JOHANNES KEPLER
UNIVERSITÄT LINZ

Submitted by
Johannes K. Kröpfl

Submitted at
Institute for Analysis

Supervisor
Dipl.-Math.
Dr. Mario Ullrich

June 2022

Using Brownian Motion and Stochastic Calculus for deriving the Black-Scholes-Merton Model of Investing

Bachelor Thesis
to obtain the academic degree of
Bachelor of Science
in the Bachelor's Program
Artificial Intelligence

JOHANNES KEPLER
UNIVERSITY LINZ
Altenbergerstraße 69
4040 Linz, Österreich
www.jku.at
DVR 0093696

Abstract

In this thesis we want to look at Brownian motion, apply principles of Stochastic Calculus and then derive the Black-Scholes-Merton model of investing. The goal of this thesis is to get insight into the theoretical underpinnings of the Black-Scholes-Merton model of investing. To achieve this goal, we will make use of the book [3] by Steven E. Shreve and the book [6] by Paul Wilmott. This thesis is mostly theoretical, but includes selected examples. In the subsection "Implied Volatility" of the last chapter we will see a practical example of how to price a European call options fairly.

Contents

1 Introduction

Random motion is an interesting concept because we cannot make predictions about the future in the present state. An example of random motion is **Brownian motion**, named after the botanist Robert Brown. He observed pollen moving seemingly at random under a microscope. The behavior of the pollen in the given environment is called Brownian motion. Later, the mathematician Norbert Wiener proved many of the properties of this process, which is why one-dimensional Brownian motion is also called the **Wiener process**. This behavior is used in areas such as option pricing in financial markets and statistical mechanics. We are interested in the financial applications. We will start out by explaining **Random Walk** and its properties, then go to the Scaled Symmetric Random Walk and derive Brownian Motion. Important properties are also covered.

When we deal with stochastic processes such as Brownian motion, we face the problem that it is nowhere differentiable. Therefore, we cannot use Standard Calculus to compute derivatives and integrals of Brownian motion. This problem was solved by the mathematician Kiyosi Itô, who introduced a new form of calculus called **Itô-Calculus**. With this new form of calculus, also called **Stochastic Calculus**, we are able to calculate derivatives and integrals of Brownian motion.

Finally, **Black, Scholes and Merton** made use of Stochastic Calculus to generate a model, that prices European options fairly. Their model of investing was awarded the Nobel Prize in 1997 and is special, because it gives a closed form solution. In practice, we don't want to estimate the value of an option, but estimate **implied volatility** to give us indications about whether an option is mispriced or not.

2 Brownian Motion

Brownian Motion can be derived from a **Symmetric Random Walk (SRW)**. Originally, Brownian Motion was derived differently, but we will end up with the same result by deriving it from the SRW. This way of deriving the Brownian Motion is probably the easiest one to grasp. All throughout this section we will use the Book [3] by Steven E. Shreve to illustrate various concepts, if not stated otherwise.

2.1 Symmetric Random Walk

We create a **Symmetric Random Walk (SRW)** by tossing a fair coin successively. There are 2 outcomes in that case, where the outcome Tails (T) has a probability of $p = \frac{1}{2}$ and the outcome Heads (H) has a probability of $q = 1 - p$. The outcome of the n^{th} toss is stored in ω_n.

Definition 2.1.1 (Symmetric Random Walk)**.** Let

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

then we can define the symmetric Random Walk M_k as

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

where $M_0 = 0$ and $k = 1, 2, 3...$.

So a SRW up to k is just the sum of all the random variables X_j (with $p = q$) up to k, where $j = 1, 2, ..., k$. The reason for calling the Random Walk symmetric lies in the fact that we've defined $p = q$, which is the condition for a fair coin in our example.We can look at a SRW visually for $k = 100$:

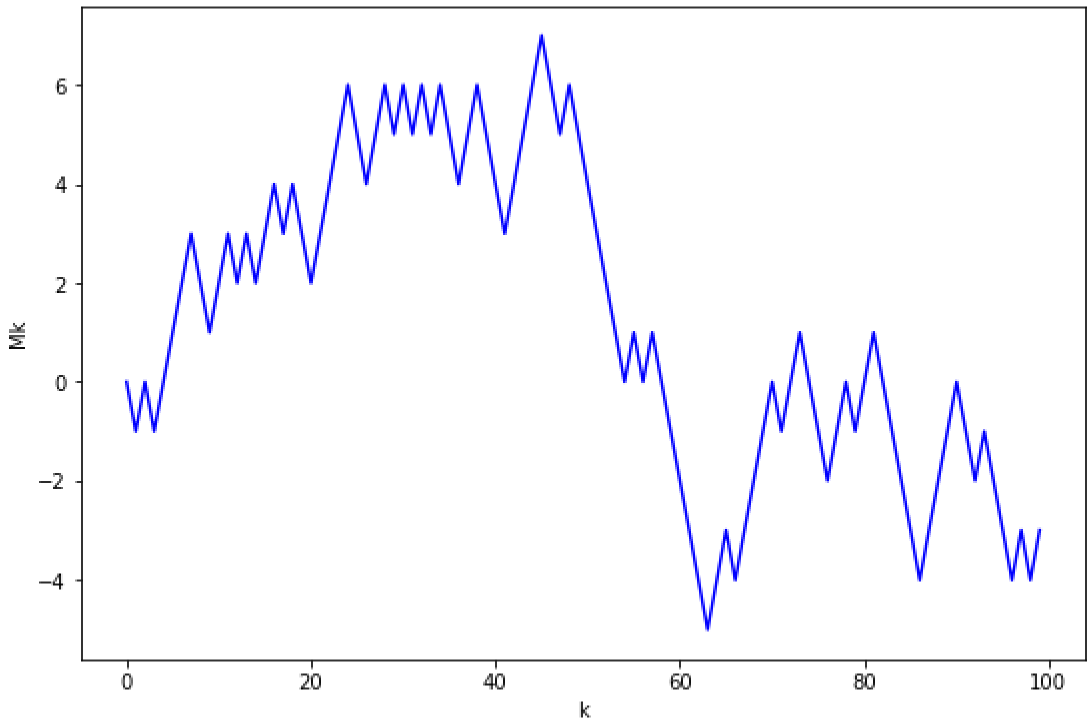

Figure 2.1: SRW up to $k = 100$

In Figure 2.1 we've now visualized that we go up by 1 if the outcome of a fair coin toss is Heads and go down by 1 if the outcome is Tails. The SRW is defined on discrete time and we cannot connect the dots. However, we used linear interpolation between the points where the SRW is defined for visualization purposes. The SRW is a discrete process, although Figure 2.1 might look like a continuous one.

Theorem 2.1.2 (Independent increments of SRW)**.** When we talk about **increments** of a SRW, then we mean the change in position of the SRW between k_i and k_{i+1}. Since we've sampled independently (in our example by fair coin toss), those increments are called **independent increments** and can be written as

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

We've defined independent increments of a SRW in Theorem 2.1.2 and are able to make some statements about the expected value and variance of those increments. We get that

$$E[M_{k_{i+1}} - M_{k_i}] = E[X_{i+1} + M_{k_i} - M_{k_i}] = E[X_{i+1}] = 0$$

and

$$Var[M_{k_{i+1}} - M_{k_i}] = \sum_{j=k_i+1}^{k_{i+1}} Var[X_j] = \sum_{j=k_i+1}^{k_{i+1}} \left(E[X_j^2] - E[X_j]^2\right) = \sum_{j=k_i+1}^{k_{i+1}} (1-0) = k_{i+1} - k_i \ .$$

So the expected value of an increment is 0 and the variance of an increment is 1. SRW has an important property, namely that it is a martingale. For a stochastic process like the SRW to be called a martingale, we need to define an associated filtration with it. Let's define a filtration, before defining a martingale. When we talk about Ω in the definition, then we mean the possible outcomes. In the coin toss example those are heads and tails. So for the coin toss example we get $\Omega = \{H, T\}$.

Definition 2.1.3 (Filtration)**.** To define a filtration, we need the nonempty set Ω. Then we can define for every $t \in [0, T]$ the σ-algebra $\mathcal{F}(t)$. Assume that there is an $s \in [0, T]$ with $s \leq t$, then $\mathcal{F}(s) \subseteq \mathcal{F}(t)$. This makes $\mathcal{F}(t)$ a collection of σ-algebras also known as filtration.

The concept of a filtration $\mathcal{F}(t)$ can be seem as the information contained in a random experiment up to t. In Definition 2.1.3 t doesn't have to be a natural number, but we could have added that as a condition if we wanted to. In that case our notation shift from $\mathcal{F}(t)$ for the continuous case to $\mathcal{F}_t$ for the discrete case. We can also say, that some stochastic process $\{X_t\}_{t=0}^T$ like SRW is adapted to a filtration $\mathcal{F}$, where $\mathcal{F}_1 \subseteq ... \subseteq \mathcal{F}_t \subseteq ... \subseteq \mathcal{F}$. This is often also called non-anticipating and just means that we cannot see into the future. A deep understanding of filtration is not necessary to follow along. One can read more on filtration on page 49 of the Book [3]. Now we can define a martingale.

Definition 2.1.4 (Martingale in the discrete case)**.** Assume we are given a stochastic process $\{X_t\}_{t=0}^T$ that is adapted to the filtration $\mathcal{F}$, where $\mathcal{F}_1 \subseteq ... \subseteq \mathcal{F}_t \subseteq ... \subseteq \mathcal{F}$. Then we call $\{X_t\}_{t=0}^T$ a **martingale** if

$$E[X_{t+1} | \mathcal{F}_t] = X_t \ .$$

Theorem 2.1.5 (SRW is a martingale)**.** With a small shift in notation (from t to k), by Definition 2.1.4 we know that $E[M_{k+1} | \mathcal{F}_k] = M_k$ has to hold in order to call SRW a martingale adapted to $\mathcal{F}$. Let's show that:

$$E[M_{k+1} | \mathcal{F}_k] = E[M_{k+1} - M_t + M_t | \mathcal{F}_k] = E[M_{k+1} - M_t | \mathcal{F}_k] + E[M_k | \mathcal{F}_k]$$

$M_{k+1} - M_k$ is just the random variable X_{k+1}, which is obviously independent of the information $\mathcal{F}_k$. In $E[M_k|\mathcal{F}_k]$ we are computing the expected value of M_k, which is just M_k. Then we finally get

$$E[X_{k+1}] + M_K = 0 + M_k = M_k \ .$$

We've now shown, that SRW is a martingale. The variance of a SRW accumulates at rate one per unit time. Assume we're given two non-negative integers k and l where $k < l$, then the variance of the increments over the interval k to l is $l - k$. This is a very important result, so we will look at it next.

Definition 2.1.6 (Quadratic Variation of SRW)**.** The **Quadratic Variation (QV)** of a SRW is computed along one realized path and can be defined as

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

In words it's just the sum over all squared increments of a SRW path, which is

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

The variance is a theoretical concept computed by taking the average over all possible paths. The Quadratic Variation is computed along one realized path only. Also, the Variance is effected by the probability of events differently than the QV. If we change the "fair coin assumption" in our example $p = q$, then the Variance will change. But since the probability (or expected value) is not considered in QV, a change in probabilities will not effect QV. The results for variance and QV in our example are only the same, because the expected value is zero for all increments.

2.2 Scaled Symmetric Random Walk

So far we've looked at the Symmetric Random Walk. Since we want to derive the Brownian Motion, we need to transform the SRW such that it becomes a continuous process. To do that, we would have to do infinitely many sampling within a finite time period. But if we make infinitely many and constant moves in a finite interval, then we will shoot off to infinity. To avoid that, we also need to introduce a scaling factor. For now we stick to the discrete case, speed up time and scale down the step size of a SRW. If we add linear interpolation between points to that, then we will transform the discrete process to a continuous one.

Definition 2.2.1 (Symmetric Random Walk)**.** We can define a **Scaled Symmetric Random Walk (SSRW)** by

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

where nt ideally is an integer and $\frac{1}{\sqrt{n}}$ is the scaling factor. If nt is not an integer, then we need to look at s as the left nearest integer of t and u as the right nearest integer of t respectively. We then approximate nt by linear interpolation like

$$y = f(x_1) + \frac{f(x_2) - f(x_1)}{x_2 - x_1}(x - x_1)$$

which in our case then is

$$nt = ns + \frac{nu - ns}{u - s}(t - s) \ .$$

If we take $W^{(n)}(t)$, take $0 \leq t \leq 4$ and make n large, then the level of "zigzagging" in the graph of the SSRW will increase. We can observe increased "zigzagging" in the graph by increasing n:

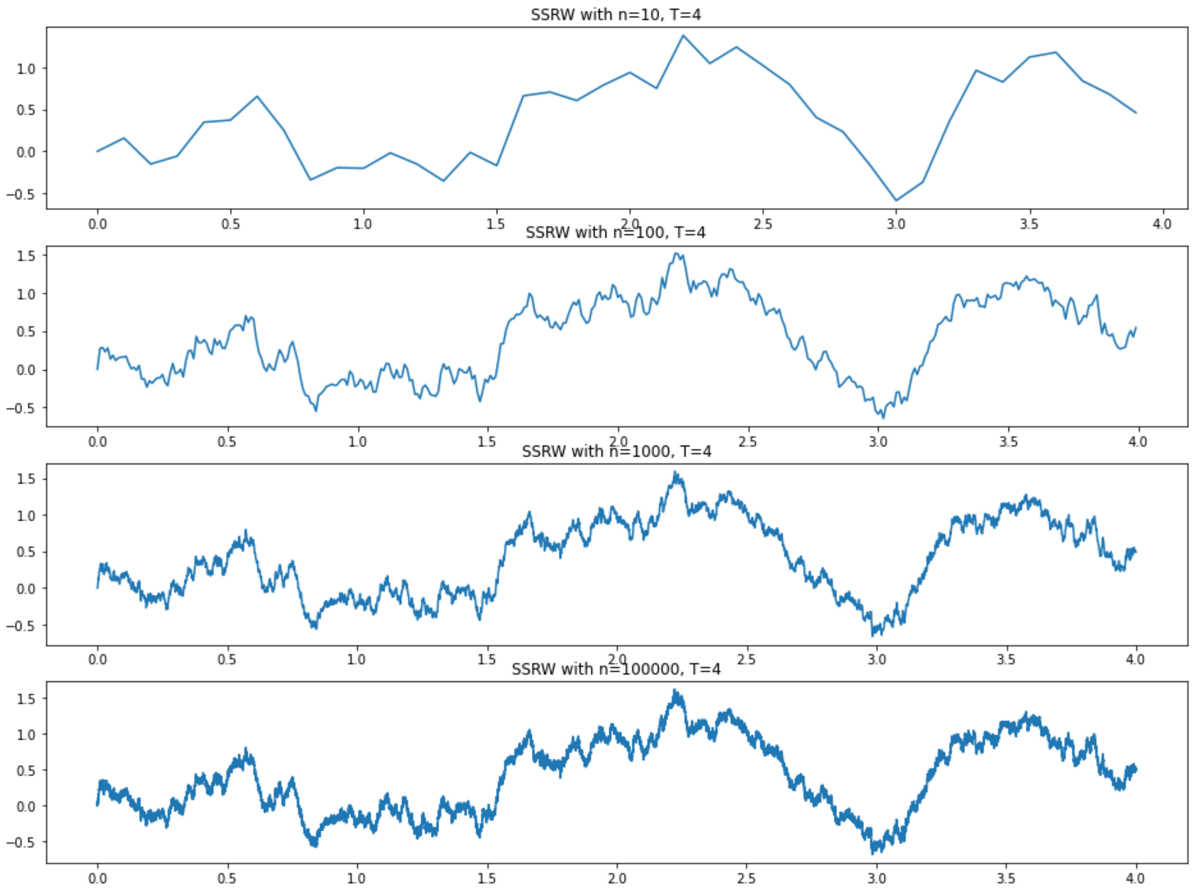

Figure 2.2: SSRW with t=4 and increasing n

In Figure 2.2 we can get an intuition about n in $W^{(n)}(t)$. We will just state here, that SSRW is a martingale and that it accumulates quadratic variation at rate 1 per unit time.
This intuition will be used in the next chapter where we will derive Brownian Motion.

2.3 Defining Brownian Motion

There's a crucial relation between the Scaled Symmetric Random Walk (SSRW) and Brownian Motion (BM). Taking a SSRW is the discrete time process $W^{(n)}(t)$ and as $n \to \infty$, we're able to transform this discrete time process to a continuous time process. The resulting continuous time process is then called **Brownian Motion (BM)**.

Definition 2.3.1 (Brownian Motion)**.** We can define the **Brownian Motion** $W(t)$ on the probability space $(\Omega, \mathcal{F}, P)$ such that for every $\omega \in \Omega$ there is a continuous function $W(t)$, with $t \geq 0$ and $W(0) = 0$, that depends on ω. Then for the Brownian Motion $W(t)$ and all $0 = t_0 \leq t_1 \leq ... \leq t_m$, the increments

$$W(t_1) - W(t_0), W(t_2) - W(t_1), ..., W(t_m) - W(t_{m-1})$$

have to follow a Normal Distribution with

$$E[W(t_{i+1}) - W(t_i)] = 0$$

$$Var[W(t_{i+1}) - W(t_i)] = t_{i+1} - t_i \ .$$

Let's clarify some things in Definition 2.3.1. As mentioned above, Brownian Motion can be derived by taking $\lim_{n\to\infty} W^{(n)}(t)$. This gives us the Brownian Motion, which is defined on the real line and is not linear anywhere almost surely. Recall, that almost surely in this context means that the probability of there being a linear part within the path of a BM, is zero. So in theory there could be a linear part in the BM path, but practically speaking this scenario will not occur.
So far we haven't talked about what a path is. Let's look at the path of a BM informally here, again with our coin toss example. Let's assume that we throw a fair coin infinitely often within a time period and get ω as the outcome of this random experiment. The resulting sequence is then the path of a BM. If we are just given the value of the path denoted by $W(t)$, then there are multiple possibilities for ω to observe $W(t)$.

BM has two important properties, namely that it is a martingale and that it accumulates quadratic variation at rate 1 per unit time. Before we tackle those important concepts in the next two chapters, we will talk about hitting times for BM to give some intuition. For that we use section 1.3 of [5]. Assume we want to access the probability of a BM hitting a before b, where $a > 0$ and $b < 0$. Then we first define $\tau = min\{t \geq 0 : W(t) \in \{a, b\}\}$ as the first time that BM hits a or b. The probability of a SRW with $R_0 = 0$ hitting a before b is given by $p_a = \frac{-b}{-b+a}$, where $\tau = min\{n \geq 0 : R_n \in \{a, b\} | R_0 = 0\}$ and $p_a = P(R_\tau = a)$. If we now do that for the SSRW, we need to pull $\sqrt{n}$ into consideration and get

$$p_a = \frac{-b\sqrt{n}}{-b\sqrt{n} + a\sqrt{n}} = \frac{-b}{-b+a}$$

for the SSRW. This is interesting, because by take $n \to \infty$, the probability of hitting a first doesn't change. Since we also know $W(t) = n \to \infty W_t^n$, we get the same p_a for the SSRW and BM. Let's restate that.

Theorem 2.3.2 (Hitting times for Brownian Motion)**.** Assume we want to access the probability of a BM hitting a before b, where $a > 0$ and $b < 0$. Then we can compute this probability p_a by

$$p_a = \frac{-b}{-b+a} \ .$$

Let's do some examples to put the learned concepts into perspective.

Example 2.3.3 (Example for BM and hitting times)**.** Let's say we look at a pollen under the microscope and keep track of its changes only in the x-direction. We do not consider the y-direction to keep the BM one-dimensional and track it w.r.t. time. The value (in our example the position) of the BM W(t) is known for $t = 3$ and given by $W(3) = 0,78$.

1. Compute $E[W(8)|W(3) = 0,78]$ and $Var[W(8)]$, given that $t = 3$.
2. What is the probability of hitting 3 before -4? (No conditional statements in this example)

Solutions:

1. Since we have independent increments, we get that $E[W(8)|W(3) = 0,78] = 0,78 + 0 = 0,78$. To get $Var[W(8)]$, given that we are at $t = 3$, we compute $Var[W(8)] - Var[W(3)] = 8 - 3 = 5$.
2. By theorem 2.3.2 we get $p_3 = \frac{-(-4)}{-(-4)+3} = \frac{4}{7} = 0,571$.

Example 2.3.3 has given us some intuition about BM and hopefully made the theory a bit more concrete. Let's have a look at the plot of a possible BM described in Example 2.3.3. We also added a, b, τ and $W(\tau)$ in the plot:

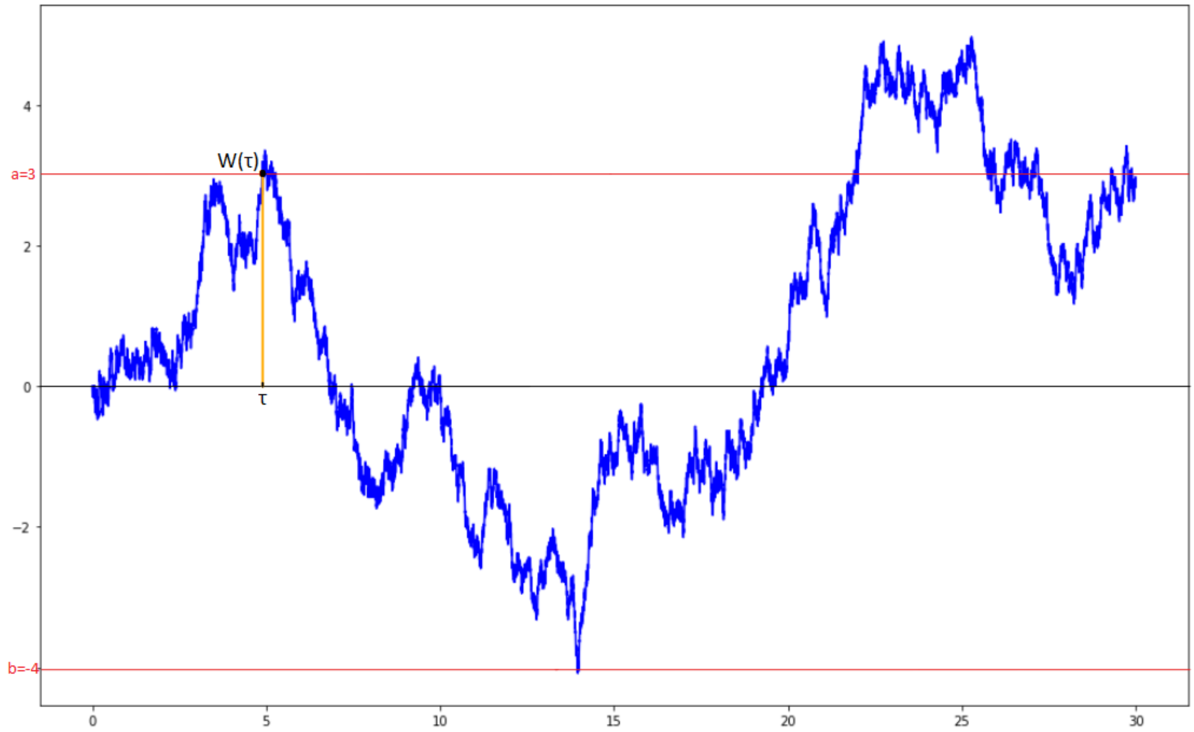

Figure 2.3: BM with hitting times

2.4 BM is a martingale

We'll now continue with the important fact, that **BM is a martingale**. For that we first define what a martingale is in the continuous case.

Definition 2.4.1 (Martingale)**.** Assume we are given a stochastic process $\{X(t)\}_{t \in R}$ that is adapted to the filtration $\mathcal{F}$, where $\mathcal{F}(1) \subseteq ... \subseteq \mathcal{F}(s) \subseteq ... \subseteq \mathcal{F}(t) \subseteq ... \subseteq \mathcal{F}$ for some $0 \leq s \leq t$. Then we call $\{X(t)\}_{t \in R}$ a **martingale** if

$$E[X(t)|\mathcal{F}(s)] = X(s) \ .$$

The definition of a martingale in the continuous case is very similar to the definition of a martingale in the discrete case 2.1.4. $\mathcal{F}(t)$ is the filtration for the BM. We will not go into further detail about $\mathcal{F}(t)$ and just informally call it "information up to t". An important thing to be noted is, that we cannot make predictions about future movements with $\mathcal{F}(t)$. One can read up more about filtrations for BM in section 3.3.3 of the Book [3]. We're now able to show that Brownian Motion is a martingale.

Theorem 2.4.2 (Brownian Motion is a martingale)**.** By Definition 2.4.1 we know that $E[W(s)|\mathcal{F}(t)] = W(t)$ has to hold in order to call BM a martingale adapted to $\mathcal{F}$. Let's show that:

$$E[W(t)|\mathcal{F}(s)] = E[W(t) - W(s) + W(s)|\mathcal{F}(s)] = E[W(t) - W(s)|\mathcal{F}(s)] + E[W(s)|\mathcal{F}(s)]$$

$[W(t) - W(s)]$ is an increment of the BM, that can be arbitrarily small or large and is independent of the information $\mathcal{F}(s)$. $E[W(t) - W(s)]$ is 0 by Definition 2.3.1 of BM. In $E[W(t)|\mathcal{F}(t)]$ we are computing the expected value of $W(t)$, which is just $W(t)$. Then we finally get

$$E[W(t) - W(s)] + W(t) = 0 + W(t) = W(t)$$

which makes BM a martingale.

2.5 Quadratic Variation

The concept of variation is a crucial concept to understand when dealing with Brownian Motion. Generally speaking, variation is just some absolute change of a function within an interval. There's different types of variation and we will look at First-Order and Second-Order or Quadratic Variation (QV).
The difference between continuously differentiable function functions and BM is, that its Quadratic Variation is non-zero, namely it accumulates at rate one per unit time. We will investigate what that means in this chapter and draw the comparison between continuously differentiable functions and BM. Since BM accumulates Quadratic Variation at rate one per unit time, we cannot use ordinary calculus to compute derivatives and integrals of BM. To do that we need to introduce Itô Calculus, which is discussed in the next chapter. Let's have look at First-Order and Quadratic Variation of continuously differentiable functions.

To give some intuition about **First-Order variation**, we can first define it in a non-technical way. Assume we're given a continuously differntiable function $f(t)$, where $t \in [0,T]$ and split the interval $[0,T]$ into n sub-intervals. Then a sub-interval is of length $\Delta t_k = \frac{T}{n}$ and given by $\left[t_{\frac{(k-1)T}{n}}, t_{\frac{kT}{n}}\right]$, where $k = 1,2,...,n$. We can then introduce First-Order Variation as

$$\lim_{n\to\infty} \sum_{k=1}^{n} \left| f\left(t_{\frac{kT}{n}}\right) - f\left(t_{\frac{(k-1)T}{n}}\right) \right| \ .$$

If we take the concept of First-Order Variation and square the summand, then we end up with **Quadratic Variation (QV)**

$$\lim_{n\to\infty} \sum_{k=1}^{n} \left| f\left(t_{\frac{kT}{n}}\right) - f\left(t_{\frac{(k-1)T}{n}}\right) \right|^2 \ .$$

Note that, if we wanted to look at an interval $[s,T]$ with $0 \leq s \leq T$ instead of $[0,T]$, then we would have to replace T by $(T-s)$ in the definitions.
Now, we'll define those concepts more formally to get rid of all those cumbersome variables. For that we use a partition.

Definition 2.5.1 (First-Order Variation of continuously differentiable functions)**.** Take the continuously differentiable function $f(t)$ with $t \in [0,T]$. Then we can define the **First-Order Variation** of f up to T as

$$FV_T(f) = \lim_{\|\pi\|_\infty \to 0} \sum_{j=0}^{n-1} |f(t_{j+1}) - f(t_j)|$$

where $\pi = \{t_0, t_1, ..., t_n\}$ is called a partition. This partition is just a way of dividing up a set into smaller disjoint subsets. Also $\|\pi\|_\infty$ denotes the infinity norm of the partition, which is defined as $\|\pi\|_\infty = \max_{j=0,1,...,(n-1)} (t_{j+1} - t_j)$. In other words, $\|\pi\|_\infty$ is just the largest sub-interval of the partition. This is important, because the sub-intervals of a partition don't necessarily have to be of equal size. So the limit $\lim_{\|\pi\|_\infty \to 0}$ just means taking the number of partitions $n \to \infty$ so that eventually the largest sub-interval of the partition $\|\pi\|_\infty$ tends to zero.

We can also rewrite $FV_T(f)$ with the help of the Mean Value Theorem. By the Mean Value Theorem we know, that for any continuously differentiable function $f(t)$ on $[t_j, t_{j+1}]$, there exists a point $s_j \in [t_j, t_{j+1}]$ such that

$$\frac{f(t_{j+1}) - f(t_j)}{t_{j+1} - t_j} = f'(s_j) \ .$$

Using that we get $f(t_{j+1}) - f(t_j) = f'(s_j)(t_{j+1} - t_j)$ and apply it to FV_T:

$$FV_T(f) = \lim_{\|\pi\|_\infty \to 0} \sum_{j=0}^{n-1} \left| f'(s_j)(t_{j+1} - t_j) \right| = \lim_{\|\pi\|_\infty \to 0} \sum_{j=0}^{n-1} \left| f'(s_j) \right| (t_{j+1} - t_j) \ .$$

Let's keep the result so far in mind. Let $f : \Omega \to R$ be continuous and consider an interval $[0, T] \subset \Omega$. Then we define by

$$\int_0^T f(t)dt := \lim_{n\to\infty} \frac{T}{n} \sum_{k=0}^{n-1} f\left(\frac{T}{n}k\right)$$

the (definite) integral of f over $[0, T]$.

In this definition of an integral, we need to make an assumptions (ass.) so that we can apply it to our result so far. We assume that all disjoint subsets of our interval are the same. This doesn't change the result as we apply the limit. Furthermore, it transforms the length of a disjoint interval $(t_{j+1} - t_j)$ into $\frac{T}{n}$, which can be pulled out of the sum since it doesn't depend on j. If we look at s_j in our result so far, we see that as the partitions get smaller s_j will change. Since we got s_j from the Mean Value Theorem, it depends on t_j and t_{j+1}, which for $\|\pi\|_\infty \to 0$ tend to the same value. Without loss of generality we can then say that since both values t_j and t_{j+1} tend to the same value, we just consider the left value t_j. We can then replace s_j by $\frac{T}{n}k$, where k starts at 0 and ends at $n-1$. We'll now apply the definition of the (definite) integral to the function $|f'(s_j)|$.

Theorem 2.5.2 (First-Order Variation using the integral)**.** Taking the result we got so far and making the mentioned assumption (ass.), we can express the First-Order Variation of a continuously differentiable function f up to T by

$$FV_T(f) = \lim_{\|\pi\|_\infty \to 0} \sum_{j=0}^{n-1} \left| f'(s_j) \right| (t_{j+1} - t_j) \overset{\text{ass.}}{=} \lim_{n\to\infty} \frac{T}{n} \sum_{k=0}^{n-1} \left| f'\left(\frac{T}{n}k\right) \right| = \int_0^T |f'(t)|dt \ .$$

We can conclude, that the First-Order Variation of the function f is the sum of the absolute values of all up and down oscillations of f from 0 to T. This can be written as $\int_0^T |f'(t)|dt$. Let's do an example on that.

Example 2.5.3 (First-Order Variation)**.** Assume we want to compute the First-Order Variation of the function $f(t) = t^3$, where $t \in [0, 4]$, by using partitions. For simplicity reasons we will use 10 equally sized partitions instead of taking the limit.
The difference between using the simpler definition introduced in the beginning of this subsection and using a more involved one like Definition 2.5.1 or Theorem 2.5.2 isn't big for that example. The advantage of the more involved ones is that sometimes it's an advantage to choose partitions of different sizes. Here it doesn't matter.
We use the following values: $T = 4$, $n = 10$.

If we go by Definition 2.5.1 then we choose $\{t_0, t_1, t_2, ..., t_{10}\}$ like $\{0, 0.4, 0.8, 1.2, 1.6, 2, 2.4, 2.8, 3.2, 3.6, 4\}$ and then have to compute $\{f(t_0), f(t_1), f(t_2)..., f(t_10)\} = \{0, \frac{8}{125}, \frac{64}{125}, ..., 64\}$. Then we compute $|f(t_1) - f(t_0)| + |f(t_2) - f(t_1)| + ... + |f(t_10) - f(t_9)| = \frac{8}{125} + \frac{56}{125} + ... + \frac{2168}{125} = 64$. So this is then the absolute change of the function between 0 and 4, partitioned into 10 equally sized pieces.

If we wanted to do the computation without partitioning and the help of an integral like in Theorem 2.5.2, then we get $\int_0^4 |f'(t)|dt = \int_0^4 |3t^2|dt = \left[t^3\right]_0^4 = 64 - 0 = 64$.

The solution to this simple example is kind of obvious, but the procedure stays the same for more complicated functions. This example is visualized in Figure 2.4 below. All 10 vertical red lines visualize the difference between $f(t_j) - f(t_{j-1})$ with $j = 0, 1, ..., 10$.

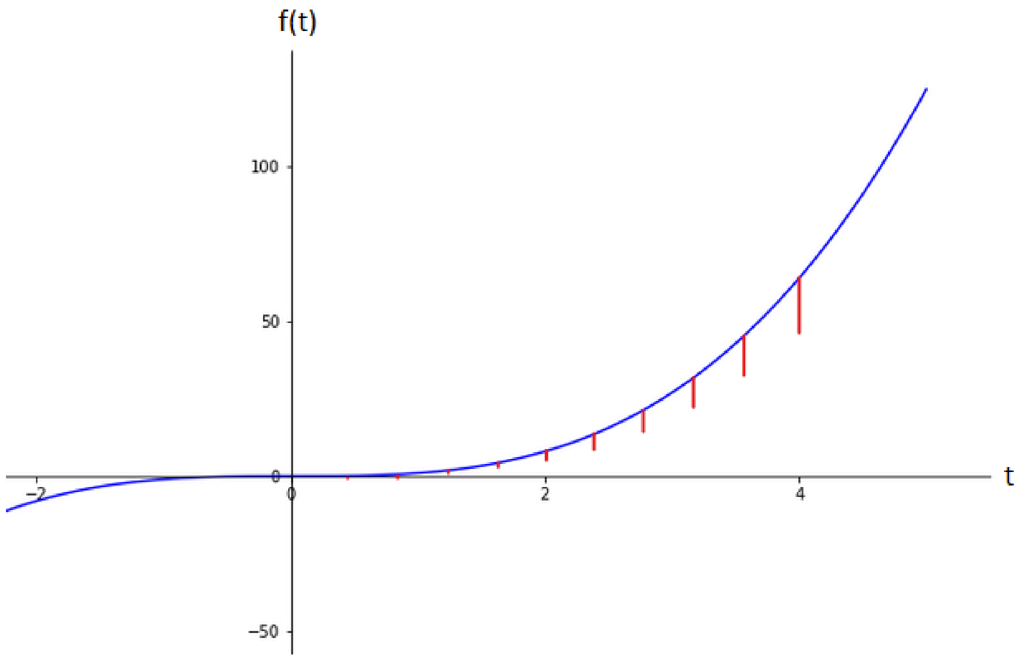

Figure 2.4: Visualizing Variation for $f(t) = t^3$ with $T = 4, n = 10$

If we add up all those 10 red lines, then we will get 64 as a result. There's a difference between First-Order and Quadratic Variation. When looking at Quadratic Variation visually, we would have to square all of those red lines before adding them up. After this example we'll now define Second-Order Variation also known as Quadratic Variation.

Definition 2.5.4 (Quadratic Variation of functions)**.** Take the function $f(t)$ with $t \in [0, T]$. Then the **Quadratic Variation** of f up to time T is defined by

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

where π is again a partition with $\pi = \{t_0, t_1, ..., t_n\}$ and $0 = t_0 < t_1 < ... < t_n = T$.

If we assume that the function is continuously differentiable, then we can use the Mean Value Theorem just like for First-Order Variation and keep in mind that $\|\pi\|_\infty$ is the largest partition, to get

$$\lim_{\|\pi\|_\infty \to 0} \sum_{j=0}^{n-1} [f(t_{j+1}) - f(t_j)]^2 = \lim_{\|\pi\|_\infty \to 0} \sum_{j=0}^{n-1} f'(s_j)^2 (t_{j+1} - t_j)^2$$

$$\leq \lim_{\|\pi\|_\infty \to 0} \sum_{j=0}^{n-1} (s_j)^2 (t_{j+1} - t_j) \|\pi\|_\infty = \lim_{\|\pi\|_\infty \to 0} \|\pi\|_\infty \lim_{\|\pi\|_\infty \to 0} (s_j)^2 (t_{j+1} - t_j) \ .$$

We can now use the same trick as in Theorem 2.5.2, to get

$$\lim_{\|\pi\|_\infty \to 0} \|\pi\|_\infty \lim_{\|\pi\|_\infty \to 0} (s_j)^2 (t_{j+1} - t_j) = \lim_{\|\pi\|_\infty \to 0} \|\pi\|_\infty \int_0^T |f'(t)|^2 dt$$

Since the integral of a continuously differentiable function is finite, we get

$$\lim_{\|\pi\|_\infty \to 0} \|\pi\|_\infty \int_0^T |f'(t)|^2 dt = 0 \int_0^T |f'(t)|^2 dt = 0 \ .$$

We can conclude, that the Quadratic Variation of a continuously differentiable function is zero. This is the reason, why we usually don't consider Quadratic Variation in ordinary calculus. To show, that Quadratic Variation of continuously differentiable functions is zero, we've used the Mean Value Theorem. This theorem can unfortunately not be applied to the path of a BM, since it is not differentiable anywhere. We still want to be able to compute Quadratic Variation of a BM path and therefore need to proof the following.

Theorem 2.5.5 (Quadratic Variation of BM)**.** Take the Brownian Motion W. Then we get the Quadratic Variation of W up to T by

$$[W, W](T) = T$$

almost surely. Almost surely here just means, that in theory there could be a BM path where $[W, W](T) = T$ doesn't hold. But this event occurs with zero probability.

In the Book [3] one can read up a mathematical rigorous way of proofing Theorem 2.5.5. Here we will just derive Quadratic Variation from the SSRW, which is a simpler way of putting things and sort of an informal proof.

Theorem 2.5.6 (Quadratic Variation of a BM path)**.** Recall from Definition 2.1.6, that $[M, M]_k = \sum_{j=1}^{k} (M_j - M_{j-1})^2 = \sum_{j=1}^{k} (X_j)^2 = k$. We also know from Definition 2.2.1, that $W^{(n)}(t) = \frac{1}{\sqrt{n}} M_{nt}$ and from Definition 2.3.1, that $W(t) = \lim_{n\to\infty} W^{(n)}(t)$. If we now use all of that, then we can derive the Quadratic Variation of a BM path like

$$[W, W](T) = \lim_{n\to\infty} \left[W^{(n)}, W^{(n)} \right](T) = \lim_{n\to\infty} \sum_{j=1}^{nT} \left[\frac{1}{\sqrt{n}} X_j \right]^2 = \lim_{n\to\infty} \sum_{j=1}^{nT} \frac{1}{n} = \lim_{n\to\infty} T = T \ .$$

We've now shown, that Brownian Motion (BM) accumulates Quadratic Variation (QV) at rate 1 per unit time. Assume we want to compute the QV of a BM path in the interval $[T_1, T_2]$, which we partition such that $T_1 = t_0 < t_1 < ... < t_m = T_2$. If we decide to go with finitely many partition, then the QV will be dependent on the BM path along which we compute it. In order to make the results for QV (of a BM) path independent, we have to create infinitely many partitions of the time interval. Only then we can symbolize Quadratic Variation in the following form

$$dW(t)dW(t) = dt \Leftrightarrow (dW(t))^2 = dt$$

where dt denotes an infinitesimally smallest change in time and $(dW(t))^2$ denotes the squared infinitesimally small change of the BM path (also known as Quadratic Variation). The change has to be infinitesimally small in order to get exact results, otherwise the result would be path dependent. The infinitesimally small change in time dt is denoted by $t_n - t_{n-1}$, where $t_{n-1} < t_n$. Remember the fact $(dW(t))^2 = dt$ since we will need this result especially for the introduction and subsection 3.3 of the next chapter.

3 Stochastic Calculus

Understanding Stochastic Calculus is crucial for someone interested in Financial Mathematics. This chapter lays the foundation for the Black-Scholes-Merton Model of Investing and should put some of the concepts from the previous chapter into context. We will see why Quadratic Variation of BM is important and how to compute differntials and integrals of BM. Those integrals are called Itô integral and used to model the value of a portfolio changing over continuous time.

3.1 Introduction

We've already mentioned, that BM is not differentiable anywhere although its paths are continuous. One can wrap their head around that by visualizing a BM and then zooming into the graph. If we zoom into the graph of a BM path, then we see that the "zigzagging" just continues and does not go away no matter how far we zoom in. We can also look at that visually:

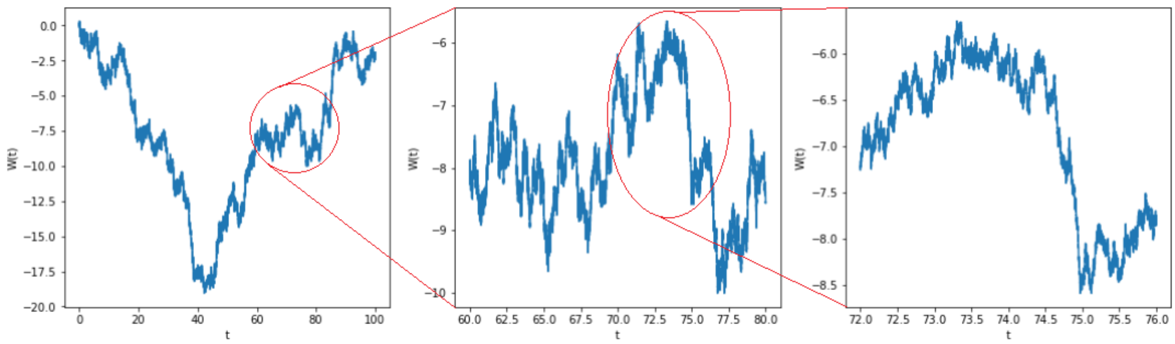

Figure 3.5: "Zigzagging" of a BM after zooming in

Before continuing, we want to show more formally that we cannot use Standard Calculus to compute the derivative of a BM.

Assume we're given a differentiable function $f(x)$ and take the BM $W(t)$ as an input to that function. We want to use Standard Calculus and so we assume $W(t)$ to be differentiable for now, even tough that is not the case. Then we would be able to use the chain rule to compute the derivative of $f(W(t))$ by

$$\frac{df(W(t))}{dt} = f'(W(t))W'(t)$$

written in a slightly different way as

$$df(W(t)) = f'(W(t))W'(t)dt = f'(W(t))dW(t) \ .$$

However, the result obtained by Standard Calculus is **wrong** for $W(t)$, because BM accumulates Quadratic Variation at rate 1 per unit time. In the formula for the chain rule above we don't see any second order terms. This makes sense in Standard Calculus, since we've derived it from Taylor's expansion and don't have to consider 2^{nd} or higher order terms. A more rigorous justification will be given in subsection 3.3, when we talk about the Itô-Doeblin formula.

For now we elaborate on that a bit more and have a closer look at Taylor's theorem. Assume we take the differentiable function $f(t)$ for some t and compute with Taylor's theorem the following for a small s:

$$f(t+s) = f(t) + f'(t)s + \frac{f''(t)}{2}s^2 + \frac{f'''(t)}{3}s^3 + \dots \ .$$

If we assume s to be a small, then we can compute a small change of f by

$$f(t+s) - f(t) = f'(t)s + \frac{f''(t)}{2}s^2 + \frac{f'''(t)}{3}s^3 + ... \ .$$

$$\Leftrightarrow \frac{f(t+s) - f(t)}{s} = f'(t) + \frac{f''(t)}{2}s + \frac{f'''(t)}{3}s^2 + ...$$

This looks very similar to the derivative of a differentiable function, which is defined by

$$\lim_{s \to 0} \frac{f(t+s) - f(t)}{s} = f'(t) \ .$$

We notice, that in Standard Calculus 2^{nd} and larger ordered terms are not considered when computing the derivative of a differentiable function. We can understand why this is the case by looking at Definition 2.5.4 and the associated computations below it. There we've shown, that Quadratic Variation is 0 when taking infinitely small time differences of a function f. However, if we replace t by the BM $W(t)$, then we won't be able to ignore the 2^{nd} order term because of Quadratic Variation (QV). We've shown why this is the case in Definition 2.5.5. This is the motivation for subsection 3.3. There we will put all of that into context by introducing the Itô-Doeblin formula in different formulations.

3.2 Itô Integral

In this subsection we want to explore the idea of Itô Integrals. More precisely we want to give meaning to

$$\int_0^T \Delta(t) dW(t) \ .$$

which is the Itô Integral up to T. In this integral $W(t)$ is a BM and $\Delta(t)$ is a simple process, which we will define now.

Definition 3.2.1 (Simple Process)**.** Assume we are given the interval $[0, T]$ and create a partition of it such that $\pi = t_0, t_1, ..., t_n$, where $0 = t_0 \leq t_1 \leq ... \leq t_n = T$. If the process $\Delta(t)$ is constant for each sub-interval $[t_j, t_{j+1}]$ for $j \in \{0, ..., n-1\}$, then we call $\Delta(t)$ a simple process.

The value of a simple process $\Delta(t)$ only depends on the information up to t. One can look at a visualization of a simple process path in Figure 3.6. When we use a simple process in our calculations, then it is depending on a BM path. If something depends on a random process, then it is random itself. Therefore $\Delta(t)$ is random.

To give further intuition about Itô Integrals and simple processes, we want to give an **example** from the investing world. Assume the BM $W(t)$ is the price per share of an asset (e.g. the share of a company) we hold at time t. For now, just ignore the fact that $W(t)$ can be negative and is a bad choice for modeling the price of an asset. Assume further, that $t_0, t_1, ..., t_{n-1}$ are trading dates and that we can only change our amount of stocks once a new date has occurred. So we can only sell or buy new stocks once the day is over and a new one begins. We change our position in the asset by either buying or selling shares. We denote by $\Delta(t_0), \Delta(t_1), ..., \Delta(t_{n-1})$ the position we take in the asset for the trading day, where $\Delta(t_0) = 0$. Then we can denote the gain from trading the asset at time t, where currently $t \in [t_2, t_3]$ holds, by

$$I(t) = \Delta(0)W(t_1) + \Delta(t_1)[W(t_2) - W(t_1)] + \Delta(t_2)[W(t) - W(t_2)] \ .$$

Looking at our gain $I(t)$ so far, where t is currently in $[t_2, t_3]$, we notice a pattern. We always multiply our position at the beginning of the day with the price change of the asset at that day. The price change over one date is denoted by $[W(t_j) - W(t_{j-1})]$, where t_j is the j^{th} date. We also have to sum up all gains from previous dates in order to get the total gain. In our example, the positions we take at the certain trading dates is $\Delta(t)$ and just a simple process. A simple process path is visualized in the following figure.

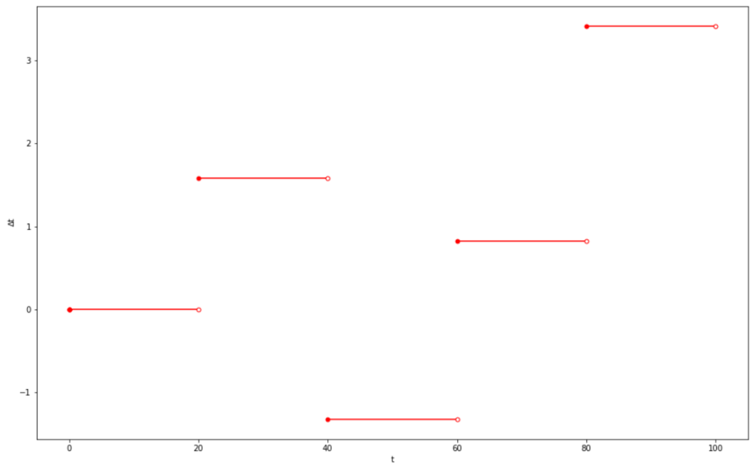

Figure 3.6: Example path of the simple process $\Delta(t)$

Definition 3.2.2 (Itô integral)**.** Assume $W(t)$ is a Brownian Motion and $\Delta(t)$ is a simple process, then we define the Itô integral for some $t \in [t_k, t_{k+1}]$ by

$$I(t) = \Delta(t_k)[W(t) - W(t_k)] + \sum_{j=0}^{k-1} \Delta(t_j)[W(t_{j+1}) - W(t_j)]$$

which can also be written as

$$I(t) = \int_0^t \Delta(u) dW(u) \ .$$

The sum definition for $I(t)$ just means adding the change in the BM path from the beginning of the time interval until now to all prior changes, which gives us the total sum up to t. The sum definition seems a bit more intuitive and we've already worked with it a bit in the example from above, where the total sum was our gain up to t.

We can also use the simple process Δt to approximate a BM path, which is visualized in the next figure.

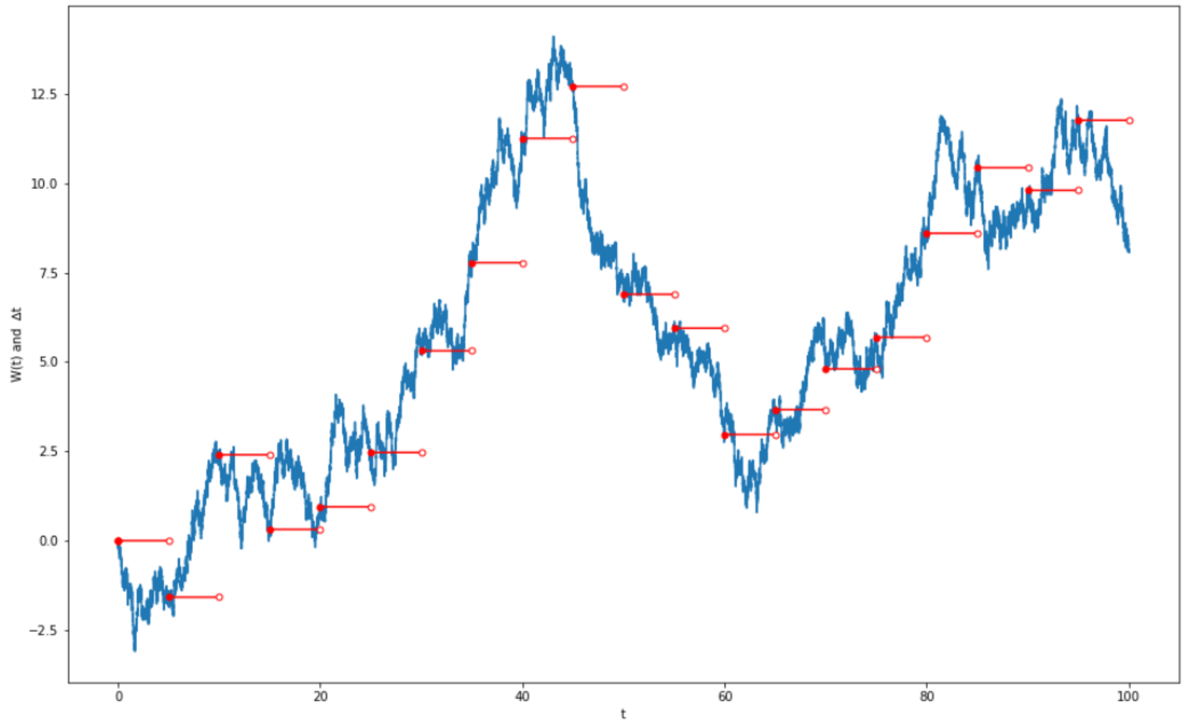

Figure 3.7: Approximating a BM path with the simple process Δt

The Itô Integral has some interesting properties, namely that the Itô integral is a martingale, has Itô isometry and its Quadratic Variation accumulates in an interesting way. One can read up on those properties in the Book [3], where also proofs are provided. We will only state that Itô integral is a martingale here.

Theorem 3.2.3 (Itô integral is a martingale)**.** The Itô integral as defined in 3.2.2 with the associated filtration $\mathcal{F}$ is a martingale and therefore satisfies

$$E[I(t)|\mathcal{F}(t)] = I(s) \ .$$

This seems somewhat intuitive. The Itô integral is the gain from trading an asset, that behaves like a BM, which is a martingale. So the Itô integral has no tendency to rise or fall, because it is depending on a martingale and is therefore also a martingale.

3.3 Itô-Doeblin formula

In the end of this subsection 3.1 we saw that we cannot use Standard Calculus for BM. But with the help of the Itô-Doeblin formula in differential and integral form, we are able to provide an alternative. We will fist look at the Itô-Doeblin formula in differential form.

Assume we're given a BM $W(t)$ with $t \in [0, T]$ and a differentiable function f, where we want to look at $f(W(t+s)) - f(W(t))$ for a small $s > 0$. Then we call the following **Itô-Doeblin formula in differential form**:

$$f(W(t+s)) - f(W(t)) =$$

$$= f'(W(t))(W(t+s) - W(t)) + \frac{f''(W(t))}{2}(W(t+s) - W(t))^2$$

$$= f'(W(t))dW(t) + \frac{f''(W(t))}{2}(dW(t))^2$$

$$\overset{\text{QV}}{=} f'(W(t))dW(t) + \frac{f''(W(t))}{2}dt$$

Note that the last part of the equation follows from $(dW(t))^2 = dt$ and denotes the Itô-Doeblin formula in differential form. Generally speaking there is no mathematically precise definition for $df(W(t))$, $dW(t)$ and dt. The d in those terms stands for a little change in t denoted by dt, that makes the formula only precise when the little change is infinitesimally small. We have talked about this a bit at the very end of chapter 2. Now we will state the Itô-Doeblin formula in its integral form, which is mathematically meaningful.

Theorem 3.3.1 (Itô-Doeblin formula in integral form)**.** Assume we're given a Brownian Motion $W(t)$ with $t \in [0, T]$ and a differentiable function f. Then we call the following **Itô-Doeblin formula in integral form**

$$f(W(t)) - f(W(0)) = \int_0^t f'(W(u))dW(u) + \frac{1}{2}\int_0^t f''(W(u))du \ .$$

Looking at the formula above, we can see that $\int_0^t f'(W(u))dW(u)$ denotes the Itô integral and $\int_0^t f''(W(u))du$ an ordinary integral.

Both Itô-Doeblin formula in differential and in integral form represent the same idea, where the differential form is easier to understand and the integral form is mathematically more meaningful. We can now state the Itô-Doeblin formula in multidimensional form, but won't do that in the general case. We will only do it for the BM.

Theorem 3.3.2 (Itô-Doeblin formula for BM)**.** Take a continuously differentiable function $f(t, x)$ and a BM $W(t)$. Then we compute $f(T, W(T))$ for every $T \geq 0$ like

$$f(T, W(T)) = f(0, W(0)) + \int_0^T f_t(t, W(t))dt + \int_0^T f_x(t, W(t))dW(t) + \frac{1}{2}\int_0^T f_{xx}(t, W(t))dt$$

where f_t is the partial derivative of f w.r.t t, f_x is the partial derivative of f w.r.t x and $f_x x$ is the partial derivative of f_x w.r.t x. We call the result above **Itô-Doeblin formula for BM**. This is more or less the same formula as Theorem3.3.1 only in higher dimensions. A proof of this formula can again be found in the Book [3].

In this chapter we've seen ways to derive a new way of taking partial derivatives and integrals of stochastic processes, especially BM. So far we haven't looked at an example, but this will change in chapter 5 since we will have a look at the Black-Scholes-Merton model of investing. In this model we use the Itô-Doeblin formula for BM to price options fairly.

4 Stochastic Differential Equations

When we talk about a **Stochastic Differential Equation (SDE)**, then we mean some differential equation where at least one term is a stochastic process. The solution to the SDE is then itself a stochastic process.

4.1 Definition

Let's now formally define an SDE. For that we stick to chapter 6 in the Book [3].

Definition 4.1.1 (Stochastic Differential Equation)**.** When we talk about a **Stochastic Differential Equation (SDE)**, then we mean an equation that satisfies

$$dX(u) = \beta(u, X(u))du + \gamma(u, X(u))dW(u)$$

where $\beta(u, x)$ is called **drift** and $\gamma(u, x)$ is called **diffusion**.

Before we show a way of solving this SDE, we have to tackle Geometric Brownian Motion.

4.2 Geometric Brownian Motion

Informally speaking, Geometric Brownian Motion (GBM) has an underlying trend that follows an exponential growth. So before we define GBM, we will have a look at population growth.

Definition 4.2.1 (Discrete population growth)**.** We define the **discrete population growth** up to t with a constant growth rate r and starting value X_0 like

$$X_t = X_0(1 + r)^t \ .$$

We can also define the discrete population growth recursively like

$$X_{t+1} = X_t(1 + r) \ .$$

To obtain X_{t+1}, we calculated $X_t(1 + r)$. Assume now, we want to increase the frequency a bit, but don't want to change the value for X_{t+1} for a given X_t and r. We can transform the discrete case to obtain $X_t = X_0(1 + \frac{r}{f})^{ft}$, with $f > 0$. By introducing a frequency term f, we are now able to define continuous population growth.

Definition 4.2.2 (Continuous population growth)**.** To obtain **continuous population growth**, we introduce a frequency term f to the discrete case and take $\lim_{f\to\infty}$ of the discrete population growth with frequency term f to get

$$X(t) = \lim_{f\to\infty} X(0)(1 + \frac{r}{f})^{ft} = X(0)e^{rt} \ .$$

We're now able to define Geometric Brownian Motion. For the remainder of this chapter, we stick to Karl Sigman's lecture notes on Geometric Brownian Motion [4].

Definition 4.2.3 (Geometric Brownian Motion)**.** We can define the **Geometric Brownian Motion (GBM)**, which we call $S(t)$ here, like

$$S(t) = S(0)e^{X(t)}$$

where $X(t) = \sigma W(t) + \mu t$ is a BM with constant σ and μ.

If we look at the path distribution of the GBM $S(t)$, then we can show that those paths need to be log-normally distributed. In other words, $ln(S(t))$ is normal with mean $\mu t + ln(S(0))$ and variance $\sigma^2 t$. We will not go into further detail on this.

In literature regarding financial mathematics, GBM is sometimes also defined differently. The reason for this is, that we can replace μ by the risk-neutral drift μ^* to get the risk-neutral version of $S(t)$. We will not use the following Definition of the Risk-neutral version of GBM. Altough it is not relevant for us, one could encounter this Definition in other literature and so we mention it.

Definition 4.2.4 (Risk-neutral version of GBM)**.** If we look at Definition 4.2.3, then we can define the **risk-neutral version of GBM** like

$$S(t) = S(0)e^{\sigma W(t) + \mu^* t}$$

where $\mu^* = r - \frac{\sigma^2}{2}$ for some fixed r and σ.

The difference between Definition 4.2.3 and 4.2.4 is that we incorporate the **correction term** $-\frac{\sigma^2}{2}$ in the definition of GBM. The correction term is part of the risk-neutral drift μ^* and just the difference between mean and median of the path-distributions of GBM, which are log-normal as already mentioned. From now on we'll mainly use the risk-neutral version of GBM and see the relevance for the next chapter.

4.3 Solution to Black-Scholes-Merton stochastic differential equation

We've talked about the Black-Scholes-Merton SDE and now we want to show what it means to "solve" it. All the steps in this subsection can again be verified from the book [3]. For simplicity reasons we stick to the Wikipedia-article on Itô's Lemma, which expresses the result elegantly and in an understandable way.

Assume we're given the price of a stock S. Then S follows a GBM if it satisfies the SDE $dS(t) = \sigma S(t)dW(t) + \mu S(t)dt$ with constant volatility σ and constant drift μ for the BM $W(t)$. We now use the Itô-Doeblin formula in differential form for $f(S(t)) = log(S(t))$ and compute

$$\begin{aligned} df &= f'(S(t))dS(t) + \frac{f''(S(t))}{2}dt \\ &= \frac{1}{S(t)}dS(t) + \frac{-S^{-2}(t)S^2(t)\sigma^2}{2}dt \\ &= \frac{1}{S(t)}(\sigma S(t)dW(t) + \mu S(t)dt) + \frac{-\sigma^2}{2}dt \\ &= \sigma dW(t) + \mu dt - \frac{-\sigma^2}{2}dt \ . \end{aligned}$$

In the lines above we've already used our assumption about $dS(t)$ and obtained a result that already has some similarities to GBM, but is missing the exponential.
We now make use of us choosing $f(S(t)) = log(S(t))$ and get

$$log(S(t)) = log(S(0)) + \sigma W(t) + \mu t - \frac{\sigma^2}{2}t \ .$$

If we now simply get rid of the log, then we end up with GBM like

$$S(t) = S(0)e^{\sigma W(t) + \mu t - \frac{\sigma^2}{2}t} \ .$$

This shows, that the GBM is a solution to the Black-Scholes-Merton SDE.

5 Black-Scholes-Merton Model of Investing

In the financial sector, there are 3 ways to make investment decisions to access the benefits and risks of a particular financial investment. The first analytical method is called fundamental analysis, where we predict the fair share price based on the intrinsic value of a company. The second analytical method is called **technical analysis** and is in contrast to the first method. It makes investment decisions based solely on historical data and patterns in the stock price. Thus, all the information is contained in the stock and not in the company. In technical analysis, the stock price reflects the value of the company. We are interested in the third method, called quantitative analysis, which assumes that stock prices are random. Under this assumption, we would need to incorporate randomness into our models and use Stochastic Calculations, stochastic differential equations, and other mathematical-statistical methods to model the behavior of financial assets.
In this chapter, we will take a closer look at the **Black-Scholes-Merton model of investing**, which falls into the realm of quantitative finance. This model was originally developed by Fischer Black and Myron Samuel Scholes. Only later did Robert C. Merton join their discussion and introduce Itô processes and stochastic differential equations into the model. This is the reason why we are interested in the Black-Scholes-Merton model for investing.

5.1 Introduction

So far we haven't talked about call and put options or about any topic concerning finance. Most financial terms used in this work will be explained in this subsection. An example will be given, too. If the terminology is unclear at any point, one can go to website of Investopedia [2] to clear things up.

When we talk about an **asset**, then we mean something that has economic value and an owner. There's a wide range of assets such as houses, cars or shares of a company. A **share** of a company is just one unit of equity ownership of a company. For publicly-traded companies one can calculate the market value of the company by multiplying the number of stocks with the price per stock. An **option** is a financial instrument that involves a buyer and a seller and is concerned with either selling or buying an asset.
A **call option** gives the option buyer the right, but not the obligation, to buy an asset to a fixed price somewhere in the future. The asset involved in the call option is called **underlying asset**. The option buyer is not obligated to buy the underlying asset within the defined time. This is convenient for the option buyer, because he or she could speculate on an investment without being the owner of the asset. Since this is too good to be true, any kind of option involves a premium. The **premium** is payed to the option seller in case the option buyer doesn't make use of his or her right to buy the asset at the predetermined price. The premium is depending on the conditions specified in the option contract.
A **put option** gives the option buyer the right, but not the obligation, to sell an asset to a fixed price somewhere in the future. The fixed price at the time of expiration in an option contract is called **strike price**.

If you hear this for the first time, then this is quite some terminology. Here is an example to improve your understanding.

Example 5.1.1 (Call option example)**.** Assume Anna owns a shop and wants to sell one shares of her small company. Ben has been interested in investing into the shop for a while now and so they meet to make a deal. The value of the shop has been steadily increasing. Ben really wants to make the deal before other investors join, but wants to wait for the company's next annual accounts to be absolutely sure about the investment. Anna and Ben decide to go with an European option.
In this example, the underlying asset is the price of one share of Anna's company. Assume they settle

the deal with a European call option contract where Ben is the buyer and Anna is the seller. Since it's mid-September, the contract expiration date is set to 3 months. They calculate the value of the call options underlying asset (here one share) and concludes that a fair price per share is somewhere around 1000€. Ben wants to speculate a bit on the investment, so he decided to fix the price at 1050€. This price at expiration is called strike price. Since this is an unlikely event to occur, Ben would be less likely to make use of the call option because the value of the underlying share would have to rise above the strike price of 1050€ to justify the purchase (premium not included). Due to the unlikelihood of the underlying asset rising above 1050€, the premium in this deal is very small.

Let's restate that more formally. We are interested in trading an option. For that we look at the underlying asset price S at time t, which we denote by $S(t)$. The option price is then calculated based on the underlying asset price S and time t, which is written as $V(S,t)$. The price of a European call option is denoted by $C(S,t)$ and the price of a European put option is $P(S,t)$. Recall, that European options can only be exercised at the expiration of the contract. The time until expiration is written as $\tau = T - t$, where T is the expiration time and current time is t. Lastly, K denotes the strike price.

5.2 Deriving the Black-Scholes-Merton partial differential equation

For the Black-Scholes-Merton PDE to deliver useful results, we have to make some assumptions about the market. In fact,we assume something called **ideal market conditions**.

Definition 5.2.1 (Ideal markte conditions)**.** When we speak of **ideal market conditions**, then we mean a market that fulfills all of the following assumption:

1) Short term interest rates are constant.
2) Stocks pay no dividends.
3) There are no transaction costs.
4) We can borrow a fraction of a stock.
5) Short selling is allowed.

Sometimes the ideal market conditions are also assumed differently, but we will stick to our definition. Explaining all those assumptions would take quite some time, so one can read up unclear words either in Investopedia [2] or in the book [3].

To derive the Black-Scholes-Merton partial differential equation we assume that the asset price $S(t)$ follows a log-normal distribution and that the option price is given by $V(S,t)$.
The main idea of the Black-Scholes-Merton PDE is to long the option (buy it) and short some portion of the asset (sell it). We can then express our position as

$$\pi = V - \Delta S$$

where Δ is the hedge factor. Our goal is to keep the position π constant. So if the asset price S changes a bit, then we want to use the hedge factor Δ to adjust for those changes to keep our position the same. If we now take the derivative of π, then we get

$$d\pi = dV - \Delta dS$$

where we can use the Itô-Doeblin formula in differential form to get

$$dV = \frac{\partial V}{\partial S} dS + \frac{\partial V}{\partial t} dt + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} dt$$

and

$$d\pi = \frac{\partial V}{\partial S}dS + \frac{\partial V}{\partial t}dt + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2}dt - \Delta dS$$

$$= \left(\frac{\partial V}{\partial S} - \Delta\right)dS + \left(\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2}\right)dt \ .$$

At this point we can choose Δ in a way, that eliminates the stochastic part of $d\pi$. We do that by choosing $\Delta = \frac{\partial V}{\partial S}$, which gives us

$$d\pi = \left(\frac{\partial V}{\partial S} - \frac{\partial V}{\partial S}\right)dS + \left(\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2}\right)dt$$

$$= \left(\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2}\right)dt \ .$$

By making the choice for Δ, we've eliminated the stochastic part in the equation and are now left with the deterministic part.
In section 6.3 of Paul Wilmott's Book [6] explains, that we call a reduction in randomness **hedging** and the prefect elimination of risk (by exploiting correlation between two instruments) **delta hedging**. The prefect hedge has to be adjusted continually. This process is called **dynamic hedging**. Delta hedging is a form of dynamic hedging.
We go back to our example and see that $d\pi$ is now a deterministic function, where we can use the **no-arbitrage principle**. This principle just says, that if investors exploit an arbitrage opportunity in the market, then the market price will move so that arbitrage is no longer possible. To state this more formally, we get

$$d\pi = r\pi dt = r(V - \delta S)dt$$

where we still have $\Delta = \frac{\partial V}{\partial S}$. Now we have expressed $d\pi$ in two different ways and set them equal to each other to get

$$\left(\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2}\right)dt = r(V - \delta S)dt$$

$$\Leftrightarrow \frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} = rV - r\frac{\partial V}{\partial S}S \ .$$

We can now state the Black-Scholes-Merton PDE by doing some simple algebraic transformations.

Definition 5.2.2 (Black-Scholes-Merton PDE)**.** The Black-Scholes-Merton PDE is given by

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} = rV$$

where S is the asset price, V is the option price, r is the risk-free return of an alternative investment and σ is the volatility of the asset.

5.3 Black-Scholes-Merton model for European options

So far we've looked at a lot of theory. In this subsection we want to look at how to use the Black-Scholes-Merton PDE 5.2.2 for European put and call options. We will stick to the notation introduced in the beginning of this chapter.

For that we first need to look at the payoff of a European option. The cash flow of a European call option can be expressed as $max\{S - K; 0\}$, where S is the price of the underlying asset and K is the strike price. The cash flow of a European put option is $max\{K - S; 0\}$. We can now define the price formula for European call and put options according to Black, Scholes and Merton.

Theorem 5.3.1 (Black-Scholes-Merton formula for European call options)**.** We can calculate the fair price for a European call option with the help of the Black-Scholes-Merton PDE. The result is then called **Black-Scholes-Merton formula for European call options** and can be expressed as

$$C(S,\tau) = SN(d_1) - Ke^{-r\tau}N(d_2)$$

where N resembles the standard normal cumulative distribution function and d_1 and d_2 are defined like

$$d_1 = \frac{ln\left(\frac{S}{K}\right) + (r + \frac{\sigma^2}{2})\tau}{\sigma\sqrt{\tau}}$$

$$d_2 = d_1 - \sigma\sqrt{\tau}\ .$$

Theorem 5.3.2 (Black-Scholes-Merton formula for European put options)**.** We can also calculate the fair price for a European put option with the help of the Black-Scholes-Merton PDE. The result is then called **Black-Scholes-Merton formula for European put options** and can be expressed as

$$P(S,\tau) = Ke^{-r\tau}N(-d_2) - SN(-d_1)$$

where N resembles the standard normal cumulative distribution function and d_1 and d_2 are the same as in Theorem 5.3.1.

Note, that the notation for the Black-Scholes-Merton formulas like expressed in 5.3.1 and 5.3.2 vary from literature to literature. Sometimes the price of the underlying S is also written as S_t or $S(t)$. In our case it's clear that we've fixed t and therefore just write S. One could also encounter notation that uses T and t instead of τ. Recall, that $\tau = T - t$. Furthermore, some people use Φ instead of N for the standard normal distribution. But this is usually specified.

If we're able to get good estimates of the variables that go into the BSM model of investing, then we're able to price European options fairly under ideal market conditions. In practice we're able to access almost all variables, that go into the BSM model for European options, reasonably well. For example, we know the price of the underlying asses S (e.g. a stock). The strike price K and expiration time T are part of the contract. We can also look up for how much a call options C or a put option P are traded in the market. The risk-free interest rate r can also be looked up. However, we're missing one more variable namely the volatility σ. Luckily for us, the volatility σ can be estimated and we will see how to do that in the next chapter.

5.4 Implied Volatility

We've looked at the Black-Scholes-Merton formula and saw how to price an option fairly under the assumption that σ is be constant. This assumption is convenient, because it allows us to make computations with the BSM method easily. However, when we assume σ to be constant, then the value for σ is depending on the time horizon we look at. For example, we could calculate σ historically based on the data of the last two weeks or of the last 5 years. The computed values for σ would differ heavily and we wouldn't really know which value for σ would yield the best result or generalize best. This is the reason why we assume σ to be non-linear in practice. Most financial models that make use of Black-Scholes-Merton's result want to estimate σ accurately. This is called estimating **implied volatility** of an option.

If we put our estimation of σ into the BSM formula for European put or call options, then we want our result to be equal to the options market price. The options market price is known and we will call it C_0.

As already stated, we also know all the parameters that go into the BSM formulas for European options except for σ. So when calculating implied volatility of an option, we want to solve

$$f(\sigma) = C(\sigma) - C_0 = 0 \ .$$

One shouldn't be confused with the shift in notation regarding $C(\sigma)$. As we've already mentioned in the end of the last chapter, the values for S, T, t, K and r are either part of the contract or are accessible to us. For that reason C is only depending on the variable σ, that we want to estimate well. All the other components are assumed to be constant. Note, that we've added σ as an input to the BSM formula for European call options. In the original Definition 5.3.1, we've assumed σ to be constant. Since we've dropped that assumption in this chapter, we've now added it as an input to the model and want to solve $f(\sigma)$ w.r.t. σ.

Let's have a look at $f(\sigma)$ visually and look at how the components change the function $f(\sigma)$. For that we assume the values $S = 194, K = 210, C_0 = 1,5, \tau = \frac{38}{365}, r = 0,01$. One can observe changes in those values in Figure 5.8. We've left out r, since there isn't a noteworthy difference for small changes in this value. We would have to change r from 0.01 to 0.1 in order to see a noteworthy difference, but this is much bigger than we assume the risk-free return to be. Also note, that the red line resembles our assumed values in all 4 plots. The assumed values are taken from a practical example and will be tackled in Example 5.4.3 in more detail.

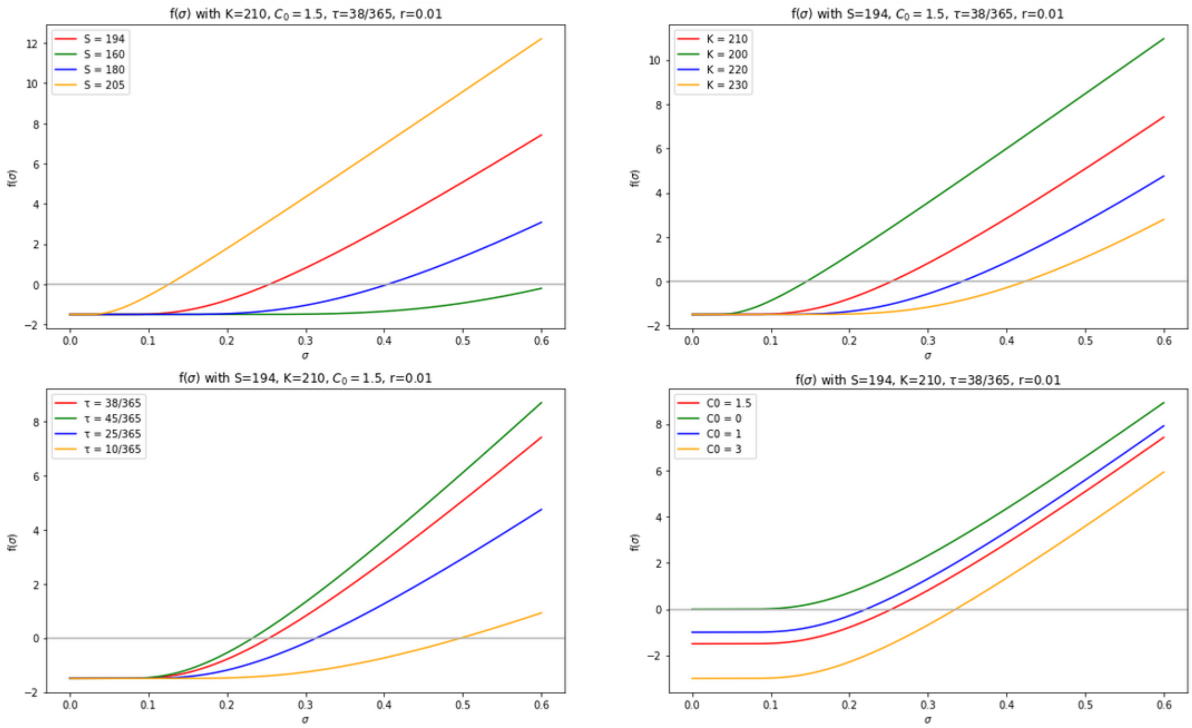

Figure 5.8: $f(\sigma)$ with different inputs

We've now specified and plotted the problem we want to solve, but haven't talked about the solution, yet. Since we want to solve to solve $f(\sigma) = 0$, we need to introduce a root finding algorithm that does the job well. The Newton–Raphson method is a very good choice for that task, since it converges very fast.

Theorem 5.4.1 (Newton–Raphson method)**.** Assume we want to solve $f(x) = 0$, where $x \in R$. This is called finding the roots of function f. The **Newton–Raphson method** is a root finding algorithm of a function $f(x)$ that solves $f(x) = 0$ iteratively until a threshold ϵ is reached. The algorithm starts with an initial guess for x_0 and approaches a solution to the problem until $0 < |x_k - x_{k+1}| < \epsilon$. We can specify the Newton–Raphson method as

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} \ .$$

The Newton–Raphson method can also be applied to $f(\sigma)$.

Theorem 5.4.2 (Newton–Raphson method for estimating implied volatility)**.** Assume we want to estimate the implied volatility of a European option by computing $f(\sigma) = C(\sigma) - C_0 = 0$. The Newton–Raphson method like defined in 5.4.1 solves the problem if $0 < |\sigma_k - \sigma_{k+1}| < \epsilon$ for some predefined ϵ is reached. This is written as

$$\sigma_{k+1} = \sigma_k - \frac{f(\sigma_k)}{\nu}$$

where $\nu = \frac{\partial f}{\partial \sigma}$. We've introduced the letter ν (Vega) as the derivative of f w.r.t. σ, which can be written as $\nu = \frac{\partial f}{\partial \delta} = SN(d_1)\sqrt{\tau}$.

The letter ν is part of "the Greeks", which is a collection of variables who make it possible to assess risk in the option market. The Greeks not only contain ν, but also σ. Although ν is part of the Greeks, it's not a Greek letter. There are many more of those letters and each of them has a different meaning. We've used some of them up to here without stating them as such explicitly. An understanding of the Greeks is not necessary to follow along the rest of this work, but it's highly advisable if someone wants to gain greater understanding of European options.

We'll now do an example and apply the Newton–Raphson method 5.4.1 to $f(\sigma)$. To do so, we will work with real data, which is visualized in the following figure and a screenshot from [1].

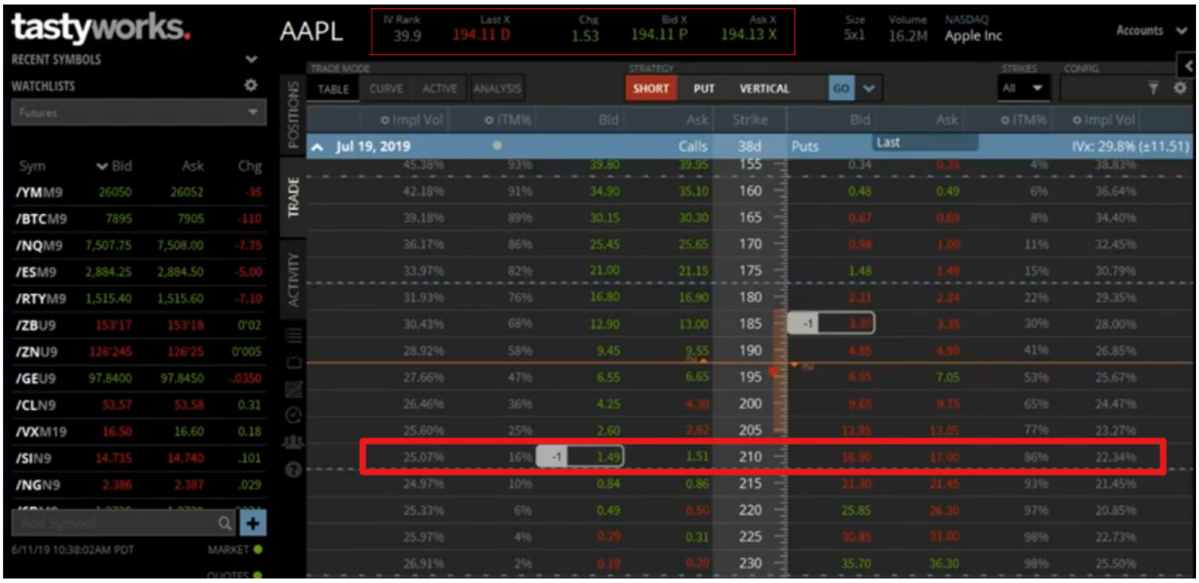

Figure 5.9: Screenshot of European option trading platform

Example 5.4.3 (Calculating implied volatility with the Newton–Raphson method)**.** Assume we want to buy a European put option. For that we have a look at the market in order to identify undervalued options. In Figure 5.9 you can see a screenshot of a real trading platform. We are interested in the contract surrounded by the red box. We start extracting the values that are presented to us and write down the relevant variables for the BSM model:

$$\tau = \frac{38}{365}\text{years} = 0,1041\text{years}$$

$$K = 210€, S = 194€, C_0 = \frac{1,49 + 1,51}{2} = 1,5€$$

In case you are searching for S in Figure 5.9, have a look at the upper and smaller red box. In the figure you can also see a column called "Bid". That's the highest price a buyer will pay for the contract. You can also see an "Ask" column. That's the lowest price a seller will sell the contract for. The difference between Bid and Ask is called spread and determines how liquid an financial instrument is. The market price C_0 is usually an agreement on the seller and buyer side. For simplicity reasons, we take the weighted average for Bid and Ask as the market price. Furthermore, we can look up a good estimation for the risk-free rate r. Let's just go with $r = 1\% = 0,01$. Recall that we want to minimize f, where $f = C(\sigma) - C_0$. We say, that the Newton Raphson method converges if $0 < |x_k - x_{k+1}| < \epsilon$, where $\epsilon = 0,0001$.

We have now gathered all the relevant information and compute $d_1 = -0,7596$, $d_2 = -0,8564$, $N(d_1) = 0,22375$, $N(d_2) = 0,1959$. We start applying 5.4.2 with the initial guess $\sigma_0 = 30\% = 0.3$ to get

$$\sigma_1 = \sigma_0 - \frac{f(\sigma_0)}{\frac{\partial f}{\partial \sigma}(\sigma_0)} = 0.3 - \frac{0,881}{18,7146} = 0,2566 \; .$$

Now we've gotten $\sigma_1 = 0,2566$. But since $|\sigma_1 - \sigma_0|$ is not smaller than ϵ, we go to the next step and compute σ_2. We do that until we converge. The next steps aren't computed by hand, since it is much easier to do that with python. Eventually the Newton–Raphson method will give us the value for the implied volatility of $\sigma = 0,2539 = 25,39\%$. According to the trading platform, the current implied volatility is $25,07\%$. So our computations worked out fine and give us a similar result as the trading platform. The results will never match perfectly, since we could easily consider other factors or methods to estimating implied volatility.

In the above Example 5.4.3 we've computed implied volatility algorithmically, but we can also show that process of convergence visually. For that have a look at Figure 5.10. We again start with an initial guess for σ_0. Here we've chosen $\sigma_0 = 60\%$, because for $\sigma_0 = 30\%$ the method converges too quickly to visualize it well. We then take the tangent line at $f(\sigma_0)$ and look where it crosses the x-axis. This intersection is our new guess for σ. We haven't converged yet, so we go up to $f(\sigma_1)$ and again take the tangent, then we look at the intersection between the tangent of $f(\sigma_1)$ with the x-axis and repeat that until we converge. This whole process is visualized in the figure below.

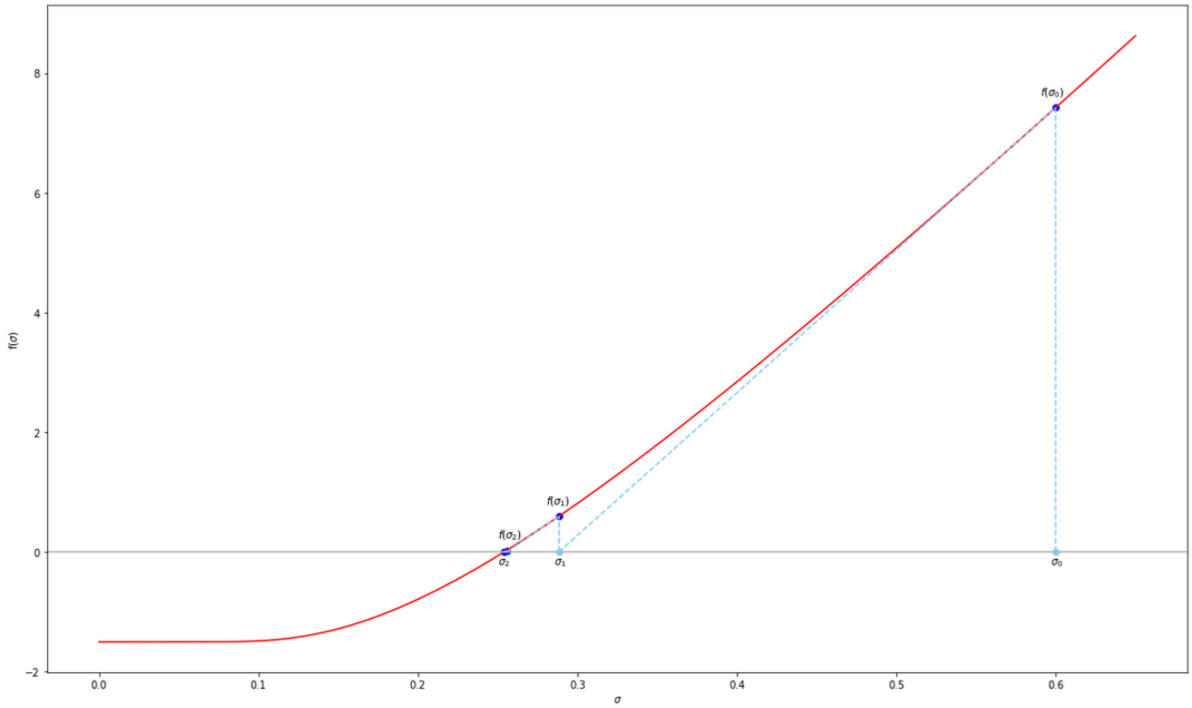

Figure 5.10: Newton-Raphson method visually

We see, that the Newton-Raphson method converges with only 3 iterations, which is very quick. From Figure 5.10 we can also see, that we get a closed form solution. This is one of the astonishing aspects of the Black-Scholes-Merton model of investing.

It should be noted that the Newton-Raphson method usually diverges if we choose σ_0 too close to 0. The reason for this is, that the slope in that point would then be also close to 0. Furthermore, our method could oscillate between 2 points and therefore never converge. All in all, the Newton-Raphson method is a good choice.

The principle of implied volatility is important, because it lays the foundation for many trading strategies. One example of that would be volatility arbitrage. Generally speaking, we want to buy volatility when we think it's low and sell volatility when it is high. All of that happens in a delta-neutral portfolio. There are many resources online, that cover all of that in detail.

5.5 Conclusion

We've seen in this chapter how to derive the Black-Scholes-Merton partial differential equation of investing and how to apply the Black-Scholes-Merton model of investing to our data. There we've introduced the Newton-Raphson method for computing implied volatility.

Generally speaking, there are some problems with the assumption that the Black-Scholes-Merton model of investing makes. For example it assumes paths to be continuous, although that's not the case in the real world. There are frequent price jumps in the market often caused by announcements of companies. Usually one would also have to pay transaction costs to make a deal. Nevertheless, the Black-Scholes-Merton model of investing gives us a closed form solution and lays the groundwork for many other models in the field.

All throughout this work we have only touched upon the concept of hedging, because we could dedicate a whole chapter to that topic. If one wants to go further into those topics, then the book [3] by Steven E. Shreve and the book [6] by Paul Wilmott are good choices to do so. Also MIT has a playlist on topics in mathematics with applications in finance, where also other topics are covered. As a general remark, don't look at quantitative finance as only one subject, it's a degree.

References

[1] Calculating Implied Volatility from an Option Price Using Python, howpublished = https://www.youtube.com/watch?v=Jpy3iCsijIU, note = Date of access: 2022-06-01.

[2] Investopedia, howpublished = `https://www.investopedia.com/`, note = Date of access: 2022-03-13.

[3] Steven E. Shreve. *Stochastic Calculus for Finance II: Continuous-Time Models*. 2004 Springer Science+Business Media Inc., 2003.

[4] Karl Sigman. 4700-07. geometric brownian motion. 2006.

[5] Karl Sigman. 4700-07. notes on brownian motion, 2006.

[6] Paul Wilmott. *Paul Wilmott Introduces Quantitative Finance*. John Wiley Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England, 2007.

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Bachelorarbeit selbstständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt bzw. die wörtlich oder sinngemäß entnommenen Stellen als solche kenntlich gemacht habe. Die vorliegende Bachelorarbeit ist mit dem elektronisch übermittelten Textdokument identisch.

Ort, Datum

Unterschrift