STOCHASTIC CALCULUS

JASON ROSS

Abstract. This expository paper is an introduction to stochastic calculus. We work up to Itô's Formula—"the fundamental theorem of stochastic calculus"—defining key items in probability, martingales in discrete time, Brownian motion, and stochastic integration. Along the way, we present important theorems associated with these concepts. This paper also seeks to provide the intuition behind these topics, walking through examples (both mathematical and conceptual), visualizations, and computer simulations.

Contents

1. Introduction

In stochastic calculus, we seek to understand random processes and how they accumulate over time. The random process in question can be incredibly complex—the movement of a particle caused by collisions with millions of other particles or the price fluctuations of an option traded on an exchange—or incredibly simple—flipping a fair coin and taking a step backwards or forwards depending on which face it lands. With stochastic calculus, we define the mathematics underlying such phenomena and use it to analyze how these processes progress.

In some cases (for example, flipping a coin), analyzing the random process at discrete time intervals may suffice. In other cases (for example, particle movements or stock price fluctuations), such a discretization may leave out important details, and we instead want our underlying model to be continuous (Brownian motion). In both cases, we will want to see how certain quantities accumulate over time (stochastic integration).

The above broad overview of stochastic calculus inspires the progression of this paper. First, in Section 2 we define key objects in probability that we will use throughout the paper. In Section 3 we examine expectation and conditional expectation, which are necessary to understand martingales and some of their accompanying theorems—the optional sampling theorem(s) and the martingale convergence theorem—which appear in Section 4. Next, in Section 5 we define Brownian motion and analyze its properties. Finally, in Section 6 we define stochastic integration for simple and continuous processes, leading us into Itô's Formula in Section 7. Along the way, we will go through many examples to see the intuition behind these concepts.

2. Fundamentals of Probability

In this section, we will define key objects in probability such as probability spaces and random variables. To define the former, we must first define an algebra and a σ-algebra of a set Σ.

Definition 2.1. (Algebra on a Set). Let Σ be a set and $\mathcal{F}$ be a collection of subsets of Σ. Then $\mathcal{F}$ is an *algebra on* Σ if it satisfies the following:
(1) $\emptyset \in \mathcal{F}$,
(2) if $S_1 \in \mathcal{F}$, then $\Sigma \backslash S_1 \in \mathcal{F}$, where $\Sigma \backslash S_1$ is the complement of S_1 relative to Σ,
(3) and if $S_1, S_2 \in \mathcal{F}$, then $S_1 \cup S_2 \in \mathcal{F}$.

Definition 2.2. (σ-Algebra on a Set). Let Σ be a set, and let $\mathcal{F}$ be an algebra on Σ. Then $\mathcal{F}$ is a *σ-algebra on* Σ if for countable $S_1, S_2, ... \in \mathcal{F}$, we have that

$$\bigcup_{i=1}^{\infty} S_i \in \mathcal{F}.$$

Observe that all σ-algebras are algebras, but not all algebras are σ-algebras. In particular, algebras are closed under pairwise unions, whereas σ-algebras are closed under countably infinite unions. As we will soon see, these definitions will be essential in understanding the event space $\mathcal{F}$ of a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 2.3. (Probability Space). A *probability space* is a triple $(\Omega, \mathcal{F}, \mathbb{P})$ that comprises the following three components:

(1) a *sample space* Ω, which is the set of all possible outcomes of a random process,
(2) an *event space* $\mathcal{F}$, which is a σ-algebra of the sample space (we will expand on this point soon),
(3) and a *probability function* $\mathbb{P} \colon \Omega \to [0, 1]$, which maps an outcome ω in the sample space Ω to a number between 0 and 1.

Note that all probabilities must sum to 1, that is to say,

$$\sum_{\omega \in \Omega} \mathbb{P}(\omega) = 1.$$

It is worth mentioning that the sample space Ω can be either discrete (finite or countably infinite) or uncountable. A discrete sample space could be the possible outcomes of flipping a coin:

$$\Omega_{\text{coin}} = \{\text{Heads, Tails}\}.$$

An uncountable sample space could be the possible percentage changes of a stock price:

$$\Omega_{\text{change in price}} = \mathbb{R}.$$

To better provide the intuition behind an event space σ-algebra, $\mathcal{F}$ is the set containing all possible sets whose elements are elements of the sample space Ω. In the above coin-flip example, the set {"Heads," "Tails," "Heads"} would be an element of the event space $\mathcal{F}$, as would the set containing one-million "Heads" or even the set with one-billion "Tails." Importantly, even for probability spaces with finite sample spaces, the event space $\mathcal{F}$ is always infinite. More subtly, the order in which elements appear in these sets (sometimes) matters. For example, the set {"Heads," "Tails"} is not necessarily "the same" as {"Tails," "Heads"} for all situations.[1]

We now move on to defining a random variable. Intuitively, a random variable X is a variable whose value depends on the outcome of a random process. We include a more formal definition below.

Definition 2.4. (Random Variable). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A *random variable* is a function $X : \Omega \to \mathbb{R}$. Letting S be a subset of $\mathbb{R}$, the probability $\mathbb{P}(X \in S)$ that X takes on some value in S is

$$\mathbb{P}(X \in S) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \in S\}) = \sum_{\substack{\omega \in \Omega \\ X(\omega) \in S}} \mathbb{P}(\omega).$$

In other words, suppose we randomly select an outcome ω from the sample space Ω according to the probability function $\mathbb{P}$. Then X maps ω to a real number.

The statement in Definition 2.4 about a random variable X taking on some value in $S \subset \mathbb{R}$ is most applicable when our sample space Ω is uncountable. In the case Ω is discrete, if we were to select $X(\omega) \in \mathbb{R}$ in the codomain of X, the probability our random variable takes on the value $X(\omega)$ is likely not zero. To illustrate, we return to our coin-flip example. Suppose

$$X(\omega) = \begin{cases} 1 & \omega = \text{ Heads} \\ 0 & \omega = \text{ Tails.} \end{cases}$$

Assume our coin is fair, in other words, that $\mathbb{P}(\text{Heads}) = \mathbb{P}(\text{Tails}) = \frac{1}{2}$. If we choose $1 = X(\text{Heads}) \in \mathbb{R}$, then the probability our random variable takes on the value 1 is $\frac{1}{2} \neq 0$.

However, when Ω is uncountable, the probability our random variable takes on a specific value $X(\omega)$ in the codomain of X may very well be zero. In our example of the percentage change of a stock price, the probability the percentage change is exactly $\sqrt{2}$ is zero. Instead, it is more worthwhile to determine the probability the percent change is within some range $S \subset \mathbb{R}$, which is likely nonzero.

Now, we define conditional probability and independence.

Definition 2.5. (Conditional Probability and Independence). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $A, B \in \Omega$. The *conditional probability* $\mathbb{P}(A|B)$, or the probability A occurs given B occurs, is

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

[1] We will see later that these two sets can be considered "the same" for a process such as random walk, where only the current value of the process matters. In this way, random walk is an example of a *Markov* process.

Notice that, if A and B are *independent* (that is, the probability A occurs does not depend on whether B occurs and vice versa) then $\mathbb{P}(A|B) = \mathbb{P}(A)$.

For the last definition of this section, we define the indicator function.

Definition 2.6. (Indicator Function). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $A \in \mathcal{F}$ be some event. Then the *indicator function* $\mathbb{1}_A$ or $\mathbb{1}(A)$ is

$$\mathbb{1}_A = \mathbb{1}(A) = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{if } A \text{ does not occur.} \end{cases}$$

3. Expectation and Conditional Expectation

We are now ready to define the expectation (or expected value) $\mathbb{E}[X]$ of a random variable X. Intuitively, if we were to simulate some random process many times, the expectation would be the average of all the outcomes' values. Formal definitions of expectation for discrete and uncountable probability spaces are below.

Definition 3.1. (Expectation/Expected Value for Discrete Probability Spaces). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a discrete probability space, and let $X : \Omega \to \mathbb{R}$ be a random variable. The *expectation* $\mathbb{E}[X]$ of X is

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} \mathbb{P}(\omega) X(\omega).$$

Notice that for probability spaces with discrete sample spaces, the expectation of a random variable may not be a possible outcome, that is to say, it is not necessarily true that $\mathbb{E}[X] \in \{X(\omega)\}_{\omega \in \Omega}$. Recall our coin-flip example where $\mathbb{P}(\text{Heads}) = \mathbb{P}(\text{Tails}) = \frac{1}{2}$ and

$$X(\omega) = \begin{cases} 1 & \omega = \text{ Heads} \\ 0 & \omega = \text{ Tails.} \end{cases}$$

It follows that

$$\mathbb{E}[X] = 0.5 \notin \{X(\omega)\}_{\omega \in \Omega}.$$

We now introduce the notion of expectation for uncountable probability spaces.

Definition 3.2. (Expectation/Expected Value for Uncountable Probability Spaces). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be an uncountable probability space, and let $X : \Omega \to \mathbb{R}$ be a random variable. The *expectation* $\mathbb{E}[X]$ of X is

$$\mathbb{E}[X] = \int_{\Omega} X d\mathbb{P}.$$

It is interesting to see how Definition 3.2 adapts Definition 3.1 to an uncountable sample space. In other words, we increment our probability function $\mathbb{P}$ infinitesimally to integrate over all possible outcomes rather than "counting" all possible outcomes.

We now move on to conditional expectation. Letting X be a random variable, if we think of $\mathbb{E}[X]$ as the best guess for X without any information on trials, then the conditional expectation $E[X]$ is the best guess for X with some information about a trial result. For example, the probability someone draws a queen from a standard deck of cards is $\frac{1}{13}$ without any additional information. But given two queens have already been removed from the deck, the probability is $\frac{2}{25}$.

More mathematically, suppose $X_1, X_2, ...$ is a sequence of random variables, with the data encoded in X_i coming in at time i. Letting Y be another random variable, we have at time n that the conditional expectation of Y given $X_1, ..., X_n$ is

$$E[Y \mid X_1, ..., X_n].$$

Recall that the event space $\mathcal{F}$ in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is the set containing all possible sets whose elements are elements of the sample space Ω. We write $X_1, ..., X_n = \mathcal{F}_n \subset \Omega$. In other words, $\mathcal{F}_n$ is a set that contains the data encoded in $X_1, ..., X_n$. Note that:

(1) $\mathcal{F}_0$ contains no information, meaning $E[Y \mid \mathcal{F}_0] = \mathbb{E}[Y]$,
(2) if Y is independent of $X_1, ..., X_n$, then $E[Y \mid \mathcal{F}_n] = \mathbb{E}[Y]$,
(3) $E[Y \mid \mathcal{F}_n]$ is itself a random variable that we say is $\mathcal{F}_n$-measureable.

We are now ready to formally define conditional expectation.

Definition 3.3. (Conditional Expectation). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $Y : \Omega \to \mathbb{R}$ be a random variable. The *conditional expectation* $E[Y \mid \mathcal{F}_n]$ is the unique random variable such that

(1) $E[Y \mid \mathcal{F}_n]$ is $\mathcal{F}_n$-measurable
(2) and for an $\mathcal{F}_n$-measureable event $A \in \mathcal{F}$, we have
$$\mathbb{E}[E[Y \mid \mathcal{F}_n]\mathbb{1}_A] = \mathbb{E}[Y\mathbb{1}_A].$$

In the following proposition, we examine some properties of conditional expectation.

Proposition 3.4. *Let $X_1, X_2, ...$ be a sequence of random variables where $\mathcal{F}_n$ gives us the information at time n. The conditional expectation $E[Y \mid \mathcal{F}_n]$ is such that*

(1) $E[Y \mid \mathcal{F}_n] = Y$ when Y is $\mathcal{F}_n$-measurable,
(2) $\mathbb{E}[E[Y \mid \mathcal{F}_n]] = \mathbb{E}[Y]$ (this property follows from (2) in Definition 3.3),
(3) if Y is independent of $X_1, ..., X_n$, then $E[Y \mid \mathcal{F}_n] = \mathbb{E}[Y]$,
(4) $E[Y \mid \mathcal{F}_n]$ satisfies linearity, that is, for random variables X, Y and constants a, b, we have
$$E[aX + bY \mid \mathcal{F}_n] = aE[X \mid \mathcal{F}_n] + bE[Y \mid \mathcal{F}_n],$$
(5) $E[Y \mid \mathcal{F}_n]$ satisfies the "tower property," that is to say, for $m < n$, we have
$$E[E[Y \mid \mathcal{F}_n] \mid \mathcal{F}_m] = E[Y \mid \mathcal{F}_m],$$
(6) and for an $\mathcal{F}_n$-measurable random variable X, we have
$$E[YX \mid \mathcal{F}_n] = XE[Y \mid \mathcal{F}_n].$$

Intuitively, these properties should make sense. For detailed proofs of these properties, see Proposition 1.1.1 in [1]. We will apply these properties in Examples 4.3 and 4.4

For our final definition of this section, we define a natural filtration. First, recall that for a sequence of random variables $X_1, ..., X_n$, we have that $\mathcal{F}_n$ holds the information contained in that sequence.

Definition 3.5. (Natural Filtration). Let $X_1, X_2, ...$ be a sequence of random variables. Then the *natural filtration* is the collection $\{\mathcal{F}_n\}$, where $\mathcal{F}_n$ is the information stored in $X_1, ..., X_n$.

Admittedly, a natural filtration's use is somewhat obscure. Intuitively, if $m < n$, then all information known at time m is still known at time n. We will invoke the concept of natural filtrations in the following section when defining martingales.

4. Discrete Time Martingales

Now that we have defined conditional expectation and examined its properties, we are equipped to examine martingales in discrete time. Intuitively, a martingale is a model of a *fair game*, in other words, a game where the probability of winning is equal to the probability of losing. After stating the formal definition of a martingale, we will return to our coin-flip example to build further intuition.

Definition 4.1. (Martingale). Let $M_0, M_1, ...$ be a sequence of random variables associated with the natural filtration $\{\mathcal{F}_n\}$. We call this sequence a *martingale with respect to the natural filtration* $\{\mathcal{F}_n\}$ if the following conditions are met:

(1) for all $n \in \mathbb{N}$, M_n is an $\mathcal{F}_n$-measureable random variable with finite expectation, that is,

$$\mathbb{E}[|M_n|] < \infty$$

(2) and for all $m, n \in \mathbb{N}$ where $m < n$, we have

$$E[M_n \mid \mathcal{F}_m] = M_m.$$

We will refer to condition (2) in Definition 4.1 as the "martingale property." Examining the martingale property, we may equivalently write

$$E[M_n - M_m \mid \mathcal{F}_n] = 0. \tag{4.2}$$

Consider our coin-flip example adapted with the following random variable function:

$$X(\omega) = \begin{cases} 1 & \text{if } \omega = \text{Heads} \\ -1 & \text{if } \omega = \text{Tails.} \end{cases}$$

Let "Heads" be a win and "Tails" be a loss. It is clear that $\mathbb{E}[X] = 0$, that is to say, the expected winnings for any particular game is 0. Putting this example in the context of the martingale property, the sum of winnings by the n^{th} game given the outcomes of the first through m^{th} games is equal to the sum of winnings by the m^{th} game, which we already know. In other words, we expect our winnings to "average" out to zero.

We can show (4.2) holds through applying the tower property of conditional expectation (condition (5) of Proposition 3.4). We want to show

$$E[M_{n+1} \mid \mathcal{F}_n] = M_n.$$

The rest follows by induction. Observe that

$$\begin{aligned} E[M_{n+2} \mid \mathcal{F}_n] &= E[E[M_{n+2} \mid \mathcal{F}_{n+1}] \mid \mathcal{F}_n] && \text{(by tower property)} \\ &= E[M_{n+1} \mid \mathcal{F}_n] \\ &= M_n. \end{aligned}$$

We now move on to constructing a discrete stochastic integral from martingales, which (as promised!) will be a nice application of Proposition 3.4.

Example 4.3. (Discrete Stochastic Integral). Let $M_0, M_1, ...$ be a martingale with respect to $\{\mathcal{F}_n\}$, and put $\Delta M_n = M_n - M_{n-1}$. Let B_i denote the "bet" on the i^{th} game. By "bet," we could mean, for example, predicting that our i^{th} coin-flip is "Tails" (in which case $B_i < 0$, which would give us a positive return on M_i). Finally, let W_m denote the winnings, that is, the sum of the outcomes. For simplicity, set $W_0 = 0$. It follows that

$$W_n = \sum_{i=1}^{n} B_i(M_i - M_{i-1}) = \sum_{i=1}^{n} B_i \Delta M_i.$$

In our coin-flip game, it is clear that B_i is bounded—generally, we will assume this is the case, that is to say, for each n there exists K_n such that

$$|B_n| \leq K_n < \infty.$$

Finally, we assume that B_n is $\mathcal{F}_{n-1}$-measureable, that is to say, that we do not know the result of a game before placing our bet. We want to prove that W_n is a martingale with respect to $\mathcal{F}_n$.

Proof. We will prove W_n satisfies the two conditions outlined in Definition 4.1.[2] We have

$$\begin{aligned}
\mathbb{E}[|W_n|] &= \mathbb{E}\left[\left|\sum_{i=1}^{n} B_i(M_i - M_{i-1})\right|\right] \\
\text{(by triangle inequality)} \qquad &\leq \sum_{i=1}^{n} \mathbb{E}[|B_i|\,|M_i - M_{i-1}|] \\
&\leq \sum_{i=1}^{n} \mathbb{E}[|K_i|\,|M_i - M_{i-1}|] \\
&< \infty.
\end{aligned}$$

This result satisfies condition (1). For proving the martingale property, observe the following progression:

$$\begin{aligned}
E[W_{n+1} \mid \mathcal{F}_n] &= E[W_n + B_{n+1}(M_{n+1} - M_n) \mid \mathcal{F}_n] \\
\text{(by linearity)} \qquad &= E[W_n \mid \mathcal{F}_n] + E[B_{n+1}(M_{n+1} - M_n) \mid \mathcal{F}_n] \\
(W_n \text{ is } \mathcal{F}_n\text{-measurable}) \qquad &= W_n + E[B_{n+1}(M_{n+1} - M_n) \mid \mathcal{F}_n] \\
\text{(condition (6) of Proposition 3.4)} \qquad &= W_n + B_{n+1}E[(M_{n+1} - M_n) \mid \mathcal{F}_n] \\
(4.2) \qquad &= W_n + B_{n+1} \cdot 0 \\
&= W_n.
\end{aligned}$$

We satisfy the martingale property, completing the proof. □

Now that we have our discrete stochastic integral, we can apply it to our coin-flip example.

Example 4.4. (Betting Strategy for a Fair Game). Recall our probability space $(\Omega, \mathcal{F}, \mathbb{P})$ where $\Omega = \{\text{Heads, Tails}\}$, $\mathbb{P}(\text{Heads}) = \mathbb{P}(\text{Tails}) = \frac{1}{2}$, and

$$X(\omega) = \begin{cases} 1 & \text{if } \omega = \text{Heads} \\ -1 & \text{if } \omega = \text{Tails.} \end{cases}$$

[2]The "pre-condition" that W_n is $\mathcal{F}_n$-measureable clearly holds.

Let $X_1, X_2, ...$ be a sequence of random variables corresponding to a sequence of coin-flips. Then $M_n = X_1 + ... + X_n$ is our martingale. Set "Heads" to a win and "Tails" to a loss.

Our betting strategy is as follows. For the i^{th} round, we bet \$$2^i$. For example, our first round bet is \$1, our second round bet is \$2, etc. Upon winning, we quit; upon losing, we continue to play. We will show that this strategy is a winning strategy, that is, M_n will be positive for some $n \in \mathbb{N}$, at which point we will quit.

Proof. Recall our discrete stochastic integral. In this case, we have $\Delta M_i = X_i$, meaning,

$$W_n = \sum_{i=1}^{n} B_i X_i.$$

We want to show that W_n is positive at some point. In fact, the point at which W_n becomes positive is such that $W_n = 1$. In essence, what we want to show is

$$2^n + \sum_{i=0}^{n-1} -(2^i) = 1$$

for all $n \in \mathbb{N}$. We will induct on n. The base case P(0) clearly holds, so we move on to our inductive step.

P(k): We have as our inductive hypothesis that P(k-1) holds, that is,

$$2^{k-1} + \sum_{i=0}^{k-2} -(2^i) = 1.$$

Starting from the above equation, we may write

$$2^{k-1} - \sum_{i=0}^{k-2} 2^i = 1.$$

Multiplying both sides by 2, we get

$$2^k - 2 \cdot \sum_{i=0}^{k-2} 2^i = 2$$

$$2^k - \sum_{i=1}^{k-1} 2^i = 2.$$

Finally, subtracting 1 from both sides, we get

$$2^k - \left(\sum_{i=1}^{k-1} 2^i \right) - 1 = 1$$

$$2^k - \sum_{i=1}^{k-1} 2^i - \sum_{i=0}^{1} 2^i = 1$$

$$2^k - \sum_{i=0}^{k-1} 2^i = 1.$$

This result satisfies the inductive step, completing our proof. $\square$

In Example 4.4, it is worth mentioning that the outlined martingale betting strategy guarantees a win only in infinite time. More mathematically, there is no time $m < \infty$ such that every sequence $X_0, X_1, \ldots$ has $X_n = 1 = X(\text{Heads})$ for $n \leq m$. This phenomenon generalizes to all martingales—you cannot beat a martingale in finite time. However, in our example, it is very likely X_n will be a win for reasonably small n. Indeed,

$$\lim_{n\to\infty} \left(\frac{1}{2}\right)^n = 0,$$

meaning the probability you flip a coin many times and exclusively get "Tails" is approximately zero.

We will now discuss the optional sampling theorem, which cements the fact that a martingale cannot be beat in finite time. To explain the intuition, suppose you purchase a stock whose price is a martingale at time 0. Then it is impossible to make money (in expectation) from the stock's maturing in finite time. In other words, the expectation of the stock's price at any time $n > 0$ is equal to the expectation of the stock's price at time 0.

Before providing the statement of the theorem, we will first define a "stopping time."

Definition 4.5. (Stopping Time). Let $T \in \mathbb{Z}$ be a non-negative random variable. Then T is a *stopping time* with respect to $\{\mathcal{F}_n\}$ if for each $n \in \mathbb{N}$, the event $\{T = n\}$ is $\mathcal{F}_n$-measureable.

Admittedly, this definition is a bit confusing. Going back to Example 4.3 on the discrete stochastic integral, think of T as some random but finite time at which all bets past T are zero, that is to say, $B_m = 0$ for all $m > T$.

Before the theorem statement, we will first provide some new notation. Recall from Example 4.3 the martingale describing the winnings:

$$W_n = M_0 + \sum_{i=1}^{n} B_i(M_i - M_{i-1}).$$

Now, letting T be a stopping time, set

$$B_i = \begin{cases} 1 & i \leq T \\ 0 & i > T. \end{cases}$$

With this betting strategy, we write the expression for the winnings as $M_{n\wedge T} = M_{\min\{n,T\}}$.

Theorem 4.6. *(Optional Sampling Theorem I). Let T be a stopping time and M_n be a martingale with respect to $\{\mathcal{F}_n\}$. Then $Y_n = M_{n\wedge T}$ is a martingale. In other words, for all $n \in \mathbb{N}$,*

$$\mathbb{E}[M_{n\wedge T}] = \mathbb{E}[M_0].$$

Also, if T is bounded, that is, there exists $K < \infty$ such that $\mathbb{P}(T \leq K) = 1$, then

$$\mathbb{E}[M_T] = \mathbb{E}[M_0].$$

Given our definition of $M_{n\wedge T}$, the proof of Theorem 4.6 is straightforward. However, in the theorem, we assume that our stopping time T is bounded by some $K < \infty$. It may not always be the case that such a K exists, that is to say, our stopping time is not necessarily bounded. If we assume that $\mathbb{P}(T < \infty) = 1$, in other words, that our process does indeed stop at M_T, we can derive a weaker result. Observe the following progression:

$$\begin{aligned}\mathbb{E}[M_0] &= \mathbb{E}[M_{n\wedge T}] \\ &= \mathbb{E}[M_T] + \mathbb{E}[M_{n\wedge T} - M_T].\end{aligned}$$

In particular, if $\lim_{n\to\infty} \mathbb{E}[M_{n\wedge T} - M_T] = 0$, we can reach the same conclusion as Theorem 4.6 with our weaker assumption. We get $M_{n\wedge T} - M_T = 0$ when $\min\{n, T\} = n \wedge T = T$, that is, when we reach our stopping time. Also,

$$M_{n\wedge T} - M_T = \mathbb{1}(T > n)[M_n - M_T].$$

Then given $\mathbb{E}[|M_T|] < \infty$, in other words, that M_T is *integrable*, we can show that

$$\lim_{n\to\infty} \mathbb{E}[|M_T|\, \mathbb{1}(T > n)] = 0.$$

Recalling Example 4.4, we have that $W_T = 1$. In particular $\mathbb{E}[W_T] < \infty$, meaning we satisfy the necessary conditions.

We restate the above conclusions in the following theorem.

Theorem 4.7. *(Optional Sampling Theorem II). Let T be a stopping time, and let M_n be a martingale with respect to $\mathcal{F}_n$. Given $\mathbb{P}(T < \infty) = 1$, $\mathbb{E}[|M_T|] < \infty$, and $\lim_{n\to\infty} \mathbb{E}[|M_n|\, \mathbb{1}(T > n)] = 0$ for all n, we may conclude that*

$$\mathbb{E}[M_T] = \mathbb{E}[M_0].$$

Notice that Example 4.4 does *not* follow Theorem 4.7's conclusion, that is to say, $\mathbb{E}[W_T] = 1 \neq 0 = \mathbb{E}[W_0]$. Indeed, Example 4.4 does not satisfy the initial conditions of the theorem. Recall that $W_n = 1 - 2^n$ when $T > n$ and $\mathbb{P}(T > n) = \frac{1}{2^n}$. Then

$$\begin{aligned}\lim_{n\to\infty} \mathbb{E}[|M_n|\, \mathbb{1}(T > n)] &= \lim_{n\to\infty} \left[|2^n - 1| \frac{1}{2^n} \right] \\ &= 1 \neq 0.\end{aligned}$$

Checking whether Example 4.4 satisfies the condition $\lim_{n\to\infty} \mathbb{E}[|M_n|\, \mathbb{1}(T > n)] = 0$ is relatively simple. However, it may not always be so simple to check. The following iteration of the optional sampling theorem provides different criteria that can be met to reach our desired conclusion.

Theorem 4.8. *(Optional Sampling Theorem III). Let T be a stopping time, and let M_n be a martingale with respect to $\mathcal{F}_n$. Suppose $\mathbb{P}(T < \infty) = 1$ and $\mathbb{E}[|M_T|] < \infty$. Also suppose there exists $C < \infty$ such that $\mathbb{E}[M_{n\wedge T}^2] \leq C$ for all n. Then we may conclude that*

$$\mathbb{E}[M_T] = \mathbb{E}[M_0].$$

The criterion that $C < \infty$ such that $\mathbb{E}[M_{n\wedge T}^2] \leq C$ for all n is called *square integrability*. For a proof of Theorem 4.8, see Theorem 1.3.3 in [1]. Theorems 4.6 and 4.7 can be found as Theorems 1.3.2 and 1.3.3, respectively, in [1].

The following example will illustrate some implications of the optional sampling theorem. Before exploring these examples, however, we will first define a random walk.

Definition 4.9. (Random Walk). Let $X_1, X_2, \ldots$ be a sequence of random variables. Then $S_n = X_1 + \ldots + X_n$ is a *simple random walk* beginning at 0.

It is clear that a random walk S_n is a martingale with respect to $X_1, ..., X_n$. In our coin flip example, our random walk is on $\mathbb{Z}$ (or the one-dimensional integer lattice), where we take a "step" of length 1 in the positive direction when we get "Heads" and a "step" of length 1 in the negative direction when we get "Tails." The notion of random walk will be our basis for introducing Brownian motion. We include a more detailed illustration of random walk with a computer simulation in Section 5.

Example 4.10. (Gambler's Ruin). The gambler's ruin problem is as such: given a winning probability p (and a losing probability $1-p$) and a starting sum of money $O > 0$, what is the probability a gambler makes K before losing O (in other words, losing all of his money)?

We will frame this problem in the context of our coin-flip example. Let $X_1, X_2, ...$ be a sequence of independent coin-tosses, and let $S_n = O + X_1 + ... + X_n$. Let $K \in \mathbb{Z}$ such that $K > O$, and let the stopping time T be the first time n where $S_n = 0$ or $S_n = K$ (it is clear that T is a random variable). It follows that $M_n = S_{n \wedge T}$ is a martingale. Applying the optional sampling theorem, we have

$$\begin{aligned} O &= M_0 \\ &= \mathbb{E}[M_T] \\ &= 0 \cdot \mathbb{P}(M_T = 0) + K \cdot \mathbb{P}(M_T = K) && \text{(by Definition 3.1)} \\ &= K \cdot \mathbb{P}(M_T = K). \end{aligned}$$

Dividing both sides by K, we get

$$\frac{O}{K} = \mathbb{P}(M_T = K).$$

If we fix our starting amount O and let $K \to \infty$, it is clear that

$$\lim_{K \to \infty} \mathbb{P}(M_T = K) = \lim_{K \to \infty} \frac{O}{K} = 0.$$

In particular, as $K \to \infty$, we have that $\mathbb{P}(M_T = 0) = 1$ (the gambler almost surely loses all his money, facing his ruin!). The property of a random walk always returning to the origin is called *recurrence*.

For the final part of this section, we discuss the martingale convergence cheorem, which describes the behavior of a martingale M_n as $n \to \infty$.

Theorem 4.11. *(Martingale Convergence Theorem). Let M_n be a martingale with respect to $\mathcal{F}_n$. Suppose there exists $K < \infty$ such that $\mathbb{E}[|M_n|] \leq K$ for all n. Then there exists a random variable M_∞ such that*

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

almost surely.

This theorem is very powerful. Intuitively, a martingale M_n converges with probability one to some well-defined random variable M_∞. Note that $\mathbb{E}[M_\infty]$ does not necessarily equal $\mathbb{E}[M_0]$. Recalling Example 4.4, we had that $\mathbb{E}[W_\infty] = 1 \neq 0 = \mathbb{E}[W_0]$.

For a proof of the martingale convergence theorem, see Theorem 1.4.1 in [1]. Intuitively, Lawler's proof shows that for $a, b \in \mathbb{R}$ where $a < b$, it is impossible for a martingale to fluctuate infinitely often below a and above b. The proof sets up a discrete stochastic integral where one "buys" a unit of a good every time the "price" (value of the martingale) falls below a and "sells" a unit of a good when the "price" passes b.[3] We can show that the value of the discrete stochastic integral is less than ∞. Then letting $a \to b$, we can narrow our discrete stochastic integral's value down to a well-defined random variable.

5. Brownian Motion

Before defining Brownian motion, we will first introduce variance, the normal distribution, and the central limit theorem. In particular, we will discuss the limit distribution of $X_1 + ... + X_n$ where $X_1, ..., X_n$ are independent, identically-distributed random variables with mean μ and variance $\sigma^2 < \infty$ (we will define the latter imminently).

Definition 5.1. (Variance). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $X : \Omega \to \mathbb{R}$ be a random variable. Then the *variance* of X is as follows:

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

Intuitively, we can consider variance as a measure of "how much" a distribution "spreads out" from its average value. Note that we do not explicitly distinguish between discrete and uncountable probability spaces in our definition of variance. Definitions of expectation, however, vary between the two types of probability spaces (see Definitions 3.1 and 3.2).

Definition 5.2. (Standard Normal Distribution Function). We denote the *standard normal distribution function* ($\mu = 0$ and $\sigma^2 = 1$) with Φ:

$$\Phi(b) = \int_{-\infty}^{b} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx.$$

We can find the approximate value of Φ at any b using a computer. For most relevant b (where *relevant* means $\frac{1}{\sqrt{2\pi}} e^{\frac{-b^2}{2}} \not\approx 0$), we can use a table of a standard normal distribution to find the approximate value of $\Phi(b)$.

In Definition 2.4, we defined a random variable with respect to a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Recall how we distinguished between discrete sample spaces and uncountable sample spaces. In particular, we discussed how for uncountable sample spaces Ω, the probability a random variable takes on a specific value $X(\omega)$ is zero. Instead, we may find the probability a random variable takes on some value in a range, which likely has positive probability. In particular, with normally distributed random variables, we are now able to better quantify such a question with the cumulative distribution function (CDF).

Definition 5.3. (Cumulative Distribution Function). Let X be a normally distributed random variable with mean μ and variance σ^2. Then the *cumulative distribution function (CDF)*, or the probability X takes on a value less than or equal to $b \in \mathbb{R}$ is as follows:

$$F_X(b) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{b} \exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right) = \mathbb{P}(X \le b).$$

[3] In other words, B_n changes to 1 when the martingale falls below a (this constitutes a "buy"). When the martingale passes b, the bet becomes 0 (this constitutes a "sell"). Our bet does not become 1 again until the martingale falls below a.

If we want to find the probability a random variable X takes on a value in the range $[a, b]$, we do the following computation:

$$\mathbb{P}(a \leq X \leq b) = F_X(b) - F_X(a).$$

The CDF pertains only to probability spaces with uncountable sample spaces. The analogue for discrete sample spaces is called the *probability mass function.* With a probability mass function, we can assign p as the probability X takes on some value x, for example, $\mathbb{P}(X = x) = p$ where $0 \leq p \leq 1$.

We now state the central limit theorem.

Theorem 5.4. *(Central Limit Theorem). Let $X_1, ..., X_n$ be independent, identically distributed random variables with mean μ and finite variance σ. Define Z_n as*

$$Z_n = \frac{(X_1 + ... + X_n) - n\mu}{\sigma\sqrt{n}}.$$

Then as $n \to \infty$, the distribution of Z_n approaches a standard normal distribution. In particular, for $a, b \in \mathbb{R}$ where $a < b$, we have

$$\lim_{n\to\infty} \mathbb{P}(a \leq Z_n \leq b) = \Phi(b) - \Phi(a).$$

Intuitively, the central limit theorem states that regardless of the distribution of X_i, as long as the variance is finite, that is, $\sigma^2 < \infty$, then the sum $X_1 + ... + X_n$ approaches a normal distribution (which can be scaled to the standard normal distribution). Importantly, note that each X_i is small when compared to the sum $\sum_{n=1}^{\infty} X_n$. This characteristic can be used to apply central limit theorem even when $X_1, ..., X_n$ are not completely independent (for example, in nature). We omit the proof of central limit theorem—one can find such a proof in [7].

For the rest of this paper, we will write $N(\mu, \sigma^2)$ to denote a distribution with mean μ and variance σ^2. For example, we denote the standard normal distribution by $N(0, 1)$, that is, the normal distribution with mean 0 and variance 1.

Now that we have defined the standard normal distribution, we are equipped to introduce Brownian motion. Brownian motion can be thought of as continuous random walk or the limit of random walk. More mathematically, recall from the previous section that for a random walk $S_n = X_1 + ... + X_n$, we have that X_i and X_{i+1} are separated by some time increment $\Delta t = 1$. In our coin-flip example, the space increment Δx also equals 1. We can instead view the process with $\Delta t = \frac{1}{N}$ for $N >>> 0$ (that is, infinitesimal time increments)—this notion is the idea behind Brownian motion.

More formally, let $B_t = B(t)$ be the value of our Brownian motion at time t. We have that B_t is a random variable for each t. Such a collection of time-indexed random variables is called a *stochastic process.* Then the function $t \to B(t)$ is a random variable whose codomain is a function $B(t)$. Note that $B_{i+\Delta t}$ is not independent of B_i. For example, if B_i is a particle's position in meters at time i, then $\mathbb{P}(B_{i+\Delta t} = 0)$ certainly varies when $B_i = 1$ meter versus when $B_i = 100$ meters.

Brownian motion has three defining characteristics:

(1) *Stationary increments:* For $s, t \in \mathbb{R}$ where $s < t$, we have that $B_t - B_s$ and $B_{t-s} - B_0$ are identically distributed.
(2) *Independent increments:* For $s, t \in \mathbb{R}$ where $s < t$, the random variable $B_t - B_s$ is independent of the values B_r for $r \leq s$.

(3) *Continuous paths:* The random variable function $t \to B_t$ is continuous.

We include the formal definition of Brownian motion below as well as a high-level overview of a proof for continuity of Brownian motion.

Definition 5.5. (Brownian Motion). We call a stochastic process B_t *one-dimensional Brownian motion* starting at the origin with *drift* μ and *variance* σ^2 if the following conditions are satisfied:

(1) $B_0 = 0$,
(2) for $s, t \in \mathbb{R}$ where $s < t$, we have that $B_t - B_s$ is normally distributed with mean $\mu(t-s)$ and variance $\sigma^2(t-s)$,
(3) for $s, t \in \mathbb{R}$ where $s < t$, the random variable $B_t - B_s$ is independent of the values B_r for $r \leq s$,
(4) and the function $t \to B_t$ is a continuous function.

If $\mu = 0$ and $\sigma^2 = 1$, we call B_t *standard Brownian motion.*

To establish the existence of Brownian motion, it suffices to prove that a standard Brownian motion exists.[4] To accomplish this goal, we can verify that conditions (1) through (4) hold. However, verifying condition (4) (continuity) holds is quite involved. We will not include a proof of existence in this paper—for a full proof of the existence of Brownian motion, see Section 2.5 in [1]. We will, however, discuss a method for proving continuity in the following two paragraphs. Alternatively, the reader can have faith that Brownian motion does indeed exist and skip this next part.

First, we want to define B_t for t in the set of dyadic rationals, that is, the set of rational numbers whose numerators are integers and denominators are powers of 2. Importantly, the dyadic rationals are a countable dense set, which will allow us to establish uniform continuity (that is, that our Brownian motion does not have arbitrarily large jumps) with relative ease. We focus initially on $t \in [0, 1]$, first defining B_0 and B_1.[5] We then subdivide and define $B_{1/2}$ and then $B_{1/4}$ and $B_{3/4}$ and so on.

Once we have B_t defined on the dyadic rationals between 0 and 1, we can prove B_t has uniformly continuous paths. Then we extend B_t to $t \in \mathbb{R}$ by continuity, at which point we will have satisfied condition (4) of Definition 5.5 for $t \in [0, 1]$. We can then extend B_t to any positive $t < \infty$ by taking a countable collection of Brownian motions on $[0, 1]$.

Before discussing some additional properties of Brownian motion, we state the following theorem.

Theorem 5.6. *(Nowhere Differentiability of Brownian Motion). Let B_t be a Brownian motion. With probability one, the function $t \to B_t$ is nowhere differentiable.*

If we look at any visualization of Brownian motion (see Figures 2 and 3), we observe that the paths are extremely jagged, and the above theorem makes intuitive sense. As a sketch of a proof, suppose an arbitrary Brownian motion B_t were

[4] We will show that any Brownian motion can be scaled to a standard Brownian motion in Proposition 5.7.

[5] Looking specifically at $t \in [0, 1]$ is allowed by the scale invariance property of Brownian motion (see Proposition 5.7). In particular, if we scale our Brownian motion to the interval $[0, 1]$, the result is still Brownian motion.

differentiable at time t. Then we could find the derivative at t by looking at B_s for $0 \leq s \leq t$. Using this derivative, we could then learn something about the increment $B_{t+\Delta t} - B_t$ for $\Delta t > 0$. But $B_{t+\Delta t} - B_t$ is independent of B_s for $0 \leq s \leq t$, leading to a contradiction. For a full proof of Theorem 5.6, see Theorem 1.30 in [2].

We will now look at a few more properties of Brownian motion: scale invariance, the strong Markov property, and the reflection principle. These three properties can be used to do computations with Brownian motion.

Proposition 5.7. *(Scale Invariance). Let B_t be a standard Brownian motion, and let $a > 0$. Then*

$$Y_t = \frac{B_{at}}{\sqrt{a}} = \frac{B(at)}{\sqrt{a}}$$

is a standard Brownian motion.

In particular, observe that when we scale time by a factor a, we scale space by a factor $\frac{1}{\sqrt{a}}$. More intuitively, if we "zoom-in" on a Brownian motion, the result is still a Brownian motion.

Proof. To prove the scale invariance property of Brownian motion, we need to show that Y_t satisfies the conditions for a standard Brownian motion. Observe the following progression:

$$\begin{aligned} \mathrm{Var}[Y_t] &= \mathrm{Var}\left[\frac{B_{at}}{\sqrt{a}}\right] \\ &= \frac{1}{a}\mathrm{Var}[B_{at}] \\ &= \frac{1}{a}(at) = t. \end{aligned}$$

The remaining conditions of stationary increments, independent increments, and continuity remain unchanged under scaling. □

Next, we discuss the strong Markov property of Brownian motion. First, recall Definition 4.5 on a stopping time T with respect to a natural filtration $\{\mathcal{F}_n\}$. Intuitively, our theorem will state that if we "stop" a Brownian motion B_t at a finite time T and then "restart" it, then the process $\{B_{t'} \mid t' \geq T\}$ is also a Brownian motion.

Proposition 5.8. *(Strong Markov Property). Let B_t be a standard Brownian motion, and let T be a stopping time such that $\mathbb{P}(T < \infty) = 1$. Suppose*

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

Then Y_t is a standard Brownian motion independent of $\{B_t \mid 0 \leq t \leq T\}$.

For a proof of the strong Markov property of Brownian motion, see Theorem 2.14 in [2]. Finally, we introduce the reflection principle of Brownian motion. Intuitively, suppose a Brownian motion B_t reaches some value a for the first time at time s, that is, $\min\{t \mid B_t = a\} = s$. Then if we reflect our Brownian motion about a, the result will be identically distributed as the non-reflected Brownian motion—in particular, the reflection will also be a Brownian motion. Figure 1 provides a graphic of a one-dimensional Brownian motion illustrating the reflection principle.

We now provide the formal statement of the reflection principle.

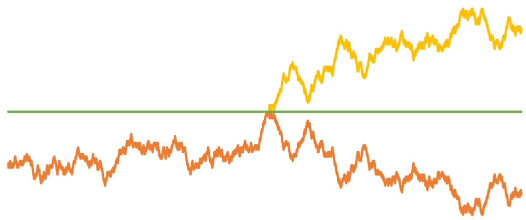

FIGURE 1. The orange line is our original Brownian motion B_t, the green line represents our value a, and the yellow line is the Brownian motion beginning at the time s reflected about a.

Proposition 5.9. *(Reflection Principle). Let B_t be a standard Brownian motion with $B_0 = 0$, and let $a > 0$. Then*

$$\mathbb{P}\left(\sup_{0\leq s\leq t} B_s \geq a\right) = 2\mathbb{P}(B_t > a) = 2\left[1 - \Phi\left(\frac{a}{\sqrt{t}}\right)\right].$$

Proof. Let T_a denote the first time our Brownian motion B_t equals a, that is,

$$T_a = \min\{s \mid B_s = a\} = \min\{s \mid B_s \geq a\}.$$

Note that we can switch out $=$ with $\geq$ as a direct consequence of Brownian motion's continuity. We have that

$$\mathbb{P}\left(\sup_{0\leq s\leq t} B_s \geq a\right) = \mathbb{P}(T_a \leq t) = \mathbb{P}(T_a < t).$$

In other words, since T_a is the first time our Brownian motion equals a, the probability $\sup_{0\leq s\leq t} B_s$ meets or exceeds a is zero when $t < T_a$ and one when $t \geq T_a$. We draw the second inequality from the fact $\mathbb{P}(T_a = t) \leq \mathbb{P}(B_t = a) = 0$, that is, the probability our Brownian motion equals any particular $a \in \mathbb{R}$ is zero. Now, because $B_{T_a} = a$, it must be that

$$\begin{aligned}
\mathbb{P}(B_t > a) &= \mathbb{P}(T_a < t \text{ and } B_t > a) \\
&= \mathbb{P}(T_a < t)\mathbb{P}(B_t - B_{T_a} > 0 \mid T_a < t) \\
&= \frac{1}{2}\mathbb{P}(T_a < t). && \text{(Proposition 5.8)}
\end{aligned}$$

This result satisfies the first equality of Proposition 5.9. For the second equality, observe

$$\begin{aligned}
\mathbb{P}(B_t > a) &= \mathbb{P}\left(B_1 > \frac{a}{\sqrt{t}}\right) && \text{(Proposition 5.7)} \\
&= 1 - \Phi\left(\frac{a}{\sqrt{t}}\right).
\end{aligned}$$

This final result yields the second equality of the proposition. □

Now that we have defined Brownian motion and analyzed several of its properties, we can discuss where Brownian motion might appear in the real world. Looking back at martingales, we had that each M_t was indexed by some integer t. For our coin-flipping example in particular, we knew the exact size of each jump was 1—the "randomness" was in the direction of each jump. Even if we were to introduce randomness in the direction *and* size of the jump, for example, have each jump be $N(0, 1)$, our martingale would still have the limitation of being indexed by discrete time values.

In the real world, analyzing random processes at discrete time intervals often fails to give a complete understanding of what is truly happening. Consider, for example, the motion of a particle. At each moment, any particular particle collides with millions of other moving particles, sending the particle into some direction at some velocity. If we were to analyze the position of the particle, say, every second (or even every hundredth or thousandth of a second), we would lose out on the position of the particle at every time in between! In particular, random processes like particle movements are *continuous*. Assuming the position of the particle is constant except at discrete time intervals is not sufficient to fully understand its movement. For every $t \in \mathbb{R}$, the position x_t of the particle is different!

As another example, consider the randomly fluctuating price of an option traded on an exchange. Suppose we assume the price is constant except at discrete intervals of time, for example, every second, when creating our trading strategy. Then we would fail to see price fluctuations at every time in between each interval, missing out on hundreds of opportunities to buy and sell! Particularly with algorithmic trading, computers are able to observe the market price, do any necessary computations, and execute trades all in a few thousandths of a second, easily beating out the human who takes a half second to even click a mouse button. We see again that ignoring continuity of real world random processes causes us to miss out on what is truly happening.

As promised, we will now look at a computer simulation of random walk, which will provide a nice visualization of Brownian motion. We will look at visualizations of one- and two-dimensional Brownian motion. The pseudocode of our program for one-dimensional Brownian motion is below (the program can be easily adapted to visualize Brownian motion in higher dimensions)—a full implementation in imperative C can be found in the Appendix.

```
// One-Dimensional Brownian Motion Psuedocode
Set int current_pos to 0

// Loop begins here
Generate a random int x

Determine whether x is congruent to 0 (mod 2) or 1 (mod 2)

   If x is congruent to 0 (mod 2), increment current_pos by 1

   If x is congruent to 1 (mod 2), decrement current_pos by 1

Print current_pos

Repeat the above loop as desired (preferably many times)
```

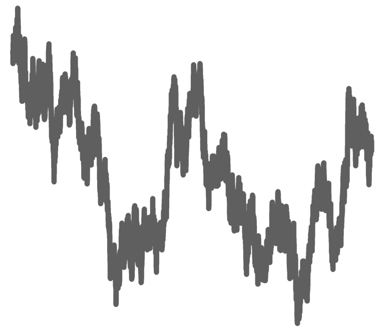

FIGURE 2. A depiction of one-dimensional Brownian motion.

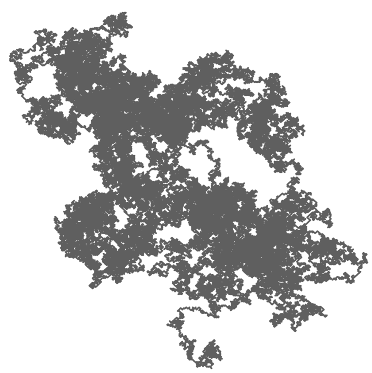

FIGURE 3. A depiction of two-dimensional Brownian motion.

When we export the output of our program as a comma separated values file, we can read it in Excel and produce visualizations such as Figures 2 and 3.

Really, our program is a simulation of a simple random walk, that is to say, we randomly decide to take a step of size one in either the positive or negative direction. The key part that allows us to produce these visualizations of Brownian motion is how many times we decide to repeat the process. In particular, when we run the simulation many times (such as tens of thousands of times) and "zoom-out," the process becomes approximately continuous, in other words, our discrete time intervals become incredibly small. In this way, we are able to recreate Brownian motion.

6. Stochastic Integration

In the final two sections of this paper, we bring together everything we have discussed to introduce stochastic integration and Itô's Formula. Before examining the mathematics of stochastic integration, we will provide the motivation in the context of the two examples discussed near the end of Section 5—particle movements and the prices of tradeable assets.

With stochastic integration, we want to see how some quantity accumulates over a random process. With the moving particle, we may want to find its displacement. In the case of an asset's price, we may want to find how much profit we can expect to make using some trading strategy. For example, with a buy-and-hold strategy, we buy a stock at time u and sell at a time $v > u$. How much money can we expect to make? In another example, suppose we buy a stock when it reaches price a and sell when it reaches price $b > a$ (in other words, stopping times). What is our expected profit? These types of questions (and far more sophisticated ones!) can be answered using stochastic integration.

We now move into mathematical definitions and properties of stochastic integration. First, we will briefly recall the Riemann integral.

Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function on $[a, b]$. We first partition $[a, b]$, that is to say, we define t_i for $0 \le i \le n$ such that $a = t_0 < t_1 < ... < t_n = b$. Then

$$\int_a^b f(x)dx = \lim_{n\to\infty} \sum_{i=1}^{n} \sup\{f(x) \mid t_{i-1} \le x \le t_i\}(t_i - t_{i-1})$$
$$= \lim_{n\to\infty} \sum_{i=1}^{n} \inf\{f(x) \mid t_{i-1} \le x \le t_i\}(t_i - t_{i-1}).$$

In particular, note that as we refine our partition, the infimum and supremum of f on each subinterval converge to one value. In this way, we can approximate $f(x)$ for each partition using a *step function*, that is,

$$f_n(x) = f(s_i), \quad t_{i-1} < x < t_i,$$

where $s_i \in [t_{i-1}, t_i]$. Then

$$\int_a^b f(x)dx = \lim_{n\to\infty} \sum_{i=1}^{n} f(s_i)(t_i - t_{i-1}).$$ [6]

For a stochastic integral, the analogue of a Riemann integral's step function is a *simple process* A_t. Intuitively, A_t is a betting strategy where we can change our bets at a finite set of predetermined times. The formal definition of a simple process is below.

Definition 6.1. (Simple Process). A process A_t is a *simple process* if there exist times $0 = t_0 < t_1 < ... < t_n < \infty$ where $t_{n+1} = \infty$ and corresponding random variables

$$\{Y_i\}_{0 \le i \le n} \text{ where } \mathbb{E}[Y_i^2] < \infty$$

that are $\mathcal{F}_{t_i}$-measurable such that

$$A_t = Y_i \text{ for } t_i \le t < t_{i+1}.$$

Notice that, since each Y_i is $\mathcal{F}_{t_i}$-measurable, we have that A_t is $\mathcal{F}_t$-measurable. Now that we have defined a simple process, we can define its stochastic integral.

[6] To be precise, the integral defined with the supremum/infimum enclosed within the summation is the Darboux integral, whereas the integral with a step function enclosed within the summation is the Riemann integral. Nonetheless, it can be shown that a function is Darboux integrable function if and only if it is Riemann integrable.

Definition 6.2. (Stochastic Integral for a Simple Process). Let A_t be a simple process, and let B_t be a standard Brownian motion with respect to $\{\mathcal{F}_t\}$. Then we define the *stochastic integral*

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

by

$$Z_{t_i} = \sum_{j=0}^{i-1} Y_j (B_{t_{j+1}} - B_{t_j}).$$

In particular, note that

$$Z_t = Z_{t_i} + Y_i(B_t - B_{t_i}) \quad \text{if} \quad t_i \le t \le t_{i+1},$$

and

$$\int_r^t A_s dB_s = Z_t - Z_r.$$

We can think of Z_t as a Brownian motion with variance A_t^2 at time t. Note that the stochastic integral $Z_t = \int_0^t A_s dB_s$ is a random variable, whereas the Riemann integral is a real number. Similarly to Example 4.3, we consider A_t as the amount "bet" at time t, where A_t is $\mathcal{F}_{t-1}$-measurable (that is, we do not know the result of the process before placing our bet).

In the next several paragraphs, we examine the differences between stochastic and Riemann integration, discussing precisely why the classical techniques of ordinary calculus cannot be applied to stochastic integration. Alternatively, the reader can accept that the two methods of integration are not the same and skip to Proposition 6.5, which describes properties of the stochastic integral for simple processes.

We define a stochastic integral with respect to infinitesimal (random) increments dB_t of Brownian motion, whereas we define a Riemann integral with respect to infinitesimal (predictable) increments dt. The nuances underlying this difference between the two infinitesimals is the principal reason why we are not able to apply techniques of non-stochastic calculus to stochastic calculus. To fully understand these nuances, we first need to define total variation.

Definition 6.3. (Total Variation). Let $a, b \in \mathbb{R}$ with $b > a$, and let $f\colon [a,b] \to \mathbb{R}$. Let $\mathcal{P} = \{P = \{x_0, ..., x_{n_P}\} \mid P \text{ is a partition of } [a,b]\}$ be the set of all partitions of $[a,b]$. We define the *total variation* $V_a^b(f)$ of f on $[a,b]$ as

$$V_a^b(f) = \sup_{\mathcal{P}} \sum_{i=0}^{n_P} \left| f(x_i) - f(x_{i-1}) \right|,$$

where the supremum runs over the set of all partitions $\mathcal{P}$ of $[a,b]$. When the supremum is infinite, we say that f has *infinite total variation*; when the supremum is finite, we say that f has *finite total variation.*

Remark 6.4. (Total Variation of a One-Dimensional Function). For a differentiable function $f\colon \mathbb{R} \to \mathbb{R}$ with derivative f', the total variation of f on $[a,b]$ is given by

$$V_a^b(f) = \int_a^b \left| f'(x) \right| dx.$$

In the Riemann integral, our infinitesimal is dt (or $t_i - t_{i-1}$). In the Riemann-Stieltjes integral, a generalization of the Riemann integral, the infinitesimal becomes $dg(t)$ (or $g(t_i)-g(t_{i-1})$), where g is an increasing function on $[a, b]$.[7] More explicitly, given a partition $P = \{t_0, ..., t_n\}$ of $[a, b]$, a function f, and an increasing function g, the Riemann-Stieltjes integral is defined to be the limit of the sum

$$\sum_{i=1}^{n} f(s_i)[g(t_i) - g(t_{i-1})] \qquad \text{where } s_i \in [t_{i-1}, t_i]$$

as the length of the largest subinterval of the partition P (that is, mesh(P)) approaches 0. We call f the *integrand* and g the *integrator*. Importantly, all increasing functions have finite total variation on $[a, b]$, making them suitable integrators for Riemann-Stieltjes integration.

With stochastic integrals, however, our integrator is a Brownian motion B_t. In particular, Brownian motion does *not* have finite total variation. We can explain this intuitively through scale invariance. No matter how much we "zoom in" on Brownian motion, we will always still see Brownian motion. Figure 2 shows how jagged Brownian motion's paths are—this jaggedness does not go away regardless of how much we "zoom in." Consequently, as the length of the longest subinterval of our partition approaches 0, the difference $|B_{t_i} - B_{t_{i-1}}|$ does *not* go to 0. In other words, the summation from Definition 6.3

$$\sum_{i=0}^{n_P} |B_{t_i} - B_{t_{i-1}}|$$

continues to grow as the partitions P become increasingly refined. Consequently, the total variation of Brownian motion is infinite, that is,

$$\sup_{\mathcal{P}} \sum_{i=0}^{n_P} |B_{t_i} - B_{t_{i-1}}|$$

does not exist. Because B_t has infinite total variation, it is *not* a suitable integrator for the Riemann-Stieltjes integral, meaning techniques for non-stochastic integration do not work for stochastic integration.[8]

In the following proposition, we will examine some properties of the stochastic integral.

Proposition 6.5. *(Properties of the Stochastic Integral for Simple Processes). Let B_t be a standard Brownian motion with respect to $\{\mathcal{F}_n\}$, and let A_t and C_t be simple processes. Then the following properties hold.*

(1) Linearity: For $a, c \in \mathbb{R}$, we have that $aA_t + cC_t$ is a simple process and

$$\int_0^t (aA_s + cC_s) dB_s = a \int_0^t A_s dB_s + c \int_0^t C_s dB_s.$$

Also, for $0 < r < t$, we have that

$$\int_0^t A_s dB_s = \int_0^r A_s dB_s + \int_r^t A_s dB_s.$$

[7] Of course, t is an increasing function on $[a, b]$.

[8] For a full proof of why Brownian motion has infinite total variation, see Proposition 3.2.7 in [1].

(2) Martingale property: The following process is a martingale with respect to $\{\mathcal{F}_t\}$*:*

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

(3) Variance rule: $\mathbb{E}[Z_t^2] < \infty$ *(square integrability) and*

$$\mathrm{Var}[Z_t] = \mathbb{E}[Z_t^2] = \int_0^t \mathbb{E}[A_s^2] ds.$$

(4) Continuity: The function $t \to Z_t$ *is almost surely a continuous function.*

We will not include a full proof of Proposition 6.5 in this paper. Linearity follows from Definition 6.2, and continuity follows from the continuity of Brownian motion (Definition 5.5). Proving the martingale property and variance rule, however, is slightly more involved. A full proof of the proposition can be found as Proposition 3.2.1 in [1].

Now, we wish to consider a *continuous process* A_t, that is to say, a process such that each A_t is $\mathcal{F}_t$-measurable. We will first consider the case where such an A_t is bounded, in other words, that there exists $K < \infty$ such that $|A_t| \leq K$ for all t with probability one. Below is the definition of a continuous process.

Definition 6.6. (Continuous Process). A process A_t with respect to $\{\mathcal{F}_n\}$ is *continuous* if each A_t is $\mathcal{F}_t$ measurable for any $t \in \mathbb{R}$. A_t is *bounded* if there exists $K < \infty$ such that $|A_t| \leq K$ for all t with probability one.

To better conceptualize the bounded, continuous process A_t, we use the following lemma to express A_t in terms of a converging sequence of simple processes $A_t^{(n)}$.

Lemma 6.7. *(Existence of Sequence of Simple Processes for Bounded, Continuous Process). Let* A_t *be a process with continuous paths with respect to* $\{\mathcal{F}_t\}$*, and suppose there exists* $K < \infty$ *such that* $|A_t| \leq K$ *for all* t *with probability one. Then there exists a sequence of simple processes* $A_t^{(n)}$ *such that for all* t*, we have*

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

Also, for all n, t*, we have that* $\left|A_t^{(n)}\right| \leq K$*.*

We will not include the proof of Lemma 6.7 in this paper. Such a proof requires some measure theory and can be found as Lemma 3.2.2 in [1]. However, the idea of the lemma makes sense: there exists some bounded sequence $A_t^{(n)}$ such that as $n \to \infty$, the difference in expectation between A_s and $A_s^{(n)}$ for $0 \leq s \leq t$ approaches 0.

With Lemma 6.7, we are now equipped to define the stochastic integral for a continuous process A_t.

Definition 6.8. (Stochastic Integral for Bounded, Continuous Processes). Let A_t be a bounded, continuous process with respect to $\{\mathcal{F}_n\}$, and let B_t be a standard Brownian motion with respect to $\{\mathcal{F}_n\}$. Let $A_t^{(n)}$ be a sequence of simple processes satisfying Lemma 6.7. Then for each t, there exists a square-integrable random variable Z_t such that

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

In particular, we define the *stochastic integral* of A_t as

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

It turns out that this stochastic integral satisfies all the properties expressed in Proposition 6.5. We restate the proposition adapted to this iteration of the stochastic integral below.

Proposition 6.9. *(Properties of the Stochastic Integral for Bounded, Continuous Processes). Let B_t be a standard Brownian motion with respect to $\{\mathcal{F}_n\}$, and let A_t and C_t be bounded, continuous processes with respect to $\{\mathcal{F}_n\}$. Then the following properties hold.*

(1) Linearity: For $a, c \in \mathbb{R}$, we have that

$$\int_0^t (aA_s + cC_s)dB_s = a\int_0^t A_s dB_s + c\int_0^t C_s dB_s.$$

Also, for $0 < r < t$, we have that

$$\int_0^t A_s dB_s = \int_0^r A_s dB_s + \int_r^t A_s dB_s.$$

(2) Martingale property: The following process is a martingale with respect to $\{\mathcal{F}_t\}$:

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

(3) Variance rule: $\mathbb{E}[Z_t^2] < \infty$ (square integrability) and

$$\mathrm{Var}[Z_t] = \mathbb{E}[Z_t^2] = \int_0^t \mathbb{E}[A_s^2] ds.$$

(4) Continuity: The function $t \to Z_t$ is almost surely a continuous function.

As was the case earlier, we will not include a proof of Proposition 6.9 in this paper (such a proof is quite involved). The reader can find a proof as Proposition 3.2.3 in [1].

In Definition 6.6, recall how we differentiated between *bounded* and *unbounded* continuous processes. We now define the stochastic integral for continuous (but not necessarily bounded) processes.

Definition 6.10. (Stochastic Integral for Continuous Processes). Let A_t be a continuous process with respect to $\{\mathcal{F}_n\}$, and let B_t be a standard Brownian motion with respect to $\{\mathcal{F}_n\}$. For each $n < \infty$, define

$$T_n = \min\{t \mid |A_t| = n\},$$

and let $A_s^{(n)} = A_{s \wedge T_n}$. Then $A_s^{(n)}$ is a bounded, continuous process, and

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

is well-defined. In particular, we define

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

To prove the limit in (*) does indeed exist, define

$$K_t = \max_{0 \le s \le t} |A_s|.$$

Then following Definition 6.10, for $n \ge K_t$, we have that $A_s^{(n)} = A_s$ when $0 \le s \le t$. In particular, $Z_t^{(n)} = Z_t$.

We now include a proposition describing the properties of the stochastic integral for a continuous process. The properties are largely the same as the stochastic integral for a bounded, continuous process minus the martingale property.

Proposition 6.11. *(Properties of Stochastic Integral for Continuous Processes). Let B_t be a standard Brownian motion with respect to $\{\mathcal{F}_n\}$, and let A_t and C_t be continuous processes with respect to $\{\mathcal{F}_n\}$. Then the following properties hold.*

(1) Linearity: For $a, c \in \mathbb{R}$, we have that

$$\int_0^t (aA_s + cC_s) dB_s = a \int_0^t A_s dB_s + c \int_0^t C_s dB_s.$$

Also, for $0 < r < t$, we have that

$$\int_0^t A_s dB_s = \int_0^r A_s dB_s + \int_r^t A_s dB_s.$$

(2) Variance rule: $\mathbb{E}[Z_t^2] < \infty$ (square integrability) and

$$\mathrm{Var}[Z_t] = \mathbb{E}[Z_t^2] = \int_0^t \mathbb{E}[A_s^2] ds.$$

(3) Continuity: The function $t \to Z_t$ is almost surely a continuous function.

As an aside, our process A_t does not have to be strictly continuous in order to be integrable. Rather, A_t must be at least *piecewise continuous* to be integrable, that is, A_t has discontinuities only at a finite number of times $t_0 < t_1 < ... < t_n$. The value of our integral does not depend on the value of A_t at these discontinuities. Stopping times are an example of such discontinuities. When we "stop" a stochastic integral, we are essentially changing our bet to 0. This aside inspires the proof for Proposition 6.11, which can be found as Proposition 3.2.5 in [1].

For the final part of this section, we will work up to a definition of quadratic variation, which we will use in our proof of Itô's Formula. First, for a standard Brownian motion B_t, consider the following integral:

$$Z_t = \int_0^t B_s dB_s.$$

We have that B_t is continuous by definition, though it is not bounded. Observe that

$$\int_0^t \mathbb{E}[B_s]^2 ds = \int_0^t s ds = \frac{t^2}{2} < \infty.$$

Thus, Z_t is a square integrable martingale. If we were to apply the fundamental theorem of (non-stochastic) calculus to Z_t, we would find that

$$Z_t = \frac{1}{2}(B_t - B_0)^2 = \frac{B_t^2}{2}.$$

Since Z_t is a martingale, we have that $\mathbb{E}[Z_t] = Z_0 = 0$. But $\mathbb{E}[B_t^2/2] = t/2 \neq 0$, so the above equality cannot possibly hold. We now see the place for Itô's Formula, which will tell us that

$$Z_t = \frac{1}{2}(B_t^2 - t).$$

Note that in the above equality, when $t > 0$, we have that $(B_t^2 - t)$ is *not* a normally distributed random variable. Although stochastic integrals are the limits of normal increments, since A_t can depend on the past (that is, our current "bet" may depend on past information), we may very well end up with non-normally distributed random variables. Now, we define quadratic variation.

Definition 6.12. (Quadratic Variation). Let A_t be a continuous process with respect to $\{\mathcal{F}_n\}$, and let B_t be a standard Brownian motion. If

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

then we define the *quadratic variation* by

$$\langle Z \rangle_t = \lim_{n \to \infty} \sum_{i \leq nt} \left(Z_{i/n} - Z_{(i-1)/n} \right)^2,$$

where we sum over all i with $i/n \leq t$.

To better conceptualize the quadratic variation, we provide the following theorem.

Theorem 6.13. *Let A_t be a continuous or piecewise continuous process with respect to $\{\mathcal{F}_n\}$, and let B_t be a standard Brownian motion. If*

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

then

$$\langle Z \rangle_t = \int_0^t A_s^2 ds.$$

Through this theorem (which can be found as Theorem 3.2.6 in [1]), we can view the quadratic variation $\langle Z \rangle_t$ as the "total amount of randomness" or the "total amount of betting" up to time t. In particular, for Brownian motion, we have a constant betting rate, so the quadratic variation grows linearly. More generally, the quadratic variation is a random variable, as our current "bet" may depend on past results. Then the quadratic variation $\langle Z \rangle_t$ at time t is a random variable with mean

$$\mathbb{E}[\langle Z \rangle_t] = \mathbb{E}\left[\int_0^t A_s^2 ds\right] = \int_0^t \mathbb{E}[A_s^2] ds.$$

Importantly, for a standard Brownian motion B_t, we have that $\langle B \rangle_t = t$. More generally, for a Brownian motion with drift m and variance σ^2, we have that $\langle B \rangle_t = \sigma^2 t$. For proofs of these facts, see Theorem 2.8.1 in [1]. These results will be vital in our proof of Itô's Formula.

7. Itô's Formula

We are now ready to state Itô's Formula—the fundamental theorem of stochastic calculus. First, recall the fundamental theorem of (non-stochastic) calculus.

Theorem 7.1. *(Fundamental Theorem of Calculus, FTC). Let f be an integrable function on $[a,b]$. Suppose there exists a function F that is continuous on $[a,b]$ and differentiable on (a,b) such $F' = f$ on (a,b). Then*

$$\int_a^b f(x)dx = F(b) - F(a) \quad \text{or} \quad F(a) + \int_a^b f(x)dx = F(b).$$

Intuitively, the fundamental theorem of calculus says that given the derivative f of a function F and a starting value $F(a)$, we can compute the value $F(b)$ of the function at any point b. We will first derive the fundamental theorem of calculus using a Taylor approximation, which will inspire our derivation of Itô's Formula.

Theorem 7.2. *(Taylor Approximation Theorem). Suppose f is a k-times differentiable function at $a \in \mathbb{R}$. Then there exists a function $o\colon \mathbb{R} \to \mathbb{R}$ such that*

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + ... + \frac{f^k(a)}{k!}(x-a)^k + o(x)(x-a)^k,$$

where

$$\lim_{x \to a} o(x) = 0.$$

A proof of the Taylor approximation theorem can be found in [8]. Using a Taylor approximation, we will now prove the fundamental theorem of calculus.

Proof of FTC. Applying Theorem 7.2 to a once-differentiable function f, we may write

$$f(b) = f(a) + f'(a)(b-a) + o(b)(b-a),$$

or

$$f(b) - f(a) = f'(a)(b-a) + o(b)(b-a), \tag{*}$$

where $o(b) \to 0$ as $b \to a$. If we partition $[0,1]$ into n equally-sized intervals, we can write f as a telescoping sum:

$$f(1) = f(0) + \sum_{i=1}^{n} \left[f\left(\frac{i}{n}\right) - f\left(\frac{i-1}{n}\right) \right].$$

Analyzing the terms in the summation, we invoke (*) to write

$$f\left(\frac{i}{n}\right) - f\left(\frac{i-1}{n}\right) = f'\left(\frac{i-1}{n}\right)\frac{1}{n} + o\left(\frac{i}{n}\right)\frac{i}{n}.$$

Then

$$\begin{aligned} f(1) &= f(0) + \lim_{n\to\infty} \sum_{i=1}^{n} \left[f'\left(\frac{i-1}{n}\right)\frac{1}{n} + o\left(\frac{i}{n}\right)\frac{1}{n} \right] \\ &= f(0) + \lim_{n\to\infty} \sum_{i=1}^{n} f'\left(\frac{i-1}{n}\right)\frac{i}{n} + \lim_{n\to\infty} \sum_{i=1}^{n} o\left(\frac{i}{n}\right)\frac{i}{n}. \end{aligned}$$

The first limit is the Riemann sum approximation of $\int_0^1 f$, and the second limit equals 0 because $\frac{i}{n} \to \frac{i-1}{n}$ as $n \to \infty$, meaning $o\left(\frac{i}{n}\right)$ approaches 0. Thus,

$$f(1) = f(0) + \int_0^1 f'(t)dt.$$

□

Observe that the fundamental theorem of calculus uses only the first derivative of the function f. For Itô's Formula, we will need both the first *and* second derivatives of f.

Theorem 7.3. *(Itô's Formula). Let f be a twice-differentiable function, and let B_t be a standard Brownian motion. Then for every t, we have that*

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

Equivalently, we may write

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

Intuitively, Itô's Formula says that the process $f(B_t)$ at time t evolves like a Brownian motion with drift $\frac{f''(B_t)}{2}$ and variance $f'(B_t)^2$. Note that since $f'(B_t)$ is continuous, the stochastic integral $\int_0^t f'(B_t)dB_s$ is well defined.

Proof. Using the Taylor approximation for the twice-differentiable function f, we may write

$$(*) \qquad f(b) - f(a) = f'(a)(b-a) + \frac{f''(a)}{2}(b-a)^2 + o(b)(b-a)^2,$$

where $o(b) \to 0$ as $b \to a$. As was the case in our proof of the fundamental theorem of calculus, we partition $[0,1]$ into n equally-sized intervals, writing f as a telescoping sum:

$$f(B_1) = f(B_0) + \sum_{i=1}^{n}\left[f\left(B_{i/n}\right) - f\left(B_{(i-1)/n}\right)\right].$$

Denote $B_{i/n} - B_{(i-1)/n}$ with $\Delta_{i,n}$. Looking specifically at the terms in the summation, as in (*) we may write

$$f\left(B_{i/n}\right) - f\left(B_{(i-1)/n}\right) = f'\left(B_{(i-1)/n}\right)\Delta_{i,n} + \frac{f''\left(B_{(i-1)/n}\right)}{2}\Delta_{i,n}^2 + o\left(B_{i/n}\right)\Delta_{i,n}^2,$$

where $o(B_{i/n}) \to 0$ as $i/n \to (i-1)/n$. Then letting $n \to \infty$, we have that $f\left(B_{i/n}\right) - f\left(B_{(i-1)/n}\right)$ equals the sum of the following three limits:

$$(1) \qquad \lim_{n\to\infty}\sum_{i=1}^{n} f'\left(B_{(i-1)/n}\right)\left(B_{i/n} - B_{(i-1)/n}\right),$$

$$(2) \qquad \lim_{n\to\infty}\frac{1}{2}\sum_{i=1}^{n} f''\left(B_{(i-1)/n}\right)\left(B_{i/n} - B_{(i-1)/n}\right)^2,$$

$$(3) \qquad \lim_{n\to\infty}\sum_{i=1}^{n} o\left(B_{i/n}\right)\left(B_{i/n} - B_{(i-1)/n}\right)^2.$$

By the stationary increments property of Brownian motion, we have that the increment $(B_{i/n} - B_{(i-1)/n})^2$ is identically distributed as

$$(B_{1/n} - B_0)^2 = (B_{1/n} - 0)^2 = (B_{1/n})^2.$$

Since B_t is $N(0,1)$, it must be that $(B_{i/n} - B_{(i-1)/n})^2 \approx \frac{1}{n}$. Examining the limit in (3), as $n \to \infty$, we have that $i/n \to (i-1)/n$, meaning $o(B_{i/n}) \to 0$, that is, the limit equals zero.

Looking at the limit in (1), we see that it is a simple process approximation to a stochastic integral, that is to say, for $n < \infty$, we can consider the process as a "step" function as in Definitions 6.1 and 6.2. Then letting $n \to \infty$, the limit becomes

$$\int_0^1 f'(B_t) dB_t.$$

Now, examining the limit in (2), we first consider the case where $f'' = b$ for some constant b. Then the limit becomes

$$\lim_{n\to\infty} \frac{b}{2} \sum_{i=1}^{n} (B_{i/n} - B_{(i-1)/n})^2 = \frac{b}{2} \langle B \rangle_1 = \frac{b}{2}.$$

For the more general case, let $h(t) = f''(B_t)$, where h is a continuous function. Then for every $\epsilon > 0$, there exists some step function $h_\epsilon(t)$ where $|h(t) - h_\epsilon(t)| < \epsilon$ for every t. For any particular ϵ, if we consider each interval where h_ϵ is constant, we see that

$$\lim_{n\to\infty} \sum_{i=1}^{n} h_\epsilon(t)(B_{i/n} - B_{(i-1)/n})^2 = \int_0^1 h_\epsilon(t) dt.$$

Also, we have that

$$\left| \sum_{i=1}^{n} [h(t) - h_\epsilon(t)](B_{i/n} - B_{(i-1)/n})^2 \right| \leq \epsilon \sum_{i=1}^{n} (B_{i/n} - B_{(i-1)/n})^2 \to \epsilon.$$

Using the above equality and inequality, we may then rewrite the limit in (2) as

$$\lim_{\epsilon\to 0} \frac{1}{2} \int_0^1 h_\epsilon(t) dt = \frac{1}{2} \int_0^1 h(t) dt = \frac{1}{2} f''(B_t) dt.$$

Thus,

$$f(B_1) = f(B_0) + \int_0^t f'(B_s) dB_s + \frac{1}{2} \int_0^t f''(B_s) ds.$$

□

There are more general statements of Itô's Formula for functions that depend on both time *and* position (for example, geometric Brownian motion). We will not include these versions in this paper. However, the same Taylor approximation strategy can be used to derive these variations (albeit with more terms in the expansion). In the final part of this section, we will walk through a couple of examples in which we apply Itô's Formula.

Example 7.4. Suppose B_t is a standard Brownian motion. We will determine an expression for

$$\int_0^t \sin(B_s) dB_s$$

that does not involve Itô integrals. The "trick" here is to choose a function $f(x)$ such that $f'(x) = \sin(x)$. It follows that $f(x) = -\cos(x)$. Applying Itô's Formula, we may write

$$-\cos(B_t) = -\cos(B_0) + \int_0^t \sin(B_s)dB_s + \frac{1}{2}\int_0^t \cos(B_s)ds.$$

Then

$$\begin{aligned}\int_0^t \sin(B_s)dB_s &= -\cos(B_t) - \cos(0) - \frac{1}{2}\int_0^t \cos(B_s)ds \\ &= 1 - \cos(B_t) - \frac{1}{2}\int_0^t \cos(B_s)ds.\end{aligned}$$

Example 7.5. We wish to find

$$\int_0^t B_s dB_s.$$

Let $f(x) = x^2$. It follows that $f'(x) = 2x$ and $f''(x) = 2$. Applying Itô's Formula, we have

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

Reordering terms, we get

$$\int_0^t B_s dB_s = \frac{1}{2}(B_t^2 - t).$$

Acknowledgments

I am deeply grateful for the help of my mentor Stephen Yearwood in writing this paper. His guidance and patience were essential in my learning and exposition of this material. I would also like to thank Gregory Lawler for his lecture series and stochastic calculus book. I certainly could not have learned this material without his wonderful resources. Last (but certainly not least!), I would like to thank Peter May for organizing UChicago's Math REU. I can hardly imagine how effort it takes to gather so many talented lecturers, mentors, and students and plan this incredible program; I am most grateful for the opportunity to have been a part of it.

References

[1] Gregory Lawler. Stochastic Calculus: An Introduction with Applications. 2014. http://www.math.uchicago.edu/~lawler/finbook.pdf
[2] Peter Mörters and Yuval Peres. Brownian Motion. https://people.bath.ac.uk/maspm/book.pdf
[3] Lawrence C. Evans. An Introduction to Stochastic Differential Equations. American Mathematical Society. 2014.
[4] Andrej Nikolaevic Kolmogorov. Foundations of the Theory of Probability. Martino Publishing. 2013.
[5] C.T. Bauer College of Business at University of Houston. Probability Theory: Introduction. https://www.bauer.uh.edu/rsusmel/phd/sR-0.pdf

[6] Ross Kravitz. How are Itô Stochastic Integrals Used in Finance. *Quora*. Quora, 2018. https://www.quora.com/How-are-Itô-stochastic-integrals-used-in-finance
[7] Yuval Filmus. Two Proofs of the Central Limit Theorem. 2010. https://www.cs.toronto.edu/~yuvalf/CLT.pdf
[8] Chinese University of Hong Kong. Proof of Taylor's Theorem. https://www.math.cuhk.edu.hk/course_builder/1516/math1010c/Taylor.pdf
[9] Walter Rudin. Principles of Mathematical Analysis. McGraw-Hill, Inc. 1953.
[10] Wikipedia. Total Variation. https://en.wikipedia.org/wiki/Total_variation
[11] Wikipedia. Cumulative Distribution Function. https://en.wikipedia.org/wiki/Cumulative_distribution_function
[12] Yale. Random Variables. http://www.stat.yale.edu/Courses/1997-98/101/ranvar.htm

Appendix

As promised, here are two full implementations of our Brownian motion simulation in imperative C. This first program simulates one-dimensional Brownian motion.

```c
// One-Dimensional Brownian Motion Simulation
#include <stdlib.h>
#include <stdio.h>
#include <time.h>

void brownian_motion(int pos, int t) {
   int r;
   while(t > 0) {
      r = rand();
      if (r % 2) {
         pos++;
      } else {
         pos--;
      }
      printf("%u\n", pos);
      t--;
   }
}

int main(int argc, char* argv[]) {
   srand(time(NULL));
   int t = atoi(argv[2]);
   int pos = 0;
   brownian_motion(pos, t);
   return 0;
}
```

This second program simulates two-dimensional Brownian motion.

```c
// Two-Dimensional Brownian Motion Simulation
#include <stdlib.h>
#include <stdio.h>
#include <time.h>

struct coord {
      int x;
```

```
    int y;
};

typedef struct coord coord;

void brownian_motion(coord pos, int t) {
  int r, c;
  while(t > 0) {
    r = rand();
    if (r % 2) {
      pos.x++;
    } else {
      pos.x--;
    }
      c = rand();
      if (c \% 2) {
      pos.y++;
    } else {
      pos.y--;
    }
    printf("%u, %u\n", pos.x, pos.y);
    t--;
  }
}

int main(int argc, char* argv[]) {
  srand(time(NULL));
  int t = atoi(argv[2]);
  coord pos = {0, 0};
  brownian_motion(pos, t);
  return 0;
}
```

For both programs, save the file with a .c extension and navigate to the appropriate directory. Then on Mac or Linux, compile the program by running the following in the terminal

`clang [filename].c`

Run the executable and send its output to a comma separated values file by running

`./a.out [filename].c -t [numtrials] > output.csv`

where `[numtrials]` is how many times you would like the program to repeat the process of randomly adding or subtracting 1. Open `output.csv` in Excel, insert the column(s) of numbers into a scatter plot, and use any chart options to format the plot as desired.