SJÄLVSTÄNDIGA ARBETEN I MATEMATIK

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET

Finding a Precursor to Rare Events in a Dynamical System

av

Olle Bergström Jonsson

2020 - No K17

Finding a Precursor to Rare Events in a Dynamical System

Olle Bergström Jonsson

Självständigt arbete i matematik 15 högskolepoäng, grundnivå

Handledare: Woosok Moon

2020

Acknowledgements

I would like to offer my sincere gratitude to my supervisor Woosok Moon for providing me with guidance and assistance throughout this project. He encouraged me when I felt overwhelmed, and helped me overcome my obstacles. I would also like to thank Ludovico Giorgini for his tips and guidance along the way.

Abstract

The goal of this thesis is to provide an understanding of the basics of stochastic calculus, in particular stochastic differential equations, and to briefly discuss the theory of path integrals, in order to explore and give a basic outline of how they can be used to find a precursor to a rare event in a dynamical system.

Contents

1 Introduction

This thesis will explore how one can use Stochastic Differential Equations (SDE) to solve the problem of predicting a rare event in a dynamical system. We will also briefly look at the formulation of path integrals, and how they can be used to optimize the path taken from one steady state to another when such an event occurs.
A dynamical system is here interpreted as the movement of a Brownian particle over time, realized by a stochastic differential equation.
We will be discussing some parts of ordinary calculus such as Ordinary Differential Equations (ODE), as well as basic probability theory to get a starting point for the discussion of the stochastic differential equations.

After introducing the SDE's, we will go on to look at the theory of the Itô calculus and see that Ito's lemma is the stochastic calculus' version of the chain rule for differentiation in ordinary calculus.

We will then discuss the theory of path integrals in physics, and lastly perform some numerical simulations of a dynamical system exhibiting rare events.

The underlying problem stems from the previous research of (Giorgini et al., 2019), presented in the text Predicting Rare Events in Stochastic Resonance. This publication was updated after this thesis was started, and the title was changed to *Precursors to Rare Events in Stochastic Resonance*, but the version used for this thesis is the old one. This theory has a wide range of applications, for example the detection of natural disasters such as earthquakes or floods.

2 Background

In stochastic calculus we encounter stochastic differential equations. To be able to discuss this subject, we will begin by introducing some of the underlying theories behind the SDE's, such as ordinary differential equations and Brownian motion.

There are many fields in which stochastic differential equations are used. In finance and insurance they play a vital role in the simulations needed to provide adequate predictions and realistic models, whereas in physics they may be used to study for example stochastic resonance, which can be described as the phenomena where one increases the strength of a weak signal by introducing white noise to it. In this thesis we will briefly discuss the application of SDE to financial matters, while the main part will cover some applications of the study of stochastic resonance, and see how we can use the SDE to model this kind of phenomena.

2.1 Ordinary Differential Equations

In ordinary calculus we encounter the Ordinary Differential Equations (ODEs) as a tool to study evolutionary processes which possess certain properties. One of these properties is determinacy, which implies that all the past and future states of the processes can be determined by their current states. As a simple example of such an evolutionary process, consider the steady increase of money deposited (one time deposit) in a bank account with a certain set interest. Denote the initial amount deposited by X_0, and let the rate of change (i.e interest) of X be $X' = a$. To answer the question of how much money we will have in our bank account after t years, we could, assuming we have a continuous interest rate, set up the very simple ordinary differential equation $\frac{dX}{dt} = aX$ which turns out to have the solution $X(t) = \frac{A}{a}e^{at}$, where $\frac{A}{a} = X_0$.

In general, a linear first order ordinary differential equation has the form

$$y'(t) + y(t)g(t) = h(t).$$

With the initial condition $y(t_0) = y_0$ this equation can be expressed

$$\frac{dy}{dt} + g(t)y = h(t), \quad y(t_0) = y_0. \tag{2.1.1}$$

One method to solve this equation is to realize that the left hand side of this equation contains the derivative of y as a function of t, and also the function y itself, which points us in the direction of the product rule of derivation. The product rule states that

$$\frac{d}{dt}(f(t)g(t)) = f'(t)g(t) + f(t)g'(t) = \frac{df}{dt}g(t) + f(t)\frac{dg}{dt}.$$

However, the left hand side of (2.1.1) is not precisely on this form yet, so we need to modify it somehow to fit the description of the product rule if we want to use this method for solving the ODE.

So we multiply our equation (2.1.1) by some arbitrary function $\mu(t)$ and obtain the new equation

$$\frac{dy}{dt}\mu(t) + \mu(t)g(t)y = \mu(t)h(t). \tag{2.1.2}$$

In order for this to be in the form of the product rule, we observe that $\mu(t)g(t)$ must be equal to $\mu'(t)$. Realizing this, we calculate $\mu(t)$ in the following way

$$\begin{aligned}
\mu(t)g(t) &= \mu'(t) \\
\Leftrightarrow g(t) &= \frac{\mu'(t)}{\mu(t)} \\
\Leftrightarrow \int g(t)dt &= \int \frac{\mu'(t)}{\mu(t)}dt \\
\Rightarrow \int g(t)dt &= \ln|\mu(t)| \\
\Leftrightarrow e^{\int g(t)dt} &= \mu(t)
\end{aligned}$$

where we choose the positive $\mu(t)$, since it is an arbitrary function. We also note that $\mu'(t) = g(t)e^{\int g(t)dt} = g(t)\mu(t)$.

Now if we let $w(t) = \mu(t)y(t)$, we see that $w'(t) = y'(t)\mu(t) + \mu'(t)y(t) = \frac{dy}{dt}\mu(t) + \mu(t)g(t)y$ which is the left hand side of (2.1.2). Thus we obtain the result

$$\begin{aligned}
w'(t) &= \mu(t)h(t) \\
\Rightarrow w(t) &= \int \mu(t)h(t)dt.
\end{aligned}$$

This can now be rewritten, using the fundamental theorem of calculus

$$w(t) = w_0 + \int_{t_0}^{t} \mu(s)h(s)ds$$

which leads us to the solution to the ODE

$$y(t) = \frac{1}{\mu(t)}\left(y(t_0)\mu(t_0) + \int_{t_0}^{t} \mu(s)h(s)ds\right). \tag{2.1.3}$$

It is this form of the solution to the ordinary differential equation that we shall use in a later section when we introduce the *stochastic differential equation*.

This method for solving an ODE is by *integrating factor*. Another example of this method follows.

Example 2.1. *Let* $\frac{\partial y}{\partial x} + \frac{y}{x} = \frac{e^x}{x}$. *We introduce the integrating factor* $exp\left\{\int \frac{1}{x}dx\right\} = e^{\ln x} = x$, *and multiply the equation with this, leading to*

$$x\frac{\partial y}{\partial x} + x\frac{y}{x} = x\frac{e^x}{x}.$$

Next we integrate both sides of the equation with respect to x, *and find the general solution* y *to the ordinary differential equation*

$$\int x\frac{\partial y}{\partial x} + ydx = \int e^x dx$$
$$xy = e^x \Leftrightarrow y = \frac{e^x}{x}.$$

□

2.2 Probability theory

2.2.1 Sample spaces

The purpose of this subsection is to provide definitions of some important terms used in later sections. This section will be quite dense with definitions, theorems and proofs (Alm and Britton, 2008), as we are just laying down the basics for the further discussions later on. If the reader is familiar with basic concepts of probability such as *random variables*, *probability spaces* and *independence*, this subsection can be skipped altogether. In probability theory we deal with random experiments, where we know for certain **that** something will happen but we cannot say exactly **what** will happen.

Definition 2.1 (Outcomes, events and sample spaces)**.** *The result of a random experiment is called an outcome, and we will denote these by* $u_1, u_2, \ldots$. *The set of all possible outcomes is called a sample space, and this we will denote by* Ω. *A specified set of outcomes is called an event, which we denote by* $A, B, ..$, *and the set of all events is called an event space. Hence, every unique outcome is its own event, and so is the whole sample space. A finite or countable infinite set of outcomes is called a discrete sample space, and the rest are called continuous sample spaces.*

Remark. *Outcomes and events are not numbers but elements or sets of elements, thus we can not add or subtract events, but instead we consider unions and intersections of events, which in turn are also events. We denote the intersection of two events* A *and* B *by* $A \cap B$ *and unions by* $A \cup B$.

Definition 2.2 (Intersections)**.** *The intersection* $A \cap B$ *of two sets (events) is the set of all elements (outcomes) that are in both sets* A *and* B, $A \cap B = \{u : u \in A \text{ and } u \in B\}$.

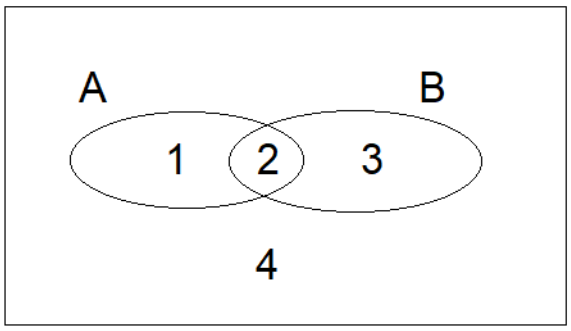


Figure 1: A Venn diagram depicting the two events A (area 1) and B (area 3), their intersection $A \cap B$ (area 2), and their union $A \cup B$ (areas 1, 2 and 3), all within the sample space Ω (areas 1, 2, 3 and 4).

Definition 2.3 (Unions)**.** *The union $A \cup B$ of two sets is the set of all elements that are in A, or in B, or in $A \cap B$, $A \cup B = \{u : u \in A \text{ or } u \in B\}$.*

Figure 1 above depicts the union and intersection of events A and B. Note that $A \cup B = B \cup A$ and $A \cap B = B \cap A$. The union and intersection of multiple events $A_1, A_2, ..., A_n$ is composed of the events $A_1 \cup A_2 \cup ... \cup A_n := \cup_{i=1}^{n} A_i$, and $A_1 \cap A_2 \cap ... \cap A_n := \cap_{i=1}^{n} A_i$ respectively.

The *complement* of the event A is denoted A^c, and is composed of all the outcomes within the sample space that does not belong to A, $A^c := \{u \in \Omega : u \notin A\}$. The complement A^c in figure 1 is composed of the areas marked 3 and 4. A special type of event is when there is no outcome. This event is called the empty set and we denote it by $\emptyset$. The complement of the empty set is hence the entire sample space, $\emptyset^c = \Omega$. If two events do not share any outcomes, we say that they are *disjoint*, i.e the intersection of two such events would be the empty set, $A \cap B = \emptyset$. If all the outcomes of the event A are also in B, we say that A is a subset of B, denoted $A \subset B$.

2.2.2 Probabilities on sample spaces

Let us now define random experiments and probabilities over sample spaces. We can describe a random experiment by defining the probabilities for all of the unique outcomes within the sample space. The probability for the event A to happen is denoted $P(A)$.

Definition 2.4 (Kolmogrovs axioms)**.** *In order for a real valued function P to be a probability function, it must fulfil the following axioms:*

1. $0 \leq P(A) \leq 1$ *for all events* $A \in \Omega$,
2. $P(\Omega) = 1$
3. *if* $A \cap B = \emptyset$, *then* $P(A \cup B) = P(A) + P(B)$.

If the sample space is infinite, axiom 3 is replaced with

3. *If $A_1, A_2, ...$ is an infinite sequence of pairwise disjoint events, i.e $A_i \cap A_j = \emptyset$ for all $i, \neq j$, then $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.*

Given these axioms, we will interpret probabilities such as $P(A) = c$ for any $c \in \mathbb{R} : 0 \leq c \leq 1$, to mean that if we repeat the same random experiment multiple times, then the proportion of experiments where

A occurs will be close to c. An event A is said to occur *almost surely* (abbreviated a.s) if $P(A) = 1$. Next we give a theorem defining the probabilities for complements and unions.

Theorem 2.1. *Let A and B be arbitrary events in the sample space Ω. Then*

1. $P(A^c) = 1 - P(A)$,
2. $P(\emptyset) = 0$,
3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Proof. The proof for the first result is quite intuitive and simple. We note that $A \cup A^C = \Omega$, and $A \cap A^c = \emptyset$, which together with the definitions given in def. 2.4 gives us $P(A) + P(A^c) = P(A \cup A^c) = P(\Omega) = 1$. The proof of the second result is also simple. Since $\emptyset = \Omega^c$, we can once again use definition 2.4 and result 1 above to see that $P(\emptyset) = P(\Omega^c) = 1 - P(\Omega) = 0$. The third result will require a bit more work. Consider the two sets A and $B \cap A^c$. These sets are disjoint, since no outcome that is in A can be in $B \cap A^c$ by definition. So $A \cup (B \cap A^c) = \emptyset$, which along with Kolomogrovs third axiom gives us

$$P(A \cup B) = P(A) + P(B \cap A^c). \tag{2.2.1}$$

Using the same axiom, we note that if we divide the event B into two disjoint events $B \cap A$ and $B \cap A^c$ we get $P(B \cap A^c) = P(B) - P(B \cap A) = P(B) - P(A \cap B)$. Substituting this into equation (2.2.1), we obtain the third result of the theorem, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. □

When talking about probability for real life events such as coin tossing or drawing random cards from a playing deck in a way that most people can grasp, we are talking about the *uniform* probability distribution.

Definition 2.5 (Uniform distribution)**.** *A random experiment is said to have uniform probability distribution if all the outcomes have the same probability, i.e if Ω has n outcomes $u_1, ..., u_n$, then the probability for each outcome is $P(u_i) = \frac{1}{n}$ for $i = 1, ..., n$.*

This definition leads us to the *classical definition of probability.*

Theorem 2.2 (Classical definition of probability)**.** *Given a discrete sample space Ω with uniform distribution, the probability of an event A occurring is equal to the amount of outcomes in A divided by the total amount of outcomes in Ω, i.e if there are m outcomes in A and n outcomes in Ω, then*

$$P(A) = \frac{m}{n}.$$

Before we give the proof of theorem 2.2, we will state and prove the following proposition.

Proposition 2.1. *If $A_1, ..., A1_n$ are disjoint events, then $P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$ for any $n \in \mathbb{N}^+$.*

The proof will be given by mathematical induction.

Proof. Let $I(n) : P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$. We want to show that $I(n)$ holds true for any $n \in \mathbb{N}^+$.
Base step.
Let $n = 2$. Then by Kolmogorov's third axiom we get

$$P(\cup_{i=1}^2 A_i) = P(A_1 \cup A_2) = P(A_1) + P(A_2) = \sum_{i=1}^{2} P(A_i),$$

since the events are disjoint by assumption.

Inductive step.

Assume $n = m$ for some $m \in \mathbb{N}^+$. Then

$$P(\cup_{i=1}^{m} A_i) + P(A_{m+1}) = \sum_{i=1}^{m} P(A_i) + P(A_{m+1}) = \sum_{i=1}^{m+1} P(A_i).$$

From Kolmogorov's third axiom, we get that

$$P(A_m \cup A_{m+1}) = P(A_m) + P(A_{m+1}),$$

thus

$$P(\cup_{i=1}^{m} A_i) + P(A_{m+1}) = P(\cup_{i=1}^{m+1} A_i),$$

which leads us to

$$P(\cup_{i=1}^{m+1} A_i) = \sum_{i=1}^{m+1} P(A_i).$$

Since both the base step and the inductive step has been shown to hold true, the statement $I(n)$ holds true for any $n \in \mathbb{N}^+$ by mathematical induction. □

Proof of Theorem 2.2. Assume the sample space Ω contains n outcomes $u_1, ..., u_n$. By Kolmogorov's axioms we have that $P(\Omega) = 1$, and $P(\Omega) = \sum_{i=1}^{n} P(U_i)$ by proposition 2.1. Since the outcomes in Ω are uniformly distributed, we have that $P(u_i) = P(u_j)$ for all $1 \leq i \leq n$ and $1 \leq j \leq n$, which implies that $P(u_i) = \frac{1}{n}$ for all $1 \leq i \leq n$. Thus for an event A consisting of m events, we get

$$P(A) = \sum_{\{i:u_i \in A\}} P(u_i) = \sum_{\{i:u_i \in A\}} \frac{1}{n} = \frac{m}{n}.$$

□

2.2.3 Conditional probability

As we will see in later sections of this thesis, we are sometimes interested in the probability of an event B occurring, given that event A has already occurred. This kind of probability is called conditional probability and we denote it by $P(B|A)$.

Definition 2.6 (Conditional probability)**.** *Let A be an arbitrary event such that $P(A) > 0$. The conditional probability that the event B will occur, given that the event A has already occurred is defined as*

$$P(B|A) := \frac{P(B \cap A)}{P(A)}.$$

This definition will be important when we define the path integral, since we will be looking at the conditional probabilities of particles passing through successive points to form a path.

Another important property events can have is independence. If two events A and B are independent, then the probability of event B occurring will not be affected by whether or not event A has already occurred.

Definition 2.7 (Independent event)**.** *Two events A and B are independent if $P(A|B) = P(A)$ given that $P(B) > 0$, and $P(B|A) = P(B)$ given that $P(A) > 0$. Another way to express this is that the events are independent if $P(A \cap B) = P(A)P(B)$.*

A set of events $\{A_1, A_2, ...\}$ is *pairwise independent* if $P(A_i \cap A_j) = P(A_i)P(A_j)$ for all $(i, j) : i \neq j$. If all the subsets of the set are also pairwise independent, i.e for all $k \geq 2$ and subsets $\{A_{i1}, .., A_{ik}\}$ where $i_1 < \cdots < i_k$ we have $P(A_{i1} \cap \cdots \cap A_{ik}) = P(A_{i1}) \cdots P(A_{ik})$, then the set is said to be *mutually independent*.

For some random experiments the probability of an event can be intuitive. Take for example the tossing of a coin, where the possible outcomes are *heads* and *tails*. Since there are only two possible outcomes, and they are clearly disjoint, we will have that $P(A \cup B) = P(\Omega) = 1 = P(A) + P(B)$. Furthermore the two events are uniformly distributed, leading us to the conclusion that $P(A) = P(B) = \frac{1}{2}$. This is however a very trivial example, and there are many scenarios in which we can not intuitively say anything about the probability of an event without some additional information. In these scenarios, the following theorem can be useful.

Theorem 2.3 (Law of total probability)**.** *For a sample space Ω consisting of disjoint events A_i with $P(A_i) > 0$ for $i = 1, ..n$ such that $\cup_{i=1}^{n} A_i = \Omega$, the probability of any event $B = \cup_{i=1}^{n}(B \cap A_i)$ occurring is given by*

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i).$$

Proof. Since the events A_i are disjoint, the events $B \cap A_i$ are also disjoint and we can apply Kolmogorov's third axiom

$$P(B) = P(\cup_{i=1}^{n}(B \cap A_i)) = \sum_{i=1}^{n} P(B \cap A_i).$$

The definition of conditional probability gives us $P(B|A_i)P(A_i) = P(B \cap A_i)$, completing the proof. □

The final theorem we will prove before moving on to defining random variables, is *Bayes' theorem*.

Theorem 2.4 (Bayes' theorem)**.** *For a sample space Ω consisting of disjoint events A_i with $P(A_i) > 0$ for $i = 1, ..n$ such that $\cup_{i=1}^{n} A_i = \Omega$, the probability of any event $B = \cup_{i=1}^{n}(B \cap A_i)$ occurring is given by*

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^{n} P(A_j)P(B|A_j)}$$

Proof. From the definition of conditional probability we have

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} \tag{2.2.2}$$

and

$$P(B|A_i) = \frac{P(B \cap A_i)}{P(A_i)} \;\Leftrightarrow\; P(B|A_i)P(A_i) = P(B \cap A_i)$$

which matches the numerator in 2.2.2 with the numerator of the theorem. Furthermore, by theorem 2.3 we get

$$P(B) = \sum_{j=1}^{n} P(A_j)P(B|A_j)$$

$$\Rightarrow P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^{n} P(A_j)P(B|A_j)} = \frac{P(A_i \cap B)}{P(B)}.$$

□

2.2.4 Probability spaces

In order to make sense of statements like "the probability of event x occurring is y", we need to define a *probability space.*

Definition 2.8 (Probability space)**.** *A probability space is made up of three components, a non-empty sample space Ω, a σ-algebra (event space) $\mathbb{F}$, and a probability function P such that $P : F \to [0, 1]$.*

Remark. *A σ algebra is a collection $\mathbb{F}$ of subsets of a set Ω such that*

- $\emptyset \in \mathbb{F}$
- *if $A \in \mathbb{F}$, then $A^c \in \mathbb{F}$*
- *if $A_1, A_2, ..$ is a countable collection of sets in $\mathbb{F}$, then their union $\cup_n A_n \in \mathbb{F}$.*

2.2.5 Random variables

A random variable, or stochastic variable as they are also called, is a variable for which the values depend on the outcome of a random experiment.

Definition 2.9 (Random variable)**.** *A random variable $X(u)$ is a function defined on a sample space, $X : \Omega \mapsto \mathbb{E}$ for some measurable set $\mathbb{E}$. Here we will be treating the most common case which is that $X(u)$ is a real valued function and hence $\mathbb{E} = \mathbb{R}$. When the random experiment has been performed and an outcome has been observed, the value of the function $X(u)$ is called an observation of the random variable. The argument u is often omitted, and we simply write X instead of $X(u)$ for random variables, and x for the observations of X.*

Since the sample space can be either a discrete or a continuous one, we will need definitions for both kinds of random variables. We will start by defining the discrete kind, and a few of its properties.

Definition 2.10 (Discrete random variables and their probability functions)**.** *A random variable X is discrete if it can assume a finite or countably infinite amount of values, $x_1, x_2, ...$ The probability distribution function p_X for a discrete random variable is defined*

$$p_X(x) := P(X = x), \quad x = x_1, x_2, ...$$

Theorem 2.5 (Properties of probability functions)**.** *For a discrete random variable, the following is true (treating the most common integer-valued random variable)*

1. $0 \leq p_X(x) \leq 1, \quad \forall k \in \mathbb{Z},$
2. $\sum_k p_X(k) = 1,$
3. $P(a \leq X \leq b) = \sum_{\{k:a \leq k \leq b\}} p_X(k),$
4. $P(X \leq a) = \sum_{\{k:k \leq a\}} p_X(k),$
5. $P(X > a) = \sum_{\{k:k>a\}} p_X(k) = 1 - \sum_{\{k:k \leq a\}} p_X(k) = 1 - P(X \leq a).$

Remark. *We treat the integer-valued discrete random variables since these are the most common kind of discrete random variables, although the same properties can be verified for real-valued discrete random variables.*

Proof. 1. Consider the subset A_k of Ω consisting of all the outcomes u such that $X(u) = k$. For this subset or event A_k, we get by Kolmogorov's axioms that $0 \leq P(A_k) \leq 1$, which proves the first statement since $p_X(k) = P(X = k) = P(A_k)$.

2. Since $X(u)$ only assumes one value for each individual u, the events $A_0, A_1, ...$ will all be disjoint, and so we can once again use Kolmogorov's axioms to prove statement 2 in the theorem,

$$\sum_k p_X(k) = \sum_k P(A_k) = P(\Omega) = 1$$

.

3. The event $\cup_{\{k:a\leq k\leq b\}} A_k$ is the same as the event $a \leq X \leq b$. Thus, again using Kolmogorov's axioms and the fact that the events A_k are disjoint, we obtain

$$P(a \leq X \leq b) = P(\cup_{\{k:a\leq k\leq b\}} A_k) = \sum_{\{k:a\leq k\leq b\}} p_X(k).$$

4. The event $\{X \leq a\}$ is identical to the event $\cup_{\{k:k\leq a\}} A_k$. Hence, by disjoint events and Kolmogorov's axioms, we obtain

$$P(X \leq a) = P(\cup_{\{k:k\leq a\}} A_k) = \sum_{\{k:k\leq a\}} p_X(k).$$

5. The event $X > a$ is identical to the event $\cup_{\{k:a<k\}} A_k$. Again by disjoint events and Kolmogorov's axioms, we obtain

$$P(X > a) = P(\cup_{\{k:a<k\}} A_k = \sum_{\{k:a<k\}} p_X(k).$$

Since we know from statement 2 that $\sum_k p_X(k) = 1$, we observe that

$$\sum_{\{k:a<k\}} p_X(k) + \sum_{\{k:k\leq a\}} p_X(k) = 1$$

$$\Leftrightarrow \sum_{\{k:a<k\}} p_X(k) = 1 - \sum_{\{k:k\leq a\}} p_X(k)$$

$$\Leftrightarrow P(X > a) = 1 - P(X \leq a).$$

□

To calculate probabilities such as $P(a \leq X \leq b)$, for a random variable X, we can also use what is called a *probability distribution function.*

Definition 2.11 (Probability distribution function)**.** *For a random variable X, the probability distribution function $F_X(t)$ is defined as*

$$F_X(t) := P(X \leq t), \quad -\infty < t < \infty.$$

So, for example, for a discrete random variable X, by property number 4 of theorem 2.5, the value of the distribution function in the point t is $F_X(t) = P(X \leq t) = \sum_{\{k:k\leq t\}} p(k)$. Next we will give a theorem concerning some properties of the distribution function.

Theorem 2.6. *Let $F_X(t)$ be the distribution function of a random variable X. Then the following statements are true*

1. $0 \leq F_X(t) \leq 1, \quad \forall\, t,$

2. $t \mapsto F_X(t)$ *is monotonically increasing and right-continuous,*

3. $\lim_{t\to-\infty} F_X(t) = 0$,

4. $\lim_{t\to\infty} F_X(t) = 1$,

5. $P(a < X \leq b) = F(b) - F(a)$,

6. $P(X > a) = 1 - F(a)$,

Proof. 1. Let $A_t := \{u : X(u) \leq t\}$. Then $P(A_t) = P(X \leq t) = F_X(t)$, and by Kolmogorov's axioms $0 \leq F_X(t) \leq 1$.

2. i) $t \mapsto F_X(t)$ **is monotonically increasing.** Consider the observations x and y where $x < y$. From the definitions of the distribution function and probability events, we have that $A_x \subseteq A_y$. We divide A_y in two disjoint subsets $A_y = A_x \cup A_{(x,y]} = \{u : x < X(u) \leq y\}$. We now obtain the following, once again using Kolmogorov's axioms

$$F_X(x) = P(A_x) \leq P(A_x) + P(A_{(x,y]}) = P(A_y) = F_X(y),$$

which proves that F_X is monotonically increasing.

ii) F_X **is right-continuous** will be proven only for the case where X is discrete and integer-valued. We fix t and let $\lfloor t \rfloor$ be the integer part of t. Then we have $\lfloor t + h \rfloor = \lfloor t \rfloor$ for $h > 0$ small enough. Then we obtain $F_X(t + h) = F_X(\lfloor t + h \rfloor) = F_X(\lfloor t \rfloor) = F_X(t)$, proving that F_X is right continuous.

3. As $t \to -\infty$, the *cardinality*[1] of the set $A_k := \{k : k \leq t\}$ will tend to zero

$$\lim_{t\to-\infty} |\{k : k \leq t\}| = 0.$$

Thus $P(A_k) \to P(\emptyset) = 0$ as $t \to -\infty$.

4. Similarly as in the proof for statement 3, as $t \to \infty$ the subset $A_k := \{k : k \leq t\}$ will tend to the whole sample space Ω, leading to $\lim_{t\to\infty} P(A_k) = P(\Omega) = 1$.

5. We have that

$$P(a < X \leq b) = \sum_{\{k:a\leq k\leq b\}} p_X(k) = \sum_{\{k:k\leq b\}} p_X(k) - \sum_{\{k:k\leq a\}} p_X(k) = F(b) - F(a).$$

6. $P(X > a) = 1 - F(a)$ follows directly from statement 5 in theorem 2.5 and the definition of $F(a)$.

□

We now move on to define the continuous random variable.

Definition 2.12 (Continuous random variables and density functions)**.** *A random variable X is continuous if there exists a function $f_X(x)$ such that for all events or subsets A*

$$P(X \in A) = \int_A f_X(t)\ dt.$$

The function $f_X(x)$ is known as the density function of the continuous random variable X.

[1] The cardinality of a set is a measure of the number of elements in the set.

The relation between the distribution functions $F_X(t)$ for continuous random variables, and the density functions $f_X(t)$ will be given by the following theorem.

Theorem 2.7. *Given a continuous random variable X with density function $f_X(\cdot)$ and distribution function $F_X(\cdot)$, the following is true for all points where $f_X(\cdot)$ is continuous*

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\ dt.$$

and conversely

$$f_X(x) = F'_X(x) = \lim_{h\to 0} \frac{F_X(x+h) - F_X(x)}{h}$$

It is also true that $\int_{-\infty}^{\infty} f_X(t)\ dt = 1$.

Proof. Consider the event $A_x := \{u : -\infty < X(u) \leq x\}$. The definition of the density function gives us $P(X \in A_x) = \int_{-\infty}^{x} f_X(t)\ dt$. The next statement is obtained using the first statement and the fundamental theorem of calculus.

The last statement $\int_{-\infty}^{\infty} f_X(t)dt = 1$ follows from the observation that $\int_{-\infty}^{\infty} f_X(t)dt = \lim_{x\to\infty} F_X(x)$, and from statement 4 in theorem 2.6 we know that $\lim_{x\to\infty} F_X(x) = 1$ for any distribution function.

□

Given the nature of random variables, it can be difficult to make intuitive predictions about their behaviour. But if we know the probability function or the density function for the random variable, we can define its expected value.

Definition 2.13 (Expected value (mean))**.** *The expected value of a random variable X, or mean as it is also called, is a real number denoted by $E(X)$. For discrete random variables it is defined as*

$$E(X) := \sum_k k p_X(k) \tag{2.2.3}$$

and for continuous random variables

$$E(X) := \int_{-\infty}^{\infty} x f_X(x)\ dx. \tag{2.2.4}$$

If the sum or the integral is infinite however, then X does not have an expected value.

In the next section we will be studying functions of random variables, and so we need to define the expected value of these types of functions as well.

Theorem 2.8 (Expected value of a function of a random variable)**.** *Let the random variable Y be defined by $Y := g(X)$, where X is a random variable and $g(\cdot)$ is a real valued function. Then*

$$E(Y) = E(g(X)) = \begin{cases} \sum_k g(k)p_X(k) & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^{\infty} g(x)f_X(x)\ dx & \text{if } X \text{ is continuous.} \end{cases}$$

Proof. We will prove the discrete case. We note that $P(g(X) = j) = \sum_{\{k:g(k)=j\}} P(X = k)$. This leads us to

$$E(Y) = \sum_j jP(Y = j) = \sum_j jP(g(X) = j)$$

$$= \sum_j \sum_{\{k:g(k)=j\}} jP(X=k) = \sum_j \sum_{\{k:g(k)=j\}} g(k)P(X=k)$$
$$= \sum_k g(k)P(X=k).$$

□

Two other useful measurements of random variables are their *variances* and *standard deviations*. These can give us a perception of how much the values deviate from the mean value.

Definition 2.14 (Variance)**.** *For a random variable X with mean value $E(X) = \mu$, the variance σ^2 is defined as $\sigma^2 = V(X) := E((X-\mu)^2)$, if μ is finite. Using theorem 2.8, this can also be expressed as*

$$V(X) = \begin{cases} \sum_k (k-\mu)^2 p_X(k), & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)\ dx, & \text{if } X \text{ is continuous.} \end{cases}$$

Definition 2.15 (Standard deviation)**.** *The standard deviation of a random variable X is defined as*

$$D(X) = \sqrt{V(X)} = \sqrt{\sigma^2} = \sigma.$$

Theorem 2.9. *For a random variable X with mean value $E(X)$ and variance $V(X)$, and constants a and b we have*

$$E(aX+b) = aE(x) + b$$
$$V(aX+b) = a^2 V(x)$$
$$D(aX+b) = |a| D(x).$$

Proof. Proof of the continuous case. Let X be a continuous random variable and a, b constants. then we have

$$E(aX+b) = \int_{-\infty}^{\infty} (ax+b) f_X(x)\ dx$$
$$= a \int_{-\infty}^{\infty} x f_X(x) + b \int_{-\infty}^{X(x)\ dx = aE(X)+b.}$$

For the variance we use the definition of variance to obtain

$$V(aX+b) = E((aX+b-(aE(X)+b))^2)$$
$$= E(a^2(X-E(X))^2) = a^2 E((X-E(X))^2) = a^2 V(X).$$

Lastly, for the standard deviation we get

$$D(aX+b) = \sqrt{V(aX+b)} = \sqrt{a^2 V(X)} = |a| D(X).$$

□

Theorem 2.10 (Calculating variance)**.** *The variance for a random variable X with mean μ is given by*

$$V(X) = E(X^2) - \mu^2 = \begin{cases} \int_{-\infty}^{\infty} x^2 f_X(x)\ dx - \mu^2, & \text{if } X \text{ is continuous,} \\ \sum_k k^2 p_X(k) - \mu^2, & \text{if } X \text{ is discrete.} \end{cases}$$

Proof. We will show the continuous case. We have from the definition of variance that

$$V(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)\ dx = \int_{-\infty}^{\infty} (x^2 - 2\mu x + \mu^2)\ dx$$

$$= \int_{-\infty}^{\infty} x^2 f_X(x)\ dx - 2\mu \underbrace{\int_{-\infty}^{\infty} x f_X(x)\ dx}_{=\mu} + \mu^2 \underbrace{\int_{-\infty}^{\infty} f_X(x)\ dx}_{=1}$$

$$= \int_{-\infty}^{\infty} x^2 f_X(x)\ dx - \mu^2.$$

□

Random variables can have a number of different probability distributions. In this the main focus will be on *normally distributed* random variables, for which we now give the definition.

Definition 2.16. *A continuous random variable X is said to be normally distributed with mean μ and variance $\sigma^2 > 0$, denoted $X \sim N(\mu, \sigma^2)$, if its density function is given by*

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty.$$

We verify that $f_X(x)$ is a density function, i.e that $\int_{-\infty}^{\infty} f_X(x)\ dx = 1$ in the following way.
We perform the variable substitution $[z = (x-\mu)/\sigma,\ \sigma dz = dx]$, and obtain the integral $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-z^2}{2}}\ dz$. So we need to show that $\int_{-\infty}^{\infty} e^{\frac{-z^2}{2}}\ dz = \sqrt{2\pi}$ or that $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{\frac{-(z^2+y^2)}{2}}\ dzdy = 2\pi$. We calculate this double integral using polar coordinates $z = r\cos(u), y = r\sin(v)$, yielding the integral

$$\int_0^{2\pi}\int_0^{\infty} r e^{-\frac{r^2}{2}}\ drdu = \int_0^{2\pi} du = 2\pi$$

$$\Rightarrow \int_{-\infty}^{\infty} e^{\frac{-z^2}{2}}\ dz = \sqrt{2\pi} \Rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-z^2}{2}}\ dz == 1,$$

showing that $f_X(x)$ is indeed a density function.

Theorem 2.11 (Calculating mean and variance for normally distributed random variables)**.** *For a normally distributed random variable $X \sim N(\mu, \sigma^2)$, the mean, variance and standard deviation is given by*

$$E(X) = \mu, \quad V(X) = \sigma^2, \quad D(X) = \sqrt{\sigma^2} = \sigma.$$

Proof. From the definition of mean value of a random variable and the density function of the normally distributed random variable, we get, using the same variable substitution $[z = (x-\mu)/\sigma,\ \sigma dz = dx]$ as above

$$E(X) = \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\ dx = \int_{-\infty}^{\infty} (\sigma z + \mu) \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}\ dz.$$

Since $f(z) = ze^{-z^2/2}$ is an odd function, its contribution to the integral will be zero, and we need only concern ourselves with the term $\mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\ dz$, which, by the verification of the density function above is equal to μ, showing that $E(X) = \mu$.

For the variance we start by calculating $E(X^2)$ using the same variable substitution as before and partial integration.

$$E(X^2) = \int_{-\infty}^{\infty} (\sigma z + \mu)^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\ dz = \int_{-\infty}^{\infty} \sigma^2 z^2 + 2\sigma\mu z + \mu^2\ dz$$

$$= \mu^2 + \sigma^2 \int_{-\infty}^{\infty} z^2 e^{-\frac{z^2}{2}}\, dz = \sigma^2 = \mu^2 + \sigma^2 \left[z \frac{-e^{-\frac{z^2}{2}}}{\sqrt{2\pi}} \right]_{-\infty}^{\infty} + \sigma^2 \int_{-\infty}^{\infty} \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}}\, dz$$
$$= \mu^2 + 0 + \sigma^2 \Rightarrow V(X) = E(X^2) - E(X)^2 = \mu^2 + \sigma^2 - \mu^2 = \sigma^2$$

□

Lastly we need to define the concept of covariance. This can be used to see how the two variables influence each other, if at all. The covariance measures the linear dependence or the joint variability of the two random variables. First we introduce multivariate random variables.

Definition 2.17 (Two dimensional random variable)**.** *A two dimensional function* $(X, Y) = (X(u), Y(u)) : \Omega \to \mathbb{R} \times \mathbb{R}$ *defined on a sample space* Ω *is called a two dimensional random variable* X, Y*. The distribution function of the two dimensional random variable is defined as*

$$F_{X,Y}(x, y) := P(X \leq x, Y \leq y).$$

Just as in the case of single random variables, we need to consider the two cases of continuous and discrete multivariate random variables.

Definition 2.18 (Continuous and discrete two dimensional random variables)**.** *If a two dimensional random variable* X, Y *only can assume a finite or a countably infinite amount of values, we say that* X, Y *is discrete and define its probability function as*

$$p_{X,Y}(j, k) := P(X = j, Y = k)$$

We say that $X, Y)$ *is continuous if there exists a function* $f_{X,Y}(x, y) > 0$*, such that for all sets* A *we have*

$$P((X, Y) \in A) = \int \int_A f_{X,Y}(x, y)\, dxdy.$$

If $f_{X,Y}$ *exists, we say that it is the density function of* X, Y*.*

Theorem 2.12 (Mean value of a function of two random variables)**.** *Let* X, Y *be a two dimensional random variable and* g *be a real-valued function. Then*

$$E(g(X, Y)) = \begin{cases} \sum_j \sum_k g(j, k) p_{X,Y}(j, k), & \text{if } (X, Y) \text{ is discrete,} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(X, Y) f_{X,Y}(x, y)\, dxdy, & \text{if } (X, Y) \text{ is continuous.} \end{cases}$$

Proof. The proof of this theorem is analogous to the proof of theorem 2.8 and will hence be omitted. □

Corollary 2.12.1 (Mean value of a sum)**.** *Let* X *and* Y *be arbitrary random variables. Then*

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

Proof. Proof of the continuous case. Let $g(X, Y) = X + Y$. Then, by theorem 2.12 we have

$$E(g(X, Y)) = E(X + Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x + y f_{X,Y}(x, y)\, dxdy$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{X,Y}(x, y)\, dxdy + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f_{X,Y}(x, y)\, dxdy$$
$$= E(X) + E(Y).$$

□

Theorem 2.13 (Variance of sum)**.** *Let X and Y be arbitrary random variables. Then*

$$V(X+Y) = V(X) + V(Y) + 2C(X,Y).$$

Proof. We have that

$$V(X+Y) = C(X+Y, X+Y)$$
$$= C(X,X) + C(X,Y) + C(Y,X) + C(Y,Y)$$
$$= V(X) + V(Y) + 2C(X,Y).$$

□

Definition 2.19 (Covariance and correlation coefficient)**.** *Let X and Y be random variables with the same distribution and finite mean values μ_X and μ_Y, and standard deviations σ_X and σ_Y. The covariance between X and Y is defined as*

$$C(X,Y) := E((X-\sigma_X)(Y-\sigma_Y)),$$

and their correlation coefficient as

$$\rho(X,Y) := \frac{C(X,Y)}{\sigma_X \sigma_Y}.$$

Theorem 2.14 (Rules for covariance calculations)**.** *Consider the random variables X, Y, Z and the constants a, b, c, d. The following rules applies for calculating covariance for X, Y, Z*

$$C(X,X) = V(X)$$

$$C(Y,X) = C(X,Y)$$

$$C(aX+b, cY+d) = acC(X,Y)$$

$$C(X+Y,Z) = C(X,Z) + C(Y,Z)$$

Proof. The proof of the first two equations are given by the definition of variance and covariance. For the third equation we get

$$C(aX+b, cY+d)$$
$$= E((aX+b-(a\mu_X+b))(cY+d-(c\mu_Y+d))$$
$$= E(a(X-\mu_X)c(Y-\mu_Y))$$
$$= acE((X-\mu_X)(Y-\mu_Y)) = acC(X,Y).$$

For the last equation, we have

$$C(X+Y,Z) = E((X+Y-(\mu_X+\mu_Y))(Z-\mu_Z))$$

$$= E(((X-\mu_X)+(Y-\mu_Y)))(Z-\mu_Z))$$

$$= C(X,Z) + C(Y,Z).$$

□

Theorem 2.15 (Calculating covariance)**.** *Let $X, Y)$ be a two dimensional random variable with mean values $E(X) = \mu_X$ and $E(Y)\mu_Y$. Then*

$$C(X,Y) = E(XY) - \mu_X \mu_Y.$$

Proof. By using the property of mean values $E(aX) = aE(X)$ and corollary 2.12.1, we obtain

$$\begin{aligned} E((X-\mu_x)(Y-\mu_Y)) &= E(XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y) \\ &= E(XY) - \mu_Y E(X) - \mu_X E(Y) + \mu_X\mu_Y \\ &= E(XY) - \mu_X\mu_Y. \end{aligned}$$

□

The final part of this section will be defining independent random variables and see how we calculate the expected value and variance for them.

Definition 2.20 (Independent and uncorrelated random variables)**.** *Let X and Y be random variables. We say that X and Y are independent if for all (x, y)*

$$p_{X,Y}(x,y) = p_X(x)p_Y(y) \quad \text{if } X \text{ and } Y \text{ are discrete,}$$

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) \quad \text{if } X \text{ and } Y \text{ are continuous.}$$

If $C(X, Y) = 0$, X and Y are uncorrelated.

Theorem 2.16. *If X and Y are independent then they are also uncorrelated.*

Proof. Proof of the continuous case. Let X and Y be independent random variables. Then

$$\begin{aligned} E(XY) &= \int\int xyf_{X,Y}(x,y)\,dxdy = \int\int xyf_X(x)f_Y(y)\,dxdy \\ &= \int xf_X(x)\,dx \int yf_Y(y)\,dy = E(X)E(Y) \\ &\Leftrightarrow^{2.15} C(X,Y) = 0. \end{aligned}$$

□

Theorem 2.17 (Variance and mean value for linear combinations)**.** *Let $X_1, ..., X_n$ be arbitrary random variables and $a_1, ..., a_n$ arbitrary constants. Then*

$$E(a_1X_1 + ... + a_nX_n) = a_1E(X_1) + ... + a_nE(X_n)$$

$$V(a_1X_1, ..., a_nX_n) = \sum_{i=1}^{n} a_i^2V(X_i) + 2\sum_{i<j} a_ia_jC(X_i, X_j).$$

If X and Y are uncorrelated we have

$$V(a_1X_1 + ... + a_nX_n) = \sum_{i=1}^{n} a_i^2V(X_i).$$

Proof. For the mean value the proof is directly obtained from theorem 2.9 and corollary 2.12.1. For the variance we get

$$\begin{aligned} V\left(\sum_{k=1}^{n} a_kX_k\right) &= C\left(\sum_{i=1}^{n} a_iX_i, \sum_{i=1}^{n} a_iX_i\right) \\ &\sum_{i=1}^{n}\sum_{j=1}^{n} a_ia_jC(X_i, X_j) \\ &= \sum_{i=1}^{n} a_i^2C(X_i, X_i) + \sum_{j\neq i} a_ia_jC(X_i, X_j) \\ &= \sum_{i=1}^{n} a_i^2V(X_i) + 2\sum_{i<j} a_ia_jC(X_i, X_j). \end{aligned}$$

□

3 Stochastic calculus

3.1 Stochastic processes

If we are interested in studying a sequence of observations of random variables, we can make use of the stochastic process, which is an assembly of random variables, each associated with an index from an index set. Such an assembly is known as a family.

Definition 3.1 (Stochastic process)**.** *A family of random variables with index t in the index set I is called a stochastic process, $\{X(t), t \in I\}$. The stochastic process assumes values in the codomain V, and the outcome observed from the stochastic process is called a realization of the process.*

Hence, for every fixed t the stochastic process $X(t)$ is a random variable. As in the case of random variables, we need to make distinctions between continuous and discrete stochastic processes.

Definition 3.2 (Continuous- and discrete-time stochastic processes)**.** *A stochastic process with the index set I is called a continuous-time process if I is a continuous set, commonly an interval of the sort $I = [0, \infty)$. If I is a discrete set, for example $I = \{0, 1, 2, ...\}$, the stochastic process is called a discrete-time process.*

Definition 3.3 (Continuous and discrete stochastic processes)**.** *A stochastic process with codomain V is called continuous if V is a continuous set, commonly an interval $V = [0, \infty)$, and conversely if V is a discrete set $V = \{0, 1, 2, ..\}$ the process is called a discrete stochastic process.*

An important tool needed to study stochastic processes is the *conditional expectation*. This can be thought of as similar to the conditional probability for events described in section 2.2, but for random variables.

Definition 3.4 (Conditioning on an event)**.** *Let X be an integrable random variable, and A be an event in the event space $\mathbb{F}$ with positive probability $P(A) > 0$. The conditional expectation of X given that A has occurred is given by*

$$P(X|A) = \frac{1}{P(A)} \int_A X \, dP.$$

We will see in section 3.4 what the meaning of integrating with respect to a function is. Next we define conditioning on a discrete random variable.

Definition 3.5 (Conditioning on a discrete random variable)**.** *Let X be an integrable random variable and let Y be a discrete random variable with possible values $y_1, y_2, ..., y_n$ such that $P(Y = y_i) > 0$ for all $i \in [0, n]$. Instead of conditioning on just one event, we want to find the conditional expectation of X given that all the events in Y has occurred, i.e a sequence of conditional expectations $E(X|Y = y_1), E(X|Y = y_2), ...$ To do so we construct the random variable $E(X|Y)(\cdot)$ on each of the sets $\{Y = y_n\}$, and we define the conditional expectation of X given Y to be the random variable*

$$E(X|Y)(u) = E(X|Y = y_n), \quad \text{if } Y(u) = y_n.$$

We demonstrate this with the following example.

Example 3.1. *Consider an experiment where we are flipping 3 coins $1kr, 2kr$ and $10kr$. We let the amount X be the sum of the value of the coins which land heads up. We now pose the question, what is*

the conditional expectation $E(X|Y)$ given the total amount Y for the flips of only the $1kr$ and $2kr$ coins? Here, Y is a discrete random variable with possible values $0, 1, 2, 3$. From the definition of conditioning on a discrete random variable, we obtain the following

$$E(X|\{Y=0\}) = 5, \quad E(X|\{Y=1\}) = 6$$

$$E(X|\{Y=2\}) = 7, \quad E(X|\{Y=3\}) = 8.$$

So our expression for the conditional expectation becomes

$$E(X|Y)(u) = \begin{cases} 5 & \text{if } Y(u) = 0 \\ 6 & \text{if } Y(u) = 1 \\ 7 & \text{if } Y(u) = 2 \\ 8 & \text{if } Y(u) = 3. \end{cases}$$

3.2 Random Walk

A random walk is a stochastic process, portraying random steps in a mathematical space.

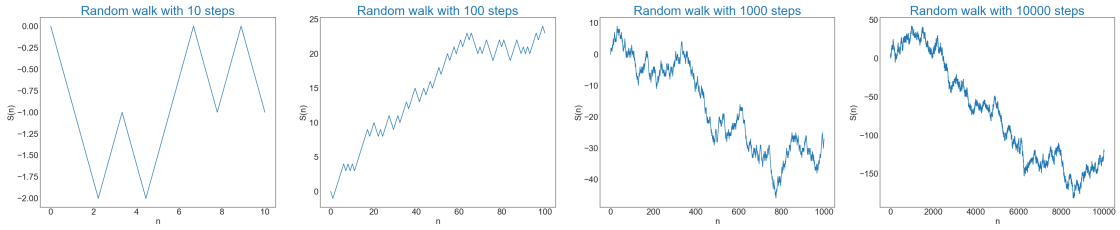
Definition 3.6. *A random walk is a sequence of random variables $\{S_n,\ n = 0, 1, 2, ...\}$ with $S_0 = 0$ which is defined by*

$$S_n := \sum_{k=1}^{n} X_k$$

where the random variables X_k are independent and from the same distribution. A random walk is said to be simple if the random variables only assume values 1 or -1, and it is symmetric and simple if $P(X_k = 1) = P(X_k = -1) = \frac{1}{2}$.

In figure 2 we can see four realizations of simple symmetric random walks where the random variable X_k represents the movement of a particle along the x-axis with steps of 1, moving between the integer points.

Figure 2: Random walks with 10, 100, 1000 and 10000 steps



3.3 Wiener process

In the natural sciences we come across a concept called Brownian motion, which is a type of random walk. It represents the random motion of particles suspended in a fluid, and was first described by Robert Brown when observing small particles of pollen immersed in a liquid through his microscope. The mathematical properties of a one-dimensional Brownian motion was first described by Norbert Wiener as a way to study continuous time *martingales*[2], and so the way to describe a Brownian motion mathematically is by defining the **Wiener process**. For simplicity purposes we will only consider one dimensional Brownian motion in this thesis, but the theory is applicable to spaces of arbitrary dimensions.

Definition 3.7 (Wiener process)**.** *A stochastic process* $\{W(t), t \geq 0\}$ *is a Brownian motion or a Wiener process if*

1. $W(0) = 0$,
2. $W(t)$ *has stationary independent increments,*
3. $W(t) \sim N(0, \sigma^2 t)$ *for all* $t > 0$.

If $\sigma = 1$*, the Wiener process is said to be standard, which is the type of Wiener processes we will be working with in this thesis.*

Remark. *A stochastic process* $W(t), t \in T$ *has independent increments if* $W(t_2) - W(t_1), ..., W(t_n) - W(t_{n-1})$ *are all independent for all* $t_1 < t_2 < ... < t_n$ *in* T*. It has stationary increments if the distribution of the random variable* $W(t) - W(s)$ *has the same distribution as* $W(t - s)$ *for any* $s < t$*. In the case of the Wiener process, we have that* $W(t) - W(s) \sim N(0, (t - s)\sigma^2)$*.*

Example 3.2. *Consider a Wiener process* W_t *with* $t_0 = 0, W(t_0) = W_0 = 0$ *and* $0 \leq t \leq T$*.*
While the time in theory flows continuously, we will discretize it for the sake of examining the properties of the Wiener process. Let $W_{t+\Delta t} - W_t = \Delta W_t$*, meaning* ΔW_t *denotes the change in* $W(\cdot)$ *over a time period beginning at time* t *with length* Δt*. The change in the Wiener process is random, so we define* ΔW_t *to depend on a random component* ε_t *where* $\{\varepsilon_0, \varepsilon_{\Delta t}, ..., \varepsilon_{T-\Delta t}\}$ *are all* $\sim N(0, 1)$ *and all uncorrelated.*

$$\Delta W_t = \varepsilon_t \sqrt{\Delta t}.$$

$$\Delta W_0 = \varepsilon_0 \sqrt{\Delta t} \Leftrightarrow W_{\Delta t} = \varepsilon_0 \sqrt{\Delta t}$$

$$\Delta W_{\Delta t} = \varepsilon_{\Delta t} \sqrt{\Delta t} \Leftrightarrow W_{2\Delta t} = W_{\Delta t} + \varepsilon_{\Delta t} \sqrt{\Delta t} = (\varepsilon_0 + \varepsilon_{\Delta t}) \sqrt{\Delta t}$$

$$\vdots$$

$$\Delta W_{T-\Delta t} = \varepsilon_{T-\Delta t} \sqrt{\Delta t} \Leftrightarrow W_T = (\varepsilon_0 + \varepsilon_{\Delta t} + \cdots + \varepsilon_{T-\Delta t}) \sqrt{\Delta t}$$

Since we want the time flow to be continuous, we scale the change ε_t *in* $W(\cdot)$ *by* $\sqrt{\Delta t}$ *instead of* Δt*. This choice ensures that the Wiener process will not freeze as* $\Delta t \to 0$*, since* $\sqrt{\Delta t}$ *goes to zero much slower than* Δt*.*

[2] A continuous-time martingale with respect to the stochastic process X_t is a stochastic process Y_t such that for all t
$\mathbf{E}(|Y_t|) < \infty$
$\mathbf{E}(Y_t | \{X_\tau, \tau \leq s\}) = Y_s \quad \forall s \leq t$
This expresses the property that the conditional expectation of an observation at time t, given all the observations up to time s, is equal to the observation at time s.

In Figure 3 below, we can see example graphs of two such Wiener processes with identical initial conditions $t_0 = 0, \quad n = 500, \quad W(t_0) = 0, \quad \Delta t = 0.02, \quad T = n\Delta t = 10$. When we go on to define the stochastic differential equation and the stochastic integral, we will think of the white noise ξ as the *time derivative* of a wiener process, $\xi(t) = \frac{d\mathbf{W}}{dt}$, even though the wiener process is nowhere differentiable, meaning it does not exist in the ordinary sense.

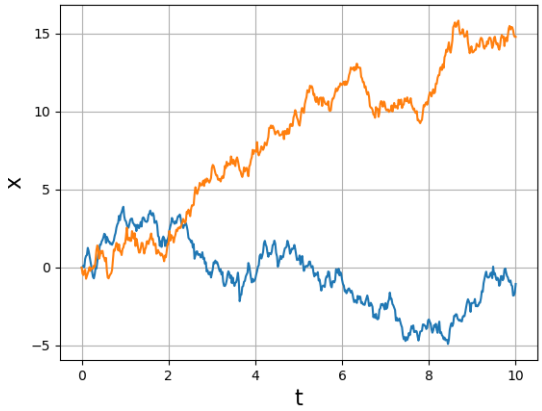


Figure 3: One-dimensional Wiener process example graph

3.4 Stochastic Integrals

Another part of the theory of stochastic differential equations is the stochastic integral. Recall that we have an integral in the the solution from (2.1.3) in section (2.1)

$$y(t) = \frac{1}{\mu(t)}\left(y(t_0)\mu(t_0) + \int_{t_0}^{t} \mu(s)h(s)ds\right). \tag{2.1.3}$$

We are going to want to find a solution on a similar form for the stochastic differential equation, but in order to do that we must first specify what a stochastic integral is. We want to be able to say something about integrals of the form (Dobrow, 2016)

$$\int_0^t B_s ds \quad \text{and} \quad \int_0^t B_s dB_s.$$

The first of the above integrals can be seen as representing the area restricted by the curve of the Brownian motion on the time interval $[0, t]$ and the horizontal axis, as the Brownian motion is integrated with respect to time. Due to the fact that the integrand is random, the integral itself is also random, a random variable. Viewed as a function of t, we can say that the integral is a stochastic process.

In the second integral however, Brownian motion is integrated *with respect to* Brownian motion.

Before we have a look at integration with respect to Brownian motion, let us first define integration of Brownian motion itself.

Let B_t be a Brownian motion process with $0 \leq t \leq 1$. Since we know that B_t is continuous for all t, it is by definition Riemann integrable. But, we still need to define what the Riemann integral of the Brownian motion actually is.

Let us first consider, for $0 \leq a < b$, the integral

$$\int_a^b B_s(\omega) ds.$$

Since we know that $B(\omega)$ is a continuous function for each ω in the probability space, we can define this integral in the ordinary sense as the limit when n tends to infinity of a Riemann sum,

$$I^{(n)}(\omega) = \sum_{k=1}^{n} B_{t_k^*}(t_k - t_{k-1})$$

for a partition $a = t_0 < t_1 < \cdots < t_{n-1} < t_n = b$ of $[a, b]$ where $t_k^* \in [t_{k-1}, t_k]$ is an arbitrary point in the subinterval $[t_{k-1}, t_k]$. This sum will be a random variable for every $n \geq 1$, and since Brownian motion is a Gaussian process, they will all be normally distributed. So if we let $n \to \infty$, we can expect the limit $\lim_{n\to\infty} I^{(n)}$ to also be normally distributed. Now if we let $I_t = \int_0^t B_s ds$ for $t \geq 0$, the mean for the random variable I_t is given by

$$E\left(I_t\right) = E\left(\int_0^t B_s ds\right) = \int_0^t E\left(B_s\right) ds = 0,$$

and so for $s \leq t$, the covariance of (I_s, I_t)

$$\begin{aligned} C\left(I_s, I_t\right) &= E\left(\left(I_s - E(I_s) * \left(I_t - E(I_t)\right)\right)\right) = E\left(I_s I_t\right) = E\left(\int_0^s B_x dx \int_0^t B_y dy\right) \\ &= \int_0^s \int_0^t E\left(B_x B_y\right) dy dx = \int_0^s \int_0^t \min\{x, y\} dy dx \\ &= \int_0^s \int_0^x y dy dx + \int_0^s \int_x^t x dy dx \\ &= \frac{s^3}{6} + \left(\frac{ts^2}{2} - \frac{s^3}{3}\right) = \frac{3ts^2 - s^3}{6} \end{aligned}$$

To find the variance of I_t, we let $s = t$ and thus we aquire $C(I_t, I_t) = V(I_t) = \frac{t^3}{3}$, leading to the conclusion that $\int_0^t B_s ds$ is a normally distributed random variable with mean 0 and variance $\frac{t^3}{3}$, $\int_0^t B_s ds \sim N(0, \frac{t^3}{3})$. In figure 4 we see examples of a few realizations of integrated Brownian motion.

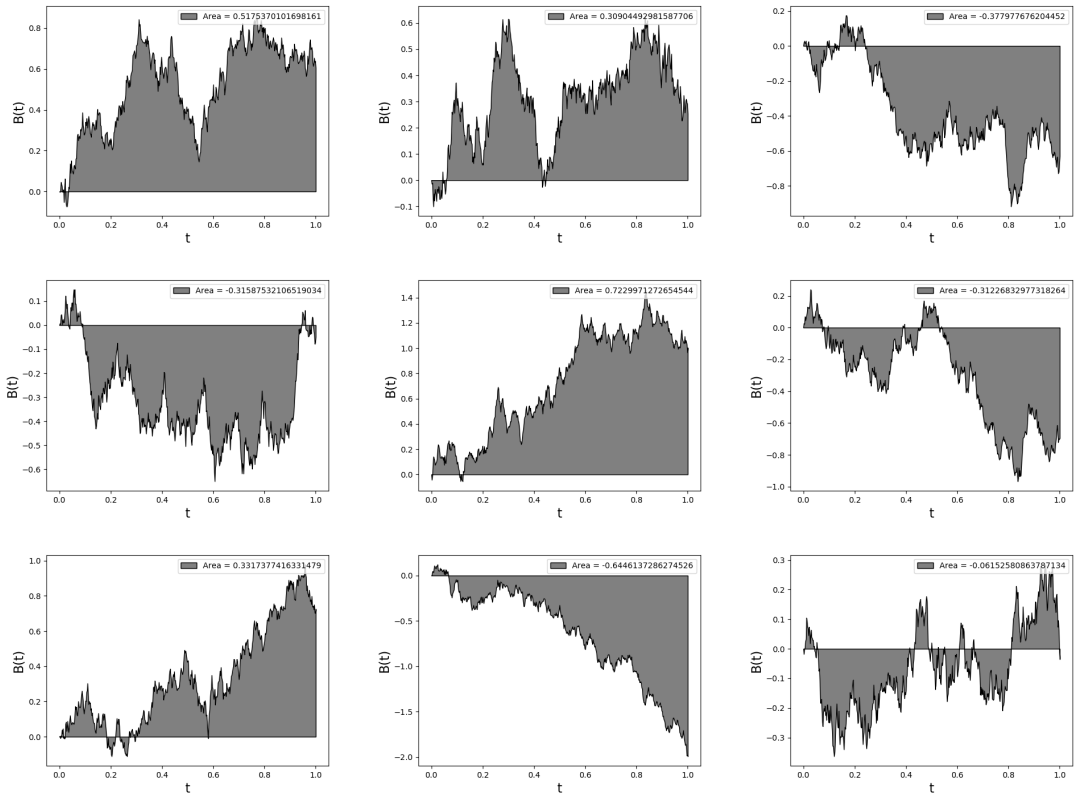


Figure 4: Plots of nine different realizations of area under a Brownian motion curve. Generated from time 0 to time 1, with 500 steps of size $\frac{1}{500}$.

Now that we have clarified what integrated Brownian motion is, we will look into what it means to integrate with respect to Brownian motion. To do this we introduce the *Riemann-Stieltjes* integral. Before we do this however, we will define some necessary properties of sets.

Definition 3.8 (Refinement)**.** *Consider an closed interval $I = [a, b]$ and let $\mathbb{P}[a, b]$ be the set of partitions of I. Let $P = \{a = x_0 < x_1 < \cdots < x_n = b\} \in \mathbb{P}[a, b]$. Then a partition $P' \in \mathbb{P}[a, b]$ such that $P \subseteq P'$ is called a refinement of P, or we say that P' is finer than P.*

Example 3.3. *Let I be the closed interval $I = [0, 10]$, and P a partition $P = \{1, 2, 4\} \in \mathbb{P}[0, 10]$. Then $P' = \{1, 2, 3, 4\} \in \mathbb{P}[0, 10]$ is a refinement of P and we say that P' is finer than P.*

Definition 3.9 (Mesh)**.** *Let $I = [a, b]$ be a closed interval and $P = \{a = x_0 < x_1 < \cdots < x_n = b\} \in \mathbb{P}[a, b]$. We denote the mesh (or norm) of the partition P as $||P||$ and define it as*

$$||P|| := \max_{k \in \{1,2,..,n\}} \{|x_k - x_{k-1}|\}.$$

Definition 3.10 (Upper and lower Riemann-Stieltjes sums)**.** *Let the functions f and g be defined on $[a, b]$, and let g be increasing on $[a, b]$, i.e for all $x, y \in [a, b] : x < y$ we have $g(x) < g(y) \Leftrightarrow g(y) - g(x) > 0$. Let $P = \{a = x_0 < x_1 < \cdots < x_n = b\}$, $M_k(f) = \sup\{f(x) : x \in [x_{k-1}, x_k]\}$, and $m_k(f) = \inf\{f(x) : x \in [x_{k-1}, x_k]\}$. We then define the upper and lower Riemann-Stieltjes sums with respect to*

the partition P as

$$U(S(P,f,g)) = \sum_{k=1}^{n} M_k(f)\Delta g_k$$

and

$$L(S(P,f,g)) = \sum_{k=1}^{n} m_k(f)\Delta g_k$$

respectively. If these sums are equal, then we say that the Riemann-Stieltjes sum on the interval $[a,b]$ is their common value.

Definition 3.11 (Upper and lower Riemann-Stieltjes integrals)**.** *Let f and g be functions defined on the interval $I = [a,b]$, where g is an increasing function. Then the upper and lower Riemann-Stieltjes integrals are defined*

$$\overline{\int_a^b} f(x)\ dg(x) := \inf\left\{U(P,f,g) : P \in \mathbb{P}[a,b]\right\},$$

and

$$\underline{\int_a^b} f(x)\ dg(x) := \sup\left\{L(P,f,g) : P \in \mathbb{P}[a,b]\right\}$$

respectively. If these integrals are equal, then we say that the Riemann-Stieltjes integral on the interval $[a,b]$ is their common value.

Definition 3.12 (Riemann-Stieltjes sum and integral)**.** *Consider the interval $I = [a,b]$ and the partition $P = \{a = x_0 < x_1 < \cdots < x_n = b\}$. For each $k \in \{1,2,\ldots,n\}$ let $t_k \in [x_{k-1}, x_k]$ and let $\Delta g_k = g(x_k) - g(x_{k-1})$. Furthermore, let f and g be functions defined on I. We denote the Riemann-Stieltjes sum with respect to P, f and g by*

$$S(P,f,g) = \sum_{k=1}^{n} f(t_k)\Delta a_k.$$

If there exists an $A \in \mathbb{R}$ such that for every $\epsilon > 0$ there exists a partition P_ϵ of $[a,b]$ such that for all partitions P such that $(P_\epsilon \subseteq P)$ and for any choice of $t_k \in [x_{k-1}, x_k]$ we have that $|S(P,f,\alpha) - A| < \epsilon$, then f is said to be a Riemann-Stieltjes integrable function, and we write

$$\int_a^b f(x)\ dg(x) = A.$$

For our definition of the integral with respect to Brownian motion however, both the integrand and the integrating function are now stochastic processes, and we consider the stochastic integral to be a generalization of the Riemann-Stieltjes integral.

For a continuous random variable X with f differentiable, the mean value can be expressed as

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)dx = \int_{-\infty}^{\infty} g(x)F'(x)dx = \int_{-\infty}^{\infty} g(x)dF(x), \tag{3.4.1}$$

where $f(x)$ is the density function, and $F(x)$ is the distribution function of X. The r.h.s of equation 3.4.1 is the Riemann-Stieltjes integral of $g(x)$ with respect to the distribution function $F(x)$. The integral with respect to Brownian motion will be defined as a result of this, namely

$$I_t = \int_0^t g(s)dB_s \tag{3.4.2}$$

where g is a bounded continuous function. With the same partition $0 = t_0 < t_1 < t_2 < \cdots < t_{n-1} < t_n = t$ and $t_k^* \in [t_{k-1}, t_k]$ this leads us to the approximating sum

$$I_t^{(n)} = \sum_{k=1}^{n} g\left(t_k^*\right)\left(B_{t_k} - B_{t_{k-1}}\right). \tag{3.4.3}$$

From the definition of Wiener process, we have that $B_{tk} - B_{tk-1}$ is normally distributed with mean 0 and variance $t_k - t_{k-1}$, and so $I^{(n)}$ is also normally distributed for all n. As n approaches infinity, equation 3.4.3 converges to the integral in equation 3.4.2, which can be shown (Dobrow, 2016) to be a normally distributed random variable. We now calculate the mean and variance for that variable, using the definitions from section 2.2. We start with the mean

$$E(I_t) = \lim_{n\to\infty} E\left(\sum_{k=1}^{n} g(t_k^*\left(B_{t_k} - B_{t_{k-1}}\right)\right)$$

$$= \lim_{n\to\infty} \sum_{k=1}^{n} g(t_k^*) E\left(B_{t_k} - B_{t_{k-1}}\right) = 0.$$

And given the independent increments of the Wiener process we get for the variance

$$V\left(I_t^{(n)}\right) = \sum_{k=1}^{n} g^2\left(t_k^*\right) V\left(B_{t_k} - B_{t_{k-1}}\right) = \sum_{k=1}^{n} g^2\left(t_k^*\right)\left(t_k - t_{k-1}\right)$$

Letting $n \to \infty$, we will get that $V(I_t^{(n)}$ converges to the integral $\int_0^t g^2(s)\ ds$, and so our stochastic integral is normally distributed with mean 0 and variance $\int_0^t g^2(s)\ ds$, $\int_0^t g(s)dB_s \sim N\left(0, \int_0^t g^2(s)ds\right)$.

3.5 Itô's Lemma

One of the most important results of stochastic calculus is the stochastic version of the chain rule of derivation, known as Itô's Lemma (Brzezniak and Zastawniak, 2000).

In order to define and prove this lemma, we start with another definition.

Definition 3.13 ($\mathbb{M}_T^2$ and $\mathbb{M}^2$ stochastic processes)**.** *Denote $\mathbb{M}_T^2$ to be the class of stochastic processes $f(t),\ t \geq 0$, such that*

$$\mathrm{E}\left(\int_0^T |f(t)|^2 dt\right) < \infty,$$

and let $\mathbb{M}^2$ be the class of stochastic processes $f(t)$ such that $f(t) \in \mathbb{M}_T^2$ for any $T > 0$.

Lemma 3.1 (Itô's lemma)**.** *Suppose that f(t, x) is a real valued function with continuous partial derivatives $F_t(t,x), F_x(t,x)$ and $F_{xx}(t,x)$ for all $t \geq 0$ and $x \in \mathbb{R}$. Assume also that the process $F_x(t, W_t)$ belongs to $\mathbb{M}^2$, and $F_x, F_t, F_{xx} \in \mathbb{L}^2$. Then $F(t, W_t)$ satisfies*

$$\begin{aligned} &F(T, W_T) - F(0, W_0) \\ &= \int_0^T \left[F_t(t, W_t) + \frac{1}{2} F_{xx}(t, W_t)\right] dt + \int_0^T F_x(t, W_t)\, dW_t, \qquad a.s. \end{aligned} \tag{3.5.1}$$

In shorter differential notation, (3.5.1) can be written as

$$dF(t, W_t) = \left[F_t(t, W_t) + \frac{1}{2} F_{xx}(t, W_t)\right] dt + F_x(t, W_t)\, dW_t \tag{3.5.2}$$

Remark

a) *Compare equation (3.5.2) with the usual chain rule*

$$dF(t, x_t) = F_t(t, x_t)\, dt + F_x(t, x_t)\, dx_t$$

for a differentiable function x_t. The additional term $\frac{1}{2} F_{xx}(t, W_t)\, dt$ in (3.5.2) is called the Ito correction.

b) *Equation (3.5.2) is often written in its shorter derivative form*

$$dF = \left(F_t + \frac{1}{2} F_{xx}\right) dt + F_x dW_t.$$

Proof. We will prove the case where F_t, F_x, F_{xx} are all bounded by some $C > 0$. Consider a partition $0 = t_0^n < t_1^n < \cdots < t_n^n = T$ of $[0, T]$, where $t_i^n = \frac{iT}{n}$. Denote $W_{t_i^n}$ by W_i^n; the increments $W_{i+1}^n - W_i^n$ by $\Delta_i^n W$; and $t_{i+1}^n - t_i^n$ by $\Delta_i^n t$. Using Taylor expansion, there is a point $\widetilde{W}_i^n$ in each interval $\left[W_i^n, W_{i+1}^n\right]$ and a point $\widetilde{t}_i^n$ in each interval $\left[t_i^n, t_{i+1}^n\right]$ such that

$$
\begin{aligned}
F(T,W_T)-F(0,W_0) &= \sum_{i=0}^{n-1}\left[F(t_{i+1}^n,W_{i+1}^n)-F(t_i^n,W_i^n)\right]\\
&= \sum_{i=0}^{n-1}\left[F(t_{i+1}^n,W_{i+1}^n)-F(t_i^n,W_{i+1}^n)\right]+\sum_{i=0}^{n-1}\left[F(t_i^n,W_{i+1}^n)-F(t_i^n,W_i^n)\right]\\
&= \sum_{i=0}^{n-1}F_t(\widetilde{t}_i^n,W_{i+1}^n)\Delta_i^n t+\sum_{i=0}^{n-1}F_x(t_i^n,W_i^n)\Delta_i^n W+\frac{1}{2}\sum_{i=0}^{n-1}F_{xx}(t_i^n,\widetilde{W}_i^n)(\Delta_i^n W)^2\\
&= \sum_{i=0}^{n-1}F_t(\widetilde{t}_i^n,W_{i+1}^n)\Delta_i^n t+\frac{1}{2}\sum_{i=0}^{n-1}F_{xx}(t_i^n,W_i^n)\Delta_i^n+\sum_{i=0}^{n-1}F_x(t_i^n,W_i^n)\Delta_i^n W+\\
&+\frac{1}{2}\sum_{i=0}^{n-1}F_{xx}(t_i^n,W_i^n)\left[(\Delta_i^n W)^2-\Delta_i^n t\right]+\frac{1}{2}\sum_{i=0}^{n-1}\left[F_{xx}(t_i^n,\widetilde{W}_i^n)-F_{xx}(t_i^n,W_i^n)\right](\Delta_i^n W)^2\\
&= S_1^n+S_2^n+S_3^n+S_4^n+S_5^n.
\end{aligned}
$$

Note that as $F_t, F_x,$ and F_{xx} are continuous bounded functions, we have

$$\lim_{n\to\infty}\sup_{i=1,\ldots,n}\sup_{t\in\left[t_i^n,t_{i+1}^n\right]}\left|F_t\left(\widetilde{t}_i^n,W_{i+1}^n\right)-F_t\left(t,W_t\right)\right|\to 0 \quad \text{a.s.} \tag{3.5.3}$$

$$\lim_{n\to\infty}\sup_{i=1,\ldots,n}\sup_{t\in\left[t_i^n,t_{i+1}^n\right]}\left|F_{xx}\left(t_i^n,W_i^n\right)-F_{xx}\left(t,W_t\right)\right|\to 0, \quad \text{a.s.} \tag{3.5.4}$$

$$\lim_{n\to\infty}\sup_{i=1,\ldots,n}\left|F_{xx}\left(t_i^n,\widetilde{W}_i^n\right)-F_{xx}\left(t_i^n,W_i^n\right)\right|\to 0 \quad \text{a.s..} \tag{3.5.5}$$

Next we will deal with each sum S_1^n to S_5^n separately.

i) From the continuity of F_t and F_{xx} in equations (3.5.3) and (3.5.5) we have the convergence of the Riemann integral

$$\lim_{n\to\infty}S_1^n=\lim_{n\to\infty}\sum_{i=0}^{n-1}F_t\left(\widetilde{t}_i^n,W_{i+1}^n\right)\Delta_i^n t=\int_0^T F_t\left(t,W_t\right)dt \quad \text{a.s., and}$$

$$\lim_{n\to\infty}S_2^n=\lim_{n\to\infty}\sum_{i=0}^{n-1}F_{xx}\left(t_i^n,W_i^n\right)\Delta_i^n t=\int_0^T F_{xx}\left(t,W_t\right)dt \quad \text{a.s.}$$

ii) From the assumption $F_x\in\mathbb{M}^2$, we have

$$\lim_{n\to\infty}S_3^n=\lim_{n\to\infty}\sum_{i=0}^{n-1}F_x\left(t_i^n,W_i^n\right)\Delta_i^n W=\int_0^T F_x\left(t,W_t\right)dW_t$$

in $\mathbb{L}^2$.

iii) We show the $\mathbb{L}^2$ convergence $\mathrm{E}\left((S_4^n)^2\right)\to 0$. To be specific,

$$\mathbb{E}\left((S_4^n)^2\right)=\mathbb{E}\left(\sum_{i=0}^{n-1}F_{xx}\left(t_i^n,W_i^n\right)\left[\left(\Delta_i^n W\right)^2-\Delta_i^n t\right]\right)^2$$

$$=\sum_{i=0}^{n-1}\mathbb{E}\left|F_{xx}\left(t_i^n,W_i^n\right)\left[\left(\Delta_i^n W\right)^2-\Delta_i^n t\right]\right|^2 \quad \text{(cross terms have expectation 0)}$$

$$=\sum_{i=0}^{n-1}\mathbb{E}\left|F_{xx}\left(t_i^n,W_i^n\right)\right|^2\mathbb{E}\left|\left(\Delta_i^n W\right)^2-\Delta_i^n t\right|^2 \quad \text{(independent increments)}$$

$$\leq C^2 \sum_{i=0}^{n-1} \mathbb{E}\left|\left(\Delta_i^n W\right)^2 - \Delta_i^n t\right|^2 \quad \text{(boundedness of } F_{xx}\text{)}$$
$$= 2C^2 \sum_{i=0}^{n-1} \left(\Delta_i^n t\right)^2 = 2C^2 \sum_{i=0}^{n-1} \frac{T^2}{n^2} = 2C^2 \frac{T^2}{n} \to 0 \quad \text{as } n \to \infty$$

iv) Note that $\sum_{i=0}^{n-1} \left(\Delta_i^n W\right)^2 \to t$ in L^2 and thus in probability since the left quantity is the quadratic variation of Brownian motion. Together with the continuity result 3.5.4, we have the following convergence in probability

$$|S_5^n| = \left|\sum_{i=0}^{n-1} \left[F_{xx}\left(t_i^n, \widetilde{W}_i^n\right) - F_{xx}\left(t_i^n, W_i^n\right)\right] \left(\Delta_i^n W\right)^2\right|$$

$$\sup_{i=1,2,\ldots,n} \left|F_{xx}\left(t_i^n, \widetilde{W}_{i+1}^n\right) - F_{xx}\left(t_i^n, W_{i+1}^n\right)\right| \sum_{i=0}^{n-1} \left(\Delta_i^n W\right)^2 \xrightarrow{P} 0$$

Note that the convergence of $S_i^n, i = 1, \ldots, 5$ involve different modes: S_1^n and S_2^n converge almost surely, S_3^n, S_4^n converge in $\mathbb{L}^2$, and S_5^n converges in probability. To combine the results, we note that convergence in $\mathbb{L}^2$ implies convergence in probability (Brzezniak and Zastawniak, 2000). Thus all S_3^n, S_4^n and S_5^n converge in probability. Note also that there is a subsequence $\{n_k\}_{k=1,2,\ldots}$ such that $\{S_3^{n_k}\}_{k=1,2,\ldots}$ converge a.s (Brzezniak and Zastawniak, 2000). Along this subsequence, we can find a further subsequence n_{k_l} such that $S_4^{n_{kl}}$ converges a.s., and so forth. Therefore, all S_j^n $j = 1, \ldots, 5$ converge a.s. with respect to some subsequence $m_1 < m_2 < \cdots$, say. Then

$$\begin{aligned}
&F\left(T, W_T\right) - F\left(0, W_0\right) \\
&= \lim_{k\to\infty} \left\{ \sum_{i=0}^{m_k-1} F_t\left(t_i^{m_k}, W_{i+1}^{m_k}\right) \Delta_i^{m_k} t + \frac{1}{2} \sum_{i=0}^{m_k-1} F_{xx}\left(t_i^{m_k}, W_i^{m_k}\right) \Delta_i^{m_k} t + \sum_{i=0}^{m_k-1} F_x\left(t_i^{m_k}, W_i^{m_k}\right) \Delta_i^{m_k} W \right. \\
&\quad + \frac{1}{2} \sum_{i=0}^{m_k-1} F_{xx}\left(t_i^{m_k}, W_i^{m_k}\right) \left[\left(\Delta_i^{m_k} W\right)^2 - \Delta_i^{m_k} t\right] \\
&\quad \left. + \frac{1}{2} \sum_{i=0}^{m_k-1} \left[F_{xx}\left(t_i^{m_k}, \widetilde{W}_i^{m_k}\right) - F_{xx}\left(t_i^{m_k}, W_i^{m_k}\right)\right] \left(\Delta_i^{m_k} W\right)^2 \right\} \\
&= \int_0^T \left[F_t\left(t, W_t\right) + \frac{1}{2} F_{xx}\left(t, W_t\right)\right] dt + \int_0^T F_x\left(t, W_t\right) dW_t, \quad \text{a.s.}
\end{aligned}$$

□

The general case where F_t, F_x and F_{xx} are not bounded is treated in (Brzezniak and Zastawniak, 2000).

Example 3.4. *For $F(t,x) = x^3$ we have $F_t(t,x) = 0, F_x(t,x) = 3x^2$ and $F_{xx}(t,x) = 6x$. By Itô's formula we obtain*

$$d\left(W(t)^3\right) = 3W(t)dt + 3W(t)^2 dW(t)$$

4 Stochastic Differential Equation

We are now ready to introduce the Stochastic Differential Equation (SDE).

In stochastic calculus we want to allow the differential equation to be influenced by a random process, or a "white noise" at any given point in time. Recall the example from 2.1 where we wanted to model the increase of money deposited in a bank account. For this simple problem, we used the ODE $\frac{dX}{dt} = aX$ which had the solution $X(t) = X_0 e^{at}$ with initial condition $X(t_0 = 0) = X_0$. Consider now instead the scenario where we want to "save" or invest money on the stock market. Since the evolutionary processes of stock market prices are not deterministic, we need to include some element of randomness in our equation.
We can do so by rewriting the equation to $dX(t) = aX(t)dt$, and then instead of the constant a, we introduce a stochastic process R such that $R = a + \sigma\frac{dB}{dt}$ (Evans, 2014).

$$\begin{cases} dX(t) = aX(t)dt + \sigma X(t)dB(t) \\ X(t_0 = 0) = X_0 \end{cases} \tag{4.0.1}$$

Equation (4.0.1) can of course also be written on the form $\frac{dX(t)}{dt} = aX(t) + \sigma X(t)\frac{dB(t)}{dt}$, but from this point of view, the addition of this white noise term $\frac{dB(t)}{dt}$ is not trivial, since this denotes the time derivative of Brownian motion, which is nowhere differentiable. Despite this, Brownian motion is often referred to as integrated white noise, implying that white noise is the derivative of Brownian motion. We will instead consider Brownian motion to be integrated white noise.

The development of the price of stocks is not linear but depends on the current price, so we will use a logarithmic scale to model the price. We let $z = \log X$, and then we apply Itô's lemma.

Since the terms dt^2 and $dtdB(t)$ tend to zero faster than dt, we set them to zero in the taylor expansion, and substitute $dB(t)^2$ with dt, which leads to

$$dX^2 = a^2X^2(t)dt^2 + 2aX(t)^2\sigma dtdB(t) + \sigma^2X^2(t)dB^2(t) = \sigma^2X^2(t)dt.$$

We now get

$$dz = d\log(X) = \frac{1}{X} + \frac{1}{2}\frac{-1}{X^2}(dX)^2 = adt + \sigma dB - \frac{1}{2X^2}\sigma^2X^2dt = \left(a - \frac{1}{2}\sigma^2\right)dt + \sigma dB$$

This SDE can be solved by stochastic integration since σ is constant, which leads us to

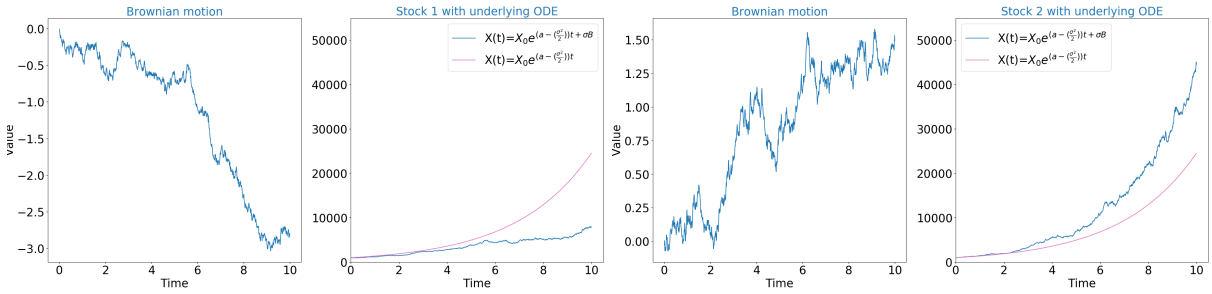
$$z(t) = t\left(a - \frac{\sigma^2}{2}\right) + \sigma(B(t) - B(0)) + z(0) = t\left(a - \frac{\sigma^2}{2}\right) + \sigma B(t) + \log(X_0).$$

Having solved $z(t)$, we can now calculate the solution to our SDE (4.0.1) by exponentiation

$$X(t) = e^{z(t)} = e^{\left(a - \frac{\sigma^2}{2}\right)t + \sigma B(t) + \log(X_0)} = X_0 e^{\left(a - \frac{\sigma^2}{2}\right)t + \sigma B(t)}$$

The graphs in figure (5) visualize the examples of ODE from section (2.1) and SDE (4.0.1). We can see the effect of the Brownian motion quite clearly since the two graphs for the stock prices are very different from each other. The graph on the left was influenced by a negatively trending BM, resulting in a weak development of the stock price, while the graph on the right was influenced by a strong positive BM yielding a much higher return on investment. Both graphs are based on the same ODE, which is shown below, and the same BM, yet they produce significantly different results.

Figure 5: SDE's with their underlying Brownian motion and ODE



4.1 Solving SDE

There are many types of SDEs, hence there is also quite a few different approaches to solving them. For instance, if the functions $b(\mathbf{X}(t)), \mathbf{B}(\mathbf{X}(t))$ are linear, we could use a technique called coefficient matching. To demonstrate how this is done, we consider the SDE

$$\begin{cases} dX_t = b(t, X_t)dt + B(t, X_t)dW_t, \quad (t > 0) \\ X(0) = x_0 \end{cases}$$

to which, as we have previously seen, the solution can be written on the form

$$X_t = x_0 + \int_0^t b\left(s, X_s\right) ds + \int_0^t B\left(s, X_s\right) dW_s. \tag{4.1.1}$$

If we want to find the *strong* solution to this SDE, that is equivalent to finding a function $f : [0, \infty) \times \mathbb{R} \to \mathbb{R}$ such that $X_t = f(t, W_t)$. To find this f, we start by applying Itô's formula

$$\begin{aligned} f\left(t, W_t\right) = f(0,0) &+ \int_0^t \left\{ \frac{1}{2} \frac{\partial^2 f}{\partial x^2}\left(s, W_s\right) + \frac{\partial f}{\partial s}\left(s, W_s\right) \right\} ds \\ &+ \int_0^t \frac{\partial f}{\partial x}\left(s, W_s\right) dW_s \end{aligned} \tag{4.1.2}$$

Now if we compare the terms in (4.1.1) and (4.1.2), we see that

$$\begin{aligned} \frac{1}{2} \frac{\partial^2 f}{\partial x^2}(s, x) + \frac{\partial f}{\partial s}(s, x) &= b(s, f(s, x)) \\ \frac{\partial f}{\partial x}(s, x) &= B(s, f(s, x)) \\ f(0, 0) &= x_0 \end{aligned}$$

As an example, consider the most basic example of an SDE

$$\begin{cases} dX_t = dt + dB_t \\ X_0 = x_0. \end{cases}$$

The solution to this SDE is $X_t = x_0 + \int_0^t dt + \int_0^t dB_t = x_0 + t + B_t$, as we know from (4.1.1). So let us check that *coefficient matching* will bring us the same result. Applying Itô's formula to $f(t, W_t) = X_t$ we get the following system of equations

$$\begin{aligned} \frac{1}{2} \frac{\partial^2 f}{\partial x^2}(s, x) + \frac{\partial f}{\partial s}(s, x) &= 1 \\ \frac{\partial f}{\partial x}(s, x) &= 1 \\ f(0, 0) &= x_0. \end{aligned}$$

Solving this system, we obtain the following

$$\begin{aligned}\frac{\partial f}{\partial x}(s,x) = 1 \Rightarrow f(s,x) = x + g(s)\\ \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(s,x) + \frac{\partial f}{\partial s}(s,x) = 1 \Rightarrow g'(s) = 1 \Rightarrow g(s) = s + c_0\\ f(0,0) = x_0 \Rightarrow c_0 = x_0.\end{aligned}$$

Next we will see how the same method can be used to solve a slightly more complex non-linear SDE

$$\begin{cases} X_t = (X_t - X_t^3 + \cos(2\pi t))dt + \sqrt{2}\sigma dW_t \\ X(0) = 1. \end{cases}$$

We set $a(x) = x - x^3 + \cos(2\pi t),\ b(x) = \sqrt{2}\sigma$ and define the stochastic process

$$Y_t = \int_{X_0}^{X_t} \frac{ds}{b(s)} = \int_{X_0}^{X_t} \frac{ds}{\sqrt{2}\sigma} = \left[\frac{s}{\sqrt{2}\sigma}\right]_{s=0}^{s=X_t} = \frac{X_t}{\sqrt{2}\sigma} - \left(\frac{X_0}{\sqrt{2}\sigma}\right) = \frac{X_t - 1}{\sqrt{2}\sigma}. \tag{4.1.3}$$

Now, we know from Itô's lemma that

$$f(X_t) - f(X_0) = \int_0^t f'(X_s)\ dX_s + \frac{1}{2}\int_0^t f''(X_s)b^2(X_s)\ ds,$$

where

$$\int_0^t f'(X_s)\ dX_s = \int_0^t f'(X_s)a(X_s)\ ds + \int_0^t f'(X_s)b(X_s)\ dW_s. \tag{4.1.4}$$

For $f(x) = \frac{x}{\sqrt{2}\sigma}$ we get

$$\begin{aligned}Y_t = \frac{X_t - 1}{\sqrt{2}\sigma} = f(X_t) - f(X_0) = \int_0^t \frac{1}{\sqrt{2}\sigma}\ dX_s + \frac{1}{2}\int_0^t 0 \cdot b^2\ ds\\ = \int_0^t \frac{1}{\sqrt{2}\sigma}\ dX_s = \int_0^t f'(X_s)\ dX_s \overset{(4.1.4)}{=} \int_0^t f'(X_s)a(X_s)\ ds + \int_0^t f'(X_s)b(X_s)\ dW_s\\ = \int_0^t \frac{1}{\sqrt{2}\sigma}(X_s - X_s^3 + \cos(2\pi s)\ ds + \int_0^t \frac{1}{\sqrt{2}\sigma}\sqrt{2}\sigma\ dW_s = \frac{1}{\sqrt{2}\sigma}\left(\frac{X_t^2}{2} - \frac{X_t^4}{4} + \frac{\sin(2\pi t)}{2\pi} - \frac{1}{4}\right) + W_t\end{aligned}$$

Combining (4.1.3) and (4.1.4) we obtain the result

$$X_t = \frac{3}{4} + \frac{X_t^2}{2} - \frac{X_t^4}{4} + \frac{\sin(2\pi t)}{2\pi} + \sqrt{2}\sigma W_t.$$

We will see later on in the numerical section of this thesis that this type of non-linear SDE can be used for simulating rare events.

5 Path integral

5.1 Background

In ordinary calculus, we come across the multi dimensional integral when posed with problems such as determining the volume of a sphere or other more abstract concepts like hypervolumes of n dimensions. The key point here being that an n-dimensional integral integrates over n variables $x_1, ..., x_n$, whereas the path integral integrates over an infinite amount of *functions* $f(x)$ of a variable x, which may be a vector or a real number. This type of integral was first introduced by Norbert Wiener who used in the fields of Brownian motion and diffusion theory. It is therefore sometimes also referred to as a Wiener integral. The reason why this type of integral is important for the study of rare events is that we can use the path integral to find the most probable path to the rare event. When a rare event occurs, the state of the system in question jumps from one steady state to another, and it can do so in an infinite number of ways, i.e take any of an infinite number of paths. Being a rare event, all of these paths are extremely unlikely to occur, however there is one path which is much more probable than all the rest of them, and that is the path we want to find.

5.2 Probability distributions and definitions

Consider again equation (**??**), and let this random walk represent the movement of a particle which moving randomly along the x-axis by steps of size ℓ and time ε. The particle can move from point $x_0 = 0$ to point $x_\varepsilon = \pm\ell$ during the first step in the time ε. As mentioned earlier, this is a discretized example of a Brownian motion or a Wiener process with equal probability for the particle to move in either direction. Seeing as the particle can only move a distance of ℓ in each step, we write the probability for the particle to travel from point $x = i\ell$ to the point $x = j\ell$ during one time step ε as

$$W(i\ell - j\ell, \varepsilon) = \begin{cases} \frac{1}{2} & \text{if } |i-j| = 1 \\ 0 & \text{otherwise} \end{cases} \qquad (i, j \in \mathbb{Z}) \; . \tag{5.2.1}$$

This discrete step random walk represents a basic example of a *Markov chain*, which can be characterized by a pair $(W(t_n), \mathbf{w}(0))$ with $W(t_n)$ as the *transition matrix*, and $\mathbf{w}(0) = w_i(0)$ as the initial probability distribution, i.e the probability of event i happening at the starting time $t = 0$ is given by $w_i(0)$, and the probability distribution $w_i(t_n)$ at the time t_n is given by $W_{ij}(t_n)$(Chaichian and Demichev, 2001):

$$w_i\left(t_n\right) = \sum_j W_{ij}\left(t_n\right) w_j(0), \quad n = 1, 2, 3...$$

and when the time $n\varepsilon$ has passed as

$$w_i(n\varepsilon) = \sum_j (W^n(\varepsilon))_{ij} w_j(0). \tag{5.2.2}$$

If we know the position $x = 0$ of the particle at the starting time $t_0 = 0$, we have $w_i(0) = 0$ for $i \neq 0$ and $w_0(0) = 1$, or, using the Kronecker delta[3],

$$\delta_{i0} = w_i(0). \tag{5.2.3}$$

[3]The Kronecker delta is a function of two variables, usually just non-negative integers. The function evaluates to 1 if the variables are equal, and 0 otherwise: $\delta_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$

Additionally, since w_i and W_{ij} are both probabilistic, we will always have

$$0 \leq w_i(0) \leq 1, \qquad \sum_i w_i(0) = 1$$
$$0 \leq W_{ij} \leq 1, \qquad \sum_i W_{ij} = 1.$$

For our example with the discrete step random walk (Brownian motion), the components of the infinite matrix $W(\varepsilon)$ are given by $W_{ij}(\varepsilon) = W(i\ell - j\ell, \varepsilon)$, and so after n time steps of duration ε the transition probabilities will be given by the product of the n matrices $W(i\ell - j\ell, n\varepsilon) = W^n(\varepsilon)_{ij}$.

Similarly, the probability distribution $w_i(n\varepsilon)$ after the time $n\varepsilon \geq 0$ will be given by

$$w_i(n\varepsilon) = \sum_j \left(W^n(\varepsilon)\right)_{ij} w_j(0).$$

The matrix notation of this is $w(n\varepsilon) = W^n w(0)$, and so viewing W^n as a function of the time variable n, this describes the time evolution of our system. So how do we define the transition matrix W? We start by introducing the two infinite operator matrices L and R which serves to change the particles position by an amount ℓ to the left and to the right respectively. The components of these matrices will be given by

$$L_{ij} = \delta_{(i+1)j}, \quad R_{ij} = \delta_{i(j+1)}$$

and thus we write L and R as

$$L = \begin{pmatrix} \ddots & \ddots & \cdots & \cdots & 0 \\ \vdots & 0 & 1 & & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ \vdots & & & 0 & 1 \\ 0 & \cdots & \cdots & \cdots & \ddots \end{pmatrix}, \quad R = \begin{pmatrix} \ddots & 0 & \cdots & \cdots & 0 \\ 1 & 0 & & & \vdots \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & & 1 & 0 & \vdots \\ 0 & \cdots & \cdots & \ddots & \ddots \end{pmatrix}.$$

A transition of the particle given by the action of L is therefore defined by

$$w_i \to w_i' = \sum_j L_{ij} w_i = w_{i+1}$$

meaning that the distribution has been shifted to the left. If we consider the distribution for the particle located at position k, $w_i(k) = \delta_{ik}$, we further notice that after letting L act on $w_i(k)$ we get $w_{i+1}(k) = \delta_{(i+1)k} = \delta_{i(k-1)}$, which tells us that the new position for the particle is at $k-1$.

Moving forward, we will require of the matrices L and R that they are invertible. If one were to take trivial examples of L and R using 4x4 matrices, one would find them to be singular and hence not invertible. Multiplying the 4x4 L and R, we get the results

$$L = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad R = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

$$LR = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad RL = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Seeing as the real matrices are of infinite size, meaning that our particle will never reach the edge of the system, we consider them to be invertible, resulting in $L = R^{-1}$ and so $RL = LR = I$. Connecting this with equation 5.2.1, we can write W as

$$W = \frac{1}{2}(R + L)$$

and find an expression for the transition matrix W^n after n transitions using the binomial formula

$$\mathrm{W}^n = \frac{1}{2^n}\sum_{k=0}^{n}\binom{n}{k}\mathrm{R}^k\mathrm{L}^{n-k} \quad = \frac{1}{2^n}\sum_{k=0}^{n}\binom{n}{k}\mathrm{R}^{2k-n} = \frac{1}{2^n}\sum_{k=0}^{n}\binom{n}{k}\mathrm{L}^{n-2k}.$$

For L and R, we can derive the components just by using ordinary matrix multiplication, leading to

$$L^n_{ij} = \delta_{(i+n)j}$$
$$R^n_{ij} = \delta_{i(j+n)}$$

Now we can expand our expression 5.2.1 for the probability of the particle to from point i to point j within one unit of time ε to define the probability of the particle moving from point i to point j during the time $n\varepsilon$

$$W(i\ell - j\ell, n\varepsilon) = \begin{cases} 0 & \text{if } |i-j| > n \\ & \text{or } (i-j)| + n \text{ is odd} \\ \frac{1}{2^n}\binom{n}{\frac{1}{2}(n+i-j)} & \text{if } |i-j| \le n \\ & \text{and } (i-j) + n \text{ is even.} \end{cases}$$

Similarly, we formulate the expression for the evolution of the probability distribution

$$w_i(n) = \begin{cases} 0 & \text{if } |i| > n \text{ or } (i+n) \text{ is odd} \\ \frac{1}{2^n}\binom{n}{\frac{1}{2}(n+i)} & \text{if } |i| \le n \text{ and } (i+n) \text{ is even} \end{cases} \tag{5.2.4}$$

In order for us to be able to define a difference equation and the time derivative of w, we write the location index i of $w_i(n)$ as an argument of w, $w(i\ell,\ n\varepsilon) \overset{\text{def}}{\equiv} w_i(n)$, and make use of a known recursion formula for binomial coefficients

$$\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1}.$$

If we let $x = i\ell, t = n\varepsilon$, we can write eq 5.2.4 as

$$w(x, t+\varepsilon) = \frac{1}{2}w(x+\ell, t) + \frac{1}{2}w(x-\ell, t).$$

and, using some algebraic manipulation as

$$\frac{w(x,t+\varepsilon) - w(x,t)}{\varepsilon} = \frac{\ell^2}{2\varepsilon}\frac{w(x+\ell,t) - 2w(x,t) + w(x-\ell,t)}{\ell^2}. \tag{5.2.5}$$

Since we are dealing with a discretized example of the continuous system, we now let $\ell \to 0, \varepsilon \to 0$ and thus fix the ratio $D = \frac{\ell^2}{2\varepsilon}$ letting the time and location variables x and t be continuous to better reflect

reality. Using D as the *diffusion constant*, we see that eq 5.2.5 is the *diffusion equation*[4]

$$\frac{\partial w(x,t)}{\partial t} = D\frac{\partial^2 w(x,t)}{\partial x^2}. \tag{5.2.6}$$

Seeing as we now have moved to continuous variables, we will naturally be dealing with integrals instead of summations and hence will be using the *Dirac delta function*[5] instead of the Kronecker delta leading to the continuous time version of the initial condition 5.2.3

$$w(x,t) \underset{t\to 0}{\longrightarrow} \delta(x).$$

We want to find a solution to the diffusion equation (5.2.6), and the way we do this is via a Fourier transform. We start by collecting the information we have

$$\begin{aligned} \frac{\partial w(x,t)}{\partial t} &= D\frac{\partial^2 w(x,t)}{\partial x^2} \\ w(x,0) = \delta(x) &= \begin{cases} w_0 \text{ if } x = 0 \\ 0 \text{ otherwise.} \end{cases} \end{aligned} \tag{5.2.7}$$

To solve this ordinary differential equation, we will use Fourier transforms. The first thing we do is write down $w(x,t)$ in terms of its Fourier transform

$$w(x,t) = \int_{-\infty}^{\infty} e^{ikx}\widetilde{w}(k,t)\ dk. \tag{5.2.8}$$

Next we take the Fourier transform of equation (5.2.6)

$$\mathbb{F}\left[\frac{\partial w}{\partial t}\right] = \mathbb{F}\left[D\frac{\partial^2 w}{\partial x^2}\right]$$

which, by the rules of derivation and convolution for Fourier transforms expands to

$$\begin{aligned} &\mathbb{F}\left[\frac{\partial w}{\partial t}\right] = \frac{\partial \widetilde{w}}{\partial t} \\ &\mathbb{F}\left[D\frac{\partial^2 w}{\partial x^2}\right] = D\cdot\mathbb{F}\left[\frac{\partial}{\partial x}\right]\mathbb{F}\left[\frac{\partial w}{\partial x}\right] = D(-ik)(-ik\widetilde{w}(k,t)) = -Dk^2\widetilde{w}(k,t) \end{aligned}$$

Our diffusion equation (5.2.6) is now reduced to the ODE

$$\frac{\partial \widetilde{w}(k,t)}{\partial t} = -Dk^2\widetilde{w}(k,t). \tag{5.2.9}$$

[4]The diffusion equation is a parabolic partial differential equation. It describes the macroscopic behavior of many micro-particles in Brownian motion, resulting from the random movements and collisions of the particles, and is related to Markov processes, such as random walks. The equation is usually written (with D constant) as the linear differential equation:

$$\frac{\partial \phi(\mathbf{r},t)}{\partial t} = D\nabla^2\phi(\mathbf{r},t)$$

where $\phi(\mathbf{r},t)$ is the density of the diffusing material at location r and time t and $D(\phi,\mathbf{r})$ is the collective diffusion coefficient for density ϕ at location r; and ∇ represents the vector differential operator del.

[5]The Dirac delta function $\delta(x)$ is a function on the real line which has the value zero everywhere except at the origin, where its value is infinite (which also satisfies $\int_{-\infty}^{\infty}\delta(x)dx = 1$)

$$\delta(x) = \begin{cases} +\infty, & x = 0 \\ 0, & x \neq 0 \end{cases}.$$

We solve this ODE using the integrating factor method, where the integrating factor I for this equation is $I = e^{\int Dk^2\ dt} = e^{Dk^2 t}$

$$\frac{\partial \widetilde{w}(k,t)}{\partial t} = -Dk^2 \widetilde{w}(k,t)$$
$$\Leftrightarrow \frac{\partial \widetilde{w}(k,t)}{\partial t} + Dk^2 \widetilde{w}(k,t) = 0$$

Multiplying this with the integrating factor and taking the integral of both sides yields

$$\int e^{Dk^2 t} \frac{\partial \widetilde{w}(k,t)}{\partial t} + e^{Dk^2 t} Dk^2 \widetilde{w}(k,t)\ dt = 0$$
$$\Leftrightarrow e^{Dk^2 t} \widetilde{w}(k,t) + \widetilde{c}_2(k) = 0$$
$$\Leftrightarrow \widetilde{w}(k,t) = \widetilde{c}_1(k) e^{-Dk^2 t}$$

The function $\widetilde{c}_1(k)$ is the Fourier transform of our initial condition (5.2.7), so we now need to find the Fourier image of $w(x,0) = \delta(x)$. The Dirac delta function is, as familiar defined as

$$\delta_\varepsilon(x-a) = \begin{cases} \frac{1}{\varepsilon} \text{ if } a < x < a+\varepsilon \\ 0 \text{ otherwise.} \end{cases}$$

Using this together with the mean value theorem for integral calculus

$$\int_a^b f(x)\ dx = (b-a) f(\xi), \quad a < \xi < b,$$

and the properties of the integral of the Dirac delta function we obtain

$$\int_{-\infty}^{\infty} \delta_\varepsilon(x-a) f(x)\ dx = \frac{1}{\varepsilon} \int_a^{a+\varepsilon} f(x)\ dx = \frac{1}{\varepsilon}(a+\varepsilon-a) f(\eta) = f(\eta), \quad a < \eta < a+\varepsilon.$$

If we now let $\varepsilon \to 0$, we find that

$$\int_{-\infty}^{\infty} \delta_\varepsilon(x-a) f(x)\ dx = f(a),$$

and so, the Fourier image of $\delta(x-a)$

$$\widetilde{\delta}(x-a) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\alpha t} \delta(x-a)\ dx = \frac{1}{2\pi} e^{i\alpha a},$$

which gives us for $a = 0$

$$\widetilde{\delta}(x) = \frac{1}{2\pi} = \widetilde{w}(k,0).$$

This leads us to the solution of our ODE (5.2.9)

$$\widetilde{w}(k,t) = \widetilde{w}(k,0) e^{-Dk^2 t} = \frac{1}{2\pi} e^{-Dk^2 t},$$

which gives us the solution to equation (5.2.8)

$$w(x,t) = \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{ikx} e^{-Dk^2 t}\ dk \tag{5.2.10}$$

If we use the well known value of the Gaussian integral

$$\int_{-\infty}^{\infty} \mathrm{e}^{-\alpha x^2} dx = \sqrt{\frac{\pi}{\alpha}} \tag{5.2.11}$$

and write the integrand of equation 5.2.10 on a common exponent

$$\mathrm{e}^{-Dk^2t}\mathrm{e}^{ikx} = \mathrm{e}^{-Dk^2t+ikx},$$

then, completing the square of the common exponent

$$-Dk^2t + ikt = -Dt\left(k^2 - \frac{ikx}{Dt}\right) = -Dt\left(\left(k - \frac{ix}{2Dt}\right)^2 + \left(-\frac{x^2}{4D^2t^2}\right)\right) = -Dt\left(k - \frac{ix}{2Dt}\right)^2 - \frac{x^2}{4Dt},$$

leads us to

$$\int_{-\infty}^{\infty} \frac{1}{2\pi}\mathrm{e}^{-Dk^2t}\mathrm{e}^{\mathrm{i}kx}dk = \exp\left\{-\frac{x^2}{4Dt}\right\}\frac{1}{2\pi}\int_{-\infty}^{\infty}\exp\left\{-Dt\left(k - \frac{ix}{2Dt}\right)^2\right\}dk.$$

If we now perform a variable substitution

$$\begin{cases} y = k - \frac{ix}{2Dt} \\ dy = dk, \end{cases}$$

we will obtain an expression for the probability distribution $w(x,t)$ (which is also the solution to the diffusion equation 5.2.6 with initial condition 5.2)

$$w(x,t) = \exp\left\{-\frac{x^2}{4Dt}\right\}\frac{1}{2\pi}\int_{-\infty}^{\infty}\exp\left\{-Dty^2\right\}dy = \exp\left\{-\frac{x^2}{4Dt}\right\}\frac{1}{\sqrt{4\pi Dt}}.$$

Now, once again using the result from equation 5.2.11, we can see that

$$\int_{-\infty}^{\infty} w(x,t)dx = \int_{-\infty}^{\infty}\frac{1}{\sqrt{4\pi Dt}}\exp\left\{-\frac{x^2}{4Dt}\right\}dx = \frac{1}{\sqrt{4\pi Dt}}\sqrt{4\pi Dt} = 1.$$

We are now working with continuous variables, so we will also need a new expression for our transitional probability $W(i\ell - j\ell,\ n\varepsilon)$. Instead of using the notations $x = i\ell, x = j\ell$, we write the infinitesimal transition probability with start and end points $(x_0,t_0),\ (x_t,t)$, corresponding to the discrete one as

$$W_{ij}^N = W(x_t,t|x_0,t_0).$$

This leads us to the expression for the probability density $w(x_t,t)$, which is the continuous version of equation 5.2.2

$$w(x_t,t) = \int_{-\infty}^{\infty} W(x_t,t|x_0,t_0)w(x_0,t_0)dx_0.$$

The transition probability now satisfies the diffusion equation

$$\frac{\partial W(x_t,t|x_0,t_0)}{\partial t} = D\frac{\partial^2 W(x_t,t|x_0,t_0)}{\partial x_t^2}, \quad t > 0$$

which, with initial condition $W\left(x_t,t|x_0,t_0\right) \underset{t\to t_0}{\longrightarrow} \delta\left(x_t - x_0\right)$ has the solution

$$W(x_t,t|x_0,t_0) = \frac{1}{\sqrt{4\pi D(t-t_0)}}\exp\left\{-\frac{(x_t-x_0)^2}{4D(t-t_0)}\right\}$$

attained by using the same methods for solving diffusion equations as above.

5.3 The path integral

Let us now consider the probability of a particle being within a certain section of the plane or within a certain "gate" consisting of endpoints A and B at the time t.

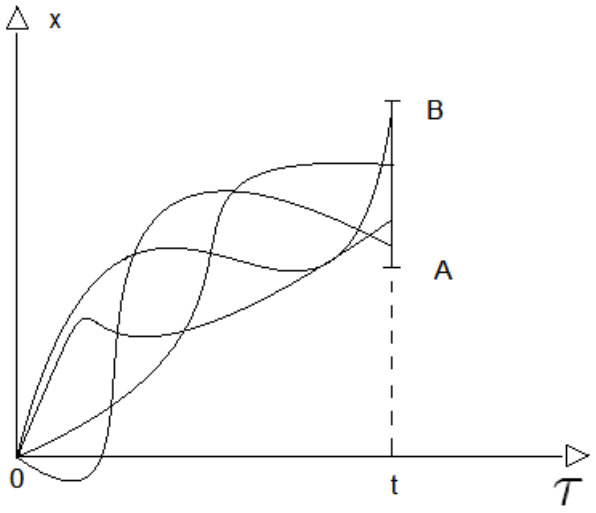


Figure 6: Trajectories of a Brownian particle starting from the origin and ending anywhere in the gate AB at the moment t.

We will denote the probability of the particle being anywhere within the section $[AB]$ as

$$\mathbb{P}\{x(t) \in [AB]\} = \int_A^B w(x,t)dx$$

which we now know can also be expressed as

$$\int_A^B \exp\left\{-\frac{x^2}{4Dt}\right\}\frac{1}{\sqrt{4\pi Dt}}dx.$$

To integrate along an entire path though, we will need to increase the amount of gates. We now consider a particle starting at (x_0, t_0), ending up somewhere in the gate $[A_nB_n]$, and passing through all the gates $[A_1B_1], [A_2B_2], \ldots, [A_{n-1}B_{n-1}]$ along the way

Given the independent probabilistic nature of the Brownian motion, we get the expression for the probability of the particle following this particular path, passing through these particular gates as the product of the individual probabilities

$$\begin{aligned}
&\mathbb{P}\left\{x\left(t_1\right) \in \left[A_1B_1\right], x\left(t_2\right) \in \left[A_2B_2\right], \ldots, x\left(t_N\right) \in \left[A_NB_N\right]\right\} \\
&\quad = \int_{A_1}^{B_1} \frac{\exp\left\{-\frac{x_1^2}{4Dt_1}\right\}}{\sqrt{4\pi Dt_1}}dx_1 \int_{A_2}^{B_2} \frac{\exp\left\{-\frac{(x_2-x_1)^2}{4D(t_2-t_1)}\right\}}{\sqrt{4\pi D\left(t_2-t_1\right)}}dx_2 \\
&\quad \times \int_{A_3}^{B_3} \frac{\exp\left\{-\frac{(x_3-x_2)^2}{4D(t_3-t_2)}\right\}}{\sqrt{4\pi D\left(t_3-t_2\right)}}dx_3 \cdots \int_{A_N}^{B_N} \frac{\exp\left\{-\frac{(x_N-x_{N-1})^2}{4D(t_N-t_{N-1})}\right\}}{\sqrt{4\pi D\left(t_N-t_{N-1}\right)}}dx_N.
\end{aligned} \tag{5.3.1}$$

Once again, since we are dealing with continuous time we will let $t_i - t_{i-1} = \Delta t_i \to 0,\ 1 \leq i \leq N$, and thus increase the amount of gates infinitely, while letting the size of each gate go to zero $B_i - A_i \to 0$. In doing so, we will naturally consider the limit as $\Delta t_i \to 0$ of equation 5.3.1 to be the probability of the particle following the path $x(\tau)$. Consider now the term

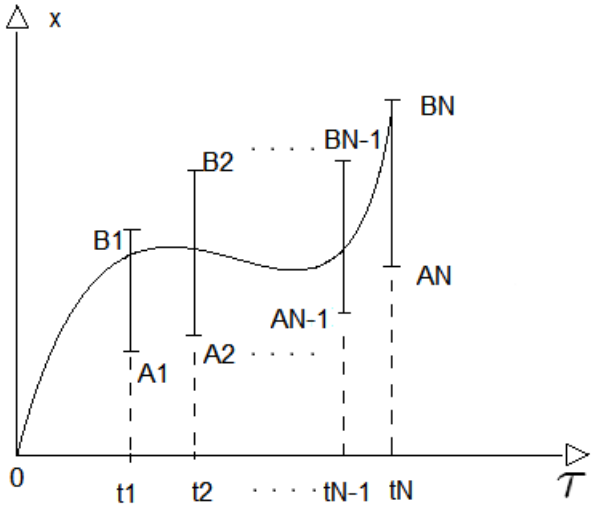


Figure 7: A trajectory of a Brownian particle starting from origin and passing through the gates A_iB_i at the times $t_i (i = 1, \ldots, N)$

$$\int_{A_{i+1}}^{B_{i+1}} \frac{\exp\left\{-\frac{(x_{i+1}-x_i)^2}{4D(t_{i+1}-t_i)}\right\}}{\sqrt{4\pi D\left(t_{i+1}-t_i\right)}} dx_{i+1},$$

which can equivalently be expressed as

$$\frac{1}{\sqrt{4\pi D\left(t_{i+1}-t_i\right)}} \int_{A_{i+1}}^{B_{i+1}} \exp\left\{-\frac{\left(x_{i+1}-x_i\right)^2}{4D\left(t_{i+1}-t_i\right)}\right\} dx_{i+1}.$$

Since the size of the gates $\to 0$,

$$\begin{aligned} &\lim_{\substack{\Delta t_i \to 0 \\ N\to\infty}} \exp\left\{-\sum_{i=1}^{N} \frac{\left(x_i - x_{i-1}\right)^2}{4D\left(t_i - t_{i-1}\right)}\right\} \prod_{i=1}^{N} \frac{dx_i}{\sqrt{4\pi D\left(t_i - t_{i-1}\right)}} \\ &= \lim_{\substack{\Delta t_i \to 0 \\ N\to\infty}} \exp\left\{-\frac{1}{4D} \sum_{i=1}^{N} \left(\frac{x_i - x_{i-1}}{t_i - t_{i-1}}\right)^2 \Delta t_i\right\} \prod_{i=1}^{N} \frac{dx_i}{\sqrt{4\pi D \Delta t_i}} \\ &\equiv \exp\left\{-\frac{1}{4D} \int_0^t \dot{x}^2(\tau) d\tau\right\} \prod_{\tau=0}^{t} \frac{dx(\tau)}{\sqrt{4\pi D d\tau}}. \end{aligned} \tag{5.3.2}$$

When analyzing the jumps caused by rare events in dynamical systems, we know the starting point (x_0, t_0), and the endpoint x_n, t_n, and in retrospect we can see the exact path the particle has taken. But if we were to try and predict which path it would take beforehand, we would have to consider all the possible paths from the starting point to the end point and then find the most probable one of them.

To do so, let us define the set $\mathcal{C}$ of paths with certain starting and end points as the following:

- $\mathcal{C}\{x_1, t_1; x_2, t_2\}$ denotes the set of trajectories starting at the point (x_1, t_1) and having the endpoint (x_2, t_2)
- $\mathcal{C}\{x_1, t_1; [AB], t_2\}$ denotes, in the one-dimensional case, the set of trajectories with the starting point $x_1 = x(t_1)$ and ending in the gate $[AB]$ at the time t_2.

So in order to get an expression for the probability of the particle ending up somewhere within the gate $[AB]$, we have to sum up all the probabilities from equation 5.3.2 for all the possible paths that end up in the gate $[AB]$ over the set $\mathcal{C}\{0, 0; [AB], t\}$

$$\begin{aligned}\mathbb{P}\{x(t) \in [AB]\} &= \int_{\mathcal{C}\{0,0;[AB],t\}} \prod_{\tau=0}^{t} \exp\left\{-\frac{1}{4D}\int_0^t d\tau \dot{x}^2(\tau)\right\} \frac{dx(\tau)}{\sqrt{4\pi D d\tau}} \\ &= \int_A^B \frac{1}{\sqrt{4\pi D t}} \exp\left\{-\frac{x^2}{4Dt}\right\} dx\end{aligned}$$

where $\int_{\mathcal{C}\{0,0;[AB],t\}}$ represents the summation over the set of paths. This kind of summation over a set of paths is what is known as a *Wiener path integral*.

6 Rare events

The type of rare event studied in this thesis is one of a Brownian particle moving in a *double-well potential*. Figure 8 shows on the left hand side an illustration of a Brownian particle resting at the minima of one of the wells in the double well potential, and the right hand side shows the trajectory of the Brownian particle as a function of time. The motion of the particle here is restricted within one well, moving up the sides and back down to the minima. The trajectory produced by the movement of the Brownian particle will follow a stable periodic orbit around the local minima specified by its initial condition, while being influenced by the noise term $\sqrt{2}\sigma\xi(t)$ and hence exhibit fluctuations of order σ. When a rare event occurs however, the particle will escape the current well and transcend the barrier leading to the other well. These wells represent the steady states of the particle's movement. To simulate a rare event in this thesis, we will run a numerical simulation for a long time until we observe a jump. The specifics of this simulation is discussed in the next section.

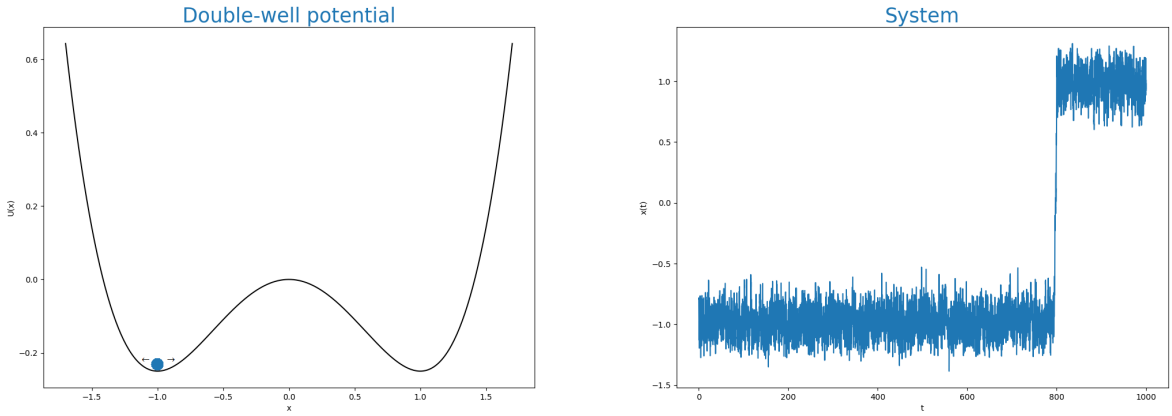


Figure 8: Brownian particle moving in a double well potential, along with visualization of the system before, during and after a rare event (jump) occurs.

The trajectory $x(t)$ here was produced by simulating the following SDE known as a *Langevin equation*[6](Giorgini et al., 2019)

$$\frac{dx}{dt} = \dot{x} = F(x(t), t) + \sqrt{2}\sigma\xi(t) \tag{6.0.1}$$

where $\sigma = \frac{0.12}{\sqrt{dt}}$ and

$$F(x, t) = -U'(x) + 0.7\cos(2\pi t),$$

where $U(x)$ is a double-well potential, $U(x) = -\frac{x^2}{2} + \frac{x^4}{4}$, $0.7\cos(2\pi t)$ is an external periodic force and $\xi(t)$ is zero mean Gaussian white noise with correlation function

$$\langle \xi(t)\xi(s) \rangle = \delta(t - s)$$

with δ being the Dirac delta function. The only non-deterministic part of equation 6.0.1 is $\xi(t)$, and so if we set $\sigma = 0$ the particle will just stay at the local minima specified in the initial condition. As we increase the value of σ, the particle will begin to oscillate around the minima within a radius proportional to σ.

[6]In physics, the Langevin equation is a stochastic differential equation describing the time evolution of a subset of the degrees of freedom. The original Langevin equation describes Brownian motion.

In order for the particle to escape the potential well it is contained in, one of two things need to happen:

1. The base oscillation is large, i.e σ needs to be large
2. The noise $\xi(t)$ deviates significantly from zero.

If we have a large base oscillation, then the event of the particle escaping the potential well is not rare and therefore this scenario is not interesting for us.

For the purpose of finding some indication that a rare event is about to occur, we will run the simulations with σ as small as possible, thus making sure that our rare event is in fact rare.

In order to find a precursor that will warn us that a rare event is about to occur, we will focus on the behaviour of the system within a short time period leading up to the jump. We want to find the most probable path between the point (x_i, t_i) which lies within one potential well, and (x_f, t_f) which lies at the maximum of the barrier separating the two wells. Once the particle reaches this point, all the possible paths it can take from there leads down into the other well. Since the system is influenced by the non-deterministic noise, we know from section 5 that there exists an infinite number of paths the particle can take between the two points, but also that there is one which is much more probable than all the others. In fact, since we are dealing with an oscillating system which progresses for a long time, there are going to be one such optimal path for each period of the system that passes between the time t_i and t_f.

The actual derivation of the construction of the optimal path is explained in further detail in (Giorgini et al., 2019) and (Lehmann et al., 2000), however this is beyond the scope and indeed level of this thesis, but we will discuss the outlines of the prediction scheme.

To find the most probable path, one uses path integrals in order to find the path of least *action*(Sussman et al., 2001). In physics, the path that is actually followed within a dynamical system is the path for which the action is minimized. Action can be described as a measurement for the overall motion of each path, and hence the most probable or most direct path will be the path with the least action.

The optimal path in each period we denote by $x_k(t)$ and the noise for the same path, i.e the optimal noise behaviour, is denoted $p_k(t)$. These paths and momenta are shown to satisfy the system of differential equations

$$\begin{aligned} \dot{x}_k(t) &= 2p_k(t) + F(x_k(t), t) \\ \dot{p}_k(t) &= -p_k(t)F'(x_k(t), t). \end{aligned} \tag{6.0.2}$$

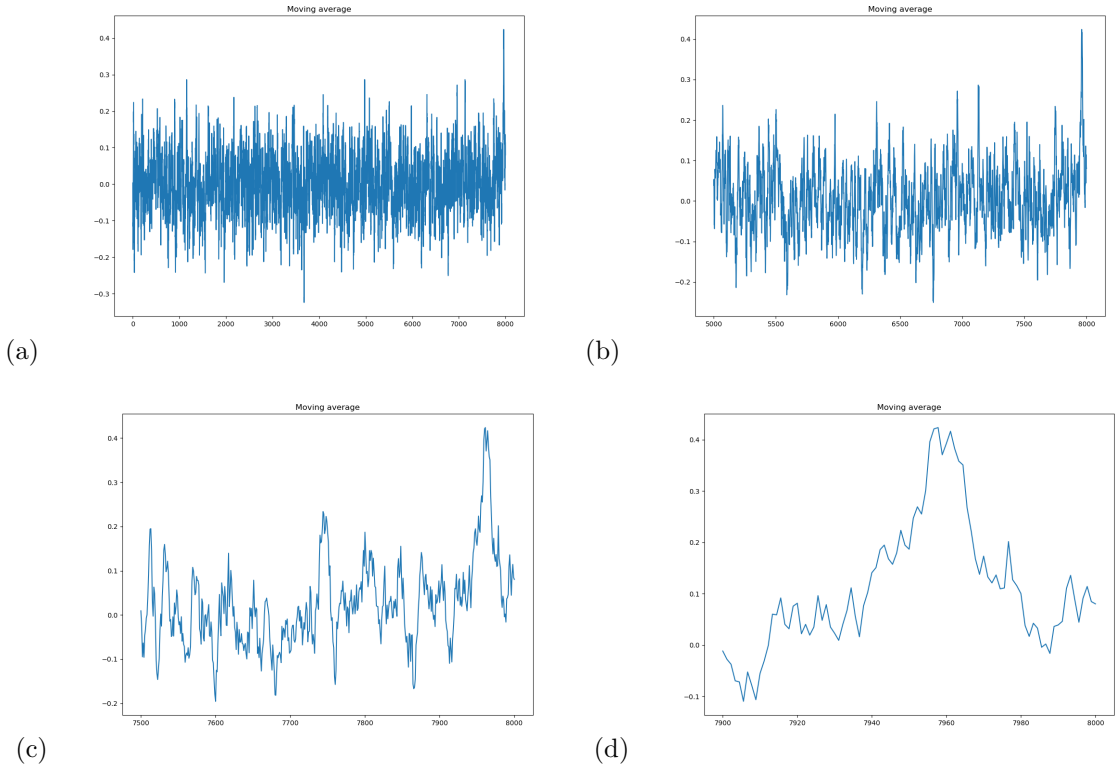
where the momenta is defined as $p_k(t) := \frac{1}{2}(\dot{x}_k(t) - F(x_k(t), t))$. Comparing equation (6.0.1) and (6), we can extract the noise $\xi(t)$

$$\xi(t) = \frac{1}{\sqrt{2}\sigma}(\dot{x}(t) - F(x(t), t)) \tag{6.0.3}$$

and see that the optimal conditions for the jump is when $\xi(t) \to \frac{\sqrt{2}p(t)}{\sigma}$. When the system approaches a rare event, the momentum of the optimal path is expected to exhibit an exponential growth according to $p(t) = p_0 e^{-\lambda_s(t-t_0)}$(Giorgini et al., 2019). If we extract the noise of $x(t)$ using equation (6.0.3), we can monitor its behaviour, so that if and when it starts to show increasing irregularities we can compare its behaviour to that of the optimal condition $\xi(t) = \frac{\sqrt{2}p(t)}{\sigma}$, we can make a prediction about the jump.

When we study the system retroactively, we know within which period the event will occur and so we only need to consider one of these optimal paths. Using this path, we extract the noise then we analyze it and use its characteristics for our prediction. One way we can analyze the noise is to construct a moving

average of varying window sizes, and try to find some sort of unique deviation. This can in turn act as a kind of indicator that a rare event is imminent. In figure (9) below we have constructed such moving averages of window size 10 for the same noise during varying stretches of the time from 0 to the time of the jump at approximately 800.



(a) (b)

(c) (d)

Figure 9: Moving averages of the same $p(t)$ plotted over times: (a) 0 to 8000, (b) 5000 to 8000, (c) 7500 to 8000, (d) 7900 to 8000. This to clarify that this spike in moving averages only occurs once, in close proximity to the jump.

7 Numerical methods

The method of integration used for the simulations in this thesis is one called *Leapfrog integration*. This is a second order method that uses the recurrence relation

$$v_{i+\frac{1}{2}} = v\left(x_i + \frac{dt}{2}, t_i + \frac{dt}{2}\right),$$
$$x_{i+1} = x_i + v_{i+\frac{1}{2}}dt, \quad i = 0, 1, 2, ...$$

In this case we will specify the initial conditions $v_{\frac{1}{2}}$ and x_0, but one can also use some sort of self starting scheme like for example the Euler method to acquire these values. As we can see from the recurrence relation, it progresses the velocity v at half step indices while the position x progresses at integer indices. This causes the process of progression for the two variables to "leapfrog" over each other, which is why the scheme is so named.

The leapfrog integration method is implemented using the SDE 6.0.1 as velocity

$$v_i = v(x_i, t_i) = \frac{dx_i}{dt_i} = F(x(t_i), t_i) + \sqrt{2}\sigma\xi(t_i) = x_i - x_i^3 + 0.7\cos(2\pi t_i) + \sqrt{2}\sigma\xi(t_i)$$

with initial conditions

$$\begin{aligned} t_0 &= 0, \\ dt &= 0.1, \\ \sigma &= \frac{0.12}{\sqrt{dt}}, \\ x(t_0) &= x_0 = -1, \\ v_{\frac{1}{2}} &= F(x_0, t_0) + \sqrt{2}\sigma\xi(t_i) \end{aligned}$$

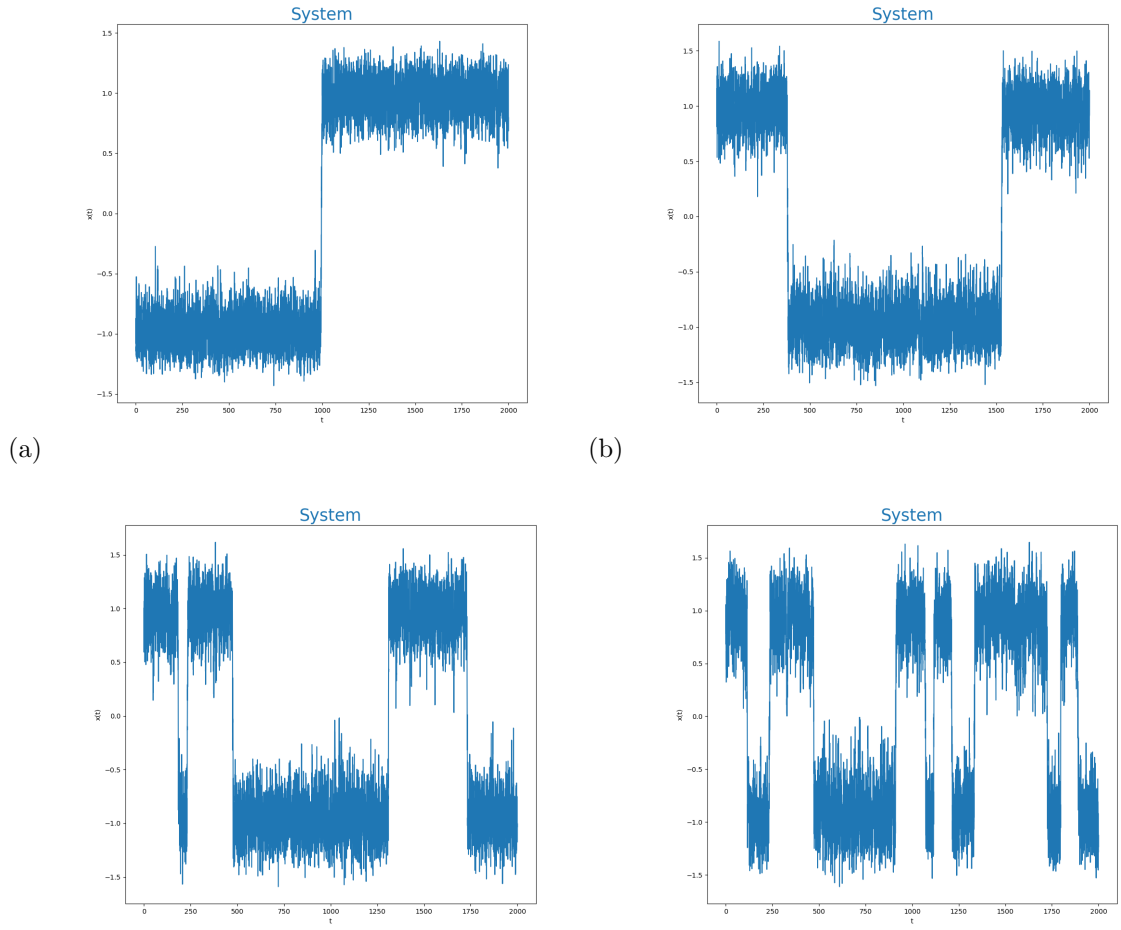
with $\xi(t_i)$ taken from a Box-Muller transform[7] (Box and Muller, 1958). We let the simulation run for a long time until we encounter a jump from one potential well to the other, at which point we plot the behaviour of the system along with the underlying noise before, during and after this jump occurs. As stated, how we choose the value of σ determines the rareness of the event, so a larger value of σ would yield a less rare event and thus would need less deviation of the noise, whereas a smaller value of σ makes the event more rare and the analysis of the noise easier.

After obtaining data from a simulation that resulted in a rare event, we can plot the noise and the system to see how they have behaved in relation to each other. We can also analyze the noise in different ways, to see if some criterion have been fulfilled for its behaviour, such as accumulation of its moving average or a spike in the data.

[7] The Box-Muller transform is a method for generating pseudo-random pairs of independent, standard, zero mean, unit variance normally distributed random numbers, given a source of uniformly distributed random numbers.

7.1 Numerical Results

Our simulations have yielded various results showing a varying amount of jumps within the same period of time, 200 periods. Values of σ between 0.38 and 0.80 have been used in numerous simulations, of which a selection of visualizations is included below in figure 10.



(a) (b)

(c) (d)

Figure 10: Visualizations of simulations of the Langevin equation (6.0.1) with various values of σ, demonstrating that larger values of σ increase the probability of the particle transitioning from one potential well to another. (a) $\sigma \approx 0.38$, (b) $\sigma \approx 0.60$, (c) $\sigma \approx 0.70$, (d) $\sigma \approx 0.80$.

In all of these cases, we can see the same kind of spike in the moving average in close proximity to the jumps. However, as we can see in figure 9, the base oscillation of the system increases as the value of σ increases. This makes the jump less likely and thus less relevant to our studies, since they concern rare events. If we look closely at figures 11-14 below, we can see that the moving average of the noise exhibits a positive spike shortly before a jump from the potential well around -1 to the well around 1, and it exhibits a negative spike shortly before a jump from the well around 1 to the well around -1. This can be expected since intuitively, if we look again at figure 8, clearly in order for the particle to leave the left well and transition to the right well, the direction of the noise impacting it would have to be positive, moving the particle to the right, and it would have to be negative to shift the particle to the left.

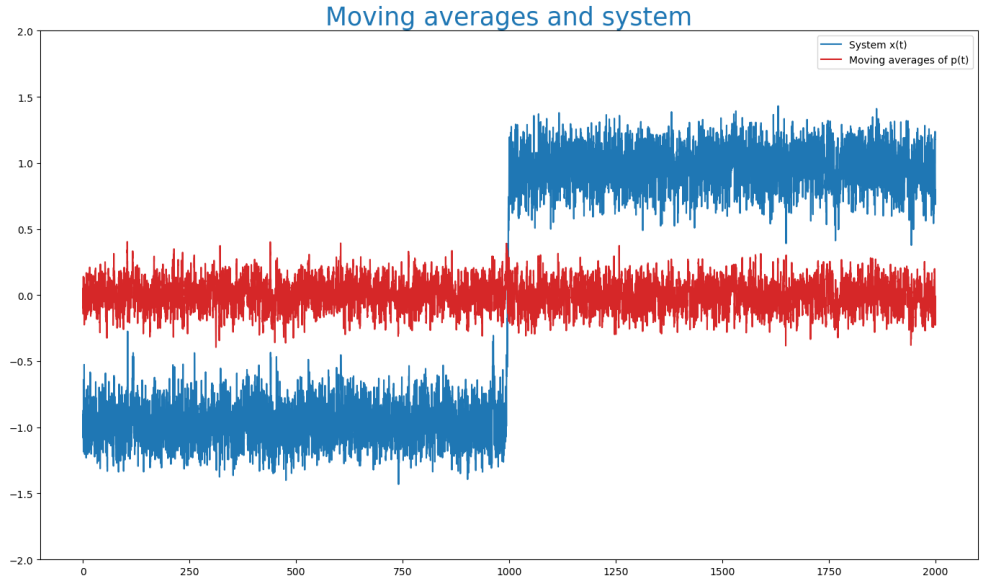


Figure 11: Plot of a system with one jump along with the moving average of the noise

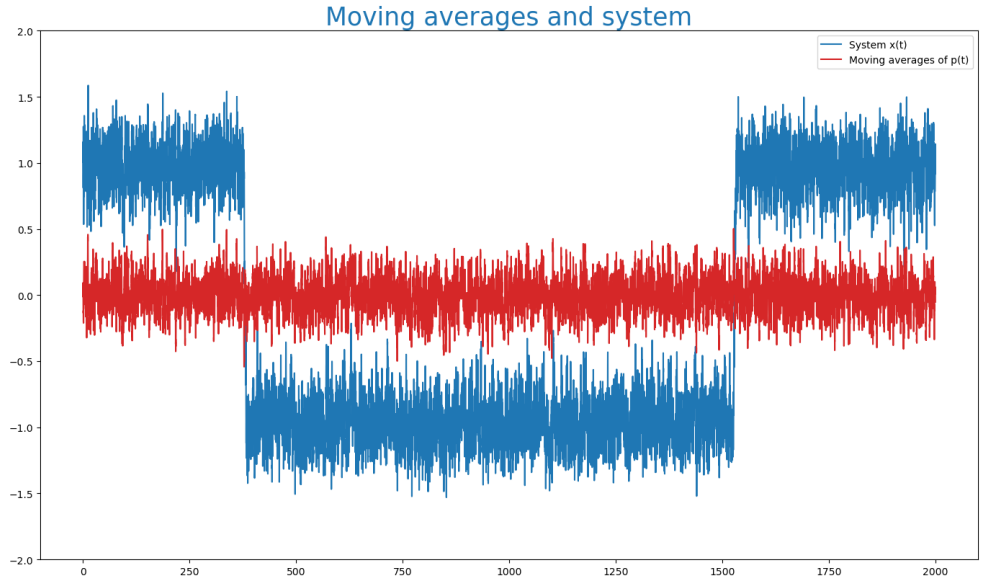


Figure 12: Plot of a system with two jumps along with the moving average of the noise

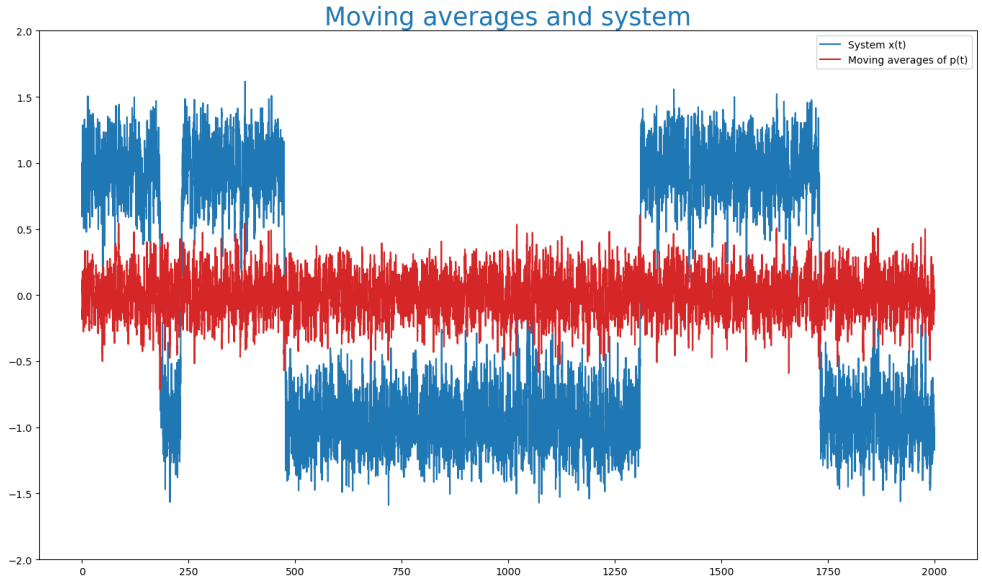


Figure 13: Plot of a system with five jumps along with the moving average of the noise

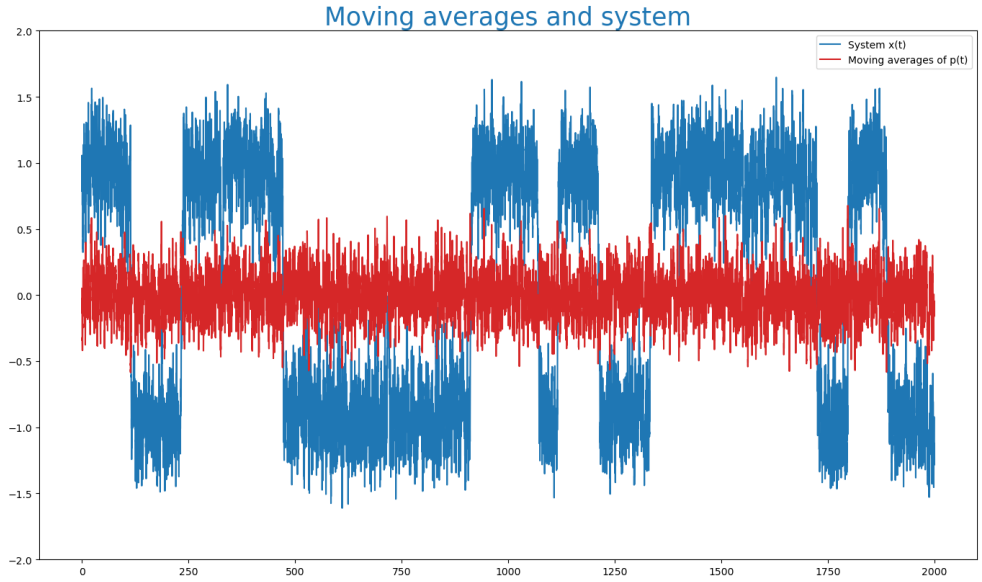


Figure 14: Plot of a system with eleven jumps along with the moving average of the noise

The above examples are made with a moving average of the noise, but one can also identify a spike in the data looking only at the behaviour of the noise itself. This is demonstrated in figure (15), where we have plotted the same system as figure (10(a)). Here we have plotted the system along with the noise scaled down by a factor of $\frac{1}{5}$ in order to make visual assessment easier.

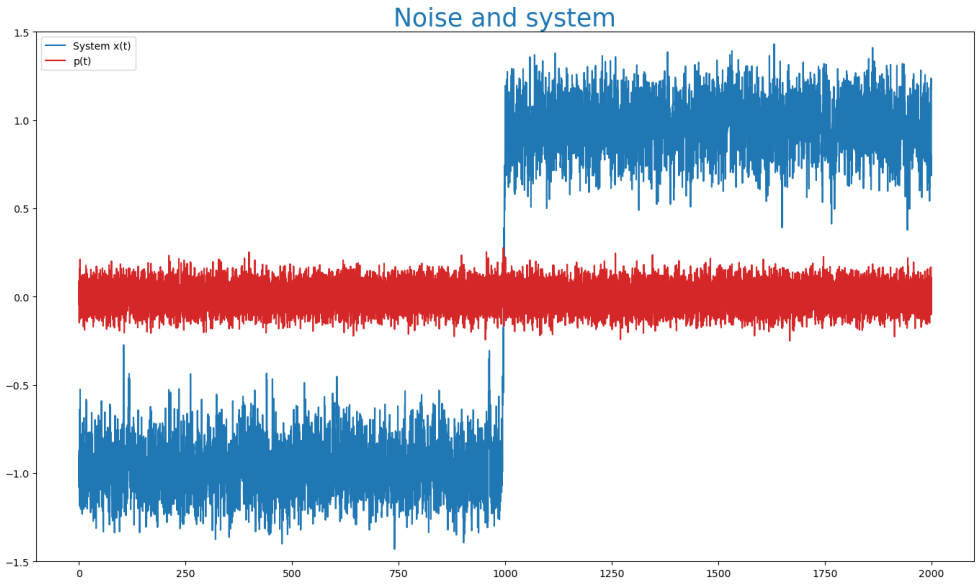


Figure 15: Visualization of the system $x(t)$ and the underlying noise $p(t)$ over 2000 time steps.

As we can see in figure 15, the scaled down noise $p(t)$ oscillates steadily around zero until it exhibits a spike shortly before the system jumps. It can seem that there are more spikes in the noise, and indeed there is, but there is only one spike here where the noise increases exponentially in accordance with the exponential growth discussed in section 6, and that is the spike immediately before the jump.

8 Conclusion

We have studied methods of finding an early warning indicator of an upcoming rare event, given a dynamic system governed by a stochastic differential equation. The methods studied have shown that there exists a way to find such an early warning indicator using stochastic calculus and path integrals. In our trials we have used varying values of σ to see how it affects the system and the derivation of the early warning indicator, where we have found that a smaller value of σ yields a lower probability for the system to experience a rare event causing the particle to travel from one potential well to another. This in turn makes the early warning indicator easier to find since it means that the deviation of the noise needs to be larger for the jump to occur.

There is certainly more to be studied within the field of path integrals, and though we have briefly discussed definitions and their part in the prediction scheme, the main theory of path integrals lies beyond the level of this thesis and we will leave it to the reader to further pursue this theory within other materials, such as (Chaichian and Demichev, 2001).

References

S. E. Alm and T. Britton. *Stokastik*. Liber, 2008.

G. E. P. Box and M. E. Muller. A note on generation of random normal deviates, 1958.

Z. Brzezniak and T. Zastawniak. *Basic stochastic processes: a course through exercises*. Springer Science & Business Media, 2000.

M. Chaichian and A. Demichev. *Path Integrals in Physics: Volume I Stochastic Processes and Quantum Mechanics*. CRC Press, 2001.

R. P. Dobrow. *Introduction to Stochastic Processes with R, First Edition*. John Wiley & Sons, inc, 2016.

L. C. Evans. *An Introduction To Stochastic Differential Equations*. American Mathematical Society, 2014.

L. Giorgini, S. Lim, W. Moon, and J. Wettlaufer. Predicting rare events in stochastic resonance, 2019.

J. Lehmann, P. Reimann, and P. Hänggi. Physical review e 62, 6282, 2000.

G. Sussman, J. Wisdom, and M. Mayer. Structure and interpretation of classical mechanics. 2001, 2001.