BROWNIAN MOTION AND ITO'S FORMULA

ETHAN LEWIS

Abstract. This expository paper presents an introduction to stochastic calculus. In order to be widely accessible, we assume only knowledge of basic analysis and some familiarity with probability. We will cover the basics of measure theoretic probability, then describe Brownian motion, then introduce stochastic integration.

Contents

1. Introduction 1
2. Measure Theoretic Probability 2
3. Brownian Motion 6
4. Quadratic Variation 9
5. Stochastic Integration 11
6. Acknowledgements 16
References 16

1. Introduction

Brownian motion aims to describe a process of a random value whose direction is constantly fluctuating. Examples of such processes in the real world include the position of a particle in a gas or the price of a security traded on an exchange. We can consider a simpler process that is constant except for jumps at discrete time intervals, where the size and direction of each jump is a random value. For example, we can think of a random walk where at each time interval, we flip a coin, and if it lands heads, we take a step right, and if tails, we take a step left. To understand Brownian motion, we consider the limit of such a process as the intervals between jumps and the size of the jumps becomes vanishingly small.

In addition, we may want to integrate with respect to such a process. As with our random walk example above, we could consider moving along a surface with a variable slope where we want to know our total vertical displacement. Once again, we can use the integral in the discrete case to understand the limit of Brownian motion.

We will begin by providing some background of measure theoretic probability and a few key theorems. Next we will define Brownian motion and describe some of its properties. The concept of quadratic variation is necessary for the theorems of stochastic calculus, so we treat it in the next section. Finally we formulate the stochastic integral and present Itô's Formula.

2. Measure Theoretic Probability

We will first introduce some definitions fundamental to measure theory.

Definition 2.1. Let S be a set and Σ_0 be a collection of subsets of S. Σ_0 is called an *algebra on S* if it satisfies the following.

(i) S is an element of Σ_0.
(ii) If $F \in \Sigma_0$, then $F^c \in \Sigma_0$, where F^c is the complement of F.
(iii) If $F \in \Sigma_0$ and $G \in \Sigma_0$, then $F \cup G \in \Sigma_0$.

Remark. Note the following properties of algebras:

- Since $\emptyset = S^c$, $\emptyset \in \Sigma_0$.
- If $F, G \in \Sigma$, then $F \cap G \in \Sigma$ since $F \cap G = (F^c \cup G^c)^c$.

Definition 2.2. Let Σ be an algebra on a set S. We call Σ a *σ-algebra on S* if for any countable subset of the algebra $\{F_n \in \Sigma : n \in \mathbb{N}\}$,

$$\bigcup_{n=1}^{\infty} F_n \in \Sigma.$$

As before, it follows that if $\{F_n \in \Sigma : n \in \mathbb{N}\}$, then

$$\bigcap_{n=1}^{\infty} F_n \in \Sigma.$$

Definition 2.3. Let $\mathcal{A}$ be a collection of subsets of a set S. The *σ-algebra generated by $\mathcal{A}$*, denoted $\sigma(\mathcal{A})$, is the smallest σ-algebra such that $\mathcal{A} \in \sigma(\mathcal{A})$. That is, for any Σ that is a σ-algebra on S such that $\mathcal{A} \in \Sigma$, $\sigma(\mathcal{A}) \subseteq \Sigma$.

Definition 2.4. For any set S, the *Borel σ-algebra on S*, denoted $\mathscr{B}(S)$, is the σ-algebra generated by the collection of open subsets of S. For the Borel σ-algebra on $\mathbb{R}$ we use the notation $\mathscr{B} = \mathscr{B}(\mathbb{R})$,

Since a closed set is the complement of an open set, a Borel σ-algebra contains all the closed sets and all the sets formed by countable unions and intersections of open sets and closed sets.

Definition 2.5. Let Σ_0 be an algebra and μ_0 be a non-negative set function on Σ_0

$$\mu_0 : \Sigma_0 \to [0, \infty].$$

μ_0 is called *countably additive* if $\mu_0(\emptyset) = 0$ and for any sequence $(F_n : n \in \mathbb{N})$ of disjoint subsets of Σ_0 such that $\bigcup F_n \in \Sigma_0$,

$$\sum_{n=1}^{\infty} \mu_0(F_n) = \mu_0\left(\bigcup_{n=1}^{\infty} F_n\right).$$

Note that the values in the above equation could be infinite.

Definition 2.6. Let S be a set, and let Σ be a σ-algebra on S. The pair (S, Σ) is called a *measurable space*. If μ is a countably additive set function on Σ then μ is a *measure* on (S, Σ). The triple (S, Σ, μ) is called a *measure space*.

Definition 2.7. Suppose (S, Σ, μ) is measure space. If $\mu(S) = 1$, then μ is called a *probability measure*, and (S, Σ, μ) is called a *probability triple*.

The standard form of a probability triple is $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is the set of all possible outcomes called the *sample space* and $\mathcal{F}$ is the collection of *events*, which are subsets of Ω, to which we can assign a probability. The *probability* of an event $E \in \mathcal{F}$ is $\mathbb{P}(E)$.

Definition 2.8. Let (S, Σ) be a measurable space, and $f : S \to \mathbb{R}$. f is called Σ-measurable if for any $A \in \mathscr{B}$, the pre-image of A is an element of Σ.

Definition 2.9. For a measurable space $(\Omega, \mathcal{F})$ that serves as our sample space and collection of events, a *real random variable* is a $\mathcal{F}$-measurable function $X : \Omega \to \mathbb{R}$.

The following two lemmas are presented without proof but can be found in [2].

Lemma 2.10. *For any measurable space (S, Σ), $f : S \to \mathbb{R}$ is Σ-measurable if for all $c \in \mathbb{R}$,*

$$\{s \in S : f(s) \leq c\} \in \Sigma.$$

This tells us that for any random variable X and any $c \in \mathbb{R}$ we can always ask the probability of X being at most c, that is $\mathbb{P}\{X \leq c\}$, since the event $\{X \leq c\}$ is an element of $\mathcal{F}$.

Note that we use the simplified notation $\{X \leq c\}$ for $\{\omega \in \Omega \ : \ X(\omega) \leq c\}$.

Lemma 2.11. *Let $m\Sigma$ be the set of Σ-measurable functions for a measurable space (S, Σ). If $\lambda \in \mathbb{R}$ and $f, g \in m\Sigma$, then*

(i) $f + g \in m\Sigma$
(ii) $fg \in m\Sigma$
(iii) $\lambda f \in m\Sigma$.

This tells us that we can scale, add, and multiply random variables and get random variables out.

For simplicity we will use the following elementary definitions of independence as they are sufficient for our discussion.

Definitions 2.12. Events $E_1, E_2, \ldots$ are *independent* if for distinct $i_1, \ldots, i_n$,

$$\mathbb{P}(E_{i_1} \cap \cdots \cap E_{i_n}) = \prod_{k=1}^{n} \mathbb{P}(E_{i_k}).$$

Random variables $X_1, X_2, \ldots, X_n$ are *independent* if

$$\mathbb{P}\Big\{X_k \leq x_k \ : \ k \in \{1, \ldots, n\}\Big\} = \prod_{k=1}^{n} \mathbb{P}\{X_k \leq x_k\}.$$

A complete treatment of the Lebesgue integral is beyond the scope of this paper and can be found in [2] or [3]. We will cover only its definition.

For the *indicator function* of a set A, we use the notation

$$1_A(x) = \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases}$$

and for the *indicator random variable* of an event E, we use the notation

$$1_E = \begin{cases} 1, & E \text{ occurs} \\ 0, & E \text{ does not occur} \end{cases}$$

Definition 2.13. Suppose (S, Σ, μ) is a measure space and s is a Σ-measurable function defined by

$$s(x) = \sum_{i=1}^{n} c_i \, 1_{E_i}(x), \tag{2.14}$$

for sets $E_i \in \Sigma$ and constants $c_i \geq 0$. If $E \in \Sigma$, define

$$I_E(s) = \sum_{i=1}^{n} c_i \, \mu(E \cap E_i).$$

If f is Σ-measurable and non-negative, define the *Lebesgue integral* of f as

$$\int_E f \, d\mu = \sup I_E(s),$$

where the supremum is taken over functions of the form given in (2.14) such that $s \leq f$. Finally if f is Σ-measurable, let

$$f^+ = \max(f, 0), \qquad f^- = \max(-f, 0).$$

Note that both are positive. If either $\int_E f^+ \, d\mu$ or $\int_E f^- \, d\mu$ is finite then we define

$$\int_E f \, d\mu = \int_E f^+ \, d\mu - \int_E f^- \, d\mu.$$

If $\int_S f \, d\mu$ is finite, we say that f is *integrable* and write $f \in L^1(S, \Sigma, \mu)$.

Definition 2.15. Suppose $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability triple and $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. The *expectation* of X is

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

If $X \geq 0$ and is Σ-measurable we define $0 \leq \mathbb{E}[X] \leq \infty$ the same way.

Definition 2.16. Suppose $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability triple, Y is an integrable random variable, and $\mathcal{G}$ is a sub-σ-algebra of $\mathcal{F}$. The *conditional expectation* $E[Y \mid \mathcal{G}]$ is defined to be the unique $\mathcal{G}$-measurable random variable such that if $A \in \mathcal{G}$,

$$\mathbb{E}[Y \, 1_A] = \mathbb{E}[E[Y \mid \mathcal{G}] \, 1_A]. \tag{2.17}$$

Here, "unique" means unique up to an event of measure zero, which is to say that if two $\mathcal{G}$-measurable random variables, Z_1 and Z_2 satisfy (2.17), then $\mathbb{P}\{Z_1 \neq Z_2\} = 0$.

Note that while expectation is a constant variable (in particular it is the weighted mean of a random variable), the conditional expectation is itself a random variable that denotes the best guess given the information in $\mathcal{G}$. This definition of conditional expectation necessitates a proof of its existence and uniqueness which can be found in [2].

To better understand the nature of conditional expectation, we will consider a simple example. Let

$$\Omega = \{(1,1),\, (1,2),\, \ldots,\, (2,1),\, (2,2),\, \ldots,\, (6,5),\, (6,6)\}$$

be the set of possible rolls of two fair, independent 6-sided dice and $\mathcal{F} = \mathcal{P}(\Omega)$ the power set of Ω. Let

$$\mathcal{G} = \sigma\left(\Big\{\{(i,1),\, \ldots,\, (i,6)\} \ : \ i \in \{1,\, \ldots,\, 6\}\Big\}\right)$$

be the sub-σ-algebra of events determined only by the value of the first die. Let X be the sum of the two dice values, so X is $\mathcal{F}$ measurable, and $\mathbb{E}[X] = 7$. On the other hand, $E[X \mid \mathcal{G}]$ is random variable determined by the value of the first die whose value is what we expect the sum to be given the value of the first die. In particular, $E[X \mid \mathcal{G}]$ is 3.5 greater than the mean value of the first die for an event in $\mathcal{G}$. One final note on conditional expectation is that we can have examples like

$$\mathbb{E}[X \mid \text{the first die is } 4] = 7.5$$

or

$$\mathbb{E}[X \mid \text{the first die is greater than } 2] = 8$$

where the expectation of X given some other event is a constant; in fact, it is a value taken by $E[X \mid \mathcal{G}]$.

Proposition 2.18. *The following are properties of conditional expectation.*

- *If Y is $\mathcal{G}$-measurable. then*

$$E[Y \mid \mathcal{G}] = Y. \tag{2.19}$$

- *If Y is independent of the events in $\mathcal{F}$, then*

$$E[Y \mid \mathcal{F}] = \mathbb{E}[Y]. \tag{2.20}$$

- **Linearity:** *If Y, Z are random variables and a, b are constants, then*

$$E[a\,Y + b\,Z \mid \mathcal{G}] = a\,E[Y \mid \mathcal{G}] + b\,E[Z \mid \mathcal{G}]. \tag{2.21}$$

- **Tower Property:** *If $\mathcal{G}$ and $\mathcal{F}$ are σ-algebras with $\mathcal{G} \subseteq \mathcal{F}$, then*

$$E[\,E[Y \mid \mathcal{F}] \mid \mathcal{G}] = E[Y \mid \mathcal{G}]. \tag{2.22}$$

Proof. We will prove only the tower property; the proof of the other properties can be found in [1] and [2]. Let

$$Z = E[\,E[Y \mid \mathcal{F}] \mid \mathcal{G}].$$

So Z is the unique (up to an event of probability zero) $\mathcal{G}$-measurable random variable such that for all $A \in \mathcal{G}$,

$$\mathbb{E}[\,E[Y \mid \mathcal{F}]\,1_A] = \mathbb{E}[Z\,1_A].$$

Similarly $E[Y \mid \mathcal{F}]$ is the unique $\mathcal{F}$-measurable random variable such that for all $B \in \mathcal{F}$,

$$\mathbb{E}[Y\,1_B] = \mathbb{E}[\,E[Y \mid \mathcal{F}]\,1_B].$$

If $A \in \mathcal{G}$, then also $A \in \mathcal{F}$, and so

$$\mathbb{E}[Y\,1_A] = \mathbb{E}[\,E[Y \mid \mathcal{F}]\,1_A] = \mathbb{E}[Z\,1_A],$$

and since Z is $\mathcal{G}$-measurable, the result follows from uniqueness of conditional expectation. $\square$

We will now give a few useful results used in probability theory.

Lemma 2.23 (Markov's Inequality)**.** *Suppose X is an integrable random variable with respect to a probability triple $(\Omega, \mathcal{F}, \mathbb{P})$. Then for all $c \geq 0$,*

$$c\,\mathbb{P}\{|X| \geq c\} \leq \mathbb{E}[|X|]$$

Proof. Let A be the event $\{|X| \geq c\}$. Then

$$c\,\mathbb{P}\{|X| \geq c\} = \int_A c\, d\mathbb{P} \leq \int_A |X|\, d\mathbb{P} \leq \int_\Omega |X|\, d\mathbb{P} = \mathbb{E}[|X|]$$

since $A \in \Omega$ and $|X| \geq 0$. $\square$

Lemma 2.24 (Chebyshev's Inequality)**.** *Suppose X is a random variable that is square integrable (that is, X^2 is integrable). Let $\mu = \mathbb{E}[X]$. For all $c \geq 0$,*

$$c^2\,\mathbb{P}\{|X - \mu| \geq c\} \leq \mathrm{Var}[X],$$

where $\mathrm{Var}[X] = \mathbb{E}\left[(X-\mu)^2\right]$ *is the variance of X.*

Proof. Since $|X - \mu| \geq c$ if and only if $|X - \mu|^2 \geq c^2$,

$$\mathbb{P}\{|X - \mu| \geq c\} = \mathbb{P}\{|X - \mu|^2 \geq c^2\},$$

and the result follows from Markov's inequality. $\square$

Lemma 2.25 (First Borel-Cantelli Lemma)**.** *If (E_n) is a sequence of events such that*

$$\sum_{n=1}^{\infty} \mathbb{P}\{E_n\} < \infty,$$

then,

$$\mathbb{P}\left\{\bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} E_n\right\} = 0.$$

In other words, the probability of infinitely many E_n occuring is zero.

The proof of the Borel-Cantelli Lemma can be found in [2].

3. Brownian Motion

Before discussing Brownian motion we will first discuss a bit about random walks. Suppose $X_1, X_2, \ldots$ are independent random variables, each with equal probability of being -1 or 1. Let

$$S_n = X_1 + X_2 + \cdots + X_n.$$

S_n is a simple random walk in $\mathbb{Z}$ which represents a discrete-time process where in each time step there is an equal probability of progressing right or left along the number line. The Central Limit theorem tells us that as n increases, S_n approaches a normal distribution with mean 0 and variance n. Equivalently, $Z_n = S_n/\sqrt{n}$ approaches a standard normal normal distribution, that is, one with mean 0 and variance 1. Put precisely, for $a < b$

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

where

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

Brownian motion can be thought of as the limit of a random walk, where instead of going farther in time, we make the time between steps and the magnitude of each step smaller, so that in the limit, the process is continuous in time. Let us denote

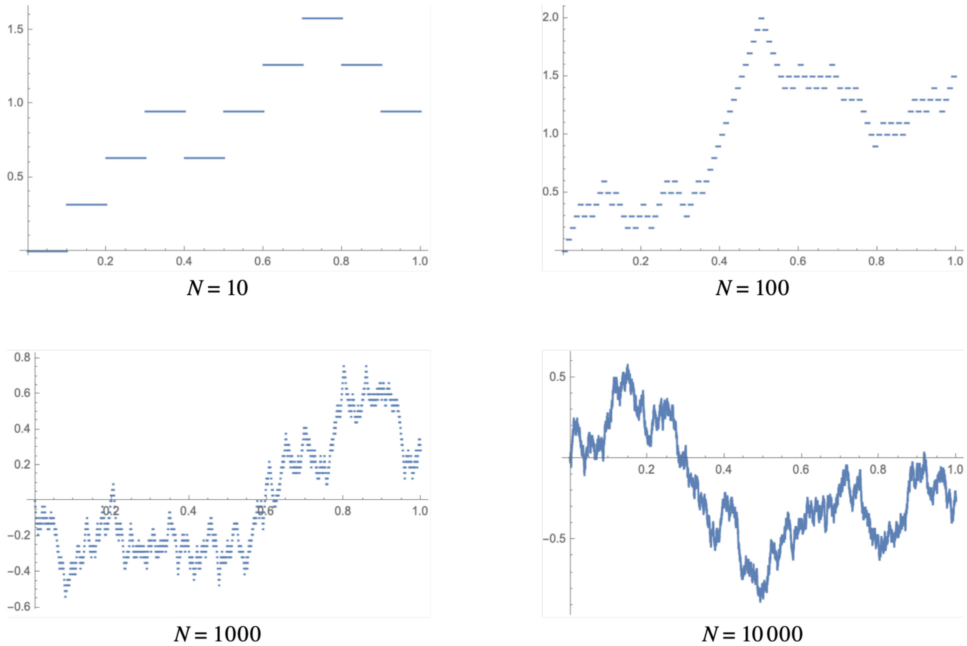

FIGURE 1. Simulations of $W_t^{(N)}$ for various values of N

the space increments as Δx and time increments as Δt and let the value of our process for these increments at a time t be

$$\sum_{0<j\Delta t\leq t} \Delta x \, X_j.$$

Our initial example of S_n fits this with $\Delta x = 1$, $\Delta t = 1$.

Now, to take the limit, we will consider time increments $\Delta t = 1/N$ and denote the resulting process as

$$W_t^{(N)} = \Delta x \, (X_1 \ + \ \cdots \ X_{\lfloor Nt \rfloor}).$$

What should be our choice of Δx for this choice of N? We want $\mathrm{Var}[W_1^{(N)}] = 1$, so

$$\begin{aligned} 1 = \mathrm{Var}[W_1^{(N)}] &= \mathrm{Var}[\Delta x \, (X_1 \ + \ \cdots \ X_N)] \\ &= (\Delta x)^2 \, (\mathrm{Var}[X_1] \ + \ \ldots \ + \mathrm{Var}[X_N]) \\ &= (\Delta x)^2 \, N \end{aligned}$$

would necessitate

$$\Delta x = \sqrt{1/N} = \sqrt{\Delta t}.$$

Figure 1 shows some simulations of possible paths of the process $W_t^{(N)}$ for various values of N. As we increase the number of steps, the process appears more like it is continuous as the size of the jumps becomes small. The process also becomes more jagged as there are many points that are not differentiable, and in fact, with probability one, Brownian motion is nowhere differentiable.

While it useful to think of Brownian motion as the limit of a random walk, we now give a more rigorous definition. Brownian motion is an example of a *stochastic* *process*, that is a collection of random variables indexed by time. The process $W_t^{(N)}$ given above is also a stochastic process.

Definition 3.1. A stochastic process B_t is called a *(one-dimensional) Brownian motion with drift* m *and variance* σ^2 starting at the origin if it satisfies the following.

(i) $B_0 = 0$.
(ii) For $s < t$ the distribution of $B_t - B_S$ is normal with mean $m(t-s)$ and variance $\sigma^2(t-s)$.
(iii) If $s < t$ the random variable $B_t - B_s$ is independent of the values of B_r for $r \leq s$.
(iv) With probability one, the function $t \mapsto B_t$ is a continuous function of t.

If B_t is a Brownian motion with drift 0 and variance 1, it is called a *standard Brownian motion*, and in fact, if

$$Y_t = mt + \sigma B_t,$$

then Y_t is Brownian motion with mean m and variance σ^2.

For $s < t$,

$$Y_t - Y_s = m\,t + \sigma\,B_t - m\,s - \sigma\,B_s = m\,(t-s) + \sigma\,(B_t - B_s),$$

which is distributed as

$$m\,(t-s) + \sigma\,N(0,\ t-s) = N(m\,(t-s),\ \sigma^2\,(t-s)),$$

where $N(\mu,\ \sigma^2)$ denotes a normal distribution with mean μ and variance σ^2. Therefore, condition (ii) is satisfied. The other conditions are immediate.

This definition does not guarantee that any such process exists. However, by the preceding discussion, the existence of a standard Brownian motion is sufficient for the existence of any Brownian motion.

It is possible to construct a standard Brownian motion by defining the process on a countable, dense subset of $\mathbb{R}$, showing that the process is continuous, and then extending the process to all the real numbers, using the result from calculus that a continuous function is uniquely determined by its values on a dense subset. This construction is lengthy and will be left out; the interested reader can consult [1].

We will now discuss a few properties of Brownian motion.

Definition 3.2. Suppose we have a set of σ-algebras $\mathcal{F}_t$ such that if $s < t$, then $\mathcal{F}_s \subseteq \mathcal{F}_t$. A stochastic process M_t is a *continuous martingale with respect to the filtration* $\{\mathcal{F}_t\}$ if:

- For all t, M_t is an integrable $\mathcal{F}_t$-measurable random variable.
- For all $s < t$, $E[M_t \mid \mathcal{F}_s] = M_s$ with probability one.

It is useful to think of a filtration as an increasing amount of information. When no filtration is given explicitly, one assumes that $\mathcal{F}_t$ is the information contained in $\{M_s\ :\ s \leq t\}$. We can extend our definition of Brownian motion to say that a process B_t is a Brownian motion with respect to a filtration $\{\mathcal{F}_t\}$ if each B_t is $\mathcal{F}_t$-measurable and B_t satisfies the conditions of Definition 3.1, replacing condition (iii) with

(iii) If $s < t$ the random variable $B_t - B_s$ is independent of $\mathcal{F}_s$.

If B_t is a Brownian motion with respect to the filtration $\{\mathcal{F}_t\}$ and with drift zero, then for $s < t$,

$$E[B_t \mid \mathcal{F}_s] = E[B_s \mid \mathcal{F}_s] + E[B_t - B_s \mid \mathcal{F}_s] = B_s + \mathbb{E}[B_t - B_s] = B_s$$

by (2.19), (2.20), and (2.21). Therefore, B_t is a continuous martingale. In other words, given any amount of information about the path of a Brownian motion up to some time, the expectation of its value at a future time is equal to its current value.

Definition 3.3. A continuous time process X_t is called *Markov* if for every t, the conditional distribution of $\{X_s : s \geq t\}$ given $\{X_r : r \leq t\}$ is the same as the conditional distribution given X_t. In other words, the future of the process is conditionally independent of the past given the present value.

Suppose B_t is a Brownian motion. For every t, if $s \geq t$, the distribution of $B_s - B_t$ is independent of $\{B_r : r \leq t\}$, so the distribution of $B_s = B_t + (B_s - B_t)$ given B_t is the same as the distribution given $\{B_r : r \leq t\}$, so B_t is a Markov process.

4. Quadratic Variation

We will now discuss a result that does not immediately appear to be useful but will serve in our discussion of Itô's Formula later on.

Suppose $X(t)$ is a stochastic process indexed by times t, and suppose Π is a partition of $[0, t]$, that is, $\Pi = \{t_0, \dots, t_n\}$ such that $0 = t_0 < t_1 < \cdots < t_n = t$, and write $\|\Pi\| = \max\limits_{1 \leq j \leq n} \{t_j - t_{j-1}\}$. Let

$$Q_X(t; \Pi) = \sum_{j=1}^{n} [X(t_j) - X(t_{j-1})]^2.$$

Definition 4.1. Suppose $X(t)$ is a stochastic process. The *quadratic variation*, denoted $\langle X \rangle_t$ is the limit of $Q_X(t; \Pi_n)$ for a sequence of partitions $\{\Pi_n\}$ such that $\|\Pi_n\| \to 0$.

This definition requires the existence of this limit, and uniqueness for all such sequences of partitions. We will prove these conditions for Brownian motion.

Let $B(t)$ be a standard Brownian motion. For the remainder of this section, we will use $Q(t; \Pi) = Q_B(t; \Pi)$. Since $B(t_j) - B(t_{j-1})$ is distributed as

$$N(0,\, t_j - t_{j-1}) = \sqrt{t_j - t_{j-1}}\, N(0,\, 1),$$

we find

$$\begin{aligned}
\mathbb{E}[Q(t; \Pi)] &= \mathbb{E}\left[\sum_{j=1}^{n} (B(t_j) - B(t_{j-1}))^2\right] \\
&= \sum_{j=1}^{n} \mathbb{E}\left[\,(B(t_j) - B(t_{j-1}))^2\right] \\
&= \sum_{j=1}^{n} \mathbb{E}\left[\,\left(\sqrt{t_j - t_{j-1}}\, N(0,\, 1)\right)^2\right] \\
&= \sum_{j=1}^{n} (t_j - t_{j-1})\, \mathbb{E}\left[\,N(0,\, 1)^2\right] \\
&= \sum_{j=1}^{n} (t_j - t_{j-1}) = t.
\end{aligned}$$

Additionally, using the fact that subsequent increments of Brownian motion are independent and that $\mathrm{Var}[N(0,\,1)^2] = 2$, we find

$$\begin{aligned}
\mathrm{Var}[Q(t;\,\Pi)] &= \sum_{j=1}^{n} \mathrm{Var}\left[\,(B(t_j) - B(t_{j-1}))^2\right] \\
&= \sum_{j=1}^{n} \mathrm{Var}\left[\,(t_j - t_{j-1})\,N(0,1)^2\right] \\
&= \sum_{j=1}^{n} (t_j - t_{j-1})^2\,\mathrm{Var}\left[\,N(0,1)^2\right] \\
&= 2\sum_{j=1}^{n} (t_j - t_{j-1})^2 \\
&\le 2\|\Pi\| \sum_{j=1}^{n} (t_j - t_{j-1}) \\
&= 2\|\Pi\| t.
\end{aligned}$$

So we find that as our partition becomes finer, $\|\Pi\|$ approaches zero, and consequently the quadratic variation with respect to Π approaches a constant value of t. This observation is made precise by the following theorem.

Theorem 4.2. *Suppose $B(t)$ is a standard Brownian motion and $\{\Pi_n\}$ is a sequence of partitions on $[0,\,t]$ with $\|\Pi_n\| \to 0$. Then $Q(t;\,\Pi_n) \to t$ in probability, that is, for all $\epsilon > 0$,*

$$\lim_{n\to\infty} \mathbb{P}(\,|Q(t;\,\Pi_n) - t| > \epsilon) = 0. \tag{4.3}$$

Moreover, if

$$\sum_{n=1}^{\infty} \|\Pi_n\| < \infty, \tag{4.4}$$

then with probability one, $Q(t;\,\Pi_n) \to t$.

Proof. By Chebyshev's inequality, for all $k \in \mathbb{N}$,

$$\mathbb{P}\left\{|Q(t;\,\Pi_n) - t| > \frac{1}{k}\right\} \le \frac{\mathrm{Var}[Q(t;\,\Pi_n)]}{(1/k)^2} \le 2k^2\,\|\Pi_n\|\,t,$$

which limits to 0 as $n \to \infty$, satisfying (4.3).

If (4.4) holds, then

$$\sum_{n=1}^{\infty} \mathbb{P}\left\{|Q(t;\,\Pi_n) - t| > \frac{1}{k}\right\} \le 2k^2\,t\,\sum_{n=1}^{\infty} \|\Pi_n\| < \infty,$$

and so by the Borel-Cantelli lemma, with probability one, there exists N such that for all $n \ge N$,

$$|Q(t;\,\Pi_n) - t| \le 1/k. \qquad \square$$

So far, we have discussed the quadratic variation on $[0,\,t]$; however, because for all $s < t$, the distribution of $B_t - B_s$ is the same as that of $B_{t-s} - B_0$, it follows that the quadratic variation of a standard Brownian motion on the interval $[s,\,t]$ is equal to $t - s$.

In the following theorem we show that for a general Brownian motion, not only is the quadratic variation a constant variable, but it depends only on the variance and not on the drift.

Theorem 4.5. *For a general Brownian motion $W_t = m\,t + \sigma\,B_t$ with mean m and variance σ^2, $\langle W\rangle_t = \sigma^2\,t$.*

Proof. Let Π_n be the partition $\{0,\ t/n,\ 2t/n,\ \ldots,\ (n-1)t/n,\ t\}$. $\{\Pi_n\}$ is a sequence of partitions such that $\|\Pi_n\| \to 0$, so $\langle W\rangle_t$ is equal to the limit as $n \to \infty$ of

$$Q_W(t;\,\Pi_n) = \sum_{j=1}^{n} \left[W\left(\frac{t\,j}{n}\right) - W\left(\frac{t\,(j-1)}{n}\right)\right]^2 = \text{(A)} + \text{(B)} + \text{(C)},$$

where

$$\begin{aligned}
\text{(A)} &= \sigma^2 \sum_{j=1}^{n} \left[B\left(\frac{t\,j}{n}\right) - B\left(\frac{t\,(j-1)}{n}\right)\right]^2, \\
\text{(B)} &= \frac{2\sigma m t}{n} \sum_{j=1}^{n} \left[B\left(\frac{t\,j}{n}\right) - B\left(\frac{t\,(j-1)}{n}\right)\right], \\
\text{(C)} &= \sum_{j=1}^{n} \frac{m^2 t^2}{n^2},
\end{aligned}$$

and as $n \to \infty$,

$$\begin{aligned}
\text{(A)} &\to \sigma^2 \langle B\rangle_t = \sigma^2\,t, \\
\text{(B)} &= \frac{2\sigma m t}{n}(B_t - B_0) \to 0, \\
\text{(C)} &= \frac{m^2 t^2}{n} \to 0.
\end{aligned}$$

□

5. Stochastic Integration

In standard calculus, a differential equation of the form

$$df(t) = C(t,\ f(t))\,dt$$

says that the rate of change with respect to t of the function $f(t)$ is equal to $C(t,\ f(t))$, which, of course, can depend on both t and the value of f at time t. One can describe the value of $f(t)$ given an initial condition by the integral

$$f(x) = f(x_0) + \int_{x_0}^{x} C(t,\ f(t)\,dt,$$

which can be calculated whether by solving the differential equation directly when possible or by numerical approximation of the integral.

Similarly, we can describe a process by a *stochastic differential equation* (SDE) of the form

$$dX_t = m(t,\ X_t)\,dt + \sigma(t,\ X_t)\,dB_t, \tag{5.1}$$

where B_t is a standard Brownian motion. This equation says that the process X_t evolves at time t like a Brownian motion with drift $m(t,\ X_t)$ and variance $\sigma(t,\ X_t)^2$.

X_t can be described as

$$X_t = X_0 + \int_0^t m(s,\, X_s)\, ds + \int_0^t \sigma(t,\, X_t)\, dB_t.$$

The first integral is the standard integral, although it is worth noting that the integrand is random; nonetheless, the definition of integration still applies. To make sense of the second term we must give a definition of

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

for a more general process A_s. The definition we will use will be that of the *Itô integral.*

To help us understand stochastic calculus, before we define the stochastic integral, we will discuss how to approximate the solution to an SDE by the *stochastic Euler method.* If X_t is described by the SDE in (5.1) then we can simulate X_t in discrete steps of size Δt by the formula

$$X_{t+\Delta t} = X_t + \Delta t\, m(t,\, X_t) + \sqrt{\Delta t}\, \sigma(t,\, X_t)\, N(0,\, 1). \tag{5.2}$$

This is equivalent to assuming m and σ are constant on the interval $[t,\, t + \Delta t)$, and thus, on that interval, X_t will increase according to a normal distribution with mean $\Delta t\, m(t,\, X_t)$ and variance $\Delta t\, \sigma(t,\, X_t)^2$. The variance function can be thought of as a "bet" on the Brownian motion, where the change in X_t is the change in B_t amplified by the value of $\sigma(t,\, X_t)$, and a negative value of $\sigma(t,\, X_t)$ is betting against the Brownian motion, that is, X_t will increase in proportion to how much B_t decreases.

We will first define the stochastic integral for simple processes. In the discussion that follows, assume B_t is a Brownian motion with respect to the filtration $\{\mathcal{F}_t\}$.

Definition 5.3. A process A_t is a *simple process* if there exist times

$$0 = t_0 < t_1 < \cdots < t_n < t_{n+1} = \infty$$

and random variables Y_j for $j = 0,\, 1,\, \ldots,\, n$ that are $\mathcal{F}_{t_j}$-measurable such that

$$A_t = Y_j, \quad t_j \le t < t_{j+1}$$

and that $\mathbb{E}[Y_j^2] < \infty$. Note that since Y_j is $\mathcal{F}_{t_j}$-measurable, A_t is $\mathcal{F}_t$-measurable.

Definition 5.4. If A_t is a simple process, we define the stochastic integral

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

by

$$Z_{t_j} = \sum_{k=0}^{j-1} Y_k\, [B_{t_{k+1}} - B_{t_k}],$$

and more generally by

$$Z_t = Z_{t_j} + Y_j\, [B_t - B_{t_j}], \quad t_j \le t < t_{j+1},$$

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

Note that this is essentially the same as (5.2), except that because this is not a simulation, instead of a generic normal distribution, we use the change in value of the Brownian motion, which is normally distributed.

To define the integral for continuous processes, we must first provide a lemma regarding approximation by simple processes. We say that a process A_t is *adapted to the filtration* $\{\mathcal{F}_t\}$ if for all t, A_t is $\mathcal{F}_t$-measurable.

Lemma 5.5. *Suppose A_t is a process with continuous paths, adapted to the filtration $\{\mathcal{F}_t\}$, and suppose that there exists $C < \infty$ such that with probability one, $|A_t| \leq C$ for all t. Then there exists a sequence of simple processes $A_t^{(n)}$ such that for all t,*

$$\lim_{n\to\infty} \int_0^t \mathbb{E}\left[|A_s - A_s^{(n)}|^2\right] ds = 0. \tag{5.6}$$

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

The following proof relies on the bounded convergence theorem, a corollary of Lebesgue's dominated convergence theorem, and Fubini's theorem. The unfamiliar reader can either trust these results or consult [2] or [3].

Proof. It suffices to show (5.6) for any fixed value of t. Assume without loss of generality that $t = 1$, and let

$$A_t^{(n)} = A(j,\, n) := n \int_{j/n}^{(j+1)/n} A_s\, ds, \qquad \frac{j}{n} \leq t \leq \frac{j+1}{n}.$$

Note that $A(j,\, n)$ is the mean of A_t on the interval $\left[\frac{j}{n},\, \frac{j+1}{n}\right]$, and so $|A_t^{(n)}| \leq C$. Because the function $t \mapsto A_t$ is continuous,

$$A_t^{(n)} \to A_t.$$

Let

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

By the bounded convergence theorem,

$$\lim_{n\to\infty} Y_n = 0.$$

Since the random variables $\{Y_n\}$ are uniformly bounded, then by the bounded convergence theorem and Fubini's theorem,

$$\lim_{n\to\infty} \int_0^t \mathbb{E}[|A_s - A_s^{(n)}|^2]\, ds = \lim_{n\to\infty} \mathbb{E}[Y_n] = \mathbb{E}\left[\lim_{n\to\infty} Y_n\right] = 0. \qquad \square$$

Definition 5.7. For a bounded, continuous process A_s adapted to the filtration $\{\mathcal{F}_s\}$, we define the stochastic integral as

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

where $A_s^{(n)}$ is a sequence of simple processes satisfying (5.6).

The existence of this limit can be shown but will not be proven here; a proof can be found in [1].

We can also define the integral for processes with piecewise continuous paths. If A_t is a bounded, adapted process with paths that are continuous on $[0,\, t]$ except at t_0, then we can define the integral (by (5.9) in the following proposition) as

$$\int_0^t A_s\, dB_s = \int_0^{t_0} A_s\, dB_s + \int_{t_0}^t A_s\, dB_s.$$

Proposition 5.8. *Suppose B_t is a standard Brownian motion with respect to a filtration $\{\mathcal{F}_t\}$, and A_t, C_t are bounded, adapted processes with piecewise continuous paths.*

- **Linearity.** *If a and b are constants, then*
$$\int_0^t (a\,A_s + b\,C_s)\,dB_s = a\int_0^t A_s + b\int_0^t C_s\,dB_s.$$
Also, if $r < t$, then
$$\int_0^t A_s\,B_s = \int_0^r A_s\,B_s + \int_r^t A_s\,B_s. \tag{5.9}$$
- **Martingale property.** *The process*
$$Z_t = \int_0^t A_s\,dB_s$$
is a martingale with respect to $\{\mathcal{F}_t\}$.
- **Variance rule.** *Z_t is square integrable and*
$$\operatorname{Var}[Z_t] = \mathbb{E}\left[Z_t^2\right] = \int_0^t \mathbb{E}\left[A_s^2\right]\,ds.$$
- **Continuity.** *With probability one, $t \mapsto Z_t$ is a continuous function.*

Suppose we have a function $f(B_t)$ and we want to determine its value at a time t given an initial value of $f(B_0)$. In standard calculus, we can often apply the fundamental theorem of calculus to calculate
$$f(x) = f(0) + \int_0^x f'(t)\,dt.$$
The stochastic analogue of the fundamental theorem of calculus is Itô's formula, which we will now prove.

Theorem 5.10 (Itô's formula)**.** *Suppose f is a C^2 function and B_t is a standard Brownian motion. Then for every t,*
$$f(B_t) = f(B_0) + \int_0^t f'(B_s)\,dB_s + \frac{1}{2}\int_0^t f''(B_s)\,ds,$$
or, written in differential form,
$$df(B_t) = f'(B_t)\,dB_t + \frac{1}{2}f''(B_t)\,dt.$$

Proof. Let $\{\Pi_n\}$ be a sequence of partitions
$$0 = t_{0,\,n} < t_{1,\,n} < \cdots < t_{k_n,\,n} = t,$$
such that
$$\sum_{n=1}^{\infty} \|\Pi_n\| < \infty, \qquad \|\Pi_n\| = \max_{1\le j\le k_n}\{t_{j,\,n} - t_{j-1,\,n}\}.$$
For any n, we can write the telescoping sum (denoting B_t as $B(t)$ and $t_{j,\,n}$ as t_j),
$$f(B(t)) - f(B(0)) = \sum_{j=1}^{k_n}[\,f(B(t_j)) - f(B(t_{j-i}))\,].$$

Let $m(j,\,n)$ and $M(j,\,n)$ be the minimum and maximum, respectively, of $f''(x)$ for $B(t_{j-1}) \leq x \leq B(t_j)$. By Taylor's theorem,

$$f(B(t_j)) - f(B(t_{j-1})) = f'(B(t_{j-1}))\,[\,B(t_j) - B(t_{j-1})\,] + \xi_n,$$

where

$$\frac{m(j,\,n)}{2}\,[\,B(t_j) - B(t_{j-1})\,]^2 \leq \xi_{j,\,n} \leq \frac{M(j,\,n)}{2}\,[\,B(t_j) - B(t_{j-1})\,]^2.$$

Hence if we let

$$Q^1(\Pi_n) = \sum_{j=1}^{k_n} f'(B(t_{j-1}))\,[\,B(t_j) - B(t_{j-1})\,],$$

$$Q^2_-(\Pi_n) = \sum_{j=1}^{k_n} \frac{m(j,\,n)}{2}\,[\,B(t_j) - B(t_{j-1})\,]^2,$$

$$Q^2_+(\Pi_n) = \sum_{j=1}^{k_n} \frac{M(j,\,n)}{2}\,[\,B(t_j) - B(t_{j-1})\,]^2,$$

then we have

$$Q^2_-(\Pi_n) \leq f(B(t)) - f(B(0)) - Q^1(\Pi_n) \leq Q^2_+(\Pi_n). \tag{5.11}$$

First we will try to understand Q^2. By Theorem 4.2, with probability one, for all $0 < q < r < t$,

$$\lim_{n\to\infty} \sum_{q \leq t_{j,n} < r} [\,B(t_{j,\,n}) - B(t_{j-1,\,n})\,]^2 = q - r.$$

On the event that this is true, by the continuity of B_t and f'', we have

$$\lim_{n\to\infty} Q^2_-(\Pi_n) = \lim_{n\to\infty} Q^2_+(\Pi_n) = \frac{1}{2}\int_0^t f''(B(s))\,ds.$$

Now we will try to understand Q^1. We will prove the theorem under the additional assumption that there exists $K < \infty$ such that $|f''(x)| \leq K$ for all x. By the mean value theorem,

$$|f'(B(s)) - f'(B(t_{j-1,\,n}))| \leq K\,|B(s) - B(t_{j-1,\,n})|.$$

Let the simple process

$$A^{(n)}(s) = f'(B(t_{j-1,\,n})), \quad t_{j-1,\,n} \leq s < t_{j,\,n}.$$

For $s \in [t_{j-1,\,n},\, t_{j,\,n})$,

$$\begin{aligned}\mathbb{E}\left[|f'(B(s)) - A^{(n)}(s)|^2\right] &\leq K^2\,\mathbb{E}\left[|B(s) - B(t_{j-1,\,n})|^2\right] \\ &= K^2\,\mathrm{Var}\,[B(s) - B(t_{j-1,\,n})] = K^2\,[s - t_{j-1,\,n}] \leq K^2\,\|\Pi_n\|,\end{aligned}$$

Therefore,

$$0 \leq \lim_{n\to\infty} \int_0^t \mathbb{E}\left[|f'(B(s)) - A^{(n)}(s)|^2\right]\,ds \leq \lim_{n\to\infty} t\,K^2\,\|\Pi_n\| = 0,$$

which satisfies (5.6), so

$$\int_0^t f'(B(s))\,ds = \lim_{n\to\infty}\int_0^t A^{(n)}(s)\,ds = \lim_{n\to\infty} Q^1(\Pi_n).$$

So by taking the limits in (5.11), we see that

$$f(B(t)) - f(B(0)) = \int_0^t f'(B(s))\,ds + \frac{1}{2}\int_0^t f''(B(s))\,ds. \qquad \square$$

Example 5.12. We can use Itô's formula to solve stochastic integrals such as $\int_0^t B_s\,dB_s$. If we let $f(x) = x^2$, then we find that

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

Therefore,

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

6. Acknowledgements

I would like to thank my mentor in the REU, Nixia Chen, for her help with writing this paper. I would also like to thank Gregory Lawler for providing lectures, texts, and discussions essential to my understanding of these topics. Finally, thank you to Peter May for making the REU work despite a forced move to remote work.

References

[1] Gregory Lawler. Stochastic Calculus: An Introduction with Applications. 2014. `http://www.math.uchicago.edu/~lawler/finbook.pdf`
[2] David Williams. Probability with Martingales. Cambridge University Press. 1991.
[3] Walter Rudin. Principles of Mathematical Analysis. McGraw-Hill. 1953.
[4] Martin Haugh. A Brief Introduction to Stochastic Calculus. 2016. `http://www.columbia.edu/~mh2078/FoundationsFE/IntroStochCalc.pdf`