STOCHASTIC CALCULUS: UNDERSTANDING BROWNIAN MOTION AND QUADRATIC VARIATION

ANNE CARLSTEIN

Abstract. This is a paper introducing Brownian motion and Ito Calculus. Brownian motion is introduced using random walks. Stochastic calculus and Ito's Lemma are motivated with a discussion of variation of Brownian motion.

Contents

1. Introduction and Some Probability

Brownian motion is a major component in many fields. In addition to its definition in terms of probability and stochastic processes, the importance of using models for continuous random motion manifests itself in physics, statistics, finance, engineering, and beyond. But, prior to using any mathematical idea for applications, it is first important to understand its derivation and fundamental characteristics. Here, we will focus on Brownian motion, and how it is used to construct stochastic integrals.

Before we dive into Brownian motion, it is important to understand the type of objects in probability that we are dealing with. The following definitions and results will give us some insight into the probabilistic background for random motion.

Definition 1.1. A **probability space** is defined with three parameters $(\Omega, \mathcal{F}, \mathbb{P})$, and the space itself has measure 1.

Definition 1.2. The **sample set** is denoted by Ω, and is the space of any possible outcomes (or combinations of outcomes) that can occur.

Definition 1.3. The letter $\mathcal{F}$ denotes a σ-algebra (defined below), and is comprised of subsets of Ω. Elements in $\mathcal{F}$ are called events.

Definition 1.4. $\mathbb{P}$ denotes the probability function, where the domain is $\mathcal{F}$, and the range is the interval from $[0, 1]$. $\mathbb{P}$ fulfills the following properties:

(1) $\mathbb{P}(\Omega) = 1$

(2) For events $E_1, E_2, ... \in \mathcal{F}$ that are disjoint, then:

$$\mathbb{P}(\cup_{E_i}) = \sum_i \mathbb{P}(E_i)$$

Definition 1.5. A σ-algebra is defined as a collection of subsets where the following are true:

(1) $\emptyset \in \mathcal{F}$
(2) For events $A_1, A_2, ... \in \mathcal{F}$:

$$\cup_{n=1}^{\infty} A_n \in \mathcal{F}$$

(3) If some event $A \in \mathcal{F}$, then the complement is also true, i.e., $\Omega \backslash A \in \mathcal{F}$

These definitions give us a large amount of information to internalize! Consider the definition of $\mathbb{P}$. Intuitively, this tells us that when we say event A has a probability of occurring, we are taking the measure of the outcomes in Ω that correspond to the event occurring. Therefore, it makes sense that the probability of any event in the total sample space should be one. Additionally, because $\mathbb{P}$ is a function with domain $\mathcal{F}$, it is important to make sure that the function is defined on all events in $\mathcal{F}$: otherwise, the probability function would not work.

Definition 1.6. A **Borel set** is the smallest σ-algebra containing all open sets of real numbers.

Definition 1.7. A **random variable** is a function whose domain is a probability space, and whose range is the real numbers.

$$X : \Omega \to (-\infty, \infty)$$

Let B denote a Borel set in $(-\infty, \infty)$. Intuitively, here, we can think of a Borel set as taking countable collections of unions or intersections of subsets of the real line.

Define $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$. The function X must fulfill the following properties:

(1) $\mathbb{P}(X = \pm\infty) = 0$
(2) For any Borel set B, $X^{-1}(B) \in \mathcal{F}$. This means that X is a measurable function.

In other words, property (2) means that any event in the real numbers must be mapped to by some possible outcomes, which in turn must be measurable events in $\mathcal{F}$.

For our purposes, to specify a random variable, we can use the **cumulative distribution function** $F_X(x)$, which measures $\mathbb{P}[X \leq x]$ for $x \in \mathbb{R}$. The cumulative distribution must fulfill the following properties:

(1) $F_X(\infty) = 1$
(2) $F_X(-\infty) = 0$
(3) F_X is non-decreasing and F_X is right-continuous.

Definition 1.8. The **expected value** for a discrete random variable X is:

$$\mathbb{E}[X] = \sum_i i\mathbb{P}(X = i)$$

For a continuous random variable, the **expected value** is defined with the following integral:

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

In other words, the expected value of a random variable calculates a weighted average: This occurs by integrating or summing over the domain of the variable, and scaling the values that the variable by the probability that the variable does take on that value.

Properties of Expectation

(1) For random variables X, Y, and constants a, b:

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$$

(2) When random variables X and Y are independent (note that this is not true in general), the following is true:

$$\mathbb{E}[XY] = \mathbb{E}[X][Y]$$

Definition 1.9. The **variance** of a random variable (denoted σ^2) is calculated with the following formula:

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

In other words, the variance calculates the average squared distance from the weighted mean that the values of random variable take.

Definition 1.10. The **normal distribution** fulfills the following cumulative distribution function:

$$F_X(x) = \int_{-\infty}^{x} \frac{e^{\frac{-(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$$

When graphed, this is a bell curve. The appearance of the normal distribution is determined by two parameters, the mean (expected value) and variance; denoted Norm(μ, σ^2). The mean μ defines the point on the x-axis at which the highest point of the curve is centered, and the variance σ^2 determines how spread out its shape is. A higher variance means that the curve is more spread out, farther away from the mean. Note that if $\sigma^2 = 0$, then the above cumulative distribution function would be undefined. (In fact, if $\sigma^2 \to 0$, this will approach the Dirac Delta function, with its peak at the mean μ).

Theorem 1.11. *The* ***Central Limit Theorem****. Given a set of independent and identically distributed random variables $X_1, X_2, X_3, ...,$ with mean μ and variance σ^2:*

$$Z_N = \frac{\sum_{i=1}^{N} X_i - \sum_{i=1}^{N} \mu}{\sqrt{N\sigma^2}} \sim \text{Norm}(0, 1)$$

We can define a random variable Z_N, by taking the sum of N random variables (the X_1, X_2, X_3... do not have to be normally distributed), subtracting the sum of their average, and scaling by the standard deviation. The Central Limit Theorem tells us that as N becomes large, the distribution of Z_N approaches the standard normal distribution, with mean 0, and variance 1.

Definition 1.12. A **filtration**, denoted $\{\mathcal{F}_t\}$ is an increasing sequence of σ-fields. In the context of random variables, given a sequence $X_1, X_2, .., X_n$ of random variables, $\mathcal{F}_n$ is the information that is contained in the sequence of random variables.

Definition 1.13. A **stochastic process** is a collection of random variables that is indexed by time (or some other mathematical set of objects). When the set is countable, the stochastic process is **discrete-time**. If the set that the process is indexed by is isomorphic to some segment of the real numbers, then the stochastic process is **continuous.**

In other words, a stochastic process gives a correspondence between some set of objects and random variables: each random variable is uniquely determined by one object in the set.

Example 1.14. An example of a stochastic process is a series of coin flips. Each time the coin is flipped, it corresponds to a separate random variable, each with probability $\frac{1}{2}$ of heads, and $\frac{1}{2}$ of tails. This is a discrete time stochastic process, and is formally known as a Bernoulli process.

Definition 1.15. We say that a stochastic process Y_t is **adapted** to a filtration $\mathcal{F}_t$ if for all times t, Y_t is measurable with respect to $\mathcal{F}_t$. For Y_t to be $\mathcal{F}_t$-measurable, this means that $Y_t(\omega)$ can be determined by the information from up until time t.

2. Random Walks

Since Brownian motion is a model for continuous random motion, it makes sense to first consider and to then generalize a discrete-time case.

Definition 2.1. For the purposes of this paper, in a **1-dimensional random walk**, a person starts at point x_0, and moves at time $t = 1, 2, 3, ...$ and thereafter can move either right or left, each with probability $\frac{1}{2}$.

We can express the location of the walker (with deterministic starting point x_0) after n steps with the following notation:

$$S_n = x_0 + X_1 + X_2 + ... + X_n \tag{2.2}$$

Here, the X_j's are discrete random variables:

$$X_j = \pm 1 \quad P(X_j = 1) = \frac{1}{2} = P(X_j = -1)$$

There are some questions that would immediately logically follow from this setup: what is the expected value of S_n? Or, in other words, after a number of iterations, what is the walker's expected location? We can calculate this, but the result $\mathbb{E}[S_n]$ is not as interesting as one might hope: As it turns out,

$$\mathbb{E}[S_n] = 0$$

This occurs because on each step, the walker has a 50-50 chance of going left versus right. There is an equal probability of contributing 1 and -1 to the sum, and clearly, $0.5(-1) + 0.5(1) = 0$.

Instead, we can calculate $\mathbb{E}[S_n^2]$, which will track the net displacement, and will allow for us to calculate the variance:

$$\mathbb{E}[S_n^2] = n \tag{2.3}$$

Proof.

$$\mathbb{E}[S_n^2] = \mathbb{E}[\sum_{j=1}^{n}\sum_{k=1}^{n} X_j X_k] = \sum_{j=1}^{n}\sum_{k=1}^{n}\mathbb{E}(X_j X_k)$$

We can split the sum up based on whether $j = k$:

$$\sum_{j=1}^{n}\sum_{k=1}^{n}\mathbb{E}(X_j X_k) = \sum_{j=1}^{n}\mathbb{E}[X_j^2] + \sum_{j\neq k}^{n}\mathbb{E}(X_j X_k)$$

Note that $(X_j)^2 = 1$ for all j, so the first component of the sum is n. For the second sum, the product $X_j X_k$ could take on 2 different values:

(1) $X_j X_k = 1$ when $X_j = X_k = 1$, or when $X_j = X_k = -1$.
(2) $X_j X_k = -1$ when $X_j = 1; X_k = -1$, or when $X_j = -1; X_k = 1$.

Both of these cases happen with probability of $\frac{1}{2}$, and the terms in this sum will cancel out, contributing zero. This leaves us with:

$$\mathbb{E}[S_n^2] = n \implies \mathrm{Var}(S_n) = n$$

□

This result is interesting, but hopefully not too surprising. This expectation calculation and variance indicates that in general, the typical distance traveled after n time steps is on the order of $\sqrt{n}$.

3. Brownian Motion

Now that we understand random walks, a logical modification would be to generalize from a discrete random walk to a continuous case. Previously, we considered a random walk with natural number time steps, that can move one step at a time; instead, we consider a random walk as the time increments decrease to zero, and at each step, the increments of distance movement are also scaled accordingly.

We can derive Brownian motion from random walks, by considering how to appropriately scale the distance moved when we divide the time intervals.

Let S_i be a standard random walk, indexed by time $t = \{0, 1, 2, ...\}$, and with $\Delta x = 1$. From before, we know that the expected value is 0, and the variance is n.

Consider how Δx should be scaled if we modify Δt. Let N be a very large number, and let $\Delta t = \frac{1}{N} = \delta$. Of course, if we just let $\Delta x = \frac{1}{N}$, then this would just be a "very small" random walk.

We want to find a scaling factor such that:

$$B_{k\delta} \approx \Delta x S_k$$

Let $k = N$. We want Brownian motion to be normalized with $\mathbb{E}[B_1^2] = 1$. For this to be true, we need:

$$\mathbb{E}[\Delta x S_N^2] = (\Delta x)^2 \mathbb{E}[S_N^2] = (\Delta x)^2 N$$

This means that the scaling factor for Δx is $\frac{1}{\sqrt{N}}$.

Note that any $t = \frac{k}{N}$ for some k. Using how we previously defined B_t in terms of S_k:

$$B_t = B_{k/N} \approx \frac{S_k}{\sqrt{N}} = \frac{S_k}{\sqrt{k}} \frac{\sqrt{k}}{\sqrt{N}} = \frac{S_k \sqrt{t}}{\sqrt{k}}$$

Additionally, remember that S_N is the sum of N random variables X_k with mean 0, and variance 1. By applying the Central Limit Theorem as $i \to N$ for some very large N, we see that $\frac{S_k}{\sqrt{k}}$ approaches a Normal distribution. Finally, the factor of $\sqrt{t}$ scales the random variable for each t, and tells us that the distribution is $B_t \sim \text{Norm}(0, t)$.

This gives us the idea for how to define Brownian motion. In addition, to axiomatize this process of continuous random motion, a few other properties are required.

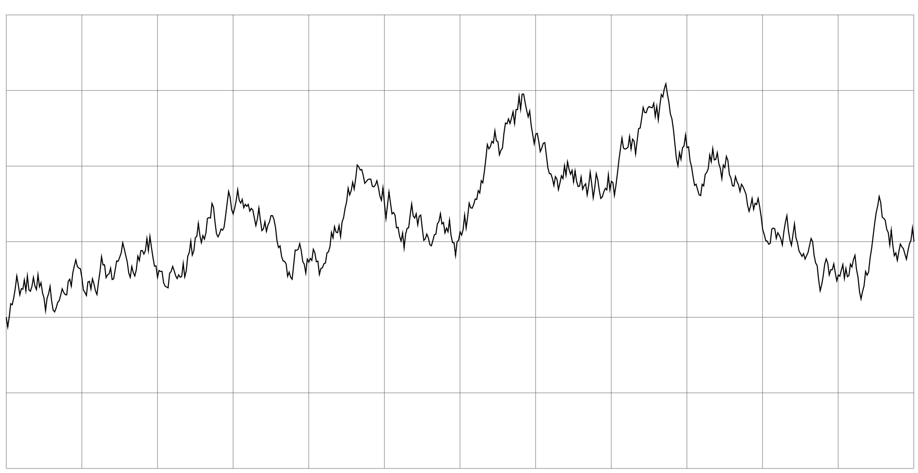
Definition 3.1. Brownian motion is a stochastic process. It is a collection of random variables that are indexed by time which satisfy the following three properties:

(1) For all times t, B_t is normally distributed with mean 0, and variance t.
(2) The random variable $B_t - B_s$ for $(0 \leq s < t < \infty$ is normally distributed with mean 0, and variance $t - s$.
(3) The random variable $B_t - B_s$ is independent of random variables $\{B_v : v \leq s\}$ for $s < t$
(4) Almost surely, the function $t \to B_t$ is continuous

Brownian motion that starts with $B_0 = 0$ is called **standard Brownian motion**.

Another way to think about this process is to imagine a random walker whose movement at every time increment is determined by assigning a normally distributed random variable. It is helpful to think about Brownian motion as a function. This means that the random walker at any given time could move any distance and (depending on the dimension with which we are working) in any direction: it is specified by selection from the Normal distribution at that point.

This graph is an example of what a one-dimensional Brownian motion looks like :



(Note: this is a pseudorandom generation of a Brownian motion path, so its appearance is not planned in advance!)

Theorem 3.2. *With probability 1, Brownian motion is nowhere differentiable.*

We will not prove this here (for full proof, see Lawler [1] pp. 49-50). But, for some intuition, we look back at how Brownian motion is defined. Additionally, note that if it is possible to differentiate a continuous function at a point t, then this means that we can determine the derivative based on how the function behaves right before that t, for some $t-\epsilon$. If this were true, it would imply that $B_{t-\epsilon}$ could "predict" B_t, which we know is not possible, because a key property of Brownian motion is independent increments. So, it makes sense that we can not differentiate it at any time t.

The fact that Brownian motion is nowhere differentiable leads to certain challenges when defining stochastic integration. Ultimately, our goal is to construct integrals that take the following form:

$$\int_0^k Y_t dB_t$$

Why would we care about integrals of this form? While it might not be obvious why one would want to integrate one stochastic process with respect to another process involving Brownian motion, this does have a logical interpretation in financial applications. For example, if Y_t is a stochastic process that tells us how many shares of a stock we own at time t, and B_t tracks the price of the stock, then this integral will tell us the net gain from time $t=0$ to time $t=k$.

However, defining this integral is not as simple as previous results we may have seen in analysis. Before we see how stochastic calculus is defined, we need to understand the concept of bounded variation.

4. Variation

Definition 4.1. The **variation** of f on $[a,b]$ over some partition $\Pi = \{a = x_0, x_1, ..., x_n = b\}$ is defined by

$$V_\Pi = \sum |f(x_i) - f(x_{i-1})|$$

$$V(f) = \lim_{|\Pi|\to 0} V_\Pi(f)$$

A function has **bounded variation** if the variation is bounded by some constant.

In other words, to have bounded variation, the sum of the difference between values that the function takes over any arbitrary partition must be finite. One way to think about this is the "wobbliness" of the function on a closed interval. Intuitively, for most smooth functions that one would consider, on a closed interval, it would seem obvious that the variation would be bounded - since we are just measuring the length of a "nice" function on a closed interval, surely this would be finite! However, there are smooth functions that do not have bounded variation.

Example 4.2. Consider the following function, on the interval $[0, \frac{1}{\pi}]$:

$$f(x) = \begin{cases} (x)(\cos(\frac{1}{x})), & \text{if } x \neq 0 \\ 0, & \text{if } x = 0 \end{cases} \tag{4.3}$$

Additionally, consider the following partition, where k is any natural number:

$$\Pi = \{x_0 = 0, x_1 = \frac{1}{2k\pi}, x_2 = \frac{1}{(2k-1)\pi}, x_3 = \frac{1}{(2k-2)\pi}, ...$$

$$..., x_{2k-2} = \frac{1}{3\pi}, x_{2k-1} = \frac{1}{2\pi}, x_{2k} = \frac{1}{\pi}\}$$

We can evaluate the value of the function at each point of the partition:

$$f(x_0) = 0$$

$$f(x_1) = \frac{1}{2k\pi}(\cos(2k\pi)) = \frac{1}{2k\pi}$$

$$f(x_2) = -\frac{1}{(2k-1)\pi}$$

$$...$$

$$f(x_2k) = \frac{-1}{\pi}$$

We can calculate the variation of this function by summing the evaluation of the function on the partition:

$$\sum |f(x_i) - f(x_{i-1)}| =$$

$$= |\frac{1}{2k\pi} - 0| + |-\frac{1}{(2k-1)\pi} - \frac{1}{2k\pi}| + |\frac{1}{(2k-2)\pi} - (-\frac{1}{(2k-1)\pi})|... + |\frac{-1}{\pi} - \frac{1}{2\pi}|$$

Note that within each absolute value expression, the two terms always have the same sign. We can simplify further:

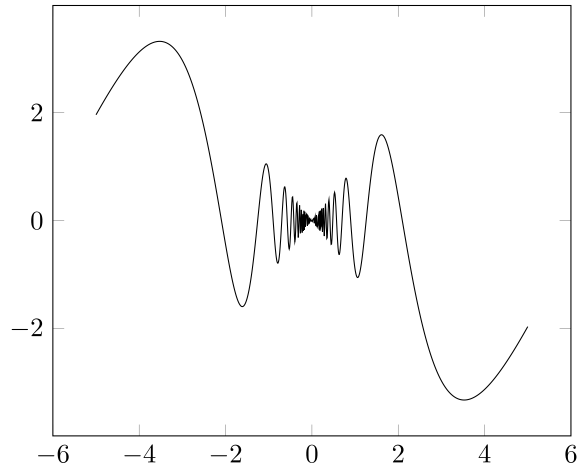
$$= \frac{1}{2k\pi} + \frac{1}{(2k-1)\pi} + \frac{1}{2k\pi} + \frac{1}{(2k-2)\pi} + \frac{1}{(2k-1)\pi} + ... + \frac{1}{\pi} + \frac{1}{2\pi}$$

$$= 2(\frac{1}{2k\pi} + \frac{1}{(2k-1)\pi} + \frac{1}{(2k-2)\pi} + \frac{1}{(2k-3)\pi} + ... + \frac{1}{2\pi} + \frac{1}{1\pi})$$

$$= \frac{2}{\pi}((\frac{1}{2k} + \frac{1}{2k-1} + \frac{1}{2k-2} + \frac{1}{2k-3} + ... + \frac{1}{2} + 1))$$

Note that the sum inside of the parentheses is the harmonic series of $\frac{1}{k}$, and as $k \to \infty$, this sum will diverge. Therefore, the function does not have bounded variation.

If we look at this graph, it should make sense that this function does not have a bounded arc length on the interval from $[0, \frac{1}{\pi}]$:



Additionally, note that the derivative of this function is:

$$\frac{d}{dx}f(x) = \cos(\frac{1}{x}) + \frac{\sin(\frac{1}{x})}{x}$$

This derivative is not bounded on the interval $[0, \frac{1}{\pi}]$ - rather, the values blow up and diverge at infinity.

This result seems rather unfortunate. If a function does not have bounded variation, is there any other type of variation that we can use? The answer lies with quadratic variation.

Definition 4.4. The **quadratic variation** of a function f on the interval $[a, b]$ over some partition $\Pi = \{a = x_0, x_1, ..., x_n = b\}$ is defined by the following.

$$Q_\Pi(f) = \sum |f(x_i) - f(x_{i-1})|^2$$

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

This definition looks very similar to how we defined variation before, but now the difference term is squared. As it turns out, this is very useful. After seeing these definitions, one might wonder what connection there is (if any) between having bounded variation and quadratic variation.

Proposition 4.5. *A differentiable function whose derivative is bounded on the interval $[a, b]$ has a quadratic variation that is equal to zero.*

Proof. Let f be a differentiable function, and consider its quadratic variation:

$$Q(f) = \lim_{\Pi \to 0} \sum_{i=1}^{\infty} |f(x_i) - f(x_{i-1})|^2$$

Recall that the **Mean Value Theorem** states that for any function g that is continuous on the closed interval $[a, b]$, and differentiable on the interval (a, b), then for some $m \in [a, b]$, the following equation holds:

$$g(b) - g(a) = g'(m)(b - a) \tag{4.6}$$

Applying the mean value theorem to the difference inside the absolute value, for some $x_i^* \in [x_i, x_{i-1}]$, we have that:

$$f(x_i) - f(x_{i-1}) = f'(x_i^*)(x_i - x_{i-1})$$

We can plug this in to the equation for quadratic variation, and see that:

$$Q(f) = \lim_{\Pi \to 0} \sum_{i=1}^{\infty} |f'(x_i^*)(x_i - x_{i-1})|^2 = \lim_{\Pi \to 0} \sum_{i=1}^{\infty} |f'(x_i^*)|^2 |x_i - x_{i-1}|^2$$

Note that for all x_n, $|x_i - x_{i-1}| \leq |\Pi|$, since to take the quadratic variation of f, we are considering the finest possible partitions on the interval $[a, b]$. This gives us the following inequality:

$$Q(f) = \lim_{\Pi \to 0} \sum_{i=1}^{\infty} |f'(x_i^*)|^2 |x_i - x_{i-1}|^2 \leq \lim_{\Pi \to 0} |\Pi| \sum_{i=1}^{\infty} |f'(x_i^*)|^2 |x_i - x_{i-1}|$$

We see that the expression inside of the sum is equivalent to the following integral:

$$\lim_{\Pi \to 0} \sum_{i=1}^{\infty} |f'(x_i^*)|^2 |x_i - x_{i-1}| = \int_a^b |f'(x)|^2 dx$$

This integral is finite over a bounded interval. The product of the integral with the size of the partitions (as $|\Pi|$ tends to zero) will also be zero:

$$\lim_{|\Pi| \to 0} |\Pi| \int_a^b |f'(x)|^2 dx \to 0 \implies Q(f) = 0$$

Since f was any arbitrary differentiable function with bounded derivative, we are done. □

We can apply a similar logic as before to show that the following proposition is true:

Proposition 4.7. *A differentiable function with a bounded derivative on the interval $[a, b]$ has bounded variation.*

Proof.

$$V(f) = \lim_{|\Pi| \to 0} \sum_{i=1}^{\infty} |f(x_i) - f(x_{i-1})| = \lim_{|\Pi| \to 0} \sum_{i=1}^{\infty} |f'(x_i^*)||x_i - x_{i-1}|$$

Again, as $|\Pi| \to 0$, we are left with an integral:

$$V(f) = \int_a^b |f'(x)| dx$$

The variation is bounded only when this integral is finite; and for this to be true, the derivative must be bounded and continuous on $[a, b]$ (except for on a set of points of measure zero). □

And finally, from the above results, the following proposition should make sense.

Proposition 4.8. *A function that has bounded variation has a quadratic variation that is equal to zero. Furthermore, we generalize this to higher dimensions. We can define variation in higher dimensions d, i.e., $V^{(d)}(f)$:*

$$V_{\Pi}^{(d)}(f) = \sum |f(x_i) - f(x_{i-1})|^p$$

$$V^{(d)}(f) = \lim_{|\Pi| \to 0} V_{\Pi}(f)$$

If the d^{th} dimensional variation of a function f is bounded, the variation in any higher dimensions is equal to zero.

We see that it is very important for a function to be differentiable, in order for it to have bounded variation. For example, as we calculated for the example above, the derivative of $x(cos(\frac{1}{x})$ diverges at zero, and it would not have a finite integral on the range $[0, \frac{1}{\pi}]$. The results of our previous calculation do make sense.

We see that for nice functions, variation is not very difficult (or that interesting) to calculate, but when we run into functions which are not differentiable in the ways described above, the variation results are much more intriguing.

Re-enter Brownian motion...

5. Variation of Brownian Motion

Now that we have introduced Brownian motion and variation, the next question that one might want to ask is what we can say about the variation of Brownian motion? Recalling what Brownian motion looks like, it is clear that it is extremely "wobbly"! And, since we know that Brownian motion is nowhere differentiable with probability 1, we should anticipate that it probably does not have bounded variation.

Theorem 5.1. *The variation of a Brownian motion path does not converge, with probability one; in other words, the following equation diverges:*

$$\lim_{n\to\infty} \sum_{i=1}^{2^n} |B(\frac{kt}{2^n}) - B(\frac{(k-1)t}{2^n})|$$

Proof. Note that the following inequality is true:

$$\sum_{i=1}^{2^n} |B(\frac{kt}{2^n}) - B(\frac{(k-1)t}{2^n})| \geq \frac{\sum_{i=1}^{2^n} |B(\frac{kt}{2^n}) - B(\frac{(k-1)t}{2^n})|^2}{\max_{n=0,1,2,...} |B(\frac{kt}{2^n}) - B(\frac{(k-1)t}{2^n})|}$$

Consider the numerator and denominator separately. In the numerator, this is the quadratic variation of Brownian motion: After this, we will show that the quadratic variation of Brownian motion is t. Meanwhile, the denominator will converge to zero, because as the time increments indexing Brownian motion become arbitrarily close, the distance between also converges to zero. So, Brownian motion does not have bounded variation. □

Definition 5.2. The quadratic variation of Brownian motion can be defined using the following:

$$\lim_{n\to\infty} \mathbb{E}[\sum_{i=1}^{2^n} |B(\frac{kt}{2^n}) - B(\frac{(k-1)t}{2^n})|^2] = t$$

Proposition 5.3. *The quadratic variation of standard Brownian motion on the interval $[a, b]$ is $a - b$.*

Proof. To prove this, for the sake of simplicity and the sanity of the reader, we will only consider the case on the interval $[0, t]$, and prove that the quadratic variation L^2 converges to t. In other words, we want to prove that

$$\mathbb{E}[(Q(B) - t)^2] = 0 \tag{5.4}$$

Note that for some partition $\{t_1, t_2, ..., t_n\}$, we can express t with the following sum:

$$t = \sum_{i=1}^{n} t_i - t_{i-1}$$

We also will use the following notation:

$$B_{t_i} - B_{t_{i-1}} = \Delta B_i$$

By definition of Brownian motion, we know that $\Delta B_i \sim \text{Norm}(0, t_i - t_{i-1})$. We have the following inequality:

$$Q(B) - t = \sum_i (\Delta B_i)^2 - (t_i - t_{i-1})$$

We want the expectation of the square of $Q(B) - t$ to converge to zero:

$$\mathbb{E}[(Q(B) - t)^2] = \sum_i \mathbb{E}[((\Delta B_i)^2 - (t_i - t_{i-1}))^2]$$

First, we square the terms inside of the expectation, and simplify. To simplify, separate the sum using linearity of expectation, and the independence of $(\Delta B_i)^2$ and $(t_i - t_{i-1})$:

$$= \sum_i \mathbb{E}[((\Delta B_i)^4 - 2(\Delta B_i)^2(t_i - t_{i-1}) + (t_i - t_{i-1})^2]$$

$$= \sum_i \mathbb{E}[((\Delta B_i)^4] - 2\sum_i \mathbb{E}[(\Delta B_i)^2](t_i - t_{i-1}) + \sum_i (t_i - t_{i-1})^2$$

Note that we are able to remove $(t_i - t_{i-1})$ from inside of the expectation, since $(t_i - t_{i-1})$ is not random. To simplify this further, remember that (ΔB_i) is normally distributed. In probability, we can find higher powers of expectation with moment generating functions. In particular, for the normal distribution, where $X \sim \text{Norm}(0, \sigma^2)$, we have that $\mathbb{E}[X^2] = \sigma^2$ and $\mathbb{E}[X^4] = 3\sigma^4$. Using this fact, we can simplify the sum more:

$$= \sum_i 3(t_i - t_{i-1})^2 - 2\sum_i (t_i - t_{i-1})(t_i - t_{i-1}) + \sum_i (t_i - t_{i-1})^2 = 2\sum_i (t_i - t_{i-1})^2$$

Consider the supremum Π^* of distances between points $(t_i - t_{i-1})$ of the partition Π. We have the following inequality (because we are partitioning a finite interval):

$$= 2\sum_i (t_i - t_{i-1})^2 \leq 2\Pi^* \sum_i (t_i - t_{i-1}) = 2\Pi^* t$$

To find the quadratic variation, we must take the limit as the length of the distances between the points in the partitions (for any partition Π) approaches zero. In particular, Π^* will approach zero, meaning that this product also converges to zero:

$$\lim_{|\Pi|\to 0} 2\Pi^* t = 0$$

And, clearly, this bound means that:

$$\mathbb{E}[(Q(B) - t)^2] = 0$$

This shows L^2 convergence. L^2 convergence also implies that the quadratic variation of standard Brownian motion converges to t with probability 1; this will directly follow from a variation on Chebyshev's Inequality:

Theorem 5.5. ***(Chebyshev's Inequality)*** *For a random variable X, for any $\epsilon > 0$:*

$$P[(X - k) \geq \epsilon] \leq \frac{\mathbb{E}[(X-k)^2]}{\epsilon^2}$$

Intuitively, Chebyshev's Inequality is a way to use standard deviation to bound the probability that a random variable will deviate beyond a certain radius of the mean. (Chebyshev's inequality is standard fare in most probability courses, and will not be proven here).

Applying this, it is clear that

$$\mathbf{P}(|Q(B) - t| > \epsilon) \leq \frac{\mathbb{E}[(Q(B)-k)^2]}{\epsilon^2} \to 0$$

This tells us that the quadratic variation of standard Brownian motion converges in probability to t. (Converging in probability is still slightly weaker than saying that it converges with probability 1, but this can be proven using the Borel-Cantelli Lemma, which is not explained in this paper: This is proven in Lawler [1] pp. 62-63).

□

Now that we have looked at the variation of Brownian motion, one might ask what the point of all of this is? As it turns out, we can use quadratic variation as a way to motivate learning about Ito's Formula.

6. Ito's Formula

Ito's Formula is at the root of stochastic calculus, and Brownian motion plays a key part. However, previously, we have shown that Brownian motion is nowhere differentiable, with probability 1. For most of us, when we started learning calculus, it was all about derivatives and integrals...so, if Brownian motion is not differentiable, how can we use it at all?

Our goal is to construct and understand integrals that have the following form:

$$\int_0^t X_s dB_s$$

First, we recall how we have defined integration in the past.

Definition 6.1. To take the **Riemann integral** of a smooth function $f(t)$ on $[0,k]$, we partition the interval $[0,k]$ into a set of $\{t_0 = 0, t_1, t_2, ..., t_n = k\}$, and approximate $f(t)$ by selecting $f(t_i^*)$ where $t_{i-1} \le t_i^* \le t_i$. The integral is found using the following sum:

$$\int_0^k f(t)dt = \lim_{n\to\infty} \sum_{i=1}^{n} f(t_i^*)(t_i - t_{i-1})$$

Definition 6.2. The **Riemann-Stieljes integral** integrates a function $f(t)$ with respect to another function $h(t)$. It is defined using the following sum:

$$\int_0^k f(t)dh(t) = \lim_{n\to\infty} \sum_{i=1}^{n} f(t_i^*)[h(t_i) - h(t_{i-1})]$$

Theorem 6.3. *If $f(t)$ is a continuous function, and $g(t)$ has bounded variation, then this integral exists.*

Note that if $h(t) = t$, this brings us back to the traditional Riemann sum.

At first glance, this looks like a definition that would be convenient to use when defining an integral with respect to Brownian motion. However, in order converge to a bounded integral, $h(t)$ must have bounded variation! And, we have seen that Brownian motion does not, so this approach must be modified.

Definition 6.4. Let X_t be a stochastic process defined for $0 \le t \le T$. X_t is a **simple adapted process** if it takes the following form:

$$X_t = \sum_{i=0}^{n} A_i \mathbb{I}_{(t_i, t_{i+1}]}(t)$$

Here, $0 = t_0 < t_1 < t_2 < .. < t_n$, and the A_i's are stochastic. The notation $\mathbb{I}$ denotes an indicator function. This means that X_t can be approximated by a step function.

Definition 6.5. The **Ito Integral** of a simple adapted process $X(t)$ is defined using the following formula:

$$I(T) = \int_0^T X(t)dB_t = \sum_{i=1}^{n} A_i(B_{t_{i+1}} - B_{t_i})$$

Proposition 6.6. ***Ito isometry*** *is a property of the Ito integral:*

$$\mathbb{E}[(\int_0^T XdB)^2] = \mathbb{E}[\int_0^T X^2 dt] \tag{6.7}$$

Proof. First, we know that these following properties are true:

(1) X is a simple process adapted to a filtration from 0 to T.
(2) $B_t - B_s$ for $t > s$ are independent (by previous property of Brownian motion). This means that $\Delta B_i = B_{t_{i+1}} - B_{t_i}$ is normally distributed.

We expand the left side of the equation:

$$\mathbb{E}[(\int_0^T XdB)^2] = \mathbb{E}[(\sum_i A_i \Delta B_i)^2] = \mathbb{E}[\sum_i \sum_j A_i A_j \Delta B_i \Delta B_j]$$

First, split up the terms of the sum based on their index:

$$= \mathbb{E}[\sum_{i \neq j} A_i A_j \Delta B_i \Delta B_j + \sum_{i=j} A_i^2 (\Delta B_i)^2]$$

Using independence of Brownian motion variables, and linearity of expectation, we can simplify these sums further. This is because $A_i A_j \Delta B_i$ is independent of ΔB_j.

$$= \sum_{i \neq j} \mathbb{E}[A_i A_j \Delta B_i]\mathbb{E}[\Delta B_j] + \sum_{i=j} \mathbb{E}[A_i^2 (\Delta B_i)^2]$$

And, because ΔB_j is normally distributed, its expectation is zero. We can eliminate this part of the sum. That leaves us with:

$$= \sum_{i=j} \mathbb{E}[A_i^2 (\Delta B_i)^2] = \sum_{i=j} \mathbb{E}[A_i^2]\mathbb{E}[(\Delta B_i)^2]$$

But, we know that the variance of ΔB_i is Δt (by definition of the normal distribution), and can be expressed by the following equation:

$$\Delta t = \text{Var}(\Delta B_i) = \mathbb{E}[(\Delta B_i)^2] - \mathbb{E}[(\Delta B_i)]^2 = \mathbb{E}[(\Delta B_i)^2] - 0 \implies \Delta t = \mathbb{E}[(\Delta B_i)^2]$$

We can substitute this into the above equation:

$$= \sum_{i=j} \mathbb{E}[A_i^2]\Delta t = \mathbb{E}[\sum_{i=j} A_i^2 \Delta t]$$

Note that Δt is not random - rather, it is a constant. By linearity of expectation, the above modification of the sum is allowed, and puts the equation into a (hopefully!) familiar form: This is just equal to the integral of a simple adapted process.

$$= \mathbb{E}[\int_0^T X^2 dt]$$

This gives us the right side of the equation, and we are done. □

Looking at the Ito Integral formula, this seems like exactly how we would want stochastic integration to be defined. The Ito Isometry seems like a nice property, and, additionally, it seems very similar to other integration formulas that we have seen. But how does it work in practice?

Example 6.8. To practice using the Ito Integral, consider $X(t) = B_t$. We want to integrate the following:

$$\int_0^T B_t dB_t$$

Using the formula for the Ito integral, we use $A_i = B_{t_i}$ to approximate the values for B_t.

We have:

$$\int_0^T B_t dB_t = \sum_{i=1}^{n} B_{t_i}(B_{t_{i+1}} - B_{t_i})$$

Note that:

$$B_{t_i}(B_{t_{i+1}} - B_{t_i}) = -\frac{1}{2}B_{t_i}^2 + \frac{1}{2}B_{t_{i+1}}^2 - \frac{1}{2}B_{t_i}^2 + B_{t_i}B_{t_{i+1}} - \frac{1}{2}B_{t_{i+1}}^2$$
$$= \frac{1}{2}(B_{t_{i+1}}^2 - B_{t_i}^2) - \frac{1}{2}(B_{t_{i+1}} - B_{t_i})^2$$

Substituting back into our initial sum and separating gives us:

$$= \frac{1}{2}\sum_{i=1}^{n}(B_{t_{i+1}}^2 - B_{t_i}^2) + \frac{1}{2}\sum_{i=1}^{n}(B_{t_{i+1}} - B_{t_i})^2$$

Clearly, the first sum simplifies to B_t^2. And, the second sum looks exactly like the formula for quadratic variation, which we know is t. This leaves us with the following:

$$\int_0^T B_t dB_t = \frac{1}{2}B_t^2 - \frac{t}{2}$$

Theorem 6.9. ***Ito's Formula***. *Let f be a function, with $f \in C^2$, and B_t denote Brownian motion. Then:*

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

Equivalently, ***Ito's Formula*** *can be expressed in integral form:*

$$f(B_t) = f(0) + \int_0^t f'(B_s)dB_s + \frac{1}{2}\int_0^t f''(B_s)ds$$

Example 6.10. From our practice with the Ito integral before, we can check that Ito's formula gives us the same result for $\int B_s dB_s$.

Consider $f(x) = x^2$. We know that $\frac{d}{dx}f = 2x$, and $\frac{d^2}{dx^2}f = 2$. Plugging in to the formula, we see that:

$$B_t^2 = B_0 + \int_0^t 2B_s dB_s + \frac{1}{2}\int_0^t 2ds = \int_0^t 2B_s dB_s + t$$

Rearranging terms, we get the same result as before:

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

Additionally, note that if we were to naively integrate $\int B_t dB_t$ in the manner in which we conduct standard, non-stochastic integrals, we would expect our result to be $\frac{B_t^2}{2}$. And this is almost the result that we get; but, there is the extra t term - which is the quadratic variation. This is like a "correcting" term that accounts for the amount of randomness that has been added up until time t.

Proof. Here, we will show an informal proof of Ito's Formula. (To see a full proof, see [1] Lawler pp. 98-100). The previous example tells us that quadratic variation is an important part of defining the stochastic integral, and it shows us how the formula works in practice. One might wonder where Ito's formula actually comes from? It should remind us of how we have defined integrals in the past, with equations like these:

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

We can get this equation by expanding a Taylor series expansion of degree 1 (as the higher order terms would correspond to $(dt)^n$ which just approach zero).

Ito's formula looks a lot like this, except there is an additional integral (which gives us the term that corresponds to quadratic variation). Instead, we consider a second order Taylor expansion:

$$f(x+a) = f(x) + f'(x)a + \frac{1}{2}a^2 f''(x) + R(x,a)$$

We know that the end behavior of $R(x,a)$ is $o(a^2)/a^2$, which approaches 0 as a approaches 0.

First, we construct a telescoping sum for $f(B_1) - f(B_0)$:

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

Also, note that we can approximate this difference using Taylor approximation:

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

Taking the sum over all i of the three above terms will give us the difference $f(B_1) - f(B_0)$. Consider each of the three summations separately. Consider the first summation:

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

For this term, this looks like how we defined a stochastic integral, if the process that we were dealing with was $f'(B_s)$. This term will give us:

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

Next, we have:

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

As we saw before, when integrating with respect to B_t^2, $f''(B_{(i-1)/n})$ was constant. This term is just t, which is the quadratic variation.

In general, let $s(t) = f''(B_t)$, which is continuous.

For some small increment δ, we can approximate $s(t)$ with the step function $s_\delta(t)$ such that $|s(t) - s_\delta(t)| < \delta$. Let us first consider the the interval where s_δ is constant. In this case, the integral is just:

$$\lim_{n\to\infty} \sum_{i=1}^{n} \frac{1}{2} s_\delta (B_{i/n} - B_{(i-1)/n})^2 = \int_0^1 s_\delta dt$$

Also, note that we chose δ and s_δ such that $|s(t) - s_\delta(t)| < \delta$. We can use these to create the following bound:

$$|\sum_{i=1}^{n}[s(t)-s_\delta(t)](B_{i/n}-B_{(i-1)/n})^2| \leq \delta\sum_{i=1}^{n}(B_{i/n}-B_{(i-1)/n})^2$$

As $n \to \infty$,

$$\sum_{i=1}^{n}(B_{i/n}-B_{(i-1)/n})^2 \to \int_0^1 dt = 1$$

$$\delta\sum_{i=1}^{n}(B_{i/n}-B_{(i-1)/n})^2 \to \delta$$

This means that as $\delta \to 0$, $s(t)$ and $s_\delta(t)$ become arbitrarily close to $s(t)$. We can use the integral that we found above:

$$\lim_{\delta\to 0}\frac{1}{2}\int_0^1 s_\delta(t)dt \sim \frac{1}{2}\int_0^1 s(t)dt = \frac{1}{2}\int_0^1 f''(B_t)dt$$

Thus, for the second part of the sum, we have that:

$$\lim_{n\to\infty}\sum_{i=1}^{n}\frac{1}{2}f''(B_{(i-1)/n})(B_{i/n}-B_{(i-1)/n})^2 = \frac{1}{2}\int_0^1 f''(B_t)dt$$

Finally, we have one more sum to consider:

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

But, these terms in the sum will just converge to $\frac{i}{n}-\frac{i-1}{n}=\frac{1}{n}$. As $n \to 0$, this sum is adding up n terms with an order that is less than $\frac{1}{n}$, this remainder term goes to zero.

This leaves us with the sum that we would anticipate:

$$f(B_1)-f(B_0) = \int_0^1 f'(B_t)dB_t + \frac{1}{2}\int_0^1 f''(B_t)dt$$

This is what we wanted to show, and tells us where Ito's Formula comes from.

Additionally, we know that if we tried to add more terms on to this Taylor series expansion, then we would be adding terms of the following form (for $m > 2$):

$$\sum_{i=1}^{n}\frac{1}{m!}f^{(m)}(B_{(i-1)/n})(B_{i/n}-B_{(i-1)/n})^m$$

But, the terms inside of this sum are just higher order variation terms. As we saw before, if the d^{th} dimensional variation of a function f is bounded, the variation in any higher dimensions is equal to zero. And, we know that the quadratic variation of Brownian motion is bounded, meaning that these higher degree variation terms will just be zero, and contribute nothing to the formula.

□

Based on what we have seen above, using quadratic variation, we can account for the inclusion of randomness to integrate stochastic processes; this is calculated using Ito's formula. This in turn allows for us to make sense of stochastic differential equations (with drift m and volatility σ), which take the form:

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

Stochastic differential equations are a topic for another time. However, we can still see the relationship between equations like this and the equations from utilizing Ito's Formula: To express a differential equation, it involves the change corresponding to time (the drift) in addition to the changes corresponding to the randomness of Brownian motion (the volatility). This demonstrates just one of the many mathematical arenas where defining and understanding the properties of Brownian motion is key.

Acknowledgments

I am grateful to Professor Greg Lawler for his lectures, and for suggesting this topic. I appreciate the helpful discussions with my mentor Drew Moore. Many thanks to Professor Peter May, for organizing the REU and for giving me this opportunity.

References

[1] Lawler, Gregory. *Stochastic Calculus: An Introduction with Applications.* https://www.math.uchicago.edu/ lawler/finbook2.pdf

[2] Lawler, Gregory. *Random Walk and the Heat Equation* http://www.math.uchicago.edu/ lawler/reu.pdf

[3] Feldman, Joel. *Functions of Bounded Variation.* http://www.math.ubc.ca/ feldman/m321/variation.pdf

[4] Dunbar, Stephen. *Stochastic Processes and Advanced Mathematical Finance.* https://www.math.unl.edu/ sdunbar1/MathematicalFinance/Lessons/BrownianMotion/QuadraticVariation/quadraticvariation.

[5] Allaart, Pieter. *The Stochastic Integral.* http://www.math.unt.edu/ allaart/6810/6810-notes-chap4.pdf.

[6] Watkins, Joseph. *Convergence Concepts.* https://www.math.arizona.edu/ jwatkins/s-convergence.pdf.

[7] Sigman, Karl. *Functions of bounded variation.* http://www.columbia.edu/ ks20/6712-14/6712-14-Notes-BMII.pdf.

[8] Lawler, Gregory. *Notes on Probability.* https://www.math.uchicago.edu/ lawler/probnotes.pdf.

[9] Silvia, Evelyn. *Riemann Stieljes Integration.* https://www.math.ucdavis.edu/ emsilvia/math127/chapter7.pdf.

[10] Weisstein, Eric W. *Borel Set.* From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/BorelSet.html