AN INTRODUCTION TO STOCHASTIC CALCULUS AND BLACK-SCHOLES OPTION PRICING

ZACHRY WANG

Abstract. This paper is an exposition of the mathematics behind the Black-Scholes model of pricing a European option. After we briefly mention the main definitions of measure-theoretic probability, we define Brownian motion and prove its key properties. Next, we introduce stochastic calculus by constructing the Itō integral and proving Itō's formula. Finally, we use these tools to derive the famous Black-Scholes partial differential equation.

Contents

1. Mathematical Probability

Here, we will define some of the basic machinery of probability.

Definition 1.1. A σ-algebra $\mathcal{F}$ on Ω is a collection of subsets of Ω such that the following are true:

(1) $\emptyset \in \mathcal{F}$.
(2) If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$.
(3) If $\{A_i\}_{i \in \mathbb{N}}$ is a countable subset of $\mathcal{F}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

We also define a sub-σ-algebra as a subcollection of $\mathcal{F}$ that is also a σ-algebra.

Definition 1.2. The function $\mathbb{P} : \mathcal{F} \to [0,1]$ is a probability measure if the following hold:

(1) If $\{A_i\}_{i \in \mathbb{N}}$ are pairwise disjoint sets in $\mathcal{F}$, then $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.
(2) $\mathbb{P}(\Omega) = 1$.

A set Ω, a σ-algebra $\mathcal{F}$, and a probability measure $\mathbb{P}$ make up a probability space, which we denote $(\Omega, \mathcal{F}, \mathbb{P})$.

Date: DEADLINES: Draft AUGUST 18 and Final version AUGUST 29, 2013.

Definition 1.3. The Borel σ-algebra $\mathcal{B}(\mathbb{R})$ is the smallest σ-algebra that contains all open sets in $\mathbb{R}$.

Definition 1.4. A real-valued function $f : \Omega \to \mathbb{R}$ is $\mathcal{F}$-measurable if for all $B \in \mathcal{B}(\mathbb{R})$,

$$f^{-1}(B) \in \mathcal{F}.$$

Definition 1.5. A random variable X is a function $X : \Omega \to \mathbb{R}$ that is $\mathcal{F}$-measurable. Note that we can also get random vectors by replacing $\mathbb{R}$ with $\mathbb{R}^n$.

Proposition 1.6. *$\sigma(X) := \{X^{-1}(B) : B \in \mathcal{B}(\mathbb{R})\}$ forms a σ-algebra.*

Proof. We see that $\emptyset = X^{-1}(\emptyset)$. If $A \in \sigma(X)$, then there exists $B_A \in \mathcal{B}(\mathbb{R})$ such that $X^{-1}(B_A) = A$. Thus, we see that $A^c = (X^{-1}(B_A))^c = X^{-1}(B_A^c) \in \sigma(X)$. If $A_1, A_2, ... \in \sigma(X)$, then there exists $B_1, B_2, ... \in \mathcal{B}(\mathbb{R})$ such that $X^{-1}(B_i) = A_i$. We get that $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty}[X^{-1}(B_i)] = X^{-1}\left(\bigcup_{i=1}^{\infty} B_i\right) \in \sigma(X)$. ☐

We call $\sigma(X)$ the σ-algebra generated by X.

Definition 1.7. The distribution measure of X is the probability measure μ_X that assigns to each $B \in \mathcal{B}(\mathbb{R})$ the measure $\mu_X(B) = \mathbb{P}\{X \in B\}$. If X is a random vector, we can get joint distributions by replacing $\mathbb{R}$ with $\mathbb{R}^n$.

Definition 1.8. The cumulative distribution function of X is the function $F : \mathbb{R} \to [0, 1]$ such that

$$F(x) = \mathbb{P}[X \leq x].$$

Definition 1.9. A random variable X has a density function $f : \mathbb{R} \to [0, \infty)$ if

$$\mathbb{P}[a \leq X \leq b] = \int_a^b f(x)dx.$$

Definition 1.10. A random variable X is a normal random variable with mean μ and variance σ^2, denoted $N(\mu, \sigma^2)$, if its probability density function is equal to

$$\phi(x) = \frac{e^{-(x-\mu)^2/(2\sigma^2)}}{\sigma\sqrt{2\pi}}.$$

In particular, a standard normal random variable is a normal random variable with mean 0 and variance 1.

Definition 1.11. The expectation of X is defined to be

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

given that X is integrable.

Definition 1.12. The variance of X is defined to be

$$\mathrm{Var}(X) = \mathrm{E}[(X - \mathrm{E}[X])^2].$$

Definition 1.13. The covariance of X and Y is defined to be

$$\mathrm{Cov}(X, Y) = \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])].$$

Definition 1.14. A random vector $(X_1, ..., X_n)$ is jointly normal if it has the joint density function

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det C}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mu)C^{-1}(\mathbf{x}-\mu)^T\right), \quad \mathbf{x} = (x_1, \cdots, x_n) \in \mathbb{R}^n,$$

where $\mu = (\mu_1, \cdots, \mu_n)$ is a row vector with $\mu_i = \mathrm{E}[X_i]$, denoted the mean vector, and C is an n by n matrix, with $c_{ij} = \mathrm{Cov}(X_i, X_j)$, denoted the covariance matrix.

Definition 1.15. We call σ-algebras $\mathcal{G}_1, \mathcal{G}_2 \subseteq \mathcal{F}$ independent if

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2)$$

for all $A_i \in \mathcal{G}_i$. We say random variables are independent if the σ-algebras generated by them are independent.

Definition 1.16. Let X be an integrable random variable with respect to $\mathcal{F}$ and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. The conditional expectation of X given $\mathcal{G}$, denoted $\mathrm{E}[X|\mathcal{G}]$, is a random variable on Ω such that

1. $\mathrm{E}[X|\mathcal{G}]$ is $\mathcal{G}$-measurable,
2. For all $A \in \mathcal{G}$,
$$\int_A \mathrm{E}[X|\mathcal{G}](\omega) d\mathbb{P}(\omega) = \int_A X(\omega) d\mathbb{P}(\omega).$$

The existence and uniqueness of conditional expectation is guaranteed by the Radon-Nikodym theorem.

The key properties of conditional expectations are stated here:

Proposition 1.17. *Let X, Y be integrable random variables. The following are true about conditional expectations:*

1. *If $a, b \in \mathbb{R}$, then* $\mathrm{E}[aX + bY|\mathcal{G}] = a\mathrm{E}[X|\mathcal{G}] + b\mathrm{E}[Y|\mathcal{G}]$.
2. *If X is $\mathcal{G}$-measurable, then:*
$$\mathrm{E}[XY|\mathcal{G}] = X\mathrm{E}[Y|\mathcal{G}].$$
3. *If $\mathcal{H}$ is a sub-σ-algebra of $\mathcal{G}$, then*
$$\mathrm{E}[\mathrm{E}[X|\mathcal{G}]|\mathcal{H}] = \mathrm{E}[X|\mathcal{H}].$$
4. *If X is independent of $\mathcal{G}$, then*
$$\mathrm{E}[X|\mathcal{G}] = \mathrm{E}[X].$$

Definition 1.18. A stochastic process is a collection of random variables $\{X_t\}$ indexed by $t \in \mathcal{T}$, where $\mathcal{T}$ is an index set. In this paper, we consider $\mathcal{T} = [0, \infty)$, which represents time. We denote a stochastic process $\{X_t\}$ or $\{X_t\}_{t \in \mathcal{T}}$.

There are typically two ways to think about a stochastic process. For each fixed t, we have a random variable. So we could think of a stochastic process as a collection of random variables indexed by $\mathcal{T}$. For each fixed ω, we have a function from $\mathcal{T}$ to $\mathbb{R}$. So, we could think of a stochastic process as a collection of functions indexed by ω. In the latter, we will call these functions *paths*.

Definition 1.19. The collection $\{\mathcal{F}_t\}$ of σ-algebras of Ω indexed by $t \in \mathcal{T}$ is a filtration if for all $t, s \in \mathcal{T}$ such that $s \leq t$, we have $\mathcal{F}_s \subseteq \mathcal{F}_t$. We denote the filtration $\{\mathcal{F}_t\}$ or $\{\mathcal{F}_t\}_{t \in \mathcal{T}}$.

Definition 1.20. A stochastic process $\{X_t\}$ is $\mathcal{F}_t$-adapted if, for each t, the random variable X_t is $\mathcal{F}_t$-measurable.

Definition 1.21. Consider an $\mathcal{F}_t$-adapted stochastic process $\{M_t\}$. If

$$\mathrm{E}[M_t|F_s] = M_s \text{ for all } 0 \leq s \leq t \leq T,$$

Then $\{M_t\}$ is a martingale.

2. Brownian Motion

Brownian motion will be our model of random motion and will be a fundamental building block as we construct our models representing underlying asset prices. We define the standard Brownian motion as follows:

Definition 2.1. A standard Brownian motion is a stochastic process $\{B_t\}_{t\in\mathcal{T}}$ with the following properties:

1. With probability 1, $B_0 = 0$.
2. For all $0 \leq t_1 \leq t_2 \leq ... \leq t_n$, the increments $B_{t_2} - B_{t_1}, B_{t_3} - B_{t_2}, ..., B_{t_n} - B_{t_{n-1}}$ are independent.
3. For $t \geq s \geq 0$, $B_t - B_s \sim N(0, t-s)$.
4. With probability 1, the function $t \mapsto B_t$ is continuous.

For a rigourous proof of existence of the standard Brownian motion using linear interpolation on the dyadic rationals, see [1].

We create a filtration to model information available at each time t in a standard Brownian motion. We typically equip a standard Brownian motion with an augmented filtration $\{\mathcal{F}_t\}_{t\in\mathcal{T}}$ which satisfies the following properties:

1. For each t, B_t is $\mathcal{F}_t$-measurable.
2. $\mathrm{E}[|B_t|] < \infty$ for all t.
3. For all $t \geq s \geq 0$, $\mathrm{E}[B_t|\mathcal{F}_s] = B_s$.

The details of the construction of $\{\mathcal{F}_t\}_{t\in\mathcal{T}}$ can be found in [5].

For our purposes, we aim to prove three important properties of standard Brownian motion that will be used later in our construction of the stochastic integral. These properties are that a standard Brownian motion is a martingale, is nowhere differentiable with probability 1, and accumulates quadratic variation at a rate of one unit per time.

Theorem 2.2. *Standard Brownian motion is a martingale.*

Proof. Fix s, t such that $0 \leq s \leq t$. Then

$$\begin{aligned}\mathrm{E}[B_t|\mathcal{F}_s] &= \mathrm{E}[B_t - B_s + B_s|\mathcal{F}_s] \\ &= \mathrm{E}[B_t - B_s|\mathcal{F}_s] + \mathrm{E}[B_s|\mathcal{F}_s] \\ &= \mathrm{E}[B_t - B_s] + B_s \\ &= B_s.\end{aligned}$$

□

Martingales are often thought of as a model of a “fair game”. In a financial context, if we model an asset’s price as a Brownian motion, then the martingale property says that the expected future price based on all information up to this point in time will be the price today. The martingale is the mathematical formulation of the Efficient Market Hypotheisis.

Theorem 2.3. *With probability 1, the paths generated by a standard Brownian motion are nowhere differentiable.*

Proof. It suffices to show that the paths are nowhere differentiable on the interval [0,1]. Our first observation is that if $f(t)$ is a continuous function on [0,1] and is differentiable at some $s \in [0,1]$, then f satisifes a Lipschitz condition at s. That is, there exists $L \in \mathbb{N}$ such that

$$|f(t) - f(s)| \leq L|t-s| \qquad (0 \leq t \leq 1).$$

This is true because

$$\lim_{h\to 0} \frac{f(s+h) - f(s)}{h} = f'(s)$$

implies $\exists \delta > 0$ s.t. for $|t-s| < \delta$, we have $|f(t) - f(s)| < (|f'(s)| + 1)\,|t-s|$. On the other hand, for $|t-s| \geq \delta$, we have $|f(t)-f(s)| \leq 2M = \frac{2M}{\delta}\delta \leq \frac{2M}{\delta}|t-s|$, where M is an upper bound of $|f(t)|$ on [0,1]. Thus,

$$|f(t) - f(s)| \leq \max\left\{ (|f'(s)| + 1), \frac{2M}{\delta} \right\} |t-s|.$$

For $L \in \mathbb{N}$, let

$$A_L := \{\omega \in \Omega : \exists s \in [0,1] \text{ s.t. } \forall t \in [0,1], |B_t(\omega) - B_s(\omega) \leq L|t-s|\}.$$

Using our previous observation, we see that $\bigcup_{L=1}^{\infty} A_L$ contains all $\omega \in \Omega$ in which the sample path is continuous and differentiable for some $s \in [0,1]$. We will show that $\mathbb{P}\left[\bigcup_{L=1}^{\infty} A_L\right] = 0$.

Fix $n \geq 3$. For each $\omega \in A_L$, there exists $0 \leq k \leq n-3$ such that for some $s \in [k/n, (k+3)/n]$, we have

$$|B_t(\omega) - B_s(\omega)| \leq L|t-s| \qquad (0 \leq t \leq 1).$$

It follows that for $j = k, k+1, k+2$,

$$\begin{aligned} \left|B_{(j+1)/n}(\omega) - B_{j/n}(\omega)\right| &\leq \left|B_{(j+1)/n}(\omega) - B_s(\omega)\right| + \left|B_s(\omega) - B_{j/n}(\omega)\right| \\ &\leq \frac{3L}{n} + \frac{3L}{n} \\ &= \frac{6L}{n}. \end{aligned}$$

However, $B_{(j+1)/n}(\omega) - B_{j/n}(\omega) \sim N(0, 1/n)$, thus

$$\mathbb{P}\left[\left|B_{(j+1)/n}(\omega) - B_{j/n}(\omega)\right| \leq \frac{6L}{n}\right] = \mathbb{P}\left[|Z| \leq \frac{6L}{\sqrt{n}}\right] \leq \frac{12L}{\sqrt{n2\pi}}.$$

By independence of Brownian increments, we see that for each $0 \leq k \leq n-3$,

$$\mathbb{P}\left(\bigcap_{k \leq j \leq k+2} \left\{ \left|B_{(j+1)/n}(\omega) - B_{j/n}(\omega)\right| \right\} \right) \leq \left(\frac{12L}{\sqrt{2\pi n}} \right)^3 < \frac{1728L^3}{n^{3/2}}.$$

Finally, by taking the union over these sets for $k = 0, 1, ..., n-3$, we see that

$$\begin{aligned} 0 \leq \mathbb{P}[A_L] &\leq \mathbb{P}\left[\bigcup_{k=0}^{n-2}\left(\bigcap_{k\leq j\leq k+2}\left\{\left|B_{(j+1)/n}(\omega) - B_{j/n}(\omega)\right|\right\}\right)\right] \\ &\leq (n-2)\frac{1728L^3}{n^{3/2}} \to 0 \text{ as } n \to \infty. \end{aligned}$$

Hence, $\mathbb{P}[A_L] = 0$. □

The idea of an everywhere continuous function that is nowhere differentible is pretty bizzare for its own sake. The first of these functions to be explicitly constructed was by Weierstrass in 1872. It can be proven that the Riemann-Stieltjes integral is defined only for an integrator of bounded variation. But a function of bounded variation is almost everywhere differentiable. The nowhere differentiability of a standard Brownian motion implies that Brownian motion is not of bounded variation, and hence, the Riemann-Stieltjes integrals cannot be applied. This is one of the primary reasons for the creation of the Itō integral.

Definition 2.4. Let $f : [0, \infty) \to \mathbb{R}$ be a function. Define

$$Q_\Pi(f, T) = \sum_{j=0}^{n-1}[f(t_{j+1}) - f(t_j)]^2,$$

where $\Pi = \{t_0, t_1, ..., t_n\}$ with $t_i = \frac{iT}{n}$. The quadratic variation of f up to time T is defined to be

$$Q(f, T) = \lim_{||\Pi||\to 0} Q_\Pi(f, T),$$

provided the limit exists.

Theorem 2.5. *For each $T \geq 0$, $Q(B_t, T) = T$ with probability 1, where B_t is a sample path of Brownian motion.*

Proof. To prove our claim, it suffices to show that $Q_\Pi(B_t, T)$, which is a random variable, has expectation T and its variance approaches 0 as $||\Pi|| \to 0$. We start by proving the first part of this by observing that

$$\begin{aligned} \mathrm{E}[Q_\Pi(B_t, T)] &= \sum_{j=0}^{n-1}\mathrm{E}[(B_{t_{j+1}} - B_{t_j})^2] \\ &= \sum_{j=0}^{n-1}\mathrm{Var}(B_{t_{j+1}} - B_{t_j}) \\ &= \sum_{j=0}^{n-1} t_{j+1} - t_j = T. \end{aligned}$$

Next, we consider the variance.

$$\begin{aligned}
&\mathrm{Var}[(B_{t_{j+1}} - B_{t_j})^2] \\
&= \mathrm{E}[((B_{t_{j+1}} - B_{t_j})^2 - (t_{j+1} - t_j))^2] \\
&= \mathrm{E}[(B_{t_{j+1}} - B_{t_j})^4] - 2(t_{j+1} - t_j)\mathrm{E}[(B_{t_{j+1}} - B_{t_j})^2] + (t_{j+1} - t_j)^2 \\
&= \mathrm{E}[(B_{t_{j+1}} - B_{t_j})^4] - 2(t_{j+1} - t_j)(t_{j+1} - t_j) + (t_{j+1} - t_j)^2 \\
&= \mathrm{E}[(B_{t_{j+1}} - B_{t_j})^4] - (t_{j+1} - t_j)^2
\end{aligned}$$

We know that $B_{t_{j+1}} - B_{t_j} \sim \sqrt{(t_{j+1} - t_j)}Z$, where $Z \sim N(0,1)$. Since $\mathrm{E}[Z^4] = 3$, we have that

$$\mathrm{E}[(B_{t_{j+1}} - B_{t_j})^4] = \mathrm{E}[(t_{j+1} - t_j)^2 Z^4] = 3(t_{j+1} - t_j)^2.$$

Hence,

$$\begin{aligned}
&\mathrm{Var}[(B_{t_{j+1}} - B_{t_j})^2] \\
&= \mathrm{E}[(B_{t_{j+1}} - B_{t_j})^4] - (t_{j+1} - t_j)^2 \\
&= 3(t_{j+1} - t_j)^2 - (t_{j+1} - t_j)^2 = 2(t_{j+1} - t_j)^2,
\end{aligned}$$

and

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

Thus, we have

$$\lim_{||\Pi|| \to 0} \mathrm{Var}[Q_\Pi(B_t, T)] = 0.$$

Applying Chebyshev's inequality gives us that for any $\lambda \geq 0$,

$$\begin{aligned}
&\mathbb{P}\left(|Q_\Pi(B_t, T) - T| \geq \lambda\right) \\
&= \mathbb{P}\left(|Q_\Pi(B_t, T) - \mathrm{E}\left[Q_\Pi(B_t, T)\right]| \geq \lambda\right) \\
&\leq \frac{1}{\lambda^2}\mathrm{E}\left[Q_\Pi(B_t, T)\right] \to 0 \quad \text{as} \quad \lambda \to 0.
\end{aligned}$$

Hence $Q_\Pi(B_t, T)$ converges to T in probability. Because of this, there exists a subsequence of $Q_\Pi(B_t, T)$ that converges to T with probability 1. That is,

$$Q(B_t, T) = \lim_{||\Pi|| \to 0} Q_\Pi(B_t, T) = T, \text{ with probability } 1,$$

and we prove the theorem. □

The significance of this calculation is that the quadratic variation is not zero. It can be shown that the quadratic variation of functions with continuous derivatives is zero. Hence, quadratic variation terms are almost never seen in standard calculus. Due to this difference, when we are developing stochastic calculus, the stochastic "fundamental theorem of calculus" (Itō's formula) will look different from the standard fundamental theorem of calculus to incorporate the effects of quadratic variation.

3. The Itō Integral

So why are we trying to build this stochastic integral? In the big picture, we will have two pieces. First, we will have a stochastic process with its accompanying filtration representing a random asset price. Second, we want to create a stochastic process that will represent a particular portfolio management strategy. The purpose of the integral will be to show how much money will be made given the strategy and a model for asset price movements. The strategy will be represented by the integrand and the movement of the asset will be represented by the integrator. Our construction will very closely follow the construction in [2].

Definition 3.1. Let $V = V(S,T)$, where $S, T \in \mathbb{R}_{\geq 0}$, be the set of functions (stochastic processes)

$$X : [0,\infty) \times \Omega \to \mathbb{R}$$

which satisfies the following properties:

(1) $X(t,\omega)$ is $\mathcal{B} \times \mathcal{F}$ measurable, where $\mathcal{B}$ is the Borel σ-algebra on $[0,\infty)$.
(2) $X(t,\omega)$ is $\mathcal{F}_t$-adapted.
(3) $\mathrm{E}\left[\int_S^T X^2(t,w)dt\right] < \infty$.

Just like the Riemann integral is built using step functions, we will first define stochastic integrals directly on a very well-behaved subset of V, which we will call elementary processes. Then, for the remaining processes in V, we will define the stochastic integral by approximation using elementary processes

Definition 3.2. We call $H_n \in V$ an elementary process if for each ω, the path can be written in the form

$$H_n(t,\omega) = \sum_{j\geq 0} e_j(\omega) I_{[t_j,t_{j+1})}(t),$$

where e_j are indexed random variables, n is a fixed positive integer, $I_{[t_j,t_{j+1})}$ is an indicator function, and

$$t_j := \begin{cases} j2^{-n} & \text{if } S \leq j2^{-n} \leq T \\ S & \text{if } j2^{-n} < S \\ T & \text{if } j2^{-n} > T. \end{cases}$$

Definition 3.3. Given the elementary process in definition 3.2, we define its Itō integral as follows:

$$\int_S^T H_n(t,\omega)dB_t := \sum_{j\geq 0} e_j(\omega)[B_{t_{j+1}} - B_{t_j}](\omega).$$

Example 3.4. Let

$$H_n(t,\omega) = \sum_{j\geq 0} B_{t_j} I_{[t_j,t_{j+1})}(t).$$

If we go back to our portfolio allocation analogy,

$$\int_S^T H_n(t,\omega)dB_t$$

represents the returns we will get from initially holding B_0 of the asset and at times $t_j, j = 1, 2, ...,$ we change to holding B_{t_j} of the asset. We see that

$$\begin{aligned}\mathrm{E}\left[\int_S^T H_n(t,\omega)dB_t\right] &= \sum_{j\geq 0} \mathrm{E}[B_{t_j}(B_{t_{j+1}} - B_{t_j})] \\ &= \sum_{j\geq 0} \mathrm{E}[B_{t_j}]\mathrm{E}[(B_{t_{j+1}} - B_{t_j})] \\ &= 0.\end{aligned}$$

The following lemma is crucial for extending the Itō integral to general integrands.

Lemma 3.5. *If $H_n(t,w) \in V$ is elementary, then*

$$\mathrm{E}\left[\left(\int_S^T H_n(t,\omega)dB_t\right)^2\right] = \mathrm{E}\left[\int_S^T (H_n(t,\omega))^2 dt\right].$$

Proof. Notice that for $i < j$,

$$\begin{aligned}&\mathrm{E}\left[e_i(B_{t_{i+1}} - B_{t_i})e_j(B_{t_{j+1}} - B_{t_j})\right] \\ &= \mathrm{E}\left[e_i(B_{t_{i+1}} - B_{t_i})e_j\right]\mathrm{E}\left[(B_{t_{j+1}} - B_{t_j})\right] \\ &= 0\end{aligned}$$

because $(B_{t_{j+1}} - B_{t_j})$ is independent of $\mathcal{F}_{t_j}$. Thus,

$$\begin{aligned}\mathrm{E}\left[\left(\int_S^T H_n(t,\omega)dB_t\right)^2\right] &= \sum_{i,j} \mathrm{E}\left[e_i(B_{t_{i+1}} - B_{t_i})e_j(B_{t_{j+1}} - B_{t_j})\right] \\ &= \sum_i \mathrm{E}\left[e_i(B_{t_{i+1}} - B_{t_i})e_i(B_{t_{i+1}} - B_{t_i})\right] \\ &= \sum_i \mathrm{E}\left[(e_i)^2\right]\mathrm{E}\left[(B_{t_{i+1}} - B_{t_i})^2\right] \\ &= \sum_i \mathrm{E}\left[(e_i)^2\right](t_{i+1} - t_i) \\ &= \mathrm{E}\left[\int_S^T (H_n(t,\omega))^2 dt\right].\end{aligned}$$

□

Lemma 3.6. *Let $H \in V$. Then there exists a sequence of elementary functions $H_n \in V$ such that*

$$\mathrm{E}\left[\int_S^T (H - H_n)^2 dt\right] \to 0$$

as $n \to \infty$.

Details of the proof can be found in [2].

Using lemmas 3.5 and 3.6, we can show that $\left\{\int_S^T H_n dB_t\right\}$ is a Cauchy sequence in $L^2(P)$ space. Since $L^2(P)$ is complete, we see that

$$\lim_{n\to\infty}\int_S^T H_n(t,\omega)dB_t$$

is a well-defined limit. This allows us to define the Itō integral.

Definition 3.7. Let $H \in V(S,T)$. We define its Itō integral as

$$\int_S^T H(t,\omega)dB_t = \lim_{n\to\infty}\int_S^T H_n(t,\omega)dB_t \qquad (\text{in } L^2(P))$$

where $\{H_n\}$ is a sequence of elementary functions such that

$$\mathrm{E}\left[\int_S^T (H(t,\omega)-H_n(t,\omega))^2 dt\right] \to 0$$

as $n \to \infty$.

4. Itō's Formula

In this section, we prove the "stochastic fundamental theorem of calculus" or Itō's formula.

Theorem 4.1. *(Itō's formula for the time-independent case with respect to Brownian motion) Suppose f is a C^2 function and $\{B_t\}_{t\in\mathcal{T}}$ is a standard Brownian motion. Then for each t,*

$$f(B_t) - f(B_0) = \int_0^t f'(B_s)dB_s + \frac{1}{2}\int_0^t f''(B_s)ds.$$

This can also be written in differential form as

$$df(B_t) = f'(B_t)dB_t + \frac{1}{2}f''(B_t)dt.$$

Proof. We define a partition of the interval $[0,T]$ by $t_i = iT/n$ for $0 \le i \le n$. Then

$$f(B_T) - f(0) = \sum_{i=1}^{n}\big(f(B_{t_i}) - f(B_{t_{i-1}})\big). \tag{4.2}$$

We will estimate each term $f(B_{t_i}) - f(B_{t_{i-1}})$ using a two-term Taylor expansion

$$f(y) - f(x) = (y-x)f'(x) + \frac{1}{2}(y-x)^2 f''(x) + r(x,y).$$

Here, $r(x,y)$ is the remainder term, which is bounded by

$$\big|r(x,y)\big| \le (y-x)^2 h(x,y),$$

where $h(x,y)$ is a nonnegative, uniformly continuous, bounded function with $h(x,x) = 0$.

We can rewrite equation 4.2 using the Taylor expansion, so we have

$$
\begin{aligned}
f(B_T) - f(0) &= \sum_{i=1}^{n} \big(f(B_{t_i}) - f(B_{t_{i-1}})\big) \\
&= \sum_{i=1}^{n} f'(B_{t_{i-1}})(B_{t_i} - B_{t_i}) \\
&+ \frac{1}{2}\sum_{i=1}^{n} f''(B_{t_{i-1}})(B_{t_i} - B_{t_{i-1}})^2 \\
&+ \sum_{i=1}^{n} r(B_{t_{i-1}}, B_{t_i}),
\end{aligned}
\tag{4.3}
$$

with the remainder terms $r(B_{t_{i-1}}, B_{t_i})$ bounded by

$$\left|\sum_{i=1}^{n} r(B_{t_{i-1}}, B_{t_i})\right| \leq \sum_{i=1}^{n} (B_{t_i} - B_{t_{i-1}})^2 h(B_{t_{i-1}}, B_{t_i}).$$

We deal with the terms of (4.3) one by one. For the first term of (4.3), since f' is continuous,

$$\sum_{i=1}^{n} f'(B_{t_{i-1}})(B_{t_i} - B_{t_i}) \to \int_0^T f'(B_t)dB_t, \text{ in probability.}$$

We rewrite the second term of (4.3) as follows:

$$
\begin{aligned}
&\frac{1}{2}\sum_{i=1}^{n} f''(B_{t_{i-1}})(B_{t_i} - B_{t_{i-1}})^2 \\
&\quad = \frac{1}{2}\sum_{i=1}^{n} \big[f''(B_{t_{i-1}})(B_{t_i} - B_{t_{i-1}})^2 - (t_i - t_{i-1})\big] \\
&\quad + \frac{1}{2}\sum_{i=1}^{n} f''(B_{t_{i-1}})(t_i - t_{i-1}).
\end{aligned}
\tag{4.4}
$$

Since f'' is continuous, the limit of the second term of (4.4) is equal to

$$\lim_{n\to\infty} \frac{1}{2}\sum_{i=1}^{n} f''(B_{t_{i-1}})(t_i - t_{i-1}) = \frac{1}{2}\int_0^T f''(B_t)dt.$$

We bound the first term of (4.4) using that fact that for $Z \sim N(0,1)$, we have that $\mathrm{E}[Z^4] = 3$, which implies that $\mathrm{Var}[Z^2] = \mathrm{E}[Z^4] - \mathrm{E}[Z^2]^2 = 3 - 1 = 2$.

Using the definition of Brownian motion and the properties of expectation and variance, we get that

$$B_{t_i} - B_{t_{i-1}} \sim N(0, \frac{T}{n}) \sim \sqrt{\frac{T}{n}}Z,$$

which implies that

$$\mathrm{E}[(B_{t_i} - B_{t_{i-1}})^2] = \mathrm{E}\left[\frac{T}{n}Z^2\right] = \frac{T}{n}\mathrm{E}[Z^2] = \frac{T}{n}$$

and

$$\mathrm{Var}((B_{t_i} - B_{t_{i-1}})^2) = \mathrm{Var}\big(\frac{T}{n}Z^2\big) = \frac{T^2}{n^2}\mathrm{Var}\big(Z^2\big) = 2\frac{T^2}{n^2}.$$

Next, we show that the summands in the first term of (4.4) have mean zero. This is true because

$$\begin{aligned}
&\mathrm{E}\left[f''(B_{t_{i-1}})\big((B_{t_i}-B_{t_{i-1}})^2-\frac{T}{n}\big)\big|\mathcal{F}_{t_{i-1}}\right]\\
&=f''(B_{t_{i-1}})\mathrm{E}\left[\big((B_{t_i}-B_{t_{i-1}})^2-\frac{T}{n}\big)\big|\mathcal{F}_{t_{i-1}}\right]\\
&=f''(B_{t_{i-1}})\mathrm{E}\left[\big((B_{t_i}-B_{t_{i-1}})^2-\frac{T}{n}\big)\right].
\end{aligned}$$

Now, we check that the summands are uncorrelated. If $i<j$, then

$$\begin{aligned}
&\mathrm{E}\left[f''(B_{t_{i-1}})\big((B_{t_i}-B_{t_{i-1}})^2-\frac{T}{n}\big)f''(B_{t_{j-1}})\big((B_{t_j}-B_{t_{j-1}})^2-\frac{T}{n}\big)\big|\mathcal{F}_{t_{j-1}}\right]\\
&=f''(B_{t_{i-1}})\big((B_{t_i}-B_{t_{i-1}})^2-\frac{T}{n}\big)f''(B_{t_{j-1}})\mathrm{E}\left[\big((B_{t_j}-B_{t_{j-1}})^2-\frac{T}{n}\big)\big|\mathcal{F}_{t_{j-1}}\right]\\
&=0,\ \text{a.s.}
\end{aligned}$$

Thus we have that

$$\begin{aligned}
&\mathrm{Var}\left(\sum_{i=1}^{n}\left[f''(B_{t_{i-1}})(B_{t_i}-B_{t_{i-1}})^2-\frac{T}{n}\right]\right)\\
&=\sum_{i=1}^{n}\mathrm{Var}\left(f''(B_{t_{i-1}})(B_{t_i}-B_{t_{i-1}})^2-\frac{T}{n}\right)\\
&\leq\|f''\|_{L^\infty}^2\sum_{i=1}^{n}\mathrm{Var}\left((B_{t_i}-B_{t_{i-1}})^2-\frac{T}{n}\right)\\
&=\|f''\|_{L^\infty}^2\sum_{i=1}^{n}\mathrm{Var}\left((B_{t_i}-B_{t_{i-1}})^2\right)\\
&=\|f''\|_{L^\infty}^2 n\frac{2T^2}{n^2}\to 0,\ \text{as } n\to\infty
\end{aligned}$$

Thus, we see that the limit of the first term of (4.4) becomes

$$\sum_{i=1}^{n}\left[f''(B_{t_{i-1}})(B_{t_i}-B_{t_{i-1}})^2-\frac{T}{n}\right]\to 0,\ \text{as } n\to\infty.$$

To prove that the third term of (4.3) goes to zero as $n \to \infty$, we use the Cauchy-Schwartz inequality and get

$$\begin{aligned}
\mathrm{E}\left[\left|\sum_{i=1}^{n} r(B_{t_{i-1}}, B_{t_i})\right|\right] &\leq \mathrm{E}\left[\sum_{i=1}^{n}(B_{t_i} - B_{t_{i-1}})^2 h(B_{t_{i-1}}, B_{t_i})\right] \\
&= \sum_{i=1}^{n} \sqrt{\mathrm{E}[(B_{t_i} - B_{t_{i-1}})^4]\mathrm{E}[h(B_{t_{i-1}}, B_{t_i})^2]} \\
&= \sum_{i=1}^{n} \sqrt{\mathrm{E}\left[\frac{T^2}{n^2} Z^4\right] \mathrm{E}[h(B_{t_{i-1}}, B_{t_i})^2]} \\
&= \sum_{i=1}^{n} \sqrt{\frac{3T^2}{n^2} \mathrm{E}[h(B_{t_{i-1}}, B_{t_i})^2]} \\
&= \sum_{i=1}^{n} \frac{\sqrt{3}T}{n} \sqrt{\mathrm{E}[h(B_{t_{i-1}}, B_{t_i})^2]}.
\end{aligned}$$

By construction, we now that for any $\epsilon > 0$, there exists $\delta > 0$ such that $|h(x, y)| < \epsilon$, given that $|x - y| < \delta$. Thus, we know that

$$\begin{aligned}
\mathrm{E}[h(B_{t_{i-1}}, B_{t_i})^2] &\leq \|h\|_{L^\infty}^2 \mathbb{P}\left(|B_{t_i} - B_{t_{i-1}}| > \delta\right) + \epsilon^2 \\
&= \|h\|_{L^\infty}^2 \mathbb{P}(\sqrt{\frac{T}{n}}|Z| > \delta) + \epsilon^2 \leq 2\epsilon^2
\end{aligned}$$

for sufficiently large n. Therefore,

$$\mathrm{E}\left[\sum_{i=1}^{n}(B_{t_i} - B_{t_{i-1}})^2 h(B_{t_{i-1}}, B_{t_i})\right] \leq \sqrt{6}T\epsilon$$

for sufficiently large n, which implies that

$$\sum_{i=1}^{n}(B_{t_i} - B_{t_{i-1}})^2 h(B_{t_{i-1}}, B_{t_i}) \to 0$$

in probability as $n \to \infty$. We thus finish the proof of the theorem. ☐

Example 4.5. Let $f(x) = x^3$. Then $f'(x) = 3x^2, f''(x) = 6x$. Thus,

$$f(B_t) = \int_0^t 3B_u^2 dB_u + \frac{1}{2}\int_0^t 6B_u du = \int_0^t 3B_u^2 dB_u + \int_0^t 3B_u du$$

We now define a standard Itō process to expand the class of stochastic processes that we can use as integrators for our stochastic integral. In our context, it allows us to use other stochastic processes for underlying price movement besides a standard Brownian motion.

Definition 4.6. We say that a process $\{X_t\}$ is a standard Itō process if

$$X_t = X_0 + \int_0^t a(\omega, s)ds + \int_0^t b(\omega, s)dB_s,$$

where $t \in [0, T]$, and a and b are adapted, measurable processes that satisfy the integrability conditions

$$\mathbb{P}\left(\int_0^T |a(\omega, s)|ds < \infty\right) = 1 \text{ and } \mathbb{P}\left(\int_0^T (b(\omega, s))^2 ds < \infty\right) = 1$$

We now state the general form of the Itō formula. The proof is similar, and will be omitted here.

Theorem 4.7. *(Itō formula for a general Itō process) Suppose* $f(t,x)$ *is a function where* $f_t(t,x), f_x(t,x)$ *and* $f_{xx}(t,x)$ *are defined and continuous and* $\{X_t\}$ *is an Itō process with the format above. Then,*

$$f(t,X_t) = f(0,0) + \int_0^t \frac{\partial f}{\partial t}(s,X_s)ds + \int_0^t \frac{\partial f}{\partial x}(s,X_s)dX_s$$
$$+\frac{1}{2}\int_0^t \frac{\partial^2 f}{\partial x^2}(s,X_s)b^2(\omega,s)ds.$$

This equation can also be written in differential form as

$$df(t,X_t) = \frac{\partial f}{\partial t}(t,X_t)dt + \frac{\partial f}{\partial x}(t,X_t)dX_t$$
$$+\frac{1}{2}\frac{\partial^2 f}{\partial x^2}(t,X_t)b^2(\omega,t)dt.$$

5. Derivation of the Black-Scholes Equation

Definition 5.1. A derivative is a financial security in which the value is dependent on the state of another financial security (the underlying security).

Definition 5.2. A call option is a derivative that gives the owner the right (but not obligation) to buy an underlying asset for a predetermined price (the strike price).

Similarly, a put option gives the owner the right to sell for a predetermined price. European options predetermine the date that the owner can exercise the option, while American options give the user freedom to exercise at anytime before a specified date.

The general approach to pricing a derivative security is to

1. Outline how much the security pays at each time given each possible state of the underlying security.
2. Create a *replicating portfolio.* That is, create a portfolio allocation strategy that has the same value at each time and at each possible state as the derivative.
3. Calculate the price of the replicating portfolio. This will be the price of the derivative security.

The philosophy for pricing derivatives this way is rooted in the *no-arbitrage* assumption. An arbitrage opportunity is a way to make riskless profit. More precisely, it is an asset allocation strategy that is guaranteed to have no negative net cash flows at any time and positive net cash flow in at least one state. This assumption is theoretically sound and for the most part empirically true. Arbitrage opportunities are often quickly discovered, and quickly exploited until prices move to eliminate the opportunity.

If the price of the derivative is any other price than the price of the replicating portfolio, then there will be an arbitrage opportunity. If the derivative is overpriced, sell it and buy the replicating portfolio. If the derivative is underpriced, doing the opposite. Either way, this will violate the no-arbitrage assumption. Thus, the derivative will have the same price as the replicating portfolio, assuming no arbitrage.

Here we will contruct a very simple example to illustrate these ideas.

Example 5.3. Suppose we live in a two-period world consisting of time 1 and time 2. Let S be a stock that is worth \$2 at time 1, and could be worth either \$4 or \$1 at time 2. Let V be a derivative security that, at time 2, is worth \$3 if the price of S moves to \$4 and is worth \$0 if the price of S moves to \$1 (V is a European call option with strike price \$1). Assume also that we lend and borrow money with no interest (using a bond, β). So the question is, what should the price of the derivative security be at time 1?

To find the replicating portfolio, we need to find weights a and b such that $aS+b\beta$ gives the same payoff as the derivative V. We know that:

1. if S is worth \$4 in period 1, the derivative is worth \$3 and the replicating portfolio is worth $4a + b$. So we need $3 = 4a + b$.
2. if S is worth \$1 in period 1, the derivative is worth \$0 and the replicating portfolio is worth $a + b$. So we need $0 = a + b$.

Thus, we get that $a = 1$ and $b = -1$ and the price of V at time 1 should be $1 \cdot 2 + (-1) \cdot 1 = 1$.

Remark 5.4. An important observation is that the price *does not* depend on the probability of the stock going up or down. As a consequence, the price does not depend on the expectation. Take for example, if S in the above example had a 1/2 probability of going up and 1/2 probability of going down. Then $\mathrm{E}[V] = 1.5 \neq 1$.

For the European call option, on the exercise date, if the price of the underlying security is less than the strike price, then the call option is worthless. In example 5.3, when the price of S is \$1 at time 2, V is worthless. Otherwise, the contract is worth the price of the underlying security minus the strike price. In example 5.3, when the price of S is \$4 at time 2, V is worth \$3. The value of the call option as a function of the stock price on the expiration date is

$$h(x) = \max\{0, x - (\text{strike price})\}$$

Now we begin our derivation of the Black-Scholes equation. Let S_t model the price of the underlying security, which is assumed to follow a geometric Brownian motion

$$dS_t = \mu S_t dt + \sigma S_t dB_t. \tag{5.5}$$

Let β_t model the price of a bond, which is assumed to be a deterministic process that increases in value exponentially

$$d\beta_t = r\beta_t dt. \tag{5.6}$$

Now we want to build our replicating portfolio. Let a_t and b_t be adapted processes that represent the combination of stock and bonds we hold at time t, respectively. At any given time, the value of the portfolio is

$$V_t = a_t S_t + b_t \beta_t$$

Since this is a replicating portfolio, we have

$$V_T = h(S_T).$$

Next, we assume that our replicating portfolio is *self-financing*. That is, any change in value in the portfolio must come from either a change in value of the stock or a change in value of the bond.

$$dV_t = a_t dS_t + b_t d\beta_t. \tag{5.7}$$

Now, substitute (5.5) and (5.6) into (5.7) to get

$$\begin{aligned} a_t dS_t + b_t d\beta_t = dV_t &= a_t\left[\mu S_t dt + \sigma S_t dB_t\right] + b_t\left[r\beta_t dt\right] \\ &= \left[a_t \mu S_t + b_t r \beta_t\right] dt + \left[a_t \sigma S_t\right] dB_t. \end{aligned} \tag{5.8}$$

On the other hand, we assume that $V_t = f(t, S_t)$ for some well-behaved, smooth function f. Thus we apply the Itō formula for a general Itō process (theorem 4.7) to get that

$$\begin{aligned} dV_t = &\; f_t(t,S_t)dt + \frac{1}{2} f_{xx}(t,S_t) dS_t \cdot dS_t + f_x(t,S_t) dS_t \\ = &\; \left[f_t(t,S_t) + \frac{1}{2} f_{xx}(t,S_t)\sigma^2 S_t^2 + f_x(t,S_t)\mu S_t\right] dt + f_x(t,S_t)\sigma S_t dB_t. \end{aligned} \tag{5.9}$$

Since (5.8) and (5.9) must be equal, the coefficients of the dt terms must be equal and the coefficients of the dB_t terms also must be equal.

When we set the coefficients of the dB_t terms equal, we get

$$a_t = f_x(t, S_t).$$

Similarly, setting the coefficients of the dt terms equal gives us

$$b_t = \frac{1}{r\beta_t}\left[f_t(t,S_t) + \frac{1}{2} f_{xx}(t,S_t)\sigma^2 S_t^2\right]. \tag{5.10}$$

Thus, we substitute (5.10) for b_t to get

$$\begin{aligned} f(t,S_t) = V_t &= a_t S_t + b_t \beta_t \\ &= f_x(t,S_t) S_t + \frac{1}{r\beta_t}\left[f_t(t,S_t) + \frac{1}{2} f_{xx}(t,S_t)\sigma^2 S_t^2\right]\beta_t. \end{aligned}$$

If we substitute x for S_t, we get

$$f_t(t,x) = -\frac{1}{2}\sigma^2 x^2 f_{xx}(t,x) - r x f_x(t,x) + r f(t,x).$$

Adding the original condition

$$f(T,x) = h(x) \text{ for all } x \in \mathbb{R}_{\geq 0}.$$

gives us the Black-Scholes model for European call options.

Solving the partial differential equation gives us that the price of the European call with a stock price of S, strike price of K, and time remaining of $\tau = T - t$ is equal to

$$S\Phi\left(\frac{\log(S/K) + (r + \frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}}\right) - Ke^{-r\tau}\Phi\left(\frac{\log(S/K) + (r - \frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}}\right),$$

where Φ is the cummulative distribution function of a standard normal random variable.

6. Assumptions and Potential Risks

Up until this point, we have carefully built up the mathematical theory. However, it is of paramount importance to understand the assumptions that were used to get the formula. Some of these assumptions include:

(1) Underlying securities move according to a geometric Brownian motion.
(2) Trading is instant and cost-free.
(3) There is a single, constant interest rate and constant volatility.
(4) Security prices move continuously.
(5) Option cannot be exercised until expiration.
(6) No dividends are paid on the security.

Understanding the assumptions allows users to be aware of and appropriately hedge risks that were assumed away in the derivation, such as tail risk, liquidity risk, or volatility risk. In addition, an understanding of the assumptions allows users to seek out the many extensions that exist and are being researched that incorporate real-world features like dividends and American options and broaden the scope of the model by incorporating bells and whistles like geometric fractional Brownian motion for underlying security movements and stochastic interest rates and volatility.

Acknowledgments. I would like to thank my mentor Jacob Perlman for his exceptional guidance and help in both learning the material and the paper-writing process. In addition, I want to thank Professor Peter May and all of the instructors in the math REU for this wonderful program that helped me to both broaden my knowledge of mathematics and increase my mathematical maturity.

References

[1] G.F. Lawler. Stochastic Calculus: An Introduction with Applications. American Mathematical Society. 2010.
[2] B. Oksendal. Stochastic Differential Equations: An Introduction with Applications. Springer-Verlag. 2003.
[3] W. Rudin. Principles of Mathematical Analysis. McGraw-Hill, Inc. 1976.
[4] S.E. Shreve. Stochastic Calculus for Finance. Springer-Verlag. 2004.
[5] J.M. Steele. Stochastic Calculus and Financial Applications. Springer-Verlag. 2001.