms, we recognize that $\mathbb{E}\left[(B_{t_{i+1}\wedge t} - B_{t_i\wedge t})^2\right]$ is the variance of the Brownian increment $(B_{t_{i+1}\wedge t} - B_{t_i\wedge t})$, which is defined to be $(t_{i+1}\wedge t) - (t_i\wedge t)$.

Thus, since all the cross products are zero,

$$\mathbb{E}\left[(I_t)^2\right] = \mathbb{E}\left[\sum_i^n \Delta_{t_i}^2 [B_{t_{i+1}\Delta t} - B_{t_i \Delta t}]^2\right] = \mathbb{E}\left[\sum_i^n \Delta_{t_i}^2 [(t_{i+1} \wedge t) - (t_i \wedge t)]\right]$$
$$= \mathbb{E}\left[\int_0^t \Delta_s^2 ds\right].$$

(3) This property is immediate from the definition.
(4) Suppose $[t_j, t_{j+1}]$ is a sub-interval on which Δ_t is constant, and let $\Pi = \{s_0, s_1, \ldots, s_m\}$ be a partition of the subinterval. Then,

$$\sum_{i=0}^{m-1} \left[I_{s_{i+1}} - I_{s_i}\right]^2 = \Delta_{t_j}^2 \sum_{i=0}^{m-1} \left[B_{s_{i+1}} - B_{s_i}\right]^2 \xrightarrow[\|\Pi\| \to 0]{} \Delta_{t_j}^2 (t_{j+1} - t_j). \tag{3.9}$$

Since $\Delta(t)$ is constant on $t_j \leq t \leq t_{j+1}$, (3.9) is equal to $\int_{t_j}^{t_{j+1}} \Delta_s^2 ds$, and so we obtain (3.7) by adding up all the time intervals. □

We can now generalize (3.3) by approximating the simple processes, loosening our definition of the stochastic integral.

Definition 3.10. Suppose $\{\Delta_t\}$ is a continuous process adapted to the filtration $\{\mathcal{F}_t\}_{t\in[0,\infty)}$, and suppose further that $\Delta_t \in L^2(\Omega, \mathcal{F}, \mathbb{P})$.[5] Let Δ_t^n be a sequence of simple process such that

$$\lim_{n\to\infty} \mathbb{E}\left[\int_0^T |\Delta_t^n - \Delta_t|^2 dt\right] = 0.$$

We define the *Itô Integral* by

$$\int_0^t \Delta_s dB_s = \lim_{n\to\infty} \int_0^t \Delta_s^n dB_s. \tag{3.11}$$

The existence of such a simple process comes from the completeness of L^2, of which the reader can explore more in [6]. We leave as a fact that this definition satisfies Proposition 3.4.

Remark 3.12. We can know define the *stochastic differential equation*, or *SDE*, to be the differential equation

$$dI_t = \Delta_t dB_t$$

satisfying

$$I_t = I_0 + \int_0^t \Delta_s dB_s.$$

Thus, an SDE means nothing more than formal notation for the stochastic integral, which is well defined.

Definition 3.13. Let B_t be a Brownian motion. We say that $\{X_t\}_{t\in\mathbb{T}}$ is an *Itô process* if it satisfies

$$X_t = X_0 + \int_0^t m_s ds + \int_0^t \sigma_s dB_s,$$

[5] We refer to L^2 as the space of Borel-measurable functions such that $\|f\|_2 = \left(\int |f(x)|^2 d\mu\right)^{\frac{1}{2}} < \infty$. One can read more about L^p spaces in [4].

where m_t and σ_t are adapted processes in L^2. In other words, X_t satisfies

$$dX_t = m_t dt + \sigma_t dB_t. \tag{3.14}$$

Remark 3.15. Note that (3.14) is a general version of the geometric Brownian motion described by (1.3). One can see this by letting $m_t = mX_t$ and $\sigma_t = \sigma X_t$. Moreover, if X_t is a geometric brownian motion, then one can use Proposition 3.4 to see that $d\langle X\rangle_t = \sigma_t^2 X_t^2$.

Note that for the following theorem, a function f being C^n in a variable x describes that the nth derivative of f with respect to x exists and is continuous.

Theorem 3.16 is a less-generalized version of the Itô-Doeblin formula. This formula is known as the fundamental theorem of stochastic calculus since it can be easily used to find solutions to stochastic differential equations. Moreover, the Black-Scholes model can be easily derived in a few lines using the Itô-Doeblin formula[6].

Theorem 3.16. *Let X_t be a Brownian motion and φ be a real valued function which is C^1 in t and C^2 in x. Then*

$$\varphi(t, X_t) = \varphi(0, X_0) + \int_0^t \partial_t \varphi(s, X_s) ds + \int_0^t \partial_x \varphi(s, X_s) dX_s + \frac{1}{2} \int_0^t \partial_{xx} \varphi(s, X_s) ds. \tag{3.17}$$

We will prove this for the single variable case, where

$$\varphi(X_t) = \varphi(X_0) + \int_0^t \varphi'(X_s) dX_s + \frac{1}{2} \int_0^t \varphi''(X_s) ds.$$

Proof. Consider that by expanding φ into a second order Taylor series, we have that

$$\varphi(y) - \varphi(x) = \varphi'(x)(y - x) + \frac{1}{2}\varphi''(x)(y - x)^2 + o\left((y - x)^2\right).$$

Let $\Pi = \{t_0, t_1, \dots, t_n\}$ be a partition on $[0, t]$, and thus

$$\begin{aligned} \varphi(X_t) &= \varphi(X_0) + \sum_{i=0}^{\infty} (\varphi(X_{t_{i+1}}) - \varphi(X_{t_i})) \\ &= \varphi(X_0) + \sum_{i=0}^{\infty} \Big[\varphi'(X_{t_i})(X_{t_{i+1}} - X_{t_i}) \\ &\quad + \frac{1}{2}\varphi''(X_{t_i})(X_{t_{i+1}} - X_{t_i})^2 + o\left((X_{t_{i+1}} - X_{t_i})^2\right)\Big]. \end{aligned} \tag{3.18}$$

If $X_t^\Pi = X_t$, then for $t \in [t_i, t_{i+1}]$, $\varphi'(X_t^\Pi) \xrightarrow[\|\Pi\| \to 0]{} \varphi'(X_t)$ a.s. This holds for φ'' as well. Note that both $\varphi'(X_t^\Pi)$ and $\varphi''(X_t^\Pi)$ are adapted processes.

For the first of the three sums in (3.18)'s right hand side, since $\varphi'(X^\Pi)$ is a continuous adapted process, then as $\|\Pi\| \to 0$, we get by Definition 3.10 that

$$\lim_{n \to \infty} \sum_{i=0}^{n} [\varphi'(X_{t_i})(X_{t_{i+1}} - X_{t_i})] = \int_0^t \varphi'(X^\Pi) dX_s \xrightarrow{P} \int_0^t \varphi'(X_s) dX_s.$$

[6] While we opt for another derivation of the model using Girsanov's theorem since it provides more insight, the simpler derivation can be found in page 143 of [5]

For the second sum in (3.18), let $h(t) = \varphi''(B_t)$. Since f is continuous, then there exists a step function $h_\epsilon(t)$ such that $|h(t) - h_\epsilon(t)| < \epsilon$ for every t. By an analogous argument to the proof for (4) in Proposition 3.4, we can say that if $\pi = \{s_1, s_1, \dots, s_m\}$ is a partition on $[t_j, t_{j+1}]$ where h_ϵ is constant, then

$$\lim_{n\to\infty} \sum_{i=1}^{m-1} h_\epsilon(t)[B_{s_{i+1}} - B_{s_i}]^2 = \int_{t_j}^{t_{j+1}} h_\epsilon(s)ds \quad \text{a.s.}$$

Moreover, consider that, as $\epsilon \to 0$

$$\left| \sum_{j=1}^{n} (h(t) - h_\epsilon(t)) [B_{t_{j+1}} - B_{t_j}]^2 \right| \leq \epsilon \sum_{j=1}^{n} [B_{t_{i+1}} - B_{t_i}]^2 \to 0.$$

Therefore,

$$\lim_{\epsilon\to 0} \frac{1}{2} \int_0^t h_\epsilon(s)ds = \frac{1}{2} \int_0^t h(s)ds = \frac{1}{2} \int_0^t \varphi''(B_s)ds \qquad \text{a.s.}$$

For the last sum, we have that, for partitions fine enough,

$$o((X_{t_{i+1}} - X_{t_i})^2)) \approx o(\frac{1}{n}) \to 0,$$

and thus the last sum converges to zero in the limit.

□

The following theorem is the differential form of (3.17), with the change that X_t is an Itô process instead of a Brownian motion and thus we change the third term in the formula. The proof is similar to Theorem 3.16, but with more terms to keep track of.

Theorem 3.19. *(Generalized Itô-Doeblin Formula) Suppose that X_t is an Itô process, B_t is a Brownian motion, and φ is C^2 in x and C^1 in t. Then*

$$d\varphi(t, X_t) = \partial_t\varphi(t, X_t)dt + \partial_x\varphi(t, X_t)dX_t + \frac{1}{2}\partial_{xx}f(t, X_t)d\langle X\rangle_t. \tag{3.20}$$

We illustrate the usefulness of the Itô-Doeblin formula by finding a strong solution to (1.3).

Example 3.21. Let B_t be a Brownian motion and $\{\mathcal{F}_t\}$ be its natural filtration. Suppose $m(t, X_t) = m_t$ and $\sigma(t, X_t) = \sigma_t$ are adapted processes in L^2. Suppose further that

$$X_t = X_0 \exp\left\{ \left(m_t - \frac{\sigma_t^2}{2} \right) t + \sigma_t B_t \right\} = \varphi(t, B_t). \tag{3.22}$$

Then

$$\partial_t\varphi(t, B_t) = \left(m_t - \frac{\sigma_t^2}{2} \right) X_t, \quad \partial_x\varphi(t, B_t) = \sigma_t X_t, \quad \partial_{xx}\varphi(t, B_t) = \sigma_t^2 X_t.$$

Theorem 3.19 gives

$$dX_t = \left[\left(m_t - \frac{\sigma_t^2}{2} \right) X_t + \frac{1}{2}\sigma_t^2 X_t \right] dt + \sigma_t X_t dB_t = m_t X_t dt + \sigma_t X_t dB_t.$$

Thus, (3.22) is a strong solution to (1.3). We use (3.22) to model an asset price that is non-negative, continuous, and driven by a single Brownian motion. We call m the drift, and we say σ is the volatility.

Remark 3.23. If $m_t \equiv 0$, then (1.3) gives

$$dX_t = \sigma_t X_t dB_t, \leftrightarrow X_t = X_0 + \int_0^t \sigma_s X_s dB_s, \tag{3.24}$$

which is a stochastic integral, and thus a martingale.[7] If $m_t \not\equiv 0$, then X_t is not a martingale, as if $m_t > 0$, then X_t would have a tendency to rise; if $m_t < 0$, then it would have a tendency to fall. In either case, $\mathbb{E}[X_{t+dt} - X_t|\mathcal{F}_t] \neq 0$, and thus X_t is not a martingale. The Itô-Doeblin formula can be used to conclude that

$$X_t = X_0 \exp\left\{\int_0^t \sigma_s dB_s - \frac{1}{2}\int_0^t \sigma_s^2 ds\right\}.$$

is a solution to (3.24).

Remark 3.25. There are a few formal rules associated with differential stochastic calculus, such as

$$dB_t dB_t = dt, \qquad dB_t dt = 0, \qquad dtdt = 0. \tag{3.26}$$

The first comes from Theorem 2.12, as

$$\int_0^t (dB_t)^2 = \langle B\rangle_t = t = \int_0^t dt.$$

Similarly, we have shown that the quadratic variation of $\langle t\rangle_t = 0$, as it is a continuous process. For the remaining rule, one can show that the quadratic co-variation $\langle B, t\rangle_t$ is equally zero.

The following theorem can be easily proved with these formal rules.

Theorem 3.27. *(Stochastic product rule) Suppose X_t, Y_t are Itô processes as in Definition 3.13. Then*

$$d(X_tY_t) = X_t dY_t + dX_t Y_t + d\langle X\rangle_t d\langle Y\rangle_t.$$

4. Girsanov's Theorem and the Risk-Neutral Measure

One considers the value of a stock in the physical world by assessing the risk attributed to it by the market. In other words, its price is associated with its risk. If one wants to consider the fair price associated with a stock without having to discount it for its individual risk profile, then using a risk-neutral probability will allow any stock to be priced by taking the expected value of its payoff. Girsanov's theorem will tell us how Brownian motion will acquire drift when we go from the physical measure $\mathbb{P}$ to a risk neutral measure $\tilde{\mathbb{P}}$ as described in Example 4.4.

Suppose M_t is a martingale given by

$$dM_t = \sigma(t, M_t)M_t dM_t, \quad M_0 = 1,$$

as in Remark 3.23. We define a probability measure $\tilde{\mathbb{P}}$ such that if E is a $\mathcal{F}_t$ measurable event, then $\tilde{\mathbb{P}}(E) = \mathbb{E}[1_E M_t]$. Thus, M_t is the Radon-Nikodym derivative

[7] In general, we can only conclude that if $m_t \equiv 0$, then X_t is a local martingale (a martingale only on specific time intervals). However, we shall make the assumption that X_t is a martingale.

of $\tilde{\mathbb{P}}$ with respect to $\mathbb{P}$. This implies that if Y is an $\mathcal{F}_t$ measurable random variable, then

$$\mathbb{E}_{\tilde{\mathbb{P}}}[Y] = \mathbb{E}[YM_t]. \tag{4.1}$$

We need an additional lemma before proving Girsanov's Theorem, the proof which can be found in the appendix.

Lemma 4.2. *Let $s < t$ and let Y be an $\mathcal{F}_t-$measurable random variable, then*

$$\mathbb{E}_{\tilde{\mathbb{P}}}[Y \mid \mathcal{F}_s] = \frac{1}{M_s}\mathbb{E}[YM_t \mid \mathcal{F}_s].$$

Theorem 4.3. *(Girsanov) Let B_t be a Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\{\mathcal{F}_t\}$ be a filtration for B_t. If*

$$\tilde{B}_t = B_t - \int_0^t \sigma(s, M_s)ds,$$

then with respect to the measure $\tilde{\mathbb{P}}$, which is defined above, $\tilde{B}_t$ is a standard Brownian motion.

In other words, if we weight $\mathbb{P}$ by the martingale, then in the new measure, the Brownian motion acquires drift $\sigma(t, M_t)$. We use Levy's characterization of a Brownian motion (Theorem 6.1) to proof this theorem in the appendix.

Example 4.4. Suppose that X_t is a geometric Brownian motion satisfying

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

under the probability measure $\mathbb{P}$, where B_t is a $\mathbb{P}-$Brownian motion. Our goal is to find a probability measure Q such that X_t satisfies

$$dX_t = r(t, X_t)X_t dt + \sigma(t, X_t)X_t dW_t \tag{4.5}$$

under Q, where W_t is a $Q-$Brownian motion. If

$$W_t = B_t + \int_0^t \frac{r(s, X_s) - m(s, X_s)}{\sigma(s, X_s)}ds,$$

then by plugging in the differential dB_t into the geometric Brownian motion, we arrive at (4.5). Thus, we must define $dQ = M_t d\mathbb{P}$, where M_t is the Radon-Nikodym derivative of Q with respect to $\mathbb{P}$ satisfying

$$dM_t = \frac{r(s, X_s) - m(s, X_s)}{\sigma(s, X_s)}M_t dB_t, \qquad M_0 = 1.$$

Thus, by Girsanov's theorem, W_t is a $Q-$Brownian motion. We call Q the *risk-neutral measure* since under Q, dX_t has a drift of r, which is the risk-free interest rate in the Black-Scholes-Merton equation.

5. Financial Applications: The Black-Scholes-Merton Model

We can now begin discussing the financial applications of stochastic calculus. Suppose the price of a stock, S_t, is a geometric Brownian motion satisfying

$$dS_t = m(t, S_t)S_t dt + \sigma(t, S_t)S_t dB_t.$$

Suppose further that we are engaged in a *European call option* in which, at the expiry time T, an agent has the right (but not the obligation), to buy a share of the stock at price K. Thus, the value of the option at time T is

$$F(S_T) = (S_t - K)_+ = \max\{S_t - K, 0\}.$$

The seller is at an obvious disadvantage here. If $S_t \leq K$, then the buyer won't exercise the option and he will have missed out on $K - S_t$ dollars, and if $S_t > K$, then he should have held on to the stock instead! The game of the Black-Scholes-Merton equation is to find the value of the option at any time $0 < t \leq T$ in order for the seller to charge a premium and not lose out on money.

More generally, if $\varphi(t, S_t)$ denotes the value of an option with payoff $F(T, S_T)$, then if $\varphi(T, S_T) < S_T$, the buyer would exercise the option and collect the payoff, and vice-versa in the other case, creating an opportunity for *arbitrage* (risk-free profit). Our model will make the assumption of an arbitrage-free market, implying that $\varphi(T, S_T) = F(S_T)$. We are interested in how we can insure that the value of our *portfolio* (a collection of stocks and bonds in this case) ends up with the same payoff as $F(S_T)$, and thus we must decide what to do with our portfolio before time T to guarantee this. If we can build a portfolio that is *self-financing* (there are no outside resource added to the portfolio) and with the same payoff as $F(S_T)$, then we could *hedge* (insure) the risk of this option before time T. If Δ_t denotes the number of shares of S_t at time t, and b_t denotes the number of risk-free bonds of price R_t held at time t, then the value of the portfolio (Δ_t, b_t) is given by

$$V_t = \Delta_t S_t + b_t R_t. \tag{5.1}$$

Another arbitrage opportunity, which we will not allow, is if $V_t < \varphi(t, S_t)$ for some $0 \leq t < T$. The strategy in this case is to sell the option for $\varphi(t, S_t)$, invest V_t dollars into the stock, and the remaining $\varphi(t, S_t) - V_t$ would go into the risk-free bond as profit. Likewise, we do not allow $V_t > \varphi(t, S_t)$ for any time t.

Therefore, since we always want to guarantee an arbitrage-free pricing model which at time T yields $V_T = F(S_T) = \varphi(T, S_T)$, then the only option is for $V_t = \varphi(t, S_t)$ at all times $0 \leq t \leq T$. Our goal then becomes in finding a strategy to handle our portfolio by switching between stocks and bonds.

Suppose the risk-free bond has a rate of $r(t, x)$. Then if R_t denotes the value of a risk-free bond at time t,

$$dR_t = r(t, S_t) R_t dt \qquad \Longrightarrow \qquad R_t = R_0 \exp\left\{ \int_0^t r(s, S_s) ds \right\}. \tag{5.2}$$

Since we are discussing the future values of options in the present, we would like to talk about the discounted stock price and discounted portfolio value, respectively given by $\tilde{S} = \frac{1}{R_t} S_t$, and $\tilde{V} = \frac{1}{R_t} V_t$ for $0 \leq t \leq T$.

We define Q as in Example 4.4 to be the risk-neutral measure. Using the product rule, we notice that under Q, the discounted stock rate satisfies

$$\begin{aligned} d\tilde{S}_t &= d(\frac{1}{R_t}S_t) \\ &= d(\frac{1}{R_t})S_t + \frac{1}{R_t}dS_t + d\langle\frac{1}{R}\rangle_t d\langle S\rangle_t \\ &= \frac{-r(t,X_t)}{R_t}S_t dt + \frac{r(t,X_t)S_t dt + \sigma(t,X_t)S_t dW_t}{R_t} \\ &= \sigma(t,X_t)\tilde{S}_t dW_t. \end{aligned}$$

Where the third equality holds because $\frac{1}{R_t}$ is a continuous function with zero quadratic variation. By Example 3.23, we know that $\tilde{S}_t$ is a $Q-$martingale.[8]

We let $\varphi(t,x)$ be the expected value of the option at time t, discounted for the interest rate:

$$V_t = \varphi(t,S_t) = \mathbb{E}_Q\left[\frac{R_t}{R_T}F(S_T) \mid S_t = x\right] = \mathbb{E}_Q\left[\frac{R_t}{R_T}F(S_T) \mid \mathcal{F}_t\right], \tag{5.3}$$

where the third equality holds because S_t is a Markov process (Theorem 2.10). We assume that ϕ is C^1 in t and C^2 in $x = S_t$.

It is not hard to show that $\tilde{V}_t$ is a $Q-$martingale, and thus by the Martingale Representation Theorem found in the appendix as Theorem 6.2, there exists an adapted process A_t such that $d\tilde{V}_t = A_t dW_t$. Thus,

$$\begin{aligned} dV_t &= d(R_t\tilde{V}_t) \\ &= d(R_t)\tilde{V}_t + R_t d(\tilde{V}_t) + d\langle R\rangle_t d\langle\tilde{V}\rangle_t \\ &= dR_t\tilde{V}_t + R_t A_t dW_t \\ &= dR_t\tilde{V}_t + \frac{A_t}{\sigma\tilde{S}_t}R_t d\tilde{S}_t \\ &= dR_t\tilde{V}_t + \frac{A_t}{\sigma\tilde{S}_t}[dS - \tilde{S}_t dR_t] \\ &= \frac{A_t}{\sigma\tilde{S}_t}dS + [\tilde{V}_t - \frac{A_t}{\sigma}]dR_t. \end{aligned}$$

Comparing to (5.1), we conclude that $\Delta_t = \frac{A_t}{\sigma\tilde{S}_t}$, and $b_t = [\tilde{V}_t - \frac{A_t}{\sigma}]$. Thus, barring the fact that A_t is unknown, we have found the portfolio, whose value at time t is given by the Black-Scholes-Merton PDE. We can apply Itô-Doeblin's formula to (5.3).

$$\begin{aligned} d\varphi(t,S_t) &= \partial_t\varphi(t,S_t)dt + \partial_x\varphi(t,S_t)dS_t + \frac{1}{2}\partial_{xx}\varphi(t,S_t)d\langle S\rangle_t \\ &= \partial_t\varphi(t,S_t)dt + \partial_x\varphi(t,S_t)[r(t,S_t)S_t dt + \sigma(t,S_t)S_t dW_t] + \frac{\sigma^2 S_t^2}{2}\partial_{xx}\varphi(t,S_t), \end{aligned}$$

[8]We only know that $\tilde{S}_t$ is a $Q-$local martingale, which is a specific type of martingale. However, we make the strong assumption here that $\tilde{S}_t$ is actually a $Q-$martingale.

Note that $x = S_t$ in all the partial derivatives. We can now compute:

$$\begin{aligned}
d\tilde{V}_t &= d\left[\frac{1}{R_t}\varphi(t, S_t)\right] \\
&= \frac{1}{R_t}\Bigg[\left(-r(t,S_t)\varphi(t,S_t) + \partial_t\varphi(t,S_t) + \partial_x\varphi(t,S_t)r(t,S_t)S_t + \frac{\sigma^2 S_t^2}{2}\partial_{xx}\varphi(t,S_t)\right)dt \\
&\quad + \partial_x\varphi(t,S_t)\sigma(t,S_t)S_t dW_t\Bigg] \\
&= \frac{1}{R_t}\left[J_t\, dt + A_t\, dW_t\right].
\end{aligned}$$

Since $\tilde{V}_t$ is a martingale, then by Remark 3.27, the dt term is zero, and so we derive the B-S-M PDE,

$$0 = J_t = -r(t,S_t)\varphi(t,S_t) + \partial_t\varphi(t,S_t) + \partial_x\varphi(t,S_t)r(t,S_t)S_t + \frac{\sigma(t,S_t)^2 S_t^2}{2}\partial_{xx}\varphi(t,S_t)$$

(5.4)

$$= -r(t,S_t)V_t + \frac{\partial V_t}{\partial t} + r(t,S_t)S_t\frac{\partial V_t}{\partial S_t} + \frac{\sigma(t,S_t)^2 S_t^2}{2}\frac{\partial^2 V_t}{\partial S_t^2}.$$

By setting the risk and volatility as constant, we can derive the explicit Black-Scholes-Merton formula as a solution to (5.4).

Theorem 5.5. *Let both the interest rate and the volatility be constant. Then a solution to the Black-Scholes-Merton PDE (5.4) in the case of a call option is given by*

$$V_t = S_t\Phi\left[\frac{\left(r+\frac{\sigma^2}{2}\right)(T-t) + \ln\frac{x}{K}}{\sigma\sqrt{T-t}}\right] - e^{-r(T-t)}K\Phi\left[\frac{\left(r-\frac{\sigma^2}{2}\right)(T-t) + \ln\frac{x}{K}}{\sigma\sqrt{T-t}}\right],$$

where Φ is the CDF of the standard normal distribution.

Proof. Assuming constant interest rate, we have that the value of the bond at time t is given by $R_t = e^{rt}$. Example 3.21 shows that we can represent the price of the stock by

$$S_t = S_0 \exp\{\left(r - \frac{\sigma^2}{2}\right)t + \sigma dW_t\}.$$

If we let $\tau = T - t$, then using the Markov property of W_t, it can be shown that

$$S_T = S_t \exp\{\left(r - \frac{\sigma^2}{2}\right)\tau + \sigma(W_T - W_t)\} = S_t \exp\{\left(r - \frac{\sigma^2}{2}\right)\tau - \sigma\sqrt{\tau}Y\},$$

where $Y = \frac{W_T - W_t}{\sqrt{T-t}}$ is a standard normal variable. Therefore, we can compute an explicit formula for (5.3) by

$$\begin{aligned}
\varphi(t,S_t) &= \mathbb{E}_Q\left[\frac{R_t}{R_T}(S_t - K)_+ \mid \mathcal{F}_t\right] \\
&= \mathbb{E}_Q\left[e^{-r\tau}(S_t\exp\{\left(r-\frac{\sigma^2}{2}\right)\tau - \sigma\sqrt{\tau}Y\} - K)_+ \mid \mathcal{F}_t\right].
\end{aligned}$$

Since S_t is $\mathcal{F}_t$ measurable and $\exp\{\left(r-\frac{\sigma^2}{2}\right)\tau-\sigma\sqrt{\tau}Y\}$ is independent of $\mathcal{F}_t$, then by letting $x=S_t$,

$$\begin{aligned}\varphi(t,x)&=\mathbb{E}_Q\left[e^{-r\tau}(x\exp\{\left(r-\frac{\sigma^2}{2}\right)\tau-\sigma\sqrt{\tau}Y\}-K)_+\right]\\&=\frac{e^{-r\tau}}{\sqrt{2\pi}}\int_{-\infty}^{\infty}(x\exp\{\left(r-\frac{\sigma^2}{2}\right)\tau-\sigma\sqrt{\tau}y\}-K)_+e^{-\frac{1}{2}y^2}dy.\end{aligned}$$

Since we only care about when the option price is positive, we can restrict the integral such that if

$$d_-(\tau,x)=\frac{1}{\sigma\sqrt{\tau}}\left[\left(r-\frac{\sigma^2}{2}\right)\tau+\ln\frac{x}{K}\right],$$

then

$$\begin{aligned}\varphi(t,x)&=\frac{e^{-r\tau}}{\sqrt{2\pi}}\int_{-\infty}^{d_-(\tau,x)}(x\exp\{\left(r-\frac{\sigma^2}{2}\right)\tau-\sigma\sqrt{\tau}y\}-K)e^{-\frac{1}{2}y^2}dy.\\&=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{d_-(\tau,x)}e^{-r\tau}x\exp\{\left(r-\frac{\sigma^2}{2}\right)\tau-\sigma\sqrt{\tau}y\}e^{-\frac{1}{2}y^2}dy.-\frac{e^{-r\tau}}{\sqrt{2\pi}}\int_{-\infty}^{d_-(\tau,x)}Ke^{-\frac{1}{2}y^2}dy.\\&=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{d_-(\tau,x)}x\exp\{\frac{\sigma^2}{2}\tau-\sigma\sqrt{\tau}y-\frac{1}{2}y^2\}dy.-e^{-r\tau}K\Phi(d_-(\tau,x))\\&=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{d_-(\tau,x)}x\exp\{\frac{-1}{2}(y+\sigma\sqrt{\tau})^2\}dy.-e^{-r\tau}K\Phi(d_-(\tau,x))\\&=\frac{x}{\sqrt{2\pi}}\int_{-\infty}^{d_-(\tau,x)+\sigma\sqrt{\tau}}\exp\{\frac{-1}{2}u^2\}du.-e^{-r\tau}K\Phi(d_-(\tau,x))\\&=x\Phi(d_+(\tau,x))-e^{-r\tau}K\Phi(d_-(\tau,x)),\end{aligned}$$

where $d_+(\tau,x))=\frac{1}{\sigma\sqrt{\tau}}\left[\left(r+\frac{\sigma^2}{2}\right)\tau+\ln\frac{x}{K}\right]$. Explicitly, we derived the Black-Scholes-Merton formula:

$$\varphi(t,S_t)=S_t\Phi\left[\frac{\left(r+\frac{\sigma^2}{2}\right)(T-t)+\ln\frac{x}{K}}{\sigma\sqrt{T-t}}\right]-e^{-r(T-t)}K\Phi\left[\frac{\left(r-\frac{\sigma^2}{2}\right)(T-t)+\ln\frac{x}{K}}{\sigma\sqrt{T-t}}\right].$$

□

6. Appendix

Proof. (Lemma 4.2) By definition, the left hand side is the conditional expectation of the right hand side, so it will suffice to show that the two characterizations of conditional expectation are met as in Remark 1.20. The first is obvious, as the right hand side is, by definition, $\mathcal{F}_s-$measurable. For the second, we must show that, if $E\in\mathcal{F}_s$, then

$$\int_E\frac{1}{M_s}\mathbb{E}_{\tilde{\mathbb{P}}}[YM_t|\mathcal{F}_s]d\tilde{\mathbb{P}}=\int_E Yd\tilde{\mathbb{P}}.$$

Consider that, if E is an $\mathcal{F}_s$ measurable set, then

$$\begin{aligned}\int_E \frac{1}{M_s}\mathbb{E}[YM_t \mid \mathcal{F}_s]d\tilde{\mathbb{P}} &= \int_\Omega \mathbb{1}_E \frac{1}{M_s}\mathbb{E}[YM_t \mid \mathcal{F}_s]d\tilde{\mathbb{P}} \\ &= \mathbb{E}_{\tilde{\mathbb{P}}}\left[\mathbb{1}_E \frac{1}{M_s}\mathbb{E}[YM_t \mid \mathcal{F}_s]\right] \\ &= \mathbb{E}\left[\mathbb{E}[\mathbb{1}_E YM_t]|\mathcal{F}_s\right] \\ &= \mathbb{E}[\mathbb{1}_E YM_s] \\ &= \mathbb{E}_{\tilde{\mathbb{P}}}[Y] \\ &= \int_E Y d\tilde{\mathbb{P}},\end{aligned}$$

where the third and fifth equality follow from (4.1), and the fourth follows from the fact that M_t is a martingale. □

Proofs for the following two theorems can be found in [5].

Theorem 6.1. *(Levy's characterization of Brownian motion) Let $M(t), t \geq 0$, be a martingale relative to a filtration $\mathcal{F}_t$, $t \geq 0$. Assume that $M(0) = 0$ has continuous paths and $\langle M \rangle_t = t$ for all $t \geq 0$. Then $M(t)$ is a Brownian motion.*

We provide a sketch of the proof for Girsanov's theorem (Theorem 4.3) using Levy's characterization of Brownian motion.

Proof. For readability, let $\sigma(t, M_t) = \sigma$. By definition, $\tilde{B}(0) = 0$. By Theorem 2.18, since the integral term, which is continuous, has zero quadratic variation, then $\langle \tilde{B} \rangle_t = t$. By Theorem 6.1, it suffices to show that $\tilde{B}_t$ is a $\tilde{P}-$martingale.

We use the stochastic product rule (Theorem 3.27) and Remark 3.25 to show that $\tilde{B}_t M_t$ is a $\mathbb{P}-$martingale.

$$\begin{aligned}d(\tilde{B}_t M_t) &= \tilde{B}_t dM_t + d\tilde{B}_t M_t + d\tilde{B}_t dM_t \\ &= \tilde{B}_t(\sigma M_t dB_t) + (dB_t - \sigma dt)M_t + (dB_t - \sigma dt)(\sigma M_t dB_t) \\ &= \tilde{B}_t \sigma M_t dB_t + dB_t M_t - \sigma M_t dt + \sigma M_t dB_t dB_t - \sigma^2 M_t dB_t dt \\ &= (\tilde{B}_t \sigma + 1)M_t dB_t.\end{aligned}$$

Since the dt term vanishes, this is just an Itô Integral, and thus the process is a $\mathbb{P}-$martingale. We finish by applying Lemma 4.2 to show that $\tilde{B}_t$ is a $\tilde{\mathbb{P}}-$martingale:

$$\mathbb{E}_{\tilde{\mathbb{P}}}[\tilde{B}_t|\mathcal{F}_s] = \frac{1}{M_s}\mathbb{E}[\tilde{B}_t M_t|\mathcal{F}_s] = \frac{1}{M_s}\tilde{B}_s M_s = \tilde{B}_s.$$

□

Theorem 6.2. *(Martingale Representation) Let W_t, be a Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $\mathcal{F}_t$ be the natural filtration of W_t. Let M_t be a martingale with respect to $\mathcal{F}_t$. Then there is an adapted process A_t, such that*

$$M_t = M_0 + \int_0^t A_s dW_t.$$

7. Acknowledgments

I would like to thank my mentor, Antonis Zitridis, for answering any/all questions and helping me find the resources necessary to write this paper. I also express gratitude to Peter May for running the UChicago Math REU program. Moreover, I would like to thank Daniil Rudenko, Elizaveta Shuvaeva, and Greg Lawler for their wonderful lectures throughout the program.

References

[1] Kiyosi Itô. Stochastic Integral. Mathematical Institute, Nagoya Imperial University. https://pdflink.to/e5e011d4/

[2] Kiyosi Itô. On A Formula Concerning Stochastic Differentials. Mathematical Institute, Nagoya University. https://tinyurl.com/2e73axun

[3] Bru, B., Yor, M. Comments on the life and mathematical legacy of Wolfgang Doeblin. Finance Stochast 6, 3–47 (2002). https://doi.org/10.1007/s780-002-8399-0

[4] Richard. F. Bass. Real Analysis for Graduate Students, Second Edition https://bass.math.uconn.edu/v43.pdf

[5] Steven E. Shreve. Stochastic Calculus for Finance II, Continuous-Time Models. https://files.owenoertell.com/textbooks/finance/stochCal2-shreve.pdf.

[6] Samuel N. Cohen, Robert J. Elliot. Stochastic Calculus and Applications, Second Edition. Springer, 2015.

[7] Gregory F. Lawler. Stochastic Calculus: An Introduction with Applications https://www.math.uchicago.edu/ lawler/finbook.pdf

[8] Joydeep Dutta. Probability and Stochastics for Finance. Indian Institute of Technology, Kanpur. https://www.youtube.com/@probabilityandstochasticsf5278/videos

[9] Friedrich Hirzebruch, Shiing-Shen Chern. World Scientific. Wolf Prize in Mathematics Volume 1 https://www.google.com/books/edition/Wolf$_Prize_in_Mathematics/cSUSUfvT3PoC?hl = engbpv = 0$