Brownian Motion, Martingales, and Markov Processes

5604425

Contents

1 Introduction **2**

2 Preliminaries **2**
2.1 Measure theory 2
2.2 Stochastic processes 5

3 Brownian motion **5**
3.1 Invariance properties 6
3.2 Nowhere differentiability 9

4 Martingales **12**

5 Markov processes **14**

1 Introduction

Once an event has transpired, it is easy to reflect and identify causes or justifications for why certain actions happened over others. However, to predict these actions before they happen — especially in a short time frame — is a great challenge for even professionals in their respective fields. This uncertainty motivates the use of probabilistic models with the objective of describing the distribution of possible behaviours, identifying potential trends, or a lack thereof.

Stochastic processes provide a mathematical description of such systems, of which Brownian motion is a famous example. Its name can be accredited to the English botanist Robert Brown in 1827, who discovered the process through his observation of small particles suspended in a stationary liquid. The foundations of associated theory were later set up by Einstein in 1905. What Einstein suggested was that the motion was random, and that the movements of a particle were continuous, time-independent, and could be modelled over a given time frame with a normal distribution [14, p. 15]. This work, one of Einstein's *annus mirabilis* papers, expedited science to what we know today, with these observations in Brownian motion providing evidence for the existence of atoms.

From a mathematical perspective, Brownian motion exhibits two of the most fundamental concepts in the theory of stochastic processes [10, p. vii]: the Markov property and the martingale property. This forms the main focus of this essay, with our objectives being to introduce Brownian motion and its remarkable properties, to explain what continuous-time Markov processes and martingales are, and to prove that Brownian motion satisfies both the Markov property and the martingale property.

2 Preliminaries

To begin to understand stochastic processes, we introduce the idea of measures, measurable spaces, and measurability, through the topic of measure theory. Necessary knowledge in basic probability theory will be assumed, including σ-algebras, expectation and variance, and the normal or Gaussian distribution, in addition to familiarity with topological spaces.

2.1 Measure theory

Measure theory is fundamental to stochastic processes, and therefore to Brownian motion, because it allows us to rigorously define stochastic processes and gain a high level of understanding of the probabilistic spaces we will be working in. The Lebesgue measure will be looked into as an example, as it has many similarities with Brownian motion.

Definition 2.1. [6, p. 8]. Let $\mathcal{F}$ be a σ-algebra on a set Ω. A **measure** is a function $\mu : \mathcal{F} \to [0, \infty]$ such that $\mu(\emptyset) = 0$, and μ is countably additive; that is, if $A_1, A_2, A_3, \ldots \in \mathcal{F}$ are disjoint, then

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

To present our example of a Lebesgue measure, we first need to introduce the concept of generating a σ-algebra, and of a Borel σ-algebra — the domain on which the measure will be defined.

Definition 2.2. [6, p. 3]. Let $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ be a family of subsets of a set Ω. The **σ-algebra generated by** $\mathcal{A}$, denoted $\sigma(\mathcal{A})$, is the smallest σ-algebra on Ω containing $\mathcal{A}$.

Example 2.3. [6, p. 4]. When $\mathcal{A} = \mathcal{P}(\Omega)$, we see that $\sigma(\mathcal{A})$ is the σ-algebra that consists of all subsets of $\mathcal{A}$. For any value of $\mathcal{A}$, the intersection of all σ-algebras which include all values in $\mathcal{A}$ is the smallest possible σ-algebra, and is therefore $\sigma(\mathcal{A})$.

Definition 2.4. [12, p. 1; 6, p. 4]. We define $\mathcal{B}(S)$, the **Borel σ-algebra** of a topological space S, as the smallest σ-algebra generated by the collection of open subsets of S. A **Borel set** in S is any set belonging to its Borel σ-algebra $\mathcal{B}(S)$.

Remark. For the Lebesgue measure, we are interested in the Borel σ-algebra of the real numbers $\mathcal{B}(\mathbb{R})$, which consists of all sets constructed from open intervals through countable unions, countable intersections, and complements. As the complement of an open set is closed, this means that $\mathcal{B}(\mathbb{R})$ includes all open subsets and closed subsets of the reals.

Example 2.5 (Lebesgue measure)**.** [11, p. 25]. There exists a measure λ which we define on the Borel σ-algebra of $\mathbb{R}$ such that for each interval $(a, b) \subset \mathbb{R}$,

$$\lambda((a, b)) = b - a.$$

Note that this does not mean the Lebesgue measure is defined only on $\mathcal{B}(\mathbb{R})$, only that we have restricted it as such. We will not prove the existence of such a measure, as this requires an explanation of outer measures and Carathéodory's extension theorem[1], but we will instead observe some of its properties.

Proposition 2.6 (Properties of the Lebesgue measure λ)**.**

- *Translation invariance: for any Borel set in the real numbers $A \in \mathcal{B}(\mathbb{R})$, and for any s in the real numbers,*

$$\lambda(A + s) = \lambda(A).$$

[1] A detailed construction can be found by Kubrusly [11, § 8].

- *Reflection invariance: for any Borel set in the real numbers $A \in \mathcal{B}(\mathbb{R})$,*

$$\lambda(-A) = \lambda(A).$$

- *Scale invariance: for any Borel set in the real numbers $A \in \mathcal{B}(\mathbb{R})$, and for any $k > 0$,*

$$\lambda(kA) = k\lambda(A).$$

Remark. These three properties of the Lebesgue measure are immediate for intervals. As the Borel σ-algebra is generated by families of open intervals, and we know the measure λ is countably additive, these properties extend to all Borel sets. To rigorously prove this fact requires additional results for measures, such as a monotone class argument, which can be seen by Florescu [8, p. 21].

Next, we will need to introduce the idea of a measurable space and a measure space, which will be used for our definition of a stochastic process.

Definition 2.7. [6, pp. 8–9]. Let E be a set and $\mathcal{E}$ a σ-algebra over E. The pair $(E, \mathcal{E})$ is called a **measurable space**. Let μ be a measure on $\mathcal{E}$. The triple $(E, \mathcal{E}, \mu)$ is known as a **measure space**.

Example 2.8. We can say that for our Lebesgue measure λ, $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is its associated measurable space. The most common measure space in the topic of Brownian motion is the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, which has the additional property that $\mathbb{P}(\Omega) = 1$.

When martingales and Markov processes are formally introduced, we will refer to the term measurable with several different contexts, so it is important we give a precise definition for what it means for a function to be measurable with respect to a σ-algebra.

Definition 2.9. [6, p. 79]. Let $(\Omega, \mathcal{F})$ and $(E, \mathcal{E})$ be measurable spaces. We say that a function $f : \Omega \to E$ is **measurable with respect to $\mathcal{E}$**, or **$\mathcal{E}$-measurable**, if the inverse of any set we map to under f is a measurable set. That is, for any $\mathcal{A}$ in $\mathcal{E}$,

$$f^{-1}(\mathcal{A}) = \{x \in \Omega : f(x) \in \mathcal{A}\} \in \mathcal{F}.$$

This ensures the properties of our measures are preserved when we apply functions to them. An example of a measurable function we are familiar with is a random variable, although the definitions we have seen before may not state this explicitly.

Example 2.10. [11, p. 38]. A *random variable* is an $\mathcal{F}$-measurable function $X : \Omega \to E$ which maps from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to a measurable space $(E, \mathcal{E})$.

2.2 Stochastic processes

In mathematics, we have encountered many different maps, which can compare objects to one another with respect to a common structure, be it linear maps between vector spaces or isometries between metric spaces. Our maps of interest, the foundation of stochastic analysis, are measurable functions between probability spaces.

When we view these maps as random variables, we can group and index them with some set T, to obtain what is known as a stochastic process, a random process which changes over T. We will think of T as a continuous set representing time, although it can be a discrete set, and could represent other quantities, such as coordinates in space.

Definition 2.11. For $(E, \mathcal{E})$ a measurable space and T a non-empty set, a **stochastic process** X indexed by T is a collection of random variables $X = \{X_t : t \in T\}$ which map values from the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ it is defined on, to $(E, \mathcal{E})$ the **state space** of X.

Remark. Here, similarly to the indexing set T, the state space of X can either be continuous or discrete, but for the reason that Brownian motion has a continuous state space, we will only consider the continuous case.

3 Brownian motion

There are several rigorous methods to construct Brownian motion and prove its existence, such as exploiting properties of Gaussian random variables [13, pp. 9–12], or indexing a Gaussian process by a Hilbert space [14, pp. 16–19], concepts which lie beyond the scope of this essay. Instead, we introduce Brownian motion through its defining properties, similar to the observations brought forward by Einstein in 1905, before deriving further properties of the process.

First, the idea of an almost surely continuous function, one of the defining properties of Brownian motion, needs to be explained. The term *almost surely* means that an event has probability 1 of occurring, which can be interpreted for functions as follows.

Definition 3.1. [14, p. 17]. Let E be a topological space and $\mathcal{E} = \mathcal{B}(E)$ the Borel σ-algebra of E. A process X with values in the measurable space $(E, \mathcal{E})$ is said to be **almost surely continuous** if the function $t \mapsto X_t(\omega)$ is continuous for $\omega \in \tilde{\Omega} \subseteq \Omega$, where $\mathbb{P}(\tilde{\Omega}) = 1$.

Remark. This is defined similarly to the *μ-almost everywhere* convergence property covered in measure theory, with our defined choice of ω often referred to as *$\mathbb{P}$-almost all* in literature. The main idea behind these definitions is to show that a certain property holds for all but a negligible set of values.

Definition 3.2. [13, p. 7; 14, p. 17]. Let $B = \{B_t : t \geq 0\}$ be a stochastic process with state space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. We say B is called a **(linear) Brownian motion** if the following holds:

a) The process B has independent increments, i.e. for any sequence $0 = t_0 < t_1 < \ldots < t_n$ the random variables $B_{t_i} - B_{t_{i-1}}$ are independent for all $i = 1, 2, \ldots, n$.

b) The increments of B are normally distributed, i.e. for all $t \geq 0$ and $h > 0$, the increments $B_{t+h} - B_t$ are normally distributed with expectation 0 and variance h.

c) The process B is almost surely continuous.

We define the initial value, a real number, $x = B_0$. When $x = 0$, we say B is a **standard** Brownian motion.

3.1 Invariance properties

Proposition 3.3. *[14, pp. 20–21]. Let B be a standard linear Brownian motion. Then, the following properties hold:*

1. *(Time-homogeneity). For any $s > 0$, the process $B_{t+s} - B_s, t \geq 0$ is a Brownian motion independent of $\sigma(B_u, u \leq s)$;*
2. *(Symmetry). The process $-B_t, t \geq 0$ is a Brownian motion;*
3. *(Scaling). For every $a > 0$, the process $\frac{1}{a}B_{a^2t}, t \geq 0$ is a Brownian motion.*

Proof. We need to show that the defined properties of Brownian motion are conserved in each case to prove they are a Brownian motion. The property of continuity follows in each of these cases from the continuity of B.

Property 1 (Time-homogeneity). For any sequence $0 = t_0 < t_1 < \ldots < t_n$, we can calculate the increments of our process $X_t = B_{t+s} - B_s$ as

$$X_{t_i} - X_{t_{i-1}} = (B_{t_i+s} - B_s) - (B_{t_{i-1}+s} - B_s) = B_{t_i+s} - B_{t_{i-1}+s}.$$

It follows from the independent increments property of Brownian motion that these increments are independent of $\sigma(B_u, u \leq s)$. Finally, to check the increments are normally distributed, we calculate that for any $h > 0$,

$$X_{t+h} - X_t = (B_{t+h+s} - B_s) - (B_{t+s} - B_s) = B_{(t+s)+h} - B_{(t+s)}.$$

As $s > 0$ and $t \geq 0$, $s + t > 0$. Substituting $s + t$ for t in our normally distributed property of Brownian motion, we get that the increments of X are normally distributed. Therefore for any $s > 0$, the process $X = \{B_{t+s} - B_s : t \geq 0\}$ is a Brownian motion independent of $\sigma(B_u, u \leq s)$, as required.

Property 2 (Symmetry). Repeating the same steps with $X_t = -B_t$, for any sequence $0 = t_0 < t_1 < \ldots < t_n$,

$$X_{t_i} - X_{t_{i-1}} = (-B_{t_i}) - (-B_{t_{i-1}}) = -(B_{t_i} - B_{t_{i-1}}).$$

From the definition of Brownian motion, $B_{t_i} - B_{t_{i-1}}$ are independent for all $i = 1, 2, \ldots, n$, and negating all of the increments preserves this property. Normally distributed increments follow from the symmetric nature of the normal distribution. Hence, we have shown $X = \{-B_t : t \geq 0\}$ is a Brownian motion.

Property 3 (Scaling). Similarly, for any sequence $0 = \tilde{t_0} < \tilde{t_1} < \ldots < \tilde{t_n}$, we calculate the increments of $X_t = \frac{1}{a}B_{a^2t}$ to be

$$X_{\tilde{t_i}} - X_{\tilde{t}_{i-1}} = \left(\frac{1}{a}B_{a^2\tilde{t}_i}\right) - \left(\frac{1}{a}B_{a^2\tilde{t}_{i-1}}\right) = \frac{1}{a}(B_{a^2\tilde{t}_i} - B_{a^2\tilde{t}_{i-1}})$$

Due to the increments of Brownian motion being independent, multiplying each $\tilde{t}_i$ in our sequence by a^2 gives us $0 = a^2\tilde{t}_0 < a^2\tilde{t}_1 < \ldots < a^2\tilde{t}_n$, which can simply be relabelled to $0 = t_0 < t_1 < \ldots < t_n$ and remain a valid sequence. Scaling these increments by $\frac{1}{a}$ also preserves the independence of increments for $a > 0$, so the increments of X are independent. To prove the property of normally distributed increments, we see that for all $h > 0$,

$$X_{t+h} - X_t = \left(\frac{1}{a}B_{a^2(t+h)}\right) - \left(\frac{1}{a}B_{a^2t}\right) = \frac{1}{a}(B_{a^2t+a^2h} - B_{a^2t}).$$

Using our defined property of Brownian motion, we can say that $a(X_t)$ is normally distributed with expectation 0 and variance a^2h. Thus, X_t is normally distributed with expectation 0 and variance h, as required.

□

Remark. We can see a strong resemblance between these three properties of Brownian motion and the invariance properties of the Lebesgue measure seen previously. The relationship between Brownian motion and the Lebesgue measure is explained in the topic of stochastic integration, where integrals of stochastic processes are defined using limits which involve Lebesgue integration[2] over time.

Our final invariance property demonstrates how the long-term and short-term behaviour of Brownian motion are connected, through a non-linear transformation.

Theorem 3.4 (Time-inversion property)**.** *Let B be a standard Brownian motion. Then the process X defined by*

$$X_t = \begin{cases} 0 & \text{for } t = 0 \\ tB_{1/t} & \text{for } t > 0 \end{cases}$$

[2]An integral defined with respect to a measure.

is a standard Brownian motion.

Proof. Because $X_0 = 0$, our process X is a standard stochastic process. To show that the increments of X are normally distributed, we see that for $h > 0$,

$$\begin{aligned} X_{t+h} - X_t &= (t+h)B_{1/(t+h)} - tB_{1/t} \\ &= (t+h)B_{1/(t+h)} - t\big(B_{1/(t+h)} + \big(B_{1/t} - B_{1/(t+h)}\big)\big) \\ &= h\big(B_{1/(t+h)} - B_0\big) - t\big(B_{1/t} - B_{1/(t+h)}\big). \end{aligned}$$

Because $\frac{1}{t+h} > 0$ and $\frac{1}{t} > \frac{1}{t+h}$, we see that the increments of X are a linear combination of independent normally distributed variables. Therefore, increments of X are normally distributed.

We now check the expectation and variance of these increments. Firstly,

$$\mathbb{E}[X_{t+h} - X_t] = (t+h)\mathbb{E}\big[B_{1/(t+h)}\big] - t \cdot \mathbb{E}\big[B_{1/t}\big] = 0.$$

Here, the first equality follows from the linearity of expectation and the second equality arises from the fact that $B_u = B_u - B_0$ when B is a standard Brownian motion, thus B_u has expectation 0 and variance u for all $u \geq 0$. The variance of these increments can be expressed as

$$\operatorname{Var}(X_{t+h} - X_t) = \operatorname{Var}(X_{t+h}) + \operatorname{Var}(X_t) - 2\operatorname{Cov}(X_{t+h}, X_t).$$

We begin by calculating the variances $\operatorname{Var}(X_{t+h})$ and $\operatorname{Var}(X_t)$ in the above equation.

$$\operatorname{Var}(X_t) = \operatorname{Var}(tB_{1/t}) = t^2 \cdot \operatorname{Var}(B_{1/t}) = t^2\big(\tfrac{1}{t}\big) = t.$$

Similarly, $\operatorname{Var}(X_{t+h}) = t + h$. To calculate the covariance [13, p. 13], we use the fact that B_u has expectation 0 and variance u for all $u \geq 0$ several times.

$$\begin{aligned} \operatorname{Cov}(X_{t+h}, X_t) &= \operatorname{Cov}((t+h)B_{1/(t+h)}, tB_{1/t}), \\ &= t(t+h)\operatorname{Cov}(B_{1/(t+h)}, B_{1/t}), \\ &= t(t+h)\Big(\mathbb{E}\big[B_{1/(t+h)}B_{1/t}\big] - \mathbb{E}\big[B_{1/t}\big]\mathbb{E}\big[B_{1/(t+h)}\big]\Big), \\ &= t(t+h)\mathbb{E}\big[B_{1/(t+h)}B_{1/t}\big], \\ &= t(t+h)\mathbb{E}\big[B_{1/(t+h)}\big(B_{1/(t+h)} + B_{1/t} - B_{1/(t+h)}\big)\big], \\ &= t(t+h)\Big(\mathbb{E}\big[B^2_{1/(t+h)}\big] + \mathbb{E}\big[B_{1/(t+h)}\big]\mathbb{E}\big[B_{1/t} - B_{1/(t+h)}\big]\Big), \\ &= t(t+h)\mathbb{E}\big[B^2_{1/(t+h)}\big], \\ &= t(t+h)\Big(\mathbb{E}\big[B^2_{1/(t+h)}\big] - \big(\mathbb{E}\big[B_{1/(t+h)}\big]\big)^2\Big), \\ &= t(t+h)\operatorname{Var}\big(B_{1/(t+h)}\big), \\ &= t\big(\tfrac{t+h}{t+h}\big) = t. \end{aligned}$$

When we expand our expectation term above, we make the choice to factor out $B_{1/(t+h)}$ over $B_{1/t}$. We do this so that our later defined increment of $B_{1/t} - B_{1/(t+h)}$ is forward in time; its expectation can be evaluated to be zero.

Combining our results, we get that $\text{Var}(X_{t+h} - X_t) = t + h + t - 2t = h$, as required.

To show the independence of increments property for X, we see that

$$\text{Cov}(X_t, X_{t+h} - X_t) = \text{Cov}(X_t, X_{t+h}) - \text{Var}(X_t) = t - t = 0.$$

Because X_{t+h} and X_t are random normal variables, their covariance being zero implies that X_t and $X_{t+h} - X_t$ are independent. Generally speaking, we have shown that any given interval is independent of any previous interval, as needed.

Lastly, the property of continuity [2, p. 179] holds for any $t > 0$ due to the continuity of B, but it is unclear whether our stochastic process X_t is continuous at $t = 0$. To show continuity here, we evaluate the limit of our process towards zero as

$$\lim_{t \to 0} X_t = \lim_{t \to 0} tB_{1/t} = \lim_{n \to \infty} \frac{B_n}{n}, \text{ where } n = \tfrac{1}{t}.$$

Writing B_n as a sum of independent increments,

$$\lim_{t \to 0} X_t = \lim_{n \to \infty} \frac{\big(B_n - B_{\lfloor n \rfloor}\big) + \big(B_{\lfloor n \rfloor} - B_{n-1}\big) + \ldots + (B_1 - B_0)}{\lceil n \rceil}.$$

Here, we use the floor and ceiling functions to ensure non-integer values of n are supported. Looking at each of these increments, $\mathbb{E}[B_i - B_j] = 0$ whenever $i \geq j$, as well as being independent by our previously proven property. Therefore, we can apply the law of large numbers[3] to get our required result of

$$\lim_{t \to 0} X_t = 0.$$

Because we have shown all required properties hold, we can conclude that X is a standard Brownian motion. □

3.2 Nowhere differentiability

A property for which Brownian motion is famous is its non-differentiability at any fixed point. To demonstrate this, we will need to define what it means for a function to be differentiable. Whilst doing so, we cover needed

[3]For $X_1, X_2, \ldots$ an infinite sequence of independent random variables,

$$\lim_{n \to \infty} \frac{X_1 + \ldots + X_n}{n} = \mathbb{E}[X].$$

results in probability theory which describe the behaviour of sequences of random events — the Hewitt-Savage zero-one law and Fatou's lemma for sets. In the following definitions, the natural numbers $\mathbb{N}$ will represent the set $\{1, 2, 3, \ldots\}$.

Definition 3.5. [13, p. 19; 2, p. 88]. Let $X_1, X_2, \ldots$ be a sequence of random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and W a set of sequences such that the event $A = \{X_1, X_2, \ldots \in W\}$ is in $\mathcal{F}$. We say that A is **exchangeable** if, for any finite permutation $\sigma : \mathbb{N} \to \mathbb{N}$,

$$\{X_1, X_2, \ldots \in W\} = \{X_{\sigma(1)}, X_{\sigma(2)}, \ldots \in W\}.$$

Lemma 3.6 (Hewitt-Savage zero-one law)**.** *[13, p. 19]. If A is an exchangeable event for an independent, identically distributed sequence of random variables $X_1, X_2, \ldots$, then $\mathbb{P}(A)$ is 0 or 1.*

This result will allow us trivialise probabilities to just two distinct possibilities. We omit the proof of this here, due to its length.[4]

Next, we focus on the limiting behaviour of sequences of events, a key component of the derivatives we seek to define.

Definition 3.7. [9, p. 4; 15, p. 109]. Let $\{A_n : n \in \mathbb{N}\}$ be a sequence of events defined on a measurable space $(\Omega, \mathcal{F})$ such that $A_n \in \mathcal{F}$ for all n in $\mathbb{N}$. Then we define the following sets:

$$\liminf_{n\to\infty} A_n = \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} A_n = \{x : x \in A_n \text{ all but finitely often}\},$$
$$\limsup_{n\to\infty} A_n = \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n = \{x : x \in A_n \text{ infinitely often}\}.$$

Lemma 3.8 (Fatou's lemma for sets)**.** *For any sequence of events $\{A_n : n \in \mathbb{N}\}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the following inequalities hold:*

$$\mathbb{P}\Big(\liminf_{n\to\infty} A_n\Big) \leq \liminf_{n\to\infty} \mathbb{P}(A_n); \quad \mathbb{P}\Big(\limsup_{n\to\infty} A_n\Big) \geq \limsup_{n\to\infty} \mathbb{P}(A_n).$$

Proof. [9, p. 56; 15, p. 109]. Let $G_n = \bigcap_{n=m}^{\infty} A_m$, so that

$$G = \lim_{n\to\infty} G_n = \liminf_{n\to\infty} A_n.$$

We observe, by the continuity of probability,

$$\lim_{n\to\infty} \mathbb{P}(G_n) = \mathbb{P}\Big(\lim_{n\to\infty} G_n\Big) = \mathbb{P}(G). \tag{1}$$

[4] We can see Durrett [7, pp. 71–72] for a proof based on the analysis of symmetric differences of sets.

Also, because G_n is a subset of A_m for all $m \geq n$,

$$\mathbb{P}(G_n) \leq \mathbb{P}(A_m) \text{ for all } m \geq n,$$
$$\text{so } \mathbb{P}(G_n) \leq \inf_{m \geq n} \mathbb{P}(A_m). \tag{2}$$

Combining results (1) and (2), we see that

$$\lim_{n\to\infty} \mathbb{P}(G_n) = \mathbb{P}\Big(\liminf_{n\to\infty} A_n \Big) \leq \liminf_{n\to\infty} \mathbb{P}(A_n),$$

as needed. The second inequality follows similarly. □

Now that we have our needed results, we are ready to define and disprove the differentiability of Brownian motion. Due to the irregular nature of the process, we must first introduce a generalisation of the derivative, over the conventional one we are familiar with.

Definition 3.9. [10, pp. 109–110]. Let $f : [0, \infty) \to \mathbb{R}$ be a continuous function. We define

$$D^{\pm} f(t) = \limsup_{h \to 0\,\pm} \frac{f(t+h) - f(t)}{h} \text{ and } D_{\pm} f(t) = \liminf_{h \to 0\,\pm} \frac{f(t+h) - f(t)}{h}$$

as the upper (right and left) and lower (right and left) **Dini derivatives**, respectively. We say f is **differentiable** at $t > 0$ if all four Dini derivatives are finite numbers and equal. At $t = 0$, f is differentiable when $D^+ f(t)$ and $D_+ f(t)$ are both finite numbers and equal.

Theorem 3.10 (Nowhere differentiability)**.** *Let $B = \{B_t : t \in T\}$ be a standard Brownian motion. At any fixed $t \geq 0$, Brownian motion is not differentiable at t, almost surely.*

We follow the argument of Mörters and Peres [13, pp. 19–20], which makes use of our invariance properties of Brownian motion, in conjunction with results from probability theory we have shown, to prove this theorem.

Proof. Given our standard Brownian B, we construct another Brownian motion X, as defined in Theorem 3.4, our time-inversion property. Then

$$\begin{aligned} D^+ X_0 &= \limsup_{h \to 0\,+} \frac{X_h - X_0}{h}, \\ &\geq \limsup_{n\to\infty} n X_{1/n}, \\ &\geq \limsup_{n\to\infty} \sqrt{n} X_{1/n} = \limsup_{n\to\infty} \frac{B_n}{\sqrt{n}}. \end{aligned}$$

We claim that the final expression above is almost surely infinite. That is,

$$\mathbb{P}\left(\limsup_{n\to\infty} \frac{B_n}{\sqrt{n}} > N \right) = 1, \tag{3}$$

for any integer $N > 0$. To prove this, we first apply Fatou's lemma for sets and multiply out our fraction to get

$$\mathbb{P}\left(\limsup_{n\to\infty} \frac{B_n}{\sqrt{n}} > N \right) \geq \limsup_{n\to\infty} \mathbb{P}\Big(B_n > N\sqrt{n}\Big). \tag{4}$$

By the scaling property proven in Proposition 3.3 with $a = \frac{1}{\sqrt{n}}$, the process $\sqrt{n} \cdot B_{(1/n)n} = B_1\sqrt{n}$ is a Brownian motion. Substituting this into (4), we see that our probability becomes $\mathbb{P}(B_1 > N)$, which is positive for all integers $N > 0$. Thus, our probability in (3) is greater than zero.

Returning to our equation (3), we use our alternative definition of lim sup to see that

$$\begin{aligned} \mathbb{P}\left(\limsup_{n\to\infty} \frac{B_n}{\sqrt{n}} > N \right) &= \mathbb{P}(B_n > N\sqrt{n} \text{ infinitely often}), \\ &= \mathbb{P}\left(\sum_{i=1}^{n} (Y_i) > N\sqrt{n} \text{ infinitely often} \right), \end{aligned} \tag{5}$$

where $Y_i = B_i - B_{i-1}$ for $i = 1, 2, \ldots, n$. Because the event in (5) is exchangeable, with $Y_1, Y_2, \ldots, Y_n$ an independent and identically distributed sequence of random variables (as intervals of a Brownian motion), we can apply the Hewitt-Savage zero-one law and evaluate our probability to be either zero or one. Since we have shown our probability is greater than zero, it must be one, so our proposed equality (3) holds.

Therefore, $D^+X_0 = \infty$ almost surely. Similarly, we get the result of $D_+X_0 = -\infty$ almost surely, so X is almost surely not differentiable at 0. By the time-homogeneity property of Brownian motion in Proposition 3.3, $X_s = B_{t+s} - B_t$ is a standard Brownian motion for any fixed $t > 0$. Hence,

$$\frac{X_s - X_0}{s} = \frac{B_{t+s} - B_t}{s}$$

shows us that the differentiability of X at zero is equivalent to the differentiability of B at t, extending our result from $t = 0$ to any fixed $t \geq 0$. Thus, Brownian motion is not differentiable at any fixed $t \geq 0$, almost surely. □

4 Martingales

Martingales are stochastic processes which have a notion of *fairness*. Their defining property is one in which the expected future value of the process, given all previous observations, is equal to the current value [3, p. 14]. To formally introduce this, we need a mathematical description for the information known at a given time, which a filtration provides.

Definition 4.1. [4, p. 2; 10, p. 3]. Given a measurable space $(\Omega, \mathcal{F})$, a collection of σ-algebras indexed by t, $\{\mathcal{F}_t : t \geq 0\}$ (which we will shorten to $\{\mathcal{F}_t\}$), on our given measurable space is called a **filtration** if

$$\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F} \text{ for every } 0 \leq s < t < \infty.$$

Equipping a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a filtration, we obtain a **filtered probability space**, the quad tuple

$$(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P}).$$

Remark. By definition, random variables are $\mathcal{F}$-measurable functions. However, this is not sufficient once we move from a probability space to a filtered probability space of a stochastic process, as our stochastic process is not guaranteed to be a measurable function with respect to $\{\mathcal{F}_t\}$.

For this reason, we introduce the concept of adaptation to a filtration.

Definition 4.2. [10, p. 4; 14, p. 42]. We say that a stochastic process $X = \{X_t : t \in T\}$ is **adapted** to a given filtration $\{\mathcal{F}_t\}$ if for every t in T, X_t is an $\mathcal{F}_t$-measurable random variable.

Example 4.3. [10, p. 3; 14, p. 42]. Given any stochastic process $X = \{X_t : t \in T\}$, the natural filtration of X, $\mathcal{F}_t^X$, also called the filtration generated by X, is the smallest filtration to which X is adapted, generated by the σ-algebra of the stochastic process up to t, i.e.,

$$\mathcal{F}_t^X = \sigma\{X_s : s \leq t\}.$$

Having introduced the concept of a filtration and how a stochastic process can be *equipped* with a filtration, we can now define a martingale.

Definition 4.4. [4, pp. 7–8; 16, pp. 71–72]. Let $X = \{X_t : t \in T\}$ be a stochastic process defined on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$. We say that X is a **martingale** if the following properties are true:

1. X is adapted to $\{\mathcal{F}_t\}$.

2. $\mathbb{E}|X_t| < \infty$ for every t in T.

3. $\mathbb{E}[X_t|\mathcal{F}_s] = X_s$ almost surely for every $0 \leq s < t$ in T.

Remark. Although not explicitly stated, when defining a martingale here, we make the assumption that $t \geq 0$ for all t in T, as this is required for our definition of a filtration.

Proposition 4.5. *Let $B = \{B_t : t \geq 0\}$ be a standard Brownian motion equipped with natural filtration $\mathcal{F}_t^B$. Then B is a martingale.*

Proof. By definition, B is adapted to its natural filtration, so the first property holds.

Since $B_t - B_0 = B_t$ for a standard Brownian motion, we know B_t is normally distributed with expectation 0 and variance t. Therefore,

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

We then use the Cauchy-Schwarz inequality[5] to give us our second property of

$$\mathbb{E}|B_t| \leq \sqrt{\mathbb{E}[B_t^2]} = \sqrt{t} < \infty.$$

For the third and final property [13, p. 54], let $0 \leq s < t$. Then

$$\mathbb{E}[B_t|\mathcal{F}_s] = \mathbb{E}[B_t + B_s - B_s|\mathcal{F}_s] = \mathbb{E}[B_t - B_s|\mathcal{F}_s] + B_s. \tag{6}$$

We now use the property of independent increments for Brownian motion on (6), which tells us that $B_t - B_s$ is independent of $\sigma(B_u : u \leq s) = \mathcal{F}_s$, to get

$$\mathbb{E}[B_t|\mathcal{F}_s] = \mathbb{E}[B_t - B_s] + B_s. \tag{7}$$

Because the increments of the Brownian motion $B_t - B_s$ are normally distributed with expectation 0, we know that $\mathbb{E}[B_t - B_s] = 0$. Applying this to (7) gives us

$$\mathbb{E}[B_t|\mathcal{F}_s] = B_s \text{ for every } 0 \leq s < t.$$

As B satisfies all three necessary properties, we conclude that standard Brownian motion, equipped with its natural filtration, is a martingale. □

5 Markov processes

Whilst the defining characteristic of a martingale is concerning the future expected values of a process, the defining property of a Markov process — the (simple) Markov property — focuses on the future distribution as a whole. A Markov process can be thought of as a *memoryless* stochastic process, in which past observations before a current value do not influence future values.

Abstracted definitions for a Markov process can be given, but to understand the underlying mathematics necessitates a more structural approach. Under a discrete time frame, the structure of a Markov process is often described by a matrix which represents the probabilities of reaching all other states from each individual state. When extended to a continuous time frame, we use a transition function in place of a matrix for this same purpose. We first define the transition probabilities which make up these functions.

[5] For X, Y random variables, $|\langle X, Y\rangle|^2 \leq \langle X, X\rangle \cdot \langle Y, Y\rangle$, with $\langle X, Y\rangle = \mathbb{E}[XY]$.

Definition 5.1. [4, p. 4; 14, pp. 79–80]. Let $(E, \mathcal{E})$ be a measurable space. A map $P : E \times \mathcal{E} \to [0, 1]$ is called a **transition probability** if:

1. For every x in E, the map $A \mapsto P(x, A)$ is a measure on $\mathcal{E}$;

2. For every A in $\mathcal{E}$, the map $x \mapsto P(x, A)$ is $\mathcal{E}$-measurable.

3. For every x in E, $P(x, E) = 1$.

Remark. A map which has the first two above properties, with codomain $[0, \infty]$, is called a *kernel*; the third property ensures our kernel is a probability measure for every x, restricting our codomain to $[0, 1]$. Consequently, transition probabilities are also referred to as probability kernels or Markov kernels.

Definition 5.2. [5, p. 6; 14, p. 80]. A **(homogeneous) transition function** is a collection of transition probabilities $\{P_t : t \in T\}$, defined on a measurable space $(E, \mathcal{E})$, with an additional consistency condition:

$$P_{t+s}(x, A) = \int_E P_t(y, A) P_s(x, dy) \text{ for all } x \text{ in } E,\ A \text{ in } \mathcal{E},\ t, s \geq 0.$$

Remark. This relation is called the Chapman-Kolmogorov equation, a consequence of the Markov property which we will define shortly.

Homogeneity of the transition function refers to the assumption that the transition probabilities depend on only the current time s and future time t through the elapsed time $t - s$, meaning that the behaviour is invariant under shifts in t.

Definition 5.3. [5, p. 6; 4, p. 3; 1, p. 8]. Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ be a filtered probability space. An adapted stochastic process $X = \{X_t : t \in T\}$ is a **(homogeneous) Markov process** with respect to $\{\mathcal{F}_t\}$, with transition function $P_t : E \times \mathcal{E} \to [0, 1]$, if it satisfies the Markov property. That is, for all A in $\mathcal{E}$ and $t > s \geq 0$,

$$\mathbb{P}(X_t \in A | \mathcal{F}_s) = P_{t-s}(X_s, A) \quad \text{almost surely.} \tag{8}$$

Proposition 5.4. *[15, p. 5]. Let $B = \{B_t : t \geq 0\}$ be a Brownian motion, equipped with natural filtration $\mathcal{F}_t^B$ and transition function*

$$P_t(x, A) = \begin{cases} \int_A p_t(x, y)\, dy & \text{for } t > 0; \\ \mathbb{1}_{\{x \in A\}} & \text{for } t = 0, \end{cases}$$

$$\text{where } p_t(x, y) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{(x - y)^2}{2t} \right),$$

for all x in $\mathbb{R}$ and A in $\mathcal{B}(\mathbb{R})$. Then B is a Markov process.

Proof. Fix $t > s \geq 0$. We will show that the Markov property (8) holds for a general bounded $\mathcal{B}(\mathbb{R})$-measurable function $f : \mathbb{R} \to \mathbb{R}$, before specifying our required $f(x) = \mathbb{1}_{\{x \in A\}}$ [5, p. 7]. Therefore, the general result we need to show is

$$\mathbb{E}[f(B_t)|\mathcal{F}_s] = P_{t-s}(B_s, f) \quad \text{almost surely,}$$

with our transition function generalising [15, p. 5] to

$$P_t(x, f) = \begin{cases} \int_{\mathbb{R}} p_t(x, y) f(y)\, dy & \text{for } t > 0; \\ f(x) & \text{for } t = 0. \end{cases}$$

We start by expanding out

$$\mathbb{E}[f(B_t)|\mathcal{F}_s] = \mathbb{E}[f(B_s + (B_t - B_s))|\mathcal{F}_s].$$

By the time-homogeneity property we proved in Proposition 3.3, $B_t - B_s$ is independent of $\sigma(B_u : u \leq s) = \mathcal{F}_s$, with expectation 0 and variance $t - s$ by definition. As our Brownian motion B is equipped with its natural filtration $\mathcal{F}_t^B$, B is an adapted stochastic process, so B_s is $\mathcal{F}_s$-measurable. Thus, we can restrict our conditional expectation and further simplify to get

$$\mathbb{E}[f(B_t)|\mathcal{F}_s] = \mathbb{E}[f(B_s + (B_t - B_s))|B_s] = \mathbb{E}[f(B_s + Z)] \quad \text{(almost surely),}$$

where Z is a normally distributed variable of expectation 0 and variance $t - s$. Calculating this expectation gives us

$$\mathbb{E}[f(B_t)|\mathcal{F}_s] = \int_{\mathbb{R}} f(B_s + x) f_Z(x) dx,$$

where $f_Z(x)$ denotes the distribution of Z. Substituting $y = B_s + x$, and writing out the probability density function of the normal distribution, we get

$$\begin{aligned} \mathbb{E}[f(B_t)|\mathcal{F}_s] &= \int_{\mathbb{R}} f(y) \frac{1}{\sqrt{2\pi(t-s)}} \exp\left(- \frac{(B_s - y)^2}{2(t-s)} \right) dy, \\ &= \int_{\mathbb{R}} f(y) p_{t-s}(B_s, y)\, dy, \\ &= P_{t-s}(B_s, f). \end{aligned}$$

Because our choice of $t > s \geq 0$ was arbitrary, specifying $f(x) = \mathbb{1}_{\{x \in A\}}$ for any $A \in \mathcal{B}(\mathbb{R})$, we get our desired result of

$$\mathbb{P}(B_t \in A|\mathcal{F}_s) = P_{t-s}(B_s, A) \quad \text{almost surely.}$$

Therefore, when equipped with a natural filtration $\mathcal{F}_t^B$ and a transition function $P_t(x, A)$ as defined, Brownian motion is a Markov process. □

References

[1] Michel Benaïm and Tobias Hurth. *Markov chains on metric spaces: a short course*. Springer, 2022.

[2] Rabi Bhattacharya and Edward C Waymire. *A basic course in probability theory*. Springer, 2007.

[3] Rabindra Nath Bhattacharya and Edward C Waymire. *Random walk, Brownian motion, and martingales*, volume 292. Springer, 2021.

[4] Andrei N Borodin and Paavo Salminen. *Handbook of Brownian motion-facts and formulae*. Birkhäuser, 2012.

[5] Kai Lai Chung and John B Walsh. *Markov processes, Brownian motion, and time symmetry*. Springer, 2005.

[6] Donald L. Cohn. *Measure theory*. Birkhäuser, 1980.

[7] Rick Durrett. *Probability: theory and examples*, volume 49. Cambridge university press, 2019.

[8] Liviu C. Florescu. *Lebesgue integral*. Birkhäuser, 2021.

[9] Allan Gut. *Probability: a graduate course*, volume 75. Springer, 2013.

[10] Ioannis Karatzas and Steven Shreve. *Brownian motion and stochastic calculus*. Springer, 2014.

[11] Carlos S. Kubrusly. *Essentials of measure theory*. Springer, 2015.

[12] Zenghu Li. *Measure-valued branching Markov processes*, volume 103. Springer, 2023.

[13] Peter Mörters and Yuval Peres. *Brownian motion*, volume 30. Cambridge University Press, 2010.

[14] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293. Springer Science & Business Media, 2013.

[15] L Chris G Rogers and David Williams. *Diffusions, Markov processes, and martingales*, volume 2. Cambridge university press, 2000.

[16] James Yeh. *Martingales and stochastic analysis*, volume 1. World Scientific, 1995.