EXISTENCE AND PROPERTIES OF BROWNIAN MOTION

HEATHER WANG

Abstract. In the physical sciences, Brownian motion is the random movement of particles suspended in a fluid. In this paper, we define a Brownian motion mathematically and provide an explicit construction of such a process through Lévy's construction. We then prove key properties of Brownian motion like the Markov property, Strong Markov property, and Martingale property. We finish with an analysis of the linear and quadratic variation of Brownian motion and a glimpse into stochastic calculus.

1. Introduction

In the physical world, Brownian motion is the random movement and tiny fluctuations of microscopic particles suspended in a fluid. It derives its name from Scottish botanist Robert Brown, who observed it in action while studying pollen particles under a microscope. By analyzing the molecules of inorganic materials like minerals, Brown verified that this random motion did not stem from a property of living organisms. Later, in 1905, Albert Einstein published a paper in which he derived the existence of Brownian motion from the constant collisions between atoms at a time when the atomic theory was still contentious. He developed a diffusion model to model the process as a random walk, and he proved the equivalence between macroscopic diffusion and Brownian motion. French physicist Jean Perrin's experiments later verified Einstein's theoretical model, winning the former a Nobel Prize in 1926 [Nel01].

The first person to rigorously examine Brownian motion through a mathematical lens was American mathematician Norbert Wiener in 1923, leading Brownian motion to also be called the Wiener process in mathematics [Wie23]. In this paper, we will stick to the name "Brownian motion." Later, mathematicians like Paul Lévy and Kiyosi Itô continued developing the mathematical theory of Brownian motion, providing an explicit construction and creating a method for integration with respect to Brownian motion respectively. Brownian motion is the foundation for stochastic calculus and is relevant in many fields like quantitative finance, physics, biology, and more.

In this paper, we begin by introducing some definitions and concepts in random variables and measure theory in Section 2 that will allow us to rigorously examine the properties of Brownian motion. Once we have built our foundation, we will define a one-dimensional Brownian motion, provide motivation for this definition, and explore basic properties like translation invariance and scaling invariance in Section 3. Next, in Sections 4 and 5, we will analyze the Markov and martingale properties respectively of Brownian motion and include generalizations to a class of random times called stopping times. In Section 6, we provide an explicit construction of Brownian motion through Lévy's construction, proving Wiener's

Date: July 13, 2026.

Theorem in 1923 that a process satisfying the properties of Brownian motion actually exists. Finally, we explore the linear and quadratic variation of Brownian motion and their implication on stochastic calculus in Section 7.

2. Preliminaries

Before we dive into Brownian motion and its properties, we have a few concepts we must introduce first. In this paper, we will stick to concepts explicitly used in defining Brownian motion and its properties. For a more comprehensive overview of probability and random variables, refer to [Ros09].

Definition 2.1. The *expected value* (or mean) of a continuous random variable X with density function $f(x)$ is defined as

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x)\, dx.$$

Definition 2.2. The *variance* of a random variable X with mean $\mathbb{E}[X] = \mu$ is

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

Definition 2.3. Two continuous random variables X and Y are *independent* if their probability density functions $f_X(x)$ and $f_Y(y)$ respectively satisfy

$$f(x,y) = f_X(x) f_Y(y)$$

for all real values of x and y, where $f(x,y)$ is the joint probability density function of X and Y (i.e. the probability that X equals x and Y equals y).

Now for something more directly connected to Brownian motion.

Definition 2.4. A continuous random variable X is *normal* with parameters μ, σ^2 if it has the following probability density function:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}.$$

We can also write $X \sim \mathcal{N}(\mu, \sigma^2)$ to indicate that X has a normal distribution. If $\mu = 0$ and $\sigma^2 = 1$, then X is called *standard normal.*

Remark 2.5. A normal distribution is also called a *Gaussian* distribution.

Normal distributions are just the standard bell-curve. In our definition above, μ and σ^2 are the mean and variance respectively of X.

Exercise. Verify this.

Normal distributions also have the following nice property:

Proposition 2.6. *Given two independent random variables $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$, we have*

$$X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

Proof. Let $Z = X + Y$. By the convolution formula, we have

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\,dx \\ &= \frac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{\infty} \exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2} - \frac{(z-x-\mu_2)^2}{2\sigma_2^2}\right) dx \\ &= \frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}} \exp\left(-\frac{(z-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}\right) \int_{-\infty}^{\infty} \frac{\sqrt{\sigma_1^2+\sigma_2^2}}{\sigma_1\sigma_2\sqrt{2\pi}} \exp\left(-\frac{(x-c)^2}{\frac{2\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}}\right) dx \\ &= \frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}} \exp\left(-\frac{(z-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}\right), \end{aligned}$$

where $c = \frac{\sigma_2^2\mu_1 + \sigma_1^2(z-\mu_2)^2}{\sigma_1^2+\sigma_2^2}$. Thus, $Z \sim \mathcal{N}(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$. ■

In order to rigorously tackle Brownian motion and other topics in probability, we need a way to filter out pathological sets, as trying to assign a probability to every set leads to paradoxes because of the existence of non-measurable sets. Measure theory and σ-algebras allow us to only look at the nice sets so that our math stays consistent.

Definition 2.7. Given a set $\mathcal{X}$, a collection $\mathscr{C}$ of subsets of $\mathcal{X}$ is a σ-algebra if it satisfies the following properties:

- $\mathcal{X} \in \mathscr{C}$
- $\mathscr{C}$ is closed under complements, i.e. for a subset A of $\mathcal{X}$, $A \in \mathscr{C}$ implies $\mathcal{X} \backslash A \in \mathscr{C}$
- $\mathscr{C}$ is closed under countable unions, i.e. if subsets A_1, A_2, ... of $\mathcal{X}$ are in $\mathscr{C}$, then their union is also in $\mathscr{C}$

We call the ordered pair $(\mathcal{X}, \mathscr{C})$ a measurable space. For completeness, we'll also define a measure.

Definition 2.8. Given a measurable space $(\mathcal{X}, \mathscr{C})$, a measure λ is a function $\lambda : \mathscr{C} \to \mathbb{R} \cup \{\infty\}$ with the following properties:

- non-negativity: $\lambda(A) \geq 0 \ \forall \ A \in \mathscr{C}$
- $\lambda(\emptyset) = 0$
- countable additivity: if A_1, A_2, ... are disjoint elements of $\mathscr{C}$, then

$$\lambda\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \lambda(A_i)$$

In the case that $\lambda(\mathcal{X}) = 1$, then λ is a probability measure, and $(\mathcal{X}, \mathscr{C}, \lambda)$ is a probability space. Such a space is typically denoted as $(\Omega, \mathcal{F}, \mathbb{P})$.

The following results are useful for proving an event occurs with probability zero or one. For more on measure theory and the Borel-Cantelli Lemmas, see [Bil95].

Lemma 2.9 (First Borel-Cantelli Lemma: Convergence)**.** *Given a sequence of events $\{A_n\}_{n\geq 1}$, if the sum of their probabilities $\sum_{n=1}^{\infty} \mathbb{P}(A_n)$ converges, then $\mathbb{P}(\limsup_{n\to\infty} A_n) = 0$.*

Here, $\limsup_{n\to\infty} A_n$ denotes the event that infinitely many of $\{A_n\}_{n\geq 1}$ occur. More formally, define

$$\limsup_{n\to\infty} A_n := \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k.$$

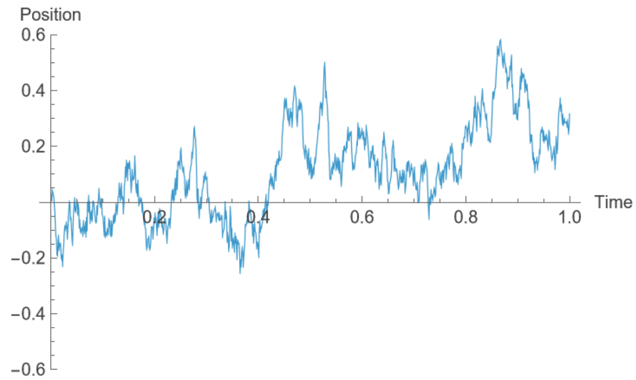

Figure 1. Sample Brownian motion path in 1-D.

Proof of Lemma 2.9. Since $\limsup_{n\to\infty} A_n \subset \bigcup_{k=i}^{\infty} A_k$, we have

$$\mathbb{P}(\limsup_{n\to\infty} A_n) \le \mathbb{P}\left(\bigcup_{k=i}^{\infty} A_k\right) \le \sum_{k=i}^{\infty} \mathbb{P}(A_k).$$

Taking the limit as i approaches ∞ and using the fact that $\sum_{n=1}^{\infty} \mathbb{P}(A_n)$ converges gives us the desired result. ■

Lemma 2.10 (Second Borel-Cantelli Lemma: Divergence)**.** *Given a sequence of independent events* $\{A_n\}_{n\ge1}$*, if the sum of their probabilities* $\sum_{n=1}^{\infty} \mathbb{P}(A_n)$ *diverges, then* $\mathbb{P}(\limsup_{n\to\infty} A_n) = 1$*.*

Proof. Showing $\limsup_{n\to\infty} A_n$ occurs with probability 1 is equivalent to showing that its complement $\left(\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k\right)^{\mathsf{C}} = \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k^{\mathsf{C}}$ occurs with probability 0. Using the inequality $1-x \le e^{-x}$ and the fact the A_n's are independent, we have

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k^{\mathsf{C}}\right) \le \sum_{n=1}^{\infty} \mathbb{P}\left(\bigcap_{k=n}^{\infty} A_k^{\mathsf{C}}\right) = \sum_{n=1}^{\infty}\prod_{k=n}^{\infty}(1-\mathbb{P}(A_k)) \le \sum_{n=1}^{\infty} e^{-\sum_{k=n}^{\infty}\mathbb{P}(A_k)}.$$

Since $\sum_{n=1}^{\infty} \mathbb{P}(A_n)$ diverges, the rightmost expression equals 0. ■

3. Definition and Basic Properties

We are now ready to define Brownian motion.

Definition 3.1. A *one-dimensional Brownian motion* $\{B(t) : t \ge 0\}$ is a stochastic process with the following properties:

(1) for all times $0 \le t_1 \le t_2 \le \ldots \le t_n$, the increments $B(t_{i+1}) - B(t_i)$ for $1 \le i < n$ are independent random variables (this is also called the property of independent increments),
(2) increments $B(t+h) - B(t)$ have a normal distribution with mean zero and variance h,
(3) the function $t \mapsto B(t)$ is continuous with probability 1.

If $B(0) = 0$, then $\{B(t)\}_{t\ge0}$ is called *standard Brownian motion.*

Why do we require these three properties? Recall the physical definition of Brownian motion as the random movement of a particle suspended in a fluid. Clearly, we would want

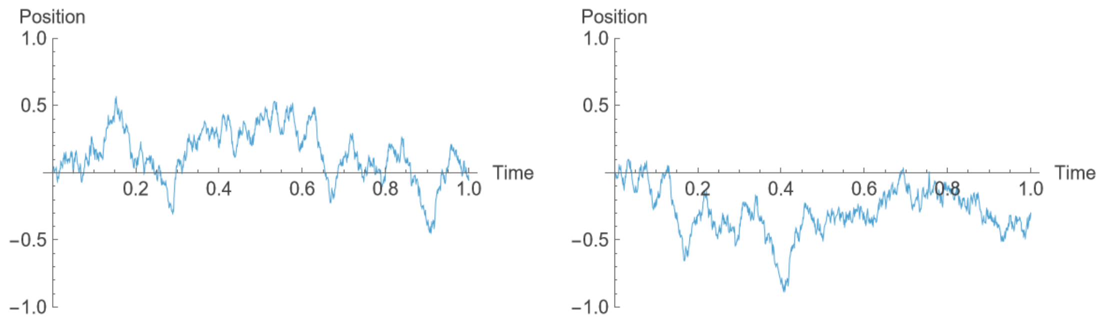

Figure 2. Sample $B(t)$ (left) and $B(t+0.5)-B(0.5)$ (right).

such a path to be continuous, motivating property 3 of Definition 3.1. In one dimension, we may expect that the path of a particle would mirror a symmetric random walk. Formally, define independent Rademacher random variables $\varepsilon_1, \varepsilon_2, \ldots$ with distribution

$$\mathbb{P}\{\varepsilon_i = 1\} = \mathbb{P}\{\varepsilon_i = -1\} = \tfrac{1}{2}$$

for all $i \in \mathbb{Z}_{>0}$, and let $S_n = \sum_{i=1}^{n} \varepsilon_i$. The Central Limit Theorem states that for independent and identically distributed random variables $X_1, X_2, \ldots$ with mean μ and variance σ^2,

$$\lim_{n\to\infty} \mathbb{P}\left\{\frac{X_1 + \ldots + X_n - n\mu}{\sigma\sqrt{n}} \leq a\right\} = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx.$$

Thus, the increments of the scaled random walk $S_n/\sqrt{n}$ approach a normal distribution as n approaches ∞, motivating property 2. Finally, the property of independent increments clearly holds for a symmetric random walk, motivating property 1. In a sense, Brownian motion is like a continuous generalization of the symmetric random walk. A sample path of a standard Brownian motion is shown in Figure 1.

With just this definition, we can already prove some basic properties of Brownian motion.

Proposition 3.2 (Translation Invariance)**.** *If $B(t)$ is a Brownian motion, then $\forall h \in \mathbb{R}_{\geq 0}$, so is $B(t+h) - B(h)$.*

Proof. Let $B'(t) = B(t+h) - B(h)$. Note that for all times $0 \leq t_1 \leq t_2 \leq \ldots \leq t_n$, we have $B'(t_i) = B(t_i + h) - B(h)$ for $1 \leq i \leq n$. Calculating the increments of $B'(t)$ in this time series, we get

$$B'(t_{i+1}) - B'(t_i) = B(t_{i+1} + h) - B(t_i + h)$$

for $1 \leq i < n$. Since increments of B are independent on $0 \leq t_1 + h \leq t_2 + h \leq \ldots \leq t_n + h$ and are normally distributed with mean 0 and variance $t_{i+1} - t_i$, the increments of B' are independent and normally distributed with the same means and variances on $0 \leq t_1 \leq t_2 \leq \ldots \leq t_n$ too. Furthermore, shifting t doesn't change continuity, so B' is a Brownian motion. ■

We can see an example of translation invariance in action in Figure 2.

Proposition 3.3 (Scaling Invariance)**.** *If B is a Brownian motion, then so is $kB(t/k^2)$ for all nonzero real k.*

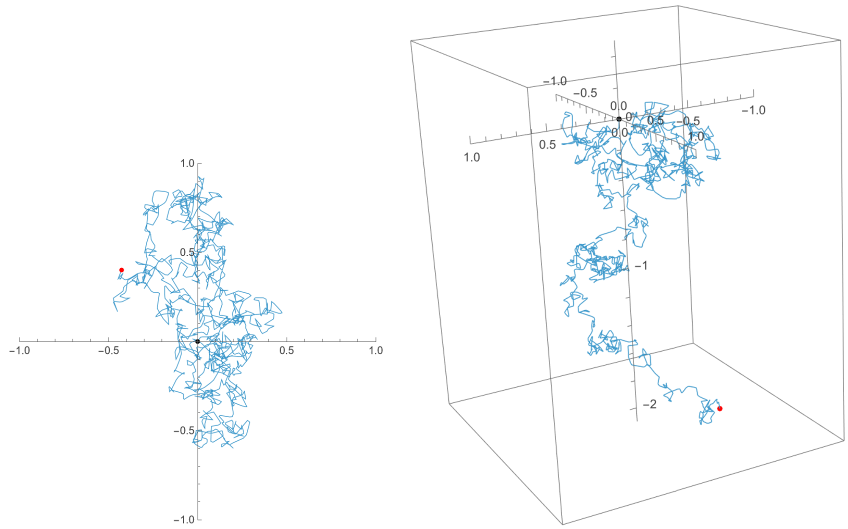

Figure 3. Sample Brownian motion paths in 2-D (left) and 3-D (right).

Proof. Again, let $B'(t) = kB(t/k^2)$. Scaling preserves continuity, and independent increments follows from considering the sequence $0 \le t_1/k^2 \le t_2/k^2 \le \ldots \le t_n/k^2$ for B. As for property 2 of Definition 3.1, observe that

$$\operatorname{Var}(kX) = \mathbb{E}[(kX - k\mu)^2] = k^2 \operatorname{Var}(X).$$

Thus, we have

$$\operatorname{Var}(B'(t+h) - B'(t)) = k^2 \operatorname{Var}(B((t+h)/k^2) - B(t/k^2)) = h. \qquad \blacksquare$$

Brownian motion doesn't only occur in one-dimension: the Brownian motion we observe in particles occurs in 3-D space. We can define Brownian motion in general in n-dimensions.

Definition 3.4. Let $B_1(t), B_2(t), \ldots, B_n(t)$ be independent one-dimensional Brownian motions. Then the stochastic process $\boldsymbol{B}(t) = \langle B_1(t), B_2(t), \ldots, B_n(t)\rangle$ is an *n-dimensional Brownian motion*.

Sample Brownian paths in two and three dimensions are shown in Figure 3. For simplicity, we'll stick to the one-dimensional case for the majority of this paper.

At this point, it's still unclear that Brownian motion exists in the first place because of the continuity property. Thankfully, it does. We'll explore constructions in Section 6.

4. Markov Properties

Now we'll tackle some more involved properties of Brownian motion. First, we'll define a couple of preliminaries to help us more rigorously tackle our later theorems.

Definition 4.1. Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a *filtration* $\{\mathcal{F}_t\}_{t\ge 0}$ is a family of increasing sub-σ-algebras such that $\mathcal{F}_{t_1} \subset \mathcal{F}_{t_2} \subset \mathcal{F}$ for all $0 \le t_1 < t_2$.

Definition 4.2. A stochastic process $\{X(t)\}_{t\ge 0}$ is *adapted* to a filtration $\{\mathcal{F}_t\}_{t\ge 0}$ if $X(t)$ is measurable on $\mathcal{F}_t$ for all $t \ge 0$.

Intuitively, a filtration tells us how much information we have at a particular time t. Filtrations are useful for calculating conditional expectations and defining stopping times, which we will examine in Definition 4.7.

Definition 4.3. The *standard filtration* with respect to Brownian motion is defined as

$$\mathcal{F}_t^B := \sigma(\{B(s)\}_{0\le s\le t}),$$

where $\sigma(\{B(s)\}_{0\le s\le t})$ is the smallest σ-algebra containing all events $\{B(s)\}$ on $0 \le s \le t$.

Standard filtrations allow us to state things relating to Brownian motion more concisely and more rigorously. Instead of saying that a process is independent of $\{B(t)\}_{0\le t\le h}$, for example, we can say it is independent of $\mathcal{F}_h^B$.

We are now ready to prove our first Markov property.

Theorem 4.4 (Markov Property)**.** *Let $\{B(t)\}_{t\ge 0}$ be a Brownian motion. Then, for all $h \in \mathbb{R}_{\ge 0}$, $\{B(t+h) - B(h)\}_{t\ge 0}$ is a standard Brownian motion and is independent of $\mathcal{F}_h^B$.*

Proof. The fact that $\{B(t+h) - B(h)\}_{t\ge 0}$ is a standard Brownian motion follows from Proposition 3.2. The independence property follows from property 1 of Definition 3.1. ■

Remark 4.5. The Markov property can be thought of as saying that Brownian motion restarts at any deterministic time t.

In fact, we can do slightly better than this. Consider the augmented filtration

$$\mathcal{F}_t^+ := \bigcap_{s>t} \mathcal{F}_s.$$

This filtration has the nice property that it is right continuous:

$$\bigcap_{s>t} \mathcal{F}_s^+ = \bigcap_{\varepsilon>0} \mathcal{F}_{t+\varepsilon}^+ = \bigcap_{\varepsilon>0}\bigcap_{\delta>0} \mathcal{F}_{t+\varepsilon+\delta} = \mathcal{F}_t^+.$$

We can think of $\mathcal{F}_t^+$ as giving ourselves an infinitely small glance into the future. Right continuity is important to ensure our flow of information is smooth and doesn't jump irregularly.

Proposition 4.6. *$\{B(t+h) - B(h)\}_{t\ge 0}$ is independent of $\mathcal{F}_h^+$ for all $h \ge 0$.*

Proof. Define a sequence $\{a_n\}_{n\ge 1}$ of positive real numbers that converges to zero. By Theorem 4.4, $\{B(t+h+a_n) - B(h+a_n)\}_{t\ge 0}$ is independent of $\mathcal{F}_{h+a_n}^B \supset \mathcal{F}_h^+$, so by continuity $\{B(t+h) - B(h)\}_{t\ge 0} = \lim_{n\to\infty}\{B(t+h+a_n) - B(h+a_n)\}_{t\ge 0}$ is independent of $\mathcal{F}_h^+$. ■

Interestingly enough, this "restarting" property of the Markov property holds not only for deterministic times but also certain random times called stopping times.

Definition 4.7. A random variable $\tau \in [0,\infty]$ defined on $\{\mathcal{F}_t\}_{t\ge 0}$ is called a *stopping time* if for all $t \ge 0$, it satisfies

$$\{\tau < t\} \in \mathcal{F}_t.$$

Example. The following are all stopping times.

- Any deterministic time t.
- The first time $B(t)$ reaches a certain value, i.e. $\inf\{t \ge 0 : B(t) = k\}$ for any particular $k \in \mathbb{R}$ (this is also called the first hitting time).
- $\tau_n = (k+1)2^{-n}$, where $k2^{-n} \le \tau < (k+1)2^{-n}$ for any stopping time τ.

Nonexample. $\inf\{t : B(t) = \inf\{B(t)\}_{0\le t\le 1}\}$, or the first time $B(t)$ hits its minimum value on $[0,1]$, is not a stopping time.

If we think of τ as the time a particular event occurs, $\{\tau < t\}$ means that that event has already occurred by time t. The stopping time properties means that using only information up to the present, we should be able to deduce whether that event has already occurred. We can also define the filtration of a stopping time as follows:

$$\mathcal{F}_\tau^+ = \{A \in \mathcal{A} : A \cap \{\tau < t\} \in \mathcal{F}_t^+ \;\forall\, t \geq 0\}.$$

We now prove a stronger version of the Markov property. We reference the proof in [MP08].

Theorem 4.8 (Strong Markov Property)**.** *Let $\{B(t)\}_{t\geq 0}$ be a standard Brownian motion and τ a stopping time on $\mathcal{F}_t^B$. Then, for $t \geq 0$, the process $B'(t) = B(t+\tau) - B(\tau)$ satisfies the following properties:*

- *$\{B'(t)\}_{t\geq 0}$ is a standard Brownian motion*
- *$\{B'(t)\}_{t\geq 0}$ is independent of $\mathcal{F}_\tau^+$ for all $t > 0$.*

Proof. Consider a sequence of stopping times $\{\tau_n\}_{n\geq 1}$ as defined previously. Define $B_k(t) = B(t + k2^{-n}) - B(k2^{-n})$ for all $k \in \mathbb{Z}_{\geq 0}$ and $B_{\tau_n}(t) = B(t+\tau_n) - B(\tau_n)$. We want to show that $B_{\tau_n}(t)$ is a Brownian motion and is independent of $\mathcal{F}_{\tau_n}^+$.

Let E be any event in $\mathcal{F}_{\tau_n}^+$ and $\{B_{\tau_n} \in A\}$ be some measurable event about the path B_{τ_n}. We have

$$\begin{aligned}
\mathbb{P}(\{B_{\tau_n} \in A\} \cap E) &= \sum_{k=0}^{\infty} \mathbb{P}(\{B_k \in A\} \cap E \cap \{\tau_n = k2^{-n}\}) \\
&= \sum_{k=0}^{\infty} \mathbb{P}(\{B_k \in A\})\mathbb{P}(E \cap \{\tau_n = k2^{-n}\}) \qquad (1) \\
&= \mathbb{P}(\{B \in A\}) \sum_{k=0}^{\infty} \mathbb{P}(E \cap \{\tau_n = k2^{-n}\}) \qquad (2) \\
&= \mathbb{P}(\{B \in A\})\mathbb{P}(E)
\end{aligned}$$

where (1) follows because $\{B_k \in A\}$ is independent of $E \cap \{\tau_n = k2^{-n}\} \in \mathcal{F}_{k2^{-n}}^+$ by Proposition 4.6 and (2) follows by the Markov Property (Theorem 4.4). By setting $E = \Omega$ in the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we get $\mathbb{P}(\{B_{\tau_n} \in A\}) = \mathbb{P}(\{B \in A\})$, so $B_{\tau_n}(t)$ is a Brownian motion. Furthermore, letting E be any event again shows that B_{τ_n} is independent of $\mathcal{F}_{\tau_n}^+$.

Now we look at $B'(t)$ over the interval $[t_1, t_2]$ for any $0 \leq t_1 \leq t_2$. Since $B_{\tau_n}(t)$ is Brownian motion independent of $\mathcal{F}_{\tau_n}^+ \supset \mathcal{F}_\tau^+$ and

$$B'(t_2) - B'(t_1) = B(t_2 + \tau) - B(t_1 + \tau) = \lim_{n\to\infty} B_{\tau_n}(t_2) - B_{\tau_n}(t_1),$$

we conclude that $B'(t)$ is also a Brownian motion, and it is independent of $\mathcal{F}_\tau^+$. ■

From the Strong Markov Property, we get the following two corollaries.

Corollary 4.9 (Splicing)**.** *Let $\{B_1(t)\}_{t\geq 0}$ and $\{B_2(t)\}_{t\geq 0}$ be standard Brownian motions, and let τ be a stopping time. Then the process*

$$B(t) = \begin{cases} B_1(t) & \text{if } 0 \leq t \leq \tau \\ B_1(\tau) + B_2(t-\tau) & \text{if } t > \tau \end{cases}$$

is a Brownian motion.

Corollary 4.10 (Reflection Principle)**.** *Let τ be a stopping time. Then the process*

$$B'(t) = \begin{cases} B(t) & \text{if } 0 \le t \le \tau \\ 2B(\tau) - B(t) & \text{if } t > \tau \end{cases}$$

is a Brownian motion. This is called the Brownian motion reflected at τ.

5. Martingale Property

We now prove another property of Brownian motion.

Definition 5.1. A continuous-time *martingale* with respect to a filtration $\mathcal{F}_t$ is a process $\{X_t\}_{t\ge 0}$ with the following properties:

(1) $\mathbb{E}[|X_t|] < \infty$
(2) X_t is adapted to $\mathcal{F}_t$
(3) $\mathbb{E}[X_t|\mathcal{F}_s] = X_s$ with probability 1 for all $0 \le s \le t$.

Theorem 5.2 (Martingale Property)**.** *The Brownian motion $\{B(t)\}_{t\ge 0}$ is a martingale with respect to the standard filtration.*

Proof. We have

$$\begin{aligned}\mathbb{E}[B(t)|\mathcal{F}_s^B] &= \mathbb{E}[B(t) - B(s) + B(s)|\mathcal{F}_s^B] \\ &= \mathbb{E}[B(t) - B(s)|\mathcal{F}_s^B] + \mathbb{E}[B(s)|\mathcal{F}_s^B] \\ &= \mathbb{E}[B(t) - B(s)] + B(s) \text{ (property of independent increments)} \\ &= B(s).\end{aligned}$$

■

Not only is $B(t)$ a martingale, so are $B(t)^2 - t$ and $e^{\lambda B(t) - \lambda^2 t/2}$ for all $\lambda \in \mathbb{C}$. For the latter, observe that

$$\begin{aligned}\mathbb{E}[e^{\lambda B(t)-\lambda^2 t/2}|\mathcal{F}_s^B] &= e^{-\lambda^2 t/2}\,\mathbb{E}[e^{\lambda[B(t)-B(s)]}e^{\lambda B(s)}|\mathcal{F}_s^B] \\ &= e^{\lambda B(s)-\lambda^2 t/2}\,\mathbb{E}[e^{\lambda[B(t)-B(s)]}] \text{ (by independent increments).}\end{aligned}$$

Let $X = B(t) - B(s)$, so X is a normal random variable with mean 0 and variance $\sigma^2 = t - s$. Then, we have

$$\begin{aligned}\mathbb{E}[e^{\lambda X}] &= \int_{-\infty}^{\infty} e^{\lambda X}\frac{1}{\sigma\sqrt{2\pi}}e^{-x^2/2\sigma^2}\,dx \\ &= e^{\lambda^2\sigma^2/2}\int_{-\infty}^{\infty}\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\lambda\sigma^2)^2/2\sigma^2}\,dx \\ &= e^{\lambda^2\sigma^2/2},\end{aligned}$$

where the last equality holds because $\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\lambda\sigma^2)^2/2\sigma^2}$ is the density function of a normal variable with mean $\lambda\sigma^2$ and variance σ^2. Thus,

$$\mathbb{E}[e^{\lambda B(t)-\lambda^2 t/2}|\mathcal{F}_s^B] = e^{\lambda B(s)-\lambda^2 t/2}e^{\lambda^2(t-s)/2} = e^{\lambda B(s)-\lambda^2 s/2}.$$

Exercise. Show that $B(t)^2 - t$ is a martingale. (Hint: The proof is similar to the proof of Theorem 5.2.)

Like how the Markov property had a random-time counterpart, the martingale property also has its random-time match in the form of Wald's First Lemma, although the correspondence is not as direct as in the Markov case. We will simply state the result here. A proof of Wald's First Lemma using the Optional Stopping Theorem can be found in [MP08], and a proof of the discrete case of the Optional Stopping Theorem can be found in [Wil91].

Theorem 5.3 (Wald's First Lemma). *Given a stopping time τ such that $\mathbb{E}[\tau]$ is finite or $\{B(t \wedge \tau)\}_{t \geq 0}$ is L^1-bounded, we have $\mathbb{E}[B(\tau)] = 0$.*

Remark 5.4. Wald's First Lemma may at first seem counterintuitive. For example, take $\tau = \inf\{t \geq 0 : B(t) = 1\}$, the first time $B(t)$ equals 1. One may expect that $\mathbb{E}[B(\tau)]$ would equal 1 in this case and violate Wald's First Lemma. This is why the $\mathbb{E}[\tau] < \infty$ condition is important: in this case, the expected value of τ does not converge.

Another way to think about this is letting $B(t)$ be a fair game, and our winnings at time t are equal to $B(t)$. Wald's First Lemma says that if we play the game until a random stopping time τ satisfying $\mathbb{E}[\tau] < \infty$ or $\{B(t \wedge \tau)\}_{t \geq 0}$ is L^1-bounded, our expected winnings are 0. If we set our strategy as playing until $\tau = \inf\{t \geq 0 : B(t) = 1\}$, we would technically expect our winnings to equal 1; however, we may have to wait infinitely long for that hitting time to occur, and we may lose infinitely much in the meantime.

Remark 5.5. Stating Theorem 5.3 as Wald's *First* Lemma seems to imply there exists a second. In fact, there does. The interested reader is encouraged to check out [MP08] for more.

6. Derivation and Existence

Up until now, we've proved properties of Brownian motion under the assumption that such a process exists in the first place. If it didn't exist, all the previous sections would go to waste; thankfully, that's not the case. The existence of Brownian motion was first proven by Wiener.

Theorem 6.1 (Wiener 1923). *Brownian motion exists.*

It's pretty clear that the properties 1 and 2 of Definition 3.1 are consistent: If $B(s) \sim \mathcal{N}(0, s)$ and $B(t-s) \sim \mathcal{N}(0, t-s)$ are independent, then $B(t) \sim \mathcal{N}(0, t)$ by Proposition 2.6. However, there's no obvious reason why such a process can also be continuous.

We will look at the construction by Lévy following the proof outlined in [MP08]. Despite its length, this method is actually quite elegant and is arguably the easiest to appreciate in its full rigor. Throughout this section, we use facts about Gaussian random vectors and their distributions, which are stated and proved in Appendix A. We start by constructing "Brownian motion" on the dyadic rationals on $[0, 1]$ and connecting consecutive points linearly.

Define $\mathcal{D}_n = \{\frac{k}{2^n}\}_{0 \leq k \leq 2^n}$, and let $\mathcal{D} = \cup_{n=0}^{\infty} \mathcal{D}_n$ be the dyadic points in $[0, 1]$. We'll define a set of random variables $\{Z_t\}_{t \in \mathcal{D}}$ such that $Z_t \sim \mathcal{N}(0, 1)$ and the Z_t's are pairwise independent for all $t \in \mathcal{D}$.

Without loss of generality, it suffices to show that standard Brownian motion exists. Let $B(0) = 0$ and $B(1) = Z_1$. Note that $B(1) - B(0) \sim \mathcal{N}(0, 1)$ as desired. Having shown the base case, we can proceed using induction.

Lemma 6.2. *Assume $B(d)$ has been defined for all $d \in \mathcal{D}_{n-1}$ such that $B(d)$ satisfies properties* 1 *and* 2 *of Definition 3.1 on $d \in \mathcal{D}_{n-1}$. For all $d \in \mathcal{D}_n \backslash \mathcal{D}_{n-1}$, define*

$$B(d) = \frac{B(d-2^{-n}) + B(d+2^{-n})}{2} + \frac{Z_d}{2^{(n+1)/2}}.$$

Then $B(d)$ satisfies properties 1 *and* 2 *of Definition* 3.1 *over $d \in \mathcal{D}_n$ as well.*

Proof. We can express $B(d)$ as follows:

$$B(d) = \frac{B(d+2^{-n}) - B(d-2^{-n})}{2} + B(d-2^{-n}) + \frac{Z_d}{2^{(n+1)/2}}.$$

By our inductive hypothesis, the three terms on the right are independent with mean zero and variances 2^{-n-1}, $d - 2^{-n}$, and 2^{-n-1} in that order. Thus, by Proposition 2.6, $B(d) \sim \mathcal{N}(0, d)$ for all $d \in \mathcal{D}_n$.

Now we will prove the property of independent increments. If we can show that all intervals of length 2^{-n} are pairwise independent, that is sufficient because we can express any interval as the sum of consecutive intervals of length 2^{-n}. Let us consider increments $B(t_1 + 2^{-n}) - B(t_1)$ and $B(t_2 + 2^{-n}) - B(t_2)$ for $0 \le t_1 < t_1 + 2^{-n} \le t_2$. In the case where $t_1 \in \mathcal{D}_{n-1}$ and $t_2 = t_1 + 2^{-n}$, we have

$$B(t_1 + 2^{-n}) - B(t_1) = \frac{B(t_1 + 2^{-n+1}) - B(t_1)}{2} + \frac{Z_{t_2}}{2^{(n+1)/2}}$$

and

$$B(t_2 + 2^{-n}) - B(t_2) = B(t_1 + 2^{-n+1}) - B(t_1 + 2^{-n}) = \frac{B(t_1 + 2^{-n+1}) - B(t_1)}{2} - \frac{Z_{t_2}}{2^{(n+1)/2}}.$$

Since $\frac{B(t_1+2^{-n+1})-B(t_1)}{2}$ and $\frac{Z_{t_2}}{2^{(n+1)/2}}$ are independent normal variables with mean zero and equal variance, we use Corollary A.5 in the appendix to prove that $B(t_1 + 2^{-n}) - B(t_1)$ and $B(t_2 + 2^{-n}) - B(t_2)$ are independent.

In all other cases, there exists some $d \in \mathcal{D}_{n-1}$ such that $t_1 + 2^{-n} \le d \le t_2$. Then, $B(t_1 + 2^{-n}) - B(t_1)$ can be expressed as a linear combination of an increment in $\mathcal{D}_{n-1}$ to the left of d and Z_{d_1} for $d_1 \in \mathcal{D}_n \backslash \mathcal{D}_{n-1}$ and $d_1 < d$, and $B(t_2 + 2^{-n}) - B(t_2)$ can be expressed as a linear combination of an increment in $\mathcal{D}_{n-1}$ to the right of d and Z_{d_2} for $d_2 \in \mathcal{D}_n \backslash \mathcal{D}_{n-1}$ and $d_2 > d$. Each of these terms are independent by our inductive hypothesis, so $B(t_1 + 2^{-n}) - B(t_1)$ and $B(t_2 + 2^{-n}) - B(t_2)$ are independent. ■

Now that we've constructed Brownian motion on the dyadic points, we connect the function linearly between consecutive points in $\mathcal{D}$ to get a continuous function. Formally, define

$$F_0(t) = \begin{cases} 0 & \text{if } t = 0 \\ Z_1 & \text{if } t = 1 \\ \text{linear} & \text{in between} \end{cases}$$

and

$$F_n(t) = \begin{cases} 0 & \text{if } t \in \mathcal{D}_{n-1} \\ Z_d/2^{(n+1)/2} & \text{if } t \in \mathcal{D}_n \backslash \mathcal{D}_{n-1} \\ \text{linear} & \text{in between.} \end{cases}$$

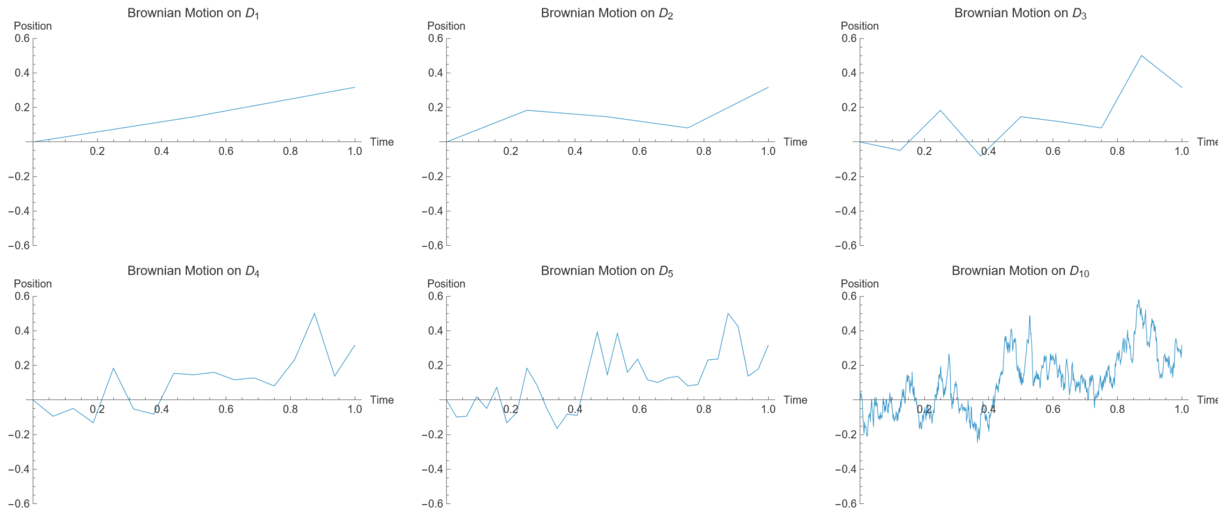

Figure 4. Constructing Brownian motion over the dyadic points.

Then let $B(t) = \sum_{i=0}^{\infty} F_i(t)$. Since the $F_n(t)$'s are continuous, so is $B(t)$. It remains to show that this sum converges and satisfies properties 1 and 2 of Definition 3.1 for all t on $[0,1]$. See Figure 4 for a visualization of this process $(\sum_{i=0}^{1} F_i(t), \sum_{i=0}^{2} F_i(t), \ldots)$.

Lemma 6.3. *$B(t)$ converges with probability 1 on $[0,1]$.*

Proof. We will show that Z_d is very large only finitely often. First, we estimate Gaussian tails. Since the probability density function of a normal distribution is difficult to compute, we will use the following bound: for a standard normal random variable X and a positive constant c,

$$P\{X \geq c\} = \int_c^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx \leq \int_c^\infty \frac{x}{c} e^{-x^2/2}\,dx = \frac{e^{-c^2/2}}{c}.$$

Thus, for any positive constant k, $P\{Z_d \geq k\sqrt{n}\} \leq e^{-k^2 n/2}$. Now we can bound the probability of Z_d taking on large values:

$$\sum_{n=0}^{\infty} P\{\exists\, d \in \mathcal{D}_n \text{ such that } |Z_d| \geq k\sqrt{n}\} \leq \sum_{n=0}^{\infty} \sum_{i=1}^{2^n} P\{|Z_{i/2^n}| \geq k\sqrt{n}\}$$
$$\leq \sum_{n=0}^{\infty} 2^n e^{-k^2 n/2}.$$

This sum is guaranteed to converge if $k > \sqrt{2\ln 2}$. Fix such a k. Then, by the Borel-Cantelli Convergence Lemma (2.9), $Z_d \geq k\sqrt{n}$ only finitely many times. Thus, there exists some positive finite integer N such that

$$\sup F_n(t) \leq \frac{k\sqrt{n}}{2^{n/2}}$$

for all $n \geq N$ and $0 \leq t \leq 1$. Since

$$B(t) \leq \sum_{n=0}^{N-1} F_n(t) + \sum_{n=N}^{\infty} \sup F_n(t)$$

for all $t \in [0,1]$ and the right hand side converges, $B(t)$ converges over $t \in [0,1]$. ■

Lemma 6.4. *$B(t)$ has the right distribution for non-dyadic points.*

Proof. Consider the time sequence $0 \le t_1 < t_2 < \ldots < t_n \le 1$ and sequences $\{t_{i,k}\}_{k\ge 0}$ for all $1 \le i \le n$ such that $\lim_{k\to\infty} t_{i,k} = t_i$ and $t_{1,k} \le t_{2,k} \le \ldots \le t_{n,k}$. Consider the random vector $\boldsymbol{X} = \langle B(t_2) - B(t_1), \ldots, B(t_n) - B(t_{n-1})\rangle$. Observe that

$$\mathbb{E}[B(t_{i+1}) - B(t_i)] = \lim_{k\to\infty} \mathbb{E}[B(t_{i+1,k}) - B(t_{i,k})] = 0$$

for all $1 \le i \le n-1$, and

$$\begin{aligned}\mathrm{Cov}(B(t_{i+1}) - B(t_i), B(t_{j+1}) - B(t_j)) &= \lim_{k\to\infty} \mathrm{Cov}(B(t_{i+1,k}) - B(t_{i,k}), B(t_{j+1,k}) - B(t_{j,k})) \\ &= \begin{cases} 0 & \text{if } i \ne j \\ t_{i+1} - t_i & \text{if } i = j. \end{cases}\end{aligned}$$

Thus, $\mathbb{E}[\boldsymbol{X}] = 0$ and $\mathrm{Cov}(\boldsymbol{X})$ is a diagonal matrix with entries $t_{i+1} - t_i$. Since the expectation and covariance matrix of a Gaussian random vector determine its distribution by Lemma A.6, the elements of $\boldsymbol{X}$ are independent, and $B(t_{i+1}) - B(t_i)$ has variance $t_{i+1} - t_i$ for all $1 \le i \le n-1$, so $B(t)$ satisfies all three properties of Definition 3.1 on $t \in [0,1]$. ■

We're now able to complete Lévy's construction and proof for Wiener's Theorem.

Proof of Theorem 6.1. Construct a set of independent Brownian motions $\{B_n(t)\}_{n>0}$ on $\mathcal{D}$ according to Lemma 6.2 and extend to all of $[0,1]$ by connecting linearly between consecutive dyadic points. By Lemmas 6.3 and 6.4, these functions converge and satisfy properties 1 and 2 of Definition 3.1. Now we extend our Brownian motion to $\mathbb{R}_{\ge 0}$ as follows:

$$B(t) = B_{\lfloor t\rfloor+1}(t - \lfloor t\rfloor) + \sum_{i=1}^{\lfloor t\rfloor} B_i(1).$$

The continuity of this function follows from its definition. The fact that it has independent increments with mean 0 and variance equal to the length of the increment follows from properties of independent normal random variables. See Figure 5 for a visualization. ■

7. Variation and Stochastic Calculus

7.1. **Total Variation.**

Definition 7.1. The *total variation* of a function $f : [0,T] \to \mathbb{R}$ is

$$\mathrm{TV}_T(f) = \lim_{n\to\infty} \sup_{\pi_n} \sum_{i=0}^{n-1} |f(t_{i+1}) - f(t_i)|,$$

where π_n is the set of partitions $0 = t_0 < t_1 < \ldots < t_n = T$ of $[0,T]$ such that the size of the largest interval goes to zero as $n \to \infty$.

If $\mathrm{TV}_T(f) < \infty$, we say f has finite total variation; otherwise, it has infinite total variation. Total variation is important in defining the Riemann-Stieltjes integral, which is a generalization of the Riemann integral where we integrate with respect to a function rather

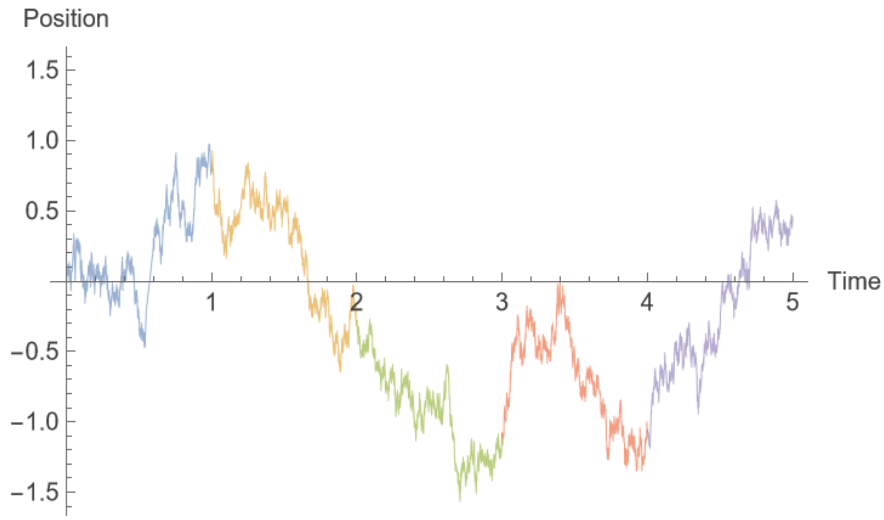

Figure 5. Lévy's Construction of Brownian Motion extended to $\mathbb{R}_{\geq 0}$.

than just dx. If we assume the Riemann-Stieltjes integral with respect to a function f is defined, we have

$$\int_0^T g(x)\, df(x) = \lim_{n\to\infty} \sup_{\pi_n} \sum_{i=0}^{n-1} g(t_i^*)(f(t_{i+1}) - f(t_i)),$$

where t_i^* is randomly chosen on $[t_i, t_{i+1}]$. In order for this integral to be defined, f must have finite total variation. We claim this is not the case with Brownian motion.

Theorem 7.2 (Infinite Variation)**.** *Brownian motion has infinite total variation with probability* 1*.*

Before we prove this theorem, we first establish some helpful inequalities that we will use in our proof.

Lemma 7.3 (Markov's Inequality)**.** *If X is a random variable that only takes on nonnegative values, then $\mathbb{P}\{X \geq a\} \leq \frac{\mathbb{E}[X]}{a}$ for all $a \in \mathbb{R}_{>0}$.*

Proof. Define a random variable Y such that

$$Y = \begin{cases} 1 & \text{if } X \geq a, \\ 0 & \text{otherwise.} \end{cases}$$

Note that we always have $Y \leq \frac{X}{a}$. Thus, $\mathbb{P}\{X \geq a\} = \mathbb{E}[Y] \leq \frac{\mathbb{E}[X]}{a}$. ∎

Lemma 7.4 (Chebyshev's Inequality)**.** *Let X be a random variable with finite mean μ and variance σ^2. Then for all $k \in \mathbb{R}_{>0}$,*

$$\mathbb{P}\{|X - \mu| \geq k\} \leq \frac{\sigma^2}{k^2}.$$

Proof. Setting X as $(X-\mu)^2$ and a as k^2 in the Markov Inequality (Lemma 7.3) gives

$$\mathbb{P}\{(X-\mu)^2 \geq k^2\} \leq \frac{\sigma^2}{k^2}.$$

The desired result follows. ∎

Lemma 7.5. *Given positive integers r, s such that $r > s$,*

$$\sup_{\pi_r} \sum_{i=0}^{r-1} |f(t_{i+1}) - f(t_i)| \geq \sup_{\pi_s} \sum_{i=0}^{s-1} |f(t_{i+1}) - f(t_i)|.$$

Proof. For any partition $0 = t_0^* < t_1^* < \ldots < t_s^* = T$ of $[0, T]$ into s intervals, we can define a partition $0 = t_0' < t_1' < \ldots < t_r' = T$ of $[0, T]$ into r such that $t_i' = t_i^*$ for all $0 \leq i < s$ and $t_{s-1}^* < t_s' < \ldots < t_r' = T$. Observe that

$$\sup_{\pi_r} \sum_{i=0}^{r-1} |f(t_{i+1}) - f(t_i)| \geq \sum_{i=0}^{r-1} |f(t_{i+1}') - f(t_i')| \geq \sum_{i=0}^{s-1} |f(t_{i+1}^*) - f(t_i^*)|.$$

The lemma follows. ■

Now we are ready to prove the infinite total variation property of Brownian motion.

Proof of Theorem 7.2. Consider the partition of $[0, T]$ into n intervals of equal lengths. Define a function

$$V_{n,T}(f) := \sum_{i=0}^{n-1} |f(\tfrac{i+1}{n} \cdot T) - f(\tfrac{i}{n} \cdot T)|$$

for any function f. For a Brownian motion $\{B(t)\}_{t \geq 0}$, note that

$$\mathbb{E}[V_{n,T}(B)] = \mathbb{E}\left[\sum_{i=0}^{n-1} |B(t_{i+1}) - B(t_i)|\right] = \sum_{i=0}^{n-1} \mathbb{E}[|B(t_{i+1}) - B(t_i)|]$$

by linearity of expectation. Let X be a random variable with mean 0 and variance σ^2. We have

$$\begin{aligned}
\mathbb{E}[|X|] &= \int_{-\infty}^{\infty} |x| \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}\, dx \\
&= 2\int_{0}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}\, dx \\
&= \left(-\sigma\sqrt{\frac{2}{\pi}} e^{-x^2/2\sigma^2}\right)\Bigg]_0^{\infty} \\
&= \sigma\sqrt{\frac{2}{\pi}}.
\end{aligned}$$

Thus, we know

$$\mathbb{E}\left[V_{n,T}(B)\right] = \sum_{i=0}^{n-1} \sqrt{\frac{2T}{n\pi}} = \sqrt{\frac{2nT}{\pi}}.$$

As $n \to \infty$, this expression goes to ∞. Moreover, because of the property of independent increments, we know that

$$\begin{aligned}\mathrm{Var}(V_{n,T}(B)) &= \sum_{i=0}^{n-1} \mathrm{Var}(|B(\tfrac{i+1}{n} \cdot T) - B(\tfrac{i}{n} \cdot T)|) \\ &= \sum_{i=0}^{n-1} \left(\mathbb{E}[(B(\tfrac{i+1}{n} \cdot T) - B(\tfrac{i}{n} \cdot T))^2] - \mathbb{E}[|B(\tfrac{i+1}{n} \cdot T) - B(\tfrac{i}{n} \cdot T)|]^2 \right) \\ &= T(1 - \frac{2}{\pi}).\end{aligned}$$

Now that we've computed the mean and variance of $V_{n,T}(B)$, we can use Chebyshev's Inequality to bound the probability that the value of $V_{n,T}(B)$ is far from the mean. By Lemma 7.4, we have

$$\mathbb{P}\{|V_{n,T}(B) - \mathbb{E}[V_{n,T}(B)]| > \tfrac{1}{2}\mathbb{E}[V_{n,T}(B)]\} \leq \frac{T(1-2/\pi)}{2nT/(4\pi)} < \frac{2\pi}{n}.$$

Let A_i denote the event $\{|V_{i^2,T}(B) - \mathbb{E}[V_{i^2,T}(B)]| > \frac{1}{2}\mathbb{E}[V_{i^2,T}(B)]\}$. Since $\sum_{i=1}^{\infty} \mathbb{P}(A_i) \leq \sum_{i=1}^{\infty} \frac{2\pi}{i^2}$ converges, by the Borel-Cantelli Convergence Lemma (2.9), $V_{i^2,T}(B) < \frac{1}{2}\mathbb{E}[V_{i^2,T}(B)] = i\sqrt{T/(2\pi)}$ only finitely many times. Thus, there exists a sufficiently large finite N such that $V_{n^2,T}(B) \geq n\sqrt{T/(2\pi)}$ for all $n \geq N$. By Lemma 7.5, we know

$$\sup_{\pi_n} \sum_{i=0}^{n-1} |f(t_{i+1}) - f(t_i)| \geq \sup_{\pi_{\lfloor\sqrt{n}\rfloor^2}} \sum_{i=0}^{\lfloor\sqrt{n}\rfloor^2 - 1} |f(t_{i+1}) - f(t_i)| \geq \lfloor\sqrt{n}\rfloor \sqrt{\frac{T}{2\pi}}.$$

As $n \to \infty$, the rightmost expression goes to infinity, giving us infinite total variation for Brownian motion. ■

As a result, we cannot define a Riemann-Stieltjes integral with respect to a Brownian motion.

7.2. **Quadratic Variation.**

Definition 7.6. The *quadratic variation* of a function $f : [0, T] \to \mathbb{R}$ is

$$\mathrm{QV}_T(f) = \lim_{n\to\infty} \sup_{\pi_n} \sum_{i=0}^{n-1} (f(t_{i+1}) - f(t_i))^2,$$

where π_n is as defined previously.

Functions with a continuous derivative have a quadratic variation of zero, which is why we don't look at the quadratic variation of such functions. Brownian motion, despite having infinite linear variation, has finite nonzero quadratic variation.

Theorem 7.7 (Quadratic Variation)**.** $\mathrm{QV}_T(B)$ *converges to* T *in* L^2.

Proof. To show convergence in L^2, we will compute the expectation and variance of $\mathrm{QV}_T(B)$. Observe that

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{E}\left[\sum_{i=0}^{n-1}(B(t_{i+1})-B(t_i))^2\right] &= \lim_{n\to\infty}\sum_{i=0}^{n-1}\mathbb{E}[(B(t_{i+1})-B(t_i))^2] \\
&= \lim_{n\to\infty}\sum_{i=0}^{n-1} t_{i+1}-t_i \\
&= T.
\end{aligned}$$

By independent increments,

$$\begin{aligned}
&\mathrm{Var}\left(\sum_{i=0}^{n-1}(B(t_{i+1})-B(t_i))^2\right) \\
&= \sum_{i=0}^{n-1}\mathrm{Var}\left((B(t_{i+1})-B(t_i))^2\right) + \sum_{0\le i,j<n, i\neq j}^{n-1}\mathrm{Cov}\left((B(t_{i+1})-B(t_i))^2,(B(t_{j+1})-B(t_j))^2\right) \\
&= \sum_{i=0}^{n-1}\mathrm{Var}\left((B(t_{i+1})-B(t_i))^2\right).
\end{aligned}$$

For a random variable $X \sim \mathcal{N}(0,\sigma^2)$,

$$\mathbb{E}[X^4] = \int_{-\infty}^{\infty}\frac{x^4}{\sigma\sqrt{2\pi}}e^{-x^2/2\sigma^2}\,dx = 3\sigma^2\int_{-\infty}^{\infty}\frac{x^2}{\sigma\sqrt{2\pi}}e^{-x^2/2\sigma^2}\,dx - \left(\frac{\sigma}{\sqrt{2\pi}}x^3e^{-x^2/2\sigma^2}\right)\Bigg]_{-\infty}^{\infty} = 3\sigma^4.$$

Thus,

$$\begin{aligned}
&\lim_{n\to\infty}\mathrm{Var}\left(\sum_{i=0}^{n-1}(B(t_{i+1})-B(t_i))^2\right) \\
&= \lim_{n\to\infty}\sum_{i=0}^{n-1}\mathrm{Var}((B(t_{i+1})-B(t_i))^2)) \\
&= \lim_{n\to\infty}\sum_{i=0}^{n-1}\mathbb{E}\left[((B(t_{i+1})-B(t_i))^2-(t_{i+1}-t_i))^2\right] \\
&= \lim_{n\to\infty}\sum_{i=0}^{n-1}\mathbb{E}\left[(B(t_{i+1})-B(t_i))^4-2(t_{i+1}-t_i)(B(t_{i+1})-B(t_i))^2+(t_{i+1}-t_i)^2\right] \\
&= \lim_{n\to\infty}\sum_{i=0}^{n-1}2(t_{i+1}-t_i)^2 \\
&\le \lim_{n\to\infty}\sum_{i=0}^{n-1}2(t_{i+1}-t_i)\max_{0\le i<n}(t_{i+1}-t_i) = \lim_{n\to\infty}2T\max_{0\le i<n}(t_{i+1}-t_i) = 0.
\end{aligned}$$

■

Remark 7.8. We've only proved convergence in L^2, but in fact $QV_T(B)$ converges to T with probability 1 without the L^2 condition as long as $0 = t_1 < t_2 < \ldots < t_n = T$ is a nested sequence (i.e. one or more partition points are added at each step). This nested condition, while capable of being replaced by other conditions, cannot be dropped fully: it's possible for $B(t)$ to have infinite quadratic variation with probability 1 over a non-nested sequence. For a more thorough discussion and proof of the quadratic variation of Brownian motion, see [MP08] and [Kal97].

7.3. **Stochastic Calculus.** Brownian motion's infinite linear variation and finite nonzero quadratic variation have important implications when attempting to integrate with respect to a Brownian motion. As we mentioned before, we cannot define a Riemann-Stieltjes integral with respect to Brownian motion due to its infinite linear variation. The extension of traditional calculus to Brownian motion takes the form of stochastic calculus.

We can see the implications of Brownian motion's nonzero quadratic variation on stochastic calculus through the Itô-Doeblin formula in differential form. Consider a differentiable function $f(x)$. The chain rule tells us that for a differentiable function $g(x)$, we have

$$\frac{d}{dx}f(g(x)) = f'(g(x))g'(x).$$

However, if the function inside f was a Brownian motion, the formula is different. The Itô-Doeblin formula states that

$$df(B(t)) = f'(B(t))dB(t) + \frac{1}{2}f''(B(t))dt.$$

The extra dt term appears because Brownian motion has nonzero quadratic variation. For our ordinary chain rule, there's no $(dx)^2$ term because quadratic variation equals zero; however, in Brownian motion, this $(dB(t))^2$ term equals dt and thus does not disappear. For a more detailed explanation of the topic, see [Shr04].

Acknowledgments

The author of this paper would like to thank Freya Edholm for her mentorship and helpful conversations during the process of writing this paper. The author would also like to thank Simon Rubinstein-Salzedo for his instruction on paper writing and for making Euler Circle possible.

References

[Bil95] Patrick Billingsley. *Probability and Measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, third edition, 1995.

[Kal97] Olav Kallenberg. *Foundations of Modern Probability*. Probability and Its Applications. Springer-Verlag, 1997.

[MP08] Peter Mörters and Yuval Peres. Brownian motion, 2008.

[Nel01] Edward Nelson. *Dynamical Theories of Brownian Motion*. Princeton Landmarks in Mathematics and Physics. Princeton University Press, second edition, 2001.

[Ros09] Sheldon M. Ross. *A First Course in Probability*. Pearson Prentice Hall, eighth edition, 2009.

[Shr04] Steven E. Shreve. *Stochastic Calculus for Finance II: Continuous-Time Models*. Springer Finance. Springer, 2004.

[Wie23] Norbert Wiener. Differential-space. *Journal of Mathematics and Physics*, 2(1–4):131–174, 1923.

[Wil91] David Williams. *Probability with Martingales*. Cambridge Mathematical Textbooks. Cambridge University Press, 1991.

Appendix A. Multivariate Gaussian Distributions

In this section, we build up to and prove lemmas on multivariate Gaussian distributions that are used to prove results in the body of the paper.

Definition A.1. A random vector $\boldsymbol{X} = \langle X_1, \ldots, X_n \rangle$ has the *n-dimensional standard Gaussian distribution* if its coordinates are all independent and standard normal random variables.

In fact, we can define Gaussian random vectors more generally.

Definition A.2. A random vector $\boldsymbol{X} = \langle X_1, \ldots, X_n \rangle$ is a *Gaussian random vector* if it can be expressed in the form of

$$\boldsymbol{X} = \boldsymbol{\mu} + A\boldsymbol{Y},$$

where $\mu \in \mathbb{R}^n$, A is an $n \times m$ matrix, and $\boldsymbol{Y}$ has a standard Gaussian distribution.

Such a Gaussian random vector has mean $\boldsymbol{\mu}$ and covariance matrix AA^T.

Definition A.3. (Multvariate Transformation Method) Let $\boldsymbol{X} = \langle X_1, \ldots, X_n \rangle$ be a random vector with joint probability density function $f_X(x)$, and let $\boldsymbol{Y} = \langle Y_1, \ldots, Y_n \rangle$ be a random vector with probability density function $f_Y(y)$ and $t_1, \ldots, t_n$ be invertible transformations such that $Y_i = t_i(X_1, \ldots, X_n)$ for all i between 1 and n. Define

$$J = \begin{pmatrix} \frac{\partial t_1^{-1}}{\partial y_1} & \cdots & \frac{\partial t_1^{-1}}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial t_n^{-1}}{\partial y_1} & \cdots & \frac{\partial t_n^{-1}}{\partial y_n} \end{pmatrix}$$

to be the Jacobian of the transformation, and let $\boldsymbol{y} = (y_1, \ldots, y_n)$. Then

$$f_Y(y) = f_X(t_1^{-1}(\boldsymbol{y}), \ldots, t_n^{-1}(\boldsymbol{y})) \cdot |\det J|.$$

Lemma A.4. *Given an n-dimensional standard Gaussian vector $\boldsymbol{X} = \langle X_1, \ldots, X_n \rangle$ and an orthogonal $n \times n$ matrix A, $A\boldsymbol{X}$ is also standard Gaussian.*

Proof. Let $\boldsymbol{x} = \langle x_1, \ldots, x_n \rangle$. Since the X_i's are independent, their joint distribution equals the product of their individual distributions, so

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-x_i^2}/2 = \frac{1}{(2\pi)^{n/2}} e^{-|\boldsymbol{x}|^2/2}.$$

By the multivariate transformation method (Definition A.3), $A\boldsymbol{X}$ has probability density function $f(A^{-1}\boldsymbol{X}) \cdot |\det A^{-1}| = f(A^{-1}\boldsymbol{X})$. Since orthogonal transformations preserve the Euclidean norm, $\boldsymbol{X}$ and $A\boldsymbol{X}$ have the same joint probability density function. ■

Corollary A.5. *Given two independent and identically distributed Gaussian random variables X and Y with mean 0 and variance σ^2, $X + Y$ and $X - Y$ are also independent and identically distributed with mean 0 and variance $2\sigma^2$.*

Proof. Since $-Y \sim \mathcal{N}(0, \sigma^2)$ as well, we have $X + Y \sim \mathcal{N}(0, 2\sigma^2)$ and $X - Y \sim \mathcal{N}(0, 2\sigma^2)$ by Proposition 2.6. The independence condition is a special case of Lemma A.4. Setting

$$A = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{X} = \begin{pmatrix} X \\ Y \end{pmatrix}$$

gives the desired result. ■

Lemma A.6. *Given two n-dimensional Gaussian vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ such that $\mathbb{E}[\boldsymbol{X}] = \mathbb{E}[\boldsymbol{Y}]$ and $\mathrm{Cov}(\boldsymbol{X}) = \mathrm{Cov}(\boldsymbol{Y})$, then $\boldsymbol{X}$ and $\boldsymbol{Y}$ have the same distribution.*

Proof. By definition, let $\boldsymbol{X} = A\boldsymbol{X}_1 + \boldsymbol{\mu}$ and $\boldsymbol{Y} = B\boldsymbol{Y}_1 + \boldsymbol{\mu}$. If $\boldsymbol{\mu}$ were nonzero, showing $\boldsymbol{X} - \boldsymbol{\mu} \stackrel{d}{=} \boldsymbol{Y} - \boldsymbol{\mu}$ implies $\boldsymbol{X} \stackrel{d}{=} \boldsymbol{Y}$, so we can assume $\boldsymbol{\mu} = 0$. Furthermore, by adding columns of zeros to A or B if necessary, we can assume that A, B are both $n \times m$ matrices and $\boldsymbol{X}_1$, $\boldsymbol{Y}_1$ are both m-dimensional vectors. The covariance condition gives us $AA^T = BB^T$

Consider the subspaces $\mathcal{A}, \mathcal{B} \subseteq \mathbb{R}^m$ generated by the rows of A, B respectively. Let $\dim(\mathcal{A}) = r$, and suppose rows $i_1, i_2, \ldots, i_r$ of A form a basis for $\mathcal{A}$. Define a linear map $L : \mathcal{A} \to \mathcal{B}$ that takes rows $i_1, i_2, \ldots, i_r$ of A to the corresponding rows of B. We want to show that L is an orthogonal isomorphism.

To show L is an isomorphism, we first show its kernel is 0. Let $e_1, \ldots, e_m$ be the standard basis. Define a vector $\boldsymbol{v} = c_1\boldsymbol{e}_{i_1}^T + \cdots + c_r\boldsymbol{e}_{i_r}^T$, such that $L(\boldsymbol{v}A) = L(c_1\boldsymbol{A}_{i_1} + c_2\boldsymbol{A}_{i_2} + \cdots + c_r\boldsymbol{A}_{i_r}) = 0$, where $\boldsymbol{A}_i$ denotes the ith row of A. By linearity, we have $\boldsymbol{v}B = 0$. Thus,

$$\|\boldsymbol{v}A\|^2 = \boldsymbol{v}AA^T\boldsymbol{v}^T = \boldsymbol{v}BB^T\boldsymbol{v}^T = 0,$$

so L is injective, giving us $\dim \mathcal{A} \leq \dim \mathcal{B}$. By swapping A and B, we can show that $\dim \mathcal{A} = \dim \mathcal{B}$, so L is an isomorphism. Furthermore, $AA^T = BB^T$ gives us that L preserves inner products, so L is orthogonal.

Now we will extend L to an orthogonal map $L_1 : \mathbb{R}^m \to \mathbb{R}^m$ by creating an orthogonal map from the orthogonal complement of $\mathcal{A}$ to the orthogonal complement of $\mathcal{B}$ in $\mathbb{R}^m$ and putting the two together. Thus, we have $\boldsymbol{Y} = B\boldsymbol{Y}_1 = AL_1^T\boldsymbol{Y}_1$. Since $L_1^T\boldsymbol{Y}_1$ is a standard Gaussian distribution by Lemma A.4, we have $\boldsymbol{X} \stackrel{d}{=} \boldsymbol{Y}$ as desired. ■