JYVÄSKYLÄN YLIOPISTO
MATEMATIIKAN JA TILASTO-
TIETEEN LAITOS

MASTER'S THESIS

On Exact Simulations of First Hitting Times of the Solutions of SDEs

Irfan Durmić

July 30, 2024

Author
Irfan Durmić

Title
On Exact Simulations of First Hitting Times of the Solutions of SDEs

Degree Program
International Master's Degree Programme in Mathematics

Date	**Page count**
July 30, 2024	51

Abstract

In this thesis, we look into exact simulations of first hitting times of the solutions to stochastic differential equations with unit diffusion coefficient. While discretization schemes do represent essential tools in SDE simulations, their inherent errors prove them to be unsuitable for fields where precision is necessary. Through our exploration of the work done in *Exact Simulation of the First-Passage Time of Diffusions (2019)* by S. Herrmann and C. Zucca, we show that it is indeed possible to construct an algorithm which exactly simulates the first hitting time of the solution of a stochastic differential equation. Our key tools are Girsanov's theorem, which allows us to shift the measure under which we are looking at the problem to put it into a frame which is more suitable for simulations, as well as the acceptance-rejection scheme which was originally proposed in *Exact Simulation of Diffusions (2005)* by A. Beskos and G. O. Roberts.

Contents

Chapter 1

Table of Notation

We provide the reader with a list of some of the most common symbols and abbreviations we use in this thesis.

- SDE - Stochastic Differential Equation (Definition 3.6.1)
- $B = (B_t)$ - (standard) Brownian motion (Definition 3.5.13)
- $(\Omega, \mathcal{F}, \mathbb{P})$ - probability space with sample space Ω, σ-algebra $\mathcal{F}$ and probability measure $\mathbb{P}$ (Definition 3.1.5)
- $\mathbb{E}_{\mathbb{P}}$ - expected value (expectation) with respect to a (probability) measure $\mathbb{P}$ (Definition 3.4.9)
- $\mathbb{1}_A$ - indicator random variable on a set A (Definition 3.1.2)
- $\mathcal{B}(E)$ - Borel σ-algebra on a topological space E (Definition 3.1.6)
- $\mathbb{R} \coloneqq (-\infty, \infty)$ - the set of real numbers
- $\overline{\mathbb{R}} \coloneqq \mathbb{R} \cup \{\pm\infty\}$ - the extended reals
- $\mathbb{N} = \{1, 2, 3, \ldots\}$ - the set of natural numbers
- $\mathbb{Q}$ - a probability measure associated with the Brownian motion (see for example Theorem 3.6.6)
- λ - the Lebesgue measure (Example 3.1.8).
- $\mathcal{L}^p$ space and $\|\cdot\|_p$ associated norm (Definition 3.4.7)
- $\mathbb{E}[X|\cdot]$ - conditional expectation with respect to some object (refer to Section 3.4 and Remark 3.4.17)
- τ_L - first hitting time (Definition 4.2.1)
- $R = (R_t)$ - a three-dimensional Bessel process/bridge (see for example Proposition 4.3.1)
- $\mathcal{L}_2^{\text{loc}}$ - see Definition 3.5.18

- In the context of SDEs σ is the diffusion coefficent and b the drift coefficient (Definition 3.6.1)
- $\mathbb{R}_+ = [0, +\infty)$
- a.s. $=$ almost surely (Definition 3.4.15)

Chapter 2

Introduction

Randomness is all around us: from deciding whether or not to take an umbrella with us to work to deciding whether or not to buy shares of a particular company. Many processes in our world can be described using *stochastic differential equations*, shorthand being SDEs, which we explore in more detail in Section 3.6. For now, we can think about them as tools that describe the evolution of random processes. There are many different applications of SDEs: for an overview, one can look at [KaSh2, So, St] or the many references provided in [HeZu1].

Given their prevalence, it is meaningful to find ways to solve stochastic differential equations, but only special classes of them have actual closed-form solutions. This is why it is valuable to find ways to solve them via simulations or, more accurately, to find ways to simulate their sample paths. Extending these different methods allows us to not only explore the different processes whose behavior is governed by SDEs, but to also explore associated objects such as stopping times and exit times.

In this thesis, we follow the work of [HeZu1] and clarify many results related to exact simulations of the first hitting times of the solutions of SDEs. While the topic of first hitting times of solutions of SDEs is interesting in its own right, it is of particular value in fields such as finance where it enables precise determination as to when a stock has reached a particular value.

We begin our work by going through some of the main concepts while providing explicit theoretical details in Chapter 3. In the simplest of words, a first hitting time (see Definition 4.2.1) is the first time the solution of an SDE, which we for the moment denote as X, reaches a particular level $L > 0$. When it is impossible to determine the first hitting time directly, we resort to *simulations*. In the context of a process X, a simulation represents any sort of numerical method through which "the computer" produces $X(\omega)$ for a fixed $\omega \in \Omega$. This can be done either using exact methods or via some approximate method. One of the most common approximation methods is the *Euler scheme* (often also called the *Euler-Maruyama approximation scheme*). Using [KlPl, Chapter 9], we give a brief overview of this method.
Let $X = (X_t)_{t\in[t_0,T]}$ be the solution of the SDE

$$\begin{aligned} \mathrm{d}X_t &= b(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}B_t \\ X_{t_0} &= x, \end{aligned}$$

where x represents the initial value and $(B_t)_{t\geq 0}$ denotes the Brownian motion (see Definition 3.5.13). We now perform a time discretization of the interval $[t_0, T]$, i.e. we break the interval up into the following time chunks:

$$t_0 = \theta_0 < \theta_1 < \ldots < \theta_n < \ldots < \theta_N = T.$$

Using these time steps gives us the continuous time stochastic process $Y = (Y_t)_{t\in[t_0,T]}$ which is defined via the following recursion:

$$Y_t \;=\; Y_n + b(\theta_n, Y_n)(t-\theta_n) + \sigma(\theta_n, Y_n)(B_t - B_{\theta_n}),\ \text{for } \theta_n \le t \le \theta_{n+1}$$

where $n = 0, 1, 2, \ldots, N-1$ and the initial value is $Y_0 = x$. Finally, note that we are using the convention that $Y_{n+1} = Y_{\theta_{n+1}}$. From a computational standpoint, this is a relatively simple procedure, but it has some drawbacks. As is beautifully illustrated in [KlPl, Sections 9.2-9.3] the method intrinsically comes with an approximation error which, according to [HeZu1, Page 1478], leads to an overestimation of the stopping time and thus seems unsuitable for estimating first hitting times. Another major issue with this method is the fact it is restricted to finite time intervals: a restriction which does not necessarily hold for first hitting times.

On the other hand, *exact simulations* do not have such limitations. In fact, this is where the name *exact* comes from: these types of simulations result in sample paths that do not have approximation errors stemming from time discretization. We present the basic idea using [SäSo, Section 8.8] and [BeRo]. One can apply this method to SDEs of the form

$$\mathrm{d}X_t \;=\; b(X_t)\mathrm{d}t + \mathrm{d}B_t, \tag{2.0.1}$$

where $b(X_t)$ can be expressed as a gradient of a scalar function such that $b(X_t) = \boldsymbol{\nabla}\Psi(X_t)$ and $(B_t)_{t\ge 0}$ is the Brownian motion as before. We note that, using the Lamperti transform (see Remark 3.6.10), we can generalize this method to SDEs with diffusion terms that are not equal to one. The key ingredient in exact simulations is *rejection sampling*. We let f, g be two density functions (see Definition 3.2.15) with respect to some measure (see Definition 3.1.4) on $\mathbb{R}^d$ and we assume there exists some $\epsilon > 0$ such that $\epsilon\frac{f}{g} \le 1$. Thus, rejection sampling works as follows:

1. We simulate a random variable Y with density g
2. We simulate a random variable U which follows the Uniform$(0,1)$ distribution
3. If $U \;<\; \epsilon\frac{f}{g}(Y)$, we accept Y. Otherwise, go back to Step 1.

For a more formal overview of rejection sampling, we refer the reader to [BeRo, Page 3]. The key insight that allows us to use rejection sampling is *Girsanov's theorem* (see Theorem 3.6.6) which allows us to view the solution of an SDE under a probability measure which reduces it to the Brownian motion. The reason why this is helpful is because of the fact that the Brownian motion, and some associated objects, are relatively easy to simulate as we show in Section 4.2. Furthermore, for a direct example of how exact simulation outperforms the Euler scheme, check out [BeRo, Page 19].

The work we explored in [HeZu1] relies on the ideas from exact simulation, but with one twist: since exact simulation are computationally rather heavy when simulating the entire path, one can choose to simulate just the skeleton of the path. In cases when one only cares about the first hitting time, the computational complexity is substantially reduced and the resulting algorithms outperform those by [BeRo]. Our main object of interest is the initial algorithm proposed in [HeZu1, Section 2.2] which we formally define in Section 4.1. The validity of this initial algorithm is justified through a detailed exploration of the key results that are formally proved in Chapter 4. For now, we state the key results of this thesis along with the most essential

prerequisites needed to understand them:
We begin with a stochastic process $X = (X_t)_{t\geq 0}$ which solves the SDE:

$$\begin{aligned} \mathrm{d}X_t &= b(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}B_t \\ X_0 &= x \in \mathbb{R}. \end{aligned} \tag{2.0.2}$$

For technical reasons, we introduce the following set of objects:

$$\beta(x) = \int_0^x b(y)\mathrm{d}y, \;\; p(x) = \int_0^x \exp\left(-\beta(y)\right)\mathrm{d}y, \;\; \text{and } \gamma(x) := \frac{b^2(x) + b'(x)}{2}. \tag{2.0.3}$$

Assumption 2.0.1 (Drift term and scale function assumptions)**.** *The drift term $b \in \mathcal{C}^1((-\infty, L])$ and $\lim_{x\to-\infty} p(x) = -\infty$.*

Definition 2.0.2 (First hitting time)**.** *We denote with τ_L the first time the process X reaches a level $L > 0$. Formally, we write:*

$$\tau_L := \inf\{t \geq 0 \colon X_t = L\}. \tag{2.0.4}$$

Proposition 2.0.3. *([HeZu1, Proposition 2.1]) Under Assumption 2.0.1, for any bounded and measurable function $\psi \colon \mathbb{R} \to \mathbb{R}$, we obtain*

$$\mathbb{E}_{\mathbb{P}}[\psi(\tau_L)\mathbb{1}_{\{\tau_L<\infty\}}] = \mathbb{E}_{\mathbb{Q}}[\psi(\tau_L)\eta(\tau_L)]\exp\left(\beta(L) - \beta(x)\right), \tag{2.0.5}$$

where $\mathbb{P}$ (respectively $\mathbb{Q}$) corresponds to the distribution of X, which was defined by Equation (2.0.2), (respectively the Brownian motion B) and

$$\eta(t) := \mathbb{E}\left[\exp\left(-\int_0^t \gamma(L - R_s)\mathrm{d}s\right)\Big| R_t = L - x\right]. \tag{2.0.6}$$

Here, $(R_t)_{t\geq 0}$ stands for a 3-dimensional Bessel process with $R_0 = 0$.

Assumption 2.0.4 ($\gamma \geq 0$ and $\tau_L < \infty$ a.s.)**.** *We assume that γ, defined in Equation (2.0.3), is non-negative and the first-hitting time τ_L is almost surely finite.*

THE ALGORITHM

Let us fix the level $L > x$.

Step 1: *Simulate a non-negative random variable T with density f_T.*

Step 2: *Simulate a 3-dimensional Bessel process (R_t) on the time interval $[0, T]$ with endpoint $R_T = L - x$. We define by $D_{R,T}$ the stochastic domain:*

$$D_{R,T} := \Big\{(t, \nu) \in [0, T] \times \mathbb{R}_+ : \; \nu \leq \gamma(L - R_t)\Big\}.$$

This domain depends on both random elements $(R_t)_{t\in[0,T]}$ and T.

Step 3: *Simulate a Poisson point process N, with Lebesgue intensity measure, on the state space $[0, T] \times [0, \infty)$, independent of the Bessel process.*

Step 4: *If $N(D_{R,T}) = 0$ then set $Y = T$. Otherwise go to Step 1.*

Outcome: *the random variable Y.*

Theorem 2.0.5. *([HeZu1, Theorem 2.3]) If γ is a non-negative function then the density f_Y of the outcome variable Y, as described in* THE ALGORITHM *above (or in Section 4.1), satisfies*

$$f_Y(t) \;=\; \frac{1}{\Xi}\eta(t)f_T(t), \tag{2.0.7}$$

where η is given by Equation (2.0.6) and Ξ stands for the normalization coefficient

$$\Xi \;\coloneqq\; \int_0^\infty \eta(t)f_T(t)\mathrm{d}t. \tag{2.0.8}$$

In particular, under Assumptions 2.0.1 and 2.0.4, if $\frac{T}{(L-x)^2}$ has the same distribution as $\frac{1}{G^2}$, where G is a standard Gaussian random variable, then Y and τ_L, defined as the first hitting time in Definition 2.0.2 are identically distributed.

What follows now, in Chapter 3, is first an introduction to all the tools from stochastics we used in this thesis. From there, we move on to presenting the mathematical framework as well as the proofs of the key results, and some of the limitations of the associated algorithm, in Chapter 4, which we wrap up with a short overview of recent results and further avenues of research. For the reader's convenience, we've also included a table of notation in Chapter 1, some key results from analysis in Appendix A, and, finally, Appendix B, gives the main distributions that were referenced in this work.

Chapter 3

Theoretical Minimum of Stochastics

We first present the necessary theoretical background following the pedagogical approach laid out by Stefan and Hannah Geiss in their Lecture Notes from the University of Jyväskylä [Ge1, Ge2]. A familiarity with basic analysis is assumed: for a quick overview of the essential results, as well as a list of appropriate resources, we refer the reader to [Ga].

3.1 Basic Results in Probability

The following is an overview of the building blocks of measure-theoretic probability theory. This section is based on [Shi, Chapter 2], [Wi1, Chapters 1-5], [Çı, Chapter 1-2], [Bi1, Chapter 2] and [Kl, Chapter 1]. At times, some inspiration in the explanations is also taken from [Mi]. For a more thorough overview of measure theory, refer to [RoFi, Ha].

Definition 3.1.1 (σ-algebra)**.** *Let Ω be a set and $\mathcal{F}$ be a non-empty set of subsets of Ω. The class of sets $\mathcal{F} \subset 2^\Omega$ is thus called a* σ-algebra *(or a σ-field) if it fulfills the following three conditions:*

(i) $\Omega \in \mathcal{F}$

(ii) $\mathcal{F}$ *is closed under complements*

(iii) $\mathcal{F}$ *is closed under countable unions*

Definition 3.1.2 (Indicator random variable)**.** *The* indicator random variable *(indicator function) on the set A is defined as*

$$\mathbb{1}_{(A)}(x) \coloneqq \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases}$$

Definition 3.1.3 (Measurable and Measure spaces)**.** *A pair $(\Omega, \mathcal{F})$ consisting of a nonempty set Ω and a σ-algebra $\mathcal{F} \subset 2^\Omega$ is called a* measurable space*. The sets $A \in \mathcal{F}$ are called measurable sets. The triplet $(\Omega, \mathcal{F}, \mu)$ is called a* measure space *if $\mu \colon \mathcal{F} \to [0, \infty)$ is a measure.*

Definition 3.1.4 (Measure)**.** *Let $(\Omega, \mathcal{F})$ be a measurable space. The mapping $\mu \colon \mathcal{F} \to [0, \infty)$ is called a* measure *if the following conditions hold:*

1. $\mu(\emptyset) = 0$

2. *for all (pairwise) disjoint sets* $A_1, A_2, \dots \in \mathcal{F}$ *it holds that:*

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

This property is known as σ-additivity.

Definition 3.1.5 (Probability Space and Probability Measure)**.** *Let* $(\Omega, \mathcal{F}, \mu)$ *be a measure space and assume that* $\mu(\Omega) = 1$*. The triplet* $(\Omega, \mathcal{F}, \mu)$ *is then called a* probability space *with the standard notation being* $(\Omega, \mathcal{F}, \mathbb{P})$*, where* $\mathbb{P}\colon \mathcal{F} \to [0,1]$ *is a* probability measure*.*

Definition 3.1.6 (Borel σ-algebra)**.** *Let* E *be a topological space. Then, the smallest* σ*-algebra containing all open subsets of* E *is called the* Borel σ-algebra *on* E*, which we shall denote as* $\mathcal{B}(E)$*. Sets of a Borel* σ*-algebra are called* Borel sets *which, in the context of a measurable space* $(E, \mathcal{B}(E))$ *are measurable sets.*

Remark 3.1.7 (Generating the Borel σ-algebra)**.** *There are many classes of sets that can be used to generate the Borel* σ*-algebra* $\mathcal{B}(R^n)$*. For a thorough discussion, check [Kl, Theorem* 1.23*].*

Example 3.1.8 (Lebesgue measure)**.** *Let* $\mu := \lambda$ *be a measure on* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$*. This measure is called the* Lebesgue measure *on* $\mathbb{R}$ *if* $\lambda(A)$ *is equal to the length of* A *for every interval* A*. For a deeper introduction to the Lebesgue measure, check out [Ha, Section 15].*

Definition 3.1.9. *(Product* σ*-algebra [Ge1, Definition 5.3.1]) Let* $I := \{1, \dots, d\}$ *or* $I := \{1, 2, \dots\}$*, i.e. we consider finite or countable product spaces at the same time.*
Assume measurable spaces $(\Omega_i, \mathcal{F}_i)$ *for* $i \in I$*, where* I *is a non-empty index set, and let*

$$\Omega := \bigtimes_{i \in I} \Omega_i = \{(\omega_i)_{i \in I} : \omega_i \in \Omega_i \text{ for } i \in I\}.$$

Then, the product σ-algebra $\bigotimes_{i \in I} \mathcal{F}_i$ *on* Ω *is generated by all sets*

$$\bigtimes_{i \in I} E_i := \{(\omega_i)_{i \in I} \in \Omega : \omega_i \in E_i\}, \text{ where } E_i \in \mathcal{F}_i$$

and where only finitely many E_i *do not coincide with* Ω_i*.*

Theorem 3.1.10. *(Products of probability spaces [Ge1, Theorem 5.3.2]) For* $I := \{1, \dots, d\}$ *or* $I := \mathbb{N}$ *assume probability spaces* $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$*,* $i \in I$*. Then, there is a unique probability measure* $\mathbb{P} = \bigotimes_{i \in I} \mathbb{P}_i$ *on* $(\Omega, \mathcal{F}) := \{\bigtimes_{i \in I} \Omega_i, \bigotimes_{i \in I} \mathcal{F}_i\}$ *such that*

$$\mathbb{P}\left(\bigtimes_{i \in I} E_i\right) = \prod_{i \in I} \mathbb{P}_i(E_i),$$

where only finitely many of the $E_i \in \mathcal{F}_i$ *differ from* Ω_i*.*
The probability space $(\Omega, \mathcal{F}, \mathbb{P})$ *is called* product of the probability spaces $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$ *and is denoted by* $\bigotimes_{i \in I} (\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$*.*

3.2 Random Variables and Distributions

We now give an overview of the essential terminology related to random variables, measurable maps, image measures and distributions. For a more thorough discussion, we refer the reader to [Kl, Sections 1.4-1.5], [GrSt, Chapter 2], [Mi, Chapter 8] which were all used in writing this section.

Definition 3.2.1 (Measurable Maps)**.** *Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ be two measurable spaces.*

(i) A map $X\colon \Omega_1 \to \Omega_2$ is said to be $\mathcal{F}_1 - \mathcal{F}_2$- measurable (or, just, measurable*) if*

$$X^{-1}(\mathcal{F}_2) \;:=\; \{X^{-1}(A_2)\colon A_2 \in \mathcal{F}_2\} \subset \mathcal{F}_1.$$

This is equivalent to: $X^{-1}(A_2) \in \mathcal{F}_1$ for any $A_2 \in \mathcal{F}_2$. If X is measurable, we write $X\colon (\Omega_1, \mathcal{F}_1) \to (\Omega_2, \mathcal{F}_2)$.

(ii) If $\Omega_2 = \mathbb{R}$ and $\mathcal{F}_2 = \mathcal{B}(\mathbb{R})$ is the Borel σ-algebra on $\mathbb{R}$, then $X\colon (\Omega_1, \mathcal{F}_1) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is called an $\mathcal{F}_1$-measurable real map.

Remark 3.2.2 (Connecting measuable maps and measures)**.** *The key reason why measurable maps play such a key role in stochastics is because they "transport" a measure from one space to another.*

Theorem 3.2.3. *(Generated σ-algebra [Kl, Theorem 1.78]) Let $(\Omega_2, \mathcal{F}_2)$ be a measurable space and $\Omega_1 \neq \emptyset$. Let $X\colon \Omega_1 \to \Omega_2$ be a map. The* preimage *$X^{-1}(\mathcal{F}_2) \;:=\; \{X^{-1}(A_2)\colon A_2 \in \mathcal{F}_2\}$ is the smallest σ-algebra with respect to which X is measurable. We say that $\sigma(X) \;:=\; X^{-1}(\mathcal{F}_2)$ is the σ-algebra on Ω_1 that is generated by X.*

Remark 3.2.4. *(Generated σ-algebra [Kl, Definition 1.79.]) Let Ω be a nonempty set. Let I be an arbitrary index set. For any $i \in I$, let $(\Omega_i, \mathcal{F}_i)$ be a measurable space and let $X_i\colon \Omega \to \Omega_i$ be an arbitrary map. Then*

$$\sigma(X_i, i \in I) \;:=\; \sigma\left(\bigcup_{i\in I} \sigma(X_i)\right) \;=\; \sigma\left(\bigcup_{i\in I} X_i^{-1}(\mathcal{F}_i)\right)$$

is called the σ-algebra on Ω that is generated by $(X_i, i \in I)$. This is the smallest σ-algebra with respect to which all X_i are measurable: If there is another σ-algebra $\mathcal{G}$ with the property that all X_i are $\mathcal{G}$-measurable, then $\sigma(X_i, i \in I) \subseteq \mathcal{G}$.

Theorem 3.2.5. *(Measurability of Continuous Maps [Kl, Theorem 1.88]) Let (Ω_1, τ_1) and (Ω_2, τ_2) be topological spaces and let $X\colon \Omega_1 \to \Omega_2$ be a continuous map. Then X is $\mathcal{B}(\Omega_1) - \mathcal{B}(\Omega_2)$-measurable.*

Theorem 3.2.6. *(Coordinate Maps are Measurable [Kl, Theorem 1.90]) Let $(\Omega, \mathcal{F})$ be a measurable space and let $X_1, \ldots, X_n\colon \Omega \to \mathbb{R}$ be maps. Define $X \;:=\; (X_1, \ldots, X_n)\colon \Omega \to \mathbb{R}^n$. Then*

$$X \text{ is } \mathcal{F}\text{–}\mathcal{B}(\mathbb{R}^n)\text{-measurable} \iff \text{each } X_i \text{ is } \mathcal{F}\text{–}\mathcal{B}(\mathbb{R})\text{-measurable}.$$

The analogous statement holds for $X_i\colon \Omega \to \overline{\mathbb{R}} \;:=\; \mathbb{R} \cup \{\pm\infty\}$.

Theorem 3.2.7. *([Kl, Theorem 1.91]) Let $(\Omega, \mathcal{F})$ be a measurable space. Let $X\colon (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and $X, Y\colon (\Omega, \mathcal{F}) \to (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ be measurable maps. Then also the maps $X + Y$, $X - Y$, $X \cdot Y$ and $\frac{X}{Y}$ are measurable.*

Remark 3.2.8 (Note about Theorem 3.2.7)**.** *Note that in Theorem 3.2.7, we set $\frac{x}{0} := 0 \,\forall\, x \in \mathbb{R}$.*

Definition 3.2.9 (Simple Function)**.** *Let $(\Omega, \mathcal{F})$ be a measurable space. A map $X\colon \Omega \to \mathbb{R}$ is called a* simple function *if there is an $n \in \mathbb{N}$ and mutually disjoint measurable sets $A_1, \dots, A_n \in \mathcal{F}$, as well as reals $\alpha_1, \dots, \alpha_n \in \mathbb{R}$, such that*

$$X = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i},$$

where $\mathbb{1}_{A_i}$ is the indicator random variable on the set A_i.

Definition 3.2.10 (Image Measure (Push Forward Measure))**.** *Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ be measurable spaces and let μ_1 be a measure on $(\Omega_1, \mathcal{F}_1)$. Further, let $X\colon (\Omega_1, \mathcal{F}_1) \to (\Omega_2, \mathcal{F}_2)$ be measurable. The image measure of μ_1 under the map X is the measure $\mu \circ X^{-1}$ on $(\Omega_2, \mathcal{F}_2)$ that is defined by*

$$\mu \circ X^{-1}\colon \mathcal{F}_2 \to [0, \infty], \quad A_2 \mapsto \mu(X^{-1}(A_2)),$$

or, written differently,

$$\mu_2(A_2) := \mu_1(X^{-1}(A_2)),$$

for $A_2 \in \mathcal{F}_2$. It thus follows that $(\Omega_2, \mathcal{F}_2, \mu_2)$ is a measure space and $\mu_1(\Omega_1) = \mu_2(\Omega_2)$.

Definition 3.2.11 (Random Variables)**.** *Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ be measurable spaces and let $X\colon \Omega_1 \to \Omega_2$ be measurable. Furthermore, let $\mathbb{P}$ be a probability measure on the measurable space $(\Omega_1, \mathcal{F}_1)$, i.e. $\mathbb{P}(\Omega_1) = 1$.*

(i) The measurable map X is called a random variable with values in $(\Omega_2, \mathcal{F}_2)$*. If $(\Omega_2, \mathcal{F}_2) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, then X is called a real random variable or simply a* random variable*. Throughout this thesis, when we say* random variable*, we are talking about real random variables, unless otherwise specified (or if it is clear from the context).*

(ii) For $A_2 \in \mathcal{F}_2$, we denote $\{X \in A_2\} := X^{-1}(A_2)$ and $\mathbb{P}(X \in A_2) := \mathbb{P}(X^{-1}(A_2))$. In particular, we let $\{X \geq 0\} := X^{-1}([0, \infty))$ and define $\{X \leq b\}$ similarly and so on.

Definition 3.2.12 (Distributions)**.** *Let X be a random variable.*

(i) The probability measure $\mathbb{P}_X := \mathbb{P} \circ X^{-1}$ is called the distribution *of X.*

(ii) For a real random variable X, the map $F_X\colon x \mapsto \mathbb{P}(X \leq x)$ is called the distribution function *of X (or, more accurately, of $\mathbb{P}_X$). This is essentially the familiar* cumulative distribution function *of the random variable X. We write $X \sim \mu$ if $\mu = \mathbb{P}_X$ and say that X has distribution μ.*

(iii) A family $(X_i)_{i \in I}$ of random variables is called identically distributed *if $\mathbb{P}_{X_i} = \mathbb{P}_{X_j}$ for all $i, j \in I$. We write $X \overset{\mathrm{d}}{=} Y$ if $\mathbb{P}_X = \mathbb{P}_Y$ (*d for distribution*).*

Remark 3.2.13 (The law of the random variable X)**.** *In the context of Definition 3.2.10, in case μ_1 is a probability measure, the measure μ_2 is called the* law *(or distribution) of the random variable X.*

Definition 3.2.14 (Borel function)**.** *A mapping $X\colon \mathbb{R}^i \to \mathbb{R}^k$ is defined to be measurable if it is $\mathcal{B}(\mathbb{R}^i) - \mathcal{B}(\mathbb{R}^k)$-measurable. Such mappings are often called* Borel functions*.*

Definition 3.2.15. *(Density of a random variable and its distribution) A random variable and its distribution have* density *f with respect to the Lebesgue measure if f is a non-negative Borel function on $\mathbb{R}^1$ and*

$$\mathbb{P}(X \in A) \;=\; \mu(A) \;=\; \int_A f(x)\mathrm{d}\lambda(x), \quad A \in \mathcal{B}(\mathbb{R}^1). \tag{3.2.1}$$

In other words, the requirement is that μ has density f with respect to the Lebesgue measure λ (recall Example 3.1.8). If we have not specified the measure, then we assume that the density is taken with respect to the Lebesgue measure. Observe that, plugging in $\mathbb{R}^1$ for A into Equation (3.2.1), gives us the familiar result that the density integrates to 1 over all space.

Remark 3.2.16 (On notation of measurable maps)**.** *Throughout this section, we denoted all measurable maps with capital letters: most often X. From now on, capital letters are reserved for random variables, unless specified otherwise.*

3.3 Independence

The concept of independence is a key idea in stochastics in general and we use it many times throughout this thesis. Below, we quote some of the essential ideas as given in [Kl, Chapter 2]

Definition 3.3.1 (Independence of Events)**.** *Let $I \neq \emptyset$ be an arbitrary index set and let $(A_i)_{i\in I}$ be an arbitrary family of events. The family $(A_i)_{i\in I}$ is called* independent *if for any finite subset $J \subset I$ the following holds:*

$$\mathbb{P}\left(\bigcap_{j\in J} A_j\right) \;=\; \prod_{j\in J} \mathbb{P}(A_j).$$

Definition 3.3.2 (Independence of Classes of Events)**.** *Let $I \neq \emptyset$ be an arbitrary index set and let $\mathcal{A}_i \subset \mathcal{F}$ be an arbitrary family of classes of events, such as σ-algebras, for all $i \in I$. The family $(\mathcal{A}_i)_{i\in I}$ is called* independent *if for any finite subset $J \subset I$ and any choice of $A_j \in \mathcal{A}_j, j \in J$ the following holds:*

$$\mathbb{P}\left(\bigcap_{j\in J} A_j\right) \;=\; \prod_{j\in J} \mathbb{P}(A_j).$$

Definition 3.3.3 (Independent random variables)**.** *Let $I \neq \emptyset$ be an arbitrary index set. For each $i \in I$, we let $(\Omega_i, \mathcal{F}_i)$ be a measurable space and let $X_i\colon (\Omega, \mathcal{F}) \to (\Omega_i, \mathcal{F}_i)$ be a random variable with generated σ-algebra $\sigma(X_i) = X^{-1}(\mathcal{F}_i)$. Now, the family $(X_i)_{i\in I}$ of random variables is said to be* independent *if the family $(\sigma(X_i))_{i\in I}$ of σ-algebras is independent.*

Remark 3.3.4 (Indicator random variables on independent events)**.** *Let $(A_j)_{j\in J} \subseteq \mathcal{F}$, where J is an arbitrary finite index set, be a collection of independent events. Then, the indicator random variables on those events, i.e. $(\mathbb{1}_{A_j})_{j\in J}$, are also independent. Actually, one could even show that this is an "if and only if" statement; refer to [Kl, Chapter 2] for details.*

3.4 The Lebesgue Integral, Expectation and Conditional Expectation

We now present the Lebesgue integral, some results associated with it, as well as its connection to the concept of expectation as can be found in [Bi1, Chapter 3-4]. We then extend our conversation to the concept of conditional expectation, following closely [Wi1, Chapter 9], from which we naturally transition into a discussion of conditional probability as presented in [Kl, Chapter 8].

Unless it is otherwise specified, throughout this section, we assume that we are operating on some measure space $(\Omega, \mathcal{F}, \mu)$, that X, Y and so on are measurable maps and that $\pm\infty$ is allowed. We also make a quick remark about calculations involving infinity.

Remark 3.4.1 (Calculations with infinity)**.** *Let $x \in (0, +\infty)$. We use the following conventions when working with infinity:*

(i) $0 \cdot \infty = \infty \cdot 0 = 0$

(ii) $x \cdot \infty = \infty \cdot x = \infty$

(iii) $\infty \cdot \infty = \infty$.

Now, what is even a Lebesgue integral and why do we care about it? In the simplest of terms, a Lebesgue integral is any integral with respect to a general measure. In our work, we will mostly be dealing with Lebesgue integrals evaluated with respect to the Lebesgue measure. Actually, this is rather common and it is no surprise that some authors specifically denote the Lebesgue integral as the integral with respect to the Lebesgue measure. You can find more details about the terminology related to the Lebesgue integral, as well as the Lebesgue measure, in [RoFi, Chapter 3-4].
As far as the importance of the Lebesgue integral is concerned, it offers us the ability to integrate functions that we could not deal with using the Riemann integral: the indicator function on the rationals is a classical example (see [Mo, Section 15.2 and Chapter 18]). We will now present the Lebesgue integral in three steps, following the convention found in [Kl, Chapter 4] and [Ge1, Chapter 6].

Definition 3.4.2 (Step 1: Non-negative simple functions)**.** *Let $Y : \Omega \to \mathbb{R}$ be a non-negative simple function with representation*

$$Y = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i}$$

for $\alpha_i \in [0, \infty)$ and $A_i \in \mathcal{F}$. We thus let

$$\int_\Omega Y \mathrm{d}\mu = \int_\Omega Y(\omega)\mathrm{d}\mu(\omega) := \sum_{i=1}^{n} \alpha_i \mu(A_i),$$

which has a unique representation (see for example [Ge1, Lemma 6.1.2]).

Definition 3.4.3 (Step 2: Non-negative functions)**.** *Let $X\colon \Omega \to \mathbb{R}$ be a measurable map such that $X(\omega) \geq 0$ for all $\omega \in \Omega$. We thus define:*

$$\begin{aligned}\int_\Omega X \mathrm{d}\mu &= \int_\Omega X(\omega)\mathrm{d}\mu(\omega) \\ &:= \sup\left\{\int_\Omega Y \mathrm{d}\mu \colon 0 \leq Y(\omega) \leq X(\omega),\ \textit{where } Y \textit{ is a simple function}\right\}\end{aligned}$$

Let X be some general measurable map. We define its *positive part* as:

$$X^+(\omega) = \begin{cases} X(\omega) & \text{if } 0 < X(\omega) \leq \infty \\ 0 & \text{if } -\infty \leq X(\omega) \leq 0 \end{cases} \tag{3.4.1}$$

and, similarly, the *negative part* as:

$$X^-(\omega) = \begin{cases} -X(\omega) & \text{if } -\infty \leq X(\omega) < 0 \\ 0 & \text{if } 0 \leq X(\omega) \leq \infty \end{cases} \tag{3.4.2}$$

Note that these functions are non-negative, measurable and that we get the following decomposition

$$X = X^+ - X^-,$$

which now allows us to define the Lebesgue integral for the general case.

Definition 3.4.4 (Step 3: The General Case)**.** *Let $X\colon \Omega \to \mathbb{R}$ be a general measurable function.*

(1) Existence
If $\int_\Omega X^+ \mathrm{d}\mu < \infty$ or *$\int_\Omega X^- \mathrm{d}\mu < \infty$, then we say that the Lebesgue integral $\int_\Omega X \mathrm{d}\mu$ exists and we set*

$$\int_\Omega X \mathrm{d}\mu := \int_\Omega X^+ \mathrm{d}\mu - \int_\Omega X^- \mathrm{d}\mu.$$

(2) Integrability
The map X is integrable if

$$\int_\Omega X^+ \mathrm{d}\mu < \infty \textit{ and } \int_\Omega X^- \mathrm{d}\mu < \infty.$$

(3) Integral over a set $A \in \mathcal{F}$
If the Lebesgue-integral $\int_\Omega X \mathrm{d}\mu$ exists *and $A \in \mathcal{F}$, then*

$$\int_A X \mathrm{d}\mu = \int_A X(\omega)\mathrm{d}\mu(\omega) := \int_\Omega X(\omega)\mathbb{1}_A(\omega)\mathrm{d}\mu(\omega).$$

Since integrability is used very often, we isolate Definition 3.4.5 which presents some of the most common ways to denote that X is integrable.

Definition 3.4.5. *(Integrability (Adapted from [Bi1, Page 206])) Let $X\colon \Omega \to \mathbb{R}$ be a general measurable function. We say that X is* integrable *if and only if:*

$$\int_\Omega |X| \mathrm{d}\mu \ < \ \infty,$$

which is equivalent to the requirement

$$\int_\Omega |X| \mathrm{d}\mu \ = \ \int_\Omega (X^+ + X^-) \mathrm{d}\mu \ < \ \infty.$$

Theorem 3.4.6. *(Linearity [Bi1, Theorem 16.1(ii)]) If X and Y are integrable and $\alpha, \beta \in \mathbb{R}$, then $\alpha X + \beta Y$ is integrable and:*

$$\int_\Omega (\alpha X + \beta Y) \mathrm{d}\mu \ = \ \alpha \int_\Omega X \mathrm{d}\mu + \beta \int_\Omega Y \mathrm{d}\mu.$$

Definition 3.4.7 ($\mathcal{L}^p$ space)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $X\colon \Omega \to \overline{\mathbb{R}}$ be a measurable map. We define the $\mathcal{L}^p$-*norm*, for $p \in [1, \infty)$ as:*

$$\|X\|_p \ = \ \left(\int_\Omega |X|^p \mathrm{d}\mu \right)^{\frac{1}{p}}.$$

The spaces on which the $\mathcal{L}^p$-norm is finite is called the $\mathcal{L}^p$ space *with the following definition:*

$$\mathcal{L}^p(\Omega, \mathcal{F}, \mu) \ = \ \mathcal{L}^p(\Omega, \mu) \ = \ \mathcal{L}^p(\mu) \ = \ \{X\colon \Omega \to \overline{\mathbb{R}} \text{ measurable and } \|X\|_p \ < \ \infty\}.$$

Remark 3.4.8 ("Defining" a norm)**.** *In Definition 3.4.7, we said that we defined $\|\cdot\|_p$ to be a norm. In reality, what we did is we defined the $\|\cdot\|_p$ of a particular object in a certain way, but we cannot just will something into being a norm. However, one can show that the object $\|\cdot\|_p$ indeed satisfies the requirements of a norm; check [Fo2, Section 4.1.] for the theoretical details that enable us to show this result.*

It turns out that the expectation is a special case of the Lebesgue integral:

Definition 3.4.9 (Expectation (expected value))**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and X a random variable on this space. Assuming that it exists, the expectation of this random variable is defined as:*

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

Note that all the previous conventions, properties and results related to the Lebesgue integral also apply to the expectation.

Remark 3.4.10 (Integrability in the language of expectation)**.** *Let $X\colon \Omega \to \mathbb{R}$ be a general measurable function. We say that X is* integrable *if and only if:*

$$\mathbb{E}[|X|] \ < \ \infty.$$

Proposition 3.4.11. *(Expectation and Distributions [Adapted from [Bi1, Page 274]]) Let X be a random variable with distribution μ (in the context of Remark 3.2.13). If g is a Borel function of a real variable, then it follows that*

$$\mathbb{E}[g(X)] \;=\; \int_{-\infty}^{\infty} g(x)\mu(\mathrm{d}x).$$

If X has density f, then we can rewrite this as:

$$\mathbb{E}[g(X)] \;=\; \int_{-\infty}^{\infty} g(x)f(x)\mathrm{d}\lambda(x).$$

This result is also often refered to as the change of variable *formula.*

Remark 3.4.12 (Remark on notation for the Lebesgue integral)**.** *Often, when it is clear from the context, we will use $\mathrm{d}x$ to mean $\mathrm{d}\lambda(x)$. This also makes sense considering how often, in our applications at least, the Riemann integral coincides with the Lebesgue integral (refer to [RoFi, Chapter 5, Section 3] for a more rigorous discussion of when this is truly the case).*

Theorem 3.4.13. *(Expectation and independence[Adapted from [Fe, Page 222]]) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and X, Y two independent and integrable random variables. The following two statements are true:*

1. $\mathbb{E}[|XY|] \;<\; \infty$
2. $\mathbb{E}[XY] \;=\; \mathbb{E}[X]\mathbb{E}[Y]$.

Proposition 3.4.14 (Expectation of a product of independent random variables)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with independent, integrable random variables $X_1, X_2, \ldots, X_n$. Then, it follows that:*

$$\mathbb{E}[X_1 \cdot X_2 \cdots X_n] \;=\; \mathbb{E}[X_1] \cdot \mathbb{E}[X_2] \cdots \mathbb{E}[X_n] \tag{3.4.3}$$

Proof of Proposition 3.4.14. Before beginning the proof, we must first check whether or not the object $X_1 \cdot X_2 \cdots X_n$ is even integrable. The easiest way to show this is via induction:

1. Base case $n = 2$
   From the first statement of Theorem 3.4.13, we know that
   $$\mathbb{E}[|X_1 \cdot X_2|] \;<\; \infty$$
2. Inductive hypothesis $n = k$, where $k \;\geq\; 2$ and $k \in \mathbb{N}$
   We assume that the object $X_1 \cdot X_2 \cdots X_k$ is integrable.
3. Inductive step
   We prove that the statement is valid for some $n = k + 1$.
   $$\begin{aligned} \mathbb{E}[|X_1 \cdot X_2 \cdots X_k \cdot X_{k+1}|] \;&\leq\; \mathbb{E}[|X_1 \cdot X_2 \cdots X_k||X_{k+1}|] \cdot \\ &\leq\; \mathbb{E}[|X_1 \cdot X_2 \cdots X_k| \cdot \mathbb{E}[|X_{k+1}|] \;<\; \infty, \end{aligned}$$

which proves integrability. In order to prove Equation (3.4.3), we will use induction and the second part of Theorem 3.4.13.

1. Base case $n = 2$
   From Theorem 3.4.13, we know that the statement of Proposition 3.4.14 holds for two independent random variables.

2. Inductive hypothesis $n = k$, where $k \geq 2$ and $k \in \mathbb{N}$
   We assume that Equation (3.4.3) is valid for some $n = k$ with properties as specified above.

3. Inductive step
   We prove that the statement is valid for some $n = k + 1$.

$$\begin{aligned} \mathbb{E}[X_1 \cdot X_2 \cdots X_k \cdot X_{k+1}] &= \mathbb{E}[X_1 \cdot X_2 \cdots X_k] \cdot \mathbb{E}[X_{k+1}] && (3.4.4) \\ &= \mathbb{E}[X_1] \cdot \mathbb{E}[X_2] \cdots \mathbb{E}[X_k] \cdot \mathbb{E}[X_{k+1}], && (3.4.5) \end{aligned}$$

where, in Equation (3.4.4) we used the fact that the product $X_1 \cdot X_2 \cdots X_k$ is independent of X_{k+1} (refer to Section 3.3 to see that this follows from the definitions) and then applied Theorem 3.4.13. Finally, in going to Equation (3.4.5), we just used the inductive hypothesis which gave us the desired result. □

Definition 3.4.15 (Almost everywhere and almost surely)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, and $\mathcal{P}(\omega)$ some sort of a property that depends on ω. We say that this property holds* almost everywhere (a.e.) *if the set $\{\omega \in \Omega : \mathcal{P} \text{ holds}\}$ is measurable with respect to $\mathcal{F}$ and $\mu(\{\omega \in \Omega : \mathcal{P} \text{ does not hold}\}) = 0$, meaning that the set where $\mathcal{P}$ is not true has measure 0. In case μ is a probability measure, we say that the property holds* almost surely (a.s.) *and often write $\mathbb{P}(\{\omega \in \Omega : \mathcal{P} \text{ holds}\}) = 1$.*

Definition and Lemma 3.4.16. *(Conditional Expectation [Wi1, Section 9.2]) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, X an integrable random variable and $\mathcal{G}$ a sub-σ-algebra of $\mathcal{F}$. Then, there exists a random variable Y such that:*

(i) *Y is $\mathcal{G}$-measurable*

(ii) *$\mathbb{E}(|Y|) < \infty$*

(iii) *for every set $G \in \mathcal{G}$, we have*

$$\int_G Y \, \mathrm{d}\mathbb{P} = \int_G X \, \mathrm{d}\mathbb{P}, \ \forall G \in \mathcal{G} \tag{3.4.6}$$

Furthermore, the conditional expectation is almost surely unique: if $\tilde{Y}$ is another random variable satisfying all of the mentioned properties, then $\tilde{Y} = Y$ a.s. A random variable Y with all of these properties is said to be a version of the conditional expectation $\mathbb{E}[X|\mathcal{G}]$ *of X given $\mathcal{G}$ and we write $Y = \mathbb{E}[X|\mathcal{G}]$. Two versions agree a.s., and, once one has become familiar with the concept, one identifies different versions and speaks of* the conditional expectation $\mathbb{E}[X|\mathcal{G}]$, *but one must always keep the* a.s. *in mind.*

Remark 3.4.17 (Conditional expectation: notation)**.** *It is common for authors to write $\mathbb{E}[X|Z]$ for $\mathbb{E}[X|\sigma(Z)]$, $\mathbb{E}[X_1|Z_1, Z_2, \ldots]$ for $\mathbb{E}[X|\sigma(Z_1, Z_2, \ldots)]$, and so on. Throughout this thesis, we will use both of these notations depending on what we want to emphasise: the former when the focus is on the behavior of a random variable at a particular time step, and the latter when the focus is on the structure of the σ-algebra being generated by the random variables.*

Proposition 3.4.18. *(Properties of conditional expectation [Wi1, Page 88]) Let X, Y be integrable random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{F}$ be sub-σ-algebras of $\mathcal{F}$. Then the following holds true:*

(1) Linearity*:*
If $\mu, \lambda \in \mathbb{R}$, then

$$\mathbb{E}[\lambda X + \mu Y|\mathcal{G}] \;=\; \lambda\mathbb{E}[X|\mathcal{G}] + \mu\mathbb{E}[Y|\mathcal{G}] \text{ a.s.}$$

(2) Monotonicity*:*
If $X \;\leq\; Y$ a.s., then $\mathbb{E}[X|\mathcal{G}] \;\leq\; \mathbb{E}[Y|\mathcal{G}]$ a.s.

(3) Positivity*:*
If $Y \;\geq\; 0$ a.s., then $\mathbb{E}[Y|\mathcal{G}] \;\geq\; 0$ a.s.

(4) Projection property *:*
If X is $\mathcal{G}$-measurable, then $\mathbb{E}[X|\mathcal{G}] \;=\; X$ a.s.

(5) Tower property*:*

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] \;=\; \mathbb{E}[\mathbb{E}[X|\mathcal{H}]|\mathcal{G}] \;=\; \mathbb{E}[X|\mathcal{H}] \text{ a.s.}$$

(6) "Take out what is known"
If $Z : \Omega \to \mathbb{R}$ is $\mathcal{G}$-measurable and $\mathbb{E}[|XZ|] \;<\; \infty$, then

$$\mathbb{E}[ZX|\mathcal{G}] \;=\; Z\mathbb{E}[X|\mathcal{G}] \text{ a.s.}$$

(7) Conditional expectation on a trivial σ-algebra
If $\mathcal{G} \;=\; (\emptyset, \Omega)$, then $\mathbb{E}[X|\mathcal{G}] \;=\; \mathbb{E}X$

(8) Conditional expectation: Independence
If for all $B \in \mathcal{B}(\mathbb{R})$ and all $A \in \mathcal{G}$, one has that

$$\mathbb{P}\left(\{X \in B\} \cap A\right) = \mathbb{P}(X \in B)\mathbb{P}(A),$$

i.e. if X is independent of $\mathcal{G}$, then $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}X$ a.s.

Definition 3.4.19 (Conditional Probability with respect to a σ-algebra)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\mathcal{G} \subseteq \mathcal{F}$ and $A \in \mathcal{F}$. We define the conditional probability of the event A with respect to the σ-algebra $\mathcal{G}$ as:*

$$\mathbb{P}(A|\mathcal{G}) \;=\; \mathbb{E}[\mathbb{1}_A|\mathcal{G}]$$

Lemma 3.4.20. *(Factorization Lemma [Kl, Corollary 1.97]) Let $(\Omega_2, \mathcal{F}_2)$ be a measurable space and let $\Omega_1 \;\neq\; \emptyset$. Let $f\colon \Omega_1 \to \Omega_2$ be a map. A map $g\colon \Omega_1 \to \overline{\mathbb{R}}$ is $\sigma(f) - \mathcal{B}(\overline{\mathbb{R}})$- measurable if and only if there is a measurable map $\phi\colon (\Omega_2, \mathcal{F}_2) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ such that $g = \phi \circ f$.*

We now present a more general definition of the conditional probability for events, as given in [Kl, Page 180]. Let $(\Omega, \mathcal{F})$ be a measurable space, X a random variable with values in the measurable space $(G, \mathcal{G})$ and Z a $\sigma(X)$-measurable (real) random variable. By Lemma 3.4.20, and setting $f = X$ and $g = Z$, it follows that there is a map $\phi\colon G \to \mathbb{R}$ such that the following conditions hold

- ϕ is $\mathcal{G} - \mathcal{B}(\mathbb{R})$-measurable
- $\phi(X) = Z$.

Definition 3.4.21 (Conditional Probability)**.** *Let $Y \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ and $X \colon (\Omega, \mathcal{F}) \to (G, \mathcal{G})$. We define the conditional expectation of Y, given $X = x$, by $\mathbb{E}[Y|X = x] := \phi(x)$, where ϕ is the function with the properties from before with $Z = \mathbb{E}[Y|X]$. Analogously, we define $\mathbb{P}[A|X = x] = \mathbb{E}[\mathbb{1}_A|X = x]$ for $A \in \mathcal{F}$.*

Example 3.4.22 (An estimate involving conditional expectation)**.** *Let $X \geq 0$ a.s. and set $\mathcal{G} = \{A, A^c, \emptyset, \Omega\}$. Let $\mathbb{P}(A) = p$. It follows that*

$$\mathbb{E}[X|A^c] \;=\; \frac{\mathbb{E}[X\mathbb{1}_{A^c}]}{1-p} \;\leq\; \frac{\mathbb{E}[X]}{1-p},$$

which we show using some tricks related to conditional expectation and conditional probability. First of all, one can interpret the object $\mathbb{E}[X|A^c] = \mathbb{E}[X|\mathcal{G}]\mathbb{1}_{A^c}$ in such a way that we restrict ourselves first to the σ-algebra $\mathcal{G}$ and then to the particular set A^c. Let us define $\mathbb{E}[X|\mathcal{G}] = Y$ and recall that, for $a, b \in \mathbb{R}$, Y can be written as

$$Y \;=\; a\mathbb{1}_A + b\mathbb{1}_{A^c}.$$

This representation is allowed since Y is $\mathcal{G}$-measurable by definition and we are allowed to write such an object as the linear combination of partitions sets; in particular, note that $\Omega = A \cup A^c$. We further write:

$$\begin{aligned} \mathbb{E}[X\mathbb{1}_{A^c}] \;=\; \mathbb{E}[Y\mathbb{1}_{A^c}] \quad &= \quad \mathbb{E}[b\mathbb{1}_{A^c}] \;=\; b\mathbb{P}(A^c) \;=\; b(1-p) \\ \implies b \quad &= \quad \frac{\mathbb{E}[X\mathbb{1}_{A^c}]}{1-p}. \end{aligned} \tag{3.4.7}$$

Similary, it follows that

$$\begin{aligned} \mathbb{E}[X\mathbb{1}_{A}] \;=\; \mathbb{E}[Y\mathbb{1}_{A}] \quad &= \quad \mathbb{E}[a\mathbb{1}_{A}] \;=\; a\mathbb{P}(A) \;=\; ap \\ \implies a \quad &= \quad \frac{\mathbb{E}[X\mathbb{1}_{A}]}{p}. \end{aligned} \tag{3.4.8}$$

Using Equation (3.4.8) and Equation (3.4.7), together with our discussion at the beginning of this example, we can write:

$$\begin{aligned} \mathbb{E}[X|A^c] \quad &= \quad \mathbb{E}[X|\mathcal{G}]\mathbb{1}_{A^c} \\ &= \quad \left(\frac{\mathbb{E}[X\mathbb{1}_A]}{p}\mathbb{1}_A\right)\mathbb{1}_{A^c} + \left(\frac{\mathbb{E}[X\mathbb{1}_{A^c}]}{1-p}\mathbb{1}_{A^c}\right)\mathbb{1}_{A^c} \\ &= \quad \frac{\mathbb{E}[X\mathbb{1}_{A^c}]}{1-p}\mathbb{1}_{A^c} \\ &\leq \quad \frac{\mathbb{E}[X]}{1-p}, \end{aligned}$$

where the last step follows from the monotonicity of the Lebesgue integral since $X\mathbb{1}_{A^c} \leq X$ a.e. It is also worth noting that in the first step of both Equation (3.4.8) and Equation (3.4.7), we used Equation (3.4.6).

Estimates such as the one in Example 3.4.22 are incredibly valuable when looking for upper bounds in a simulation to check for efficiency. This exact estimate is crucial in [HeZu1, Theorem 3.6].

3.5 Stochastic Analysis

Following [ChWi, Chapter 1], [IkWa, Chapter 1], [KaSu, Chapter 1-2], [LaLa, Section 3.4], [Du, Chapter 3], and [CoTa, Chapter 2], we present the key results in stochastic analysis used in this thesis. For the most part, we restrict ourselves to the 1-dimensional case while noting that the results naturally extend into higher dimensions. Throughout this section, unless specified otherwise or it is clear from the context, we assume that we are operating on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 3.5.1 (Stochastic Process)**.** *Let $I = [0, T]$ for some $T \in (0, \infty)$ or $I = [0, \infty)$ and let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A collection of random variables $X = (X_t)_{t \in I}$, where $X_t \colon \Omega \to \mathbb{R}$, is said to be a* stochastic process *with index set I and* state space $\mathbb{R}$.

Definition 3.5.2 (Index set)**.** *We say that I is the index set if either $I = [0, T]$ for some $T \in (0, \infty)$ or $I = [0, \infty)$.*

From now on, in this section we assume that our index set is defined as in Definition 3.5.2.

Definition 3.5.3 (The trajectory of a stochastic process)**.** *For each fixed $\omega \in \Omega$, the function $t \mapsto X_t(\omega)$ is known as the* sample path *or* trajectory *associated with ω.*

Definition 3.5.4 (Filtration (history))**.** *The* filtration $\mathbb{F} = (\mathcal{F}_t)_{t \in I}$ *is a collection of σ-algebras on Ω for which $\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}$ whenever $s \leq t$ and $s, t \in I$. In slightly simpler terms, a filtration is an increasing family of sub-σ-algebras.*

Definition 3.5.5 (Natural history)**.** *The* natural history *of a stochastic process X is the history (filtration) $\mathbb{F}^X = (\mathcal{F}_t^X)_{t \in I}$ and $\mathcal{F}_t^X$ is the smallest σ-algebra for which X_s is $\mathcal{F}_t^X$-measurable for all $s \leq t$ with $s, t \in I$. See Remark 3.2.4 for additional details.*

Definition 3.5.6 (Stochastic Basis)**.** *The quadruplet $(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F})$ represents a* stochastic basis*. The terms* probability field *and* filtered probability space *are also used.*

Definition 3.5.7 (Martingales and related processes)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F})$ be a stochastic basis and let $X = (X_t)_{t \in I}$ be a stochastic process, where $X_t : \Omega \to \mathbb{R}$ and $s, t \in I$. A process X is called an ($\mathbb{F}$-)*martingale *if*

(a) X_t *is $\mathcal{F}_t$-measurable for all $t \in I$ (we say that X is* adapted *or $\mathbb{F}$-adapted),*

(b) $\mathbb{E}[|X_t|] < \infty$ *for all $t \in I$ and*

(c) $\mathbb{E}[X_t | \mathcal{F}_s] = X_s$ *for all $0 \leq s \leq t$, with $s, t \in I$.*

If we keep conditions (a) and (b) as well as the original setup as given in the definition, but modify (c) as follows

(c^{sub}) $\mathbb{E}[X_t \mid \mathcal{F}_s] \geq X_s$*, then X is a* submartingale *and*

(c^{sup}) $\mathbb{E}[X_t \mid \mathcal{F}_s] \leq X_s$*, then X is a* supermartingale.

Definition 3.5.8 (Finite-dimensional distribution)**.** *Let $X = (X_t)_{t\in I}$ and $Y = (Y_t)_{t\in I}$ be stochastic processes on $(\Omega, \mathcal{F}, \mathbb{P})$ and $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$, respectively. We say that X and Y have the same* finite-dimensional distributions *if*

$$\mathbb{P}((X_{t_1}, \dots, X_{t_n}) \in B) = \mathbb{P}_1((Y_{t_1}, \dots, Y_{t_n}) \in B)$$

for all $0 \leq t_1 < \dots < t_n \in I$, where $n = 1, 2, \dots$ and $B \in \mathcal{B}(\mathbb{R}^n)$.

Definition 3.5.9 (Progressively measurable)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F})$ be a stochastic basis and $X = (X_t)_{t\in I}$ a stochastic process where $X_t \colon \Omega \to \mathbb{R}$. We say that such a process is* progressively measurable *with respect to the filtration $\mathbb{F}$ if, for each $s \in I$, the map*

$$(t, \omega) \mapsto X_t(\omega)$$

from $[0, s] \times \Omega$ to $\mathbb{R}$ is measurable with respect to $\mathcal{B}([0, s]) \otimes \mathcal{F}_s$. We usually denote this by writing $X \colon (\Omega \times [0, s], \mathcal{F}_s \otimes \mathcal{B}([0, s])) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Definition 3.5.10 (Null sets)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. A set $A \in \mathcal{F}$ is called a μ-null set, or simply a* null set*, if $\mu(A) = 0$. By $\mathcal{N}_\mu$ we denote the class of all subsets of μ-null sets.*

Definition 3.5.11 (Complete Probability Space)**.** *A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called* complete *if $\mathcal{N}_\mathbb{P} \subset \mathcal{F}$.*

Definition 3.5.12 (Usual conditions)**.** *We say that a stochastic basis $(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F})$ satisfies the* usual conditions *if the following conditions are satisfied*

(i) The probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is complete,

(ii) $A \in \mathcal{F}_t$ for all $A \in \mathcal{F}$ with $\mathbb{P}(A) = 0$ and $t \in I$,

(iii) the filtration $(\mathcal{F}_t)_{t\in I}$ is right-continuous,i.e. $\mathcal{F}_t = \bigcap_{s>t, s\in I} \mathcal{F}_s$ for all $t \in I$.

The concepts of Definition 3.5.12 are crucial. Without them, much of the theory would fail. For a deeper overview of this, check out [Kl, Chapter 21, Chapter 25].

Definition 3.5.13. *(Brownian Motion [KaSu, Definition 2.1.1]) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space and $B = (B_t)_{t\in I}$ a stochastic process. The process B is called a* standard Brownian motion *if it satisfies the following conditions:*

(i) Initial condition:
$B_0(\omega) = 0$ a.s.,

(ii) Independent increments property:
for all $0 \leq s < t < \infty$ the random variable $B_t - B_s$ is independent of $(B_u)_{u\in[0,s]}$, which means that for all $0 \leq s_1 \leq \dots \leq s_n \leq s$ and $A, A_1, \dots, A_n \in \mathcal{B}(\mathbb{R})$ it holds that

$$\mathbb{P}(B_{s_1} \in A_1, \dots, B_{s_n} \in A_n, B_t - B_s \in A) = \mathbb{P}(B_{s_1} \in A_1, \dots, B_{s_n} \in A_n)\mathbb{P}(B_t - B_s \in A),$$

(iii) Stationary Increments Property:
for all $0 \leq s < t < \infty$ and for all $A \in \mathcal{B}(\mathbb{R})$ one has that

$$\mathbb{P}(B_t - B_s \in A) = \frac{1}{\sqrt{2\pi(t-s)}} \int_A \exp\left(-\frac{x^2}{2(t-s)}\right) \mathrm{d}\lambda(x),$$

(iv) Continuity of paths:
The sample paths $t \mapsto B_t(\omega)$ are almost surely continuous.

Definition 3.5.14 (Stopping time)**.** *The random variable $\tau : \Omega \to I \cup \{\infty\}$ is the* stopping time (or $\mathbb{F}$-stopping time) *if $\{\tau \leq t\} \in \mathcal{F}_t$ for all $t \in I$. Note that a stopping time is allowed to take the value ∞.*

Definition 3.5.15 (First Hitting Time)**.** *Let $X = (X_t)_{t \in I}$ be continuous, in the sense that it has continuous paths, and adapted and let H be a non-empty subset of the real line. We define the* first hitting time *to be*

$$\tau_H \;=\; \inf\{t \geq 0 \colon X_t \in H\}.$$

Theorem 3.5.16. *([Du, (3.4)Theorem]) If H in Definition 3.5.15 is a closed set, then the hitting time is actually a stopping time.*

Proposition 3.5.17. *(Recurrence of the Brownian motion [JeYoCh, Proposition 1.4.2.1]) Let B be a Brownian motion and τ_L the first hitting time of L, defined as $\tau_L = \inf\{t > 0 \colon B_t = L\}$. Then, it follows that $\mathbb{P}(\tau_L < \infty) = 1$ for every $L \in \mathbb{R}$.*

We now present *Itô's formula* and some necessary definitions. Since we only care about the key result, i.e. the formula itself, we skip over any formal introductions of the stochastic integral: the interested reader should refer to [Kl, Chapter 25] or [KaSu, Chapter 5].

Definition 3.5.18 ($\mathcal{L}_2^{\text{loc}}$)**.** *We define the space $\mathcal{L}_2^{\text{loc}}$ as:*

$$\mathcal{L}_2^{\text{loc}} \;:=\; \{H = (H_t)_{t \geq 0} \colon H \text{ is progressively measurable and } \int_0^t H_u^2(\omega)\mathrm{d}u < \infty \text{ a.s. for all } t \geq 0\}$$

Proposition 3.5.19. *([KaSu, Page 96]) Let $K, H \in \mathcal{L}_2^{\text{loc}}$ and $\alpha, \beta \in \mathbb{R}$. It then follows that*

$$\int_0^t (\alpha K_u + \beta H_u)\mathrm{d}B_u \;=\; \alpha \int_0^t K_u \mathrm{d}B_u + \beta \int_0^t H_u \mathrm{d}B_u \;\text{ a.s.}$$

Definition 3.5.20 (Itô process)**.** *A continuous and adapted process $X = (X_t)_{t \geq 0}$, $X_t : \Omega \to \mathbb{R}$, is called an* Itô process *if there exists a process $H \in \mathcal{L}_2^{\text{loc}}$ and a progressively measurable process $a = (a_t)_{t \geq 0}$ where*

$$\int_0^t |a_u(\omega)|\mathrm{d}u \;<\; \infty \text{ for all } t \geq 0 \text{ and } \omega \in \Omega,$$

and $x_0 \in \mathbb{R}$ such that

$$X_t(\omega) \;=\; x_0 + \left(\int_0^t H_u dB_u\right)(\omega) + \int_0^t a_u(\omega) du$$

for $t \;\geq\; 0$, a.s.

Definition 3.5.21 (The space $C^{1,2}([0,\infty) \times \mathbb{R})$)**.** *A continuous function $f \colon [0,\infty) \times \mathbb{R} \to \mathbb{R}$ belongs to the space $C^{1,2}([0,\infty) \times \mathbb{R})$ assuming all the partial derivatives $\frac{\partial f}{\partial t}, \frac{\partial f}{\partial x}$ and $\frac{\partial^2 f}{\partial x^2}$ exist on $(0,\infty) \times \mathbb{R}$, are continuous, and can be continuously extended to $[0,\infty) \times \mathbb{R}$. Of course, in order for the partial derivatives to make sense, we require that $(t,x) \mapsto f(t,x)$.*

Theorem 3.5.22. *(Itô's formula [LaLa, Theorem 3.4.10]) Let X be as in Definition 3.5.20 and let $f \in C^{1,2}([0,\infty) \times \mathbb{R})$. It follows that*

$$\begin{aligned} f(t, X_t) \;=\; f(0, X_0) \;&+\; \int_0^t \frac{\partial f}{\partial u}(u, X_u)\,\mathrm{d}u + \int_0^t \frac{\partial f}{\partial x}(u, X_u)\, H_u \mathrm{d}B_u \\ &+\; \int_0^t \frac{\partial f}{\partial x}(u, X_u)\, a_u \mathrm{d}u + \frac{1}{2}\int_0^t \frac{\partial^2 f}{\partial x^2}(u, X_u)\, H_u^2 \mathrm{d}u \end{aligned}$$

for $t \geq 0$ a.s. Note that $f(t, X_t)$ is also itself an Itô process.

We end this section with the definition of a Radon measure and the Poisson random measure. Random measures are beyond the scope of this thesis so we refer the interested reader to [Ka, Chapter 10].

Definition 3.5.23 (Radon measure)**.** *Let $E \subset \mathbb{R}^d$. A* Radon measure *on $(E, \mathcal{B}(E))$ is a measure μ such that for every compact measurable set $A \in \mathcal{B}(E)$, $\mu(A) < \infty$.*

Definition 3.5.24. *(Poisson random measure [CoTa, Section 2.6.1]) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $E \subset \mathbb{R}^d$, and μ a given (positive) Radon measure on a measurable space $(E, \mathcal{B}(E))$. For simplicity, let's denote $\mathcal{B}(E)$ with $\mathcal{E}$. A* Poisson random measure *on E with intensity measure μ is an integer-valued random measure:*

$$M : \Omega \times \mathcal{E} \to \mathbb{N}$$

$$(\omega, A) \mapsto M(\omega, A),$$

such that

1. *For (almost all) $\omega \in \Omega$, $M(\omega, \cdot)$ is an integer-valued Radon measure on E: for any bounded measurable $A \subset E$, $M(A) < \infty$ is an integer-valued random variable.*
2. *For each measurable set $A \subset E$, $M(\cdot, A) = M(A)$ is a Poisson random variable with parameter $\mu(A)$:*

$$\forall k \in \mathbb{N}, \quad P(M(A) = k) \;=\; e^{-\mu(A)} \frac{(\mu(A))^k}{k!}. \tag{3.5.1}$$

3. *For disjoint measurable sets $A_1, \ldots, A_n \in \mathcal{E}$, the variables $M(A_1), \ldots, M(A_n)$ are independent.*

3.6 Stochastic Differential Equations (SDEs)

We close our theoretical minimum with the presentation of *stochastic differential equations*, further denoted as *SDEs*, and some associated results, as they are given in [KaSu, Chapter 6], [JeYoCh, Chapter 1], [KlPl], [ReYo, Chapters 3-4], [Wi2], [ScPaBö, Chapter 6] and [KaSh1, Chapter 5]. Throughout this section we assume that we are working on a stochastic basis $(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0})$, that the usual conditions hold, and that $B = (B_t)_{t\geq 0}$ denotes a Brownian motion. From the setup, it follows that B is $\mathbb{F}$-adapted.

Definition 3.6.1. *(Strong solution of an SDE [KaSu, Adapted from Definition 6.1.1]) Let $x \in \mathbb{R}$, $D \subseteq \mathbb{R}$ be an open set, and let $b, \sigma \colon [0,\infty) \times D \to \mathbb{R}$ be continuous. We say that a stochastic process $(X_t)_{t\geq 0}$ is a* strong solution of the SDE

$$\mathrm{d}X_t \;=\; b(t, X_t)\mathrm{d}t + \sigma(t, X_t)\mathrm{d}B_t, \tag{3.6.1}$$

with initial condition $X_0 = x$, if the following assertions hold:

(i) $X_t(\omega) \in D$ *for all* $t \geq 0$ *and* $\omega \in \Omega$

(ii) X *is* $\mathbb{F}$*-adapted with (a.s.) continuous sample paths*

(iii) For all $t \geq 0$

$$X_t \;=\; x + \int_0^t b(u, X_u)\mathrm{d}u + \int_0^t \sigma(u, X_u)\mathrm{d}B_u \;\; a.s.$$

Note that from the continuity of X and (b, σ), it follows that

$$\int_0^t \left(|b(u, X_u(\omega))| + |\sigma(u, X_u(\omega))|^2\right) \mathrm{d}u \;<\; \infty$$

for all $\omega \in \Omega$ and that $\sigma(u, X_u) \in \mathcal{L}_2^{\mathrm{loc}}$. Note that σ is known as the diffusion *coefficient and b as the* drift *coefficient.*

Definition 3.6.2 (Strong uniqueness)**.** *The SDE in (3.6.1) with initial condition $X_0 = x$, has a unique strong solution if, for any two strong solutions $X = (X_t)_{t\geq 0}$ and $Y = (Y_t)_{t\geq 0}$ on $(\Omega, \mathcal{F}, \mathbb{P})$, one has*

$$\mathbb{P}\left\{\omega \in \Omega \colon X_t(\omega) = Y_t(\omega)\ \forall t \geq 0\right\} \;=\; 1.$$

This notion of uniqueness is called strong uniqueness *of solutions.*

Theorem 3.6.3. *(Conditions for Strong Solution [LaLa, Theorem 3.5.3]) Given all the assumptions so far, and letting there be some $K < \infty$, we can impose two conditions on b and σ to guarantee that Equation (3.6.1) has a unique solution:*

(i) Lipschitz condition:

$$|b(t,x) - b(t,y)| + |\sigma(t,x) - \sigma(t,y)| \;\leq\; K|x - y| \tag{3.6.2}$$

(ii) Linear Growth Bound:

$$|b(t,x)| + |\sigma(t,x)| \;\leq\; K(1 + |x|), \tag{3.6.3}$$

for $t \geq 0$ and $x, y \in \mathbb{R}$.

Some of the conditions in Theorem 3.6.3 can be relaxed. For us, the most important one is that we can replace the global Lipschitz condition in Equation (3.6.2) with a local one. This means that the fixed K can be replaced with, possibly different constants, K_N for $|x|, |y| \leq N$ and each $N > 0$. For more details, refer to [KlPl, Section 4.5].

Remark 3.6.4 (Note about weak solutions)**.** *In the theory of SDEs, we also distinguish* weak solutions. *We will not really deal with them in this work, but the basic idea behind them is that, given a problem, our goal is to construct an appropriate stochastic basis on which the problem can be solved and find the solution itself. The only relevant piece of information for this work is that, if a result holds for weak solutions, it will also hold for strong solutions. For more details, refer to [KaSu, Section 6.4].*

Definition 3.6.5 ($\mathcal{L}_2$ space)**.** *Let $\mathcal{L}_2$ be the set of all progressively measurable processes $H = (H_t)_{t\geq 0}$ where $H_t \colon \Omega \to \mathbb{R}$, such that*

$$\|H\|_{\mathcal{L}_2,T} \coloneqq \left(\mathbb{E}\int_0^T H_t^2 \mathrm{d}t\right)^{\frac{1}{2}} < \infty \text{ for all } T > 0.$$

Theorem 3.6.6. *(Girsanov's theorem [PaRă, Adapted from Theorem 2.51]) Let $H = (H_t)_{t\geq 0} \in \mathcal{L}_2$ and assume that the process $(\mathcal{E}_t)_{t\geq 0}$ defined by*

$$\mathcal{E}_t \coloneqq \exp\left(-\int_0^t H_u \mathrm{d}B_u - \frac{1}{2}\int_0^t H_u^2 \mathrm{d}u\right)$$

is a martingale. Let $T > 0$ and

$$\mathrm{d}\mathbb{Q}_T \coloneqq \mathcal{E}_T \mathrm{d}\mathbb{P}.$$

Then $(W_t)_{t\geq 0}$ with

$$W_t \coloneqq B_t + \int_0^t \mathbb{1}_{[0,T]}(u) H_u \mathrm{d}u$$

defines a Brownian motion $(W_t)_{t\geq 0}$ with respect to the stochastic basis $(\Omega, \mathcal{F}, \mathbb{Q}_T, (\mathcal{F}_t)_{t\geq 0})$.

Definition 3.6.7. *(Explosion Time [JeYoCh, Definition 1.5.4.10]) Suppose that X is a solution of an SDE with locally Lipschitz coefficients. Then, a localisation argument allows to define unambiguously, for every n, $(X_t)_{t\leq \tau_n}$, where τ_n is the first exit time from $[-n, n]$. Let $S = \sup \tau_n$. When $S < \infty$, we say that X* explodes *at time S.*

Proposition 3.6.8. *(Girsanov's Theorem for stopping times [JeYoCh, Proposition 1.7.5.4]) Let S be the explosion time of the solution of the SDE*

$$\mathrm{d}X_t = b(X_t)\mathrm{d}t + \mathrm{d}B_t.$$

Then, for any stopping time $\tau \leq S$ and for any bounded function $F \colon C[0,\infty) \to \mathbb{R}$. which is $(\mathcal{B}(C[0,\infty)), \mathcal{B}(\mathbb{R}))$-measurable, it holds

$$\mathbb{E}_{\mathbb{P}}\left[F(X_t, t \leq \tau)\right] = \mathbb{E}_{\mathbb{Q}}\left[\exp\left(\int_0^{\hat{\tau}} b(B_u)\mathrm{d}B_u - \frac{1}{2}\int_0^{\hat{\tau}} b^2(B_u)\mathrm{d}u\right) F(B_t, t \leq \hat{\tau})\right],$$

where $\hat{\tau}$ is the associated stopping time for $(B_t)_{t\geq 0}$.

Remark 3.6.9 (How to prove Proposition 3.6.8?)**.** *One could prove Propostion 3.6.8 by adapting the proof of* Theorem 3.2 *which can be found in [KaRu, Page 1033].*

Remark 3.6.10 (Lamperti transform)**.** *In Proposition 3.6.8, we restricted ourselves to SDEs where the diffusion term is equal to* 1*. In practice, we are able to transform SDEs into SDEs with diffusion term* 1 *using the* Lamperti transform*. The details can be found in [SäSo, Section 7.1].*

What follows now is the Feller test, along with some necessary results, as given in [KaSh1, Chapter 5].

Definition 3.6.11 (Weak solution up to an explosion time)**.** *A* weak solution up to an explosion time *of the equation* $\mathrm{d}X_t = b(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}B_t$ *is a triple* $(X, B), (\Omega, \mathcal{F}, \mathbb{P}), \mathbb{F}$*, where*

(i) $(\Omega, \mathcal{F}, \mathbb{P})$ *is a probability space and* $\mathbb{F}$ *is a filtration of sub-*σ*-algebras of* $\mathcal{F}$ *satisfying the usual conditions,*

(ii) $X = \{X_t, \mathcal{F}_t; 0 \leq t < \infty\}$ *is a continuous, adapted,* $\mathbb{R} \cup \{\pm\infty\}$*-valued process with* $|X_0| < \infty$ *a.s., and* $\{B_t, \mathcal{F}_t; 0 \leq t < \infty\}$ *is a standard, one-dimensional Brownian motion*

(iii) with

$$S_n \coloneqq \inf\{t \geq 0; |X_t| \geq n\}$$

we have

$$\mathbb{P}\left(\int_0^{t\wedge S_n} \left\{|b(X_u)| + \sigma^2(X_u)\right\} \mathrm{d}u < \infty\right) = 1; \forall\, 0 \leq t < \infty \tag{3.6.4}$$

and

(iv)

$$\mathbb{P}\Big(X_{t\wedge S_n} = X_0 + \int_0^t b(X_u)\mathbb{1}_{\{u\leq S_n\}}\mathrm{d}u + \int_0^t \sigma(X_u)\mathbb{1}_{\{u\leq S_n\}}\mathrm{d}B_u; \forall\, 0 \leq t < \infty\Big) = 1 \tag{3.6.5}$$

valid for every $n \geq 1$*.*

We refer to

$$S \coloneqq \lim_{n\to\infty} S_n,$$

as the explosion time *for* X*. The assumption of continuity of* X *in the extended real numbers implies that*

$$S = \inf\{t \geq 0 \colon X_t \notin \mathbb{R}\} \text{ and } X_S = \pm\infty \text{ a.s. on } \{S < \infty\}.$$

We stipulate that $X_t = X_S; S \leq t < \infty$*.*

When dealing with explosions in a general setting, we start with an interval on the real line

$$I = (l, r) \text{ where } -\infty \leq l < r \leq \infty, \tag{3.6.6}$$

and we assume that the coefficents $\sigma \colon I \to \mathbb{R}$ and $b \colon I \to \mathbb{R}$ satisfy

(i) Nondegeneracy

$$\sigma^2(x) \ > \ 0; \ \forall x \in I \tag{3.6.7}$$

(ii) Local integrability

$$\forall x \in I, \ \exists \epsilon > 0 \text{ such that } \int_{x-\epsilon}^{x+\epsilon} \frac{1+|b(y)|}{\sigma^2(y)} \mathrm{d}y \ < \ \infty \tag{3.6.8}$$

Definition 3.6.12 (Weak solution in an interval)**.** *A weak solution in the interval $I = (l, r)$ of the equation $\mathrm{d}X_t = b(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}B_t$ is a triple $(X, B), (\Omega, \mathcal{F}, \mathbb{P}), \mathbb{F}$, where*

(i) *$(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $\mathbb{F}$ is a filtration of sub-σ-algebras of $\mathcal{F}$ satisfying the usual conditions,*

(ii) *$X = \{X_t, \mathcal{F}_t; 0 \leq t < \infty\}$ is a continuous, adapted, $[l, r]$-valued process with $X_0 \in I$ a.s., and $\{B_t, \mathcal{F}_t; 0 \leq t < \infty\}$ is a standard, one-dimensional Brownian motion,*

(iii) *with $(l_n)_{n=1}^{\infty}$ and $(r_n)_{n=1}^{\infty}$ strictly monotone sequences satisfying $l < l_n < r_n < r$, $\lim_{n\to\infty} l_n = l$, $\lim_{n\to\infty} r_n = r$, and*

$$S_n \ := \ \inf\{t \geq 0 \colon X_t \notin (l_n, r_n)\}; \ n \geq 1,$$

the Equations (3.6.4) and (3.6.5) hold.

On the interval I, we define the exit time *as:*

$$S \ = \ \inf\{t \geq 0 \colon X_t \notin (l, r)\} = \lim_{n\to\infty} S_n. \tag{3.6.9}$$

Proposition 3.6.13. *(Feller Test I [KaSh1, 5.22 (c) Proposition]) Assume that nondegeneracy and local integrability hold, that X is a weak solution in the context of Definition 3.6.12, on the interval I, and that $X_0 = x_0 \in I$. Furthermore, for a fixed real number c and some $x \in I$, we define the* scale function *as*

$$p(x) \ := \ \int_c^x \exp\left(-2\int_c^{\xi} \frac{b(\zeta)\mathrm{d}\zeta}{\sigma^2(\zeta)}\right)\mathrm{d}\xi. \tag{3.6.10}$$

Given all of this, if $p(l+) = -\infty$ and $p(r-) < \infty$, then it holds that

$$\mathbb{P}\left(\lim_{t\uparrow S} X_t = r\right) \ = \ \mathbb{P}\left(\inf_{0\leq t<S} X_t > l\right) \ = \ 1. \tag{3.6.11}$$

Proposition 3.6.14 (Feller Test II)**.** *Assume that nondegeneracy and local integrability hold. We have that $\mathbb{P}(S < \infty) = 1$ if and only if one of the following conditions holds:*

(i) *$v(r-) < \infty$ and $v(l+) < \infty$*

(ii) *$v(r-) < \infty$ and $p(l+) = -\infty$*

(iii) *$v(l+) < \infty$ and $p(r-) = \infty$,*

where p is the scale function as before and v is defined, for a fixed real c and some $x, y \in I$, as:

$$v(x) = \int_c^x p'(y) \int_c^y \frac{2\mathrm{d}z}{p'(z)\sigma^2(z)}\mathrm{d}y.$$

Moreover, in Case (i), one has that $\mathbb{E}[S] < \infty$.

Definition 3.6.15. *(Squared Bessel Process [LeGa, Page 227]) Let $m \geq 0$ be a real number. The m-dimensional* squared Bessel process *is the real process taking non-negative values that solves the SDE*

$$\begin{aligned} \mathrm{d}X_t &= 2\sqrt{X_t}\mathrm{d}B_t + m\mathrm{d}t \\ X_0 &= x. \end{aligned} \tag{3.6.12}$$

If $d \geq 1$ is an integer and $\beta = (\beta^1, \ldots, \beta^d)$ is a d-dimensional Brownian motion, an application of Itô's formula shows that the process

$$|\beta_t|^2 = (\beta_t^1)^2 + \ldots + (\beta_t^d)^2$$

is a d-dimensional squared Bessel process. For more details, see for example [La, Section 1.8.3, Section 1.10] or [ReYo, Chapter 11].

Definition 3.6.16 (Bessel process)**.** *The m-dimensional Bessel process is obtained by setting $Y_t = \sqrt{X_t}$ in Equation (3.6.12), and, when $m = d$ is a positive integer, it corresponds to the Euclidean norm of a d-dimensional Brownian motion. One can get an explicit SDE for the Bessel process when $m > 1$, but it is less tractable than (3.6.12); see for Example [LeGa, Exercise 8.13].*

For completion, we give the formal definition of a Bessel bridge as can be found in [ReYo, Chapter XI]. For our purposes, however, it is enough to understand that a Bessel bridge is simply a stochastic object which has a known beginning and an end.

Remark 3.6.17 (Bessel Bridge)**.** *Formally defining a Bessel bridge would require introducing many concepts that are not necessary for this work so we give an intuitive definition. Let us say that we have a Bessel process on a time interval $[0, T]$. A Bessel bridge is basically a Bessel process with fixed locations at 0 and T, i.e., we condition the Bessel process to be at certain locations at those times. This is a practical tool when we have information about the starting point and the ending point of a process. For a more formal introduction to Bessel bridges, one can refer to [ReYo, Chapter XI]; in particular [ReYo, (3.1) Definition on Page 463].*

Proposition 3.6.18. *([ReYo, Page 317] or [Wi2, Section 3.6]) Let X be a three-dimensional Bessel process starting at 0, i.e., $X_0 = 0$ and B a Brownian motion starting at $b > 0$, i.e., $B_0 = b > 0$. Then, if $L_b = \sup\{t \colon X_t = b\}$ and $T_0 = \inf\{t > 0 \colon X_t = 0\}$, the processes $\{X_{L_b - t}, 0 \leq t \leq L_b\}$ and $\{B_t, 0 \leq t \leq T_0\}$ have the same law.*

This concludes our theoretical overview: we are finally ready to discuss the main results of Chapter 4.

Chapter 4

The Framework and Key Results

In this chapter, we present the key results: Proposition 2.0.3 and Theorem 2.0.5. We begin by presenting the modified acceptance-rejection scheme we are working with (Section 4.1), followed by a discussion of the general framework under which we are operating (Section 4.2), using the terminology of Section 3.6. After that, we give a detailed proof of Proposition 2.0.3 in Section 4.3 and Theorem 2.0.5 in Section 4.4. We complete the chapter with a discussion of the algorithm's shortcomings, the proposed solutions in Section 4.5 and a very brief overview of recent research in Section 4.6. Crucially, throughout this entire chapter, we adopt the conventions of Chapter 3 and, in particular, the framework and assumptions of Section 3.6.

4.1 The Acceptance Rejection Protocol

The key goal of this work is to "unpack" an acceptance-rejection protocol, given as *Algorithm* A_0 in [HeZu1, Page 1482] and inspired by the work of [BeRo]. What follows is an intuitive introduction to the algorithm; the skipped mathematical details are discussed in the next sections.

THE ALGORITHM

Let us fix the level $L > x$.

Step 1: *Simulate a non-negative random variable* T *with density* f_T.

Step 2: *Simulate a* 3*-dimensional Bessel process* (R_t) *on the time interval* $[0, T]$ *with endpoint* $R_T = L - x$. *We define by* $D_{R,T}$ *the stochastic domain:*

$$D_{R,T} := \Big\{(t, \nu) \in [0, T] \times \mathbb{R}_+ : \ \nu \leq \gamma(L - R_t)\Big\}. \tag{4.1.1}$$

This domain depends on both random elements $(R_t)_{t\in[0,T]}$ *and* T.

Step 3: *Simulate a Poisson point process* N*, with Lebesgue intensity measure, on the state space* $[0, T] \times [0, \infty)$*, independent of the Bessel process.*

Step 4: *If* $N(D_{R,T}) = 0$ *then set* $Y = T$. *Otherwise go to Step 1.*

Outcome: *the random variable* Y.

Let us walk through the algorithm step by step. The goal is to exactly simulate the first time the solution of an SDE, denoted by X and starting at x, hits some positive level L. Hence, we fix

$L > x$.
In the *first step*, we simulate a non-negative random variable T with a particular density, f_T. It being non-negative makes sense since, in this universe at least, time is non-negative.
In the *second step*, we use the 3-dimensional Bessel process to construct the stochastic domain in Equation (4.1.1). One of the reasons the Bessel process is used here is because it is relatively easy to simulate; its connection to X is explored further in Section 4.3. This stochastic domain is crucial in defining a success, but only once we pair it up with a Poisson point process, which we introduce in the *third step*.
Using the Poisson point process, we are able to encode a decaying exponential, an object that will show up multiple times in the calculations, into our simulation. Furthermore, it also enables the insight [HeZu1] had about using skeleton paths instead of the full sample path: since the goal is to generate something which hits a level L, everything that is within the stochastic domain should be rejected since it does not go high enough; this is the intuition behind $N(D_{R,T}) = 0$, i.e. no points are in the stochastic domain, being the desired event indicating success. This is exactly the point of the *fourth step*.
We now move on to the mathematical framework of our problem.

4.2 The Mathematical Framework

We begin with a stochastic process $X = (X_t)_{t\geq 0}$ which solves the SDE:

$$\begin{aligned} \mathrm{d}X_t &= b(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}B_t \\ X_0 &= x \in \mathbb{R}. \end{aligned} \tag{4.2.1}$$

Throughout, we will set $\sigma \equiv 1$ which is justified using the Lamperti transform that we discussed in Remark 3.6.10. The goal is to find the exact simulation of the first time the process X reaches the level $L > 0$. We formalize this now with a definition:

Definition 4.2.1 (First hitting time)**.** *We denote with τ_L the first time the process X reaches a level $L > 0$. Formally, we write:*

$$\tau_L := \inf\{t \geq 0 \colon X_t = L\}. \tag{4.2.2}$$

Let us observe that, in some cases, we can get an explicit expression for the density of τ_L. In particular, when we set $b(X_t) = 0$, we are reducing ourselves to the case of the Brownian motion which has the explicit solution given by $X_t = x + B_t$. Note that one can interpret this solution as the standard Brownian motion shifted by x.

Proposition 4.2.2 (Density of first hitting time of a Brownian motion starting in x)**.** *Let $B_0 = x$ and $G \sim \mathcal{N}(0,1)$. Our claim is that:*

$$\tau_L \overset{\mathrm{d}}{=} \frac{(L-x)^2}{G^2}$$

and that

$$f_{\tau_L}(t) = \frac{L-x}{\sqrt{2\pi t^3}} \exp\left(-\frac{(L-x)^2}{2t}\right).$$

Proof of Proposition 4.2.2. Without loss of generality, we first assume that $x = 0$. This keeps the algebra cleaner and the end result can be easily adjusted with a "shift" in x. Since our goal is to determine the density, we begin with the distribution function since it is an easier object to analyse: $\mathbb{P}(\tau_L < t)$, where t plays the role of a sort of dummy variable. The event $\{\tau_L < t\}$ can actually be broken up as follows:

$$\{\tau_L < t\} \;=\; \{\{\tau_L < t\} \cap \{B_t < L\}\} \cup \{\{\tau_L < t\} \cap \{B_t > L\}\},$$

which gives

$$\begin{aligned}
\mathbb{P}(\tau_L < t) &= \mathbb{P}(\{\{\tau_L < t\} \cap \{B_t < L\}\} \cup \{\{\tau_L < t\} \cap \{B_t > L\}\}) \\
&= \mathbb{P}(\{\{\tau_L < t\} \cap \{B_t < L\}\}) + \mathbb{P}(\{\{\tau_L < t\} \cap \{B_t > L\}\}) \\
&= \mathbb{P}(\tau_L < t, B_t < L) + \mathbb{P}(\tau_L < t, B_t > L),
\end{aligned}$$

where we used σ-additivity since the events were disjoint and we rewrote the final result to keep the notation cleaner. Now, we take a look at two different cases.

1. Case 1: $\tau_L < t$ and $B_t < L$
   In this case, there is not really much to unpack about these events beyond the fact that the Brownian motion hit L for the first time at τ_L and then continued to wiggle around the level L. At this point, one needs to take a bit of an intuitive leap. We exploit the fact that a Brownian motion starting at level L is symmetric around the level L. Let us say that B reached L and then it moved around a little bit till it reached the level M. Going from level L to level M is equally as likely as going from level L to the level $2L - M$. This implies that every path above L has a "twin" below L. Of course, this is not a rigorous argument: the probability of a Brownian motion following any one particular path is zero. However, it does lead us to the right idea that:
   $$\mathbb{P}(\tau_L < t, B_t < L) = \mathbb{P}(\tau_L < t, B_t > L),$$
   which we exploit to write
   $$\mathbb{P}(\tau_L < t) \;=\; 2\mathbb{P}(\tau_L < t, B_t > L). \tag{4.2.3}$$
   We have now reduced ourselves to exploring only one case.

2. Case 2: $\tau_L < t$ and $B_t > L$
   In this case, the Brownian motion hits L and, at the end of the day, the Brownian motion is above L. Essentially, the event $\{\tau_L < t\}$ becomes irrelevant in determining the probability of $B_t > L$ i.e. $\mathbb{P}(\tau_L < t, B_t > L) = \mathbb{P}(B_t > L)$. Plugging this into Equation (4.2.3) gives
   $$\mathbb{P}(\tau_L < t) \;=\; 2\mathbb{P}(B_t > L).$$

From here we write:

$$\begin{aligned}
\mathbb{P}(\tau_L < t) &= 2\mathbb{P}(B_t > L) \\
&= 2(1 - \mathbb{P}(B_t \leq L)) \\
&= 2 - 2\mathbb{P}(B_1\sqrt{t} \leq L) \\
&= 2 - 2\mathbb{P}(B_1 \leq \frac{L}{\sqrt{t}}) \\
&= 2 - \sqrt{\frac{2}{\pi}} \int_{-\infty}^{\frac{L}{\sqrt{t}}} \exp\left(-\frac{y^2}{2}\right) \mathrm{d}y,
\end{aligned} \tag{4.2.4}$$

where the crucial insight was that $B_1 \sim \mathcal{N}(0,1)$ and that $B_t \overset{\mathrm{d}}{=} B_1\sqrt{t}$. Equation (4.2.4) gives us a useful form of the distribution function which we now differentiate, using Theorem A.0.5 to get the density:

$$\begin{aligned} f_{\tau_L}(t) &= \frac{\mathrm{d}}{\mathrm{d}t}\left(-\sqrt{\frac{2}{\pi}}\int_{-\infty}^{\frac{L}{\sqrt{t}}}\exp\left(-\frac{y^2}{2}\right)\mathrm{d}y\right) \\ &= -\sqrt{\frac{2}{\pi}}\frac{\mathrm{d}}{\mathrm{d}t}\left(\int_{-\infty}^{\frac{L}{\sqrt{t}}}\exp\left(-\frac{y^2}{2}\right)\mathrm{d}y\right) \\ &= -\sqrt{\frac{2}{\pi}}\left[\left(\int_{-\infty}^{\frac{L}{\sqrt{t}}}\frac{\partial}{\partial t}\exp\left(-\frac{y^2}{2}\right)\mathrm{d}y\right) + \exp\left(-\frac{L^2}{2t}\right)\left(-\frac{1}{2}Lt^{-\frac{3}{2}}\right)\right] \\ &= \sqrt{\frac{2}{\pi}}\frac{1}{2}\frac{L}{\sqrt{t^3}}\exp\left(-\frac{L^2}{2t}\right) \\ &= \frac{L}{\sqrt{2\pi t^3}}\exp\left(-\frac{L^2}{2t}\right). \end{aligned} \tag{4.2.5}$$

Our last task is to show that, under the simplification that $x = 0$,

$$\tau_L \overset{\mathrm{d}}{=} \frac{L^2}{G^2}.$$

A clever way to show this is as follows

$$\begin{aligned} \mathbb{P}\left(\frac{L^2}{G^2} \le t\right) &= \mathbb{P}\left(\frac{L^2}{t} \le G^2\right) \\ &= \mathbb{P}\left(G^2 \ge \frac{L^2}{t}\right) \\ &= \mathbb{P}\left(|G| \ge \frac{L}{\sqrt{t}}\right) \\ &= 2\mathbb{P}\left(G \ge \frac{L}{\sqrt{t}}\right) \\ &= 2\left(1 - \mathbb{P}\left(G < \frac{L}{\sqrt{t}}\right)\right) \end{aligned}$$

where, in the second to last step, we used the symmetry of the Gaussian distribution. After differentiating both sides with respect to t and observing the connection to the calculations associated with Equation (4.2.5), it follows that

$$\tau_L \overset{\mathrm{d}}{=} \frac{(L-x)^2}{G^2}.$$

Remark 4.2.3 (Strict inequality?)**.** *The strict inequality in the way we defined the distribution function in this problem seems to contradict our Definition 3.2.12. Sometimes, the strict inequality is better for illustrating a proof, as was the case in this problem. The reason why it does not mathematically matter is because, after we take the derivative to find the density, the endpoint essentially becomes irrelevant. One could also argue that a non-strict inequality is actually redundant since, in any case, the probability that a continuous random variable takes on any specific value is zero anyway.*

□

The result given in Proposition 4.2.2 is fairly standard, but it serves to show the computational usefulness of the Brownian motion. For a different approach to this problem, refer to [LeGa, Section 2.4], [BhWa, Page 191] or [JeYoCh, Page 140]. Some of these books look at the reflection principle from a much more formal standpoint than we did. For another intuitive approach to the reflection principle, presented wonderfully in the discrete-time case and with an application to American options, refer to [Shr1, Section 5.2-5.3].
We now delve a bit deeper into the machinery we are using in dealing with this task. The first step is the introduction of three key objects:

$$\beta(x) \;=\; \int_0^x b(y)\mathrm{d}y, \tag{4.2.6}$$

$$p(x) \;=\; \int_0^x \exp\left(-\beta(y)\right)\mathrm{d}y, \tag{4.2.7}$$

$$\gamma(x) \;:=\; \frac{b^2(x)+b'(x)}{2}. \tag{4.2.8}$$

The similarity between Equation (4.2.7) and Equation (3.6.10) is no accident and the importance will become clear shortly.

Assumption 4.2.4 (Drift term and scale function assumptions)**.** *The drift term* $b \in \mathcal{C}^1((-\infty, L])$ *and* $\lim_{x\to-\infty} p(x) = -\infty$.

Remark 4.2.5 (Negative upper limit of integration?)**.** *The careful reader notices that the latter part of Assumption 4.2.4 leads to a negative upper limit of integration in Equation (4.2.7). It turns out that this is not an issue: the clearest way to show this is via a* u*-substitution. We begin first with the assumption that* $x < 0$.

$$\begin{aligned}
p(x) \quad &= \int_0^x \exp\left(-\beta(y)\right)\mathrm{d}y \\
&= \int_0^x \exp\left(-\int_0^y b(z)\mathrm{d}z\right)\mathrm{d}y \\
&\qquad \text{Substitution} \qquad \boxed{\begin{array}{ll} y=-u, \textit{ If } & y=x \implies u=-x>0 \\ \mathrm{d}y=-\mathrm{d}u, \textit{ If } & y=0 \implies u=0 \end{array}} \\
&= \int_0^{-x} \exp\left(-\int_0^{-u} b(z)\mathrm{d}z\right)\mathrm{d}(-u) \\
&= -\int_0^{-x} \exp\left(-\int_0^{-u} b(z)\mathrm{d}z\right)\mathrm{d}u \\
&= -\int_0^{-x} \exp\left(-\beta(-u)\right)\mathrm{d}u,
\end{aligned}$$

which gives us an integral with "familiar" limits.

Given Assumption 4.2.4, we now look at its implications.

Claim 4.2.6 (Strong solution)**.** *Assumption 4.2.4 implies that Equation (4.2.1) has a strong solution in the sense of Definition 3.6.1 and Theorem 3.6.3.*

Proof of Claim 4.2.6. We focus first on $b \in \mathcal{C}^1((-\infty, L])$ which means that b is continuously differentiable on the interval $(-\infty, L]$ and $b\colon (-\infty, L] \to \mathbb{R}$. From [KlPl, Page 134], or via Theorem 3.6.3, we know that Local Lipschitz conditions are sufficient for the solution to be strong. Thus, our job is to show that the assumption $b \in \mathcal{C}^1((-\infty, L])$ guarantees local Lipschitz conditions in the context of Equation (4.2.1). Given that $\sigma \equiv 1$, our problem is reduced to showing that there exists some constant $K > 0$ such that, for some $x, y \in (-\infty, L]$,

$$|b(x) - b(y)| \ \leq \ K|x - y|.$$

Now, let $x, y \in (-\infty, L]$. From Theorem A.0.6, we get that there is a $c \in (x, y)$ such that

$$|b'(c)| \ = \ \left|\frac{b(y) - b(x)}{y - x}\right|.$$

Now, b' is a continuous function on a compact interval and thus bounded (Theorem A.0.8), i.e. there exists some $K \in (0, \infty)$ such that $|b'| \leq K$, so it follows that

$$\begin{aligned} \left|\frac{b(y) - b(x)}{y - x}\right| &\leq K \\ \implies |b(y) - b(x)| &\leq K|y - x|, \end{aligned}$$

so local Lipschitz conditions hold. Hence, the solution X satisfies the conditions of strong uniqueness. Of course, being that we are working with local conditions, this is only valid until the process reaches L. Fortunately for us, this is all we need since we do not care about the process after it passes L. □

Claim 4.2.7 (Nondegeneracy and local integrability)**.** *Equation (3.6.7) and Inequality (3.6.8) are satisfied for $b \in \mathcal{C}^1((-\infty, L])$ and $\sigma \equiv 1$.*

Proof of Claim 4.2.7. Since we will need this later for the Feller test, we look at the complete setup. We are working on the interval $I = (-\infty, L)$ and $\sigma, b\colon (-\infty, L) \to \mathbb{R}$. Nondegeneracy is satisfied since $\sigma^2(x) \equiv 1 > 0$ for all $x \in I$. For local integrability, we want to show that for all $x \in I$ there exists $\epsilon > 0$ such that Inequality (3.6.8) is satisfied. This follows from

$$\begin{aligned} \int_{x-\epsilon}^{x+\epsilon} \frac{1 + |b(y)|}{\sigma^2(y)} \mathrm{d}y &= \int_{x-\epsilon}^{x+\epsilon} (1 + |b(y)|) \mathrm{d}y \\ &= (x + \epsilon - x + \epsilon) + \int_{x-\epsilon}^{x+\epsilon} |b(y)| \mathrm{d}y \\ &= 2\epsilon + \int_{x-\epsilon}^{x+\epsilon} |b(y)| \mathrm{d}y, \end{aligned}$$

and now we just have to show that $\int_{x-\epsilon}^{x+\epsilon} |b(y)| \mathrm{d}y$ is finite. It is possible to choose $\epsilon > 0$ such that $[x + \epsilon, x - \epsilon] \subset (-\infty, L)$ so b is continuous on that compact interval and thus bounded which we can denote as $|b(y)| \leq M < \infty$ (Theorem A.0.4 and Theorem A.0.8). We thus rewrite the previous integral and get

$$\begin{aligned} \int_{x-\epsilon}^{x+\epsilon} |b(y)| \mathrm{d}y &\leq \int_{x-\epsilon}^{x+\epsilon} M \mathrm{d}y \\ &= 2M\epsilon. \end{aligned}$$

Finally, it follows that

$$\int_{x-\epsilon}^{x+\epsilon} \frac{1+|b(y)|}{\sigma^2(y)} \mathrm{d}y \quad \leq \quad 2(M+1)\epsilon < \infty,$$

for the right choice of ϵ so local integrability holds. □

The results of Claim 4.2.7 allow us to perform the Feller test in the spirit of Proposition 3.6.13. This form of the Feller test is meaningful for us because of the setup: the goal is for the process to escape the interval in such a way that, as it exists the interval, it hits the level $L > 0$. Given Assumption 4.2.4, it is clear that $p(l^+) = p(-\infty) = -\infty$. Hence, what we are left with is to show that $p(r^-) = p(L) < \infty$ in order to satisfy all the conditions for the Feller test.

Claim 4.2.8 ($p(L) < \infty$)**.**

Proof of Claim 4.2.8. From Equation (4.2.6) and Equation (4.2.7), it follows that

$$p(L) \coloneqq \int_0^L \exp\left(-\int_0^y b(z)\mathrm{d}z\right) \mathrm{d}y.$$

Note that this is exactly Equation (3.6.10), where we used the fact that $\sigma \equiv 1$ and we dropped the 2 in the exponential term since it will not have an impact on the limiting behavior of the function. Furthermore, and looking again at Equation (3.6.10), we set $c \equiv 0$. The rest boils down to doing the necessary algebra while keeping in mind that, again, $b(z)$ is bounded on $[0, L]$, let us say by some $C > 0$: this means that $-C \leq b(z) \leq C$ for $z \in [0, L]$. It follows that

$$\begin{aligned}
p(L) \quad &\leq \quad \int_0^L \exp\left(-\int_0^y (-C)\mathrm{d}z\right) \mathrm{d}y \\
&= \quad \int_0^L \exp\left(Cy\right) \mathrm{d}y \\
&= \quad \frac{1}{C} \exp\left(Cy\right) \Big|_0^L \\
&= \quad \frac{1}{C} \left(\exp\left(CL\right) - \exp\left(C \cdot 0\right)\right) \\
&= \quad \frac{1}{C} \left(\exp\left(CL\right) - 1\right) \\
&= \quad \frac{\exp\left(CL\right) - 1}{C} < \infty,
\end{aligned}$$

since all of the involved quantities are themselves finite. □

Finally, this leads us to the following key claim:

Claim 4.2.9 (The solution of Equation (4.2.1) explodes and $\tau_L < S$ a.s.)**.** *The solution of Equation (4.2.1) explodes in some finite time S, defined in Definition 3.6.12, which is greater than the first hitting time.*

Proof of Claim 4.2.9. The proof follows directly from the calculations done in Claim 4.2.8 and Proposition 3.6.13. In particular, Claim 4.2.8 shows that the conditions for explosion in

Proposition 3.6.13 are satisfied. Crucially, the result shows that the process exits through the right side of the interval I, in our case the level $L > 0$. Explicitly:

$$\mathbb{P}\left(\lim_{t\uparrow S} X_t = L\right) \;=\; \mathbb{P}\left(\inf_{0\le t<S} X_t > -\infty\right) \;=\; 1.$$

It follows directly from $v(L) < \infty$, which is shown in Claim 4.4.2, and Assumption 4.2.4, that the conditions for Proposition 3.6.14 (ii) are met so $S < \infty$ a.s. Finally, from all of this, it follows that $\tau_L < S < \infty$ a.s. □

The result of Claim 4.2.9 is crucial since it allows us to use Girsanov's theorem in the form of Proposition 3.6.8. Granted, Proposition 3.6.8 was stated only for stopping times, but we observe that the event $\{X_t = L\}$ is a closed set so, according to Theorem 3.5.16, τ_L is also a stopping time so we are allowed to use Proposition 3.6.8. We exploit all of these results in Section 4.3 and formally prove Proposition 2.0.3.

4.3 An Application of Girsanov's Theorem

For the reader's convenience, we first restate Proposition 2.0.3.

Proposition 4.3.1. *([HeZu1, Proposition 2.1]) Under Assumption 4.2.4, for any bounded and measurable function $\psi\colon \mathbb{R} \to \mathbb{R}$, we obtain*

$$\mathbb{E}_{\mathbb{P}}[\psi(\tau_L)\mathbb{1}_{\{\tau_L<\infty\}}] \;=\; \mathbb{E}_{\mathbb{Q}}[\psi(\tau_L)\eta(\tau_L)]\exp\left(\beta(L)-\beta(x)\right), \tag{4.3.1}$$

where $\mathbb{P}$ (respectively $\mathbb{Q}$) corresponds to the distribution of X, which was defined by Equation (4.2.1), (respectively the Brownian motion B) and

$$\eta(t) \;:=\; \mathbb{E}\left[\exp\left(-\int_0^t \gamma(L-R_s)\mathrm{d}s\right)\middle| R_t = L - x\right]. \tag{4.3.2}$$

Here, $(R_t)_{t\ge 0}$ stands for a 3-dimensional Bessel process with $R_0 = 0$.

Before delving into the proof, it is worth mentioning why we are interested in the bounded and measurable function ψ. As is often the case in mathematics, it tends to be easier to prove a result for a general class of functions instead of one particular function. In the context of stochastics, one interesting function that is within the class of bounded and measurable functions is an indicator function of the form $\mathbb{1}_{(-\infty,x]}(X)$. Taking the expectation of $\mathbb{1}_{(-\infty,x]}(X)$ gives us exactly the distribution function of the associated random variable X. This is exactly why we care about ψ; it allows us to "pin down" the distribution function. We are now ready to proceed with the proof.

Proof of Proposition 4.3.1.

$$\mathbb{E}_{\mathbb{P}}[\psi(\tau_L)\mathbb{1}_{\{\tau_L<\infty\}}] \overset{\text{Prop. 3.6.8}}{=} \mathbb{E}_{\mathbb{Q}}\left[\psi(\hat{\tau}_L)\exp\left(\int_0^{\hat{\tau}_L} b(B_u)\mathrm{d}B_u - \frac{1}{2}\int_0^{\hat{\tau}_L} b^2(B_u)\mathrm{d}u\right)\right]. \tag{4.3.3}$$

Note that, under the measure $\mathbb{Q}$, the process B is a one-dimensional Brownian motion starting at the point x (follows from the $b \equiv 0$ and $\sigma \equiv 1$ case in Equation 4.2.1). Furthermore, observe that we have changed the "lens" under which we are looking at the first hitting time. Under the

measure $\mathbb{P}$, the first hitting time was exactly the one given in Equation (4.2.2), meaning that we were looking at the first time the process X hits level L. However, once we change the measure, the first hitting time is observed under the measure $\mathbb{Q}$ which interprets it as the first time the Brownian motion hits the level L which is why, after the application of Proposition 4.3.1, we use $\hat{\tau}_L$. Crucially, $\hat{\tau}_L = \inf\{t \geq 0 \colon B_t = L\} < \infty$ a.s. according to Proposition 3.5.17 and $B_{\hat{\tau}_L} = L$ by construction. The next step now is to apply Itô's formula, in the sense of Theorem 3.5.22, to the function $\beta(z)$, with z playing the role of a dummy variable, where we choose our process to be $B_{\hat{\tau}_L}$. It is worth remembering that a Brownian motion can be rewritten in such a way to make it explicit that it is an Itô process in the sense of Definition 3.5.20:

$$B_t = 0 + \int_0^t 1 \cdot \mathrm{d}B_u + \int_0^t 0 \cdot \mathrm{d}u, \tag{4.3.4}$$

which is useful since it directly gives us that $H_u \equiv 1$ and $a_u \equiv 0$. We now write

$$\begin{aligned} \beta(B_{\hat{\tau}_L}) = \beta(x) &+ \int_0^{\hat{\tau}_L} \frac{\partial \beta}{\partial u}(u, B_u)\,\mathrm{d}u \\ &+ \int_0^{\hat{\tau}_L} \frac{\partial \beta}{\partial z}(u, B_u)\,\mathrm{d}B_u + 0 + \frac{1}{2}\int_0^{\hat{\tau}_L} \frac{\partial^2 \beta}{\partial z^2}(u, B_u)\,\mathrm{d}u. \end{aligned}$$

Evaluating the derivatives and using Theorem A.0.7 gives

$$\begin{aligned} \beta(B_{\hat{\tau}_L}) &= \beta(x) + \int_0^{\hat{\tau}_L} b(B_u)\mathrm{d}B_u + \frac{1}{2}\int_0^{\hat{\tau}_L} b'(B_u)\mathrm{d}u \\ \implies \int_0^{\hat{\tau}_L} b(B_u)\mathrm{d}B_u &= \beta(B_{\hat{\tau}_L}) - \beta(x) - \frac{1}{2}\int_0^{\hat{\tau}_L} b'(B_u)\mathrm{d}u. \end{aligned} \tag{4.3.5}$$

We now go back to Equation (4.2.8) and integrate that from 0 to $\hat{\tau}_L$ and get:

$$\begin{aligned} \int_0^{\hat{\tau}_L} \gamma(B_u)\mathrm{d}u &= \int_0^{\hat{\tau}_L} \frac{b^2(B_u)}{2}\mathrm{d}u + \int_0^{\hat{\tau}_L} \frac{b'(B_u)}{2}\mathrm{d}u \\ \implies -\int_0^{\hat{\tau}_L} \frac{b^2(B_u)}{2}\mathrm{d}u &= \int_0^{\hat{\tau}_L} \frac{b'(B_u)}{2}\mathrm{d}u - \int_0^{\hat{\tau}_L} \gamma(B_u)\mathrm{d}u. \end{aligned} \tag{4.3.6}$$

Adding up together Equation (4.3.5) and Equation (4.3.6) gives

$$\int_0^{\hat{\tau}_L} b(B_u)\mathrm{d}B_u - \frac{1}{2}\int_0^{\hat{\tau}_L} b^2(B_u)\mathrm{d}u = \beta(B_{\hat{\tau}_L}) - \beta(x) - \int_0^{\hat{\tau}_L} \gamma(B_u)\mathrm{d}u,$$

which, after plugging it into Equation (4.3.3), leads to

$$\mathbb{E}_{\mathbb{P}}[\psi(\tau_L)\mathbb{1}_{\{\tau_L<\infty\}}] = \mathbb{E}_{\mathbb{Q}}\left[\psi(\hat{\tau}_L)\exp\left(\beta(B_{\hat{\tau}_L}) - \beta(x) - \int_0^{\hat{\tau}_L}\gamma(B_u)\mathrm{d}u\right)\right]. \tag{4.3.7}$$

The truly beautiful thing about going from Equation (4.3.3) to Equation (4.3.7) is that we went from having to deal with stochastic integrals to standard Lebesgue integrals. Using this insight, along with the standard properties of the Lebesgue integral, we can simplify Equation (4.3.7) even further:

$$\mathbb{E}_{\mathbb{P}}[\psi(\tau_L)\mathbb{1}_{\{\tau_L<\infty\}}] = \exp\left(\beta(B_{\hat{\tau}_L}) - \beta(x)\right)\mathbb{E}_{\mathbb{Q}}\left[\psi(\hat{\tau}_L)\exp\left(-\int_0^{\hat{\tau}_L}\gamma(B_u)\mathrm{d}u\right)\right], \tag{4.3.8}$$

where we used the fact that $\exp\left(\beta(B_{\hat{\tau}_L}) - \beta(x)\right)$ is just a scalar and can thus be pulled out. We shift our focus now just to the expectation with respect to the measure $\mathbb{Q}$.

$$\begin{aligned}
&\mathbb{E}_{\mathbb{Q}}\left[\psi(\hat{\tau}_L)\exp\left(-\int_0^{\hat{\tau}_L}\gamma(B_u)\mathrm{d}u\right)\right] \;=\; \mathbb{E}_{\mathbb{Q}}\left[\psi(\hat{\tau}_L)\exp\left(-\int_0^{\hat{\tau}_L}\gamma(B_u)\mathrm{d}u\right)\Big|\{\emptyset,\Omega\}\right]\\
=\quad &\mathbb{E}_{\mathbb{Q}}\left[\mathbb{E}_{\mathbb{Q}}\left[\psi(\hat{\tau}_L)\exp\left(-\int_0^{\hat{\tau}_L}\gamma(B_u)\mathrm{d}u\right)\Big|\sigma(\hat{\tau}_L)\right]\Big|\{\emptyset,\Omega\}\right]\\
=\quad &\mathbb{E}_{\mathbb{Q}}\left[\psi(\hat{\tau}_L)\mathbb{E}_{\mathbb{Q}}\left[\exp\left(-\int_0^{\hat{\tau}_L}\gamma(B_u)\mathrm{d}u\right)\Big|\sigma(\hat{\tau}_L)\right]\right], \qquad (4.3.9)
\end{aligned}$$

where we first used Proposition 3.4.18 (7) (conditional expectation on a trivial σ-algebra), then Proposition 3.4.18 (5) (the tower property) and finally Proposition 3.4.18 (6) and (7) (take out what is known and conditional expectation on a trivial σ-algebra). By the definition of conditional expectation, $\mathbb{E}_{\mathbb{Q}}\left[\exp\left(-\int_0^{\hat{\tau}_L}\gamma(B_u)\mathrm{d}u\right)|\sigma(\hat{\tau}_L)\right]$ is $\sigma(\hat{\tau}_L)$-measurable. Thus, by the Factorization Lemma (3.4.20), there exists an η such that

$$\eta(\hat{\tau}_L) \;=\; \mathbb{E}_{\mathbb{Q}}\left[\exp\left(-\int_0^{\hat{\tau}_L}\gamma(B_u)\mathrm{d}u\right)\Big|\sigma(\hat{\tau}_L)\right],$$

which, once we condition on the event $\{\hat{\tau}_L = t\}$ gives

$$\eta(t) \;:=\; \mathbb{E}_{\mathbb{Q}}\left[\exp\left(-\int_0^{t}\gamma(B_u)\mathrm{d}u\right)\Big|\{\hat{\tau}_L = t\}\right]. \qquad (4.3.10)$$

Plugging this back into Equation (4.3.9), combining it with Equation (4.3.8) and using that $B_{\hat{\tau}_L} = L$ gives us

$$\mathbb{E}_{\mathbb{P}}[\psi(\tau_L)\mathbb{1}_{\{\tau_L<\infty\}}] \;=\; \exp\left(\beta(B_{\hat{\tau}_L}) - \beta(x)\right)\mathbb{E}_{\mathbb{Q}}\left[\psi(\hat{\tau}_L)\eta(\hat{\tau}_L)\right], \qquad (4.3.11)$$

which means that, essentially, we get all the information about the distribution of τ_L from the functional form of η. However, the form of Equation (4.3.10) is not that practical since it is being conditioned on what we are trying to determine: the first hitting time. We now rewrite Equation (4.3.10) in such a way that we can exploit the two pieces of information we have about the process: where it starts, x, and where it ends, L.
Let $\rho^{0,x}$ denote a three-dimensional Bessel process that, at time 0 is at point x. We recall from Definition 3.6.16 that one can connect the Bessel process and the Brownian motion which we exploit now to write:

$$\rho_s^{0,x} \;=\; \left\|\tilde{B}_s^{0,a}\right\|,$$

where $\|\cdot\|$ represents the Euclidean norm and $\|a\| = x \geq 0$. Writing it out component-wise, in the spirit of Definition 3.6.15, we can say that

$$\tilde{B}_s^{0,a} \;=\; a + (B_s^1, B_s^2, B_s^3) \text{ and } a \in \mathbb{R}^3.$$

We further define $(R_s^{x\to L})_{s\in[0,t]}$ to be the Bessel bridge from a point x to L . Note that we are working with respect to the measure $\mathbb{Q}$ so recall that now $\hat{\tau}_L = \inf\{s \geq 0\colon B_s^{0,x} = L\}$. In order to apply Proposition 3.6.18, we now shift the Bessel bridge to begin at 0 and end at $L - x$, where

we recall that, naturally, $L > x$. Furthermore, applying a time reversal allows us to go from a supremum to the infimum in Proposition 3.6.18 so we can connect the Bessel bridge to the Brownian motion. Explicitly, defining also $\tilde{\tau}_L := \{s \geq 0 \colon \rho_s^{0,x} = L\}$, we write:

$$\begin{aligned}(R_s^{0\to L-x})_{s\in[0,t]} &= \{L - \left\|\tilde{B}_{t-s}^{0,a}\right\| \colon 0 \leq s \leq \tilde{\tau}_L\} \text{ given } \{\tilde{\tau}_L = t\} \\ &\overset{\mathrm{d}}{=} \{L - B_{t-s} \colon 0 \leq s \leq \hat{\tau}_L\} \text{ given } \{\hat{\tau}_L = t\},\end{aligned}$$

which essentially says that, given $\{\tau_L = t\}$,

$$R_s := L - B_{t-s}, \tag{4.3.12}$$

is a Bessel bridge going from 0 to $L - x$ for $s \in [0, t]$. Note again that $B_0 = x$ in our case. Plugging Equation (4.3.12) into Equation (4.3.10) gives us

$$\eta(t) := \mathbb{E}_{\mathbb{Q}}\left[\exp\left(-\int_0^t \gamma(L - R_u)\mathrm{d}u\right)\middle| R_t = L - x\right], \tag{4.3.13}$$

which concludes our proof. □

Remark 4.3.2 (On notation for the first hitting time)**.** *In this section, we made sure to differentiate between first hitting times on the basis of the different measures under which we are observing it. However, the common practice is to use a single notation and let the reader deduce from the context with respect to which measure we are viewing the first hitting time: outside of this section, as well as in the statements of the main results, this is the approach we adopt.*

The result of Propositon 4.3.1 thus shows that simulating the first hitting time becomes straightforward once the function η is well-known. However, this is not the case in practice, which is why we need an algorithm in order to deal with η. This is explored in Section 4.4 in which we also prove Theorem 2.0.5.

4.4 Justifying the Algorithm

Throughout this section, we will be operating under the following assumption:

Assumption 4.4.1 ($\gamma \geq 0$ and $\tau_L < \infty$ a.s.)**.** *We assume that γ, defined in Equation (4.2.8), is non-negative and the first-hitting time τ_L is almost surely finite.*

We have already shown through Claim 4.2.9 that $\tau_L < \infty$ a.s. for a particular set of conditions. However, we can show this claim holds true also for the following:

Claim 4.4.2 ($\tau_L < \infty$ a.s. for $\lim_{x\to-\infty} v(x) < \infty$)**.** *Let $v(x)$ be defined as in Proposition 3.6.14. Choosing $c = 0$, and working on the interval $I = (-\infty, L)$ while being mindful that $\sigma \equiv 1$, we get that:*

$$v(x) := \int_0^x \int_0^y \frac{2p'(y)}{p'(z)}\mathrm{d}z\mathrm{d}y. \tag{4.4.1}$$

It follows that, if $\lim_{x\to-\infty} v(x) < \infty$, then $\tau_L < \infty$ a.s.

Proof of Claim 4.4.2. We will show this is true by using Proposition 3.6.14 (i). First of all, note that we have already demonstrated through the proof of Claim 4.2.7 that nondegeneracy and local integrability hold for our set of conditions. Furthermore, from the assumption, it holds that $v(-\infty) < \infty$. Hence, we only have to show that $v(L) < \infty$. We first recall the definition of $p(x)$ which we slightly rewrite to make it clear what are the "dummy variables" in the integration.

$$p(x) \;=\; \int_0^x \exp\left(-\int_0^\zeta b(\theta)\mathrm{d}\theta\right)\mathrm{d}\zeta.$$

We now use Theorem A.0.7 and the fact that for $x, y, z \in I$, as before, b can be bounded by some $M > 0$, to write:

$$\begin{aligned} p'(y) &\leq \exp\left(-\int_0^y (-M)\mathrm{d}\theta\right) = \exp\left(My\right) \\ p'(z) &\leq \exp\left(-\int_0^z (-M)\mathrm{d}\theta\right) = \exp\left(Mz\right), \end{aligned}$$

from which it follows that

$$\frac{2p'(y)}{p'(z)} \;\leq\; 2\exp\left(My\right)\exp\left(-Mz\right),$$

which, after plugging into Equation (4.4.1) gives:

$$\begin{aligned} v(L) &\leq \int_0^L \left(\int_0^y 2\exp\left(My\right)\exp\left(-Mz\right)\mathrm{d}z\right)\mathrm{d}y \\ &= -\frac{2}{M}\int_0^L \left(1-\exp\left(My\right)\right)\mathrm{d}y \\ &= -\frac{2}{M}\left[L - \frac{1}{M}\left(\exp\left(ML\right)-1\right)\right] < \infty, \end{aligned}$$

due to all the objects and their functions being bounded. Finally, from Proposition 3.6.14 it follows that $S < \infty$ a.s. and thus $\tau_L < \infty$ a.s. which completes our proof. □

Claim 4.4.3 ($\eta(t)$ is a probability of rejection in the algorithm)**.** *Given $\gamma \geq 0$, it follows that $\eta(t) \in [0, 1]$.*

Proof of Claim 4.4.3. We refer to Equation (4.3.13):

$$\eta(t) \;\coloneqq\; \mathbb{E}_{\mathbb{Q}}\left[\exp\left(-\int_0^t \gamma(L-R_u)\mathrm{d}u\right)\middle| R_t = L - x\right].$$

In particular, our object of interest is $\exp\left(-\int_0^t \gamma(L-R_u)\mathrm{d}u\right)$. Since $\gamma \geq 0$, it follows that we have a decaying exponential for $t \geq 0$, thus $\eta(t) \in [0, 1]$. □

We move now to proving Theorem 2.0.5 which we now restate for the reader's convenience as Theorem 4.4.4.

Theorem 4.4.4. *([HeZu1, Theorem 2.3]) If γ is a non-negative function then the density f_Y of the outcome variable Y, as described in* THE ALGORITHM *of Section 4.1, satisfies*

$$f_Y(t) \;=\; \frac{1}{\Xi}\eta(t)f_T(t), \tag{4.4.2}$$

where η is given by Equation (2.0.6) and Ξ stands for the normalization coefficient

$$\Xi \coloneqq \int_0^\infty \eta(t) f_T(t) \mathrm{d}t. \tag{4.4.3}$$

In particular, under Assumptions 4.2.4 and 4.4.1, if $\frac{T}{(L-x)^2}$ has the same distribution as $\frac{1}{G^2}$, where G is a standard Gaussian random variable, then Y and τ_L, defined as the first hitting time in Definition 4.2.1 are identically distributed.

Example 4.4.5 (The special case)**.** *Before we move to the formal proof, it is wortwhile to look at the normalization coefficient when $T \sim \frac{(L-x)^2}{G^2}$ which is a case familiar to us from Proposition 4.2.2. We refer to Equation (4.3.1) where we set $\psi \equiv 1$ from which it follows that*

$$\begin{aligned} \mathbb{E}_P\left[\mathbb{1}_{\{\tau_L < \infty\}}\right] &= \mathbb{E}_{\mathbb{Q}}[\eta(\tau_L)] \exp\left(\beta(L) - \beta(x)\right) \\ \implies \mathbb{P}(\tau_L < \infty) &= \Xi \exp\left(\beta(L) - \beta(x)\right) \\ \implies \Xi &= \exp\left(\beta(x) - \beta(L)\right), \end{aligned}$$

where we used the definition of the expected value, as well as Equation (4.4.3), and our result from the Feller Test (Claim 4.4.2).

We now proceed to the proof.

Proof of Theorem 4.4.4. Let $\mathcal{I}$ be the number of iterations of THE ALGORITHM. For each iteration $\mathcal{I} = i \in \mathbb{N}$, we denote the following generated random variables:

1. $T^{(i)}$ - a non-negative random variable with density f_T
2. $R_t^{(i)}$ - a 3-dimensional Bessel process on the interval $[0, T]$ and with endpoint $R_T = L - x$
3. $N^{(i)}$ - a Poisson point process with the Lebesgue intensity measure.

It follows from the construction that each iteration results in random variables that are independent of the other iterations. As in Proposition 4.3.1, we let $\psi \colon \mathbb{R} \to \mathbb{R}$ be a bounded and measurable function. We recall also the following identity:

$$1 = \mathbb{1}_\Omega(\omega) = \sum_{i=1}^\infty \mathbb{1}_{\{\mathcal{I} = i\}},$$

which we use below

$$\begin{aligned} \mathbb{E}[\psi(Y)] &= \mathbb{E}\left[\psi(Y) \sum_{i=1}^\infty \mathbb{1}_{\{\mathcal{I} = i\}}\right] \\ &= \mathbb{E}\left[\sum_{i=1}^\infty \psi(Y) \mathbb{1}_{\{\mathcal{I} = i\}}\right] \\ &= \sum_{i=1}^\infty \mathbb{E}\left[\psi(Y) \mathbb{1}_{\{\mathcal{I} = i\}}\right], \end{aligned} \tag{4.4.4}$$

where we used the fact that $\psi(Y)$ is bounded and measurable, with no dependence on i, so we can push it into the sum. In the final step, we used the linearity of expectation to simplify the

expression. Recall that, in the algorithm, we said that we have generated Y if $N(D_{R,T}) = 0$, so let us define the set E_i which indicates our failure to generate the random variable at the i^{th} step:

$$E_i \;\coloneqq\; \{N^{(i)}(D_{R^{(i)},T^{(i)}}) \neq 0\}, \tag{4.4.5}$$

where $D_{R^{(i)},T^{(i)}}$ is the stochastic domain generated at the i^{th} step (refer to Equation (4.1.1)). Note that the family of sets $(E_i)_{i\in\mathcal{I}}$ is independent due to the nature of the iterations and the setup. A set that signifies the successful generation of the random variable Y at the i^{th} step is the complement set:

$$E_i^{\mathsf{c}} \;=\; \Omega \setminus E_i.$$

In the following, we assume that we successfully generated the random variable Y for the first time on the i^{th} try. This then implies that $Y = T^{(i)}$. Using this insight, along with plugging Equation (4.4.5) into Equation (4.4.4), gives us:

$$\mathbb{E}[\psi(Y)] \quad = \quad \sum_{i=1}^{\infty} \mathbb{E}\left[\psi(T^{(i)})\mathbb{1}_{E_1 \cap E_2 \cap \ldots \cap E_{i-1} \cap E_i^{\mathsf{c}}}\right],$$

where the event $E_1 \cap E_2 \cap \ldots \cap E_{i-1} \cap E_i^{\mathsf{c}}$ "encodes" our previous statement that the first successful iteration was the i^{th} one whereas all the other ones resulted in failure. It further follows from Remark 3.3.4 that $\mathbb{1}_{E_1 \cap E_2 \cap \ldots \cap E_{i-1} \cap E_i^{\mathsf{c}}} = \mathbb{1}_{E_1} \cdot \mathbb{1}_{E_2} \cdots \mathbb{1}_{E_{i-1}} \cdot \mathbb{1}_{E_i^{\mathsf{c}}}$. Finally, using Proposition 3.4.14, we get:

$$\mathbb{E}[\psi(Y)] \quad = \quad \sum_{i=1}^{\infty} \mathbb{E}\left[\psi(T^{(i)})\mathbb{1}_{E_i^{\mathsf{c}}}\right] \mathbb{P}(E_1) \cdots \mathbb{P}(E_{i-1}).$$

Now, observe that each of the events $(E_j)_{j=1}^{i-1}$ is equally likely since, every time we iterate, we restart the simulation. A good analogy is with tossing a coin: every particular time we toss a coin, there is always a 50% chance to get heads. Using this idea, we rewrite our expression as:

$$\begin{aligned}
\mathbb{E}[\psi(Y)] \quad &= \quad \sum_{i=1}^{\infty} \mathbb{E}\left[\psi(T^{(1)})\mathbb{1}_{E_1^{\mathsf{c}}}\right] \mathbb{P}(E_1)^{i-1} \\
&= \quad \mathbb{E}\left[\psi(T^{(1)})\mathbb{1}_{E_1^{\mathsf{c}}}\right] \sum_{i=1}^{\infty} \mathbb{P}(E_1)^{i-1} \\
&= \quad \mathbb{E}\left[\psi(T^{(1)})\mathbb{1}_{E_1^{\mathsf{c}}}\right] \frac{1}{1-\mathbb{P}(E_1)} \\
&= \quad \mathbb{E}\left[\psi(T^{(1)})\mathbb{1}_{E_1^{\mathsf{c}}}\right] \frac{1}{\mathbb{P}(E_1^{\mathsf{c}})} \\
&= \quad \frac{1}{\mathbb{P}(N(D_{R^{(1)},T^{(1)}} = 0))} \mathbb{E}\left[\psi(T^{(1)})\mathbb{1}_{E_1^{\mathsf{c}}}\right] \\
&= \quad \frac{1}{\mathbb{P}(N(D_{R,T} = 0))} \mathbb{E}\left[\psi(T)\mathbb{1}_{E^{\mathsf{c}}}\right],
\end{aligned} \tag{4.4.6}$$

where we used the geometric series formula and the fact that $\mathbb{P}(E_1) \in (0,1)$ to simplify the expression. In the last step, we dropped the iteration number for simplicity. We now

focus on $\mathbb{E}\left[\psi(T)\mathbb{1}_{E^{\mathsf{c}}}\right]$ which we can simplify using similar methods as we did with $\eta(t)$ in Equation (4.3.9).

$$\begin{aligned}
\mathbb{E}\left[\psi(T)\mathbb{1}_{E^{\mathsf{c}}}\right] &\overset{\text{Proposition 3.4.18 (7)}}{=} \mathbb{E}\left[\psi(T)\mathbb{1}_{E^{\mathsf{c}}}\middle|\{\emptyset,\Omega\}\right] \\
&\overset{\text{Proposition 3.4.18 (5)}}{=} \mathbb{E}\left[\mathbb{E}\left[\psi(T)\mathbb{1}_{E^{\mathsf{c}}}\middle|\sigma(T)\right]\middle|\{\emptyset,\Omega\}\right] \\
&\overset{\text{Proposition 3.4.18 (6)}}{=} \mathbb{E}\left[\psi(T)\mathbb{E}\left[\mathbb{1}_{E^{\mathsf{c}}}\middle|\sigma(T)\right]\middle|\{\emptyset,\Omega\}\right] \\
&\overset{\text{Definition 3.4.19}}{=} \mathbb{E}\left[\psi(T)\mathbb{P}(N_{D_{R,T}}=0|T)\right]. \qquad (4.4.7)
\end{aligned}$$

We now simplify $\mathbb{P}(N_{D_{R,T}}=0|T)$ even further.

$$\begin{aligned}
\mathbb{P}(N_{D_{R,T}}=0|T) &\overset{\text{Definition 3.4.21}}{=} \mathbb{E}\left[\mathbb{1}_{\{N_{D_{R,T}}=0\}}\middle|\sigma(T)\right] \\
&\overset{\text{Proposition 3.4.18 (5)}}{=} \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_{\{N_{D_{R,T}}=0\}}\middle|\sigma(R,T)\right]\middle|\sigma(T)\right] \\
&\overset{\text{Definition 3.4.19}}{=} \mathbb{E}\left[\mathbb{P}(N_{D_{R,T}}=0|R,T)\middle|T\right] \\
&\overset{\text{Definition 3.5.24}}{=} \mathbb{E}\left[\exp\left(-\lambda(D_{R,T})\right)\middle|T\right],
\end{aligned}$$

so now we have to find the Lebesgue measure of the stochastic domain in Equation (4.1.1). This is

$$\begin{aligned}
\lambda(D_{R,T}) &= \int_0^T \left(\int_0^{\gamma(L-R_t)} 1\mathrm{d}\lambda(\nu)\right)\mathrm{d}\lambda(t) \\
&= \int_0^T \gamma(L-R_t)\mathrm{d}t.
\end{aligned}$$

This further implies that

$$\begin{aligned}
\mathbb{P}(N_{D_{R,T}}=0|T) &= \mathbb{E}\left[\exp\left(-\int_0^T \gamma(L-R_t)\mathrm{d}t\right)\middle|T\right] \\
&= \mathbb{E}\left[\exp\left(-\int_0^T \gamma(L-R_t)\mathrm{d}t\right)\middle|R_T=L-x,T\right] \\
&= \eta(T), \qquad (4.4.8)
\end{aligned}$$

where the last step follows directly from the setup of the algorithm where it was assumed that $R_T = L - x$ so we naturally construct a Bessel bridge and notice the connection to Equation (4.3.13). Finally, plugging Equation (4.4.8) into Equation (4.4.7) and combining with Equation (4.4.6), we get:

$$\mathbb{E}[\psi(Y)] = \frac{1}{\mathbb{P}(N(D_{R,T}=0))}\mathbb{E}[\psi(T)\eta(T)].$$

Notice now when we set $\psi \equiv 1$, we get exactly:

$$\mathbb{P}(N(D_{R,T}=0)) = \mathbb{E}[\eta(T)],$$

which, together with Example 4.4.5, leads us to Equation (4.4.2) as well as the theorem's claim. □

Remark 4.4.6 (Did Theorem 4.4.4 give us τ_L?)**.** *It might have been easy to miss it, but the crucial idea behind Theorem 4.4.4 is that it is indeed possible to use* THE ALGORITHM *to generate a random variable which is distributed as the first hitting time. This is a meaningful note to make: we did not* get *the actual first hitting time, but rather a random variable which is* distributed *like it.*
Going back to the coin analogy, one can think about it as follows. Let us say you are in a group of friends and there are two options to choose from, but you cannot really make up your mind so it comes down to you to toss a coin. You look into your pocket expecting to find a standard US quarter, but for some reason, you find one Finnish markka in your pocket. Do you give up on the coin toss? Absolutely not. The Finnish markka can perfectly "simulate" a US quarter despite not being one.

While we have shown in Section 4.3 and Section 4.4 that it is possible to construct an algorithm that exactly simulates the first hitting time of Equation (4.2.1), we still need to consider the feasibility of practically using this algorithm: this is the topic we briefly discuss in Section 4.5.

4.5 Algorithm Limitations and Proposed Solutions

The problem with THE ALGORITHM is the fact that it is rather difficult to achieve in practice. The problem is how do we deal with $D_{R,T}$ since the stochastic domain depends on the whole path of $(R_t)_{t\geq 0}$?
Herrmann and Zucca propose solutions in the forms of Algorithm (A1) and Algorithm (A2) in [HeZu1, Page 1482-1486]. While we will not go into the details, we do mention the key insight that allows us to modify THE ALGORITHM in a way that it becomes practically usable. The big idea is to reduce the domain to a bounded one which resolves the problem of having to simulate the Poisson point process on something which is essentially an infinite state space. As for the second difficulty, the only thing we actually need to know is whether or not a particular point is within the stochastic domain. Through a modification of how we simulate the Poisson point process, one can avoid having to deal with the full paths of the Bessel process $(R_t)_{t\geq 0}$. Finally, we refer the interested reader to [HeZu1, Section 3,6] for a discussion of the efficacy of the different algorithms.

4.6 Conclusion and Further Avenues of Research

Due to its inherent applicability in a wide range of fields, there is very active research being done when it comes to exact simulations. An interesting contribution to the field would be a generalization of the problem to stochastic processes with jumps. Furthermore, as proposed in [HeZu1, Pag 1482], one could also modify the algorithm to not only deal with time-homogeneous processes, but also a wider range of processes. For an overview of the work currently being done in the field, we refer the reader to [HeMa1, HeMa2, HeZu2, HeZu3].

Appendix A

An Overview of Standard Results from Analysis

We present for the reader some essential results in analysis that are implicitly or explicitly used throughout this thesis. Most of these results are taken from [Fo1, Kl, Mo, RoFi, Ru] which, on their own, provide an excellent overview of many tools used in modern analysis.

Definition A.0.1 (Topology)**.** *Let $\Omega \neq \emptyset$ be an arbitrary set. A class of sets $\tau \subset 2^{\Omega}$ is called a* topology *on Ω if the following properties hold:*

1. $\emptyset, \Omega \in \tau$
2. $A \cap B \in \tau$ *for any* $A, B \in \tau$
3. $(\bigcup_{A \in \mathcal{F}} A) \in \tau$ *for any* $\mathcal{F} \subset \tau$

The pair (Ω, τ) is called a topological space*. The sets $A \in \tau$ are called* open*, and the sets $A \subset \Omega$ with $A^{\mathsf{c}} \in \tau$ are* closed*.*

Definition A.0.2 (Metric space)**.** *A metric space is a nonempty set X equipped with a function*

$$d \colon X \times X \to [0, \infty)$$

(a metric) *that satisfies the following three conditions:*

(i) $d(x, y) = d(y, x)$ *for all* $x, y \in X$.

(ii) (The triangle inequality) $d(x, z) \leq d(x, y) + d(y, z)$ *for all* $x, y, z \in X$.

(iii) $d(x, y) = 0$ *if and only if* $x = y$.

Definition A.0.3 (Totally bounded)**.** *A subset E of a metric space X is called* totally bounded *if for every $\epsilon > 0$, E is contained in the union of finitely many balls of radius ϵ.*

Theorem A.0.4. *(On compactness [Fo2, 1.10 Theorem]) If E is a subset of a metric space X, the following conditions are equivalent.*

(a) *E is complete and totally bounded.*

(b) *Every sequence in E has a subsequence that converges to a point of E.*

(c) *For every collection $\mathcal{U}$ of open sets in X such that $E \subset \bigcup_{U\in\mathcal{U}} U$, there is a finite subcollection $U_1, \ldots, U_n$ such that $E \subset \bigcup_{j=1}^{n} U_j$. (In brief: every open cover of E has a finite subcover.)*

The fact that (a) $\Longrightarrow$ (b) is the Bolzano-Weierstrass theorem, and the fact that (a) $\Longrightarrow$ (c) is the Heine-Borel theorem. A set E that possesses the properties (a)–(c) is called compact.

Theorem A.0.5. *(Leibniz's Rule [Mo, Page 82]) Suppose that $\int_{a(t)}^{b(t)} f(x,t)\mathrm{d}x$ exists, that $a(t), b(t)$ and $f(x,t)$ are all continuously differentiable with respect to t, and that there is a function $g(x) \geq \left|\frac{\partial f}{\partial t}(x,t)\right|$ with $\int_{a(t)}^{b(t)} g(x,t)\mathrm{d}x$ bounded. Then:*

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{a(t)}^{b(t)} f(x,t)\mathrm{d}x \;=\; \int_{a(t)}^{b(t)} \frac{\partial f}{\partial t}(x,t)\mathrm{d}x + f(b(t),t)b'(t) - f(a(t),t)a'(t).$$

We allow $a = -\infty$ or $b = +\infty$.

Theorem A.0.6. *(Mean Value Theorem [Ru, 5.10 Theorem]) If f is a real continuous function on $[a,b]$ which is differentiable in (a,b), then there is a point $c \in (a,b)$ at which*

$$f(b) - f(a) \;=\; (b-a)f'(c)$$

Theorem A.0.7. *(Fundamental Theorem of Calculus [Mo, Theorem 16.1]) let f be a continuous function on $[a,b]$. The two following statements hold true:*

1. $\frac{\mathrm{d}}{\mathrm{d}b}\int_a^b f(x)\mathrm{d}x \;=\; f(b)$
2. *If $f(x) \;=\; F'(x)$, then $\int_a^b f(x)\mathrm{d}x \;=\; F(x)|_a^b \;=\; F(b) - F(a)$.*

Theorem A.0.8. *(Extreme Value Theorem [Si, Theorem 4.2.9]) Let A be a compact subset of $\mathbb{R}$ and let $f\colon A \to \mathbb{R}$ be a continuous function. Then, f has a maximum and a minimum value on A.*

Appendix B

Relevant Distributions

B.1 Uniform Distribution

Let Ω be a finite nonempty set. By

$$\mu(A) \coloneqq \frac{\#A}{\#\Omega},$$

for $A \subset \Omega$, we define a probability measure on $\mathcal{F} = 2^{\Omega}$. This μ is called the *uniform distribution* on Ω. For this distribution, we introduce the symbol $\mathcal{U}_{\Omega} \coloneqq \mu$. ([Kl, Example 1.30(ii)]).

B.2 Gaussian (Normal) Distribution on $\mathbb{R}$

A Gaussian (Normal) distribution with mean $m \in \mathbb{R}$ and variance σ^2 is denoted by $\mathcal{N}(m, \sigma^2)$. Its density function is

$$f(x) \coloneqq \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right).$$

Bibliography

[BeRo] A. Beskos and G. O. Roberts, *Exact Simulation of Diffusions*, The Annals of Applied Probability Vol. 15 No.4 2422-2444, 2005

[BhWa] R. Bhattacharya and E. C. Waymire, *Random Walk, Brownian Motion, and Martingales*, Springer, 2021

[BoSa] A. N. Borodin and P. Salminen, *Handbook of Brownian Motion - Facts and Formulae*, Springer, Second Edition, 2002

[Bi1] P. Billingsley, *Probability and Measure*, John Wiley and sons ltd, Third Edition, 2008

[Bi2] P. Billingsley, *Convergence in Probability Measures*, John Wiley and Sons, INC, Second Edition, 1999

[Çı] E. Çınlar, *Probability and Stochastics*, Springer, 2011

[CoTa] R. Cont and P. Tankov, *Financial Modelling with Jump Processes*, Chapman & Hall/CRC, 2004

[ChWi] K. L. Chung and R.J. Williams, *Introduction to Stochastic Integration*, Springer, Second Edition, 1990

[Du] R. Durrett, *Stochastic Calculus A Practical Introduction*, CRC Press, 1996

[Fe] W. Feller, *An Introduction to Probability Theory and Its Applications: Volume 1*, John Wiley and Sons, INC, Second Edition, 1968

[Fo1] G. B. Folland, *Real Analysis: Modern Techniques and Their Applications*, John Wiley and Sons, INC, Second Edition, 1999

[Fo2] G. B. Folland, *A Guide To Advanced Real Analysis*, Mathematical Association of America, Washington, DC, 2009

[Ga] T. A. Garrity, *All the Math you Missed (But Need to Know for Graduate School)*, Cambridge University Press, Second Edition, 2021.

[Ge1] H. Geiss and S. Geiss, *An Introduction to Probability Theory*, University of Jyväskylä, 2023

[Ge2] S. Geiss, *Stochastic Differential Equations*, Lecture Notes, University of Jyväskylä, 2023

[Gi] I. V. Girsanov, *On Transforming a Certain Class of Stochastic Processes by Absolutely Continuous Substition of Measures*, Theory Probab. Appl. **5** (285-301), 1960

[GrSt] G. Grimmett and D. Stirzaker, *Probability and Random Processes*, Oxford University Press, Third Edition, 2001

[Ha] P. R. Halmos, *Measure Theory*, Springer-Verlag New York, 1974

[HeMa1] S. Herrmann and N. Massin, *Approximation of exit times for one-dimensional linear diffusion processes*, Computers & Mathematics with Applications. An International Journal. (2020) **80**:1668-1682, 2020

[HeMa2] S. Herrmann and N. Massin, *Exact simulation of the first passage time through a given level of jump diffusions*, Mathematics and Computers in Simulation, (2023) **203**: 553-576, 2023

[HeZu1] S. Herrmann and C. Zucca, *Exact Simulation of the First-Passage Time of Diffusions*, Journal of Scientific Computing (2019) **79**:1477-1504, 2019

[HeZu2] S. Herrmann and C. Zucca, *Exact simulation of first exit times for one-dimensional diffusion processes*, ESAIM. Mathematical Modelling and Numerical Analysis (2020) **54**:811-844, 2020

[HeZu3] S. Herrmann and C. Zucca, *Exact simulation of diffusion first exit times: algorithm acceleration*, Journal of Machine Learning Research (JMLR) (2022) **23**: Paper No. [138], 20, 2022

[IcKa] T. Ichiba and C. Kardaras, *Efficent Estimation of One-Dimensional Diffusion First Passage Time Densities via Monte Carlo Simulations*, Journal of Applied Probability **48**, 699-712, 2011

[IkWa] N. Ikeda and S. Watanabe, *Stochastic Differential Equations and Diffusion Processes*, North-Holland Publishing Company and Kodanska Ltd. Tokyo, Second Edition, 1989

[JeYoCh] M. Jeanblanc, M. Yor, M. Chesney, *Mathematical Methods for Financial Markets*, Springer London, 2009

[Ka] O. Kallenberg, *Foundations of Modern Probability*, Springer, 1997

[KaRu] I. Karatzas and J. Ruf, *Distribution of the time to explosion for one-dimensional diffusions*, Probability Theory and Related Fields (2016) **164**: 1027-1069, 2016

[KaSh1] I. Karatzas and S.E. Shreve, *Brownian Motion and Stochastic Calculus*, Springer, Second Edition, 1991

[KaSh2] I. Karatzas and S.E. Shreve, *Methods of Mathematical Finance*, Springer, Second Edition, 1998

[KaSu] G. Kallianpur and P. Sundar, *Stochastic Analysis and Diffusion Processes*, Oxford University Press, 2014

[Kl] A. Klenke, *Probability Theory: A Comprehensive Course*, Springer, Second Edition, 2014

[KlPl] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer, 1995

[La] G.F. Lawler, *Conformally Invariant Processes in the Plane*, American Mathematical Society, 2005

[LaLa] D. Lamberton and B. Lapeyre, *Introduction to Stochastic Calculus Applied to Finance*, Chapman & Hall/CRC, Second Edition, 2008

[LeGa] J-F. Le Gall, *Brownian Motion, Martingales, and Stochastic Calculus*, Springer, 2016

[Mi] S. J. Miller, *The Probability Lifesaver: All the tools you need to understand chance*, Princeton University Press, 2017

[Mo] F. Morgan, *Real Analysis*, American Mathematical Society, 2005

[PaRă] E. Pardoux and A. Răşcanu, *Stochastic Differential Equations, Backward SDEs, Partial Differential Equations*, Springer, 2014

[PeSh] G. Peskir and A. Shiryaev, *Optimal Stopping and Free-Boundary Problems*, Birkhäuser Verlag, 2006

[RoFi] H. L. Royden, P. M. Fitzpatrick, *Real analysis*, Pearson Education, Fourth Edition, 2010

[ReYo] D. Revuz and M. Yor, *Continuous Martingales and Brownian Motion*, Springer, Third Edition, 2005

[RoWi1] L. C. G. Rogers and D. Williams, *Diffusions, Markov Processes, and Martingales. Volume 1: Foundations*, John Wiley and sons, Second Edition, 1994

[RoWi2] L. C. G. Rogers and D. Williams, *Diffusions, Markov Processes, and Martingales. Volume 2: Itô Calculus*, John Wiley and sons, Second Edition, 2000

[Ru] W. Rudin, *Principles of Mathematical Analysis*, McGraw-Hill, Inc, Third Edition, 1976

[ScPaBö] R. L. Schilling, L. Partzsch and B. Böttcher, *Brownian Motion An Introduction to Stochastic Processes*, De Gruzter Graduate, 2012

[Si] C. E. Silva, *Invitation to Real Analysis*, American Mathematical Society, 2019

[Shi] A. N. Shiryaev, *Probability*, Springer, Second Edition, 1996

[Shr1] S. E. Shreve, *Stochastic Calculus for Finance I: The Binomial Asset Pricing Model*, Springer, 2003

[Shr2] S. E. Shreve, *Stochastic Calculus for Finance II: Continuous-Time Models*, Springer, 2004

[So] K. Sobczyk, *Stochastic Differential Equations with Applications to Physics and Engineering*, Springer, 1991

[St] J. M. Steele, *Stochastic Calculus and Financial Applications*, Springer, 2001

[SäSo] S. Särkkä and A. Solin, *Applied Stochastic Differential Equations*, Cambridge University Press, 2019

[Wi1] D. Williams, *Probability with martingales*, Cambridge Mathematical Textbooks, Cambridge University Press, 1991

[Wi2] D. Williams, *Path Decomposition and Continuity of Local Time for One-Dimensional Diffusions, I*, Proc. London Math. Soc. (3) 28 738-768, 1974