AN INTRODUCTION TO STOCHASTIC CALCULUS

VALERIE HAN

ABSTRACT. This paper will introduce the Ito integral, one type of stochastic integral. We will discuss relevant properties of Brownian motion, then construct the Ito integral with analogous properties. We end with the stochastic calculus analogue to the Fundamental Theorem of Calculus, that is, Ito's Formula.

CONTENTS

1. INTRODUCTION

Stochastic calculus is used in a number of fields, such as finance, biology, and physics. Stochastic processes model systems evolving randomly with time. Unlike deterministic processes, such as differential equations, which are completely determined by some initial value and parameters, we cannot be sure of a stochastic process's value at future times even with full knowledge of the state of the system and its past. Thus, to study a stochastic process, we study its distribution and the behavior of a sample path. Moreover, traditional methods of calculus fail in the face of real-world data, which is noisy. The stochastic integral, which is the integral of a stochastic process with respect to another stochastic process, thus requires a whole different set of techniques from those used in calculus.

The most important and most commonly used stochastic process is the one that models random continuous motion: Brownian motion. We discuss in Section 3 how Brownian motion naturally arises from random walks, the discrete-time analogue of Brownian motion, to provide one way of understanding Brownian motion. In Section 4, we actually define Brownian motion and discuss its properties. These properties make it very interesting to study from a mathematical standpoint, are also useful for computations, and will allow us to prove Ito's formula in Section 7.

However, some of Brownian motion's properties, such as its non-differentiability, make it difficult to work with. We discuss its non-differentiability in Section 5 to provide some motivation for the construction of the Ito integral in Section 6. Just as with Riemann integrals, computing with the definitions themselves is often tedious. Ito's formula discussed in Section 7 is often referred to as the stochastic calculus analogue to the Fundamental Theorem of Calculus or to the chain rule. It makes computations of the integral much easier as well as being a useful tool in understanding how the Ito integral is different from the Riemann integral.

This paper will assume familiarity with the basics of measure theory and probability theory. We will be closely following Lawler's book [1] on the subject.

2. Preliminaries

We begin by reviewing a few important terms and discussing how they can be understood as mimicking information.

Definition 2.1. A family S of subsets of a nonempty set X is a *σ-algebra* on X if

(i) $\varnothing \in S$.
(ii) If $A \in S$, then $X \setminus A \in S$.
(iii) If $A_n \in S$ for all $n \in \mathbb{N}$, then $\cup_n A_n \in S$.

Remark 2.2. A σ-algebra represents known information. By being closed under the complement and countable unions, the set of information becomes fully filled out; given some information, we have the ability to make reasonable inferences, and the σ-algebra reflects that.

Definition 2.3. Let A, B be σ-algebras on X. The σ-algebra A is a *sub-σ-algebra* of B if $A \subseteq B$.

Definition 2.4. A *filtration* is a family $\{F_t\}$ of sub-σ-algebras of F with the property that $F_s \subseteq F_t$, for $0 \leq s < t$.

Remark 2.5. Note that a filtration can be understood as the information over time. For each t, F_t is a σ-algebra, which mimics known information as we discussed in Remark 2.2. Moreover, just as information (theoretically) cannot be lost, $F_s \subseteq F_t$ for $s < t$.

Definition 2.6. Let F_t be a sub-σ-algebra of σ-algebra F, and let X be a random variable with $\mathbb{E}[X] < \infty$. The *conditional expectation* $\mathrm{E}(X \mid F_t)$ of X with respect to F_t is the random variable Y such that

(i) $\mathrm{E}(X \mid F_t)$ is F_t-measurable.
(ii) For every F_t-measurable event A, $\mathbb{E}[\mathrm{E}(X \mid F_t)\mathbb{1}_A] = \mathbb{E}[X\mathbb{1}_A]$, where $\mathbb{1}$ denotes the indicator function.

Notation 2.7. We distinguish conditional expectation from expectation by denoting conditional expectation $\mathrm{E}(\cdot)$ and expectation $\mathbb{E}[\cdot]$.

Note that the conditional expectation always exists and is unique in that any two random variables that fulfill the definition are equal almost surely. See proof of Proposition 10.3 in [3] for the proof.

Since the definition of conditional expectation is difficult to work with, the following properties are helpful for actually computing the conditional expectation. See proof of Proposition 10.5 in [3] for the proof.

Proposition 2.8. *Let X, Y be random variables such that $\mathbb{E}[X], \mathbb{E}[Y] < \infty$. Let F_s, F_t be sub-σ-algebras of σ-algebra F such that $s < t$. Then,*

(i) (Linearity) If $\alpha, \beta \in \mathbb{R}$, then $E(\alpha X + \beta Y \mid F_t) = \alpha E(X \mid F_t) + \beta E(Y \mid F_t)$.
(ii) (Projection or Tower Property) $E(E(X \mid F_t) \mid F_s) = E(X \mid F_s)$.
(iii) (Pulling out what's known) If Y is F_t-measurable and $\mathbb{E}[XY] < \infty$, then $E(XY \mid F_t) = YE(X \mid F_t)$.
(iv) (Irrelevance of independent information) If X is independent of F_t, then $E[X \mid F_t] = \mathbb{E}[X]$.

Remark 2.9. Each of these four properties is true almost surely. Here and in the rest of the paper, statements about conditional expectation contain an implicit "up to an event of measure zero."

Definition 2.10. Let $T \subseteq \mathbb{R}^{>0}$ be a set of times. A *stochastic process* $\{X_t\}_{t \in T}$ is a collection of random variables defined on the same probability space indexed by time.

Remark 2.11. T is most often $\mathbb{N}$ or $\mathbb{R}^{>0}$, and these are what we will use for random walk and Brownian motion, respectively.

Notation 2.12. Note that a stochastic process X_t has two "inputs" as a function, one of which is usually omitted from notation. (The other is t.) Since X_t is a random variable, any depiction of X_t is of a *realization* or *sample path* of X_t, where we in fact mean $X_t(\omega)$ for some ω in the sample space Ω.

Definition 2.13. A stochastic process $\{X_t\}$ is *adapted* to filtration $\{F_t\}$ if X_t is F_t-measurable for all t.

Remark 2.14. Recall that the random variable X being F-measurable means that the preimage under X of any Borel set is in F. We will see F-measurability appear throughout this paper because we want to measure the probability of a random variable X belonging to the nice sets in $\mathbb{R}$ − the Borel sets. Since probabilities only make sense (ie. are countably additive) for sets in the σ-algebra F, we need $\{X \in B\} \in F$ for all Borel sets B.

3. Random Walk

Since discrete time processes are often easier to understand and describe, we will start by considering the discrete time analogue to Brownian motion: the random walk.

Definition 3.1. Let $X_1, \dots, X_n$ be independent random variables such that $P(X_i = 1) = P(X_i = -1) = \frac{1}{2}$. The random variable $S_n := X_1 + \cdots + X_n$ is a *random walk*.

As the name suggests, a random walk can be understood by considering the position of a walker who does the following every time increment: He walks forwards one step if a flipped coin turns up heads and backwards if tails.

We can make a few observations immediately. By linearity of expectation, $\mathbb{E}[S_n] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n] = 0$. Since the random variables X_i are independent, $\mathrm{Var}[S_n] = \mathrm{Var}[X_1] + \cdots + \mathrm{Var}[X_n] = \mathbb{E}[X_1^2] + \cdots + \mathbb{E}[X_n^2] = n$.

This matches our intuition that although it should be equally likely for the walker to end up in a positive or negative position, the higher n is, that is, the higher the number of steps taken, the higher the likelihood he will be far away from the origin, spreading out the distribution.

Let Δt denote the time increment and $\pm\Delta x$ denote the position change each time increment. (The position change above is 1.) Suppose $\Delta t = 1/N$, and $B_{n\Delta t} = \Delta x(X_1 + \cdots + X_n)$. What happens when N grows very large and Δx becomes small?

The graph of $B_{n\Delta t}$ against t begins to look like the graph of a continuous function. Moreover, by the Central Limit Theorem, the distribution of $\frac{S_N}{\sqrt{N}}$ approaches the standard normal distribution. Both of these qualities are qualities of Brownian motion.

In fact, it can be shown that Brownian motion is the limit of random walks, but the proof is complicated. The interested reader is encouraged to read the proof of Theorem 5.22 in [4]. This fact legitimizes the intuition that Brownian motion and random walk have similar properties. Moreover, in order to simulate Brownian motion, one must simulate random walks as we have done here with time and space increments being very small.

4. Brownian Motion

Brownian motion is one of the most commonly used stochastic processes. It is used to model anything from the random movement of gas molecules to the often (seemingly) random fluctuations of stock prices.

In order to model random continuous motion, we define Brownian motion as follows. For simplicity, we only discuss standard Brownian motion.

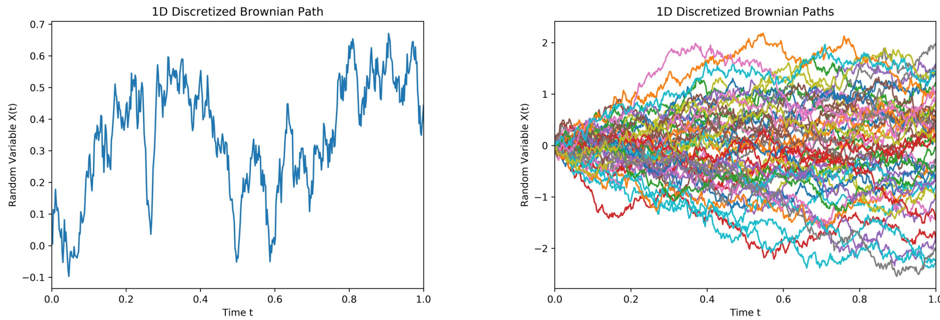
Definition 4.1. A stochastic process $\{B_t\}$ is a *(standard) Brownian motion* with respect to filtration $\{F_t\}$ if it has the following three properties:

(i) For $s < t$, the distribution of $B_t - B_s$ is normal with mean 0 and variance $t - s$. We denote this by $B_t - B_s \sim N(0, t - s)$.
(ii) If $s < t$, the random variable $B_t - B_s$ is independent of F_s.
(iii) With probability one, the function $t \mapsto B_t$ is a continuous function of t.

Brownian motion is often depicted as, and understood as, a sample path, as in Figure 1(A). However, Figure 1(B) depicts the random nature of the random variable B_t, exposing the hidden ω in $B_t(\omega)$. The density of the sample paths gives a sense of Brownian motion's normal distribution, as well as showing that, like the random walk, the paths of Brownian motion spread out the further they are from $t = 0$.

It may not be clear from the definition why (or if) Brownian motion exists. In fact, Brownian motion does exist. We can show this by first defining the process $\{B_t\}$ for the dyadic rationals t. After proving that with probability one the function $t \mapsto B_t$ is continuous on the dyadics, we can extend B_t to all real t by continuity. The interested reader can see Section 2.5 of [1] for the full proof and construction.

We now introduce a property of Brownian motion that is useful in computing the conditional expectation: the martingale property. Note that the first two properties in the definition of Brownian motion mean that we can "start over" Brownian motion at any time t. The value B_t for $s < t$ does not depend on knowing the previous values B_r where $r < s$, and the distribution of B_t, given the value of B_s, is identical except shifted vertically by the value B_s. The martingale property partially captures this characteristic. Given the stream of information up to time s,



(A) A realization of Brownian motion.

(B) Multiple (differently colored) realizations of Brownian motion.

FIGURE 1. Brownian motion simulated with random walk as described in Section 3, adopted from [7].

the best guess of any future value of a martingale is precisely the random variable's value at time s.

Definition 4.2. Let $\{M_t\}$ be a process adapted to filtration $\{F_t\}$ such that $\mathbb{E}[M_t] < \infty$. The process $\{M_t\}$ is a *martingale* with respect to $\{F_t\}$ if for each $s < t$, $\mathrm{E}(M_t \mid F_s) = M_s$.

Remark 4.3. If no filtration is specified, we assume that the filtration is the one generated by $\{M_s : s \le t\}$.

Definition 4.4. A martingale M_t is a *continuous martingale* if with probability one the function $t \mapsto M_t$ is continuous.

Proposition 4.5. *Brownian motion is a continuous martingale.*

Proof. The function $t \to B_t$ is continuous with probability one by definition, so all we need to show is $\mathrm{E}(B_t \mid F_s) = B_s$, which easily follows from the properties of conditional expectation.

$$\begin{aligned}\mathrm{E}(B_t \mid F_s) &= \mathrm{E}(B_s \mid F_s) + \mathrm{E}(B_t - B_s \mid F_s) \text{ by Proposition 2.8(i)}\\ &= B_s + \mathbb{E}[B_t - B_s] \text{ by Proposition 2.8(iii),(iv)}\\ &= B_s \text{ since } B_t - B_s \sim N(0, t-s).\end{aligned}$$

□

Now we turn to another property of Brownian motion that is useful in computations and will come into use when proving Ito's formula in Section 7. While we shall see in Section 5 that Brownian motion has unbounded variation, its quadratic variation is bounded.

Definition 4.6. Let $\{X_t\}$ be a stochastic process. The *quadratic variation* of X_t is the stochastic process

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

Proposition 4.7. *The quadratic variation of B_t is the constant process $\langle B\rangle_t = t$.*

Proof. For ease of notation, we prove the Proposition for $t = 1$. Thus, we want to show that the variance of $\langle B\rangle_t$ is 0 and the expectation is 1.

Consider

$$Q_n = \sum_{j=1}^{n} \left[B_{j/n} - B_{(j-1)/n}\right]^2.$$

In order to more easily compute the variance and expected value, we express Q_n in terms of a distribution we are familiar with. Let

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

Then,

$$Q_n = \frac{1}{n}\sum_{j=1}^{n} Y_j,$$

and the random variables $Y_1, \ldots, Y_n$ are independent, each with distribution Z^2, where Z is the standard normal distribution. Thus, $\mathbb{E}[Y_j] = \mathbb{E}[Z^2] = 1$, $\mathbb{E}[Y_j^2] = \mathbb{E}[Z^4] = 3$, and $\mathrm{Var}[Y_j] = \mathbb{E}[Y_j^2] - \mathbb{E}[Y_j]^2 = 3 - 1 = 2$.

We can now easily calculate the expectation and variance of Q_n.

$$\mathbb{E}[Q_n] = \frac{1}{n}\sum_{j=1}^{n} \mathbb{E}[Y_j] = 1.$$

$$\mathrm{Var}[Q_n] = \frac{1}{n^2}\sum_{j=1}^{n} \mathrm{Var}[Y_j] = \frac{2}{n}.$$

Thus,

$$\mathbb{E}[\langle B\rangle_t] = \lim_{n\to\infty} \mathbb{E}[Q_n] = 1, \text{ and}$$

$$\mathrm{Var}[\langle B\rangle_t] = \lim_{n\to\infty} \mathrm{Var}[Q_n] = 0,$$

as desired. □

Remark 4.8. Informally, this Proposition tells us that the squared difference of the values of Brownian motion on a small interval is just the length of the interval. This gives us the intuition $(dB_t)^2 = dt$, which is often used in stochastic integral calculations.

5. Motivating the Stochastic Integral

In Section 4, we discussed some of the properties that make Brownian motion easy to work with. In contrast, the property we discuss in this section is the non-differentiability of Brownian motion.

Theorem 5.1. *With probability one, the function $t \to B_t$ is nowhere differentiable.*

Remark 5.2. Note that this is stronger than the statement "For every t, the derivative does not exist at t with probability one."

Rather than including a formal proof of this fact, which is located in many different places (e.g. see proof of Theorem 2.6.1 in [1]), we give a more intuitive explanation for why this fact should not be surprising.

Suppose B_t were differentiable at time s. Then, the limit $\lim_{t\to s} \frac{B_t - B_s}{t-s}$ would exist, so both the right and left hand limits would exist and be equal. However, this would mean that knowing B_r for $0 \leq r \leq s$ would tell us something about $B_{s+\Delta t} - B_s$ for Δt sufficiently small, which seemingly contradicts with the fact that $B_{s+\Delta t} - B_s$ is independent of the information given by B_r for $r \leq s$.

The non-differentiability of Brownian motion also implies that Brownian motion has unbounded variation (see Corollary 25 in [5]), which means we cannot simply use the Riemann-Stieltjes integral to define an integral with respect to B_t.

6. Construction of Ito Integral

As explained in Section 5, due to Brownian motion's non-differentiability and unbounded variation, we cannot simply use Riemann-Stieltjes integration for an integral of the form $\int_0^t A_s dB_s$. However, we can use a similar strategy as the one used to define the Lebesgue integral.

Recall that, in the construction of the Lebesgue integral, we first defined the integral for simple functions and then approximated other, more general, functions with simple functions. Similarly, for the Ito integral, we first define the integral for simple processes, and then approximate other functions with simple processes.

Definition 6.1. A process A_t is a *simple process* if there exist a finite number of times $0 = t_0 < t_1 < \cdots < t_n < \infty$ and associated F_{t_j}-measurable random variables Y_j such that $A_t = Y_j$ on the associated interval $t_j < t < t_{j+1}$. We denote $t_{n+1} = \infty$.

Simple processes, like the simple functions used in the construction of the Lebesgue integral, are an extension of step functions. Any realization of a simple process looks like a step function with a finite number of intervals, with the last interval's right "endpoint" being ∞, as in Figure 2.

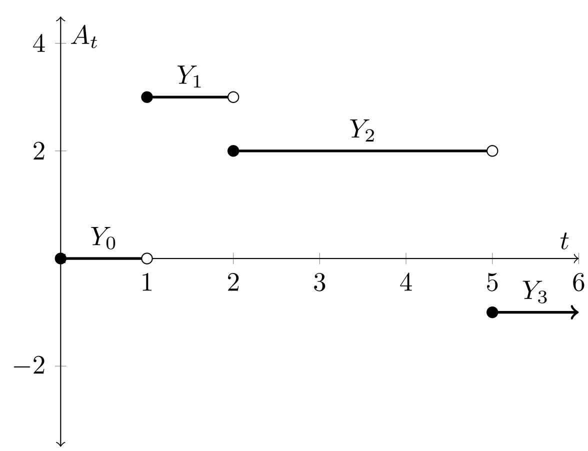


Figure 2. A realization of a simple process A_t with $t_0 = 0$, $t_1 = 1$, $t_2 = 2$, and $t_3 = 5$.

We can now define the integral of a simple process, much like how we would define the Riemann-Stieltjes integral. However, unlike with Riemann-Stieltjes integration, where the endpoint chosen does not affect the properties of the integral, the endpoint chosen for the stochastic integral is significant. As we shall see in 6.4, the martingale property of the integral relies on choosing the left endpoint. Moreover, from an intuitive standpoint, given the use of stochastic processes to model real world systems evolving with time, we may want to only use "past" values, which corresponds to the left endpoint, rather than "future" ones.

Definition 6.2. Let A_t be a simple process as in Definition 6.1. Let $t > 0$ such that $t_j \leq t < t_{j+1}$. Define the *stochastic integral of a simple process* to be

$$\int_0^t A_s dB_s = \sum_{i=0}^{j} Y_i (B_{t_{i+1}} - B_{t_i}) + Y_j (B_t - B_{t_j}).$$

Remark 6.3. More generally, to start integrating from a non-zero time, we define

$$\int_r^t A_s dB_s = \int_0^t A_s dB_s - \int_0^r A_s dB_s$$

.

This newly defined integral has a number of properties we'd expect from an integral, such as linearity and continuity. However, it also has other useful properties that come into use for computations, as well as for proving Ito's formula in Section 9.

Proposition 6.4. *Let B_t be a standard Brownian motion with respect to a filtration $\{F_t\}$, and let A_t, C_t be simple processes. Then, the following properties hold.*

(i) (Linearity) Let $a, b \in \mathbb{R}$. Then, $aA_t + bC_t$ is also a simple process, and

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

Moreover, if $0 < r < t$,

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

(ii) (Continuity) With probability one, the function $t \mapsto \int_0^t A_s dB_s$ is a continuous function.

(iii) (Martingale) The process $Z_t = \int_0^t A_s dB_s$ is a martingale with respect to $\{F_t\}$.

(iv) (Ito isometry) The process $Z_t = \int_0^t A_s dB_s$ has variance

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

Proof. Let A_t be a simple process as in Definition 8.1. Let $t_j \leq t < t_{j+1}$.

(i) Linearity easily follows from the definition of a simple process and the definition of the stochastic integral for simple processes.

(ii) By Definition 8.3, $\int_0^t A_s dB_s = \sum_{i=0}^{j} Y_i (B_{t_{i+1}} - B_{t_i}) + Y_{t_j} (B_t - B_{t_j})$. Then, since the function $t \mapsto B_t$ is continuous with probability one, the function $t \mapsto \int_0^t A_s dB_s$ is also continuous with probability one.

(iii) We show $\mathrm{E}(Z_t \mid F_s) = Z_s$ for $s < t$.

Let $s < t$. We only consider the case of $s = t_j$ and $t = t_k$, where the meat of the argument lies. (Note that since $s < t$, $j < k$.)

Then, we have

$$Z_t = Z_s + \sum_{i=j}^{k-1} Y_i[B_{t_{i+1}} - B_{t_i}].$$

We will now use the properties of conditional expectation to compute $\mathrm{E}(Z_t \mid F_s)$.

$$\begin{aligned}\mathrm{E}(Z_t \mid F_s) &= \mathrm{E}(Z_s \mid F_s) + \sum_{i=j}^{k-1} \mathrm{E}(Y_i[B_{t_{i+1}} - B_{t_i}] \mid F_s) \text{ by Proposition 2.8(i)} \\ &= Z_s + \sum_{i=j}^{k-1} \mathrm{E}(Y_i[B_{t_{i+1}} - B_{t_i}] \mid F_s) \text{ by Proposition 2.8(iii)}\end{aligned}$$

Consider the sum. Since $j \leq i \leq k-1$, $t_i \geq s$. Hence, by Proposition 2.8(ii), each term of the sum can be expressed as

$$\mathrm{E}(Y_i[B_{t_{i+1}} - B_{t_i}] \mid F_s) = \mathrm{E}(\mathrm{E}(Y_i[B_{t_{i+1}} - B_{t_i}] \mid F_{t_i}) \mid F_s).$$

Recall that Y_i is F_{t_i} measurable by definition of a simple process and $B_{t_{i+1}} - B_{t_i}$ is independent of F_{t_i} by definition of Brownian motion.

$$\begin{aligned}\mathrm{E}(Y_i[B_{t_{i+1}} - B_{t_i}] \mid F_{t_i}) &= Y_i \mathrm{E}(B_{t_{i+1}} - B_{t_i} \mid F_{t_i}) \text{ by Proposition 2.8(iii)} \\ &= Y_i \mathrm{E}(B_{t_{i+1}} - B_{t_i}) \text{ by Proposition 2.8(iv)} \\ &= 0 \text{ since } B_{t_{i+1}} - B_{t_i} \sim N(0, t_{i+1} - t_i).\end{aligned}$$

Thus, all the terms of the sum are 0, giving us $\mathrm{E}(Z_t \mid F_s) = Z_s$.

(iv) We compute $\mathrm{Var}[Z_t] = \mathbb{E}[Z_t^2] - \mathbb{E}[Z_t]^2$ again using the properties of conditional expectation.

We again only consider the case $t = t_j$, giving us

$$Z_t = \sum_{i=0}^{j-1} Y_i[B_{t_{i+1}} - B_{t_i}] \text{ and}$$

$$Z_t^2 = \sum_{i=0}^{j-1} \sum_{k=0}^{j-1} Y_i[B_{t_{i+1}} - B_{t_i}] Y_k[B_{t_{k+1}} - B_{t_k}].$$

We first show $\mathbb{E}[Z_t] = 0$, which gives us $\mathrm{Var}[Z_t] = \mathbb{E}[Z_t^2]$.

$$\begin{aligned}\mathbb{E}[Z_t] &= \sum_{i=0}^{j-1} \mathbb{E}[Y_i[B_{t_{i+1}} - B_{t_i}]] \text{ by linearity of expectation} \\ &= \sum_{i=0}^{j-1} \mathbb{E}[\mathrm{E}(Y_i[B_{t_{i+1}} - B_{t_i}] \mid F_{t_i})] \text{ by Proposition 2.8(ii)} \\ &= \sum_{i=0}^{j-1} \mathbb{E}[Y_i \mathrm{E}([B_{t_{i+1}} - B_{t_i}] \mid F_{t_i})] \text{ by Proposition 2.8(iii)} \\ &= \sum_{i=0}^{j-1} \mathbb{E}[Y_i \mathbb{E}[[B_{t_{i+1}} - B_{t_i}]]] \text{ by Proposition 2.8(iv)} \\ &= 0 \text{ since } B_{t_{i+1}} - B_{t_i} \sim N(0, t_{i+1} - t_i).\end{aligned}$$

Now we compute $\mathbb{E}[Z_t^2]$. By linearity of expectation, we have

$$\mathbb{E}[Z_t^2] = \sum_{i=0}^{j-1}\sum_{k=0}^{j-1} \mathbb{E}[Y_i[B_{t_{i+1}} - B_{t_i}]Y_k[B_{t_{k+1}} - B_{t_k}]].$$

Each term with $i < k$ can be expressed as

$$\mathbb{E}[Y_i[B_{t_{i+1}} - B_{t_i}]Y_k[B_{t_{k+1}} - B_{t_k}]] = \mathbb{E}[\mathrm{E}(Y_i[B_{t_{i+1}} - B_{t_i}]Y_k[B_{t_{k+1}} - B_{t_k}] \mid F_{t_k})].$$

Note that for $i < k$, Y_i, $B_{t_{i+1}} - B_{t_i}$, and Y_k are F_{t_k}-measurable, and $B_{t_{k+1}} - B_{t_k}$ is independent of F_{t_k}, allowing us to compute the conditional expectation inside each term.

$$\begin{aligned}\mathrm{E}(Y_i[B_{t_{i+1}} - B_{t_i}]Y_k[B_{t_{k+1}} - B_{t_k}] \mid F_{t_k}) &= Y_i[B_{t_{i+1}} - B_{t_i}]Y_k\mathrm{E}([B_{t_{k+1}} - B_{t_k}] \mid F_{t_k}) \\ &\quad \text{by Proposition 2.8(iii)} \\ &= Y_i[B_{t_{i+1}} - B_{t_i}]Y_k\mathbb{E}[[B_{t_{k+1}} - B_{t_k}]] \\ &\quad \text{by Proposition 2.8(iv)} \\ &= 0 \text{ since } B_{t_{k+1}} - B_{t_k} \sim N(0, t_{k+1} - t_k).\end{aligned}$$

The same argument holds for $i > k$, so we are left with

$$\mathbb{E}[Z_t^2] = \sum_{i=0}^{j-1} \mathbb{E}[Y_i^2(B_{t_{i+1}} - B_{t_i})^2].$$

By Proposition 2.8(ii), we can express each term as

$$\mathbb{E}[Y_i^2(B_{t_{i+1}} - B_{t_i})^2] = \mathbb{E}[\mathrm{E}(Y_i^2[B_{t_{i+1}} - B_{t_i}]^2 \mid F_{t_i})].$$

Since Y_i^2 is F_{t_i}-measurable and $(B_{t_{i+1}} - B_{t_i})^2$ is independent of F_{t_i}, we can again use the properties of conditional expectation.

$$\begin{aligned}\mathrm{E}(Y_i^2[B_{t_{i+1}} - B_{t_i}]^2 \mid F_{t_i}) &= Y_i^2\mathrm{E}([B_{t_{i+1}} - B_{t_i}]^2 \mid F_{t_i}) \text{ by Proposition 2.8(iii)} \\ &= Y_i^2\mathbb{E}[(B_{t_{i+1}} - B_{t_i})^2] \text{ by Proposition 2.8(iv)} \\ &= Y_i^2(t_{i+1} - t_i) \text{ since } \mathbb{E}[(B_{t_{i+1}} - B_{t_i})^2] = \mathrm{Var}[B_{t_{i+1}} - B_{t_i}].\end{aligned}$$

Thus, we have

$$\mathbb{E}[Z_t^2] = \sum_{i=0}^{j-1} \mathbb{E}[Y_i^2](t_{i+1} - t_i) \tag{6.5}$$

Note that the function $s \mapsto \mathbb{E}[A_s^2]$ is a step function with the value $\mathbb{E}[Y_i^2]$ for $t_i \le s < t_{i+1}$. Thus, (6.5) gives us $\mathbb{E}[Z_t^2] = \int_0^t \mathbb{E}[A_s^2]ds$. □

Now that we have the stochastic integral for simple processes, we consider a more general process. We show that we can approximate bounded processes with continuous paths with a sequence of simple processes that are also bounded.

Lemma 6.6. *Let A_t be a process with continuous paths, adapted to the filtration F_t, such that there exists $C < \infty$ such that with probability one $|A_t| \le C$ for all t. Then there exists a sequence of simple processes $A_t^{(n)}$ such that for all t,*

$$\lim_{n\to\infty} \int_0^t \mathbb{E}[|A_s - A_s^{(n)}|^2]ds = 0. \tag{6.7}$$

Moreover, with probability one, for all n, t, $|A_t^{(n)}| \leq C$.

Proof. We define $A_t^{(n)}$ as a simple process approximation of A_t by letting $A_t^{(n)} = A(j,n)$ for $\frac{j}{n} \leq t < \frac{j+1}{n}$, where $A(0,n) = A_0$ and $A(j,n) = n\int_{(j-1)/n}^{j/n} A_s ds$. Note that by construction, $A_t^{(n)}$ are simple processes, and with probability one $|A_t^{(n)}| \leq C$.

Let $Y_n = \int_0^1 [A_t^{(n)} - A_t]^2 dt$.

Note that $A_t^{(n)}$ is essentially a step function approximation to A_t, which is continuous with probability one, so $A_t^{(n)} \to A_t$. Then, by the bounded convergence theorem applied to Lebesgue measure, $\lim_{n\to\infty} Y_n = 0$.

Since the random variables Y_n are uniformly bounded,

$$\lim_{n\to\infty} \mathbb{E}\left[\int_0^1 [A_t^{(n)} - A_t]^2 dt\right] = \lim_{n\to\infty} \mathbb{E}[Y_n] = 0.$$

□

Given the Lemma, we can now define the stochastic integral of bounded processes with continuous paths in terms of the integrals of simple processes. Fortunately, because of the constructive nature of the proof, we can actually find the sequence described in the Lemma and not merely know that it exists.

Definition 6.8. Let A_t and $A_t^{(n)}$ be as in Lemma 8.7. Define

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

The same properties as those in Proposition 6.4 hold because all the properties are preserved under L^2 limits. See proof of Proposition 3.2.3 in [1] for the full proof.

Remark 6.9. This definition easily extends to piecewise continuous paths. Let $t > 0$. If A_s has discontinuities at $0 \leq t_0 < t_1 < \cdots < t_n \leq t$, then we can define

$$\int_0^t A_s dB_s = \int_0^{t_1} A_s dB_s + \int_{t_1}^{t_2} A_s dB_s + \cdots + \int_{t_n}^{t} A_s dB_s.$$

Similarly, to Lebesgue integration, we can extend the definition of the stochastic integral to unbounded processes, by pushing the bounds further and further and then taking the limit. Suppose A_t is an unbounded adapted process with continuous paths. We have two options. The option that may seem more straightforward is cutting off everything above C on A_t, such that it becomes bounded by C, and then taking the limit as $C \to \infty$. This is the approach that is often taken for constructing the Lebesgue integral. For example, see Chapter 2 of [6]. However, the intuition for stochastic integral is different. Since stochastic integrals model a quantity evolving with time, we instead want to "stop time" at a certain point T, and then take the limit as $T \to \infty$. Note that we still have the upper bound on A_t monotonically increasing.

Definition 6.10. Let A_s be adapted with continuous paths. For each $n \in \mathbb{N}$, let $T_n = \min\{t : |A_t| = n\}$, and let $A_s^{(n)} = A_{s\wedge T_n}$, where $s \wedge T_n = \min\{s, T_n\}$. Define

$$\int_0^t A_s dBs = \lim_{n\to\infty} \int_0^t A_s^{(n)} dBs. \tag{6.11}$$

By making the path have the constant value n after time T_n, we have "stopped" the process and obtained bounded, continuous processes $A_s^{(n)}$. Thus, the integral $\int_0^t A_s^{(n)} dBs$ is well defined.

Moreover, the limit in (6.11) is also well-defined, as we shall now show.

Let $K_t = \max_{0 \leq s \leq t} |A_s|$, which exists because $[0, t]$ is compact and A_s is continuous.

Then, for all $n \geq K_t$, A_s is bounded by K_t for $0 \leq s \leq t$, so $A_s^{(n)} = A_s$ for $0 \leq s \leq t$. Thus, $\int_0^t A_s dBs = \int_0^t A_s^{(n)} dBs$ for $n \geq K_t$, so the limit is well-defined.

Note that K_t is a random variable that depends on the path. Although the other three properties hold, this integral may not satisfy the martingale property. (See Section 4.1 in [1] for more details.)

7. Ito's Formula

Now that we've defined the stochastic integral, we might wonder if there exists a stochastic calculus analogue to the Fundamental Theorem of Calculus. Indeed, there is: Ito's formula.

Theorem 7.1. *(Ito's formula I) Suppose f is a C^2 function. Then for every t,*

$$f(B_t) = f(B_0) + \int_0^t f'(B_s) dB_s + \frac{1}{2} \int_0^t f''(B_s) ds \tag{7.2}$$

Alternatively, we can write

$$df(B_t) = f'(B_t) dB_t + \frac{1}{2} f''(B_t) dt \tag{7.3}$$

Remark 7.4. Ito's formula is also called the chain rule of stochastic calculus, referencing the differential form (7.3).

Proof. This proof omits some of the technical details in favor of focusing on the bigger ideas. See the proof of Theorem 3.3.1 in [1] for a similar proof with the technical details added. Alternatively, for an even more technical proof, see proof of Theorem 8.6.1 in [2] for a different proof.

For ease of notation, we prove the Theorem for $t = 1$. Consider $f(B_1) - f(B_0)$.

First write this as the telescoping sum

$$f(B_1) - f(B_0) = \sum_{j=1}^{n} f(B_{j/n}) - f(B_{(j-1)/n}) \tag{7.5}$$

so that we can examine the infinitesimal changes in f as well as put it in a form that is closer to the stochastic integral of a simple process.

Now let us look at the second-order Taylor expansion for each term in the sum. Ordinarily, for normal calculus, the second-order term goes to 0, but because of the added randomness, this does not happen here.

Let $\Delta_{j,n} = B_{j/n} - B_{(j-1)/n}$. Then, the second-order Taylor expansion gives us

$$f(B_{j/n}) - f(B_{(j-1)/n}) = f'(B_{(j-1)/n}) \Delta_{j,n} + \frac{1}{2} f''(B_{(j-1)/n}) \Delta_{j,n}^2 + o(\Delta_{j,n}^2). \tag{7.6}$$

Since (7.5) and (7.6) hold for all $n \in \mathbb{N}$, we have that $f(B_1) - f(B_0)$ is the sum of the following three limits

$$\lim_{n \to \infty} \sum_{j=1}^{n} f'(B_{(j-1)/n}) \Delta_{j,n}, \tag{7.7}$$

$$\frac{1}{2}\lim_{n\to\infty}\sum_{j=1}^{n} f''(B_{(j-1)/n})\Delta_{j,n}^2, \text{ and} \tag{7.8}$$

$$\lim_{n\to\infty}\sum_{j=1}^{n} o(\Delta_{j,n}^2) \tag{7.9}$$

The first limit is just a simple process approximation to a stochastic integral. Thus, (7.7) is equal to $\int_0^1 f'(B_t)dB_t$.

The third limit is actually just the quadratic variation, and so we have that (7.9) is $\lim_{n\to\infty} o(1) = 0$.

Now consider the second limit. Comparing with the statement of the Theorem, we can see that we'd like to show that (7.8) is $\frac{1}{2}\int_0^1 f''(B_t)dt$.

Recall again the quadratic variation. Note that if we could somehow pull $f''(B_{(j-1)/n})$ out of the sum, the limit of the sum would again be the quadratic variation. Yet, the second derivative need not be constant, so we cannot quite do that. However, because f is C^2, we can approximate f'' by a step function, which allows us to do this trick on the intervals where the step function is constant.

Call $f''(B_t) = h(t)$. Since h is continuous, we can approximate $h(t)$ by the step function $h_\epsilon(t)$ such that for all t, $|h(t) - h_\epsilon(t)| < \epsilon$.

On each interval where h_ϵ is constant, we can pull out h_ϵ, leaving us with a Riemann sum. Thus, by the linearity of the Riemann integral, we get

$$\frac{1}{2}\sum_{j=1}^{n} h_\epsilon((j-1)/n)\Delta_{j,n}^2 = \frac{1}{2}\int_0^1 h_\epsilon(t)dt.$$

Now we show that the sums with h and h_ϵ are sufficiently close. This is easy to show because of how how we defined h_ϵ.

$$\left|\sum_{j=1}^{n}[h(t) - h_\epsilon(t)]\Delta_{j,n}^2\right| \le \epsilon\sum_{j=1}^{n}\Delta_{j,n}^2 \to \epsilon.$$

Thus, we have

$$\lim_{\epsilon\to 0}\left(\lim_{n\to\infty}\sum_{j=1}^{n} h_\epsilon(t)\Delta_{j,n}^2\right) = \lim_{\epsilon\to 0}\int_0^1 h_\epsilon(t)dt = \int_0^1 h(t)dt = \int_0^1 f''(B_t)dt,$$

giving us that (7.8) is $\frac{1}{2}\int_0^1 f''(B_s)ds$, as desired. □

We give an example that emphasizes the difference between the Ito integral and the normal calculus integral.

Example 7.10. Let $f(x) = x^2$. Then, $f'(x) = 2x$ and $f''(x) = 2$. Thus,

$$\begin{aligned} B_t^2 &= B_0^2 + \int_0^t f'(B_s)dB_s + \frac{1}{2}\int_0^t f''(B_s)ds \text{ by Ito's formula} \\ &= 0 + \int_0^t 2B_s dB_s + \frac{1}{2}\int_0^t 2ds \\ &= 2\int_0^t B_s dB_s + t, \end{aligned}$$

giving us

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

Note that this is different from what we might naively expect without seeing Ito's formula, as normal calculus would have us incorrectly guess

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

Ito's formula also has other more general forms. We state one of these here. The interested reader is encouraged to read Theorem 3.4.2 and 3.4.3 in [1] for two even more general versions.

Theorem 7.11. *(Ito's formula II) Suppose $f(t,x)$ is a function that is C^1 in t and C^2 in x. Then for every t,*

$$f(t,B_t) = f(0,B_0) + \int_0^t \frac{\partial f}{\partial x}(s,B_s)dB_s + \int_0^t \left[\frac{\partial f}{\partial s}(s,B_s) + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(s,B_s)\right] ds \tag{7.12}$$

Again, we can also write this as

$$df(t,B_t) = \frac{\partial f}{\partial x}(s,B_s)dB_s + \left[\frac{\partial f}{\partial s}(s,B_s) + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(s,B_s)\right] ds \tag{7.13}$$

The proof is similar to Ito's formula I, except that the Taylor expansion gives us another term. See the proof of Theorem 3.4.2 in [1] for more details.

We end with an example that uses Ito's formula II, this time using the differential form of Ito's formula.

Example 7.14. Let $f(t,x) = e^{at+bx}$. Then, $\frac{\partial f}{\partial t}(t,x) = af(t,x)$, $\frac{\partial f}{\partial x}(t,x) = bf(t,x)$, and $\frac{\partial^2 f}{\partial x^2}(t,x) = b^2 f(t,x)$.

Let $X_t = f(t,B_t) = e^{at+bB_t}$. Then,

$$\begin{aligned} dX_t &= \frac{\partial f}{\partial x}(s,B_s)dB_s + \left[\frac{\partial f}{\partial s}(s,B_s) + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(s,B_s)\right] ds \text{ by Ito's formula II} \\ &= bX_t dB_t + \left(a + \frac{b^2}{2}\right) X_t dt. \end{aligned}$$

Acknowledgments

I'd like to thank my mentor Yucheng Deng for answering my questions and reviewing my paper. I would also like to thank Abhimanyu Choudhary for helping me understand filtrations. Finally, I'd like to thank Professor Peter May for organizing the program.

References

[1] Gregory Lawler. Stochastic Calculus: An Introduction with Applications. 2014.
[2] Rick Durrett. Probability: Theory and Examples. Cambridge University Press. 2013.
[3] Gordan Zitkovic. Theory of Probability, Parts I and II. https://web.ma.utexas.edu/users/gordanz/lecture_notes_page.html.
[4] Peter Morters and Yuval Peres. Brownian Motion. https://www.stat.berkeley.edu/ãldous/205B/bmbook.pdf. 2008.
[5] Christopher Heil. Functions of Bounded Variation. http://people.math.gatech.edu/ heil/6337/fall07/section3.5a.pdf

[6] Elias Stein and Rami Shakarchi. Real Analysis: Measure Theory, Integration, and Hilbert Spaces. Princeton University Press. 2005.

[7] Vivi Andasari. http://people.bu.edu/andasari/courses/stochasticmodeling/lecture5/stochasticlecture5.html