BROWNIAN MOTION

A Bachelor Thesis for Mathematics and Applications on the construction and properties of Brownian motion

Written by

Hannah Oudshoorn

Supervised by

Dr. K. Dajani

January 15, 2021

Universiteit Utrecht

Preface

I would like to start this thesis by thanking my supervisor Dr. Karma Dajani for her guidance through each state of the process, her expertise, and her very welcome motivational words given to me. Without her help, I would not have come so far!

Contents

Chapter 1

Introduction

Almost 200 years ago, in 1827, Scottish botanist Robert Brown was the first to describe the physical phenomenon of the random motion of particles suspended in a liquid or gas. Looking through his microscope at pollen grain suspended in water, he notices that the pollen undergo a type of random walk: if we were to plot the motion of one such particle on a graph, random zig-zagging motion is observed. This phenomenon, which we now refer to as *Brownian motion*, provoked a lot of discussion at the time of discovery of what caused these particles to move and remained uncertain for a long time. It wasn't until 1905 that Albert Einstein showed that Brownian motion is the result of the erratic collision of solvent particles against solution particles.

Unaware of the discovery of Roberts Brown, Louis Bachelier worked out a theory for the fluctuations of the stock market that involved the development of a type stochastic process called a Brownian motion. He did so for his Ph.D thesis in 1900 at the University of Paris under the supervision of the mathematician Poincaré. Even though Bachelier's work went largely unnoticed for nearly a century, he did discover two deep and fundamental facts about Brownian motion: given the position of the process at time t, you do not need the prior positions to predict or simulate the future behavior of the process; and two, that it has a reflection property. [1]

In 1923, the mathematician Norbert Wiener proved the existence of Brownian motion and set down a firm mathematical foundation for its further development and analysis. Finally, in 1939, the classical development of Brownian motion was completed by Paul Lévy.

Mathematically, Brownian motion can be thought of as a continuous time process in which over every infinitely small time interval $\Delta(t)$, the entity under consideration moves one "step" in a certain direction, which suggests that Brownian motion can be viewed as a random walk process. The aim of this thesis is to gain a thorough understanding of the construction and properties of the one-dimensional Brownian motion. [11]

The outline of this thesis is as follows. We first propose a theoretical framework consisting of relevant measure and probability theory in Chapter 2. Then, in Chapter 3 we introduce the one-dimensional symmetric random walk and show how we can use this symmetric random walk to construct the Brownian motion.
This is followed by Chapter 4, in which we introduce the formal definition of a one-dimensional Brownian motion and look at some basic and distributional properties of Brownian motion. In addition to this, Chapter 4 will also address the non-differentiability of Brownian motion. In Chapter 5, we take a careful look at the Markov property for Brownian motion, some of its properties and how we can use this property on the local and global maxima of Brownian motion. In addition to this, we discuss the strong Markov property and show how we can use this to prove the reflection principle. This thesis will be concluded in Chapter 6, which will contain an overview of the entire thesis and

[1] Both of these facts about Brownian motion will be discussed in Chapter 5 of this thesis.

Chapter 2

Theoretical framework

The aim of this chapter is to establish background knowledge for this thesis. It is based on the book of Schreve (2010) ([2]) and the book of Schilling (2005) ([3]). If the reader is familiar with measure theory and probability theory, the first three sections of this chapter can safely be passed. If the reader is familiar with filtrations and martingales, the fourth section of this chapter may be skipped.

2.1 Measure Theory

Definition 2.1.1. *(σ - **algebra**) Let Ω be a nonempty set and let $\mathcal{F}$ be a collection of subsets of Ω. We say that $\mathcal{F}$ is a σ-algebra provided that:*

i. the empty set $\emptyset$ belongs to $\mathcal{F}$,

ii. whenever a set A belongs to $\mathcal{F}$, its complement A^c also belongs to $\mathcal{F}$,

iii. Whenever a sequence of sets $A_1, A_2, \ldots$ belongs to $\mathcal{F}$, then their union and intersection $\bigcup_{i=1}^{\infty} A_i$, $\bigcap_{i=1}^{\infty} A_i$ respectively, also belongs to $\mathcal{F}$.

Definition 2.1.2. *A measure μ on Ω is a function $\mu : \mathcal{F} \to [0, \infty) \cup \{+\infty\}$ defined on a σ-algebra $\mathcal{F}$ satisfying the following conditions:*

i. $\mu(\emptyset) = 0$,

ii. For any countable family of pairwise disjoints sets $(A_j)_{j\in\mathbb{N}} \subset \mathcal{F}$,

$$\mu(\bigcup_{j\in\mathbb{N}} A_j) = \sum_{j\in\mathbb{N}} \mu(A_j)$$

which we call σ-additivity

When $\mu(\Omega) = 1$, then μ is called a probability measure and we denote the measure $\mathbb{P}$.

Definition 2.1.3. *Let Ω be a set and $\mathcal{F}$ be a σ-algebra on Ω. The pair $(\Omega, \mathcal{F})$ is called measurable space. If μ is a measure on Ω, $(\Omega, \mathcal{F}, \mu)$ is called measure space. If $\mathbb{P}$ is a probability measure, we call the triple $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space. In this case, we call Ω a sample space.*

Definition 2.1.4 (**Almost surely**)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. If a set $A \in \mathcal{F}$ satisfies $\mathbb{P}(A) = 1$, we say that the event A occurs almost surely.*

Definition 2.1.5 (**Borel σ-algebra**)**.** *The Borel σ-algebra is the smallest σ-algebra containing all the open subsets of $\mathbb{R}^n$. We denote the Borel σ-algebra by $\mathcal{B}(\mathbb{R})$. We call B a Borel set if $B \in \mathcal{B}$.*

Definition 2.1.6 (**Measurable function**)**.** *Let $(\Omega, \mathcal{F})$ be a measurable space. A function $f : \Omega \to \mathbb{R}$ is said to be an $\mathcal{F}$-measurable if $f^{-1}(B) \in \mathcal{F}$ for every $B \in \mathcal{B}(\mathbb{R})$, where we define $\{f \in B\} := f^{-1}(B) = \{\omega \in \Omega : f(\omega) \in B\}$*

Definition 2.1.7 (Indicator function). *The indicator function of a set $A \subset \Omega$ is a function*

$$\chi_A : \Omega \to \mathbb{R}$$

defined as

$$\chi_A(x) := \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A. \end{cases} \tag{2.1}$$

Example 2.1.8. *Let $(\Omega, \mathcal{F})$ be a measurable space. If $A \in \mathcal{F}$, then the indicator function is a measurable function. In order to see this, we take look at all the possible pre-images of the indicator function:*

$$\chi_A^{-1}(\emptyset) = \emptyset, \chi_A^{-1}(\mathbb{R}) = \Omega$$

$$\chi_A^{-1}(\{1\}) = A, \chi_A^{-1}(\{0\}) = A^c.$$

Now note that by Definition 2.1.1 that all the above values are contained in the σ-algebra $\mathcal{F}$. Therefore, we can conclude that the indicator function is a measurable function.

2.1.1 Additional results from Measure Theory

The following properties follow immediately from Definition 2.1.2. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $A, B \in \mathcal{F}$. Then

i. If $A \subset B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$.

ii. $\mathbb{P}(A \cup B) + \mathbb{P}(A \cap B) = \mathbb{P}(A) = \mathbb{P}(B)$

iii. $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$.

iv. If $A_1, A_2, ..., A_N$ are finitely many disjoint sets in $\mathcal{F}$, then

$$\mathbb{P}\Big(\bigcup_{n=1}^{N} A_n\Big) = \sum_{n=1}^{N} \mathbb{P}(A_n).$$

v. If $A_1, A_2, ...$ is a sequence in $\mathcal{F}$, then

$$\mathbb{P}\Big(\bigcup_{n=1}^{\infty} A_n\Big) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n)$$

i.e. $\mathbb{P}$ is σ-subadditive.

Theorem 2.1.9. *Let $(\Omega, \mathcal{F})$ be a measurable space. A function $\mathbb{P} : \mathcal{F} \mapsto [0, 1]$ is a probability measure if and only,*

i. $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(\Omega) = 1$.

ii. $\mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B)$ *for all* $A, B \in \mathcal{F}$ *with* $A \cap B = \emptyset$

iii. (Continuity of measures from below)
For any increasing sequence $(A_j)_{j \in \mathbb{N}} \subset \mathcal{F}$ with $A_j \uparrow A \in \mathcal{F}$ with limit $A = \cup_j A_j$, we have

$$\mathbb{P}(A) = \lim_{j \to \infty} \mathbb{P}(A_j)$$

Since $\mathbb{P}(A) < \infty$ for all $A \in \mathcal{F}$, the last item is equivalent with the following condition

(Continuity of measures from above)
For any decreasing sequence $(A_j)_{j \in \mathbb{N}} \subset \mathcal{F}$ with $A_j \downarrow A \in \mathcal{F}$ with limit $A = \cap_j A_j$, we have

$$\mathbb{P}(A) = \lim_{j \to \infty} \mathbb{P}(A_j)$$

Lemma 2.1.10 (Borel-Cantelli lemma). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(A_n)_n$ a sequence in $\mathcal{F}$, i.e., a sequence of events. Suppose that $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$. Then*

$$\mathbb{P}\Big(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m\Big) = 0$$

i.e., the event $\{\omega \in \Omega : \omega \in A_n$ for infinitely many $n\}$ has probability zero. So almost surely, $\omega \notin A_m$ for all m sufficiently large.

Proof. We start with the fact that $\cup_{m=n}^{\infty} A_m$ is a decreasing sequence. To see this, note that if you increase n, you take smaller and smaller unions. By continuity of any probability measure from below and since $\mathbb{P}$ is σ-subadditive, we have

$$\mathbb{P}\Big(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m\Big) = \lim_{n\to\infty} \mathbb{P}\Big(\bigcup_{m=n}^{\infty} A_m\Big) \tag{2.2}$$

$$\leq \lim_{n\to\infty} \sum_{m=n}^{\infty} \mathbb{P}(A_m) = 0 \tag{2.3}$$

Where we used in the that if $\sum_{n=1}^{\infty} a_n < \infty$, then $\lim_{m\to\infty} \sum_{n=m}^{\infty} a_n = 0$. ■

2.2 Probability Theory

Definition 2.2.1 (Random variable). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A random variable X is an $\mathcal{F}$-measurable subset of Ω. See Figure 2.1.*

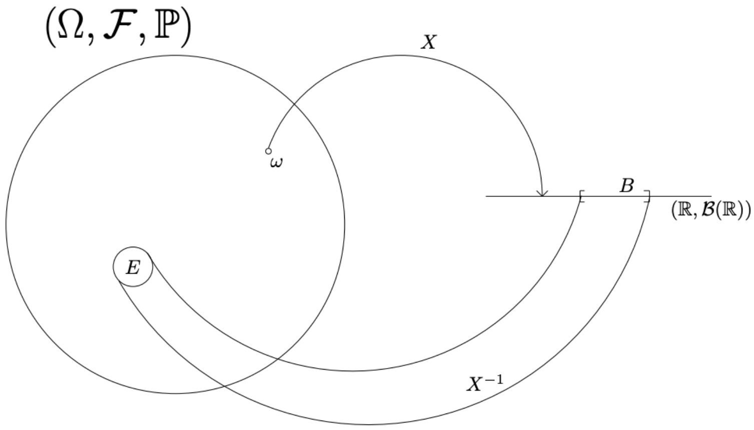

Figure 2.1: A random variable $X : \Omega \mapsto \mathbb{R}$. The pre-image of a Borel set B is an event $E \in \Omega$.[6]

Definition 2.2.2. *Let X be a random variable defined on a nonempty sample space Ω. The σ-algebra generated by X, denoted by $\sigma(X)$ is the smallest σ-algebra on Ω making X measurable. We have*

$$\sigma(X) = \big\{\{X \in B\} : B \in \mathcal{B}(\mathbb{R})\big\}$$

Definition 2.2.3. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let X be a random variable in Ω. Let $\mathcal{G} \subseteq \mathcal{F}$ be a sub-σ-algebra. If $\sigma(X) \subseteq \mathcal{G}$, then we say that X is $\mathcal{G}$-measurable.*

Heuristically, A random variable X is $\mathcal{G}$-measurable if and only if the information in $\mathcal{G}$ is sufficient to determine the value of X.

Definition 2.2.4. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space, and let* $\mathcal{G}$ *and* $\mathcal{H}$ *be sub-*σ*-algebras of* $\mathcal{F}$ *(i.e., the sets in* $\mathcal{G}$ *and the sets in* $\mathcal{H}$ *are also in* $\mathcal{F}$*). We say these two* σ*-algebras are independent if*

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B), \ \forall A \in \mathcal{G}, B \in \mathcal{H}.$$

Let X *and* Y *be random variables on* $(\Omega, \mathcal{F}, \mathbb{P})$ *. We say these two random variables are independent if the* σ*-algebras they generate,* $\sigma(X)$ *and* $\sigma(Y)$*, are independent. We say that the random variable* X *is independent of the* σ*-algebra* $\mathcal{G}$ *if* $\sigma(X)$ *and* $\mathcal{G}$ *are independent.*

2.3 Conditional expectations

Definition 2.3.1 (Expectation). *Let* X *be a random variable on a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *. The expectation (or expected value) of* X *is defined to be the Lebesgue integral*

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

provided that X *is integrable, i.e., if*

$$\mathbb{E}[|X|] = \int_{\Omega} |X(\omega)| d\mathbb{P}(\omega) < \infty$$

or if $X \geq 0$ *almost surely. In the latter case,* $\mathbb{E}[X]$ *take on the value* ∞*.*

Definition 2.3.2 (Conditional expectation given a σ-algebra $\mathcal{G}$). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$, and let X be a random variable that is either nonnegative or integrable. The conditional expectation of X given $\mathcal{G}$, denoted $\mathbb{E}[X|\mathcal{G}]$, is an almost surely unique random variable that satisfies*

i. $\mathbb{E}[X|\mathcal{G}]$ *is $\mathcal{G}$-measurable, and*

ii. $\int_A \mathbb{E}[X|\mathcal{G}](\omega)d\mathbb{P}(\omega) = \int_A X(\omega)d\mathbb{P}(\omega)$ *for all $A \in \mathcal{G}$*

Theorem 2.3.3. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let be a sub-σ-algebra of $\mathcal{F}$. Then the following properties hold:*

i. ***(Linearity of conditional expectations)*** *If X and Y are integrable random variables and c_1 and c_2 are constants, then*

$$\mathbb{E}[c_1X + c_2Y|\mathcal{G}] = c_1\mathbb{E}[X|\mathcal{G}] + c_2\mathbb{E}[Y|\mathcal{G}].$$

ii. ***(Taking out what is known)*** *If X and Y are integrable random variables, Y and XY are integrable, and X is $\mathcal{G}$-measurable, then*

$$\mathbb{E}[XY|\mathcal{G}] = X\mathbb{E}[Y|\mathcal{G}].$$

iii. ***(Iterated conditioning)*** *If $\mathcal{H}$ is a sub-σ-algebra of $\mathcal{G}$ and X is an integrable random variable, then*

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}].$$

iv. ***(Independence)*** *If X is integrable and independent of $\mathcal{G}$, then*

$$\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X].$$

v. ***(Conditional Jensen's inequality)*** *If $\phi(x)$ is a convex function of a dummy variable x and X is integrable, then*

$$\mathbb{E}[\phi(X)|\mathcal{G}] \geq \phi(\mathbb{E}[X|\mathcal{G}])$$

2.4 Stochastic processes

The modelling of real world phenomena often requires the description of a quantity that evolves over time in a random manner. Examples of these phenomena are calls arriving at a help desk, the number of people waiting in a queue or the price of a stock market.

Mathematically, such an object is described by a stochastic process as follows:

Definition 2.4.1 (Stochastic process). *A stochastic process is a parameterized collection of random variables*

$$\{X(t) : t \in T\}$$

indexed by a set T, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and assuming values in $\mathbb{R}^n$

- *If T is a discrete set, the process is called a discrete-time stochastic process.*
- *If $T = \mathbb{R}_+$ or a sub-interval of this, the process is called a continuous-time stochastic process*
- *For any $\omega \in \Omega$ we call the map $t \mapsto X(t)(\omega)$ a trajectory of the process.*
- *For $s < t$ we call $X(t) - X(s)$ the increment of the process over the time interval (s, t). We call these increments stationary if the distribution of $X(t + s) - X(s)$ only depends on t.*

Now we will discuss the concept of a filtration. Informally speaking, A filtration $\mathcal{F}(t)$ contains any information that could possibly be asked and answered for the considered stochastic process at time t. As a random experiment progresses, and new information becomes available, you know which part of the σ-algebra you already know, so some part of the σ-algebra will be completely revealed. Let us first illustrate this with an example, before giving a formal definition.

Example 2.4.2. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and suppose Ω is the set of the eight possible outcomes of three fair coin tosses, i.e.*

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\},$$

and $\mathcal{F} = \mathcal{P}(\Omega)$, where $\mathcal{P}(\Omega)$ denotes the power set of Ω, i.e. the set of all subsets of Ω. A friend has tossed the coin three times, observing the true $\omega \in \Omega$. Before the friend tells us anything about his throws, no specific information about the true value of w is resolved. We denote

$$\mathcal{F}(0) = \{\emptyset, \Omega\}.$$

Now we are told the outcome of the first throw, for example heads. Now we know only one of the following four events can be our true ω: $\{HHH\}, \{HHT\}, \{HTH\}, \{HTT\}$. We denote the set that contains these for events A_H. In a similair way, if the first coin is tail, the four possibilities are $\{THH\}, \{THT\}, \{TTH\}, \{TTT\}$, which we denote by A_T. Note that A_T is the complement of A_H.
These two revealed sets together with the information that is always resolved, form the σ-algebra

$$\mathcal{F}(1) = \{\emptyset, \Omega, A_H, A_T\}$$

and it contains the information learned by observing the fist coin toss. So, if we know the outcome of the first coin toss, we can tell of each set in $\mathcal{F}(1)$ whether or not ω is an element of it.
If we are also told the second coin toss, we obtain a finer resolution. In particular, the four sets

$$A_{HH} = \{HHH, HHT\}, A_{HT} = \{HTH, HTT\},$$
$$A_{TH} = \{THH, THT\}, A_{TT} = \{TTH, TTT\},$$

are revealed. Whenever a set is resolved, so is its complement. Whenever two sets are resolved, so is their union. The triple unions are also resolved. We find 16 resolved sets that together form a σ-algebra we call $\mathcal{F}(2)$; i.e.,

$$\mathcal{F}(2) = \{\emptyset, \Omega, A_H, A_T, A_{HH}, A_{HT}, A_{TH}, A_{TT}, A^c_{HH}, A^c_{HT}, A^c_{TH},$$
$$A^c_{TT}, A_{HH} \cup A_{TH}, A_{HH} \cup A_{TT}, A_{HT} \cup A_{TH}, A_{HT} \cup A_{TT}\}$$

Keeping in mind Definition 2.1.1, we see that this is indeed a σ-algebra. We shall think of this σ-algebra as containing the information learned by observing the first two coin tosses.
If we were told all three coin tosses, we would know the true ω and every subset of Ω is revealed. Therefore,

$$\mathcal{F}(s) = \mathcal{P}(\Omega).$$

Note that $\mathcal{F}(0) \subset \mathcal{F}(1) \subset \mathcal{F}(2) \subset \mathcal{F}(3)$.

Let us now define a filtration formally.

Definition 2.4.3 (Filtration). *[4]*
Let $(\Omega, \mathcal{F})$ be a measurable space. Consider a collection of sub-σ-algebras $\{F(t) : t \in I\}$ of $\mathcal{F}$, indexed by $I \subset \mathbb{R}$.(For example, $I = \{0, 1, \ldots, n\}, I = \mathbb{N}, I = [0, T]$ and $I = [0, \infty[$. The index t represents time in general). If $\mathcal{F}(s) \subset \mathcal{F}(t) \subset \mathcal{F}$ for $s, t \in I$, such that $s \leq t$, then $\{\mathcal{F}(t) : t \in I\}$ is called a filtration. We assume that 0 is the smallest element in I and $\mathcal{F}(0) = \{\emptyset, \Omega\}$.

A filtration is interpreted as a monotone increment of information as time passes by.

Definition 2.4.4 (Adapted process). *[4]*
Consider a filtration $\{F(t) : t \in I\}$ and a stochastic process $\{X(t) : t \in T\}$. If $X(t)$ is measurable with respect to $\mathcal{F}(t)$ for every t, then $\{X(t) : t \in I\}$ is said to be adapted to the filtration $\{\mathcal{F}(t) : t \in I\}$

The importance of adapted processes lies in the following definition:

Definition 2.4.5 (**Martingale**). *[4]*
Suppose that a stochastic process $\{X(t) : t \in I\}$ is adapted to a filtration $\{F(t) : t \in I\}$, and that $X(t)$ is integrable for every t. If

$$\mathbb{E}[M(t)|\mathcal{F}(s)] = M(s) \textit{ for all } 0 \leq s \leq t,$$

We say this $\{X(t) : t \in T\}$ is a martingale with respect to $\{\mathcal{F}(t) : t \in I\}$. It has no tendency to rise or fall. If

$$\mathbb{E}[M(t)|\mathcal{F}(s)] \geq M(s) \textit{ for all } 0 \leq s \leq t,$$

then it is called a submartingale. It has no tendency to fall; it may have a tendency to rise. If

$$\mathbb{E}[M(t)|\mathcal{F}(s)] \leq M(s) \textit{ for all } 0 \leq s \leq t \leq T,$$

then it is a supermartingale. It has no tendency to rise; it may have a tendency to fall.

If the process $\{X(t) : t \geq 0\}$ is a martingale with respect to a filtration $\{\mathcal{F}(t) : t \geq 0\}$, then $\mathbb{E}[X(t)] = \mathbb{E}[X(0)]$ for every t.

Lastly, we want to define a way to measure the smoothness or peaks behaviour of a function on small intervals. We will do so with the help of quadratic variation.

Definition 2.4.6 (**Quadratic variation**). *Let $f : [0, T] \to \mathbb{R}$ be a function. Let $\Pi_n = \{0 = t_0, t_1, ..., t_n = T\}$ be a partition of $[0, T]$ and let $\|\Pi_n\| = \max_{0 \leq j \leq n-1}(t_{j+1} - t_j)$. The quadratic variation of f is defined by*

$$[f, f](T) := \lim_{\|\Pi_n\| \to 0} \sum_{i=0}^{n-1} [f(t_{i+1}) - f(t_i)]^2$$

Theorem 2.4.7. *Any continuously differentiable function $f : [0, T] \to \mathbb{R}$ has zero quadratic variation.*

Proof. Since the function f has a continuous derivative, i.e., $f'(t)$ is a continuous function, there exists some $t_j^* \in (t_j, t_{j+1}$ such that $f(t_{j+1} - f(t_j) = f'(t_j^*(t_{j+1} - t_j)$ by the Mean Value Theorem. Therefore,

$$\begin{aligned}\sum_{i=0}^{n-1} [f(t_{j+1}) - f(t_j)]^2 &= \sum_{i=0}^{n-1} [f'(t_j^*)^2 (t_{j+1} - t_j)^2] \\ &\leq \|\Pi_n\| \sum_{i=0}^{n-1} [f'(t_j^*)(t_{j+1} - t_j)]\end{aligned}$$

Where

$$\sum_{i=0}^{n-1} [f'(t_j^*)(t_{j+1} - t_j)]$$

is precisely the Riemann sum of the continuous function $(f'(t_j^*))^2$ and is therefore finite. Now,

$$[f, f](T) = \lim_{\|\Pi_n\| \to 0} \sum_{i=0}^{n-1} [f(t_{j+1}) - f(t_j)]^2 \leq \lim_{\|\Pi_n\| \to 0} \|\Pi_n\| \sum_{i=0}^{n-1} \int_0^T (f'(x))^2 dx = 0$$

∎

Chapter 3

Random walks

The aim of this chapter is to give the reader some intuition of how we can create a Brownian motion. All we need for this is a simple coin. We start by using this coin to construct a one-dimensional symmetric random walk and discuss some of its relevant properties for our later discussion of Brownian motion.
Afterwards, we are going to change the rules of the coin toss to introduce the scaled random walk show and how we can apply the Central Limit Theorem to this scaled random walk to construct a one-dimensional Brownian motion.
This chapter is based on the book by Schreve (2010) ([2])

3.1 Symmetric random walk

Let us suppose we are going to take steps on the integer number line, $\mathbb{Z}$. We start at zero and we will take steps of size one. The direction in which we take our steps are decided by repeatedly tossing a fair coin. If the coin lands on head, we take a step in the positive direction and if the coin lands on tail, we will take a step in the negative direction. Note that since we are throwing a fair coin, these events have equal probability.

To define the symmetric random walk formally, we introduce independent, identically distributed random variables $X_1, X_2, ...$ for which

$$X_j = \begin{cases} 1, & \text{if } \omega_j = H \\ -1, & \text{if } \omega_j = T \end{cases}$$

Here ω_j denotes the outcome of the j-th toss. We denote the successive outcomes of the tosses by $\omega = \omega_1\omega_2\omega_3...$, such that ω is an infinite sequence of tosses We note that $\mathbb{E}(X_j) = 1 \cdot \frac{1}{2} + -1 \cdot \frac{1}{2} = 0$ and $\mathrm{Var}(X_j) = 1$ for all j

We now define the stochastic process $\{M_k\}_{k \geq 0}$ with $M_0 = 0$ and

$$M_k = \sum_{j=1}^{k} X_j, \ k = 1, 2, ...$$

and call it the *symmetric random walk*
In Figure 3.1, we see a sample path of symmetric random walk of ten steps or coin tosses. We need to keep in mind that there are many different sample paths for each amount of coin tosses. We illustrate this with an example for two coin tosses.

Example 3.1.1. *If we were to throw the coin twice, there are four possibilities for our walk, i.e., four possible sample paths:*

- *We throw heads twice. Therefore, by the rules of our random walk, we two times take a step in the positive direction and we end up at two.*

- *We first throw heads and then tails, (or first tails and then heads). In both cases we first take a step in the positive (or negative) direction, and then back to zero. Hence, (in both cases,) we end up where we started: at zero.*
- *We throw tails twice. We take two successive steps in the negative direction and end up at minus two.*

Since each of the sample paths have an equal probability of happening, we will end up at (minus) two with probability $\frac{1}{4}$ and at zero with probability $\frac{1}{2}$.

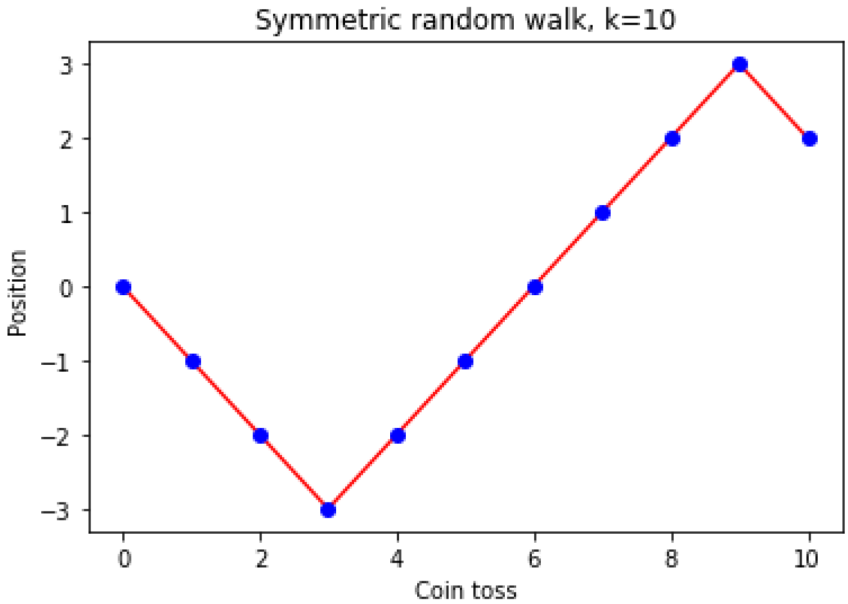

Figure 3.1: A sample path of a symmetric random walk of ten steps; $\{\omega = TTTHHHHHHT\}$.

Remark 3.1.2. *We are only considering symmetric random walks. The theory for (non-symmetric) random walks is similar, the only difference being that the coin is not fair anymore. In other words, if we denote the probability of throwing heads or tail by p, q respectively, then $p \neq q$ is the coin is unfair. As a result, the the probability of a step in the positive direction is not the same to the probability of a step in the negative direction anymore.*

3.1.1 Interesting properties of the Symmetric Random Walk

Increments of the Symmetric Random Walk

A random walk has *independent increments*. This means that if $0 = k_0 < k_1 < ... < k_m$ are integers, then the random variables

$$(M_{k_1} - M_0), (M_{k_2} - M_{k_1}), \ldots, (M_{k_m} - M_{k_{m-1}})$$

are independent. Each of these random variables,

$$M_{k_{i+1}} - M_{k_i} = \sum_{j=k_i+1}^{k_{i+1}} X_j, \tag{3.1}$$

is called an *increments* of the random walk. The increments over non-overlapping time intervals are independent because they depend on different coin tosses. We calculate the expected value and variance of each increment,

$$\begin{aligned}\mathbb{E}(M_{k_{i+1}} - M_{k_i}) &= \mathbb{E}(\sum_{j=k_i+1}^{k_{i+1}} X_j) \\ &= \sum_{j=k_i+1}^{k_{i+1}} \mathbb{E}(X_j) \\ &= 0\end{aligned}$$

and

$$\begin{aligned}\mathrm{Var}(M_{k_{i+1}} - M_{k_i}) &= \mathrm{Var}(\sum_{j=k_i+1}^{k_{i+1}} X_j) \\ &= \sum_{j=k_i+1}^{k_{i+1}} \mathrm{Var}(X_j) \\ &= \sum_{j=k_i+1}^{k_{i+1}} \mathbb{E}[X^2] \\ &= \sum_{j=k_i+1}^{k_{i+1}} 1 \\ &= k_{i+1} - k_i\end{aligned}$$

Where we used the independence of X_j in the second equality. The variance of the symmetric random walk accumulates at rate one per unit time, so that the variance of the increment over any time interval k to l for nonnegative integers $k < l$ is $l - k$.

Martingale Property

To show that the symmetric random walk is a martingale, we recall from chapter two that this is the case if

$$\mathbb{E}[M_l|\mathcal{F}_k] = M_k$$

where $k < l$ are nonnegative integers and $\mathcal{F}_k$ is the σ-algebra generated by $X_1, ..., X_k$, in this case the information of the first k tosses. Now,

$$\begin{aligned}\mathbb{E}[M_l|\mathcal{F}_k] &= \mathbb{E}[(M_l - M_k) + M_k|\mathcal{F}_k] \\ &= \mathbb{E}[(M_l - M_k)|\mathcal{F}_k] + \mathbb{E}[M_k|\mathcal{F}_k] \\ &= \mathbb{E}[M_l - M_k] + \mathbb{E}[M_k|\mathcal{F}_k] \\ &= 0 + M_k \\ &= M_k\end{aligned}$$

Here we have used the linearity of conditional expectations for the second equality. For the third equality, since $k < l$, the increment $M_l - M_k$ is independent of $\mathcal{F}_k$, as $\mathcal{F}_k$ does not have information on anything beyond time k. The second-last equation is a result of the fact that M_k is $\mathcal{F}_k$-measurable.

Quadratic Variation

The last property of symmetric random walks we will consider is its quadratic variation. The quadratic variation up tot time k equals

$$[M, M]_k = \sum_{j=1}^{k}(M_j - M_{j-1})^2 = \sum_{j=1}^{k}(\pm 1)^2 = k.$$

Remark 3.1.3. *Note that the quadratic variation is computed path-by-path. It has the same value as Var(M_k), but the computations of these two quantities are quite different. The variance of (M_k) is computed taking an average over all paths, taking into account their probabilities. If the random walk were not symmetric, this would affect Var(M_k), while the quadratic variation up to time k remains unchanged.*

3.2 Scaled Random Walk

The aim of this chapter was to approximate Brownian motion. Thus far, we have only looked at a symmetric random walk. A symmetric random walk however does not suffices. That is why we are going to change the 'rules' of the coin toss: we speed up time and we scale down the step size of a symmetric random walk to create a process called the *scaled symmetric random walk*,

$$W^{(n)}(t) = \frac{1}{\sqrt{n}} M_{nt} = \frac{1}{\sqrt{n}} \sum_{j=1}^{nt} X_j,$$

where n is a fixed positive integer, $t \geq 0$ and nt is required to be an integer.
If it is the case that nt is not an integer, we take the two nearest values s, u to the left and to the right of t for which st, su are integers. We use these integers to calculate $W^{(s)}(t), W^{(u)}(t)$ and use linear interpolation to approximate the value of $W^{(t)}(t)$.

As the scaled random walk is a continuous function of the symmetric random walk, the properties that we derived for the symmetric random walk still hold for the scaled random walk. We shall now review these properties specified to the scaled random walk.

Increments of the Scaled Random Walk

Each random variable of the form

$$W^{(n)}(t_{i+1}) - W^{(n)}(t_i) = \frac{1}{\sqrt{n}} \sum_{j=nt_i+1}^{nt_{i+1}} X_j,$$

is called an increment. The scaled random walk has *independent increments*.
The expected value of the scaled random walk equals zero.
The variation of the random walk is given by,

$$\begin{aligned}
\mathrm{Var}(W^{(n)}(t_{i+1}) - W^{(n)}(t_i)) &= \mathrm{Var}(\frac{1}{\sqrt{n}} \sum_{j=nt_i+1}^{nt_{i+1}} X_j) \\
&= \frac{1}{n}\mathrm{Var}(\sum_{j=nt_i+1}^{nt_{i+1}} X_j) = \frac{1}{n} \sum_{j=nt_i+1}^{nt_{i+1}} \mathrm{Var}(X_j) \\
&= \frac{1}{n} \sum_{j=nt_i+1}^{nt_{i+1}} \mathrm{Var}(1) = \frac{1}{n}(nt_{i+1} - nt_i) \\
&= t_{i+1} - t_i.
\end{aligned}$$

Martingale Property

The scaled random walk is a martingale,

$$\mathbb{E}[W^{(n)}(t)|\mathcal{F}(s)] = W^{(n)}(s),$$

for $0 \leq s \leq t$ such that ns and nt are integers, where $\mathcal{F}(s)$ is the σ-algebra of information available at time s, i.e., is the knowledge of the first ns coin tosses.

Quadratic Variation

The quadratic variation of the scaled random walk equals,

$$
\begin{aligned}
[W^{(n)}, W^{(n)}](t) &= \sum_{j=1}^{nt} \left[W^{(n)}(\frac{j}{n}) - W^{(n)}(\frac{j-1}{n}) \right]^2 \\
&= \sum_{j=1}^{nt} \left[\frac{1}{\sqrt{n}} X_j \right]^2 \\
&= \sum_{j=1}^{nt} \frac{1}{n} \\
&= t.
\end{aligned}
$$

To conclude the section on scaled random walks, let us visualize a sample path of a scaled random walk. For this, see Figure 3.2, which shows a sample path of $W^{(100)}$ up to time 4. This was generated by 400 coin tosses with a step up or down of size 1/10 at each coin toss.

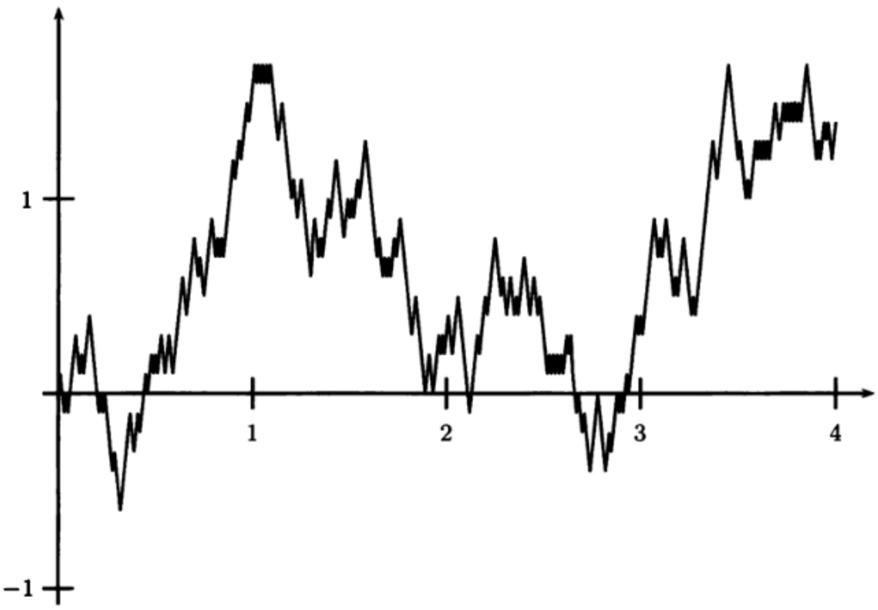

Figure 3.2: A sample path of $W^{(100)}$.[2]

3.2.1 Limiting Distribution of the Scaled Random Walk

If we compare Figure 3.2 to Figure 3.1, we see that the sample path look already much more like that of a Brownian motion. If we keep on increasing the value of n , the approximation of Brownian motion will keep on getting better. Let us formalize this.

Theorem 3.2.1. *Fix $t \geq 0$. As $n \to \infty$, the distribution of the scaled random walk $W^{(n)}(t)$ evaluated at time (t) converges to the normal distribution with mean zero and variance t.*

Chapter 4

Brownian motion

We are now ready to define a Brownian motion. We quickly remind ourselves that in the previous section, we constructed a scaled random walk from the symmetric random walk and obtained Brownian motion as the limit of the scaled random walks. Whereas the scaled random walk has a natural time step of $1/n$ and is linear between these time steps, Brownian motion has no linear pieces. In addition to this, while the scaled random walk is only approximately normal for each t, Brownian motion is exactly normal as a consequence of the Central Limit Theorem. Nevertheless, Brownian motion inherits properties from these random walk.

In this chapter, we will first give the definition of Brownian motion and then look at various properties, for which we were inspired by the structure of the article written by Ermogenous (2006)([5]). Some basic properties are discussed, followed by distributional properties of Brownian motion.
Next, we will take a look at the quadratic variation of a Brownian motion, which will motivate us to prove the non-differentiability of a Brownian motion as stated in the book of Mörters and Peres (2010) [1].

4.1 Formal definition of Brownian motion

Definition 4.1.1 (Standard Brownian motion). *A standard Brownian motion is a Wiener stochastic process. A Wiener process is a stochastic process $W(t)$ with values in $\mathbb{R}$ defined for $t \in [0, \infty)$ such that the following conditions hold:*

i. *W(0)=0.*

ii. *If $0 < s < t$ then $W(t) - W(s)$ has a normal distribution with mean 0 and variance $(t - s)$.*

iii. *If $0 \leq s \leq t \leq u \leq v$ (i.e., the two intervals $[s, t]$ and $[u, v]$ do not overlap, except possibly at the boundary) then $W(t) - W(s)$ and $W(v) - W(u)$ are independent random variables.*

iv. *The sample paths $t \mapsto W(t)$ are almost surely continuous.*

The probability density function of $W(t)$ is $f_{W(t)}(x) = \frac{1}{\sqrt{2\pi t}} e^{\frac{-x^2}{2t}}$.

Since $W(0) = 0$, we refer to this Wiener stochastic process as a *standard* Brownian motion. It is also allowable that $W(0) = x, x \in \mathbb{R}$. In this thesis however, we will focus on standard Brownian motions.

Remark 4.1.2. *This stochastic process is named in honor of the American mathematician Norbert Wiener, who investigated the mathematical properties of the one-dimensional Brownian motion, hence the choice of notation $W(t)$. Throughout this thesis, we will generally call the stochastic process denoted by $W(t)$ a Brownian motion, but we could equivalently call it a Wiener process.*

In addition to Brownian motion itself, we will need some notation for the amount of information about the process available at each time. We will do so with a filtration.

Definition 4.1.3. *[2] Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space on which is defined a Brownian motion $\{W(t) : t \geq 0\}$. A filtration for the Brownian motion is a collection of σ-algebras $\{\mathcal{F}(t) : t \geq 0\}$, satisfying:*

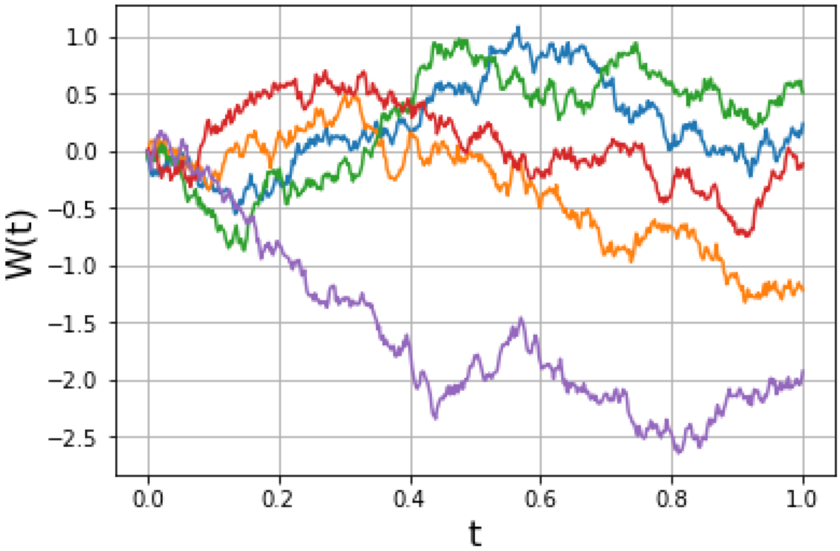

Figure 4.1: five possible sample paths of a standard Brownian motion of the interval $[0, 1]$.

- ***(Information accumulates)*** *For* $0 \leq s < t$, *every set in* $\mathcal{F}(s)$ *is also in* $\mathcal{F}(t)$. *In other words, there is at least as much information available at the later time* $\mathcal{F}(t)$ *as there is at the earlier time* $\mathcal{F}(s)$.
- ***(Adaptivity)*** *For eacht* $t \geq 0$, *the Brownian motion* $W(t)$ *at time* t *is* $\mathcal{F}(t)$*-measurable. In other words, the information available at time* t *is sufficiient to evaluate the Brownian motion* $W(t)$ *at that time.*
- ***(Independence of future increments)*** *For* $0 \leq t < u$, *the increment* $W(u) - W(t)$ *is independent of* $\mathcal{F}(t)$. *In particular,* $W(u) - W(t)$ *is independent of* $W(s)$ *for all* $s \leq t$.

4.2 Basic Properties of Brownian Motion

Throughout this section we assume that $\{W(t) : t \geq 0\}$ is a Brownian motion on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and that $\{F(t) : t \geq 0\}$ is a filtration for $\{W(t) : t \geq 0\}$.

Martingale property for Brownian Motion

Theorem 4.2.1. *Brownian motion is a martingale.*

Proof.

$$\begin{aligned}
\mathbb{E}[W(t)|\mathcal{F}(s)] &= \mathbb{E}[(W(t) - W(s)) + W(s)|\mathcal{F}(s)] \\
&= \mathbb{E}[W(t) - W(s)|\mathcal{F}(s)] + \mathbb{E}[W(s)|\mathcal{F}(s)] \\
&= \mathbb{E}[W(t) - W(s)] + W(s) \\
&= W(s).
\end{aligned}$$

The argumentation of this proof is analogous to the proof that a symmetric random walk is a martingale. We refer to the subsection *Martingale Property* of 3.1.1 for justification of this proof. ■

Covariance of Brownian Motion

Theorem 4.2.2. *For any two times* $s, t \geq 0$, *the covaraince of* $W(s)$ *and* $W(t)$ *is given by*

$$Cov(W(s), W(t)) = min\{s, t\}.$$

Proof. First assume that $s < t$. Since $\mathbb{E}[W(s)] = \mathbb{E}[W(t)] = 0$, we have

$$\begin{aligned}
Cov(W(s), W(t)) &= \mathbb{E}[W(s)W(t)] - \mathbb{E}[W(s)]\,\mathbb{E}[W(t)] \\
&= \mathbb{E}[W(s)[(W(t) - W(s)) + W(s)]] \\
&= \mathbb{E}[W(s)(W(t) - W(s))] + \mathbb{E}[W^2(s)] \\
&= \mathbb{E}[W(s)]\,\mathbb{E}[(W(t) - W(s))] + \mathrm{Var(W(s))} \\
&= 0 + s = s,
\end{aligned}$$

where we used linearity of conditional expectations in the third equality and independence of the increments in the fourth equality. By symmetry, for $t < s$ we find $Cov(W(s), W(t)) = t$. Therefore, $Cov(W(s), W(t)) = min\{s, t\}$.

■

4.3 Distributional Properties of Brownian Motion

In this section, distributional properties of Brownian motion will be discussed.

Example 4.3.1. *If $\{W(t) : t \geq 0\}$ is a Brownian motion, then $\{-W(t) : t \geq 0\}$ is also a Brownian motion. Independence and continuity are clearly preserved by a negative multiplication. Furthermore, the normal distribution of the Brownian motion is symmetric around zero. Hence, all the increments have the correct means and variances.*

Theorem 4.3.2 (Scaling invariance). *If $\{W(t) : t \geq 0\}$ is a standard Brownian motion and $c > 0$, then $\{\sqrt{c}W(\frac{t}{c}) : t \geq 0\}$ is also a standard Brownian motion.*

Proof. We define the random variable $X(t) = \sqrt{c}W(\frac{t}{c})$ and show that $X(t) : t \geq 0$ satisfies the four conditions of a standard Brownian motion, see Definition 4.1.1. Since $X(t)$ is a continuous function of $W(t)$, $X(t)$ is continuous and the independence of the increments still hold. In addition, $X(0) = \sqrt{c}W(\frac{0}{c}) = 0$. Lastly, for any increment,

$$\begin{aligned}
X(t_{j+1}) - X(t_j) &= \sqrt{c}W(t_{j+1}/c) - \sqrt{c}W(t_j/c) \\
&= \sqrt{c}[W(t_{j+1}/c) - W(t_j/c)]
\end{aligned}$$

which is distributed $\mathcal{N}(0, t_{j+1} - t_j)$ as desired. ■

Theorem 4.3.3 (Time inversion). *Let $\{W(t) : t \geq 0\}$ be a standard Brownian motion. The process $\{Z(t) : t \geq 0\}$ defined by*

$$Z(t) = \begin{cases} 0, & t = 0 \\ tW(\frac{1}{t}) & t > 0 \end{cases}$$

is also a Brownian motion.

Proof. [7]
We first note that $Z(0) = 0$. Now we will show that $Z(t)$ has normal increments with the right means and variances. Take any $0 < s < t$, which implies that $\frac{1}{t} < \frac{1}{a}$. write,

$$(Z(t) - Z(s) = (tW(1/t) - sW(1/s)) = (t - s)W(1/t) + s(W(1/t) - W(1/s)). \tag{4.1}$$

The first part of the right hand side of (4.1) is normally distributed with

$$(t - s)W(1/t) \sim \mathcal{N}(0, (t - s)^2(1/t)).$$

Since $\frac{1}{t} < \frac{1}{s}$, $W(1/s) - W(1/t)$ is independent of $W1/t)$. In particular, $s(W(1/t) - W(1/s))$ is independent of $(t - s)W(1/t)$. Therefore, the second part of the right hand side of (4.1) is also normally distributed with

$$s(W(1/t) - W(1/s)) \sim \mathcal{N}(0, s^2(1/s - 1/t)).$$

Since the increment $Z(t) - Z(s) = (tW(1/t) - sW(1/s))$ is the sum of two independent normally distributed random variables, this increment itself is also normally distributed with mean 0 and variance

$$(t-s)^2(1/t) + s^2(1/s - 1/t) = t - s$$

as desired.
To establish independence of the increments of $Z(t)$, we take a look at the covariance between two increments, $t > s \geq 0$.

$$\begin{aligned}
\text{Cov}(Z(s), (Z(t) - Z(s)) &= \text{Cov}(Z(s), Z(t)) - \text{Cov}(Z(s), Z(s)) \\
&= \text{Cov}(sW(1/s), tW(1/t)) - \text{Var}(Z(s)) \\
&= st \cdot \frac{1}{t} - s \\
&= 0.
\end{aligned}$$

Since both the random variables $Z(s), (Z(t) - Z(s))$ are normal and uncorrelated, they are independent.
Lastly, we will prove continuity. Continuity for $t > 0$ is clear. To prove continuity in $t = 0$, we set $t = 1/n$ and we consider $tW(1/t) = W(n)/n$. Since

$$W(n) = \sum_{i=1}^{n}(W(i) - W(i-1))$$

is a sum of independent, identically distributed standard normal random variables, by the Strong Law of Large Numbers,

$$\lim_{t \to 0} tW(1/t) = \lim_{n \to \infty} W(n)/n = \mathbb{E}[W(1) - W(0)] = 0 \text{ almost surely.}$$

■

4.4 Quadratic Variation of Brownian Motion

Heuristically, Theorem 4.3.2 states that as we decrease c, i.e., as we 'zoom' in, the average features of the function do not change, i.e. the fractal image still looks like (and is) a Brownian motion, no matter how long we keep on zooming in. If we think about this, it might suggest that Brownian motion is not a smooth function, as you would then expect to see a straight line once we have zoomed in far enough. We expect the Brownian motion to always have sharp peaks. This motivates us to look at the quadratic variation of Brownian motion.

We first recall the definition of quadratic variation (Definition 2.4.6) to see that the quadratic variation of a Brownian motion $W(t) : t \geq 0$ is given by

$$[W,W](T) = \lim_{|\Pi_n| \to 0} \sum_{j=0}^{n-1} (W(t_{j+1}) - W(t_j))^2$$

Theorem 4.4.1. *Let $W(t)$ be a Brownian motion. Then $[W,W](T) = T$ in $\mathcal{L}^2$ and almost surely for all $T \geq 0$.*

Proof. Let $\Pi = \{0 = t_0 < t_1 < \cdots < t_n = T\}$ We define the *sampled quadratic variation* corresponding to this partition to be

$$\mathcal{Q}(\Pi_n) = \sum_{j=0}^{n-1} (W(t_{j+1}) - W(t_j))^2$$

We will first show $\mathcal{L}^2$-convergence, i.e., proof that the following holds:

$$\lim_{|\Pi_n| \to 0} \mathbb{E}[(\mathcal{Q}(\Pi_n) - T)^2] = 0,$$

We can write

$$\mathcal{Q}(\mathbf{\Pi_n}) - T = \sum_{j=0}^{n-1} [(W(t_{j+1}) - W(t_j))^2 - (t_{j+1} - t_j)] \tag{4.2}$$

and call the summand $\mathcal{I}_j$. We can now rewrite equation (4.2) as

$$\mathcal{Q}(\mathbf{\Pi_n}) - T = \sum_{j=0}^{n-1} \mathcal{I}_j. \tag{4.3}$$

We remark that each $\mathcal{I}_j, j = 0, ..., n-1$, is a sum of the square of a normally distributed random variable and a constant. Hence $I_0, ..., I_{n-1}$ are independent and normally distributed with

$$\begin{aligned}
\mathbb{E}[\mathcal{I}_j] &= \mathbb{E}[(W(t_{j+1}) - W(t_j))^2] - (t_{j+1} - t_j) \\
&= \mathrm{Var}(W(t_{j+1}) - W(t_j)) - (t_{j+1} - t_j) \\
&= (t_{j+1} - t_j) - (t_{j+1} - t_j) \\
&= 0.
\end{aligned}$$

and,

$$\begin{aligned}
\mathrm{Var}(\mathcal{I}_j) &= \mathbb{E}[\mathcal{I}_j^2] \\
&= \mathbb{E}[(W(t_{j+1} - W(t_j))^4] - 2(t_{j+1} - t_j)\mathbb{E}[(W(t_{j+1}) - W(t_j))^2] + (t_{j+1} - t_j)^2 \\
&= 3(t_{j+1} - t_j)^2 - 2(t_{j+1} - t_j)^2 + (t_{j+1} - t_j)^2 \\
&= 2(t_{j+1} - t_j)^2.
\end{aligned}$$

In the second last step we used that if $X \sim N(0, \sigma^2)$ then $\mathbb{E}[X^4] = 3\sigma^2$. Now, since $\mathbb{E}[\mathcal{I}_i] = \mathbb{E}[\mathcal{I}_j] = 0$,

$$\begin{aligned}
\mathrm{Var}(\mathcal{Q}(\Pi_n) - T) = \mathbb{E}[(\mathcal{Q}(\Pi_n) - T)^2] &= E[(\sum_{j=0}^{n-1} \mathcal{I}_j)^2] \\
&= \mathbb{E}[\sum_{j=0}^{n-1} \mathcal{I}_j^2] + 2\sum_i \sum_j \mathbb{E}[\mathcal{I}_i \mathcal{I}_j] \\
&= \mathbb{E}[\sum_{j=0}^{n-1} \mathcal{I}_j^2] + 2\sum_i \sum_j \mathbb{E}[\mathcal{I}_i]\mathbb{E}[\mathcal{I}_j] \\
&= \mathbb{E}[\sum_{j=0}^{n-1} \mathcal{I}_j^2] \\
&= 2\sum_{j=0}^{n-1} (t_{j+1} - t_j)^2.
\end{aligned}$$

We now conclude the proof by making the following estimation

$$\begin{aligned}
\lim_{|\Pi_n| \to 0} \mathbb{E}[(\mathcal{Q}(\Pi_n) - T)^2] &= \lim_{|\Pi_n| \to 0} 2\sum_{j=0}^{n-1} (t_{j+1} - t_j)^2 \\
&\leq \lim_{|\Pi_n| \to 0} 2|\Pi_n| \sum_{j=0}^{n-1} (t_{j+1} - t_j) \\
&= \lim_{\|\Pi_n\| \to 0} 2\|\Pi_n\| T \\
&= 0.
\end{aligned}$$

Hence, $\mathcal{Q}(\Pi_n)$ converges to T in $\mathcal{L}^2$.
To conclude the proof, we will show that $[W, W](T) = T$ for all $T \geq 0$ almost surely. For this, we introduce a sequence $\varepsilon_n \to 0$ such that $\|\Pi_n\| = \frac{\varepsilon_n}{n^2}$. By Chebychev's inequality,

$$\mathbb{P}((\mathcal{Q}(\|\Pi_n\| - T) > 2\varepsilon_n) \leq \frac{\mathbb{E}[(\mathcal{Q}(\|\Pi_n\| - T)^2]}{2\varepsilon_n} \leq \frac{2\|\Pi_n\| T}{2\varepsilon_n} = \frac{T}{i^2}.$$

Now, since $\sum_{n \geq 0} \frac{T}{i^2} < \infty$, we can apply the Borel-Cantelli Lemma (Lemma 2.1.10), the probability that $\mathcal{Q}(\|\Pi_n\| - T) \geq 2\varepsilon_n$ for infinitely many n is equal to zero. By construction, $\varepsilon_n \to 0$. Therefore we can conclode that

$$\mathbb{P}\Big(\big\{\omega \in \Omega : \lim_{\|\Pi_n\| \to 0} \mathcal{Q}(\|\Pi_n\|(w) = T\big\}\Big) = 1$$

i.e. $[W, W](T) = T$ almost surely for all $T \geq 0$. [12] ■

4.5 Non-differentiability of Brownian motion

We have seen in Theorem 2.4.7 that any continuously differentiable function f has zero quadratic variation. As Brownian motion accumulates quadratic variation at rate one per unit time, it cannot be continuously differentiable. In fact, the paths of a Brownian motion are almost surely nowhere differentiable. We formalize the non-differentiability of Brownian motion in this in this section, but before we do so we define the upper and lower right derivatives for a function $f : [0, 1) \to \mathbb{R}$ by

$$D^* f(t) = \limsup_{h \downarrow 0} \frac{f(t+h) - f(t)}{h}, \ \ D_* f(t) = \liminf_{h \downarrow 0} \frac{f(t+h) - f(t)}{h}$$

respectively. We remind ourselves that a function f is differentiable in t if and only if

$$-\infty < \limsup_{h \downarrow 0} \frac{f(t+h) - f(t)}{h} = \liminf_{h \downarrow 0} \frac{f(t+h) - f(t)}{h} < \infty$$

Now let us look at the non-differentiabiliy of Brownian motion.

Theorem 4.5.1 (**Paley, Wiener and Zygmund 1933**). *Almost surely, Brownian motion $\{W(t) : t \geq 0\}$ is nowhere differentiable.*

Proof. Suppose that there is a $t_0 \in [0, 1]$ such that $-\infty < D_* B(t_0) \leq D^* B(t_0) < \infty$. Then

$$\limsup_{h \downarrow 0} \frac{|W(t_0 + h) - W(t_0)|}{h} < \infty.$$

This implies that for some finite constant M there exixts t_0 with

$$\sup_{h \in [0,1]} \frac{|W(t_0 + h) - W(t_0)|}{h} \leq M \tag{4.4}$$

Note that

$$\Big\{ -\infty < D^* W^(t_0) < \infty \Big\} \subseteq \bigcup_{M=1}^{\infty} \left\{ \sup_{h \in [0,1]} \frac{|W(t_0 + h) - W(t_0)|}{h} \leq M \right\}$$

Which implies by σ-additivity that is suffices to show that

$$\mathbb{P}\Big(\Big\{ \sup_{h \in [0,1]} \frac{|W(t_0 + h) - W(t_0)|}{h} \leq M \Big\} \Big) = 0$$

for any M.
From now on fix M. Note that the sets $[(k-1)/2^n, k/2^n], k = 1, ..., 2^n$ are a partition of $[0, 1]$. In particular, for any $t_0 \in [0, 1]$ and n sufficiently large, we can pick $k \in \{1, ..., 2^n - 3\}$ such that

$$t_0 \in \left[\frac{k-1}{2^n}, \frac{k}{2^n} \right].$$

So, if $t_0 \in [(k-1)/2^n, k/2^n]$ for $n > 2$, then for all $1 \leq j \leq 2^n - k$, we find that

$$\big| W(k+j)/2^n) - W(k+j-1/2^n) \big| \leq M \frac{2j+1}{2^n}$$

Where we used that $t_0 \geq \frac{k-1}{2^n}$ and the triangle inequality in addition to applying (4.4) to see that

$$\frac{\Big| W\Big(t_0 + \Big(\frac{k+j}{2^n} - t_0\Big)\Big) - W(t_0) \Big|}{\Big(\frac{k+j}{2^n} - t_0\Big)} \leq M \implies \Big| W\Big(\frac{k+j}{2^n}\Big) - W(t_0) \Big| \leq M \Big(\frac{k+j}{2^n} - t_0\Big)$$

$$\frac{\Big| W(t_0) - W\Big(t_0 + \Big(\frac{k+j-1}{2^n} - t_0\Big)\Big) \Big|}{\Big(\frac{k+j-1}{2^n} - t_0\Big)} \leq M \implies \Big| W(t_0) - W\Big(\frac{k+j-1}{2^n}\Big) \Big| \leq M \Big(\frac{k+j-1}{2^n} - t_0\Big)$$

We now define the following event

$$\Omega_{n,k} := \big\{ \big| W(k+j)/2^n) - W(k+j-1/2^n) \big| \le M \frac{2j+1}{2^n} \text{ for } j = 1,2,3 \big\}$$

This is the event that Brownian motion on the three binary intervals to the right of the interval in which t_0 is contained are bounded by $M(\frac{2j+1}{2^n})$. Now,

$$\begin{aligned}
\mathbb{P}(\Omega_{n,k}) &= \prod_{j=1}^{3} \mathbb{P}\Big\{ \big| W(k+j)/2^n) - W(k+j-1/2^n) \big| \le M \frac{2j+1}{2^n} \Big\} \\
&\le \prod_{j=1}^{3} \mathbb{P}\Big\{ \big| \sqrt{2^n}[W(k+j)/2^n) - W(k+j-1/2^n] \big| \le M \frac{2j+1}{\sqrt{2^n}} \Big\} \\
&= \prod_{j=1}^{3} \mathbb{P}\Big\{ \big| W(1) \big| \le M \frac{2j+1}{\sqrt{2^n}} \Big\} \\
&\le \mathbb{P}\Big\{ \big| W(1) \big| \le \frac{7M}{\sqrt{2^n}} \Big\}^3
\end{aligned}$$

Where we used Theorem 4.3.2 and the independence of the increments. Now since the normal density is bounded by 1/2,

$$\begin{aligned}
\mathbb{P}(|X| \le t) &= \mathbb{P}(t \le X \le t) \\
&= \int_{-t}^{t} \frac{1}{\sqrt{2\pi}} e^{-1/2x^2} dx \\
&\le \int_{-t}^{t} \frac{1}{2} dx = t
\end{aligned}$$

And therefore,

$$\mathbb{P}\big(\Omega_{n,k}\big) \le \Big(\frac{7M}{\sqrt{2^n}} \Big)^3$$

Now, by countable additivity,

$$\begin{aligned}
\mathbb{P}\Big(\bigcup_{k=0}^{2^n-3} \Omega_{n,k} \Big) &\le \sum_{k=0}^{2^n-3} \mathbb{P}\big(\Omega_{n,k}\big) \\
&\le (7M)^3 2^{-n/2}
\end{aligned}$$

Since $\mathbb{P}\Big(\bigcup_{k=0}^{2^n-3} \Omega_{n,k} \Big) < \infty$ we see by Lemma 2.1.10 that

$$\mathbb{P}\Big(\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} \bigcup_{k=0}^{2^n-3} \Omega_{n,k} \Big) = 0.$$

To conclude the proof, we note that

$$\begin{aligned}
&\mathbb{P}\Big\{ \text{there is } t_0 \in [0,1] \text{ with } \sup_{h \in [0,1]} \frac{|W(t_0 + hW(t_0)|}{h} \le M \Big\} \\
&\le \mathbb{P}\Big(\bigcup_{k=0}^{2^n-3} \Omega_{n,k} \text{ for infinitely many n } \Big) = 0
\end{aligned}$$

■

We have just proven that Brownian motion is nowhere differentiable, even though it's paths are continuous. This is one of the properties that makes Brownian motion so unique.

To summarize this chapter, we have seen that Brownian motion inherits properties from the scaled random walk as a result of the Central Limit Theorem. We also looked at distributional properties of Brownian motion, the most important being the scaling invariance of Brownian motion, which informally states that given a standard Brownian motion, and we zoom in on any part of the process, it will always be a Brownian motion. This motivated us to look at the quadratic variation of Brownian motion, which was non-zero as expected and implied that Brownian motion cannot be continuously differentiable. We concluded this section by proving that it is in fact the case that Brownian motion is nowhere differentiable.

On a last note for this chapter, ordinary calculus is based on the fact that a continuously differentiable $g(t)$ have zero quadratic variation. Therefore, we can define

$$\int_0^T \Delta(t)dg(t) = \int_0^T \Delta(t)g'(t)dt$$

where $\Delta(t)$ is an adapted stochastic process.
As for now, we do however not know how to make sense of such an integral for Brownian motion,

$$\int_0^T \Delta(t)dW(t), \tag{4.5}$$

As Brownian motion has non-zero quadratic variation and is therefore not differentiable. This problem will give rise to Stochastic calculus, in which we use that we can write $dW(t)dW(t) = dt$ to symbolize the fact that the amount of quadratic variation of Brownian motion accumulates in an interval is equal to the length of the interval, regardless of the path along which we do the computation. We define (4.5) as the *Itô integral.*

Stochastic calculus won't be discussed further in this thesis, but the reader should be aware of the connection between Brownian motion and stochastic calculus.

Chapter 5

The (Strong) Markov Property for Brownian Motion

In this final chapter of this thesis, we will look at Markov property and strong Markov property, which informally both states that at certain times $s > 0$, the Brownian motion starts anew, see Figure 5.1. We first look at the Markov property and how we can use this to predict the behaviour of a Brownian motion infinitesimally close to zero. We will then apply all this to study the local and global maxima of a Brownian motion.
Afterwards, we improve the Markov Property by looking at the strong Markov property, in which we extend the set of times for which the Markov property hold to random times called stopping times. We conclude this chapter and henceforth this thesis by a discussion of the reflection principle.

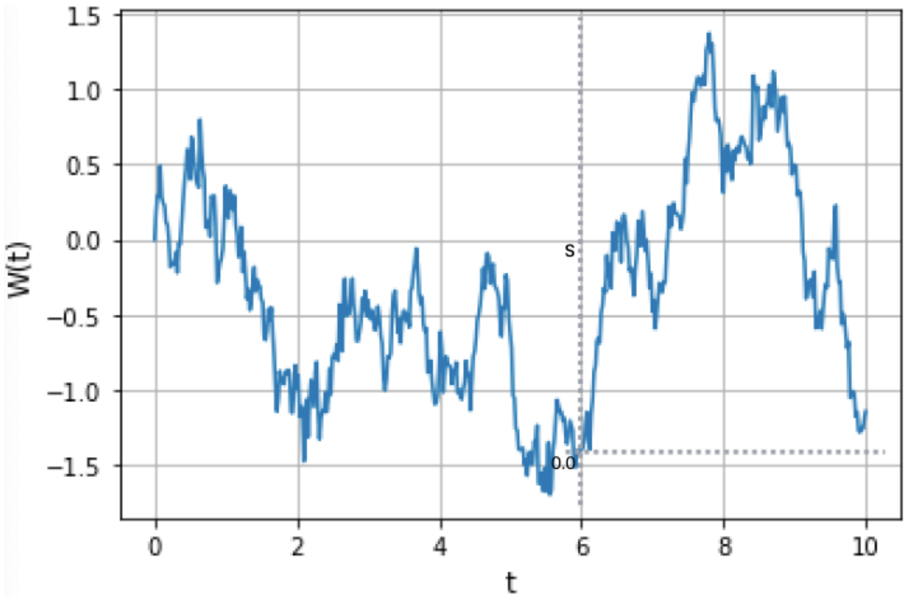

Figure 5.1: Brownian motion starts afresh at fixed time s.[1]

5.1 The Markov Property

Suppose that $\{W(t) : t \geq 0\}$ is a standard Brownian motion, and we know the behaviour of this process on the interval $[0, s]$ for a fixed s. Our goal is to predict the future of the process, i.e. $\{W(t) : t \geq s\}$. To do so, according to the Markov property which we are going to discuss in this section, it is as useful to know just the value at the endpoint $W(s)$ as knowing the entire process from 0 up until time s. In this sense, you could say that process is memoryless: The evolution in the future of a Brownian motion only depends on the past through its present state.

In this section, we will first state and the Markov property for a standard Brownian motion. In addition to this, we will look at Blumenthal's 0-1 law and the behaviour of a standard Brownian motion infinitesimally close to zero.

Remark 5.1.1. *In this thesis, we will only focus on the (strong) Markov property for Brownian motion. The Markov property can however be defined for stochastic processes other than Brownian motion.*

Theorem 5.1.2 (Markov property). *Let $\{W(t) : t \geq 0\}$ be a standard Brownian motion and fix $s \geq 0$. Then the process $\{W(t+s) - W(s) : t \geq 0\}$ is again a Brownian motion started in the origin and it is independent of $\{W(t) : 0 \leq t \leq s\}$.*

Proof. Let $X(t) = W(t+s) - W(s)$. The proof that $\{X(t) : t \geq 0\}$ is a standard Brownian motion will be analogously to the proofs given in Section 4.3. The independence property follows from Definition 4.1.1. ■

We now introduce some useful terminology with which we will be able to slightly improve the Markov property.

Suppose we have a Brownian motion $\{W(t) : t \geq 0\}$ defined on some probability space. Then we can define a filtration $\{\mathcal{F}^0(t) : t \geq 0\}$ by letting

$$\mathcal{F}^0(t) = \sigma\{W(s) : 0 \leq s \leq t\}$$

be the σ-algebra generated by the random variables $W(s)$, for $0 \leq s \leq t$. Intuitively, this σ-algebra contains all the information available from observing the process up to time t.

In addition to this, we define a slightly larger σ-algebra $\mathcal{F}^+(s)$ by

$$\mathcal{F}^+(s) = \bigcap_{t>s} \mathcal{F}^0(t).$$

where $\{\mathcal{F}^+(t) : t \geq 0\}$ is again a filtration and $\mathcal{F}^0(s) \subset \mathcal{F}^+(s)$.

Remark 5.1.3. *This second σ-algebra has the advantage of being right-continuous, that is,*

$$\bigcap_{\epsilon>0} \mathcal{F}^+(t+\epsilon) = \bigcap_{n=1}^{\infty} \bigcap_{n=1}^{\infty} \mathcal{F}^0(t + 1/n + 1/k) = \mathcal{F}^+(t).$$

Theorem 5.1.4. *For every $s \geq 0$, the process $\{W(t+s) - W(s) : t \geq 0\}$ is independent of the σ-algebra $\mathcal{F}^+(s)$.*

Proof. By continuity,

$$W(t+s) - W(s) = \lim_{n\to\infty} W(s_n + t) - W(s_n),$$

where $(s_n)_{n\in\mathbb{N}}$ is a strictly decreasing sequence converging to s. Now by Theorem 5.1.2, for any fixed $s_j, j \in \mathbb{N}$, the process $\{W(s_j + t) - W(s_j) : t \geq 0\}$ is independent of $\mathcal{F}^0(s_j) \supset \mathcal{F}^+(s)$ for all $j \in \mathbb{N}$. Therefore, $\{W(s_j + t) - W(s_j) : t \geq 0\}$ is independent of $\mathcal{F}^+(s)$ and so it the limit. We can conclude that the process $\{W(t+S) - W(s) : t \geq 0\}$ is independent of $\mathcal{F}^+(s)$ for every $s \geq 0$. ■

An alternative way of stating this is that conditional on $\mathcal{F}^+(s)$ the process, $\{W(t+s) - W(s) : t \geq 0\}$ is a Brownian motion started in $W(s)$.

We will now take a look at the *germ σ-algebra* $\mathcal{F}^+(0) = \bigcap_{s>0} \mathcal{F}^0(s)$, which heuristically comprises all events defined in terms of Brownian motion on an infinitesimal small interval to the right of the origin.

Theorem 5.1.5 (Blumenthal's 0-1 law). *Let $A \in \mathcal{F}^+(0)$. Then $\mathbb{P}(A) \in \{0, 1\}$*

Proof. Using Theorem 5.1.4 for $s = 0$ we see that any $B \in \sigma(W(t) : t \geq 0)$ is independent of $\mathcal{F}^+(0)$. Let us now take $A \in \mathcal{F}^+(0) \subset \sigma(W(t) : t \geq 0)$, then A is independent of $\mathcal{F}^+(0)$, and therefore A is independent of itself, that is,

$$\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)^2$$

Which implies that $\mathbb{P}(A) \in \{0, 1\}$ ∎

The following theorem is an application of Blumenthal's 0-1 law.

Theorem 5.1.6. *Suppose* $\{W(t) : t \geq 0\}$ *is a standard Brownian motion. Define* $\tau = \inf\{t > 0 : W(t) > 0\}$ *and* $\omega = \inf\{t > 0 : W(t) = 0\}$*. Then*

$$\mathbb{P}\{\tau = 0\} = \mathbb{P}\{\omega = 0\} = 1$$

Proof. We will first look at $\mathbb{P}\{\tau = 0\}$. The event $\{\tau = 0\}$ occurs if in an infinitesimal small interval to the right of zero, there are infinitely many points $t_i \searrow 0$ such that $B(t_i) > 0$, i.e.,

$$\{\tau = 0\} = \bigcap_{n=1}^{\infty} \Big\{ \text{ there is an } 0 < \epsilon < 1/n \text{ such that } B(\epsilon) > 0 \Big\}.$$

We note that $\mathcal{F}^+(0) = \bigcap_{n=1}^{\infty} \mathcal{F}^+(\frac{1}{n})$ and thus we see that $\{\tau = 0\} \in \mathcal{F}^+(0)$. Hence, by Theorem 5.1.5,

$$\mathbb{P}\{\tau = 0\} \in \{0, 1\}.$$

We will now show that $\mathbb{P}[\tau = 0]$ is positive and must therefore be equal to one. For this, note that

$$\mathbb{P}(\tau \leq t) \geq \mathbb{P}(W(t) > 0) = \frac{1}{2} \text{ for } t > 0. \tag{5.1}$$

By the fact that $W(t)$ is normally distributed with mean zero which is symmetric. Note that

$$\mathbb{P}\{\tau = 0\} = \mathbb{P}(\bigcap_{n=1}^{\infty} \{\tau \leq \frac{1}{n}\}) = \lim_{n \to \infty} \mathbb{P}(\tau \leq \frac{1}{n}) \geq \frac{1}{2}.$$

By Theorem 2.1.9 and (5.1), which yields the desired result.

We will now prove the second claim, $\mathbb{P}\{\omega = 0\} = 1$. We will be using the same argument after remarking that $\{\omega = 0\}$ occurs when in an infinitesimal close interval to the right of zero, we need to cross the t-axis infinitely many times, i.e.,

$$\begin{aligned}\{\omega = 0\} = &\bigcap_{n=1}^{\infty} \Big\{ \text{ there is a } 0 < t_i < 1/n \text{ such that } B(t_i) > 0 \Big\} \\ &\cap \bigcap_{n=1}^{\infty} \Big\{ \text{ there is a } 0 < s_i < 1/n \text{ such that } B(s_i) < 0 \Big\}\end{aligned}$$

Since the Brownian motion has continuous paths and is symmetric, we take the intersection of two events with a probability of one. Therefore, $\mathbb{P}\{\omega = 0\} = 1$. ∎

5.2 Local and global maxima of a Brownian motion

As an application of our discussion the Markov property for Brownian motion, we will study the local and global extrema of a Brownian motion. Before we do so, we will first look at whether Brownian motion has any intervals of monotonicity in the following Theorem.

Theorem 5.2.1. *Almost surely, for all* $0 < a < b < \infty$*, Brownian motion is not monotone on the interval* $[a, b]$

Proof. First fix a non-degenerate interval $[a, b]$, i.e. $a < b$. If it is an interval of monotonicity, i.e. $W(s) \leq W(t)$ for all $a \leq s \leq t \leq b$, then we divide the interval $[a, b]$ into n sub-intervals $[a_i, a_{i+1}]$ for $a = a_1 \leq a_2 \leq \cdots \leq a_{n+1} = b$. By assumption the interval $[a, b]$ is an interval of monotonicity, and therefore each of the increments $W(a_i) - W(a_{i-1})$ has to have the same sign.
We note that $\mathbb{P}\big(W(a_i) - W(a_{i+1}) > 0\big) = \mathbb{P}\big(W(a_i) - W(a_{i+1}) < 0\big) = \frac{1}{2}$ as the increments are normally distributed with mean zero. The increments are also independent. From this we see that the event that each increment $W(a_i) - W(a_{i-1})$ has the same sign has probability $2^{-n} + 2^{-n}$. Taking $n \to \infty$ shows that the probability that $[a, b]$ is an interval of monotonicity must be zero.

By taking the countable union over all intervals with rational endpoints, we see that almost surely $\{W(t) : a \leq t \leq b\}$ is non-monotonic on any interval with rational endpoints. By the density of $\mathbb{Q} \subset \mathbb{R}$, there is an interval with rational endpoints contained within every real interval. So every interval contains a sub-interval which is almost surely non monotonic, and therefore every interval is almost surely non monotonic. ([8]) ■

Theorem 5.2.1 shows us one manifestation of why Brownian motion is erratic. Now let us conclude our discussion of the Markov property by exploiting it to study the local and global extrema of a Brownian motion.

Definition 5.2.2 (Local Maximum)**.** *A point c in the domain of some function f is called a local maximum if there exists some $\delta > 0$ such that*

$$f(x) \leq f(c)$$

for all $x \in (c - \delta, c + \delta)$. Furthermore, if $f(x) < f(c)$ we call this local maximum a strict local maximum.

Theorem 5.2.3. *For a standard Brownian motion $\{B(t) : 0 \leq t \leq 1\}$, almost surely,*

i. every local maximum is a strict local maximum;

ii. the set of times where the local maxima are attained is countable and dense;

iii. the global maximum is attained at a unique time.

Proof. Let $a_1 < b_1 \leq a_2 < b_2$ so that $[a_1, b_1]$ and $[a_2, b_2]$ are non-degenerate, non-overlapping intervals, i.e. the intervals have disjoint interiors, with

$$m_1 = \max_{a_1 \leq t \leq b_1} W(t) \text{ and } m_2 = \max_{a_2 \leq t \leq b_2} W(t).$$

We will first show that one can assume with no loss of generality that $b_1 < a_2$. By the Markov property (Theorem 5.1.2), $\big\{W(t + a_2) - W(a_2) : t \geq 0\big\}$ is a Brownian motion starting in 0, which is independent of the process $\{W(t) : 0 \leq t \leq a_2\}$. Now, if we consider $\tau = \tau_{a_2}$ as given in Theorem 5.1.5, namely

$$\tau_{a_2} = \inf\{t > 0 : W(t + a_2) - W(a_2) > 0\} = \inf\{t > 0 : W(t + a_2) > W(a_2)\},$$

we have seen that

$$\mathbb{P}_0\big(\tau_{a_2} = 0\big) = 1,$$

which means that there are infinitely many $t > 0$ such that $W(t + a_2) > W(a_2)$, implying that $W(a_2)$ cannot be a local maximum, so $W(a_2) < m_2$. Since the paths of a Brownian motion are continuous, we can find an n such that m_2 is the maximum on the interval $[a_2 + 1/n]$.
So if is the case that $b_1 = a_2$, then we can replace the interval $[a_2, b_2]$ by $[a_2 + 1/n, b_2]$ without changing the maxima and now $b_1 < a_2 + 1/n$, so we may assume that $b_1 < a_2$.

We will proceed by showing that the maxima on our two intervals are different almost surely. Applying Theorem 5.1.2 at time b_1, we see that the random variable $W(a_2) - W(b_1)$ is independent of $m_1 - W(b_1)$. Now using Theorem 5.1.2 at time a_2, we see that $m_2 - W(a_2)$ is also independent of both these variables. The event $\{m_1 = m_2\}$ can be written as

$$W(a_2) - W(b_1) = m_1 - W(b_1) - (m_2 - (W(a_2)).$$

Now, if we condition on the values of the random variables $X = m_1 - W(b_1)$ and $Y = m_2 - W(a_2)$, then

$$\begin{aligned}\mathbb{P}(m_1 = m_2) = &\int\int \mathbb{P}(W(a_2) - W(b_1) = X - Y \big| X = x, Y = y) f_{X,Y}(x,y)dydx = \\ &\int\int \mathbb{P}(W(a_2) - W(b_1) = X - Y \big| X = x, Y = y) f_X(x) f_Y(y)dydx = \\ &\int\int \mathbb{P}(W(a_2) - W(b_1) = x - y) f_X(x) f_Y(y)dydx = \\ &\int\int 0 \cdot f_X(x) f_Y(y)dydx = 0\end{aligned}$$

Where we used in the first equality that X and Y are independent random variables. In the second equality we used that X and Y are independent of $W(a_2) - W(a_1)$. Lastly we used that $W(a_2) - W(a_1)$ is a continuous random variable.

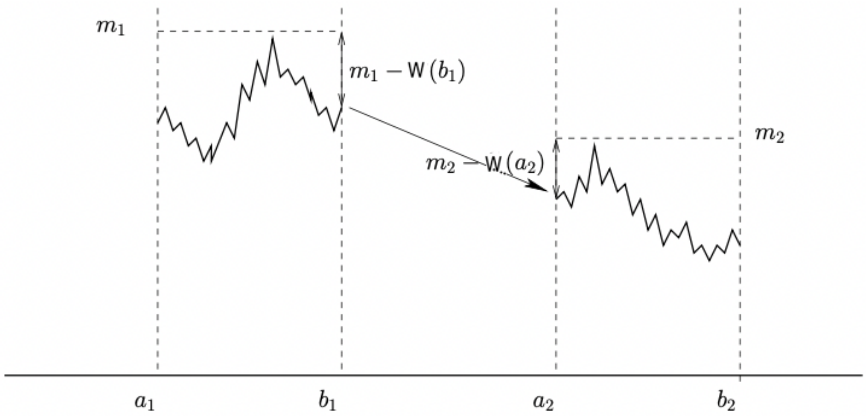

Figure 5.2: The random variables $m_1 - W(b_1)$ and $m_2 - W(b_2)$ are independent of the increment $W(a_2) - W(b_1)$.[1]

Let us do a short recap. Thus far we have shown that given two non-degenerate, non-overlapping intervals $[a_1, b_1]$ and $[a_2, b_2]$, we may assume that without loss of generality that $b_1 < a_2$. Furthermore, we also proved that the maxima on these intervals, m_1, m_2, respectively are different almost surely.
We will now proceed with the proof of $(i.)$. By the statement we just proved, almost surely, all non-overlapping pairs of non-degenerate compact intervals with rational endpoints have different maxima. If Brownian motion were to have a non-strict local maximum, there would exist two such intervals where the Brownian motion has the same maximum, which contradicts what we have previously proven.

For the proof of $(ii.)$, we note that in particular, almost surely, the maximum over any non-degenerate compact interval with rational endpoints is not attained at an endpoint. Since the rationals are countable and dense in $\mathbb{R}$, and almost surely all local maxima are strict by *(i.)*, we see that the set of times where to local maxima are attained is countable and dense.

We conclude with the proof of $(iii.)$. As we have seen, almost surely, for any rational $q \in (0, 1)$, we maximum on the non-overlapping intervals $[0, q]$ and $[q, 1]$ are different. So if the global maximum is not unique, i.e., it is attained at $0 < t_1 < t_2 < 1$, then since the rationals are dense, we can find a rational number $t_1 < q < t_2$ for which the maximum in $[0, q]$ and $[q, 1]$ agree, contradicting $(i.)$. ■

5.3 The Strong Markov Property

Thus far we have discussed the Markov property, which informally states that at a *fixed* time $s > 0$, the Brownian motion starts afresh, meaning that it can be seen as a standard Brownian motion and hence independent of $W(t)$ for $0 \leq t < s$.
In this chapter, we would like to extend the Markov property for Brownian motion to random times as well, a crucial property. It turns out that for the class of random times called *stopping times*, the Markov property also holds.

We will start this section by defining and shortly discussing stopping times. After this we will state the strong Markov property and give the reader some intuition of why it is important to consider stopping time in stead of any random time. Last but not least we will look at an interesting application of the Strong Markov property, the reflection principle.

Definition 5.3.1 (Stopping time). *A random variable $T \in [0, \infty]$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with filtration $\{\mathcal{F}(t) : t \geq 0\}$ is called a stopping time if the event $\{T \leq t\} \in \mathcal{F}(t)$ for every $t > 0$.*

Now we will look at some fact and properties of stopping times associated with Brownian motions.

- Every deterministic time $s \leq 0$ is a stopping time with respect to every filtration $\{\mathcal{F}(t) : t \geq 0\}$
- If $\{T_n : n = 1, 2, ...\}$ is an increasing sequence of stopping times with respect to $\{\mathcal{F}(t) : t \geq 0\}$ and $T_n \uparrow T$, then T is also a stopping time with respect to $\{\mathcal{F}(t) : t \geq 0\}$. To see this, note that

$$\{T \leq t\} = \bigcap_{n=1}^{\infty} \{T_n \leq t\} \in \mathcal{F}(t)$$

- Every stopping time T with respect to $\{\mathcal{F}^0(t) : t \geq 0\}$ is a stopping time with respect to $\{\mathcal{F}^+(t) : t \geq 0\}$ as $\mathcal{F}^0(t) \subset \mathcal{F}^+(t)$ for every $t \geq 0$.

The following result indicates the technical advantage of right-continuous filtrations as mentioned in Remark 5.1.3

Lemma 5.3.2. *Suppose a random variable T on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with values in $[0, \infty]$ satisfies $\{T < t\} \in \mathcal{F}(t)$ for every $t \geq 0$. If $\{\mathcal{F}(t) : t \geq 0\}$ is right-continuous, then T is a stopping time with respect to $\{\mathcal{F}(t) : t \geq 0\}$*

Definition 5.3.3. *The first hitting time of a value $b \neq 0$ is given by*

$$T_b = \inf\{t \geq 0 : W(t) = b\}$$

Lemma 5.3.4. *The first hitting time of a value b is a stopping time with respect to $\{\mathcal{F}^0(t) : t \geq 0\}$.*

Proof. Note that

$$\left\{T_b \leq t\right\} = \left\{\max_{0 \leq s \leq t} W(s) \geq b\right\} \in \mathcal{F}^0(t)$$

■

We will illustrate with an example that we need to be mindful about the assumption that in order for the strong Markov Property to hold, we need to have a stopping time T rather than just any random time.

Example 5.3.5. *Let S be a random time where the Brownian motion $\{W(t) : t \geq 0\}$ attains a strict local maximum. Therefore, there exists some $\delta > 0$ such that for all $r \in (S - \delta, S + \delta) : W(r) < W(S)$. Write $\{X(t) : t \geq 0\}$ for the Brownian motion defined by $X(t) = W(t + S) - W(S)$. If S were a stopping time, Theorem 5.3.8 would suggest that $\{X(t) : t \geq 0\}$ is a Brownian motion independent of $\mathcal{F}^+(S)$. This is however false. To see this, we write*

$$\mathbb{P}(X(\delta/2) - X(0) < 0) = \mathbb{P}(W(s + \delta/2) - W(s) < 0) = 1$$

Implying that the increment $X(\delta/2) - X(0)$ is not normally distributed and therefore $\{X(t) : t \geq 0\}$ cannot be a Brownian motion.

This example clarifies why the assumption of a stopping time is so important. We want to point out that intuitively, it makes sense that $\{X(t) : t \geq 0\}$ from Example 5.3.5 is not a Brownian motion.
Given our heuristic understanding of $\mathcal{F}^+(S)$ as the collection of events that happened before the stopping time S. However, to know whether $W(t)$ attains a strict local maximum at time S, we need to check whether $W(s + \delta) < W(s)$ for $\delta > 0$. This is information not contained in $\mathcal{F}^+(S)$.

Lemma 5.3.6 (**Dyadic stopping times**). *Let T be a stopping time defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with respect to the filtration $\{\mathcal{F}(t) : t \geq 0\}$. Define times T_n by*

$$T_n = \frac{(m+1)}{2^n} \text{ if } \frac{m}{2^n} \leq T < \frac{(m+1)}{2^n}$$

Then T_n is a stopping time with respect to $\{\mathcal{F}(t) : t \geq 0\}$.

Proof. For every $t \geq 0$ we can find a $l \in \mathbb{N}$ such that $\frac{l}{2^n} \leq t < \frac{l+1}{2^n}$. This implies that

$$\{T_n \leq t\} = \{T_n \leq \frac{l}{2^n}\} = \{T < \frac{l}{2^n}\} = \bigcup_{q < \frac{l}{2^n}, q \in \mathbb{Q}} \{T \leq q\}$$

Since T is a stopping time by assumption, $\{T \leq q\} \in \mathcal{F}(q) \subset \mathcal{F}(\frac{l}{2^n}) \subset \mathcal{F}(t)$. Therefore, $\{T_n \leq t\} \in \mathcal{F}(t)$ which is the desired result. ■

For every stopping time T we define the σ-algebra

$$\mathcal{F}^+(T) = \big\{A \in \mathcal{A} : A \cap \{T \leq t\} \in \mathcal{F}^+t) \text{ for all } t \geq 0\big\}$$

Heuristically, this is the collection of events that happened before the stopping time T.

Remark 5.3.7. *In previous chapters we denoted the triple $(\Omega, \mathcal{F}, \mathbb{P})$ as the probability space. We are now looking at the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ to illustrate the difference between the σ-algebra $\mathcal{A}$ and the filtration $\{\mathcal{F}(t) : t \geq 0\}$.*

We are now ready to look at the Strong Markov property.

Theorem 5.3.8 (**Strong Markov Property**). *For every almost surely finite stopping time T, the process $\{W(t + T) - W(T), t \geq 0\}$ is a standard Brownian motion independent of $\mathcal{F}^+(T)$.*

Proof. We will first show that the statement holds for $T_n = \frac{m+1}{2^n}$ if $\frac{m}{2^n} \leq T < \frac{m+1}{2^n}$, which are stopping times by Lemma 5.3.6.
We will first look at the process defined by $W_* = \{W_*(t) : t \geq 0\}$ where $W_*(t) = W(t+T_n) - W(T_n)$ and show that this is a standard Brownian motion independent of $\mathcal{F}^+(T_n)$. We write $W_k = \{W_k(t) : t \geq 0\}$ for the Brownian motion defined by $W_k(t) = W(t + k2^{-n}) - W(k2^{-n})$.
Suppose that $E \in \mathcal{F}^+(T_n)$. Then for every event $\{W_*(t) \in B\}, B \in \mathcal{B}(\mathbb{R})$, we have

$$\begin{aligned}
\mathbb{P}\big(\{W_*(t) \in B\} \cap E\big) &= \mathbb{P}\Big(\bigcup_{k=0}^{\infty} \{W_k \in B\} \cap \{T_n = \frac{k}{2^n}\} \cap E\Big) \\
&= \sum_{k=0}^{\infty} \mathbb{P}\Big(\{W_k(t) \in B\} \cap \{T_n = \frac{k}{2^n}\} \cap E\Big) \\
&= \sum_{k=0}^{\infty} \mathbb{P}\{W_k \in B\} \mathbb{P}\big(E \cap \{T_n = \frac{k}{2^n}\}\big)
\end{aligned}$$

Where we used that Theorem 5.1.4 to see that $\{W_k \in B\}$ is independent of $E \cap \big\{T_n = k2^{-n}\big\} \in \mathcal{F}^+(k2^{-n})$, noting that $k2^{-n}$ is a deterministic time. We now remark that the process $W_k(t) = W(t + k2^{-n}) - W(k2^{-n})$ is $\mathcal{N}(0, t)$ distributed and hence has the same distribution as $W(t) = \{W(t) : t \geq 0\}$. Therefore,

$$
\begin{aligned}
\mathbb{P}\big(\{W_*(t) \in B\} \cap E\big) &= \sum_{k=0}^{\infty} \mathbb{P}(W(t) \in B)\mathbb{P}\big(\{T_n = k2^{-n}\} \cap E\}\\
&= P(W(t) \in B)\mathbb{P}\Big(\bigcup_{k=0}^{\infty} \{T_n = k2^{-n}\} \cap E\Big)\\
&= \mathbb{P}(W(t) \in B)\mathbb{P}(E)
\end{aligned}
$$

Which shows that W_* is independent of E and hence of $\mathcal{F}^+(T_n)$. If we now take $E = \Omega$, we see that $W_*(t)$ and $W(t)$ have the same distribution. Therefore, $W_*(t)$ is a standard Brownian motion independent of $\mathcal{F}^+(T_n)$.

To conclude the proof, we have to generalise the above result to stopping times T. Note that $T_n \downarrow T$ and $T_n \geq T$ for all $n \in \mathbb{N}$. Since $\mathcal{F}^+(T) \subset \mathcal{F}^+(T_n)$,the standard Brownian motion $\{W(s+T_n)-W(T_n) : s \geq 0\}$ is also independent of $\mathcal{F}^+(T)$. Hence the increments

$$
W(s+t+T) - W(t+T) = \lim_{n\to\infty} W(s+t+T_n) - W(t+T_n) \tag{5.2}
$$

of the process $\{W(r+T) - W(T) : r \geq 0\}$ are independent and normally distributed with mean zero and variance s. Since the process is almost surely continuous, it is a Brownian motion. Lastly, all of the increments as in 5.2, and thus the process itself, are independent of $\mathcal{F}^+(T)$. ■

5.3.1 The Reflection Principle

As an application of the Strong Markov property we will take a look at the reflection principle. This interesting property states that the Brownian motion reflected at some stopping time T is still a Brownian motion. In this section we will prove this and afterwards we will show how the reflection principle can be used to determine the distribution of the event $\{M(t) \geq a\}$, where $M(t) = \max_{0\leq s\leq t} W(s)$.
We will first look at the reflection principle, see Figure 5.3.This principle states that if the path of a Brownian motion $\{W(t) : t \geq 0\}$ reaches a stopping time T, the subsequent path after time T has the same distribution as the reflection of this path in the value $W(T)$.

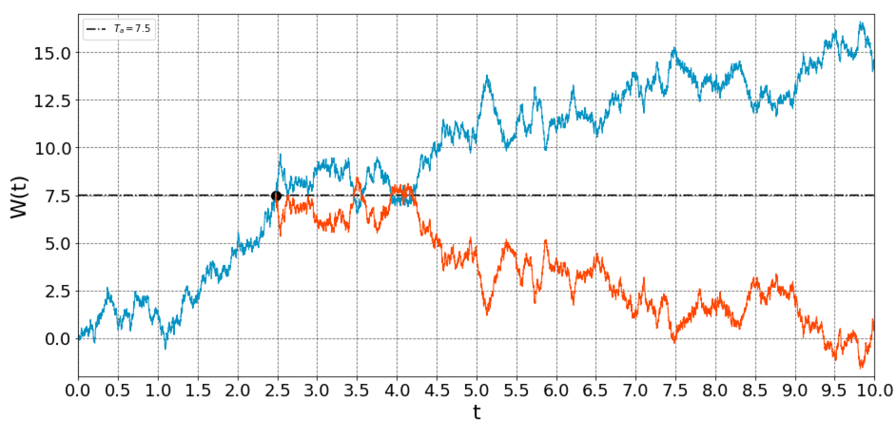

Figure 5.3: The Reflection Principle, where T_a is the first hitting time of the value 7.5.

Theorem 5.3.9 (Reflection principle). *Let $\{W(t) : t \geq 0\}$ be a standard Brownian motion and let T be a stopping time. Then the process $\{W^T(t) : t \geq 0\}$ called Brownian motion reflected at T and defined by*

$$
W^T(t) = W(t)1_{\{t\leq T\}} + (2W(T) - W(t))1_{\{t>T\}}
$$

is also a standard Brownian motion.

Proof. [9]
If T is an almost surely finite stopping time, by the strong Markov property the path

$$\big\{W(t+T) - W(T) : t \geq 0\big\}$$

is a Brownian motion started in zero and independent of $\{W(t) : 0 \leq t \leq T\}$ As in Example 4.3.1, $\big\{-(W(t+T) - W(T)) : t \geq 0\big\}$ has the same distribution as $\big\{W(T) - W(t+T) : t \geq 0\big\}$ and is therefore also a standard Brownian motion independent of $\{W(t) : 0 \leq t \leq T\}$.

Note that we can write $W(t) = W(t) + W(T) - W(T)$. Then

$$\begin{aligned}\mathbb{P}\{W(t), t > T\} &= \mathbb{P}\{W(T) + W(t) - W(T), t > T\}\\ &= \mathbb{P}\{W(T) + W(T) - W(t), t > T\}\\ &= \mathbb{P}\{2W(T) - W(t), t > T\},\end{aligned}$$

which yields the desired result.

■

Lastly we will give an application of the reflection principle. Let $M(t) = \max_{0 \leq s \leq t} W(s)$. Our goal is to determine the distribution of this random variable, which is possible as a result of the reflection principle.

Theorem 5.3.10. *If $a > 0$ then $\mathbb{P}(M(t) > a) = 2\mathbb{P}(W(t) \geq a)$.*

Proof. Let $T_a = \inf\{t \geq 0 : W(t) = a\}$. Then T_a is a stopping time and therefore by Theorem 5.3.9 the Brownian motion reflected at T_a is of the form

$$W^{T_a}(t) = \begin{cases} W(t) \text{ if } t \leq T_a \\ 2a - W(t) \text{ if } t > T_a \end{cases}$$

Consider the event

$$\begin{aligned}\{M(t) > a\} &= \{M(t) > a, W(t) > a\} \cup \{M(t) > a, W(t) \leq a\}\\ &= \{W(t) > a\} \uplus \{M(t) > a, W(t) \leq a\}\\ &= \{W(t) > a\} \uplus \{W^{T_a}(t) \geq a, T_a < t\}.\end{aligned}$$

So

$$\begin{aligned}\mathbb{P}(M(t) > a) &= \mathbb{P}(W(t) > a) + \mathbb{P}(W^{T_a} \geq a, T_a < t)\\ &= \mathbb{P}(W(t) \geq a) + \mathbb{P}(2a - W(t) \geq a)\\ &= 2\mathbb{P}(W(t) \geq a).\end{aligned}$$

■

Chapter 6

Conclusion

In this thesis we took a deep dive into the theory behind the one-dimensional Brownian motion. We first derived how we could use the simple type of stochastic processes, random walks, to construct Brownian motion, giving us intuition behind the definition of a standard Brownian. (see Definition 4.1.1). As Brownian motion is a martingale, we know that it has no natural tendency to go up or down. Looking at the distributional properties of Brownian motion, we saw that Brownian motion is scaling invariant, implying that if we look at an arbitrary small part of a Brownian motion, it will always be a Brownian motion. The fact that Brownian motion has non-zero quadratic variation then motivated us to prove that Brownian motion is nowhere differentiable.

Next, we proved the Markov property as well as the strong Markov property, which states that that for any fixed time $s > 0$ or any stopping time T, respectively, the process started in these points is again a standard Brownian motion and is independent of everything that happened up until s respectively T. The Markov property allowed us to predict the behaviour of a Brownian motion on an infinitesimally small interval to the right of zero, which we used to show that every local maximum of a standard Brownian motion is a strict local maximum and the global maximum is attained at a unique time.
At last, we applied the strong Markov property to prove the reflection principle, which informally states that after the reaches some stopping time T, the subsequent path of the Brownian motion can be reflected on the value of the Brownian motion at T which will gives us another standard Brownian motion.

Brownian motion finds it applications in many things, such as in the modeling of the stock market, medical imaging and decision making [11]. These applications were beyond the scope of this thesis, but that does not mean that they are not worth the attention of the reader. Brownian motion is truly a wonderful topic, as it ties in so well the randomness of the world we live in. As humans, we desire to control everything, even the thing we cannot control. That is why probability and statistics are such a favorable tool and the beauty of Brownian motion is that from the randomness that characterizes the process, we can make fairly accurate predictions about how the process will behave later, which can help us predict our unpredictable world.

Bibliography

[1] Mörters, P. & Peres, Y. (2010). *Brownian Motion (Cambridge Series in Statistical and Probabilistic Mathematics)*. New York, USA: Cambridge University Press.

[2] Schreve, S. (2010). *Stochastic Calculus for Finance II*. New York, USA: Springer Publishing.

[3] Schilling, R.L. (2005). *Measures, Integrals and Martingales*. Cambridge, UK: Cambridge University Press.

[4] Choe, G. H. (2016). *Stochastic Analysis for Finance with Simulations*.New York, USA: Springer Publishing

[5] Ermogenous, A. (2006) *Brownian Motion and Its Applications in The Stock Market*. Undergraduate Mathematics Day, Electronic Proceedings. Paper 15. Retrieved on 29-11-20 from https://ecommons.udayton.edu/mth_epumd/15/

[6] Jagannathan, K. (2015) *Lecture 11: Random Variables*. (Lecture notes from Probability Foundations for Electrical Engineers.) Indian Institute of Technology Madras. Retrieved on from https://nptel.ac.in/content/storage2/courses/108106083/lecture11_rvs.pdf

[7] Gamarnik, D. (2013) *Lecture 6 : Introdcution to Brownian motion*. (Lecture notes from Advanced Stochastic Processes). Massachusetts Institute of Technology. Retrieved on 29-11-20 from https://ocw.mit.edu/courses/sloan-school-of-management/15-070j-advanced-stochastic-processes-fall-2013/lecture-notes/MIT15_070JF13_Lec6.pdf

[8] Dahl, A. (2010) *A Rigorous Introduction to Brownian Motion*(Paper). University of Chicago. Retrieved on 29-11-20 from http://www.math.uchicago.edu/~may/VIGRE/VIGRE2010/REUPapers/Dahl.pdf

[9] Huang, Y. (2014) *Lecture 23 Hitting Time, Maximum, Reflection Principle.*(Lecture notes from Introduction to Probability Models). The University of Chicago. Retrieved from http://www.stat.uchicago.edu/~yibi/teaching/stat317/2014/Lectures/Lecture23_6up.pdf

[10] Copur, Z. (2015)*Handbook of Research on Behavioral Finance and Investment Strategies: Decision Making in the Financial Industry* Ankara, Turkey: Hacettepe University.

[11] Duncan, T. (2007)*Brownian Motion: A Study of Its Theory and Applications*. BSc thesis, Boston College, Boston. Retrieved on 29-11-20 from https://dlib.bc.edu/islandora/object/bc-ir%3A102098/datastream/PDF/view

[12] Gamarnik, D. (2013) *Lecture 8: Quadratic variation property of Brownian motion*. (Lecture notes from Advanced Stochastic Processes). Massachusetts Institute of Technology. Retrieved on 13-01-21 from https://ocw.mit.edu/courses/sloan-school-of-management/15-070j-advanced-stochastic-processes-fall-2013/lecture-notes/MIT15_070JF13_Lec8.pdf