SJÄLVSTÄNDIGA ARBETEN I MATEMATIK

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET

Kolmogorov's Theorem, Haar Functions and Brownian Motion
A Measure Theoretic Introduction

av

Samuel Lockman

2022 - No K31

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET, 106 91 STOCKHOLM

Kolmogorov's Theorem, Haar Functions and Brownian Motion
A Measure Theoretic Introduction

Samuel Lockman

Självständigt arbete i matematik 15 högskolepoäng, grundnivå

Handledare: Salvador Rodríguez-López

2022

Abstract

In this text, we seek to give a measure theoretic introduction to Brownian Motion and prove two of its main properties. Part of it is done through the continuity theorem of Kolmogorov, which in itself can be considered a central result. Furthermore, the paper also includes a construction of a Brownian motion via the Haar functions by first proving it is an orthonormal basis of $L^2[0, 1]$.

Contents

Chapter 1

Introduction

Brownian motion originates as a physical phenomena discovered by Robert Brown, from whom the name has its clear origin, in 1827 while he was examining pollen particles of a plant submerged in water. He describes the movement of each particles as very irregular, and seemingly independent of other particles and its previous path. This notion was further elaborated upon by Albert Einstein, where he described how the pollen particles were being moved by individual water molecules. Interestingly, this was one of Einstein's first scientific contributions only at 26 years of age. As a consequence, this would finally settle the scientific debate on the existence of atoms. This was done through the work of Jean Perrin, who verified Einstein's theoretical work experimentally and was even able to determine the size of certain atoms. It was due to this work that Perrin was awarded the 1926 Nobel price in physics. Mathematically, it was first given a fully rigorous construction by Norbert Wiener, who used the language of measure theory to formalize precedent mathematical efforts of such a construct. [4, Chapter 1]

As a mathematical object, Brownian motion has many intriguing properties and we shall go on to examine its path regularities. We will state and prove the almost sure Hölder continuity of the sample paths and the iterated logarithm law. The first of which will be proven through Kolmogorov's continuity theorem, which in itself is a theorem we shall state and give a full proof of. Furthermore, the text also includes a construction of a Brownian motion via the Haar functions, by first showing that those functions is an orthonormal basis of $L^2([0,1])$. These statements relies heavily on the language of measure theory, functional analysis, probability theory and stochastic processes, which we aim to give a solid presentation of.

Excluding this introductory chapter, the thesis shall be divided into four chapters; each of which discussing one of the topics mentioned above. We begin by trying to present the reader with an overview of what is needed for our main results in later chapters along with some side results. The following chapter is devoted to presenting the full statement and proof of Kolmogorov's continuity theorem along with relevant information regarding Hölder-continuity. For our later construction of a Brownian motion, we shall as mentioned before, need the fact that the Haar functions is an orthonormal basis of $L^2([0,1])$; something we have devoted an entire chapter for. The last chapter will be discussing the central part of the thesis, Brownian motion. We begin by a construction via the Haar functions to, in some sense, show the existence of a Brownian motion. Further, we state and prove two of its path properties in the final section, which allows us to understand one of its smoothness properties and one of its limiting behaviours. If time had allowed it, a further topic included in this chapter would have been the almost sure nowhere differentiability of the sample paths which may be found in the book of Mörters & Peres [3].

Chapter 2

Preliminaries

Throughout the text, we seek to make rigorous and precise statements, and to do so, this first chapter will be used as a building block for being able to make such statements.

2.1 Measure Theory and Functional Analysis

The proof of Theorem 2.1.2 is of the lines of the proof found in Friedman for the same theorem [2, Theorem 6.4.3], but the rest of proofs in this section are either somewhat trivial or my own creation.

Definition 2.1.1. Given a set Ω, we say that a family, $\mathcal{A}$, of subsets of Ω defines a σ-algebra on Ω if the following three conditions hold

$$\begin{aligned}
&\text{i) } \emptyset \in \mathcal{A},\\
&\text{ii) } A \in \mathcal{A} \Longrightarrow A^c \in \mathcal{A},\\
&\text{iii) } A_1, A_2, \ldots \in \mathcal{A} \Longrightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}.
\end{aligned}$$

Remark 2.1.1. Note that for any countable set of elements $A_1, A_2, \ldots \in \mathcal{A}$, both $A_1 \setminus A_2 \in \mathcal{A}$ and $\cap_{n=1}^{\infty} A_n \in \mathcal{A}$ since we can write

$$\begin{aligned}
A_1 \setminus A_2 &= A_1 \cap A_2^c = (A_1^c \cup A_2)^c, \text{ and}\\
\cap_{n=1}^{\infty} A_n &= ((\cap_{n=1}^{\infty} A_n)^c)^c = (\cup_{n=1}^{\infty} A_n^c)^c.
\end{aligned}$$

Lemma 2.1.1. *Given any non-empty family $\mathcal{U}$ of subsets of a set Ω, the family given by*

$$\sigma(\mathcal{U}) = \cap\{H \mid H \ \sigma\text{-algebra of } \Omega,\ \mathcal{U} \subset H\} \tag{2.1}$$

defines a σ-algebra on Ω. Any sigma algebra that contains $\mathcal{U}$ also contains $\sigma(\mathcal{U})$

Proof. Since the power set of Ω is a σ-algebra, $\sigma(\mathcal{U})$ is non-empty. Since properties ii) and iii) holds for all such σ-algebras H defined above, it also holds for $\sigma(\mathcal{U})$. The last part of the statement follows trivially since any intersection of sets $\cap_{n=1}^{\infty} H_n$ is a subset of each H_i, $i = 1, 2, \ldots$. □

Definition 2.1.2. We say that $\sigma(\mathcal{U})$ above is the σ-algebra generated by $\mathcal{U}$. When the σ-algebra is generated by all the open sets of a topological space, then we call it the Borel σ-algebra. In this text, $\mathcal{B}(\mathbb{R}^n)$ will denote the Borel σ-algebra on $\mathbb{R}^n$, where $\mathbb{R}^n$ is given the standard topology.

Remark 2.1.2. When we say the standard topology on $\mathbb{R}^n$, we mean that a neighbourhood is defined via the Euclidean metric.

Lemma 2.1.2. *For any set $A \subset \mathbb{R}^n$, the following statements are equivalent.*

i) A is open,
ii) A can be written as a countable union of open balls of radius less than or equal to 1,
iii) A can be written as a countable union of open squares.

Proof. We intend to show that $i) \Rightarrow ii) \Rightarrow iii) \Rightarrow i)$.

Suppose A is open. Then, for every $x \in A$ there exists an open ball $B_{r_x}(x)$ centered at x of radius $r_x < 1$ that is contained in A. We may thus write $A = \bigcup_{x \in A} B_{r_x}(x)$. Now, for each $x \in A$, there exists a point $y \in \mathbb{Q}^n \cap A$ such that $\|x - y\| < \frac{r_x}{3}$, which in other words mean that both $x \in B_{\frac{r_x}{3}}(y)$ and $y \in B_{\frac{r_x}{3}}(x)$. Furthermore, for each $y' \in \mathbb{Q} \cap A$, let

$$R(y') := \{x \in A \mid x \in B_{\frac{r_x}{3}}(y')\}$$

and let $r_{x'} = \sup_{x \in R(y')} r_x$, which is finite by our above assumption on r_x. Define the ball $B^{(y')}$ as

$$B^{(y')} := B_{\frac{r_{x'}}{3}}(y'). \tag{2.2}$$

There exists an $x^* \in R(y')$ such that

$$B^{(y')} \subset B_{r_{\frac{x^*}{2}}}(y'),$$

since otherwise $r_{x'}$ would not be the supremum. Finally, since for any point $z \in B_{r_{\frac{x^*}{2}}}(y')$, we have that

$$\|z - x^*\| \leq \|z - y\| + \|y - x^*\| < \frac{r_{x^*}}{2} + \frac{r_{x^*}}{2} = r_{x^*},$$

which implies that

$$B^{(y')} \subset B_{r_{\frac{x^*}{2}}}(y') \subset B_{r_{x^*}}(x) \subset A.$$

By our above argument, every point in A is in some $B^{(y')}$, so in fact

$$A \subset \bigcup_{y \in \mathbb{Q}^n \cap A} B^{(y)} \subset A$$

and we have proved that $i)$ implies $ii)$.

Next, suppose we can write a set $A \subset \mathbb{R}^n$ as a countable union of open balls with radius less than or equal to 1, i.e $A = \bigcup_{j=1}^{\infty} B_{r_j}(x_j)$ where each $r_j \leq 1$. It is sufficient to show that each $B_{r_j}(x_j)$ can be written as a countable union of open squares. Fix some $B_r(t)$ from the union above. For each $x \in B_r(t)$ there exists a point $y \in \mathbb{Q}^n \cap B_r(t)$ and an open square of the form $(a, b)^n$ centered at y such that

$$\sqrt{n}\frac{b-a}{2} = \frac{r - \|x - t\|}{3},$$

where the left hand side above denotes the Euclidean distance from the center of the square to each of its corners. Denote this square by $K_{\frac{r-\|x-t\|}{3}}(y)$. The existence of such a square comes from the fact that the square $K_{\frac{r-\|x-t\|}{3}}(x)$ contains x. But since this square is open there exists at least one point $y \in \mathbb{Q}^n \cap B_r(t)$ that is contained in the square, so we may shift the square $K_{\frac{r-\|x-t\|}{3}}(x)$ to $K_{\frac{r-\|x-t\|}{3}}(y)$ and thus be sure that it contains x. Furthermore, for each $y' \in \mathbb{Q}^n \cap B_r(t)$ let

$$R(y') := \{x \in B_r(t) \mid x \in K_{\frac{r-\|x-t\|}{3}}(y')\}$$

and let $h(y') = \sup_{x \in R(y')} r - \|x - t\|$. Define the square $K^{(y')}$ as

$$K^{(y')} := K_{h(y')}(y')$$

and note that there exists a point $x^* \in R(y')$ such that

$$K^{(y')} \subset K_{\frac{r-\|x^*-t\|}{2}}(y'),$$

since otherwise $h(y')$ would not be the supremum. Finally, for any point $z \in K_{\frac{r-\|x^*-t\|}{2}}(y')$ we have that

$$\begin{aligned}\|z-t\| &\leq \|z-x^*\| + \|x^*-t\| \\ &< \frac{r-\|x^*-t\|}{2} + \|x^*-t\| \\ &= \frac{r+\|x^*-t\|}{2} < r,\end{aligned}$$

which allows us to conclude that

$$B_r(t) \subset \bigcup_{y \in \mathbb{Q}^n \cap B_r(t)} K^{(y')} \subset B_r(t).$$

Lastly, suppose A can be written as a countable union of open squares. Since any countable union of open sets is an open set, the statement follows. □

Definition 2.1.3. For any sequence of points $\{x_n\} \subset \mathbb{R}$, we define $\overline{\lim}_n x_n$ and $\underline{\lim}_n x_n$ as

$$\begin{aligned}\overline{\lim_n} x_n &= \inf_{n\geq 0}(\sup_{m\geq n} x_m), \\ \underline{\lim_n} x_n &= \sup_{n\geq 0}(\inf_{m\geq n} x_m).\end{aligned}$$

Similarly, for any sequence of sets $\{\mathcal{F}_n\}$, we define

$$\begin{aligned}\overline{\lim_n}\{\mathcal{F}_n\} &= \bigcap_{n=0}^{\infty}(\bigcup_{m=n}^{\infty}\{\mathcal{F}_m\}), \\ \underline{\lim_n}\{\mathcal{F}_n\} &= \bigcup_{n=0}^{\infty}(\bigcap_{m=n}^{\infty}\{\mathcal{F}_m\}).\end{aligned}$$

Lemma 2.1.3. *A point x is in $\overline{\lim}_n \mathcal{F}_n$ if and only if x is in $\mathcal{F}_n$ for infinitely many n.*

Proof. First, suppose x is in $\mathcal{F}_n$ for infinitely many n, then for any n, x will be in the set

$$\bigcup_{m=n}^{\infty} \mathcal{F}_m,$$

and so $x \in \overline{\lim}_n \mathcal{F}_n$. On the contrary, suppose that $x \in \overline{\lim}_n \mathcal{F}_n$ and suppose for a contradiction that x is not in infinitely many $\mathcal{F}_n$. Then there exists an n such that x is not in the set

$$\bigcup_{m=n}^{\infty} \mathcal{F}_m,$$

so $x \notin \overline{\lim}_n \mathcal{F}_n$ and we have established a contradiction. □

Definition 2.1.4 (Real Hilbert Space)**.** A non-empty set H is called a real Hilbert space if H is a linear space over the real field, together with a real-valued mapping, $\langle \cdot, \cdot \rangle$ with domain H^2 such that for all $f, g, h \in H$ and $c \in \mathbb{R}$,

i) $\langle f, f \rangle \geq 0$, and $\langle f, f \rangle = 0$ if and only if $f = 0$,
ii) $\langle f, g \rangle = \langle g, f \rangle$,
iii) $\langle f+g, h \rangle = \langle f, h \rangle + \langle g, h \rangle$,
iv) $\langle cf, g \rangle = c\langle f, g \rangle$,
v) if $\{f_n\}$ is a sequence in H such that $\langle f_n - f_m, f_n - f_m \rangle \to 0$ as $n, m \to \infty$, then there exists an element $f \in H$ such that $\lim_n \langle f_n - f, f_n - f \rangle = 0$.

Remark 2.1.3. Since this text will only include real Hilbert spaces, we shall omit the word "real" when referring to such spaces. However, extending this to the complex field only requires making use of complex conjugation, so indeed, everything proven in this chapter will be applicable to complex Hilbert spaces.

Remark 2.1.4. Property v) is the property corresponding to completeness of a normed linear space, but as we shall see, every Hilbert space H is actually also a complete normed linear space.

Definition 2.1.5. For a Hilbert space H, we say that the mapping $\|\cdot\|$ defined via

$$\|f\| = \langle f, f\rangle^{\frac{1}{2}}$$

for any $f \in H$, is the norm of f, induced by H.

Theorem 2.1.1 (Cuachy-Schwarz Inequality)**.** *For any elements f, g in a Hilbert space H*

$$|\langle f, g\rangle| \leq \|f\| \, \|g\|$$

Proof. If $f = g = 0$ then the statement follows trivially. Suppose therefore that f and g are non-trivial and note that for any t, we have that

$$0 \leq \|f - tg\|^2 = \|f\|^2 - 2t\langle f, g\rangle + t^2 \|g\|^2 .$$

Put $t = \frac{\langle f,g\rangle}{\|g\|^2}$, so that we get

$$0 \leq \|f\| - \frac{\langle f, g\rangle^2}{\|g\|^2},$$

which proves the statement. ☐

Lemma 2.1.4. *The norm induced by a Hilbert space H makes H into a complete normed real linear space.*

Proof. Recall that the norm of a normed real linear space X need to have the property that for any $f, g \in X$ and $c \in \mathbb{R}$,

$$\begin{aligned} &\text{i)} \;\; \|f\| \geq 0, \text{ and } \|f\| = 0 \text{ if and only if } f = 0, \\ &\text{ii)} \;\; \|cf\| = |c| \, \|f\| , \\ &\text{iii)} \;\; \|f + g\| \leq \|f\| + \|g\| . \end{aligned}$$

Suppose therefore that $f, g \in H$ and $c \in \mathbb{R}$. The first two properties follows by property i) and iv) of the definition of a real Hilbert space. Next, note that by the Cuachy-Schwarz Inequality, we have that

$$\|f + g\|^2 = \|f\| + \|g\| + 2\langle f, g\rangle \leq \|f\| + \|g\| + 2 \|f\| \, \|g\| = (\|f\| + \|g\|)^2$$

and the last property follows. For the completeness part, suppose that $\{f_n\}$ is a Cauchy sequence in H. By the definition of a norm induced by a Hilbert space, we have that

$$\|f_n - f_m\|^2 = \langle f_n - f_m, f_n - f_m\rangle \to 0 \text{ as } n, m \to \infty.$$

By property v) of the definition of a Hilbert space, we have that there exists an $f \in H$ such that

$$0 = \lim_n \langle f_n - f, f_n - f\rangle = \lim_n \|f_n - f\|^2 ,$$

so every Cauchy sequence in H converges to a point in H and thus the statement follows. ☐

Lemma 2.1.5. *The norm of a normed linear space X is a continuous function.*

Proof. Fix $\varepsilon > 0$ and a point $f \in X$. Let g be any point in X such that

$$\|f - g\| < \varepsilon,$$

so that by the reverse triangle inequality we get that

$$| \, \|f\| - \|g\| \, | \leq \|f - g\| < \varepsilon,$$

which proves the statement. ☐

Lemma 2.1.6. *Let S be a continuous function from a normed linear space X into another normed linear space Y and suppose $\{x_n\} \subset X$ is a sequence such that $\lim_n x_n = x$. Then*

$$\lim_n S(x_n) = S(x).$$

Proof. Fix $\varepsilon > 0$ and note that since S is continuous, there exists a $\delta > 0$ such that

$$\|y - x\| < \delta \Rightarrow \|S(y) - S(x)\| < \varepsilon.$$

Since $\lim_n x_n = x$, there exists a natural number N such that for all $n > N$

$$\|x_n - x\| < \delta,$$

which then implies that

$$\|S(x_n) - S(x)\| < \varepsilon$$

and the statement follows. □

Corollary 2.1.1. *Let $\{x_n\}$ be a sequence in a normed linear space that converges to x, then*

$$\lim_n \|x_n\| = \|x\|.$$

Proof. This follows by Lemma 2.1.5 and Lemma 2.1.6. □

The following remark is only a side note and is not needed for the coherence of the remainder of this text.

Remark 2.1.5. The reader might be familiar with the concept of a measure space $(X, \mathcal{A}, \mu)$ and the space $L^1(X, \mathcal{A}, \mu)$ consisting of all all real valued $\mathcal{A}$-measurable functions f such that $|f|$ is integrable. This is a normed linear space with the norm given by

$$\|f\| = \int_X |f| d\mu.$$

Furthermore, by the above corollary, it follows that for any sequence of non-negative functions f_n that converges to f, it holds that

$$\lim_n \int_X f d\mu = \int_X f d\mu,$$

which is a special case of the dominated convergence theorem.

Definition 2.1.6. A sequence $\{f_n\}$ in a Hilbert space H is called orthonormal if

$$\langle f_n, f_m \rangle = \begin{cases} 0 \text{ if } n \neq m \\ 1 \text{ otherwise }. \end{cases}$$

Theorem 2.1.2. *Let $\{f_n\}$ be an orthonormal sequence in a Hilbert space and let $\{c_n\}$ be any sequence of real numbers. Then the series $\sum c_n f_n$ converges if and only if $\sum c_n^2$ converges, and in that case*

$$\left\| \sum_{n=0}^{\infty} c_n f_n \right\|^2 = \sum_{n=0}^{\infty} c_n^2.$$

Proof. By completeness, the first series in the statement converges if and only if, for $N > M$

$$\left\| \sum_{n=M}^{N} c_n f_n \right\|^2 \to 0 \text{ as } N, M \to \infty.$$

Consider the computations

$$\left\|\sum_{n=M}^{N} c_n f_n\right\|^2 = \langle \sum_{n=M}^{N} c_n f_n, \sum_{j=M}^{N} c_j f_j\rangle = \sum_{n=M}^{N} c_n \langle f_n, \sum_{j=M}^{N} c_j f_j\rangle$$
$$= \sum_{n=M}^{N} c_n \sum_{j=M}^{N} c_j \langle f_n, f_j\rangle,$$

and since $\langle f_n, f_j\rangle$ equals 1 if and only if $n = j$ and otherwise equals 0 we get that the above equals

$$\sum_{n=M}^{M} c_n^2,$$

so the first part of the statement follows. For the second part, suppose both series converges. Then, by Corollary 2.1.1, we have that

$$\left\|\lim_N \sum_{n=0}^{N} c_n f_n\right\|^2 = \lim_N \left\|\sum_{n=0}^{N} c_n f_n\right\|^2,$$

so by our previous computations, the theorem follows. □

2.2 Probability Theory

The proofs of this section was mostly just clarified from what was found in Baldi and Friedman [1], [2]. However, some proofs are solely my own creation; for example Lemma 2.2.4 and Theorem 2.2.1. Both the Examples in this section were created by myself.

Definition 2.2.1 (Probability space)**.** Given a set Ω and a family $\mathcal{A}$ of subsets of Ω such that $\mathcal{A}$ is a σ-algebra, we say that the space $(\Omega, \mathcal{A})$ is a measurable space. Further, a function $\mathbb{P}$ from $\mathcal{A}$ into the unit interval is called a probability measure on $(\Omega, \mathcal{A})$ if it maps the empty set to 0, the whole space Ω to 1 and has the property that for any mutually disjoint, countable family of sets $\{A_n\} \subset \mathcal{A}$,

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

We say that the space $(\Omega, \mathcal{A}, \mathbb{P})$ is a probability space. Further, we say that a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is complete if for any A in $\mathcal{A}$ such that $\mathbb{P}(A) = 0$ and for any $B \subset A$, we must have that B is also in $\mathcal{A}$. In this text, $(\Omega, \mathcal{A}, \mathbb{P})$, will denote a complete probability space. The reason for working with a complete probability space will become clear once we have defined the notion of measurability; see Remark 2.2.2.

Since the measure $\mathbb{P}$ measures sets in Ω, every time we talk about the measure $\mathbb{P}$, it is implied that it measures sets in Ω. We will write

$$\mathbb{P}(\mathrm{Cond}_1(\omega), \mathrm{Cond}_2(\omega), \ldots) := \mathbb{P}(\omega \in \Omega \mid \mathrm{Cond}_1(\omega), \mathrm{Cond}_2(\omega), \ldots),$$

where $\mathrm{Cond}_i(\omega)$ is some condition on ω. However, in some cases this notation might be somewhat ambiguous; in such cases, more detailed notation will be used. Furthermore, when a condition is satisfied by all points ω in a set A, where $\mathbb{P}(A)$ equals 1, we shall say that the condition is true almost surely, or *a.s.* for short. Conversely, we call sets of probability 0 negligible.

Remark 2.2.1. For the sake of compactness of the text, we sometimes will use non-trivial, but nevertheless standard statements regarding the probability measure $\mathbb{P}$ that may all be found in the book of Friedman [2, Chapter 1.2].

Definition 2.2.2. Given a probability (or more generally a measure) space $(\Omega, \mathcal{A}, \mathbb{P})$ and a function X from Ω into $\mathbb{R}^n$, we say that X is $\mathcal{A}$-measurable, if the preimage of any Borel set in $\mathbb{R}^n$ is in $\mathcal{A}$. Such measurable functions are also called random variables. Every random variable induces a probability measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, called the law of the random variable, defined via

$$\mu_X(B) = \mathbb{P}(X^{-1}(B)),$$

where B is any Borel set in $\mathbb{R}^n$. In this way, we have actually constructed a new probability space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mu_X)$. Further, we can also consider the j:th marginal law, μ_{X_j}, defined via

$$\mu_{X_j}(D) = \mathbb{P}(X_j^{-1}(D)),$$

where D is any borel set in $\mathbb{R}$ and X_j is the j:th component of X.

For every random variable, and thus for every induced probability measure on $\mathbb{R}^n$, we can consider the distribution function $F_X(\mathbf{x})$, defined as

$$F_X(\mathbf{x}) = \mu_X(C),$$

where $\mathbf{x} = (x_1, \ldots, x_n)$ and $C = (-\infty, x_1] \times \ldots \times (-\infty, x_n]$. Hence, the distribution function may be written as

$$F_X(\mathbf{x}) = \mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n).$$

Actually, as we shall see in the following lemma, our definition of measurability is equivalent to just considering sets of the same form as C.

Remark 2.2.2. Now, one will be able to understand the importance of having a complete probability space. Suppose we do not have a complete probability space, i.e there exists some set $B \subset \Omega$ such that $B \subset A$ and $\mathbb{P}(A) = 0$, where $A \in \mathcal{A}$. Further, suppose we have a random variable X from Ω into $\mathbb{R}^n$ and a function Y from Ω into $\mathbb{R}^n$ such that $Y(\omega) = X(\omega)$ for all $\omega \in \Omega \setminus B$. However, Y need not be a random variable on this space since the set $B \notin \mathcal{A}$. This is of course unwanted since they are equal almost surely.

Lemma 2.2.1. *A function X from Ω into $\mathbb{R}^n$ is a random variable if and only if for any set $C \subset \mathbb{R}^n$ of the form $C = (-\infty, x_1], \times \ldots \times (-\infty, x_n]$, the pre-image of C under X is in $\mathcal{A}$.*

Proof. First, assume X is a random variable. Since $C \in \mathcal{B}(\mathbb{R}^n)$, the statement follows. On the contrary, suppose for any C on the form above that the preimage of C under X is in $\mathcal{A}$. For any real numbers $a < b$

$$(a, b) = \bigcup_{n=1}^{\infty} ((a, b - \frac{1}{n}]) = \bigcup_{n=1}^{\infty} ((-\infty, b - \frac{1}{n}] \setminus (-\infty, a]), \tag{2.3}$$

and thus, any set of the form $A = (a, b)^n$ may be written as a Cartesian product of sets which the right hand side of (2.3) has. By Lemma 2.1.2, any set in $\mathcal{B}(\mathbb{R}^n)$ may be written as a countable union of open squares. Hence the preimage of any set in $\mathcal{B}(\mathbb{R}^n)$ under X is in $\mathcal{A}$. □

Remark 2.2.3. Some immediate consequences of the previous lemma are the following.

The above lemma can be stated the same way but with the sets C being of the form $C = (-\infty, x_1) \times \ldots \times (-\infty, x_n)$. This is just because we can write $(a, b) = (-\infty, b) \setminus (-\infty, a]$.

The Borel σ-algebra on $\mathbb{R}^n$ is also generated by the n-fold Cartesian products of sets of the form $(-\infty, c_j]$.

A function X from Ω into $\mathbb{R}^n$ is a random variable if and only if each function X_j for $j = 1, \ldots, n$ from Ω into $\mathbb{R}$ is a random variable.

If X is a random variable and if $k \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$, then $\lambda X + k$ and $\lambda X - k$ are both random variables. Indeed since if $\lambda \neq 0$ we may write the set $\{\lambda X_1 + k_1 < c_1, \ldots, \lambda X_n + k_n < c_n\} = \{X_1 < \frac{c_1 - k_1}{\lambda}, \ldots, X_n < \frac{c_n - k_n}{\lambda}\}$ and if $\lambda = 0$, the result is trivial.

Lemma 2.2.2. *If X, Y are random variables into $\mathbb{R}^n$, then the set*

$$K = \{X_1 < Y_1, \ldots, X_n < Y_n\}$$

is in $\mathcal{A}$.

Proof. Since the rationals are countable, there exists a sequence $\{r_k\}$ that is the sequence of all rational numbers. Since if $X_1(\omega) < Y_1(\omega)$ there exists a rational number r such that $X_1(\omega) < r < Y_1(\omega)$; we may thus write the set K above as

$$\bigcup_{k_1=1}^{\infty} \dots \bigcup_{k_n=1}^{\infty} (\{X_1 < r_{k_1}\} \cap \dots \cap \{X_n < r_{k_n}\} \cap \{Y_1 > r_{k_1}\} \cap \dots \cap \{Y_n > r_{k_n}\})$$

which is in $\mathcal{A}$ by the previous remark. □

Corollary 2.2.1. *If X, Y are random variables into $\mathbb{R}^n$ then $X + Y$ and $X - Y$ are random variables.*

Proof. By Lemma 2.2.2, if X and Y are random variables, the set $K = \{X_1 < Y_1, \dots, X_n < Y_n\}$ is in $\mathcal{A}$ and since we may rewrite sets of the form

$$\{X_1 + Y_1 < c_1, \dots, X_n + Y_n < c_n\} = \{X_1 < c_1 - Y_1, \dots, X_n < c_n - Y_n\}$$

and similarly for $X - Y$, the statement follows by Remark 2.2.3. □

Remark 2.2.4. In the case when the random variable X takes values in $\mathbb{R}^n$ for $n \geq 2$, we sometimes call it a random vector. Note that a random vector $X = (X_1, \dots, X_n)$ can be viewed as a list of length n, where each X_j is a one-dimensional random variables on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

Definition 2.2.3. For a probability (or more generally a measure) space $(\Omega, \mathcal{A}, \mathbb{P})$ and any positive number p, we define the space $L^p(\Omega, \mathcal{A}, \mathbb{P})$ to be the set of equivalence classes of real valued random variables X such that $|X|^p$ is integrable. The equivalence relation " $\sim$ " is defined via $X_1 \sim X_2$ if $X_1 = X_2$ a.s.. We define a norm on this space via

$$\|X\|_{L^p} = (\int_\Omega |X|^p d\mu)^{\frac{1}{p}},$$

for any $X \in L^p(\Omega, \mathcal{A}, \mathbb{P})$.

Remark 2.2.5. Usually, we omit the phrase "in some equivalence class of" when we say that a random variable X is in L^p.

Remark 2.2.6. As shown in Friedman, for $p \geq 1$, this is in fact a complete, metric space [2, Theorem 3.2.3]. Since metric spaces are normed spaces, we may say that for $p \geq 1$, the above is in fact a complete normed linear space.

Lemma 2.2.3. *The space $L^2(\Omega, \mathcal{A}, \mathbb{P})$ is a Hilbert space with the inner product defined as*

$$\langle f, g \rangle = \int_\Omega fg d\mathbb{P}.$$

Proof. properties i) to iv) of Definition 2.1.4 follow by the linearity of integration and the final property follows by the above remark. □

First time readers of formal probability theory may find the use of the word random to be confusing, since there is nothing random about a random variable; it is merely a function. The "randomness" occurs in the probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Thus, to help the reader acquire some intuition for the concept of probability spaces and random variables, we shall consider two examples.

Example 2.2.1. Suppose we want to measure the height of random citizens of the earth. Let Ω be the set of all people on the globe and let $\mathcal{A}$ be the power set of Ω. Define the probability measure $\mathbb{P}$ via $\mathbb{P}(\omega) = \frac{1}{|\Omega|}$, for every $\omega \in \Omega$, i.e every person on earth has equal likelihood to be chosen in this experiment. Now, define the random variable X from Ω into $\mathbb{R}$ which maps every person ω to their height. As we can see, there is nothing random about this random variable; its "randomness" comes from the uncertainty of which ω is "picked".

Furthermore, we could construct many more random variables on this probability space that could for example map each person ω to their favourite colour, number of kids, or whatever else you may want to measure.

An example that we shall revisit in a later section is the following

Example 2.2.2. Consider a particle that starts at the point 0 and at each second changes position to a neighbouring integer; i.e after the first second it will go to either 1 or -1 and so on. The particle will do this for n seconds and then stop. A probability space may be constructed as follows. Let Ω be the set of all sequences of points $\omega_0, \omega_1, \ldots, \omega_n$ such that $\omega_0 = 0$ and for any $i \in \{1, \ldots, n\}$, $|\omega_i - \omega_{i-1}| = 1$. Let the σ-algebra $\mathcal{A}$ be the power set of Ω and let the probability measure $\mathbb{P}$ be defined as $\mathbb{P}(\omega) = \frac{1}{|\Omega|} = \frac{1}{2^n}$, i.e it maps each $n+1$ sequence of points to equal probability.

As before, we may wish to construct random variables on this probability space. Let X_i be the mapping that maps the i:th element ω_i of the sequence $\omega \in \Omega$ to $|\omega_i|$. In other words, it measures the particles distance from the origin after i seconds. Note that this leads to a sequence of random variables $\{X_i\}_{i \in \{0,1,\ldots,n\}}$, on $(\Omega, \mathcal{A}, \mathbb{P})$, something we shall explore further in a later chapter.

Similar to how we can generate a σ-algebra from a family of subsets of Ω, we can generate a σ-algebra from a family random variables:

Definition 2.2.4. For an arbitrary family of random variables $\{X_t\}$, we define $\sigma\{X_t\}$ to be the σ-algebra on Ω with the property that for any σ-algebra $\mathcal{F}$ such that every X_t is measurable, $\sigma\{X_t\}$ is contained in $\mathcal{F}$. We call this σ-algebra the σ-algebra generated by $\{X_t\}$.

Remark 2.2.7. It is of course equivalent to say that $\sigma\{X_n\}$ is the intersection over all $\mathcal{F}$ defined as above.

Lemma 2.2.4. *For a random variable X from Ω into $\mathbb{R}^n$ and $A \in \sigma\{X\}$, there exists a set $B \in \mathcal{B}(\mathbb{R}^n)$ such that $A = X^{-1}(B)$.*

Proof. Suppose for a contradiction that for some set $A \in \sigma\{X\}$, we have the property that for every $B \in \mathcal{B}(\mathbb{R}^n)$, $A \neq X^{-1}(B)$. Then for A^c the same has to be true since if $A^c = X^{-1}(C)$, for some $C \in \mathcal{B}(\mathbb{R}^n)$, then $A = X^{-1}(C^c)$. Further, for any sets $A_1, A_2, \ldots$ such that $A = \cup_{n=1}^{\infty} A_n$, there cannot exist sets $D_1, D_2, \ldots$ such that $X^{-1}(D_i) = A_i$ for every i, since then, we could write $A = X^{-1}(\cup_{n=1}^{\infty} D_i)$.

By these observations, we wish to construct a new family as follows. Take $\sigma\{X\}$ and first remove the set A and A^c. Then, consider all sequences of sets $A_1, A_2, \ldots$, such that $A = \cup_{n=1}^{\infty} A_n$ or $A^c = \cup_{n=1}^{\infty} A_n$. For each such sequence, consider the sets A_i, which have the property that for all $B \in \mathcal{B}(\mathbb{R}^n)$, $A_i \neq X^{-1}(B)$. The existence of such a set is guaranteed by the above. Now, remove all such sets A_i together with their respective complements from $\sigma\{X\}$. Let this new family of subsets we have constructed be denoted by $\mathcal{E}$. This family is in fact a σ-algebra for which X is measurable. Indeed, the empty set is in $\mathcal{E}$, and for any set we removed from $\sigma\{X\}$ we also removed its complement. The last property of the σ-algebra holds since for any set A we removed, we also removed at least one set from every possible union of A. The measurability property comes from the fact that for each set A we removed from $\sigma\{X\}$, it was a set which had the property that for all $B \in \mathcal{B}(\mathbb{R}^n)$, $A \neq X^{-1}(B)$. We have now established a contradiction since we have constructed a σ-algebra $\mathcal{E}$ for which X is measurable but $\mathcal{E}$ does not contain $\sigma\{X\}$. ☐

Definition 2.2.5 (Independence)**.** For a family $\{X_i\}$ of k random variables, we say that they are independent if for every choice of $B_1 \in \mathcal{B}(\mathbf{R}^n), \ldots, B_k \in \mathcal{B}(\mathbf{R}^n)$,

$$\mathbb{P}(X_1^{-1}(B_1) \cap \ldots \cap X_k^{-1}(B_k)) = \mathbb{P}(X_1^{-1}(B_1)) \cdot \ldots \cdot \mathbb{P}(X_k^{-1}(B_k)).$$

Similarly, for a collection $\{\mathcal{A}_i\}$ of k families of sets such that $\mathcal{A}_i \subset \mathcal{A}$ for every i, we say that they are independent if for every choice of $A_1 \in \mathcal{A}_1, \ldots A_k \in \mathcal{A}_k$,

$$\mathbb{P}(A_1 \cap \ldots \cap A_k) = \mathbb{P}(A_1) \cdot \ldots \cdot \mathbb{P}(A_k).$$

Further, when any of these families are infinite, we shall instead define them being independent if any finite collection of them are independent. Lastly, we say that a random variable X is independent of the σ-algebra $\mathcal{U}$ if $\sigma\{X\}$ and $\mathcal{U}$ are independent.

Lemma 2.2.5. *If two events $A_1, A_2 \subset \mathcal{A}$ are such that*

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2),$$

then

$$\mathbb{P}(A_1 \cap A_2^c) = \mathbb{P}(A_1)\mathbb{P}(A_2^c).$$

Proof. Suppose $A_1, A_2 \subset \mathcal{A}$ are such that $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2)$. Since $A_1 \cap A_2$ and $A_1 \cap A_2^c$ are disjoint, it follows that

$$\begin{aligned}\mathbb{P}(A_1 \cap A_2^c) &= \mathbb{P}(A_1) - \mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)(1 - \mathbb{P}(A_2)) \\ &= \mathbb{P}(A_1)\mathbb{P}(A_2^c),\end{aligned}$$

which proves the statement. □

Corollary 2.2.2. *Any finite family $\{X_i\}$ of k random variables are independent if and only if the σ-algebras $\{\sigma(X_i)\}$ are independent.*

Proof. This follows directly from Lemma 2.2.4. □

Definition 2.2.6 (Expected value)**.** for any random variable $X = (X_1, \ldots, X_n)$ such that X_j is in the space $L^1(\Omega, \mathcal{A}, \mathbb{P})$, then we define the expected value of X_j with respect to $\mathbb{P}$ as

$$E[X_j] := \int_\Omega X_j(\omega) d\mathbb{P}(\omega), \tag{2.4}$$

and if the random variable X_j is not in $L^1(\Omega, \mathcal{A}, \mathbb{P})$, we say that X_j has infinite absolute expectation. Further, we define the expected value of X as

$$E[X] = (E[X_1], \ldots, E[X_n]).$$

In this text, the notation $\mathbf{1}_A$ will be used to denote the characteristic function, defined as

$$\mathbf{1}_A(x) = \begin{cases} 1 \text{ if } x \in A \\ 0 \text{ otherwise.} \end{cases}$$

Remark 2.2.8. Similarly to what is mentioned in Remark 2.2.1, we wish to also use non-trivial, yet standard statements regarding the Lebesgue integral. These may also be found in the book of Friedman [2, Chapter 2.5-2.6].

Theorem 2.2.1. *Any two independent random variables X, Y with finite absolute expectation satisfies the property that*

$$E[XY] = E[X]E[Y].$$

Proof. First, we intend to show that XY is integrable. This will be done using the definition in Friedman, which states that a real-valued, measurable function f is integrable if there exists a sequence of simple functions f_n that converges point-wise almost everywhere to f and is Cauchy in the mean [2, Definition 2.6.1]. Recall that a sequence $\{f_n\}$ is Cauchy in the mean if $\int_\Omega |f_n - f_m| d\mathbb{P} \to 0$ as $n, m \to \infty$.

By this definition, since X and Y are assumed integrable, there exists sequences of simple functions X_n and Y_n that converges point-wise almost everywhere to X and Y respectively and are Cauchy in the mean. X_n and Y_n may be written as

$$\begin{aligned}X_n &= \sum_{j=0}^{n} x_j \mathbf{1}_{E_j} \\ Y_n &= \sum_{j=0}^{n} y_j \mathbf{1}_{F_j},\end{aligned}$$

where E_j and F_j are sets in $\mathcal{A}$. Since we can write the functions X and Y of the form $X = X^+ - X^-$, where both X^+ and X^- are positive, it follows by the triangle inequality that we can pick X_n and Y_n to converge absolutely. Hence, the Cauchy product of $\lim_n X_n$ and $\lim_n Y_n$ converge to XY, i.e XY may be written as

$$\begin{aligned}XY &= \lim_n \sum_{j=0}^{n} x_j \mathbf{1}_{E_j} \lim_n \sum_{j=0}^{n} y_j \mathbf{1}_{F_j} \\ &= \lim_n \sum_{j=0}^{n} \sum_{k=0}^{j} x_k y_{j-k} \mathbf{1}_{E_k \cap F_{j-k}}.\end{aligned}$$

Now we wish to show that X_nY_n is Cauchy in the mean; without loss of generality, we may assume that $n > m$, so that we can compute

$$\begin{aligned}
E[(X_nY_n - X_mY_m)] &= E[(X_n - X_m)Y_n + (Y_n - Y_m)X_m] \\
&= E[(X_n - X_m)Y_n] + E[(Y_n - Y_m)X_m] \\
&= E[\sum_{j=m+1}^{n}\sum_{k=0}^{n} x_j y_k \mathbf{1}_{E_j \cap F_k}] + E[\sum_{j=m+1}^{n}\sum_{k=0}^{n} y_j x_k \mathbf{1}_{E_k \cap F_j}] \\
&= \sum_{j=m+1}^{n}\sum_{k=0}^{n} x_j y_k \mathbb{P}(E_j \cap F_k) + \sum_{j=m+1}^{n}\sum_{k=0}^{n} y_j x_k \mathbb{P}(E_k \cap F_j) \\
&= \sum_{j=m+1}^{n}\sum_{k=0}^{n} x_j y_k \mathbb{P}(E_j)\mathbb{P}(F_k) + \sum_{j=m+1}^{n}\sum_{k=0}^{n} y_j x_k \mathbb{P}(E_k)\mathbb{P}(F_j) \\
&= E[X_n - X_m]E[Y_n] + E[Y_n - Y_m]E[X_m],
\end{aligned}$$

which tends to 0 since X_n and Y_n are Cauchy in the mean; hence, XY is integrable. By the definition mentioned above, we may therefore write

$$\begin{aligned}
E[XY] &= E[\lim_n \sum_{j=0}^{n}\sum_{k=0}^{j} x_k y_{j-k} \mathbf{1}_{E_k \cap F_{j-k}}] \\
&= \lim_n \sum_{j=0}^{n}\sum_{k=0}^{j} x_k y_{j-k} \mathbb{P}(E_k \cap F_{j-k}) \\
&= \lim_n \sum_{j=0}^{n}\sum_{k=0}^{j} x_k y_{j-k} \mathbb{P}(E_k)\mathbb{P}(F_{j-k}) \\
&= E[X]E[Y],
\end{aligned}$$

which completes the proof. □

Definition 2.2.7. Let X be a random variable from Ω into $\mathbb{R}$ such that it is in the space $L^2(\Omega, \mathcal{A}, \mathbb{P})$. Then we define the variance of X as

$$\text{Var}(X) = E[(X - E[X])^2]. \tag{2.5}$$

Remark 2.2.9. By linearity of the integral, we can rewrite the above as

$$\begin{aligned}
E[(X - E[X])^2] &= E[X^2 - 2XE[X] + E[X]^2] \\
&= E[X^2] - E[2XE[X]] + E[E[X]^2] \\
&= E[X^2] - 2E[X]^2 + E[X]^2 = E[X^2] - E[X]^2.
\end{aligned}$$

This is a very convenient rewriting for computational purposes. Also note that for any $a \in \mathbb{R}$

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

Corollary 2.2.3. *For any two independent random variables with $X, Y \in L^2(\Omega, \mathbb{P})$, we have that*

$$Var(X + Y) = Var(X) + Var(Y).$$

Proof. By the triangle inequality, the sum of any L^2-functions is an L^2-function, so we can consider

$$\begin{aligned}
\text{Var}(X + Y) = E[(X + Y - E[X + Y])^2] &= E[(X - E[X] + Y - E[Y])^2] \\
&= E[(X - E[X])^2 + 2(X - E[X])(Y - E[Y]) + (Y - E[Y])^2] \\
&= \text{Var}(X) + \text{Var}(Y) + 2E[(X - E[X])(Y - E[Y])],
\end{aligned}$$

but by Theorem 2.2.1

$$\begin{aligned} E[(X - E[X])(Y - E[Y])] &= E[XY - XE[Y] - YE[X] + E[X]E[Y]] \\ &= E[XY] - E[X]E[Y] - E[Y]E[X] + E[X]E[Y] \\ &= 0, \end{aligned}$$

so the statement follows. □

Definition 2.2.8. A random variable X from Ω into $\mathbb{R}$ is called Gaussian (or normal) with expected value μ and variance σ if its induced distribution function is given by

$$F_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt,$$

and we write $X \sim N(\mu, \sigma)$ to denote that X is Gaussian.

Remark 2.2.10. As many probably know from elementary probability theory, the integrand (along with the constant) above is known as the density of the Gaussian distribution. This is in fact an application of the Radon-Nikodym theorem, which we shall not delve into in this text. Nevertheless, it is important to know that not all random variables have a density function, but the Gaussian distribution has and it is given by

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}.$$

One reason why it is useful is that we can compute expected values with it. This can be stated as, for a Gaussian random variable X and a measurable function g, we have that

$$E[g(X)] = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} g(x) e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx.$$

A useful fact about the Gaussian is the following lemma

Lemma 2.2.6.

$$\int_{\mathbb{R}} e^{-x^2} dx = \sqrt{\pi}.$$

Proof. First, consider

$$\begin{aligned} \int_{-\infty}^{\infty} |e^{-x^2}| dx &= \int_{-\infty}^{-1} e^{-x^2} dx + \int_{-1}^{1} e^{-x^2} dx + \int_{1}^{\infty} e^{-x^2} dx \\ &\leq \int_{-\infty}^{-1} -xe^{-x^2} dx + \int_{-1}^{1} e^{-x^2} dx + \int_{1}^{\infty} xe^{-x^2} dx \\ &\leq \frac{1}{2e} + 2e + \frac{1}{2e} < \infty, \end{aligned}$$

So the integral is finite, which means that

$$(\int_{-\infty}^{\infty} e^{-x^2} dx)^2$$

is also finite. Hence by applying Fubini's Theorem, we can compute as

$$\begin{aligned} (\int_{-\infty}^{\infty} e^{-x^2} dx)^2 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(x^2+y^2)} dx dy = \int_{0}^{2\pi} \int_{0}^{\infty} e^{-r^2} r dr d\theta \\ &= \int_{0}^{2\pi} [-\frac{e^{-r^2}}{2}]_0^\infty d\theta = \int_{0}^{2\pi} \frac{1}{2} d\theta = \pi \end{aligned}$$

and the statement follows. □

Theorem 2.2.2 (Markov's Inequality). *For all positive numbers δ and β and for any random variable X*

$$\mathbb{P}(|X| \geq \delta) \leq \frac{E[|X|^\beta]}{\delta^\beta}.$$

Proof. Since $\frac{|X|}{\delta} < 1$ if and only if $\mathbf{1}_{\{|X| \geq \delta\}} = 0$, it follows that $\frac{|X|}{\delta} \geq \mathbf{1}_{\{|X| \geq \delta\}}$. From this, we deduce that $(\frac{|X|}{\delta})^\beta \geq \mathbf{1}_{\{|X| \geq \delta\}}$, since either the left hand side is greater than 1 or the right hand side is equal to 0. Hence

$$E[\mathbf{1}_{\{|X| \geq \delta\}}] = \mathbb{P}(|X| \geq \delta) \leq \frac{E[|X|]^\beta}{\delta^\beta},$$

which completes the proof. □

2.3 Stochastic Processes

The proofs and the example of this section are my own creation, expect for the Second Borel-Cantelli lemma, for which I got the tip to use the logarithm by a fellow student.

Below, we will use the notation that T is some subset of $\mathbb{R}_+$.

Definition 2.3.1. Given a set Ω and a σ-algebra $\mathcal{A}$ on Ω, we say that a family $\{\mathcal{A}_t\}_{t \in T}$, of σ-algebras such that for all $0 \leq i < j$, $A_i \subset A_j \subset \mathcal{A}$, is a filtration of $(\Omega, \mathcal{A})$.

Definition 2.3.2. Given an indexed family $\{X_t\}_{t \in T}$ of random variables and a filtration $\{\mathcal{A}_t\}_{t \in T}$ of $(\Omega, \mathcal{A})$ we say that $\{X_t\}_{t \in T}$ is $\{\mathcal{A}_t\}_{t \in I}$-adapted if for every t in T, X_t is $\mathcal{A}_t$-measurable.

Lemma 2.3.1. *Given an indexed family of random variables $\{X_t\}_{t \in T}$ and a filtration $\{\mathcal{A}_t\}_{t \in T}$, for which $\{X_t\}_{t \in T}$ is adapted to, then for any $s, u, v \in T$ such that $v \leq u \leq s$, we have that both $\sigma\{X_u - X_v\} \subset \mathcal{A}_s$ and $\sigma\{X_u + X_v\} \subset \mathcal{A}_s$.*

Proof. By the definition of a filtration, both X_u and X_v is $\mathcal{A}_s$-measurable. Then, by Corollary 2.2.1, both $X_u + X_v$ and $X_u - X_v$ is $\mathcal{A}_s$-measurable, and so by the definition of the generated sigma algebra, it follows that $\sigma\{X_u - X_v\} \subset \mathcal{A}_s$ and $\sigma\{X_u + X_v\} \subset \mathcal{A}_s$. □

Definition 2.3.3. Let $(\Omega, \mathcal{A}, \mathbb{P})$, be a probability space. If $\{\mathcal{A}_t\}_{t \in T}$ is a filtration of $(\Omega, \mathcal{A})$ and a family of random variables $\{X_t\}_{t \in T}$ are $\{\mathcal{A}_t\}_{t \in T}$-adapted, we say that the quintuple $(\Omega, \mathcal{A}, \{\mathcal{A}_t\}_{t \in T}, \{X_t\}_{t \in T}, \mathbb{P})$ is a stochastic process.

To help the reader better understand the concept of a filtration we shall give a rather simple extension of Example 2.2.2.

Example 2.3.1. Consider the same probability space and notation as given in Example 2.2.2. Let $T = \{0, 1, \ldots, n\}$ and define a sequence of random variables $\{X_t\}_{t \in T}$ via $X_t(\omega) = |(\omega_t)|$. We shall now construct a filtration $\{\mathcal{A}_t\}_{t \in T}$ such that $\{X_t\}_{t \in T}$ is adapted to that filtration.

Let $A_0 = \{\emptyset, \Omega\}$. This is a σ-algebra and since $X_0(\omega) = 0$ for every $\omega \in \Omega$, X_0 is $\mathcal{A}_0$-measurable. Furthermore, since $X_1(\omega) = 1$ for every $\omega \in \Omega$, we shall again let $\mathcal{A}_1 = \{\emptyset, \Omega\}$, which is by the same reason as before a σ-algebra such that X_1 is $\mathcal{A}_1$-measurable. However, since X_2 equals either 0 or 2, we can no longer consider the trivial σ-algebra. Let $\mathcal{U} \subset \Omega$ be the set of all sequences $\omega = \omega_0, \omega_1, \ldots, \omega_n$, with ω_2 equal to 2 or -2. Thus, we shall let $\mathcal{A}_2 = \{\emptyset, \Omega, \mathcal{U}, \mathcal{U}^c\}$. Because $\mathcal{U}^c$ will be the set of sequences ω, with $\omega_2 = 0$. this is in fact a σ-algebra such that X_2 is $\mathcal{A}_2$ adapted since for any set $B \in \mathcal{B}(\mathbb{R})$

$$X_2^{-1}(B) = \begin{cases} \emptyset & \text{if } B \cap (\{0\} \cup \{2\}) = \emptyset \\ \Omega & \text{if } B \cap (\{0\} \cup \{2\}) = \{0\} \cup \{2\} \\ \mathcal{U} & \text{if } B \cap (\{0\} \cup \{2\}) = \{2\} \\ \mathcal{U}^c & \text{if } B \cap (\{0\} \cup \{2\}) = \{0\}. \end{cases}$$

Further, since X_3 may take the values 1 or 3, we may let $\mathcal{V}$ be the set of all sequences $\omega \in \Omega$ such that ω_3 equals 3 or -3. Note that $\mathcal{V}^c$ is the set of sequences $\omega \in \Omega$ with ω_3 equal to 1 or -1. By the definition of a filtration, we want $\mathcal{A}_2 \subset \mathcal{A}_3$, so we need to let $\mathcal{A}_3$ equal $\sigma(\mathcal{A}_2 \cup \mathcal{V})$. Continuing in this fashion, for $\mathcal{A}_4, \ldots, \mathcal{A}_n$, one may construct the whole filtration $\{\mathcal{A}_t\}_{t\in T}$, making $(\Omega, \mathcal{A}, \{\mathcal{A}_t\}_{t\in T}, \{X_t\}_{t\in T}, \mathbb{P})$ into a stochastic process.

As one might have recognized, constructing a filtration in this explicit manner is rather cumbersome. Therefore, we shall make the following definition.

Definition 2.3.4. Given an arbitrary family of random variables $\{X_t\}_{t\in T}$, on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, we can always consider a filtration on $(\Omega, \mathcal{A})$ with respect to $\{X_t\}_{t\in T}$, defined as

$$\mathcal{A}_t = \sigma\{X_s, s \leq t\}.$$

This is called the natural filtration and it is easily seen that this is the coarsest filtration such that $\{X_t\}_{t\in T}$ is $\{A_t\}_{t\in T}$-adapted.

Remark 2.3.1. As mentioned before, the filtration constructed in Example 2.3.1 is the natural filtration, but it is certainly not the only filtration that could have been constructed. Consider for example the trivial filtration $\mathcal{A}_t = \mathcal{A}$ for every t. Nevertheless, the natural filtration shall be the most common filtration used in this text.

Lemma 2.3.2 (First Borel-Cantelli Lemma)**.** *For any sequence of sets $(\mathcal{F}_n)_{n\in\mathbb{N}} \subset \mathcal{A}$, such that*

$$\sum_{n=1}^{\infty} \mathbb{P}(\mathcal{F}_n) < \infty$$

then it holds that $\mathbb{P}(\overline{\lim}_n \mathcal{F}_n) = 0$.

Proof. Since $\overline{\lim}_n \mathcal{F}_n = \bigcap_{k=0}^{\infty} \bigcup_{j=k}^{\infty} \mathcal{F}_j$, it follows that

$$\mathbb{P}(\overline{\lim_n} \mathcal{F}_n) \leq \mathbb{P}(\bigcup_{j=k}^{\infty} \mathcal{F}_j) \leq \sum_{j=k}^{\infty} \mathbb{P}(\mathcal{F}_j),$$

for every natural number k. Hence, for every $\varepsilon > 0$, there exists a k_ε such that

$$\mathbb{P}(\overline{\lim_n} \mathcal{F}_n) \leq \sum_{j=k_\varepsilon}^{\infty} \mathbb{P}(\mathcal{F}_j) < \varepsilon,$$

which proves the statement. □

Lemma 2.3.3. *For $x \in [0, 1]$*

$$-\log(1 - x) \geq x$$

Proof. Let $x \in [0, 1)$ and Consider the power series expansion of $-\log(1 - x)$

$$-\log(1 - x) = -\sum_{n=1}^{\infty} \frac{(-1)^{n-1}(-x)^n}{n} = \sum_{n=1}^{\infty} \frac{x^n}{n} = x + \frac{x^2}{2} + \frac{x^3}{3} + \ldots \geq x,$$

since all the terms in the series are positive. Lastly, $-\log(1 - 1) = \infty > 1$ so the statement follows. □

Lemma 2.3.4 (Second Borel-Cantelli Lemma)**.** *For any sequence of independent sets $(\mathcal{F}_n)_{n\in\mathbb{N}} \subset \mathcal{A}$ such that*

$$\sum_{n=1}^{\infty} \mathbb{P}(\mathcal{F}_n) = \infty,$$

it is also true that $\mathbb{P}(\overline{\lim}_n \mathcal{F}_n) = 1$

Proof. Suppose $(\mathcal{F}_n)_{n\in\mathbb{N}} \subset \mathcal{A}$ is a sequence of independent sets such that $\sum_{n=1}^{\infty} \mathbb{P}(\mathcal{F}_n) = \infty$. We want to show that

$$\mathbb{P}(\overline{\lim_n} \mathcal{F}_n) = 1 \iff \mathbb{P}((\overline{\lim_n} \mathcal{F}_n)^c) = 0.$$

Since if $x \in \cap_{n=N}^{\infty} F_n^c$, then $x \in \cap_{n=N+1}^{\infty} F_n^c$ for all N, it follows that

$$\begin{aligned}\mathbb{P}((\overline{\lim_n} \mathcal{F}_n)^c) &= \mathbb{P}((\cap_{N=1}^{\infty} \cup_{n=N}^{\infty} F_n)^c) = \mathbb{P}((\cup_{N=1}^{\infty} \cap_{n=N}^{\infty} F_n^c)) \\ &= \mathbb{P}(\lim_N \cap_{n=N}^{\infty} F_n^c) = \lim_N \mathbb{P}(\cap_{n=N}^{\infty} F_n^c),\end{aligned}$$

as the sequence $\cap_{n=N}^{\infty} F_n^c$ is monotone increasing, the last equality is an application of what is written in Remark 2.2.1. Note that since the range of any probability measure is a subset of $[0, 1]$,

$$\begin{aligned}\lim_N \mathbb{P}(\cap_{n=N}^{\infty} F_n^c) = 0 &\iff \lim_N \log(\mathbb{P}(\cap_{n=N}^{\infty} F_n^c)) = -\infty \\ &\iff \lim_N - \log(\mathbb{P}(\cap_{n=N}^{\infty} F_n^c)) = \infty.\end{aligned}$$

Now, since $\lim_N - \log(\mathbb{P}(\cap_{n=N}^{\infty} F_n^c)) = \lim_N - \log(\mathbb{P}(\lim_K \cap_{n=N}^{K} F_n^c))$ and $\cap_{n=N}^{K} F_n^c$ is monotonically decreasing with K we have again by Remark 2.2.1, that

$$\begin{aligned}\lim_N - \log(\mathbb{P}(\cap_{n=N}^{\infty} F_n^c)) &= \lim_N \lim_K - \log(\mathbb{P}(\cap_{n=N}^{K} F_n^c)) \\ &= \lim_N \lim_K - \log(\mathbb{P}(F_N^c) \cdot \ldots \cdot \mathbb{P}(F_K^c)) \\ &= \lim_N \lim_K \sum_{n=N}^{K} - \log(\mathbb{P}(F_n^c)) \\ &= \lim_N \lim_K \sum_{n=N}^{K} - \log(1 - \mathbb{P}(F_n)),\end{aligned}$$

as by Lemma 2.2.5, the sets F_n^c are independent. Lastly, by Lemma 2.3.3, we have that

$$\lim_N \lim_K \sum_{n=N}^{K} - \log(1 - \mathbb{P}(F_n)) \geq \lim_N \lim_K \sum_{n=N}^{K} \mathbb{P}(F_n) = \infty$$

so the statement follows. □

Chapter 3

Kolmogorov's Continuity Theorem

In this chapter, $|\cdot|$ will denote the euclidean norm in $\mathbb{R}^n$. Which n it refers to will be implicit from the argument.

3.1 Hölder Continuity

All the proofs in this section are my own.

Definition 3.1.1. For $\gamma > 0$, we say that a function f from $D \subset \mathbb{R}^n$ into $\mathbb{R}^m$ is locally γ-Hölder continuous if there exists $\varepsilon > 0$ such that

$$\sup_{0<|x-y|<\varepsilon} \frac{|f(x)-f(y)|}{|x-y|^\gamma} < \infty.$$

Furthermore, we say that f is γ-hölder continuous on a subset $A \subset D$ if

$$\sup_{\substack{x,y\in A,\\ x\neq y}} \frac{|f(x)-f(y)|}{|x-y|^\gamma} < \infty.$$

Lemma 3.1.1. *If a function f from $D \subset \mathbb{R}^n$ into $\mathbb{R}^m$ is locally γ-Hölder contiuous, then f is continuous at each point $x \in D$.*

Proof. Suppose f is locally γ-Hölder continuous. This means that there exists $\xi > 0$ such that

$$\sup_{0<|x-y|<\xi} \frac{|f(x)-f(y)|}{|x-y|^\gamma} \leq C_\xi,$$

For some C_ξ. Fix $\varepsilon > 0$ and a point $x \in D$. First, for all points y such that $|x-y| < \xi$, we have that

$$|f(x)-f(y)| \leq C_\xi |x-y|^\gamma < C_\xi \xi^\gamma.$$

Now if, $C_\xi \xi^\gamma < \varepsilon$, we are done. Otherwise, we have that

$$C_\xi \xi^\gamma \geq \varepsilon \Longleftrightarrow \xi \geq (\frac{\varepsilon}{C_\xi})^{\frac{1}{\gamma}},$$

so by picking $\delta = (\frac{\varepsilon}{C_\xi})^{\frac{1}{\gamma}}$, we have that

$$|x-y| < \delta \Longrightarrow |f(x)-f(y)| \leq C_\xi |x-y|^\gamma < \varepsilon,$$

which proves the statement. □

Lemma 3.1.2. *If a function f from $D \subset \mathbb{R}^n$ into $\mathbb{R}^m$ is locally γ-Hölder continuous, then f is γ-Hölder continuous on each compact subset of D.*

Proof. Suppose f is locally γ-Hölder continuous and let ε be given as in Definition 3.1.1. Fix a compact subset $K \subset D$ and consider any two points $x, y \in K$ with $0 < |x-y| < L$. If $L \leq \varepsilon$, the statement follows and if $L > \varepsilon$, we have that since continuous functions are bounded on compact sets, we trivially have that

$$\sup_{|x-y|>L} \frac{|f(x)-f(y)|}{|x-y|^\gamma} < \infty$$

since the numerator and the denominator is both bounded. □

3.2 Kolmogorov's Continuity Theorem

The proofs of this section was found in Baldi [1, Section 2.2]. Nevertheless, there were some mistakes and skipped steps, for which I have filled in the details.

Theorem 3.2.1 (Kolmogorov's Continuity Theorem)**.** *Let $D \subset \boldsymbol{R}^m$ be an open set and let $\{X_y\}_{y \in D}$ be a family of $d-$ dimensional random variables on $(\Omega, \mathcal{A}, I\!P)$ such that there exist positive numbers α, β, c satisfying*

$$E[|X_y - X_z|^\beta] \leq c|y-z|^{m+\alpha}, \quad y, z \in D. \tag{3.1}$$

Then there exists a family $\{X'_y\}_{y \in D}$ of d-dimensional random variables such that $X_y = X'_y$ $I\!P$-a.s for every $y \in D$. Furthermore, for every $\omega \in \Omega$, the map $y \mapsto X'_y(\omega)$ is γ-Hölder continuous on every compact subset of D for every $\gamma < \frac{\alpha}{\beta}$.

Remark 3.2.1. The map $y \mapsto X'_y(\omega)$ is also called the sample paths of X or in the context of a stochastic process, we refer to it as the sample paths of the stochastic process.

To prove this theorem, we shall first need the following lemma

Lemma 3.2.1. *Let the assumptions of Kolmogorov's continuity theorem be given and let D_B be the set of dyadic points in D. Then, for every $\gamma < \frac{\alpha}{\beta}$, there exists a negligible set N such that for all $\omega \in N^c$, the restriction of the map $y \mapsto X_y(\omega)$ to D_B is locally γ-Hölder continuous.*

Remark 3.2.2. Dyadic points are points where the coordinates are give by rational numbers with multiples of 2:s in the denominator. Later on in the text, we shall use the notion of dyadic *intervals*, which are similar but not to be confused with. Furthermore, the set of dyadic points in the reals are dense in the reals since every real number has a binary expansion.

Proof. By Lemma 2.1.2, the open set D may be written as a countable union of open squares, so it is sufficient to prove the statement for open squares. Furthermore, by scaling and translation, it is actually enough to prove the lemma for the open unit square. Thus, let $D = (0,1)^m$ and for a fixed natural number n, let $A_n \subset D_B$ be the set of all points in D_B of the form $k2^{-n}$ where k is a natural number. Let $\gamma < \frac{\alpha}{\beta}$ and define

$$\Gamma_n = \{\omega \in \Omega \mid \exists y, z \in A_n, \text{ s.t } |y-z| = 2^{-n} \text{ and } |X_y(\omega) - X_z(\omega)| > 2^{-n\gamma}\}.$$

Now, for fixed $y, z \in A_n$ such that $|y-z| = 2^{-n}$ we have, by Markov's inequality, and by assumption (3.1) in the Continuity Theorem, that

$$\begin{aligned} I\!P(\{\omega \in \Omega \mid |X_y - X_z| > 2^{-n\gamma}\}) &\leq 2^{n\beta\gamma} E[|X_y - X_z|^\beta] \\ &\leq 2^{n\beta\gamma} c|y-z|^{m+\alpha} \\ &= c2^{n(\beta\gamma - m - \alpha)}. \end{aligned}$$

The set of pairs $y, z \in A_n$ such that $|y-z| = 2^{-n}$ has less than $2m2^{nm}$ elements because for each point $y \in A_n$ there is at most $2m$ number of points z such that $|y-z| = 2^{-n}$ and since the set A_n has $(2^n-1)^m$ number of elements we obtain the desired bound for the number of elements. Hence we get the inequality

$$I\!P(\Gamma_n) \leq c2m2^{nm+n(\beta\gamma-\alpha-m)} \leq d2^{-n\mu},$$

where $d = c2m$ and $\mu = \alpha - \beta\gamma$, which is positive by the assumption on γ. Hence

$$\sum_{n=1}^{\infty} \mathbb{P}(\Gamma_n) \leq d \sum_{n=1}^{\infty} 2^{-n\mu} < \infty$$

and so by the first Borel-Cantelli lemma, $\mathbb{P}(\overline{\lim}_n \Gamma_n) = 0$. Let $N = \overline{\lim}_n \Gamma_n$ and note that for each $\omega \in N^c$, there exists an n such that $\omega \in \Gamma_k^c$ for every $k > n$ since otherwise ω would be in N, by Lemma 2.1.3. Now, fix $\omega \in N^c$ and pick such an n as above. First, assume that $m = 1$ and let $y \in D_B$, so for $v > n$ and $y \in [i2^{-v}, (i+1)2^{-v})$, there exists an $r \geq v + 1$ such that

$$y = i2^{-v} + \sum_{j=v+1}^{r} a_j 2^{-j}$$

where the coefficients a_j are either 0 or 1. Now, for $k \in \{v, v+1, \dots, r\}$ consider the sequence

$$y_k = \begin{cases} i2^{-v} + \sum_{j=v+1}^{k} a_j 2^{-j} & \text{if } k \geq v+1 \\ i2^{-v} & \text{if } k = v. \end{cases}$$

Note that $y_r = y$, so that we may write

$$|X_y - X_{i2^{-v}}| = |\sum_{k=v}^{r-1} X_{y_{k+1}} - X_{y_k}|,$$

which by the triangle inequality yields

$$|X_y - X_{i2^{-v}}| \leq \sum_{k=v}^{r-1} |X_{y_{k+1}} - X_{y_k}|.$$

Since $|y_{k+1} - y_k| = a_{k+1} 2^{-(k+1)}$ but if $a_{k+1} = 0$, then $y_{k+1} = y_k$ so there is nothing to prove; suppose therefore that $a_{k+1} = 1$. Since $w \in \Gamma_{k+1}^c$ and $|y_{k+1} - y_k| = 2^{-(k+1)}$, by the definition of Γ_n, it has to be the case that

$$|X_y - X_{i2^{-v}}| \leq \sum_{k=v}^{r-1} 2^{-(k+1)\gamma}.$$

Put $l = k - v$ to obtain that the right hand side above equals

$$\begin{aligned} \sum_{l=0}^{r-v-1} 2^{-(l+v+1)\gamma} \leq 2^{-v\gamma} \sum_{l=0}^{r-v-1} 2^{-l\gamma} = 2^{-v\gamma} \frac{1 - 2^{-\gamma(r-v)}}{1 - 2^{-\gamma}} \\ \leq 2^{-v\gamma} \frac{1}{1 - 2^{-\gamma}}, \end{aligned}$$

where the last inequality comes from the fact that $0 < 2^{-\gamma(r-v)} < 1$. Now, let $y, z \in D_B$ be such that $|y - z| \leq 2^{-v}$, for some $v > n$. There are two cases; either there exists an i such that $(i-1)2^{-v} \leq y \leq i2^{-v} \leq z < (i+1)2^{-v}$ or $y, z \in [i2^{-v}, (i+1)2^{-v}$. For the first case, we have by the above that

$$\begin{aligned} |X_y - X_z| &\leq |X_z - X_{i2^{-v}}| + |X_y - X_{(i-1)2^{-v}}| + |X_{i2^{-v}} - X_{(i-1)2^{-v}}| \\ &\leq 2^{-v\gamma}(1 + \frac{2}{1 - 2^{-\gamma}}), \end{aligned}$$

and for the other case, we have that

$$|X_y - X_z| \leq |X_y - X_{i2^{-v}}| + |X_z - X_{i2^{-v}}| \leq \frac{2}{1 - 2^{-\gamma}} 2^{-v\gamma}.$$

Hence we have shown that for any points $y, z \in D_B$ with $|y - z| \leq 2^{-v}$, we obtain the following desired inequality

$$\frac{|X_y - X_z|}{|x - y|^\gamma} \leq 1 + \frac{2}{1 - 2^{-\gamma}},$$

which proves the statement for $m = 1$. Now, let $m > 1$ and Let S be any square of dyadic points of the form

$$S = \{x \in D_B \mid x \in (c_1, c_2)^m,\ 0 < |c_2 - c_1| \leq 2^{-v}\}. \tag{3.2}$$

For $y, z \in S$ and $i = 0, \ldots, m$, define $x^{(i)} \in S$ as

$$x_j^{(i)} = \begin{cases} y_j & \text{if } j \leq i \\ z_j & \text{otherwise} \end{cases}$$

and note that $x^{(0)} = z, x^{(m)} = y$ and for $i = 1, \ldots, m$, $x^{(i-1)}$ and $x^{(i)}$ differ at only one coordinate. Hence, we get that

$$|X_y - X_z| = |\sum_{i=1}^{m} X_{x^{(i)}} - X_{x^{(i-1)}}| \leq \sum_{i=1}^{m} |X_{x^{(i)}} - X_{x^{(i-1)}}|,$$

but for each i, we can view $X_{x^{(i)}} - X_{x^{(i-1)}}$ as a one dimensional random variable, since it is non-vanishing at only one coordinate. Thus, by our previous computations, we obtain the inequality

$$\begin{aligned}\sum_{i=1}^{m} |X_{x^{(i)}} - X_{x^{(i-1)}}| &\leq (1 + \frac{2}{1 - 2^{-\gamma}}) \sum_{i=1}^{m} |x^{(i)} - x^{(i-1)}|^{\gamma} \\ &\leq m(1 + \frac{2}{1 - 2^{-\gamma}})|y - z|^{\gamma},\end{aligned}$$

as $|x^{(i)} - x^{(i-1)}| \leq |y - z|$. Hence we have shown that for every $\gamma < \frac{\alpha}{\beta}$ there exists a negligible set N such that for all $\omega \in N^c$ the restriction of the map $y \mapsto X_y(\omega)$ to D_B is locally γ-Hölder continuous. Indeed, by the above, if we let $\varepsilon = \sqrt{m}2^{-v}$, we have that

$$\sup_{0<|y-z|<\varepsilon} \frac{|X_y - X_z|}{|y - z|^{\gamma}} < \infty,$$

since if $|y - z| < \varepsilon$ there exists a square S of the form (3.2) such that both y and z are points of S. □

Remark 3.2.3. If the reader is not convinced about the choice of ε above, ask yourself "what is the furthest distance between any two points of a square?"

Remark 3.2.4. In the following, we will use the notation $C = 1 + \frac{2}{1-2^{-\gamma}}$.

We shall now prove the Continuity Theorem

Proof. By the previous lemma, we have that for $\gamma < \frac{\alpha}{\beta}$, there exists a negligible event N such that for $\omega \in N^c$, the restriction of the map $y \mapsto X_y(\omega)$ to D_B is locally γ-Hölder continous, where D_B is the set of dyadic points in D. Now, we wish to construct a family of d-dimensional random variables $\{X'_y\}_{y \in D}$ such that $X_y = X'_y$ a.s. for every $y \in D$ and that for every $\omega \in \Omega$ the map $y \mapsto X'_y(\omega)$ is γ-Hölder continuous on every compact subset of D, for all $\gamma < \frac{\alpha}{\beta}$.

For $\omega \in N^c$, we wish to define

$$X'_y(\omega) = \lim_n X_{y_n}(\omega), \tag{3.3}$$

where $\{y_n\}$ is some sequence of dyadic points in D converging to y. However, in order to do so, we need to justify that the right hand side converges and is independent of our choice of sequence. Hence, let $\{y_n\}$ be a sequence of dyadic points in D that converges to y. By local γ-Hölder continuity of the previous lemma, we have that there exists a natural number N such that for all n and k greater than N, we have that

$$|X_{y_n}(\omega) - X_{y_k}(\omega)| \leq C|y_n - y_k|^{\gamma}$$

and since $\{y_n\}$ is convergent, it is also a Cauchy sequence. Therefore, the sequence $\{X_{y_n}\}$ is also a Cauchy sequence, so it converges. For the other part, let $\{z_n\}$ be another sequence of dyadic points that converges to y. Again, by local γ-Hölder continuity of the previous lemma, we have that there exists a natural number N' such that for all n larger than N' we have that

$$|X_{y_n} - X_{z_n}| \leq C|y_n - z_n|^{\gamma},$$

and since the right hand side converges to 0, so does the left hand side. Hence, for $\omega \in N^c$, we may define $X'_y(\omega)$ as in (3.3) and for $\omega \in N$, we simply put $X'_y(\omega) = 0$. We shall now show that for $\omega \in \Omega$, the map $y \mapsto X'_y(\omega)$ is locally γ-Hölder continuous. This is trivial if $\omega \in N$; hence fix $\omega \in N^c$ and let $\varepsilon > 0$ be the ε from the proof of the previous lemma. Consider two distinct points y, z in D such that $|y - z| < \varepsilon$. By Corollary 2.1.1, we get that

$$\begin{aligned} |X'_y - X'_z| &= |\lim_n (X_{y_n} - X_{z_n})| = \lim_n |X_{y_n} - X_{z_n}| \\ &\leq \lim_n C|y_n - z_n|^\gamma = C|y - z|^\gamma, \end{aligned}$$

so for each $\omega \in \Omega$, the map $y \mapsto X'_y(\omega)$ is in fact locally γ-Hölder continuous on D. By Lemma 3.1.2, we have that for each $\omega \in \Omega$, the map $y \mapsto X'_y(\omega)$ is in fact γ-Hölder continuous on each compact subset of D. Lastly, to show that $X_y = X'_y$ $\mathbb{P}$-a.s. we shall make use of Fatou's lemma, which can be found in Friedman [2, Theorem 2.10.5]. By (3.1), we have that

$$\begin{aligned} E[|X_y - X'_y|^\beta] &= E[\lim_n |X_y - X_{y_n}|^\beta] \leq \varliminf_n E[|X_y - X_{y_n}|^\beta] \\ &\leq \varliminf_n |y - y_n|^{m+\alpha} = \sup_{n\geq 0}(\inf_{m\geq n} |y - y_n|^{m+\alpha}) \end{aligned}$$

but for every $n \geq 0$, $\inf_{m\geq n} |y - y_n|^{m+\alpha} = 0$ since y_n converges to y so we have that

$$E[|X_y - X'_y|^\beta] = 0, \tag{3.4}$$

and as shown in Friedman, (3.4) is true if and only if $|X_y - X'_y|^\beta = 0$ $\mathbb{P}$-a.s. [2, Theorem 2.7.3]. This simplifies to $X_y = X'_y$ $\mathbb{P}$-a.s. and we may finish the proof. □

Remark 3.2.5. When we say "let $\varepsilon > 0$ be the ε from the proof of the previous lemma", one could have also said, "let $\varepsilon > 0$ be the ε from Definition 3.1.1, for which the restriction of the map $y \mapsto X_y(\omega)$ to D_B is locally γ-Hölder continuous."

Chapter 4

The Haar Functions

4.1 The Haar Functions

The content of the following section is inspired by the lecture notes made by Lenya Ryzhik [5]. Also, similar to as been mentioned twice before, we shall use non-trivial statements regarding the theory of linear transformations and orthonormal sets, which may be found in Friedman [2, Chapter 4.4 & Chapter 6.4].

Definition 4.1.1 (Haar functions)**.** The functions defined on $[0, 1]$ via

$$\begin{aligned} h_0(t) &= 1, \\ h_n^k(t) &= 2^{\frac{n}{2}} \mathbf{1}_{[\frac{2k}{2^{n+1}}, \frac{2k+1}{2^{n+1}})}(t) - 2^{\frac{n}{2}} \mathbf{1}_{[\frac{2k+1}{2^{n+1}}, \frac{2k+2}{2^{n+1}})}(t), \end{aligned}$$

where n is any non-negative integer and k is any element of $\{0, 1, \ldots, 2^n - 1\}$, are called *Haar functions* on the unit interval.

In order to to simplify some computations, we state an alternative, yet equivalent definition of the Haar functions.

Lemma 4.1.1. *Let* $\rho(t) = \mathbf{1}_{[0,\frac{1}{2})}(t) - \mathbf{1}_{[\frac{1}{2},1)}(t)$*. We may then write the Haar functions as*

$$\begin{aligned} h_0(t) &= 1, \\ h_n^k(t) &= 2^{\frac{n}{2}} \rho(2^n t - k). \end{aligned}$$

Remark 4.1.1. Note that $\rho(t)$ is defined to be $h_0^0(t)$. This function $\rho(t)$ is from now on exclusively used to denote $\mathbf{1}_{[0,\frac{1}{2})}(t) - \mathbf{1}_{[\frac{1}{2},1)}(t)$.

Proof. Since $t \in [\frac{2k}{2^{n+1}}, \frac{2k+1}{2^{n+1}})$ if and only if $(2^n t - k) \in [0, \frac{1}{2})$ and $t \in [\frac{2k+1}{2^{n+1}}, \frac{2k+2}{2^{n+1}})$ if and only if $(2^n t - k) \in [\frac{1}{2}, 1)$ the statement follows. □

Lemma 4.1.2. *Any Haar function except* $h_0(t)$ *has the property that*

$$\int_0^1 h_n^k(s) ds = 0$$

Proof.

$$\begin{aligned} \int_0^1 h_n^k(s) ds &= 2^{\frac{n}{2}} \Big(\int_0^1 \mathbf{1}_{[\frac{2k}{2^{n+1}}, \frac{2k+1}{2^{n+1}})} ds - \int_0^1 \mathbf{1}_{[\frac{2k+1}{2^{n+1}}, \frac{2k+2}{2^{n+1}})} ds \Big) \\ &= 2^{\frac{n}{2}} \Big(\frac{1}{2^{n+1}} - \frac{1}{2^{n+1}} \Big) = 0. \end{aligned}$$

□

Lemma 4.1.3. *The set of Haar functions are orthonormal in* $L^2([0, 1], \mathcal{B}([0, 1]), \lambda)$*, where* λ *is the Lebesgue measure.*

Proof. We begin to show that $\langle h_n^k, h_{n'}^{k'} \rangle = 0$ if either $n \neq n'$, or if $k \neq k'$. First, if $n = n'$ and $k \neq k'$, h_n^k and $h_n^{k'}$ have disjoint support, so their inner product is trivially 0. For the other case, suppose that $n \neq n'$ and without loss of generality we may also assume that $n < n'$ and consider

$$\langle h_n^k, h_{n'}^{k'} \rangle = \int_0^1 h_n^k(s) h_{n'}^{k'}(s) ds = 2^{\frac{n+n'}{2}} \int_0^1 \rho(2^n s - k)\rho(2^{n'} s - k')$$
$$= 2^{\frac{n+n'}{2} - n'} \int_{k2^{-n}}^{(1+k)2^{-n}} \rho(t)\rho(2^{n'-n}(t+k) - k')dt,$$

where the last equality comes from the substitution $t = 2^n s - k$. Note that $k2^{-n} = \frac{2k}{2^{n+1}}$ and $(k+1)2^{-n} = \frac{2k+2}{2^{n+1}}$, so with $C = 2^{\frac{n+n'}{2} - n'}$, we get that the above equals

$$C(\int_0^{\frac{1}{2}} \rho(2^{n'-n}(t+k) - k')dt - \int_{\frac{1}{2}}^1 \rho(2^{n'-n}(t+k) - k')dt)$$
$$= 2^{n-n'} C(\int_{2^{n'-n}k-k'}^{2^{n'-n}(k+\frac{1}{2})-k'} \rho(r)dr - \int_{2^{n'-n}(k+\frac{1}{2})-k'}^{2^{n'-n}(k+1)-k'} \rho(r)dr),$$

as we made the substitution $r = 2^{n'-n}(t+k) - k'$. Now, by our assumption of $n' > n$, we have that $2^{n'-n}$ is some natural number that is a multiple of 2. It follows that all the bounds of integration above are integers and thus, by the support of $\rho(t)$ and Lemma 4.1.2, $\langle h_n^k, h_{n'}^{k'} \rangle = 0$. Lastly,

$$\langle h_n^k, h_n^k \rangle = \int_0^1 (h_n^k(s))^2 = 1$$

and the statement follows. □

Definition 4.1.2. We say that the inner product

$$\langle f, h_n^k \rangle = \int f(x) h_n^k(x) dx =: c_{nk}$$

are the Haar coefficients with respect to $f \in L^2([0,1], \mathcal{B}([0,1]), \lambda)$ and we define the Haar series of f as

$$h(f)(x) := \langle f, h_0 \rangle h_0 + \sum_{n=0}^{\infty} \sum_{k=0}^{2^n - 1} c_{nk} h_n^k(x).$$

Remark 4.1.2. We shall prove the convergence of the Haar series later on.

Definition 4.1.3. Let D_n be the set of dyadic intervals of length 2^{-n} in $[0,1)$. We use the notation $D_n = \{I_{n,k} \mid k \in \{0, 1, \ldots, 2^n - 1\}\}$, where

$$I_{n,k} = [\frac{2k}{2^{n+1}}, \frac{2k+2}{2^{n+1}}).$$

Further, we shall use the notation $m_I(f)$ to denote the average value of f over the interval I; i.e

$$m_I(f) = \frac{1}{|I|} \int_I f(s) ds.$$

Lastly, for any $f \in L^2([0,1], \mathcal{B}([0,1]), \lambda)$, where λ denotes the Lebesgue measure, we say that

$$P_n(f)(x) = \sum_{I \in D_n} m_I(f) \mathbf{1}_I(x)$$

is the dyadic projection of f in $L^2([0,1], \mathcal{B}([0,1]), \lambda)$.

Remark 4.1.3. Clearly, for each n, all intervals in D_n are disjoint.

Lemma 4.1.4. *The set $\{P_n\}$ of dyadic projections on $L^2([0,1], \mathcal{B}([0,1]), \lambda)$, is a family of bounded linear operators with norms less than or equal to 1.*

Proof. The fact that it is a linear operator is a direct consequence of the properties of integration. To prove that it is bounded, we use the above remark to write

$$\|P_n(f)\|_{L^2}^2 = \int_0^1 |P_n(f)(s)|^2 ds = \sum_{I \in D_n} \int_I |P_n(f)(s)|^2 ds.$$

By the definition of $P_n(f)$, we have that for $s \in I$, $P_n(f)(s) = m_I(f)$, which is a constant, so we get that the above equals

$$\begin{aligned}\sum_{I \in D_n} \int_I |m_I(f)|^2 ds &= \sum_{I \in D_n} |m_I(f)|^2 |I| = \sum_{I \in D_n} \frac{1}{|I|} |(\int_I f(t)dt)^2| \\ &\leq \sum_{I \in D_n} \frac{1}{|I|} (\int_I |f(t)|dt)^2 = \sum_{I \in D_n} \frac{1}{|I|} (\|f\|_{L^1[I]})^2 \\ &\leq \sum_{I \in D_n} \frac{1}{|I|} (\|1\|_{L^2[I]} \|f\|_{L^2[I]})^2 \\ &= \sum_{I \in D_n} \int_I |f(t)|^2 dt = \int_0^1 |f(t)|^2 dt,\end{aligned}$$

where the second inequality is an application of Hölder's inequality. We have thus obtained that

$$\|P_n(f)\|_{L^2} \leq \|f\|_{L^2},$$

which proves the statement. □

Lemma 4.1.5. *For any $f \in L^2([0,1], \mathcal{B}([0,1]), \lambda)$, $P_j(f)(x)$ equals the $(j-1)$:th partial sum of the Haar series of f.*

Proof. First, note that each interval $I_{n,k}$ may be written as a disjoint union in the following way

$$\begin{aligned}I_{n,k} =& [\frac{2k}{2^{n+1}}, \frac{2k+2}{2^{n+1}}) = [\frac{2k}{2^{n+1}}, \frac{2k+1}{2^{n+1}}) \cup [\frac{2k+1}{2^{n+1}}, \frac{2k+2}{2^{n+1}}) \\ =& [\frac{4k}{2^{n+2}}, \frac{4k+2}{2^{n+2}}) \cup [\frac{4k+2}{2^{n+2}}, \frac{4k+4}{2^{n+1}}) = I_{n+1,2k} \cup I_{n+1,2k+1}.\end{aligned}$$

Define $I_{n,k}^l := I_{n+1,2k}$ and $I_{n,k}^r := I_{n+1,2k+1}$. Clearly, $|I| = 2|I^l| = 2|I^r|$, so by linearity of integration we can rewrite $m_I(f)$ as

$$m_I(f) = \frac{1}{2}(m_{I^l}(f) + m_{I^r}(f)).$$

With this in mind, fix an interval $I_{n,k} \in D_n$ and suppose first that $x \in I_{n,k}^l$, then

$$\begin{aligned}P_{n+1}(f)(x) - P_n(f)(x) &= \sum_{I \in D_{n+1}} m_I(f)\mathbf{1}_I(x) - \sum_{I \in D_n} m_I(f)\mathbf{1}_I(x) \\ &= m_{I_{n,k}^l}(f) - m_{I_{n,k}}(f) \\ &= m_{I_{n,k}^l}(f) - \frac{1}{2}(m_{I_{n,k}^l}(f) + m_{I_{n,k}^r}(f)) \\ &= \frac{1}{2}(m_{I_{n,k}^l}(f) - m_{I_{n,k}^r}(f)).\end{aligned}$$

For $x \in I_{n,k}^r$, we have by similar calculations, that

$$P_{n+1}(f)(x) - P_n(f) = -\frac{1}{2}(m_{I_{n,k}^l}(f) - m_{I_{n,k}^r}(f)).$$

Since for any non negative integer n and for any k in $\{0, 1, \ldots, 2^n - 1\}$, $|I_{n,k}| = 2^{-n}$, the Haar function h_n^k may be rewritten as

$$h_n^k(x) = \frac{1}{\sqrt{|I_{n,k}|}}(\mathbf{1}_{I_{n,k}^l}(x) - \mathbf{1}_{I_{n,k}^r}(x)).$$

Thus, for $x \in I_{n,k}$

$$P_{n+1}(f)(x) - P_n(f)(x) = \langle f, h_n^k \rangle h_n^k(x).$$

Hence, for any $x \in [0,1]$,

$$P_{n+1}(f)(x) - P_n(f)(x) = \sum_{k=0}^{2^n-1} \langle f, h_n^k \rangle h_n^k(x),$$

since if $x = 1$ both sides are 0. Furthermore, for any $j \geq 1$ and $x \in [0,1]$, we can write

$$\begin{aligned} P_j(f)(x) - P_0(f)(x) &= \sum_{n=0}^{j-1} (P_{n+1}(f)(x) - P_n(f)(x)) \\ &= \sum_{n=0}^{j-1} \sum_{k=0}^{2^n-1} \langle f, h_n^k \rangle h_n^k(x), \end{aligned}$$

and since $P_0(f) = \langle f, h_0 \rangle h_0$, the statement is proved. □

Theorem 4.1.1. *For any $f \in L^2([0,1], \mathcal{B}([0,1]), \lambda)$, the Haar series of f converges to f in the L^2-norm.*

Proof. Since the set of continuous function on the unit interval, $C[0,1]$, is dense in $L^2[0,1]$, we begin by considering functions $g \in C[0,1]$. Recall that continuous functions on compact intervals are uniformly continuous, so we use the fact that for every $\varepsilon > 0$, there exists a natural number N_ε such that for all $n \geq N_\varepsilon$

$$|x - y| \leq 2^{-n} \implies |g(x) - g(y)| < \varepsilon.$$

By the mean value theorem, for every dyadic interval $I \subset [0,1]$, there exists an interior point c_I of I such that

$$m_I(g) = \frac{1}{|I|} \int_I g(s) ds = g(c_I).$$

Now, fix $\varepsilon \geq 0$ and pick an N_ε such that for all $n \geq N_\varepsilon$ and for every $x \in [0,1]$, there exists a point $c_{I_{n,k}} \in I_{n,k}$ for some k such that

$$|x - c_{I_{n,k}}| \leq 2^{-n} \implies |g(x) - g(c_{I_{n,k}})| = |g(x) - m_{I_{n,k}}(g)| = |g(x) - P_n(g)(x)| < \varepsilon,$$

where the last equality comes from the fact that the intervals $I_{n,k} \in D_n$ are disjoint. This means that for any $g \in C[0,1]$, the sequence of functions $x \mapsto P_n(g)(x)$ converges uniformly to g, so by Lemma 4.1.5, the Haar series of g converges uniformly to g. This implies L^2-convergence in $[0,1]$ since

$$\begin{aligned} \|P_n(g) - g\|_{L^2[0,1]}^2 &= \int_0^1 |P_n(g)(s) - g(s)|^2 ds \leq 1 \sup_{s \in [0,1]} |P_n(g)(s) - g(s)|^2 \\ &= (\sup_{s \in [0,1]} |P_n(g)(s) - g(s)|)^2 \end{aligned}$$

which tends to 0 by our previous argumentation.

Now, we want to extend this to the whole space $L^2([0,1], \mathcal{B}([0,1]), \lambda)$. Fix $\varepsilon > 0$ and let $f \in L^2([0,1], \mathcal{B}([0,1]), \lambda)$, since $C[0,1]$ is dense in $L^2[0,1]$, there exists a function $g \in C[0,1]$ such that

$$\|f - g\|_{L^2} < \frac{\varepsilon}{3}.$$

By our previous discussion on uniform continuity, there exists a natural number N such that for all $n \geq N$

$$\|P_n(g) - g\|_{L^2} < \frac{\varepsilon}{3}.$$

Thus, by Lemma 4.1.4, in combination with Minkowski's inequality, we have that

$$\begin{aligned} \|P_n(f) - f\|_{L^2} &\leq \|P_n(f) - P_n(g)\|_{L^2} + \|P_n(g) - g\|_{L^2} + \|f - g\|_{L^2} \\ &= \|P_n(f - g)\|_{L^2} + \|P_n(g) - g\|_{L^2} + \|f - g\|_{L^2} \\ &\leq \|P_n\| \, \|f - g\|_{L^2} + \|P_n(g) - g\|_{L^2} + \|f - g\|_{L^2} \leq 3\frac{\varepsilon}{3} = \varepsilon \end{aligned}$$

so by 4.1.5, the statement follows. □

Corollary 4.1.1. *The set of Haar functions forms an orthonormal basis for* $L^2([0,1], \mathcal{B}([0,1]), \lambda)$.

Proof. By Lemma 4.1.3, the set of Haar functions are orthonormal in $L^2([0,1], \mathcal{B}([0,1]), \lambda)$ and by Theorem 4.1.1, every function in $f \in L^2([0,1], \mathcal{B}([0,1]), \lambda)$ may be written as

$$f(x) = \langle f, h_0 \rangle h_0 + \sum_{n=0}^{\infty} \sum_{k=0}^{2^n - 1} c_{nk} h_n^k(x),$$

which means that the Haar functions span $L^2([0,1], \mathcal{B}([0,1]), \lambda)$. $\square$

Chapter 5

Brownian motion

5.1 A Construction of a Brownian motion

The outline of the following construction was found in the lecture notes of Sanz-Solé [6, Section 2.2].

Definition 5.1.1. A real valued stochastic process $B = (\Omega, \mathcal{A}, \{\mathcal{A}_t\}_{t\in T}, \{B_t\}_{t\in T}, \mathbb{P})$ is said to be a one-dimensional Brownian motion if the following conditions are satisfied

$$\begin{aligned} &\text{i)}\ B_0 = 0 \text{ a.s.,}\\ &\text{ii)}\ \text{if } 0 \le s \le t,\ \text{then } B_t - B_s \text{ is independent of } \mathcal{A}_s,\\ &\text{iii)}\ \text{if } 0 \le s \le t,\ \text{then } B_t - B_s \text{ is Gaussian with mean 0 and variance } t - s. \end{aligned}$$

Lemma 5.1.1. *Let $B = (\Omega, \mathcal{A}, \{\mathcal{A}_t\}_{t\in T}, \{B_t\}_{t\in T}, \mathbb{P})$ be a Brownian motion. Then, for any $0 \le u \le v \le s \le t$, $B_t - B_s$ and $B_v - B_u$ are independent.*

Proof. By Corollary 2.2.1, we have that for any $D \in \mathcal{B}(\mathbb{R}^n)$, $(B_v - B_u)^{-1}(D) \in \mathcal{A}_v$. Since $\mathcal{A}_v \subset \mathcal{A}_s$, we have by property ii) of the above definition that for any $D_1, D_2 \in \mathcal{B}(\mathbb{R}^n)$

$$\mathbb{P}((B_t - B_s)^{-1}(D_1) \cap (B_v - B_u)^{-1}(D_2)) = \mathbb{P}((B_t - B_s)^{-1}(D_1))\mathbb{P}((B_v - B_u)^{-1}(D_2)),$$

which proves the statement. □

Lemma 5.1.2. *Let $f \in L^2([0,1], \mathcal{B}([0,1]), \lambda)$ and let $\{N_0, N_n^k\}$ be a family of independent Gaussian random variables with mean 0 and variance 1. The transform I_N from $L^2([0,1], \mathcal{B}([0,1]), \lambda)$ to $L^2(\Omega, \mathcal{A}, \mathbb{P})$*

$$I_N(f) = \langle f, h_0\rangle N_0 + \sum_{n=0}^{N}\sum_{k=0}^{2^n-1} \langle f, h_n^k\rangle N_n^k,$$

converges in $L^2(\Omega)$ to

$$I(f) = \langle f, h_0\rangle N_0 + \sum_{n=0}^{\infty}\sum_{k=0}^{2^n-1} \langle f, h_n^k\rangle N_n^k.$$

as N tends to ∞.

Proof. We begin by showing that $I_N(f)$ is a Cauchy sequence. Let

$$S_N(f) = \sum_{n=0}^{N}\sum_{k=0}^{2^n-1} |\langle f, h_n^k\rangle|^2.$$

By Remark 2.2.9, we have that

$$\begin{aligned} \|I_N(f) - I_M(f)\|_{L^2}^2 &= E[(I_N(f) - I_M(f))^2]\\ &= Var(I_N(f) - I_M(f)) + E[I_N(f) - I_M(f)]^2. \end{aligned}$$

Since the family $\{N_0, N_n^k\}$ consists of independent random variables, we may by Corollary 2.2.3 compute the variance above as

$$\begin{aligned} Var(I_N(f) - I_M(f)) =& Var(\sum_{n=}^{N}\sum_{k=0}^{2^n-1}\langle f, h_n^k\rangle N_n^k) = \sum_{n=}^{N}\sum_{k=0}^{2^n-1}\langle f, h_n^k\rangle^2 Var(N_n^k) \\ =& S_N(f) - S_M(f), \end{aligned}$$

and since

$$\begin{aligned} E[I_N(f) - I_M(f)] &= E[\sum_{n=}^{N}\sum_{k=0}^{2^n-1}\langle f, h_n^k\rangle N_n^k] \\ &= \sum_{n=}^{N}\sum_{k=0}^{2^n-1}\langle f, h_n^k\rangle E[N_n^k] = 0, \end{aligned}$$

we get that

$$\|I_N(f) - I_M(f)\|_{L^2}^2 = S_N(f) - S_M(f),$$

which tends to 0 as N and M goes to infinity by Theorem 4.1.1. By Corollary 2.1.1 we have that

$$\lim_N \|I_N(f)\| = \left\|\lim_N I_N(f)\right\| = \|I(f)\|,$$

which proves the statement. □

Remark 5.1.1. We may now for any $f \in L^2([0,1], \mathcal{B}([0,1]), \lambda)$ refer to $I(f)$ as the transform above.

Corollary 5.1.1. *The transform I is an isometry between $L^2([0,1], \mathcal{B}([0,1]), \lambda)$ and $L^2(\Omega, \mathcal{A}, I\!P)$.*

Proof. To shorten the notation we shall use the notation $L^2([0,1])$ and $L^2(\Omega)$ to denote the two spaces of interest. By the computations of the proof of the previous lemma, we have that

$$\begin{aligned} \|I(f)\|_{L^2(\Omega)}^2 &= \lim_N \|I_N(f)\|_{L^2(\Omega)}^2 = \lim_N E[I_N(f)^2] \\ &= \lim_N (Var(I_N(f)) + E[I_N(f)]^2) \\ &= \langle f, h_0\rangle^2 Var(N_0) + \lim_N \sum_{n=0}^{N}\sum_{k=0}^{2^n-1} < f, h_n^k >^2 Var(N_n^k) \\ &= \langle f, h_0\rangle^2 + \sum_{n=0}^{\infty}\sum_{k=0}^{2^n-1} < f, h_n^k >^2 = \|f\|_{L^2([0,1])}^2, \end{aligned}$$

where the last equality comes from Theorem 2.1.2. □

Theorem 5.1.1. *The stochastic process $(\Omega, \mathcal{A}, \{\mathcal{A}_t\}_{t\in[0,1]}, \{B_t\}_{t\in[0,1]}, I\!P)$, where the family $\{B_t\}_{t\in[0,1]}$ of random variables is defined via $B_t = I(\mathbf{1}_{[0,t]})$ and $\{\mathcal{A}_t\}_{t\in[0,1]}$ is given by the natural filtration, defines a one-dimensional Brownian motion.*

Proof. First, note that for $0 \le s \le t \le 1$, $B_t - B_s = I(\mathbf{1}_{(s,t]})$. We check the conditions of Definition 5.1.1. Trivially, $B_0 = 0$ so the first condition is satisfied. For the proof of the second and third condition we shall refer to Schilling & Partzsch [4, Section 3.1]. □

Remark 5.1.2. The proof made in to Schilling & Partzsch above is done using characteristic functions, (i.e the Fourier Transform), which we haven't discussed in this text. However, it uses the fact that if a random vector (X, Y) is jointly Gaussian and the components are uncorrelated, i.e that $E[XY] = E[X]E[Y]$, then X and Y are independent. Hence, if one shows that $(I(\mathbf{1}_{(s,t]}), I(\mathbf{1}_{(0,r]}))$, is jointly Gaussian for any $r \le s$, then we simply use the isometry property to calculate $E[I(\mathbf{1}_{(s,t]})I(\mathbf{1}_{(0,r]})] = \langle I(\mathbf{1}_{(s,t]}), I(\mathbf{1}_{(0,r]})\rangle = 0$. Nevertheless, we still need to show that $I(\mathbf{1}_{(u,v]})$ have law $N(0, v-u)$ for any $u \le v$ to prove both properties.

5.2 Regularity of the Brownian Motion

The first Theorem below is of the lines of what is given in the lecture notes of Sanz-Solé [6, Section 2.2]. The outline of the proof for the Iterated Logarithm Law along with the relevant lemmas were found in Baldi [1, Section 3.4].

Theorem 5.2.1. *For a one-dimensional Brownian motion $B = (\Omega, \mathcal{A}, \{\mathcal{A}_t\}_{t\in T}, \{X_t\}_{t\in T}, I\!P)$, the sample paths of $\{X_t\}_{t\in T}$ are almost surely γ-Hölder continous on any compact subset of T, for any $\gamma \in (0, \frac{1}{2})$.*

Proof. We intent to use Kolmogorov's Continuity Theorem, so we need to show that there exists $\alpha, \beta, c > 0$ such that for every $s, t \in T$,

$$E[|X_t - X_s|^\beta] \leq c|t-s|^{1+\alpha}.$$

In this proof we shall use what is mentioned in Remark 2.2.10. Hence, for any natural number k, and with the notation that $L = \frac{1}{(t-s)\sqrt{2\pi}}$, we can compute as follows

$$\begin{aligned} E[|X_t - X_s|^{2k}] &= E[(X_t - X_s)^{2k}] = L(\int_{\mathbb{R}} x^{2k} e^{-\frac{1}{2}(\frac{x}{(t-s)})^2} dx) \\ &= L([x^{2k-1}(-(t-s)^2 e^{-\frac{1}{2}(\frac{x}{(t-s)})^2})]_{-\infty}^{\infty} \\ &+ \int_{\mathbb{R}} (2k-1)x^{2k-2}(t-s)^2 e^{-\frac{1}{2}(\frac{x}{(t-s)})^2} dx) \\ &= L(2k-1)(t-s)^2 \int_{\mathbb{R}} x^{2k-2} e^{-\frac{1}{2}(\frac{x}{(t-s)})^2} dx. \end{aligned}$$

Continuing in this fashion and using Lemma 2.2.6, we obtain

$$L(2k-1)(2k-3)(2k-5)\cdot\ldots\cdot 3\cdot 1(t-s)^{2\frac{k}{2}} \int_{\mathbb{R}} e^{-\frac{1}{2}(\frac{x}{(t-s)})^2} = \frac{(2k)!}{2^k k!}(t-s)^k.$$

Therefore, with $\alpha = k-1, \beta = 2k, c = \frac{(2k)!}{2^k k!}$, we find that

$$E[|X_t - X_s|^\beta] \leq c|t-s|^{1+\alpha}.$$

Since for any $\gamma \in (0, \frac{1}{2})$ there exists a k such that $\gamma < \frac{\alpha}{\beta}$. Indeed, since when k tends to infinity, $\frac{\alpha}{\beta}$ tends to $\frac{1}{2}$. Hence the statement follows by Kolmogorov's Continuity Theorem. □

Corollary 5.2.1. *The sample paths of a one-dimensional Brownian motion are almost surely continuous.*

Proof. This follows by the previous theorem and Lemma 3.1.1. □

Lemma 5.2.1. *Let $(\Omega, \mathcal{A}, \{\mathcal{A}_t\}_{t\in T}, \{X_t\}_{t\in T}, I\!P)$ be a Brownian motion, then for any real non-negative numbers x and T,*

$$I\!P(\sup_{0\leq t\leq T} X_t > x) \leq 2I\!P(X_T > x).$$

Proof. Let $I = \{t_0, t_1, \ldots, t_n\}$, be an increasing sequence of points in $[0, T]$ such that $t_n = T$ and such that $\{t_j \mid X_{t_j}(\omega) > x\}$, if no such sequence exists, then the statement follows trivially. Assume therefore that such a sequence exists and let $\tau(\omega) = \inf\{t_j \mid X_{t_j}(\omega) > x\}$. If $X_T(\omega) > x$, then $\tau(\omega) \leq T$, which implies that $\{\omega \in \Omega \mid X_T(\omega) > x\} \subset \{\omega \in \Omega \mid \tau(\omega) \leq T\}$. Furthermore, the sets of the form

$$\{\omega \in \Omega \mid \tau(\omega) = t_j, X_T > x\}$$

where j is an integer such that $0 \leq j \leq n$, are disjoint. We can thus proceed with the following computations

$$\begin{aligned} \mathbb{P}(X_T(\omega) > x) &= \mathbb{P}(\tau(\omega) \leq T, X_T > x) \\ &= \sum_{j=0}^{n} \mathbb{P}(\tau(\omega) = t_j, X_T(\omega) > x) \\ &\geq \sum_{j=0}^{n} \mathbb{P}(\tau(\omega) = t_j, X_T(\omega) - X_{t_j}(\omega) \geq 0), \end{aligned}$$

as when $\tau(\omega) = t_j$, we have that $X_{t_j} > x$. By the definition of $\tau(\omega)$, we may write

$$\{\omega \in \Omega \mid \tau(\omega) = t_j\} = \{\omega \in \Omega \mid X_{t_0}(\omega) \leq x, \dots, X_{t_{j-1}}(\omega) \leq x, X_{t_j}(\omega) > x\},$$

which by Remark 2.2.3 and the definition of a stochastic process is an element of $\mathcal{A}_{t_j}$. By Lemma 2.2.1 we also have that

$$\{\omega \in \Omega \mid X_T - X_{t_j} \geq 0\} \in \sigma(X_T - X_{t_j}).$$

By definition 2.2.8 and the result of Lemma 2.2.6, we easily compute that $\mathbb{P}(X_T - X_{t_j} \geq 0) = \frac{1}{2}$. By the definition of Brownian motion, we have that $\mathcal{A}_{t_j}$ and $\sigma(X_T - X_{t_j})$ are independent, so we may thus compute as

$$\begin{aligned}
\mathbb{P}(X_T(\omega) > x) &\geq \sum_{j=0}^{n} \mathbb{P}(\tau(\omega) = t_j)\mathbb{P}(X_T(\omega) - X_{t_j}(\omega) \geq 0) \\
&= \frac{1}{2}\sum_{j=0}^{n} \mathbb{P}(\tau(\omega) = t_j)
\end{aligned}$$

which by disjointedness of the sets $\{\omega \in \Omega \mid \tau(\omega) = t_j\}$ equals

$$\begin{aligned}
\frac{1}{2}\mathbb{P}(\cup_{j=0}^{n}\{\tau(\omega) = t_j\}) &= \frac{1}{2}\mathbb{P}(\cup_{j=0}^{n}\{X_{t_0}(\omega) \leq x, \dots, X_{t_{j-1}}(\omega) \leq x, X_{t_j}(\omega) > x\}) \\
&= \frac{1}{2}\mathbb{P}(\{X_{t_0}(\omega) > x\} \cup \{X_{t_0}(\omega) \leq x, X_{t_1}(\omega) > x\} \\
&\cup \{X_{t_0}(\omega) \leq x, X_{t_1}(\omega) \leq x, X_{t_2}(\omega) > x\} \cup \dots \\
&\cup \{X_{t_0}(\omega) \leq x, \dots, X_{t_{n-1}}(\omega) \leq x, X_{t_n}(\omega) > x\}) \\
&= \frac{1}{2}\mathbb{P}(\{\omega \in \Omega \mid \sup_{t \in I} X_t(\omega) > x\}),
\end{aligned}$$

where one can see this last equality by picking an ω and checking if $X_{t_0}(\omega) > x$ and if not, checking X_{t_1} and so on.

Now, let I_n be a sequence of finite subsets of $[0, T]$ that increases to $\mathbb{Q} \cap [0, T]$, such that for each n the last point of I_n equals T and where the set $\{t_j \mid X_{t_j} > x\}$ is non-empty, where the t_j:s denotes the points of I_n. The existence of such a sequence comes from Corollary 5.2.1. Since the sets $\{\omega \in \Omega \mid \sup_{t \in I_n} X_t(\omega) > x\}$ increases with n, we obtain that

$$\begin{aligned}
\mathbb{P}(\sup_{t \in [0,T]} X_t > x) = \mathbb{P}(\sup_{t \in \mathbb{Q} \cap [0,T]} X_t > x) &= \lim_n \mathbb{P}(\sup_{t \in I_n} X_t > x) \\
&\leq \lim_n 2\mathbb{P}(X_T > x) = 2\mathbb{P}(X_T > x)
\end{aligned}$$

and the statement follows. □

Lemma 5.2.2. *for $x > 0$*

$$(x + \frac{1}{x})^{-1} e^{-\frac{x^2}{2}} \leq \int_x^\infty e^{-\frac{y^2}{2}} dy \leq \frac{1}{x} e^{-\frac{x^2}{2}}.$$

Proof. Since we are integrating over the interval (x, ∞), we have that

$$\int_x^\infty e^{-\frac{y^2}{2}} dy \leq \frac{1}{x}\int_x^\infty y e^{-\frac{y^2}{2}} dy = \frac{1}{x}\Big[- e^{-\frac{y^2}{2}} \Big]_x^{+\infty} = \frac{1}{x} e^{-\frac{x^2}{2}},$$

which proves the right hand side inequality. Furthermore, since

$$\frac{d}{dx}\frac{1}{x} e^{-\frac{x^2}{2}} = -(1 + \frac{1}{x^2}) e^{-\frac{x^2}{2}},$$

and because $(1 + \frac{1}{x^2})$ decreases for positive x, we have

$$\frac{1}{x} e^{-\frac{x^2}{2}} = \int_x^\infty (1 + \frac{1}{y^2}) e^{-\frac{y^2}{2}} dy \leq (1 + \frac{1}{x^2}) \int_x^\infty e^{-\frac{y^2}{2}} dy,$$

which by dividing by $(1 + \frac{1}{x^2})$ on both sides yields the left hand side of the statement. □

Theorem 5.2.2 (Iterated Logarithm law)**.** *Let $(\Omega, \mathcal{A}, \{\mathcal{A}_t\}_{t\in T}, \{X_t\}_{t\in T}, I\!P)$ be a Brownian motion, then for any decreasing sequence $\{t_n\} \subset T$, which converges to 0, it holds that*

$$I\!P(\overline{\lim_n} \frac{X_{t_n}}{(2t_n \log\log \frac{1}{t_n})^{\frac{1}{2}}} = 1) = 1.$$

Remark 5.2.1. Note that by Corollary 5.2.1 a one-dimensional Brownian motion is continuous, so if the statement is true for one such sequence $\{t_n\}$ as above, it holds for all such sequences $\{t_n\}$ above.

Proof. We begin to show that

$$\overline{\lim_n} \frac{X_{t_n}}{(2t_n \log\log \frac{1}{t_n})^{\frac{1}{2}}} \leq 1 \text{ a.s.}$$

Let $\phi(t_n) = (2t_n \log\log \frac{1}{t_n})^{\frac{1}{2}}$, fix $\delta > 0$ and define

$$A_n = \{\omega \in \Omega \mid X_{t_n}(\omega) > (1+\delta)\phi(t_n)\}.$$

If we can show that $\mathbb{P}(\overline{\lim}_n A_n) = 0$, then the desired inequality is proven. Indeed, by Lemma 2.1.3 the set of points that are in infinitely many A_n:s have probability 0, so for every $\delta > 0$ there exists an n such that

$$\mathbb{P}(\sup_{m>n}\{\frac{X_{t_n}}{\phi(t_n)} > 1+\delta\}) = 0,$$

which implies that

$$\overline{\lim_n} \frac{X_{t_n}}{\phi(t_n)} \leq 1 \text{ a.s..}$$

Since ϕ is increasing,

$$A_n \subset \{\sup_{0\leq t\leq t_n} X_t > (1+\delta)\phi(t_{n+1})\},$$

and since $\frac{X_{t_n}}{\sqrt{t_n}} \sim N(0,1)$, it follows by Lemma 5.2.1 and Lemma 5.2.2, that for $t_{n+1} < \frac{1}{e}$, we have that

$$\begin{aligned}
\mathbb{P}(A_n) &\leq \mathbb{P}(\sup_{0\leq t\leq t_n} X_t \geq (1+\delta)\phi(t_{n+1})) \leq 2\mathbb{P}(X_{t_n} \geq (1+\delta)\phi(t_{n+1}) \\
&= 2\mathbb{P}(\frac{X_{t_n}}{\sqrt{t_n}} \geq (1+\delta)(2\frac{t_{n+1}}{t_n}\log\log\frac{1}{t_{n+1}})^{\frac{1}{2}}) = \frac{2}{\sqrt{2\pi}}\int_{x_n}^{+\infty} e^{-\frac{y^2}{2}}dy \\
&\leq \sqrt{\frac{2}{\pi}}\frac{1}{x_n}e^{-\frac{x_n^2}{2}},
\end{aligned}$$

where

$$x_n = (1+\delta)(2\frac{t_{n+1}}{t_n}\log\log\frac{1}{t_{n+1}})^{\frac{1}{2}}.$$

By Remark 5.2.1 we can restrict ourselves to the case when $t_n = q^n$ for some q between 0 and 1, such that $\lambda = q(1+\delta)^2 > 1$. Now if we write $\alpha = \log\frac{1}{q}$, which is positive, we get that

$$x_n = (1+\delta)(2q\log[(n+1)\log\frac{1}{q}])^{\frac{1}{2}} = [2\lambda\log(\alpha(n+1))]^{\frac{1}{2}}.$$

Note that, for each q there exists a k such that $x_j > 1$ for all $j \geq k$. Fix such a k, so that by the computations above, we obtain the upper bound

$$\begin{aligned}
\sum_{j=k}^{\infty}\mathbb{P}(A_j) &\leq \sum_{j=k}^{\infty}\sqrt{\frac{2}{\pi}}\frac{1}{x_j}e^{-\frac{x_j^2}{2}} \leq \sum_{j=k}^{\infty}\sqrt{\frac{2}{\pi}}e^{\log(\alpha(j+1))^{-\lambda}} \\
&= \alpha\sqrt{\frac{2}{\pi}}\sum_{j=k}^{\infty}\frac{1}{(j+1)^{\lambda}} < +\infty
\end{aligned}$$

since $\lambda > 1$, so the desired inequality follows by the first Borel-Cantelli lemma. We now intend to prove the reverse inequality, namely

$$\overline{\lim_{t\to 0+}} \frac{X_t}{(2t\log\log\frac{1}{t})^{\frac{1}{2}}} \geq 1 \text{ a.s.}$$

Again, let $\phi(t)$ be defined as before and let $\{t_n\}$ be a sequence in T converging to 0. For any $\delta > 0$, define the set

$$F_n = \{\omega \in \Omega \mid X_{t_n}(\omega) > (1-\delta)\phi(t_n)\}.$$

In a similar fashion as before the above inequality is equivalent to showing that for any $\delta > 0$, $\mathbb{P}(\overline{\lim}_n F_n) = 1$. However, as will become clear later, we first intend to show that for every $\varepsilon > 0$ the set

$$G_n = \{\omega \in \Omega \mid X_{t_n} - X_{t_{n+1}} > (1-\varepsilon)\phi(t_n)\}$$

has the property that $\mathbb{P}(\overline{\lim}\, G_n) = 1$. Fix $\varepsilon > 0$ and note that by the definition of Brownian motion and by Lemma 5.1.1, $Z_n = X_{t_n} - X_{t_{n+1}}$ are independent random variables with law $N(0, t_n - t_{n+1})$ for any n; in particular, the family $\{G_n\}$ is an independent family. Further, by Remark 2.2.9, we have that $\frac{Z_n}{\sqrt{t_n - t_{n+1}}}$ are $N(0,1)$ random variables. Hence, by Lemma 5.2.2, we have that for every $x > 1$

$$\begin{aligned}
\mathbb{P}(\frac{Z_n}{\sqrt{t_n - t_{n+1}}} > x) &= \frac{1}{\sqrt{2\pi}} \int_x^{+\infty} e^{-\frac{z^2}{2}} dz \\
&\geq \frac{x}{x^2+1} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \\
&> \frac{1}{2x\sqrt{2\pi}} e^{-\frac{x^2}{2}},
\end{aligned}$$

as for $x > 1$,

$$1 > \frac{1}{x^2} \iff \frac{x}{x^2+1} > \frac{1}{2x}.$$

Now, let $t_n = q^n$, where q is any number between 0 and 1 such that

$$\beta = \frac{2(1-\varepsilon)^2}{1-q} < 2$$

and let $\alpha = \log\frac{1}{q}$. Put

$$\begin{aligned}
x_n = (1-\varepsilon)\frac{\phi(t_n)}{\sqrt{t_n - t_{n+1}}} &= \frac{1-\varepsilon}{\sqrt{1-q}} \sqrt{2\log(n\log\frac{1}{q})} \\
&= \sqrt{\frac{2(1-\varepsilon)^2}{1-q}\log(n\log\frac{1}{q})} = \sqrt{\beta\log(\alpha n)}
\end{aligned}$$

since $\alpha > 0$ for any q, there exists a natural number k such that $x_n > 1$ for every $n \geq k$. Hence fix such a k and plug in x_n in the above calculations to obtain the inequality

$$\mathbb{P}(Z_n > (1-\varepsilon)\phi(t_n)) = \mathbb{P}(G_n) \geq \frac{c}{n^{\frac{\beta}{2}}\sqrt{\log n}}.$$

Which means that

$$\sum_{n=k}^{\infty} \mathbb{P}(G_n) \geq \sum_{n=k}^{\infty} \frac{c}{n^{\frac{\beta}{2}}\sqrt{\log n}} = \infty,$$

so by the second Borel-Cantelli lemma we have that $\mathbb{P}(\overline{\lim}_n G_n) = 1$. Fix $\delta > 0$ recall that we wish to show that

$$\mathbb{P}(\overline{\lim_n} F_n) = \mathbb{P}(\overline{\lim_n}\{\omega \in \Omega \mid X_{t_n}(\omega) > (1-\delta)\phi(t_n)\}) = 1.$$

Furhter, recall the set A_n from above and consider the Brownian motion induced by $\{-X_t\}_{t\in T}$, then we have that for any $\xi > 0$

$$\mathbb{P}(\overline{\lim_n}\{-X_{t_{n+1}} > (1+\xi)\phi(t_{n+1})\}) = 0$$

which is equivalent to saying that

$$\mathbb{P}(\overline{\lim_n}\{-X_{t_{n+1}} \leq (1+\xi)\phi(t_{n+1})\}) = 1.$$

Since for any $\xi > 0$, $\mathbb{P}(\overline{\lim}_n G_n) = 1$, we get that

$$\begin{aligned}
1 &= \mathbb{P}(\overline{\lim_n}(\{X_{t_n} - X_{t_{n+1}} > (1-\xi)\phi(t_n)\} \cup \{-X_{t_{n+1}} \leq (1+\xi)\phi(t_{n+1})\})) \\
&\leq \mathbb{P}(\overline{\lim_n}\{X_{t_n} + (1+\xi)\phi(t_{n+1}) > (1-\xi)\phi(t_n)\}) \\
&= \mathbb{P}(\overline{\lim_n}\{X_{t_n} > \phi(t_n)((1-\xi-(1+\xi)\frac{\phi(t_{n+1})}{\phi(t_n)})\}).
\end{aligned}$$

By Remark 5.2.1, we may pick the sequence t_n again as $t_n = q^n$ for any $0 < q < 1$. Since $\log\log\frac{1}{q^n} = \log n + \log\log\frac{1}{q}$ and $\lim_n \frac{\log(n+1)}{\log n} = 1$, we obtain

$$\lim_n \frac{\phi(t_{n+1})}{\phi(t_n)} = \lim_n \frac{(2q^{n+1}\log\log\frac{1}{q^{n+1}})^{\frac{1}{2}}}{(2q^n\log\log\frac{1}{q^n})^{\frac{1}{2}}} = \sqrt{q},$$

so for every $\rho > 0$ there exists a natural number N such that for $n > N$,

$$\phi(t_n)(1-\xi-(1+\xi)\frac{\phi(t_{n+1})}{\phi(t_n)}) < \phi(t_n)((1-\xi-(1+\xi)(\sqrt{q}-\rho))$$

and thus we obtain the bound

$$\begin{aligned}
1 &= \mathbb{P}(\overline{\lim_n}(\{X_{t_n} > \phi(t_n)(1-\xi-(1+\xi)\frac{\phi(t_{n+1})}{\phi(t_n)})\}) \\
&\leq \mathbb{P}(\overline{\lim_n}(\{X_{t_n} > \phi(t_n)((1-\xi-(1+\xi)(\sqrt{q}-\rho))\}) \\
&= \mathbb{P}(\overline{\lim_n}(\{X_{t_n} > \phi(t_n)(1-(\xi+\sqrt{q}+\xi\sqrt{q}-\rho-\xi\rho)\}).
\end{aligned}$$

Lastly, we may pick $\rho = \xi = \delta$ and $q = (\delta - c)^2$, for some $c > 0$ such that $q \in (0,1)$ since then

$$\begin{aligned}
1-(\xi+\sqrt{q}+\xi\sqrt{q}-\rho-\xi\rho) &= 1-(\delta-c+\delta(\delta-c)-\delta^2)) \\
&= (1-(\delta-c+\delta(\delta-c)-\delta^2) \\
&= 1-(\delta-c-c\delta) > 1-\delta.
\end{aligned}$$

This allows us to finally write

$$\begin{aligned}
1 &= \mathbb{P}(\overline{\lim_n}(\{X_{t_n} > \phi(t_n)(1-(\xi+\sqrt{q}+\xi\sqrt{q}-\rho-\xi\rho)\}) \\
&\leq \mathbb{P}(\overline{\lim_n}(\{X_{t_n} > \phi(t_n)(1-\delta\}) = \mathbb{P}(\overline{\lim_n} F_n)
\end{aligned}$$

which proves the second inequality and subsequently the whole theorem. □

Bibliography

[1] Paolo Baldi, *Stochastic calculus*, Universitext, Springer, Cham, 2017. An introduction through theory and exercises.

[2] Avner Friedman, *Foundations of modern analysis*, Dover Publications, Inc., New York, 1982. Reprint of the 1970 original.

[3] Peter Mörters & Yuval Peres, *Brownian Motion*, Cambridge University Press, 2010.

[4] Lothar Partzsch & René L. Schilling, *Brownian Motion*, De Gruyter, 2010. An Introduction to Stochastic Processes.

[5] Lenya Ryzhik, *The Haar functions and the Brownian motion*, Stanford University. Lecture notes.

[6] Marta Sanz-Solé, *An Introduction to Stochastic Calculus*, University of Barcelona, 2017. Lecture notes for a PHD-course.