

THE MARTINGALE APPROACH TO FINANCIAL MATHEMATICS

A Thesis

presented to

the Faculty of California Polytechnic State University,

San Luis Obispo

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Mathematics

by

Jordan Rowley

June 2019

© 2019  
Jordan Rowley  
ALL RIGHTS RESERVED

## COMMITTEE MEMBERSHIP

TITLE: The Martingale Approach to Financial  
Mathematics

AUTHOR: Jordan Rowley

DATE SUBMITTED: June 2019

COMMITTEE CHAIR: Dana Paquin, Ph.D.  
Professor of Mathematics

COMMITTEE MEMBER: Robert Easton, Ph.D.  
Professor of Mathematics

COMMITTEE MEMBER: Erin Pearse, Ph.D.  
Professor of Mathematics

## ABSTRACT

### The Martingale Approach to Financial Mathematics

Jordan Rowley

In this thesis, we will develop the fundamental properties of financial mathematics, with a focus on establishing meaningful connections between martingale theory, stochastic calculus, and measure-theoretic probability. We first consider a simple binomial model in discrete time (Chapter 1), and assume the impossibility of earning a riskless profit, known as arbitrage. Under this no-arbitrage assumption alone, we stumble upon a strange new probability measure  $Q$ , according to which every risky asset is expected to grow as though it were a bond. As it turns out, this measure  $Q$  also gives the arbitrage-free pricing formula for every asset on our market. In considering a slightly more complicated model over a finite probability space (Chapter 2), we see that  $Q$  once again makes its appearance. Finally, in the context of continuous time (Chapter 3), we build a framework of stochastic calculus to model the trajectories of asset prices on a finite time interval. Under the absence of arbitrage once more (Chapter 4), we see that  $Q$  makes its return as a Radon-Nikodym derivative of our initial probability measure. Finally, we use the properties of  $Q$  and a stochastic differential equation that models the dynamics of the assets of our market, known as the Itô formula, in order to derive the classic Black-Scholes Equation.

# TABLE OF CONTENTS

	Page
CHAPTER	
1 THE BINOMIAL MODEL . . . . .	1
1.1 Introduction . . . . .	1
1.2 The Absence of Arbitrage . . . . .	2
1.3 Stumbling Upon $Q$ . . . . .	3
1.4 Pricing Contingent Claims . . . . .	5
2 THE ONE PERIOD MODEL . . . . .	8
2.1 Building a More Complicated Market . . . . .	8
2.2 Farkas' Lemma . . . . .	9
2.3 A Surprising Result . . . . .	13
3 STOCHASTIC CALCULUS . . . . .	15
3.1 The Wiener Process . . . . .	15
3.2 Random Processes as Wiener Functionals . . . . .	17
3.3 The Stochastic Integral . . . . .	20
3.4 The Itô Formula . . . . .	27
3.5 Some examples using Itô's Formula . . . . .	33
3.6 The Martingale Representation Theorem . . . . .	35
4 A CONTINUOUS TIME MODEL . . . . .	42
4.1 Portfolio Dynamics . . . . .	42
4.2 The Search for $Q$ . . . . .	43
4.3 The First Fundamental Theorem of Asset Pricing . . . . .	43
4.4 Characterizing $Q$ as an Equivalent Probability Measure . . . . .	51

4.5 The Black-Scholes Formula . . . . .	63
BIBLIOGRAPHY . . . . .	65
APPENDIX . . . . .	66

# 1 THE BINOMIAL MODEL

## 1.1 Introduction

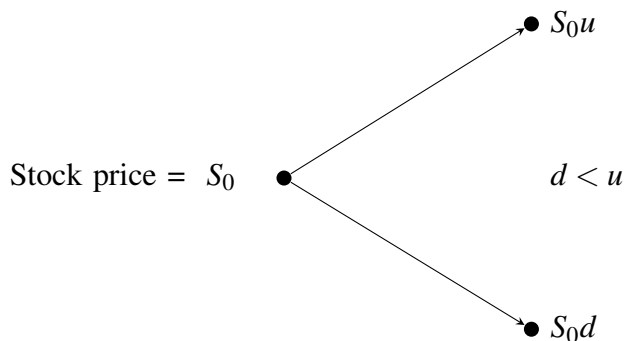
Consider a simple market consisting of two assets: a riskless bond and a risky stock. We are interested in how these assets change after a single (discrete) increment of time from  $t = 0$  to  $t = 1$ . If  $B_0$  denotes the amount of money we have initially invested in bonds, then we are absolutely certain that when  $t = 1$ :

$$B_1 = (1 + r)B_0,$$

where  $r$  denotes the periodic risk-free interest rate. For simplicity, we will consider the bond of this market to have an initial value of \$1. While we may predict the behavior of  $B_0$ , the behavior of a share of stock is not predictable: if  $S_i$  denotes the value of a share at time  $i$ , then

$$S_1 = \begin{cases} S_0u & \text{with probability } p_u \\ S_0d & \text{with probability } p_d \end{cases}.$$

This is given by the binomial tree



Definition: A portfolio is simply a collection of stocks and bonds. A portfolio consisting of  $x$  bonds and  $y$  shares of stock will be denoted by  $h(x,y)$ . The value process of this portfolio at time  $t$ , denoted  $V_t^h$ , is defined by

$$V_t^h = (1+r)^t \cdot x + S_t \cdot y,$$

where  $t$  is either 0 or 1.

## 1.2 The Absence of Arbitrage

Perhaps the most basic assumption in mathematical finance is the absence of arbitrage. This simply means that there is no way to make a riskless profit at a rate higher than the risk-free interest rate. We'll understand this in the context of portfolios:

Definition: An arbitrage portfolio  $h$  is a portfolio whose value process satisfies

$$V_0^h = 0,$$

$$V_1^h > 0 \text{ with probability 1.}$$

It is important to remember that arbitrage does not simply mean earning a riskless profit. Rather, it means that there is an investment strategy whose rate of return is guaranteed to be higher than the risk-free interest rate  $r$ .

Now what sort of portfolio would have to exhibit this? If  $h(x,y)$  denotes an arbitrage portfolio of  $x$  shares of bonds and  $y$  shares of stock, then it must be true that

$$x + S_0 \cdot y = 0$$

$$(1+r) \cdot x + S_{0d} \cdot y > 0$$

$$(1+r) \cdot x + S_{0u} \cdot y > 0.$$

Since  $y \neq 0$ , we will first consider when  $y < 0$ . Combining the first equation with the first inequality gives  $(1+r) > u$ . In addition, we can combine the case in which  $y > 0$



with the second inequality to obtain  $(1+r) < d$ . Therefore, the assumption of arbitrage opportunities implies that either  $(1+r) > u$  or  $(1+r) < d$ . Contraposition leads us to the following theorem.

Theorem: The binomial model is arbitrage free if and only if

$$d \leq 1+r \leq u.$$

Proof: We have already established the backwards direction. Showing the forward direction consists in reversing the above calculations.  $\square$

### 1.3 Stumbling Upon $Q$

Given the inequality presented by the above theorem, we can find two unique real numbers  $q_1$  and  $q_2$  such that

$$uq_1 + dq_2 = 1+r,$$

where

$$q_1 + q_2 = 1.$$

Convexity implies that this pair is the unique solution to this system of equations. In fact  $q_1$  and  $q_2$  are given by

$$q_1 = \frac{(1+r) - d}{u - d}; q_2 = \frac{u - (1+r)}{u - d}.$$

It is tempting to understand  $q_1$  and  $q_2$  as probabilities from some alternate probability space  $Q$ , and if we multiply the first equality by  $S_0$ , we see the expected value of  $S_1$  under  $Q$ . Therefore, according to this new probability measure,

$$\mathbf{E}^Q[S_1] = S_0(1+r)$$

In this strange world of  $Q$ , even the risky asset is expected to grow as though it were a bond. This type of probability measure is called a martingale measure. As we will see, the existence of such a measure, in market models as simple as this binomial model, to ones built on continuous time, is equivalent to the absence of arbitrage.

Definition: A probability measure  $Q$  is a martingale measure if

$$S_0 = \frac{1}{1+r} \mathbf{E}^Q[S_1].$$

The martingale measure  $Q$  may also be given the equivalent equation

$$\frac{S_0}{B_0} = \mathbf{E}^Q \left[ \frac{S_1}{B_1} \right].$$

While this might seem to be a matter of taste, the above equation conveys a feature of martingale measures that we will see throughout the later sections of this thesis. By the above equation, the normalized asset  $\frac{S_0}{B_0}$ , which is simply the risky asset divided by the bond, is not expected to change according to  $Q$ . In the parlance to be used in the upcoming sections, we say that the martingale measure perceives the normalized asset to “exhibit no drift,” in which case we say that the normalized asset is a  $Q$  - martingale. One of the goals of this thesis is to investigate how this theme recurs even in much more complicated markets.

In any event, there exists an essential connection between the existence of  $Q$  and the absence of arbitrage.

Theorem: The binomial market model is free of arbitrage if and only if there exists a martingale measure  $Q$ .

Proof: From the first theorem of this section, a market is free of arbitrage if and only if it satisfies the inequality  $d \leq 1+r \leq u$ , and this inequality itself begot the probability measure  $Q$ . □

## 1.4 Pricing Contingent Claims

Now this measure  $Q$ , which seems to have arisen out of nowhere, will play a very important role in the pricing of certain other assets on our market. In addition to stocks and bonds, our market can contain assets whose values depend entirely on whether the stock has gone up or down. These assets are known as contingent claims.

Definition: A contingent claim is a function  $X$  of the form  $X = \Phi(S_t)$ .

An example of this is an option, which may be understood as a formal promise to allow the purchaser of the option to buy or sell the stock at a stipulated price  $K$ , known as the strike price, at  $t = 1$ .

One of the simplest claims in the market is a European call option. As an example, consider the case in which the future value of our stock is given by the possibilities

$$S_0d = \$20$$

$$S_0u = \$40$$

Now suppose there is also a European call option that promises the buyer that he/she can buy the stock for a price of \$30 at  $t = 1$ . We can find possible values of this option at this moment in time by considering the two possible outcomes. If  $S_1 = S_0d = \$20$ , then the option is absolutely worthless. Why would you buy something for \$30 when its current market value is \$20? On the other hand, if  $S_1 = S_0u = \$40$ , then the holder of the option has struck it big. He/She can buy the stock for \$30, then proceed to sell it for its current value of \$40, securing a profit of \$10.

In general, the function defining the price of this option at time  $t = 1$  is given by

$$\Phi(S_0u) = S_0 - K,$$

$$\Phi(S_0d) = 0.$$

Now let's denote the price of a contingent claim  $X$  at time  $t$  by  $\Pi(t; X)$ . We know the possible values that this claim can assume at time  $t = 1$ , but how are we to price it at  $t = 0$ ?

Since the only options we know have definite value at  $t = 0$  are stocks and bonds, perhaps we can assemble a portfolio whose value at  $t = 1$  is exactly that of  $\Pi(1;X)$ . This process is known as hedging.

Definition: A contingent claim  $X$  is reachable if there exists a portfolio  $h$  such that  $V_1^h = \Pi(1;X)$ .

Now if such a portfolio existed, then it must be true that this portfolio provides the arbitrage-free price from this claim. After all, if another pricing method existed for the claim  $X$ , which we will denote by  $\Lambda(t,X)$ , then the absence of arbitrage implies that  $\Lambda(t,X) = \Pi(t;X)$  for both  $t = 0$  and  $t = 1$ .

So how can we construct a hedging portfolio for the claim  $X = \Phi(S_t)$ ? Any such portfolio  $h(x,y)$  would have to satisfy the equations

$$(1+r) \cdot x + S_0 d \cdot y = \Phi(S_0 d)$$

$$(1+r) \cdot x + S_0 u \cdot y = \Phi(S_0 u)$$

Solving for  $x$  and  $y$ , we obtain

$$x = \frac{1}{1+r} \cdot \frac{u\Phi(S_0 d) - d\Phi(S_0 u)}{u-d}$$

$$y = \frac{1}{S_0} \cdot \frac{\Phi(S_0 u) - \Phi(S_0 d)}{u-d}$$

The portfolio  $h(x,y)$  provides the pricing formula for our contingent claim. After a little arithmetic, the price of  $X$  at time  $t = 0$  is therefore equal to

$$\Pi(0;X) = V_0^h = x + S_0 \cdot y = \frac{1}{1+r} \left[ \frac{(1+r) - d}{u-d} \cdot \Phi(S_0 u) + \frac{u - (1+r)}{u-d} \cdot \Phi(S_0 d) \right].$$

What a mess! However, we can re-express the above equation to be of the form

$$\Pi(0;X) = \frac{1}{1+r} [q_1 \Phi(S_0 u) + q_2 \Phi(S_0 d)].$$

The right hand side is simply the expected value of  $\Phi(S_t)$  under  $Q$ ! This probability measure  $Q$  has given us a pricing formula for  $X$  at time  $t = 0$  defined by

$$\Pi(0;X) = \frac{1}{1+r} \mathbf{E}^Q[X].$$

We may rewrite this as

$$\frac{\Pi(0;X)}{B_0} = \mathbf{E}^Q \left[ \frac{\Pi(1;X)}{B_1} \right].$$

We have discovered the primary function of the probability measure  $Q$ . The contingent claims of our market, whose initial price we knew nothing about, now have pricing processes that are given by this martingale measure. Moreover, the pricing process provided by  $Q$  for a given claim is the only one consistent with the absence of arbitrage. The power of this strange probability measure  $Q$  seems quite astonishing, given the fact that it arose out of entertaining the innocent idea that  $q_1$  and  $q_2$  may be understood as probabilities.

Now we will consider more complicated market models in order to see whether this measure  $Q$  returns as it did in this section.

## 2 THE ONE PERIOD MODEL

We considered the simple case in which a future stock price has only two possible states. By assuming the absence of arbitrage, we encountered this strange new probability  $Q$  that views stocks to behave as though they were bonds. But was  $Q$  just a mere curiosity of the simple binomial model? To explore a more complicated case, let's create a more general market model over a single discrete period of time in which both the number of available assets and possible outcomes is finite but arbitrary.

### 2.1 Building a More Complicated Market

Suppose that  $S^1, S^2, \dots, S^{N-1}$  denote the  $N - 1$  risky assets in our market, and our state space  $\Omega = \{\omega_1, \omega_2, \dots, \omega_M\}$  consists of  $M$  possible outcomes. We will still maintain that there is a risk-free asset  $B$  whose value at time  $t = 1$  is known with absolute certainty. The value of risky asset  $i$  at time  $t = 1$ , denoted by  $S_1^i$ , is a random variable that will assume one of  $M$  possible outcomes, all of which are captured by the set

$$\{S_1^i(\omega_1), S_1^i(\omega_2), \dots, S_1^i(\omega_M)\}.$$

If we organize these sets for each asset into an  $N$  by  $M$  matrix, we will have the full story of whatever could happen in the market at time  $t = 1$ . We will denote this matrix by  $\mathcal{D}$ , so that

$$\mathcal{D} = \begin{bmatrix} B_1(\omega_1) & B_1(\omega_2) & \dots & B_1(\omega_M) \\ S_1^1(\omega_1) & S_1^1(\omega_2) & \dots & S_1^1(\omega_M) \\ \vdots & \vdots & \ddots & \vdots \\ S_1^{N-1}(\omega_1) & S_1^{N-1}(\omega_2) & \dots & S_1^{N-1}(\omega_M) \end{bmatrix}.$$

As in the previous section, a portfolio  $h$  is simply a linear combination of assets that may be understood as an  $N$ -dimensional vector of the form

$$h = (h_0, h_1, \dots, h_{N-1}).$$

If we equate the risk free asset  $B$  with  $S^0$ , then  $h_i$  denotes the (possibly negative) amount of asset  $S^i$  owned by the portfolio. The value process of the portfolio at time  $t$ , denoted by  $V_t^h$ , is given by

$$V_t^h = \sum_{i=0}^{N-1} h_i S_t^i.$$

If we let  $S_t$  denote the vector of random variables of the form  $S_t^i$ , so that

$$S_t = (S_t^1, S_t^2, \dots, S_t^{N-1}),$$

then the portfolio value process at time  $t$  is simply the dot product  $h \cdot S_t$ .

In this section, we will work with a seemingly different but equivalent definition of an arbitrage portfolio.

Definition: An arbitrage portfolio  $h$  is a portfolio whose value process is defined by

$$V_0^h < 0$$

$$V_1^h \geq 0 \text{ with probability } 1$$

The reason why this definition uses a portfolio whose initial value is negative will be made clear with the following important lemma.

## 2.2 Farkas' Lemma

Lemma: Farkas' Lemma. Let  $K \subseteq \mathbb{R}^n$  denote the convex cone spanned by the vectors  $u_1, u_2, \dots, u_M$ , and let  $u_0 \in \mathbb{R}^n \setminus K$ . Then there exists a vector  $h$  such that

$$h \cdot u_0 < 0$$

$$h \cdot u_i \geq 0 \forall i = 1, 2, \dots, M$$

This lemma gives us a geometrical interpretation of arbitrage in our market: if we let the vector  $S_0$  consist of the initial values of all of our assets, then unless  $S_0$  lies within the

cone spanned by the column vectors of our market matrix  $\mathcal{D}$ , our market will have arbitrage opportunities!

To prove this lemma, we will need to define the concept of a hyperplane.

Definition: A hyperplane in  $\mathbb{R}^n$  is any subset  $H(\delta, v)$  defined by

$$H(\delta, v) = \{u \in \mathbb{R}^n : u \cdot v = \delta\}$$

for some  $v \in \mathbb{R}^n$ ,  $\delta \in \mathbb{R}$ .

Proof: We will show that there exists a hyperplane that separates  $u_0$  from  $K$ . It can be shown that the cone  $K$  is closed, so there exists a unique element  $x_0 \in K$  whose distance from  $u_0$  is minimal.

To create a separating hyperplane, we first let  $v_1 = x_0 - u_0$  denote the vector connecting  $x_0$  to  $u_0$ . We will choose the midpoint  $\frac{1}{2}(x_0 + u_0)$  to lie on our plane, so let

$$\delta = \frac{1}{2}(x_0 + u_0) \cdot (x_0 - u_0) = \frac{1}{2}(|x_0|^2 - |u_0|^2).$$

The set  $H(\delta, v_1)$  is a candidate for our separating hyperplane. This set induces a partition of  $\mathbb{R}^n$  into two pieces:

$$H_1 = \{u \in \mathbb{R}^n : u \cdot v_1 < \delta\}$$

$$H_2 = \{u \in \mathbb{R}^n : u \cdot v_1 \geq \delta\}.$$

We need to show that the cone  $K$  is completely contained within one of these subsets. In order to find out which one, let's find out which set contains  $u_0$ . A little backwards thinking gives us the inequality

$$(x_0 - u_0) \cdot u_0 < \frac{1}{2}(|x_0|^2 - |u_0|^2) = \delta,$$

whence

$$(x_0 - u_0) \cdot x_0 \geq \delta.$$



Now if we can show that for any  $y \in K$ ,  $(y - x_0) \cdot (x_0 - u_0) \geq 0$ , then this would imply that

$$(x_0 - u_0) \cdot y - (x_0 - u_0) \cdot x_0 \geq 0$$

$$(x_0 - u_0) \cdot y \geq (x_0 - u_0) \cdot x_0 \geq \delta.$$

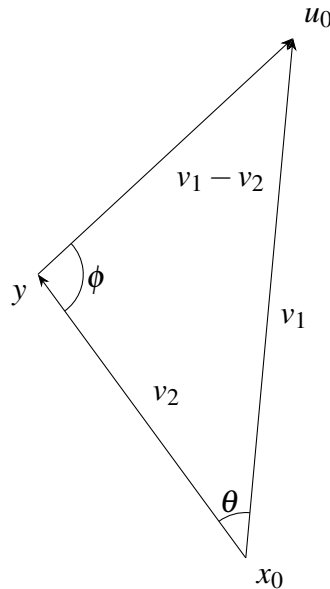
Intuitively, this means that it is sufficient to prove that for any other point  $y$  on our cone, the angle between the vectors connecting  $x_0$  to  $y$  (denoted  $v_2$ ) and  $x_0$  to  $u_0$  ( $v_1$ ) is *not* acute.

Seeking a contradiction, let's suppose this angle is indeed acute, so that

$$(y - x_0) \cdot (u_0 - x_0) \geq 0.$$

We will divide into two cases pertaining to the angle between the vectors extending from  $y$  to  $x_0$  and  $y$  to  $u_0$ , denoted  $\phi$ .

Case 1:  $\phi$  is obtuse



As suggested by our diagram, we want to derive a contradiction by showing that the distance from the point  $y$  to  $u_0$  is less than from  $x_0$  to  $u_0$ .

The obtuseness of  $\phi$  implies that

$$(v_1 - v_2) \cdot (-v_2) < 0$$

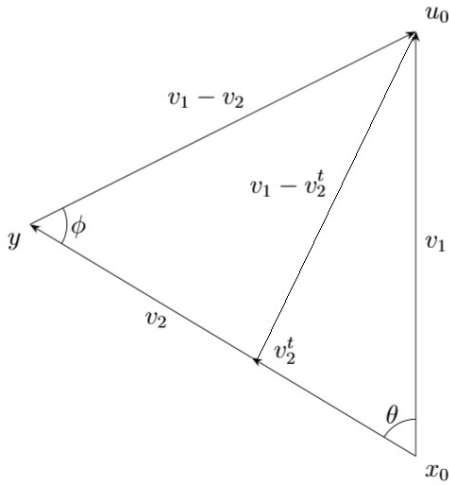
$$|v_2|^2 - 2(v_1 \cdot v_2) < 0$$

$$|v_1|^2 - 2(v_1 \cdot v_2) + |v_2|^2 < |v_1|^2$$

$$|v_1 - v_2|^2 < |v_1|^2,$$

which contradicts the minimality of  $v_1$ .

Case 2:  $\phi$  is acute



For this case, let

$$y(t) = y - (y - x_0)t, \text{ where } t \in [0, 1].$$

The vector  $v_2^t = y(t) - x_0$  may be understood as travelling from  $y$  to  $x_0$  as time ranges from 0 to 1. Now when  $t = 0$ ,

$$-v_2^0 \cdot (v_1 - v_2^0) = -(y(0) - x_0) \cdot (v_1 - (y(0) - x_0)) = -(v_2) \cdot (v_1 - v_2) \geq 0.$$

However, for some time  $t \in [0, 1]$ ,

$$-v_2^t \cdot (v_1 - v_2^t) < 0.$$

By continuity, there is some time  $t_0 \in [0, 1]$  such that

$$-v_2^{t_0} \cdot (v_1 - v_2^{t_0}) = 0.$$

This implies that

$$|v_1|^2 = |(v_1 - v_2^{t_0}) + v_2^{t_0}|^2 = |(v_1 - v_2^{t_0})|^2 + |v_2^{t_0}|^2.$$

Once more this contradicts minimality of  $v_1$ .

By contradiction, we have reached the result that

$$\forall y \in K, (y - x_0) \cdot (x_0 - u_0) \geq 0,$$

which, as mentioned before, implies that

$$(x_0 - u_0) \cdot y \geq \delta.$$

This tells us that every point on the cone  $K$  is on the side of the hyperplane that is opposite of  $u_0$ . To complete the proof, let's characterize the constant  $\delta$ .

As a cone,  $K$  must contain the origin. Therefore,  $\delta \leq 0$ . Moreover, if there exists a vector  $w$  in  $K$  such that  $v_1 \cdot w < 0$ , then there exists a natural number  $n$  such that  $v_1 \cdot nw < \delta$ , which is a contradiction. From all of this, we have finally arrived at the following results:

$$\forall x \in K, x \cdot v_1 \geq 0$$

$$u_0 \cdot v_1 < 0.$$

□

### 2.3 A Surprising Result

Farkas' lemma gives a remarkable implication for our market. If we denote the column vectors of  $\mathcal{D}$  by  $d_1, d_2, \dots, d_M$ , the initial value vector  $S_0$  is given by

$$S_0 = \sum_{i=1}^M \lambda_i d_i, \text{ where } \lambda_i \geq 0.$$

The fact that each coefficient of  $d_i$  is nonnegative allows us to construe this sum as an expected value! If we normalize the coefficients by letting

$$q_i = \frac{\lambda_i}{\beta}, \text{ where } \beta = \sum_i \lambda_i,$$

then we may rewrite  $S_0$  as

$$S_0 = \beta \mathbf{E}^\Lambda[S_1],$$

where  $\Lambda$  is some probability measure. Since  $S_0$  is simply the vector containing the initial price of each asset in our market, then if we set the initial price of the risk-free asset  $B$  to be \$1, then the first coordinate of each of column vectors of  $\mathcal{D}$  is simply  $1 + r$ . Therefore,

$$1 = \beta(1 + r),$$

so that

$$S_0 = \frac{\mathbf{E}^\Lambda[S_1]}{1 + r}.$$

The probability measure  $\Lambda$  is precisely the martingale measure  $Q$  that we saw in the previous section! Once more, we see that according to  $Q$ , for any asset  $S^i$ , we have

$$\frac{S_0^i}{B_0} = \mathbf{E}^Q \left[ \frac{S_1^i}{B_1} \right].$$

So the mysterious probability measure  $Q$  has returned, even when we constructed a market with many assets and many possible outcomes. As was the case in the previous section, the martingale measure  $Q$  functions to give us an arbitrage-free pricing formula for contingent claims, but we will not develop this idea for this market any further, because the concept remains nearly the same as that of the previous section.

Instead, a natural question to ask is whether  $Q$  will also make its appearance in the context of continuous time? The main topic of this thesis will consist in building a market in continuous time in order to search for the martingale measure  $Q$  once more.

### 3 STOCHASTIC CALCULUS

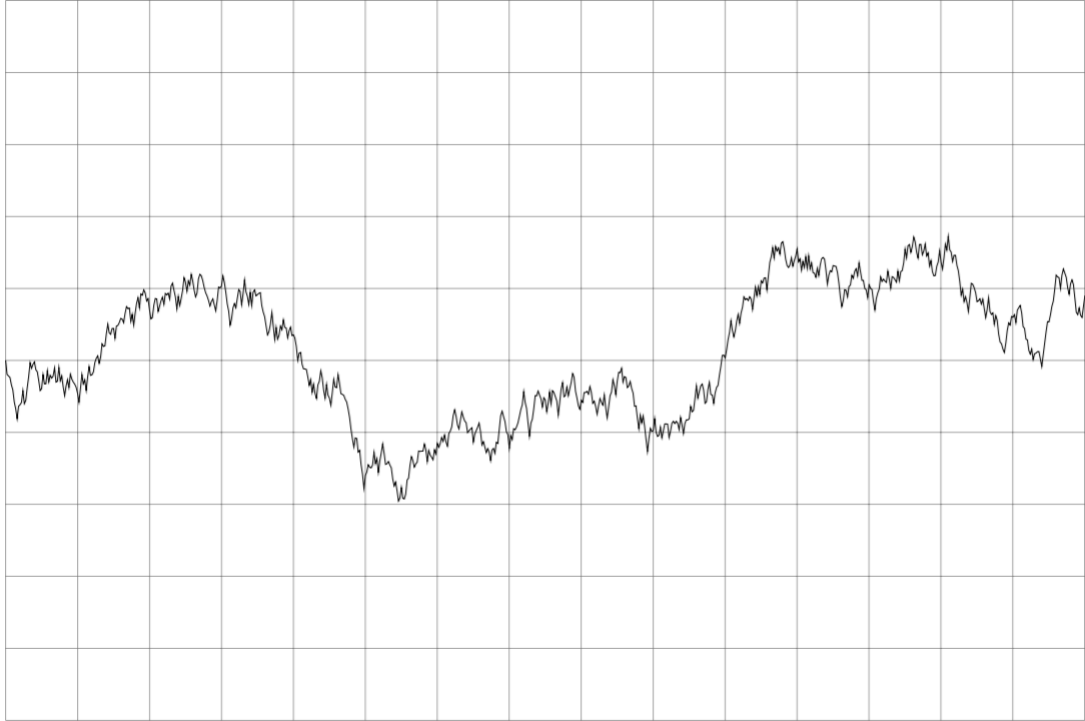
If we are to venture into the context of continuous time with the hope of creating a model for market behavior, we should start by characterizing the type of random processes that we will encounter. The trajectory of a risky asset, such as a stock, is random in the sense that its value in the future appeals only to its present value, while its past values are considered irrelevant. While certain statistical properties of the random process followed by the stock price, such as volatility, may be estimated based upon its past values, the particular path followed by the asset up until the present time has no relevance on its future trajectory.

#### 3.1 The Wiener Process

In the context of continuous time, the stochastic process that drives this Markovian behavior of risky assets is known as a Wiener Process.

**Definition:** A Wiener Process is a stochastic process  $W(t)$  defined by

1.  $W(0) = 0$
2.  $\forall r < s \leq t < u$ , the random variables  $W(u) - W(t)$  and  $W(s) - W(r)$  are independent.
3. for  $s < t$ , the random variable  $W(t) - W(s)$  is normally distributed with mean 0 and variance  $(t - s)$ .
4.  $W(t)$  is continuous with probability 1.



An example of a Wiener trajectory

As mentioned above, a risky financial asset is driven by an underlying Wiener process in the sense that the random behavior of the asset  $X(t)$  is due to the Wiener process alone. Of course, this does not imply that  $X(t)$  is itself a Wiener process: if we examined the change in  $X(t)$  over the interval  $[t, t + \Delta t]$ , we may find that the local dynamics of  $X(t)$  may exhibit some sort of systematic “drift” in addition to the “noise” churned out by the underlying Wiener process. That is, the change undergone by  $X$  from  $t$  to  $t + \Delta t$  may be given by

$$X(t + \Delta t) - X(t) = \mu(t, X(t))\Delta t + \sigma(t, X(t))\Delta W(t).$$

The term  $\mu(t, X(t))\Delta t$  may be understood as the local drift term, while  $\sigma(t, X(t))\Delta W(t)$  denotes stochastic noise amplified by the process  $\sigma(t, X(t))$ .

To frame this equation in the context of integration, we will let  $\Delta t \rightarrow 0$  in order to formally obtain

$$dX(t) = \mu(t, X(t))dt + \sigma(t, X(t))dW(t).$$

Perhaps this expression is nothing more than shorthand for

$$X(t) = \int_0^t \mu(s, X(s))ds + \int_0^t \sigma(s, X(s))dW(s).$$

But then what exactly does the second integral represent? What does it mean to integrate with respect to a Wiener process? Tackling this issue requires us to lay out a fair amount of fancy machinery, but first let's focus on understanding  $X(t)$  as a random variable.

### 3.2 Random Processes as Wiener Functionals

From this formulation, we see that the process  $X$  is actually a functional of a Wiener trajectory. Given a time  $t$ , the presence of the second integral in the above equation implies that the value of  $X(t)$  appeals to the past values of the Wiener process leading up to that particular moment.

On the other hand, if the present value of  $X(t)$  appealed to future states of the Wiener process, we would have an asset that looks into the future of the Wiener trajectory in order to derive its present value. We have reason to believe that such oracular assets do not actually exist. Therefore, it must be true that for any given time  $t$ , the value of  $X(t)$  appeals to the trajectory of the underlying Wiener process only over the interval  $[0, t]$ .

Now we'd like to understand  $X(t)$  at some time  $t$  as some random variable over the state space  $\Omega$  consisting of Wiener trajectories. Furthermore, this random variable  $X$  needs to be measurable in the sense that we can assign a probability to an even such as  $\{X(t) > 2\} \subseteq \Omega$ . With this in mind, we'll understand the process  $X$  as a function not only of time  $t$ , but also of an underlying Wiener trajectory leading up to  $t$ .

In this sense, the random variable  $X(t, \omega)$  denotes the value of the asset at time  $t$  under a Wiener trajectory  $\omega$  over the interval  $[0, t]$ . But then what do the measurable sets of this

probability space look like? If we are interested in the probability distribution of  $X(t, \omega)$ , then our measurable sets should consist of events involving  $X$  at time  $t$ , not some other time in the future. Essentially, we should be able to decide whether or not a given event has occurred solely based upon the trajectory of a Wiener process on the interval  $[0, t]$ . If  $\omega_1, \omega_2$  denote two Wiener trajectories that agreed until time  $t$ , then  $X(t, \omega_1) = X(t, \omega_2)$  precisely because our random variable does not look into the future. Our  $\sigma$ -algebra may therefore be understood as the “flow of information” pertaining to the underlying Wiener trajectory  $W$  that is available up until time  $t$ , which we will denote by  $\mathcal{F}_t^W$ . The exact presentation of measurable sets in this  $\sigma$ -algebra can be made explicit through creating a dense sequence of times  $\{t_i\}_{i=1}^\infty \subseteq [0, t]$  and using the continuity of a Wiener process. This is further discussed in the proof of the Martingale Representation Theorem (page 35).

So the exact subsets of the state space  $\Omega$  that are measurable is decided by the moment in time at which we are viewing a Wiener functional. If we considered some later time  $u \geq t$ , then the  $\sigma$ -algebra of this probability space  $\mathcal{F}_u^W$  will consist of all the information regarding a Wiener trajectory over the interval  $[0, u]$ . Since this already contains all of the information about a Wiener trajectory over the subinterval  $[0, t]$ , we have the relation  $\mathcal{F}_t^W \subseteq \mathcal{F}_u^W$ .

Therefore, as time unfolds, we have more information to decide whether more complicated events have occurred. This ordering of  $\sigma$ -algebras is known as a filtration.

**Definition:** A filtration of a probability space on the  $\sigma$ -algebra  $\mathcal{F}$  is defined by an increasing set of  $\sigma$ -algebras  $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \dots \subseteq \mathcal{F}$ .

In our case, we have used time to index an increasing set of  $\sigma$ -algebras, but there are multiple ways to create such a set. Another way is to consider a dense sequence  $\{t_i\}_{i=1}^\infty \subseteq [0, T]$ , and index a  $\sigma$ -algebra  $\mathcal{G}_i^W$  to each member of this sequence, such that

$$\mathcal{G}_i^W = \sigma\{W(t_1) \dots W(t_i)\}.$$



In this sense, the  $\sigma$ -algebra is generated by where the Wiener process was at the moments  $t_1, t_2, \dots, t_i$ . As mentioned before, we will develop this approach in our proof of the Martingale Representation Theorem (page 35), but for now we will work with the initial filtration that was indexed by time.

Now the random processes  $X$  we would like to consider are such that for any fixed time  $s$ , the value of the random variable  $X(s, \omega)$  can be completely determined by observing a Wiener trajectory over the interval  $[0, s]$ .  $\mathcal{F}_s^W$  is the coarsest  $\sigma$ -algebra that contains all of the information we would need to give a full account of  $X(s, \omega)$ . In this case we say that  $X(s, \omega)$  is “ $\mathcal{F}_s^W$ -measurable,” and write

$$X(s, \omega) \in \mathcal{F}_s^W.$$

When a random process  $X$  is such that for any  $t \geq 0$ ,  $X(t, \omega) \in \mathcal{F}_t^W$ , we say that the process is adapted to the filtration  $\{\mathcal{F}_t^W\}_{0 \leq t \leq T}$

Certainly  $X(s, \omega)$  is also  $\mathcal{F}_t^W$ -measurable for any  $t \geq s$ , but the reverse is not the case. We cannot give a full account of the random variable  $X(t, \omega)$  if we are restricted to working with  $\mathcal{F}_s^W$ , intuitively because this  $\sigma$ -algebra gives us information about a Wiener trajectory only over the interval  $[0, s]$ . We are not getting the full story about where  $X(t, \omega)$  could be, because our  $\sigma$ -algebra is too coarse to capture the differences that  $X(t, \omega)$  could exhibit when given two Wiener trajectories  $\omega_1$  and  $\omega_2$  that differ only after time  $s$ .

However, we can at least form an expectation of its possible values given the information we have available. We will denote this conditional expectation by

$$\mathbf{E}[X(t, \omega) \mid \mathcal{F}_s^W].$$

This expression may be understood as “the expected value of the process  $X$  at time  $t$ , given the information we have available up until time  $s$ .” The conditional expectation

has many interesting properties that will be used in the next section. Most importantly, it characterizes martingales in continuous time!

Recall from the previous sections that a martingale was a random variable, in particular, a normalized asset from our market, that was not expected to change under some probability measure. In the context of continuous time, we say that a random process  $M(t)$  is a martingale when its conditional expectation is equal to its present value.

Definition: Given a filtration  $\{\mathcal{F}_t^W\}_{0 \leq t \leq T}$ , an adapted process  $M(t)$  is an  $(\mathcal{F}_t^W)$ -martingale if

1.  $\mathbf{E}[|M(t)|] < \infty$
2.  $\forall s \leq t$  in  $[0, T]$ ,  $\mathbf{E}[M(t) | \mathcal{F}_s^W] = M(s)$ .

Martingales as random processes will play an absolutely essential role in constructing an arbitrage free market in continuous time, and this importance is due to the connection between a martingale and a stochastic integral, which we now investigate.

### 3.3 The Stochastic Integral

Now that we have an account of how we may view our probability space as a measure space with a changing  $\sigma$ -algebra, let us return to the strange integral we saw in the previous section:

$$\int_0^t \sigma(s, \omega) dW_s.$$

Perhaps we could interpret this expression as a Riemann-Stieltjes integral, so that for a sequence of partitions  $\{\mathcal{P}_i\}_{i=1}^\infty$  whose mesh tends to zero, we have

$$\int_0^t \sigma(s, \omega) dW_s = \lim_{n \rightarrow \infty} \sum_{k=1}^N \sigma(t_{n_k}^*, \omega) [W(t_{n_k}) - W(t_{n_{k-1}})],$$

where  $t_{n_k}^* \in [t_{n_{k-1}}, t_{n_k}]$ . Unfortunately, this limit will not converge because a Wiener process is not of bounded variation! If we chose  $t_{n_k}^*$  to be the midpoint of the interval, then we will get a different value for the integral than if we chose the left endpoint instead. However,

based upon our initial formulation of the local dynamics of the process  $X$  defined above, we will choose  $t_{n_k}^*$  to be the left endpoint of our interval, so that

$$\int_0^t \sigma(s, \omega) dW_s = \lim_{n \rightarrow \infty} \sum_{k=1}^N \sigma(t_{n_{k-1}}, \omega) [W(t_{n_k}) - W(t_{n_{k-1}})].$$

But at this point we have gone too far ahead of ourselves. We need to further characterize the random process  $\sigma(t, \omega)$  in order to make further sense of this stochastic integral.

First of all, the integral itself is a random variable that is “well-behaved” in the sense that

$$\mathbf{E} \left[ \int_0^t |\sigma(s, \omega)|^2 ds \right] = \int_{\Omega} \left( \int_0^t |\sigma(s, \omega)|^2 ds \right) dP < \infty.$$

The reason why  $\sigma$  is now squared and our integral is now with respect to time will be made clear in an upcoming theorem. In any event, maintaining the above equation implies that

$$P \left[ \left\{ \omega \in \Omega : \int_0^t |\sigma(s, \omega)|^2 ds = \infty \right\} \right] = 0.$$

This means with probability 1, the function  $\sigma(t, \omega)$  is in  $L^2([0, T], \mathcal{B}, \mu)$ , where  $\mathcal{B}$  denotes the  $\sigma$ -algebra generated by the Borel sets of  $[0, T]$  and  $\mu$  the corresponding Borel measure. Now if we are to use the definition of integrability to help us make sense of the stochastic integral, we will need to come up with a way to approximate these random processes with something like the simple functions we saw in real analysis.

When we said a real-valued nonnegative function  $f : \mathbb{R} \rightarrow \mathbb{R}$  was in  $L^1$ , this meant that we could construct a sequence of simple functions that point-wise monotonically converged to  $f$ . This consisted in partitioning the range of  $f$  into intervals of the form  $A_1, A_2, \dots, A_n$ , and defining a simple function approximation of  $f$  by

$$\phi_n = \sum_{i=1}^n \min(A_i) 1_{f^{-1}(A_i)}$$

so that

$$\int \phi_n d\mu = \sum_{i=1}^n \min(A_i) \mu(f^{-1}(A_i)).$$

Finer partitions of the range give rise to better approximations, and as  $n$  increases, the punchline is that

$$\int \phi_n d\mu \rightarrow \int f d\mu.$$

Importantly, it could also be shown that we may reach this same result from simple functions of the form  $\phi_n = \sum_i^n \alpha_i 1_{I_i}$  where  $I_i$  is an interval!

Can we apply this same sort of reasoning to the random variable  $\sigma(t, \omega)$ ? The only caveat is that this is now a function of two variables, so we will need to approximate  $\sigma(t, \omega)$  with simple processes.

Definition: Given a partition  $\mathcal{P} = \{t_0, t_1, \dots, t_n\}$  of  $[0, T]$ , a simple process is a function of the form

$$\phi(t, \omega) = \sum_{i=0}^{n-1} \alpha_i(t_i, \omega) 1_{[t_i, t_{i+1})},$$

where  $\alpha_i(t, \omega) \in \mathcal{F}_{t_i}^W$

Now we are in a position to formulate how we can understand the process  $\sigma(t, \omega)$  through simple processes.

We will define the set  $L^2(\Omega, \mathcal{F}_T^W, P \times \mu)$  to consist of all processes  $\sigma(t, \omega)$  such that

1.  $\forall s \in [0, T], \sigma(s, \omega) \in \mathcal{F}_s^W,$
2.  $\mathbf{E} \left[ \int_0^t |\sigma(s, \omega)|^2 ds \right] < \infty.$

For ease of notation, we will refer to this set as  $\mathcal{L}$ .

Theorem: Let  $\sigma(t, \omega)$  denote a process in  $\mathcal{L}$ . Then there exists a sequence of simple processes  $\{\phi_n(t, \omega)\}_{n=1}^\infty$  in  $\mathcal{L}$  such that

$$\mathbf{E} \left[ \int_0^t |\sigma(s, \omega) - \phi_n(s, \omega)|^2 ds \right] \rightarrow 0.$$

Proof: We will prove the case in which  $\sigma(t, \omega)$  is uniformly bounded and Riemann integrable. For the full proof, see Øksendal [1].

Let  $\{\mathcal{P}_i\}_{i=1}^\infty$  denote a sequence of partitions such that  $\mathcal{P}_i$  uniformly divides  $[0, T]$  into  $2^i$  subintervals of the form  $I_{i_1}, I_{i_2}, \dots, I_{i_{2^i}}$ , where  $I_{i_k} = [a_{i_k}, b_{i_k}]$  and we will denote the simple process  $\phi_i$  corresponding to this partition by

$$\phi_i(t, \omega) = \sum_{k=1}^{2^i} \sigma(a_{i_k}, \omega) 1_{I_{i_k}}.$$

Certainly,  $\phi_i(t, \omega)$  satisfies properties (1) and (2). Now for any  $\omega \in \Omega$ ,  $\phi_n(t, \omega) \rightarrow \sigma(t, \omega)$ , almost everywhere because the set of discontinuities of  $\sigma(s, \omega)$  is of measure zero. Therefore,

$$\lim_{n \rightarrow \infty} \int_0^T |\sigma(t, \omega) - \phi_n(t, \omega)|^2 dt \rightarrow 0.$$

Since  $\sigma(t, \omega) \leq M$  for each  $\omega \in \Omega$ , it follows that

$$\lim_{n \rightarrow \infty} \mathbf{E} \left[ \int_0^T |\sigma(t, \omega) - \phi_n(t, \omega)|^2 dt \right] = 0.$$

□

We are almost in a position to understand the stochastic integral

$$\int_0^t \sigma(s, \omega) dW_s$$

through the simple processes defined above. We will now draw a connection between time integrals and stochastic integrals.

Theorem: Itô's Isometry. Let  $\sigma(t, \omega)$  denote a process in  $\mathcal{L}$ . Then for any  $t \in [0, T]$ ,

$$\mathbf{E} \left[ \left( \int_0^t \sigma(s, \omega) dW_s \right)^2 \right] = \mathbf{E} \left[ \int_0^t |\sigma(s, \omega)|^2 ds \right].$$

The reason why this theorem is termed an isometry is that both sides of this equality are norms: the right side denotes the norm of  $\sigma$  in the space  $\mathcal{L}$  while the left side denotes the norm of  $\int_0^t \sigma(s, \omega) dW_s$  over the probability space  $L^2(\Omega, \mathcal{F}_T^W, P)$ .

Proof: Appealing to the previous theorem, it is sufficient to prove this isometry for the case of simple processes.

Let  $\phi$  be a simple process defined on the partition  $\mathcal{P} = \{I_1, \dots, I_n\}$ , where  $I_i = [a_i, b_i)$ , so that

$$\phi(t, \omega) = \sum_{i=1}^n c_i(a_i, \omega) 1_{I_i},$$

where  $c_i \in \mathcal{F}_{a_i}^W$ . It follows that

$$\int_0^t \phi(s, \omega) dW_s = \sum_{i=1}^n c_i(a_i, \omega) (W_{b_i} - W_{a_i}).$$

Therefore,

$$\mathbf{E} \left[ \left( \int_0^t \phi(s, \omega) dW_s \right)^2 \right] = \mathbf{E} \left[ \left( \sum_{i=1}^n c_i(a_i, \omega) (W_{b_i} - W_{a_i}) \right)^2 \right].$$

The independence of Wiener increments implies that for any  $i \neq j$ ,

$$\mathbf{E}[(W_{b_i} - W_{a_i})(W_{b_j} - W_{a_j})] = \mathbf{E}[(W_{b_i} - W_{a_i})] \mathbf{E}[(W_{b_j} - W_{a_j})] = 0.$$

In addition, since  $c_i \in \mathcal{F}_{a_i}^W$ , this means that  $c_i(a_i, \omega)$  and  $(W_{b_i} - W_{a_i})$  are independent events. Moreover, the difference

$$(W_{b_j} - W_{a_j})$$

is normally distributed with mean 0, which means

$$\mathbf{E}[(W_{b_j} - W_{a_j})^2] = \text{Var}[(W_{b_j} - W_{a_j})] = (b_j - a_j).$$

Hence

$$\mathbf{E} \left[ \left( \sum_{i=1}^n c_i(a_i, \omega)(W_{b_i} - W_{a_i}) \right)^2 \right] = \mathbf{E} \left[ \sum_{i=1}^n c_i(a_i, \omega)^2 (b_i - a_i) \right] = \mathbf{E} \left[ \int_0^t |\phi(s, \omega)|^2 ds \right].$$

Since this holds for simple processes, so also must it hold for any  $\sigma(t, \omega)$  in  $\mathcal{L}$ . To see this, let  $\{\phi_n\}_{n=1}^\infty$  denote a sequence of simple processes closing in on  $\sigma(t, \omega)$ , then the fact that Itô's Isometry preserves distances implies that the sequence

$$\mathbf{E} \left[ \left( \int_0^t \phi_n(s, \omega) dW_s \right)^2 \right]$$

is Cauchy.

With Itô's Isometry in hand, we are now finally able to characterize stochastic integrals in terms of simple processes. The first theorem of this section implies that for the arbitrary process  $\sigma(t, \omega)$  in  $\mathcal{L}$ , there exists a sequence of simple processes  $\phi_n(t, \omega)$  in  $\mathcal{L}$  such that

$$\lim_{n \rightarrow \infty} \mathbf{E} \left[ \int_0^t |\sigma(s, \omega) - \phi_n(s, \omega)|^2 ds \right] = 0.$$

Itô's Isometry therefore tells us that

$$\lim_{n \rightarrow \infty} \mathbf{E} \left[ \left( \int_0^t (\sigma(s, \omega) - \phi_n(s, \omega)) dW_s \right)^2 \right] = 0.$$

Now mean-square convergence implies that there is a subsequence of simple processes  $\phi_{n_k}(t, \omega)$  for which

$$\lim_{k \rightarrow \infty} \int_0^t \phi_{n_k}(s, \omega) dW_s = \int_0^t \sigma(s, \omega) dW_s.$$

This allows us to understand a stochastic integral as a limit of Stieltjes-type integrals involving simple processes. Hence,

$$\int_0^t \sigma(s, \omega) dW_s = \lim_{n \rightarrow \infty} \sum_{i=0}^n \sigma(t_i, \omega) (W_{i+1} - W_i).$$

Now that we understand how to characterize integration involving random processes through estimation by simple processes, we can unravel more basic properties of the stochastic integral, such as its expected value.

Lemma: Let  $f \in \mathcal{L}$ . Then for the stochastic integral  $\int_0^t f(s, \omega) dW_s$ , we have

$$\mathbf{E} \left[ \int_0^t f(s, \omega) dW_s \right] = 0.$$

Proof: Let's first show that this is the case for simple processes. If  $\phi(t, \omega)$  denotes such a process, then for some partition  $\mathcal{P}_n$

$$\int_0^t \phi(s, \omega) dW_s = \sum_{\mathcal{P}_n} \phi(t_k, \omega) (W_{t_{k+1}} - W_{t_k}).$$

Now

$$\mathbf{E} \left[ \int_0^t \phi(s, \omega) dW_s \right] = \mathbf{E} \left[ \sum_{\mathcal{P}_n} \phi(t_k, \omega) (W_{t_{k+1}} - W_{t_k}) \right].$$

Since the Wiener increment  $(W_{t_{k+1}} - W_{t_k})$  is independent of  $\phi(t_k, \omega)$ , the expected value of this integral is zero.

Now to prove the case for a general process  $f$  in  $\mathcal{L}$ , it is tempting to simply approximate a  $f$  by a sequence of simple functions  $\{\phi_n\}_{i=1}^\infty$ , so that

$$\mathbf{E} \left[ \int_0^t f(s, \omega) dW_s \right] = \lim_{n \rightarrow \infty} \mathbf{E} \left[ \int_0^t \phi_n(s, \omega) dW_s \right] = 0.$$

But we may only conclude that this is the case if we can pass the limit through the expected value! Luckily, Itô's Isometry tells us that

$$\lim_{n \rightarrow \infty} \mathbf{E} \left[ \left( \int_0^t (f(s, \omega) - \phi_n(s, \omega)) dW_s \right)^2 \right] = 0.$$

Since the expected value is just an integral, applying the Cauchy-Schwartz inequality yields, for all  $n$ ,

$$\mathbf{E} \left[ \left( \int_0^t (f(s, \omega) - \phi_n(s, \omega)) dW_s \right)^2 \right] \geq \left( \mathbf{E} \left[ \int_0^t (f(s, \omega) - \phi_n(s, \omega)) dW_s \right] \right)^2,$$



whence

$$\lim_{n \rightarrow \infty} \mathbf{E} \left[ \int_0^t (f(s, \omega) - \phi_n(s, \omega)) dW_s \right] = 0.$$

Therefore,

$$\mathbf{E} \left[ \int_0^t f(s, \omega) dW_s \right] = 0.$$

□

The reason for this lemma is that it offers a connection between a martingale and a stochastic integral. If we take a stochastic integral of the form, say,

$$\int_0^t g(t, \omega) dW_s,$$

where  $g(t, \omega)$  is in  $\mathcal{L}$ , then for any  $s \leq t$

$$\mathbf{E} \left[ \int_0^t g(u, \omega) dW_u \mid \mathcal{F}_s^W \right] = \int_0^s g(u, \omega) dW_u + \mathbf{E} \left[ \int_s^t g(u, \omega) dW_u \mid \mathcal{F}_s^W \right].$$

However, a previous lemma implies that the right expectation is equal to zero. Therefore,

$$\mathbf{E} \left[ \int_0^t g(u, \omega) dW_u \mid \mathcal{F}_s^W \right] = \int_0^s g(u, \omega) dW_u.$$

This shows us that every stochastic integral is a martingale! This important result will be used frequently in the upcoming sections. Towards the end of this section, we will also prove the powerful result that the converse is also true, which means that every martingale is a stochastic integral.

### 3.4 The Itô Formula

Now that we have given an account of integration involving random processes, how can we proceed in the opposite direction and tackle differentiation of functions of random processes?

Consider a stochastic process  $X(t, \omega)$  given by

$$X(t, \omega) = x_0 + \int_0^t \mu(s, \omega) ds + \int_0^t \sigma(s, \omega) dW_s,$$

where  $x_0$  is some real number and  $\mu, \sigma$  are adapted processes. A shorthand expression for the above equation is reminiscent of a differential equation:

$$\begin{aligned} dX(t, \omega) &= \mu(t, \omega) dt + \sigma(t, \omega) dW_t, \\ X(0, \omega) &= x_0 \end{aligned}$$

In this case we say that the stochastic differential of  $X$  is given by the first equation above, the second above equation denotes an initial condition.

Our next goal is to find the stochastic differential of a function  $f(t, X(t, \omega)) \in C^{1,2}$ , where the dynamics of the random process  $X(t, \omega)$  are given by

$$dX(t, \omega) = \mu(t, \omega) dt + \sigma(t, \omega) dW_t,$$

where  $\mu$  and  $\sigma$  are adapted processes.

Theorem: Itô's Formula. Let  $f(t, X(t, \omega))$  be a function in  $C^{1,2}$ , and dynamics of  $X(t, \omega)$  be given by the previous equation. The stochastic differential  $df(t, \omega)$  is given by

$$df(t, \omega) = \left( \frac{\partial f}{\partial t}(t, \omega) + \mu(t, \omega) \frac{\partial f}{\partial x}(t, \omega) + \frac{1}{2} (\sigma(t, \omega))^2 \frac{\partial^2 f}{\partial x^2} \right) dt + \sigma(t, \omega) \frac{\partial f}{\partial x}(t, \omega) dW_t.$$

Proof: We will prove the case when  $\mu(t, \omega)$  is Riemann integrable with respect to  $t$ , and  $\sigma(t, \omega)$  exhibits continuous trajectories. For the proof of the general case, see for example, Steele [2].

For ease of notation, we will denote expressions of the form  $\mu(s, \omega) \frac{\partial f}{\partial t}(s, \omega)$  by  $\mu \frac{\partial f}{\partial s}$ , and  $X(t, \omega)$  by  $X(t)$ .

Now it is sufficient to show that with probability 1, the following equation holds:

$$f(T, X(T, \omega)) - f(0, X(0, \omega)) = \int_0^T \left( \frac{\partial f}{\partial t} + \mu \frac{\partial f}{\partial x} + \frac{1}{2} \sigma^2 \frac{\partial^2 f}{\partial x^2} \right) dt + \int_0^T \sigma \frac{\partial f}{\partial x} dW_t.$$

Let  $\mathcal{P}_n = \{0 = t_1, t_2, t_3, \dots, t_{2^n} = T\}$  denote a uniform partition of the interval  $[0, T]$  into  $2^n$  subintervals of length  $\Delta t$ . We will carve  $f$  into increments along this partition. Define  $\Delta X_k$  and  $\Delta f_k$  by

$$\begin{aligned}\Delta X_k &= X(t_{k+1}) - X(t_k), \\ \Delta f_k &= f_t(t_k, X_k)\Delta t + f_x(t_k, X_k)(\Delta X_k) + \frac{1}{2}f_{xx}(t_k, X_k)(\Delta X_k)^2.\end{aligned}$$

Since  $f(t, X(t, \omega))$  is  $C^{1,2}$ , Taylor's Theorem implies

$$\lim_{n \rightarrow \infty} \sum_{\mathcal{P}_n} \Delta f_k = f(t, X(t)) - f(0, X(0)).$$

However, there is so much information in this sum that we have yet to unpack. Since

$$X(t) = \int_0^t \mu ds + \int_0^t \sigma dW_s,$$

then

$$\Delta X_k = \int_{t_k}^{t_{k+1}} \mu ds + \int_{t_k}^{t_{k+1}} \sigma dW_s.$$

The Riemann integrability of  $\mu$  and  $\sigma$  implies that

$$\Delta X_k = \mu \Delta t + \sigma \Delta W_k + R_k,$$

where  $R_k$  is a remainder such that, as  $n$  approaches infinity,

$$\sum_{\mathcal{P}_n} R_k \rightarrow 0$$

Given this definition of  $\Delta X_k$ , we may unpack  $\Delta f_k$  as

$$\Delta f_k = A_k + B_k + C_k + D_k + E_k,$$

where

$$A_k = \{f_t(t_k, X(t_k)) + \mu f_x(t_k, X(t_k))\} \Delta t$$

$$B_k = \sigma f_x(t_k, X(t_k)) \Delta W_k$$

$$C_k = \mu^2 f_{xx}(t_k, X(t_k)) (\Delta t)^2$$

$$D_k = \sigma f_x(t_k, X_k) (\Delta W_k)^2$$

$$E_k = \mu \sigma f_{xx}(t_k, X(t_k)) \Delta t \Delta W_k.$$

The idea is to examine how the sum of each term on the left-hand side of each of these equations behaves as we let  $n$  approach infinity. Each term in the sum will give us a fragment of the Itô formula.

Immediately, we see that

$$\lim_{n \rightarrow \infty} \sum_{\mathcal{P}_n} A_k = \int_0^t (f_t(s, X(s)) + \mu f_x(s, X(s))) ds$$

$$\lim_{n \rightarrow \infty} \sum_{\mathcal{P}_n} B_k = \int_0^t \sigma f_x(s, X(s)) dW_s$$

$$\lim_{n \rightarrow \infty} \sum_{\mathcal{P}_n} C_k = \lim_{n \rightarrow \infty} \sum_{\mathcal{P}_n} E_k = 0$$

The final piece of the puzzle is  $D_k$ . We need understand how  $(\Delta W_s)^2$  behaves as we refine our underlying partition of  $[0, T]$ .

Definition: The quadratic variation of a process  $X$  on the interval  $[0, T]$ , denoted  $\langle X \rangle$ , is defined by

$$\lim_{\|\mathcal{P}\| \rightarrow 0} \sum_k (X(t_{k+1}) - X(t_k))^2$$

Where  $\|\mathcal{P}\|$  denotes the mesh of any partition of  $[0, T]$ .

Interestingly, the quadratic variation of a random variable may itself be deterministic. The variation of a Wiener process over a finite interval  $[0, t]$  is one such case.

Lemma: The quadratic variation of a Wiener process  $W$  over the interval  $[0, t]$  equals  $t$ .

Proof: Let  $\{\mathcal{P}_i\}_{i=1}^{\infty}$  denote a sequence of partitions with mesh tending to zero.

Denote  $Q_n$  by

$$Q_n = \sum_{\mathcal{P}_n} (W(t_{k+1}) - W(t_k))^2.$$

Then the properties of expectation and the normal distribution imply that

$$\mathbf{E}[Q_n] = \sum_{\mathcal{P}_n} (t_{k+1} - t_k) = t \text{ for any } n.$$

Now let's examine the variance of  $Q_n$ . Since the increments of the sum  $Q_n$  are independent, the variance is linear. That is,

$$\text{Var}[Q_n] = \sum_{\mathcal{P}_n} \text{Var}[W(t_{k+1}) - W(t_k)]^2.$$

By appealing the properties of the normal distribution once more,

$$\text{Var}[W(t_{k+1}) - W(t_k)]^2 = 2(t_{k+1} - t_k)^2.$$

Therefore, as  $n$  tends to infinity,  $\text{Var}[Q_n] \rightarrow 0$ . This means that as  $n$  increases,  $Q_n$  becomes more and more deterministic! Since its expected value was shown to tend to  $t$ , it follows with probability 1 that

$$\lim_{n \rightarrow \infty} Q_n = \langle W \rangle = t.$$

□

Frequently, textbooks on financial mathematics will use this fact to “justify” the rule

$$(dW_t)^2 = dt.$$

However, such justification is purely motivational. To fully tackle the final summand  $D_k$ , we let  $\varepsilon > 0$ , fix  $\omega \in \Omega$ , and approximate the continuous function  $\sigma f_x(t, X(t))$  over the (sufficiently fine) uniform partition  $\mathcal{P}_n$  so that

$$\sum_{\mathcal{P}_n} \sigma f_x(t_k, X(t_k)) 1_{[t_k, t_{k+1})} - \varepsilon \leq \sigma f_x(t, X(t)) \leq \sum_{\mathcal{P}_n} \sigma f_x(t_k, X(t_k)) 1_{[t_k, t_{k+1})} + \varepsilon$$

for all  $t \in [0, T]$ . Denoting  $\sigma f_x(t_k, X(t_k)) 1_{[t_k, t_{k+1})}$  by  $\phi_k$ , and  $\sum \sigma f_x(t_k, X(t_k)) 1_{[t_k, t_{k+1})}$  by  $\phi$ , it follows that

$$\sum_{\mathcal{P}_n} (\phi_k - \varepsilon) (\Delta W_k)^2 \leq \sum_{\mathcal{P}_n} \sigma f_x(t, X(t)) (\Delta W_k)^2 \leq \sum_{\mathcal{P}_n} (\phi_k + \varepsilon) (\Delta W_k)^2.$$

Now let's examine what happens when we refine the underlying partition  $\mathcal{P}_n$  without changing  $\phi$ . If we zoom in on a subinterval  $[t_i, t_{i+1})$  on which  $\phi_k$  is constant, and denote the part of the partition underlying this subinterval by  $\mathcal{I}_i$ , then continually refining  $\mathcal{I}_i$  amounts to calculating the limit

$$\lim_{n \rightarrow \infty} C_i \sum_{\mathcal{I}_n} (\Delta W_k)^2$$

for some constant  $C_i$ . However, for a fixed  $\omega \in \Omega$ , a previous lemma allows us to conclude that this sum converges to the length  $(t_{i+1} - t_i)$  with probability 1. Since this result holds for all subintervals on which  $\phi$  is constant, it follows that

$$\lim_{n \rightarrow \infty} \sum_{\mathcal{P}_n} (\phi_k \pm \varepsilon) (\Delta W_k)^2 = \sum_k (\phi_k \pm \varepsilon) (t_{k+1} - t_k) = \int_0^T (\phi \pm \varepsilon) dt.$$

Therefore,

$$\int_0^T \phi dt \approx_{T\varepsilon} \lim_{n \rightarrow \infty} \sum_{\mathcal{P}_n} \sigma f_x(t, X(t)) (\Delta W_k)^2.$$

However, it is also true that inequality above implies

$$\int_0^T \phi dt \approx_{T\varepsilon} \int_0^T \sigma f_x(t, X(t)) dt.$$

The arbitrariness of  $\varepsilon$  therefore implies that with probability 1,

$$\lim_{n \rightarrow \infty} \sum_{\mathcal{P}_n} D_k = \lim_{n \rightarrow \infty} \sum_{\mathcal{P}_n} \sigma f_x(t_k, X_k) (\Delta W_k)^2 = \int_0^T \sigma f_x(t, X(t)) dt.$$

This gives us the final fragment of the Itô formula.

□

### 3.5 Some examples using Itô's Formula

Let's see this formula in action! Consider the classic integral

$$\int_0^T W(t) dW_t$$

If we can think of a differentiable function  $f(t, X(t))$  such that  $f_x = W(t)$ , then we would be golden. Consider

$$f(t, W(t)) = W(t)^2.$$

Applying Itô's formula with  $\mu = 0$ ,  $\sigma = 1$  yields

$$df = dt + 2W(t)dW_t,$$

whence

$$W(t)^2 = t + 2 \int_0^T W(t) dW_t, \text{ and}$$

$$\int_0^T W(t) dW_t = \frac{W(t)^2}{2} - \frac{t}{2}.$$

So working with these sorts of stochastic processes gives us a different result than if we went about this using techniques from ordinary calculus.

One of the building blocks of modelling the trajectory of an asset is given by the stochastic differential equation

$$dX(t) = \mu X(t) dt + \sigma X(t) dW_t$$

$$X(0) = x_0.$$

Where  $\mu$ ,  $\sigma$  are constants. Insofar as  $X(t)$  is positive, this stochastic differential equation describes Geometrical Brownian Motion.

To solve for  $X(t)$ , we will apply Itô's formula to the stochastic variable  $Z(t) = \ln(X(t))$  to obtain

$$dZ(t) = \left( \mu \frac{1}{X(t)} X(t) - \frac{1}{2} \sigma^2 \frac{1}{X(t)^2} X(t)^2 \right) dt + \sigma \frac{1}{X(t)} X(t) dW_t.$$

This cleans up nicely to the stochastic differential equation

$$dZ(t) = \left( \mu - \frac{1}{2} \sigma^2 \right) dt + \sigma dW_t$$

$$Z_0 = \ln(x_0).$$

Since this is just shorthand for

$$Z(t) = \ln(x_0) + \left( \mu - \frac{1}{2} \sigma^2 \right) t + \sigma W(t),$$

this tells us that the variable  $Z(t) = \ln(X(t))$  is normally distributed with mean  $(\mu - \frac{1}{2} \sigma^2)t$  and variance  $\sigma^2 t$ . In this case, we say that the random process  $X(t)$  is lognormally distributed.

Solving for  $X(t)$ , we obtain

$$X(t) = x_0 e^{(\mu - \frac{1}{2} \sigma^2)t + \sigma W(t)}.$$

Now what is the expected value of this process? From the outset, we know that

$$\mathbf{E}[X(t)] = x_0 e^{(\mu - \frac{1}{2} \sigma^2)t} \mathbf{E}[e^{\sigma W(t)}].$$

The question then reduces down to finding  $\mathbf{E}[e^{\sigma W(t)}]$ . With the aim of using Itô's formula once more, let  $Y(t) = e^{\sigma W(t)}$ . Then

$$dY(t) = \frac{1}{2} \sigma^2 e^{\sigma W(t)} dt + \sigma e^{\sigma W(t)} dW_t,$$

which we may write as

$$Y(t) = \frac{\sigma^2}{2} \int_0^t Y(s) ds + \sigma \int_0^t e^{\sigma W(s)} dW_t.$$



Appealing to a previous lemma, we may take the expected value of both sides and apply Fubini's Theorem to obtain

$$\mathbf{E}[Y(t)] = \frac{\sigma^2}{2} \int_0^t \mathbf{E}[Y(s)] ds.$$

Notice that taking the expected value has reduced this problem into an ordinary differential equation! Using standard methods, we obtain

$$\mathbf{E}[Y(t)] = e^{\frac{\sigma^2}{2}t}.$$

Therefore,

$$\mathbf{E}[X(t)] = x_0 e^{\mu t}.$$

Many pricing models in continuous time are built upon the assumption that the price of a stock follows Geometrical Brownian Motion, including Black's formula for a European call option.

The main reason why we have developed the Itô calculus consists in allowing us to characterize martingales in continuous time. This powerful result, known as the Martingale Representation Theorem, states that every martingale in  $\mathcal{F}_t^W$  may be understood as a stochastic integral!

### 3.6 The Martingale Representation Theorem

**Theorem: The Martingale Representation Theorem.** Consider a  $d$ -dimensional Wiener process  $W$  whose value at time  $t$  is given by  $\{W_1(t), W_2(t), \dots, W_d(t)\}$ , and let  $M$  be any  $\mathcal{F}_t^W$ -adapted martingale, where  $t \in [0, T]$ . Then there exists uniquely determined processes  $h_1, \dots, h_d$  in  $\mathcal{F}_t^W$  for which  $M$  has the representation

$$M(t) = M(0) + \sum_{i=1}^d \int_0^t h_i(u) dW_i(u), t \in [0, T].$$

This result will demand us to roll out a fair amount of fancy machinery that will grant us insight about the general behavior of martingales. However, instead of unpacking claim after claim without any clear-cut motivation, we will first pursue a more basic result from which The Martingale Representation Theorem is an immediate corollary.

Theorem: Any Wiener functional  $X$  that is adapted to the filtration  $\mathcal{F}_T^W$ , where  $W$  denotes a  $d$ -dimensional Wiener process, admits of the representation

$$X(T) = \mathbf{E}[X] + \sum_{i=1}^d \int_0^T h_i(u) dW_i(u)$$

for uniquely determined adapted processes  $h_1, \dots, h_d$ .

Proof: We'll consider the case in which the Wiener process is one-dimensional; the multidimensional case follows by similar reasoning.

The idea for this proof is to consider a variant of a simple stochastic variable we have already encountered: that of geometric Brownian motion. After we have proved that this variable admits of this representation, we will use Fourier techniques to deduce that all Wiener functionals may be represented in this light.

Now, consider the type of Geometric Brownian motion given by the dynamics

$$dX(t) = \sigma X(t) dW(t)$$

$$X_0 = 1,$$

where  $\sigma$  is constant. We know that the above expression is just shorthand for

$$X(t) = 1 + \int_0^t \sigma X(u) dW(u).$$

From the example in the previous section, we know that  $X(t)$  is of the form

$$X(t) = e^{-\frac{1}{2}\sigma^2 t + \sigma W(t)}.$$

Using this form and the integral from above, some arithmetic gives us

$$e^{\sigma W(t)} = e^{\frac{1}{2}\sigma^2 t} + \sigma \int_0^t e^{-\frac{1}{2}\sigma^2(u-t) + \sigma W(u)} dW(u).$$

If we repeated this process for the stochastic variable  $X(t) - X(s)$ , where  $s \leq t$ , then we will end up with

$$e^{\sigma(W(t)-W(s))} = e^{\frac{1}{2}\sigma^2(t-s)} + \sigma \int_s^t e^{-\frac{1}{2}\sigma^2(u-t+s) + \sigma W(u)} dW(u).$$

What this tells us is that we may understand any process of the form

$$Y = e^{\sigma(W(t)-W(s))},$$

whose expected value  $\mathbf{E}[Y]$  we have seen to be equal to  $e^{\frac{1}{2}\sigma^2(t-s)}$ , as represented by

$$Y = \mathbf{E}[Y] + \int_0^T h(u) dW(u),$$

where  $h$  is a process with compact support  $[s, t]$ . It must be true that  $h$  is a uniquely determined process, for if  $g$  were an adapted process such that

$$\int_0^T (g(u) - h(u)) dW(u) = 0 \text{ } P\text{-almost surely,}$$

then an application of Itô's Isometry yields

$$\mathbf{E} \left[ \int_0^T (g(u) - h(u)) dW(u) \right]^2 = \mathbf{E} \left[ \int_0^T (g(u) - h(u))^2 du \right] = \int_0^T \mathbf{E}[(g(u) - h(u))^2] du = 0.$$

The positive integrand implies that  $g = h$  almost everywhere which respect to the measure  $P \times \mu$ . Therefore, the process  $h$  is uniquely determined.

Since we are able to uniquely represent random variables of the same form as  $Y$ , let's use these variables as building blocks for other variables of the form

$$Z = \prod_{i=1}^n e^{\sigma_i[W(T_i) - W(T_{i-1})]}$$

where  $0 \leq t_0 \leq t_1 \leq \dots \leq t_n \leq T$ , and  $\sigma_i \in \mathbf{R}$ . As a first attempt to find the representation for this variable, we will multiply the representations of its building blocks:

$$\prod_{i=1}^n e^{\sigma_i[W(T_i) - W(T_{i-1})]} = \prod_{i=1}^n Z_i$$

where

$$Z_i = \mathbf{E}[Z_i] + \int_0^T h_i(u) dW(u),$$

$h_i$  has compact support  $[t_{i-1}, t_i]$ .

To get an idea of how this product will appear, we need only to consider the case  $Z_i Z_j$  where  $i \neq j$ . To see how  $Z_i(t)Z_j(t)$  is represented, we will apply the Itô formula:

$$d(Z_i Z_j) = Z_i(dZ_j) + (dZ_i)Z_j + (dZ_i)(dZ_j).$$

Since  $dZ_i(t) = h_i dW(t)$ ,  $dZ_j(t) = h_j dW(t)$ , and  $Z_i, Z_j$  are independent, we have

$$Z_i(t)Z_j(t) = \mathbf{E}[Z_i]\mathbf{E}[Z_j] + \int_0^t (Z_i h_i(u) + Z_j h_j(u)) dW(u).$$

Therefore, the variable  $Z$  defined above may be represented by

$$Z = \mathbf{E}[Z] + \int_0^T h(u) dW(u).$$

In order to prove that every Wiener functional admits of a representation of this form, we should prove as a preliminary that the set of all functionals that have such a representation is closed in  $L^2(\Omega, \mathcal{F}_T^W, P)$  (hereafter denoted  $L^2$ )

Theorem: Suppose that each member of the sequence  $\{X_n\}_{n=1}^\infty \subseteq L^2$  is such that

$$X_n = \mathbf{E}[X_n] + \int_0^T h_n(t) dW(t)$$

where  $h_n \in L^2$ . If  $X_n \rightarrow X$  in  $L^2$ , then there exists a process  $h$  in  $L^2$  such that  $h_n \rightarrow h$  in  $L^2$  and

$$X = \mathbf{E}[X] + \int_0^T h(t) dW(t).$$

Proof: Since  $\mathbf{E}[(X - X_n)^2] \rightarrow 0$  almost surely with respect to  $P$ , it follows that

$$\mathbf{E}[X_n] \rightarrow \mathbf{E}[X],$$

which means that the sequence

$$\{X_n - \mathbf{E}[X_n]\}_{n=1}^{\infty}$$

is Cauchy in  $L^2$ . Therefore, the sequence of integrals  $\left\{ \int_0^T h_n dW(t) \right\}_{n=1}^{\infty}$  is also Cauchy in  $L^2$ , which means that

$$\lim_{n \rightarrow \infty} \mathbf{E} \left[ \left( \int_0^T h_n(t) dW(t) \right)^2 \right] = \lim_{n \rightarrow \infty} \mathbf{E} \left[ \int_0^T (h_n(t))^2 dt \right].$$

This implies that  $h_n \rightarrow h$  for some process  $h$ . It only takes a bit of arithmetic to show that

$$\lim_{n \rightarrow \infty} \mathbf{E} \left[ \left\{ \left( \mathbf{E}[X_n] + \int_0^T h_n(t) dW(t) \right) - \left( \mathbf{E}[X] + \int_0^T h(t) dW(t) \right) \right\}^2 \right] = 0.$$

$$\text{Therefore, } X = \mathbf{E}[X] + \int_0^T h(t) dW(t). \quad \square$$

Now that we have shown that the set of functions who admit of such a representation are closed, we will now show that this set is also dense in  $L^2$ .

**Theorem:** The set of representable functions is dense in  $L^2$ .

**Proof:** Consider a sequence  $\{t_i\}_{i=0}^{\infty}$  whose elements form a dense subset of  $[0, T]$ , where  $t_0 = 0$ , and let  $\mathcal{G}_i$  denote the  $\sigma$ -algebra generated by the value of the Wiener process at points  $t_1, t_2, \dots, t_n$ , that is,

$$\mathcal{G}_i = \sigma(W(t_1), W(t_2), \dots, W(t_n)).$$

Immediately  $\mathcal{G}_{i-1} \subseteq \mathcal{G}_i$  for every natural number  $i$ , so  $\mathcal{G}_{i=1}^{\infty}$  forms a filtration. Since  $\{t_i\}_{i=0}^{\infty}$  is dense in  $[0, T]$ , then the information about a given Wiener trajectory contained in  $\mathcal{G}_{\infty}$  is “essentially” the same as that contained in  $\mathcal{F}_T^W$  because a Wiener trajectory is continuous with probability 1. In other words, for any  $A$  in  $\mathcal{F}_T^W$ , there exists a  $B$  in  $\mathcal{G}_{\infty}$  such that  $A \subseteq B$  and  $P[A] = P[B]$ .

Now given  $g \in L^2$ , the Martingale Convergence Theorem implies that the limit

$$\lim_{n \rightarrow \infty} \mathbf{E}[g | \mathcal{G}_n] = \mathbf{E}[g | \mathcal{G}_{\infty}] = g$$

holds  $P$ -almost surely. Therefore, we have a sequence of martingales that converges in  $L^2$  to  $g$ . However, for any natural number  $n$ , the conditional expectation  $\mathbf{E}[g | \mathcal{G}_n]$  is a  $\mathcal{G}_n$ -measurable function of the form  $g_n : \mathbb{R}^n \rightarrow \mathbb{R}$ . If  $\mathcal{B}_n$  denotes the measurable functions of  $L^2$  that may be expressed in this form, then the arbitrariness of  $g$  implies that  $\bigcup_{i=1}^{\infty} \mathcal{B}_i$  is dense in  $L^2$ .

Appealing to the dense sequence defined above, let  $\mathcal{S}_n = \text{span} \left\{ \prod_{i=1}^n e^{\sigma_i [W(t_i) - W(t_{i-1})]} \right\}$ . Certainly  $\mathcal{S}_n \subseteq \mathcal{B}_n$ , and if we can show that this span of functions is dense in  $\mathcal{B}_n$ , then we may conclude that this span is dense in  $L^2$ .

A helpful fact about Hilbert spaces and orthogonal projections is given by the following lemma: Lemma: If  $\mathcal{H}$  denotes a closed subspace of  $L^2$ , and  $K \subseteq \mathcal{H}$ , then

$$\text{If } \mathcal{H} \cap K^\perp = \{0\}, \text{ then } \bar{K} = \mathcal{H}.$$

Proof: See lemma A.3 in the Appendix. □

With this lemma in mind, let  $f \in \mathcal{B}_n$ , and assume that it is orthogonal to every  $g$  in  $\mathcal{S}_n$ , that is,  $\mathbf{E}[fg] = 0$ . It follows that  $f = 0$ , then using the preceding lemma completes the proof.

Let's define a new function  $F : \mathbb{R}^n \rightarrow \mathbb{R}$  by

$$F(\sigma_1, \dots, \sigma_n) = \int_{\Omega} f(W(t_1), \dots, W(t_n)) e^{\sigma_1 W(t_1) + \dots + \sigma_n W(t_n)} dP.$$

Now  $F$  is not a very interesting function since it maps any element of  $\mathbb{R}^n$  to zero, but it is analytic, and we can see this by approximating  $f$  by a sequence of pointwise-converging simple functions  $\{\phi_i\}_{i=1}^k$ . If we denote

$$F_n = \sum_{k=1}^n \phi_{n_k} e^{\sigma_1 W(t_1) + \dots + \sigma_n W(t_n)},$$

then since  $F$  is bounded,  $F_n \rightarrow F$  uniformly. Since  $\frac{\partial F_n}{\partial \sigma_i} = \sigma_i F_n$ , we see that the derivatives of all orders also converge. Therefore,  $F$  is infinitely differentiable and thereby admits of an analytic extension into  $\mathbb{C}^n$ .

The analytic extension of  $F$ , denoted  $\tilde{F}$ , is defined by

$$\tilde{F}(z_1, \dots, z_n) = \int_{\Omega} f(W(t_1), \dots, W(t_n)) e^{z_1 W(t_1) + \dots + z_n W(t_n)} dP,$$

where  $(z_1, \dots, z_n) \in \mathbb{C}^n$ . Since  $F = 0$  on  $\mathbb{R}^n$ ,  $\tilde{F} = 0$  on  $\mathbb{C}^n$ . In particular, if we choose any  $(\sigma_1 i, \dots, \sigma_n i) \in \mathbb{C}^n$ , where  $\sigma_i \in \mathbb{R}$ , then  $F(\sigma_1 i, \dots, \sigma_n i) = 0$ . This means that the Fourier Transform of  $f$  is zero. By the uniqueness of the Fourier transform,  $f = 0$ .

By application of the previous lemma, we see that  $\mathcal{S}_n$  is dense in  $\mathcal{B}_n$ . Since  $\bigcup_{i=1}^{\infty} \mathcal{S}_n$  is dense in  $\bigcup_{i=1}^{\infty} \mathcal{B}_n$ , which is in turn dense in  $L_2$ , it follows that  $\bigcup_{i=1}^{\infty} \mathcal{S}_n$  is dense in  $L_2$ . □

Therefore, any Wiener functional in  $L^2$  admits a representation of the form previously described. □

From this, we immediately obtain the Martingale Representation Theorem, which we restate:

**Theorem:** Let  $W$  denote a  $d$ -dimensional Wiener process, and  $M$  denote an  $\mathcal{F}_t^W$ -adapted martingale. Then there exist uniquely determined  $\mathcal{F}_T^W$ -adapted processes  $h_1, \dots, h_d$  such that  $M$  has the representation

$$M(t) = M(0) + \sum_{i=1}^d \int_0^t h_i(u) dW_i(u)$$

where  $t \in [0, T]$ .

**Proof:** Using a previous theorem, we certainly know that

$$M(T) = M(0) + \sum_{i=1}^d \int_0^T h_i(u) dW_i(u)$$

Hence for  $t \leq T$ ,

$$\mathbf{E}[M(T) | \mathcal{F}_t^W] = M(t) = M(0) + \sum_{i=1}^d \int_0^t h_i(u) dW_i(u).$$

□

We now have all the machinery we need to build our model in continuous time.

## 4 A CONTINUOUS TIME MODEL

### 4.1 Portfolio Dynamics

We continue our quest for this strange probability measure  $Q$  in the context of continuous time. Our market will consist of  $N + 1$  assets of the form  $S_0, S_1, \dots, S_N$ , where  $S_0$  denotes the risk-free asset that we may interchangeably call  $B$ . In addition, we have a  $d$ -dimensional Wiener process  $W(t) = (W_1(t), W_2(t), \dots, W_d(t))$  that drive the random behavior exhibited by the  $N$  risky assets of market. In fact, these risky assets individually appeal to each component of this Wiener process, such that the dynamics of these risky assets assume the general form

$$dS_i(t, \omega) = \alpha_i(t, \omega)S_i(t, \omega)dt + S_i(t, \omega) \sum_{j=1}^d \sigma_{ij}(t, \omega)dW_j(t),$$

where  $\alpha_i$  and  $\sigma_{ij}$  are adapted processes that exhibit continuous trajectories. The dynamics of the risk-free interest rate are given by

$$dB(t) = rB(t)dt,$$

where  $r$  denotes the risk-free interest rate.

As in the previous sections, we will be working with hedging portfolios, but since we are working in the context of continuous time, the value process of such a portfolio  $h$  is a random process  $V_t^h$  with dynamics given by

$$dV^h(t) = \sum_{i=0}^N h_i(t)dS_i(t, \omega),$$

where  $h_i(t)$  denotes the amount of asset  $S_i$  owned at time  $t$ .

The dynamics of this value process tell us that any change exhibited in the value of the portfolio is due to changes in the values of the underlying assets. Since there is no influx of money into this portfolio other than from the assets themselves, we say that this portfolio is self-financing. In addition, we will consider the initial value of such portfolios to be zero,



so that  $V_t^h$  is given by

$$V^h(t) = \int_0^t \sum_{i=0}^N h_i dS_i(t, \omega).$$

This characterizes all of the hedging portfolios that we will deal with in this section.

#### 4.2 The Search for $Q$

Now from the outset, we know that in order for  $Q$  to be of any use, it must assign the same certainty to financially significant events such as the value of a hedging portfolio for a claim  $X$  maturing at  $T$ . In other words, if  $V^h(T)$  denotes a self-financing portfolio for which

$$P[\{V^h(T) - X(T) = 0\}] = 1,$$

then we certainly require that  $Q$  also assigns probability 1 to this same event. A sufficient condition for this is that

$$\forall A \in \mathcal{F}_T^W, P[A] = 1 \rightarrow Q[A] = 1.$$

Since  $P$  and  $Q$  are finite measures, this also holds for events of  $P$  - probability 0. When  $P$  and  $Q$  satisfy these conditions, we say that  $Q$  is absolutely continuous with respect to the measure  $P$ , denoted  $Q \ll P$ .

#### 4.3 The First Fundamental Theorem of Asset Pricing

We will prove that a market on the finite interval  $[0, T]$  is absent of arbitrage (in addition to “almost arbitrage,” which we will discuss), if and only if there exists a probability measure  $Q$  for which not only  $Q \ll P$ , but also  $P \ll Q$ . In this case we say that  $Q$  and  $P$  are equivalent measures.

In order to prove this, however, we will work with a slightly simpler market.

Definition: Consider a market model with assets  $S_0(t, \omega), S_1(t, \omega), \dots, S_N(t, \omega)$ , where  $S_0(t, \omega) = B(t)$ . The normalized market consists of assets of the form  $U_i(t, \omega) = \frac{S_i(t, \omega)}{B(t)}$  for  $i = 0, 1, 2, \dots, N$ .

Comparing the normalized market to the original, we see that important characteristics of portfolio dynamics have not changed.

It can be shown that a portfolio  $h$  is self-financing in the original market if and only if it is self-financing in the normalized market. If  $V_t^{h,O}$  denotes the value process of  $h$  in the original market, while  $V_t^{h,N}$  denotes the value process in the normalized market, then these processes are related by

$$V_t^{h,N} = \frac{1}{B(t)} V_t^{h,O}.$$

This implies that the original market is arbitrage free if and only if the normalized market is arbitrage free.

As we saw in previous sections, this probability measure  $Q$  will also be such that each asset of the normalized market is a  $Q$ -martingale. To make sense of what this actually implies in terms of the dynamics of the normalized asset  $U_i(t, \omega)$ , the Itô formula applied to the normalized asset  $U_i(t, \omega) = \frac{S_i(t, \omega)}{B(t)}$  yields

$$dU_i(t, \omega) = \frac{S_i(t, \omega)(\alpha_i(t, \omega) - r)}{B(t)} dt + \frac{S_i(t, \omega)}{B(t)} \sum_{j=1}^d \sigma_{ij}(t, \omega) dW_j(t).$$

For the normalized asset to be a  $Q$ -martingale, we know from the Martingale Representation Theorem that the dynamics of  $dU_i$  under  $Q$  would exhibit no drift. Effectively, then, we are after an equivalent probability measure  $Q$  under which the dynamics of a stock are such that its drift is the risk-free interest rate  $r$ . Finding this probability measure will culminate in the following remarkable theorem.

**Theorem: First Fundamental Theorem of Asset Pricing.** A market is arbitrage free if and only if there exists an equivalent martingale measure  $Q$  under which each asset of its normalized market  $U_i(t, \omega)$  is a  $Q$ -martingale.

We will first prove that the existence of such a martingale measure  $Q$  implies the absence of arbitrage.

**Proof:** Consider a self-financing portfolio  $h$  in the normalized market whose value process  $V_t^h$  is such that

$$V_0^h = 0$$

$$P[V_T^h \geq 0] = 1.$$

Let's use the definition of an arbitrage portfolio from Section 1 (page 2). To prove that this market is free of arbitrage, it is sufficient to show that  $P[\{V_T^h > 0\}] = 0$ .

If we can show that

$$Q[V_T^h > 0] = 0, \text{ then the result follows by equivalence of } P.$$

At any time  $t$ , the dynamics of the self-financing portfolio  $h$  are given by

$$dV_t^h = \sum_{i=1}^N h_i(t) dU_i(t, \omega).$$

Since the normalized asset  $U_i(t, \omega)$  is a  $Q$ -martingale, the Martingale Representation Theorem implies that

$$dU_i(t, \omega) = \sum_{j=1}^d \theta_j(t, \omega) dW_j^Q(t),$$

where  $\theta_j$  denotes some adapted process and  $W^Q(t)$  is some  $d$ -dimensional process that is a Wiener process according to  $Q$ . By substitution, we see that the process  $V_t^h$  is a stochastic integral, and therefore also a  $Q$  martingale. Hence,

$$\mathbf{E}^Q[V_T^h] = V_0^h = 0.$$

Since  $V_T^h$  is assumed to be nonnegative,

$$\mathbf{E}^Q[V_T^h] \geq \sum_{n=1}^{\infty} \frac{1}{n+1} Q \left[ \left\{ V_T^h \in \left( \frac{1}{n+1}, \frac{1}{n} \right) \right\} \right] + \sum_{n=1}^{\infty} n Q \left[ \left\{ V_T^h \in (n, n+1) \right\} \right].$$

This implies that  $Q[V_T^h > 0] = 0$ . Therefore, the normalized market is arbitrage free. □

Now we will outline the proof of the much more difficult part of this theorem. In order to show that the absence of arbitrage in our normalized market implies the existence of an equivalent martingale measure  $Q$ , we will need a few results from functional analysis.

If we understand the set of all portfolio value processes as a vector space of integrable functions over the probability space  $(\Omega, \mathcal{F}_T^W, P)$ , we can use results from functional analysis, in particular the powerful Kreps-Yan Separation Theorem, to prove the existence of  $Q$ . Before this, let's lay out a few definitions:

Let  $L^1$  denote the space of all integrable functions under a measure  $\mu$ , and let  $L^\infty$  denote the set of all measurable functions that are bounded almost everywhere. That is,

$$f \in L^\infty \iff \exists C \geq 0 : \mu(|f| > C) = 0.$$

We can see that both  $L^1$  and  $L^\infty$  are vector spaces, with the norm of  $L^\infty$  given by:

$$\|f\|_\infty = M_0 \in \mathbb{R} : \mu(\{|f| > M_0\}) = 0 \text{ and } \forall M < M_0, \mu(\{M < |f| < M_0\}) > 0.$$

The number  $M_0$  is known as the essential supremum of  $f$ .

Now any vector space  $X$  has a corresponding dual space, denoted  $X^*$ , which is defined as the vector space consisting of all continuous linear functionals that operate on the elements of  $X$ .

Lemma: A linear functional  $\Lambda$  is continuous with respect to the norm topology if and only if it is bounded.

Proof: Suppose that  $\Lambda : X \rightarrow \mathbb{R}$  is bounded, that is, there exists some  $M \in \mathbb{R}$  such that for any  $x$  in  $X$ ,  $|\Lambda(x)| < M\|x\|_X$ . If  $\varepsilon > 0$ , then for any  $y$  in  $X$  such that  $\|x - y\| < \frac{\varepsilon}{M}$ , it follows that

$$|\Lambda(x - y)| \leq M\|x - y\|_X \leq M \frac{\varepsilon}{M} = \varepsilon.$$

Since  $\Lambda$  was arbitrary, this shows that any bounded linear functional is continuous.

Now consider the case in which  $\Lambda$  is continuous. To show that it is bounded, first note that  $\Lambda$  must be continuous at the zero vector  $0$  in  $X$ . Once more, let  $\varepsilon > 0$ . There exists  $\delta > 0$  such that for any  $x \in X$ , if  $\|x\|_X < \delta$ , then  $|\Lambda(x)| < \varepsilon$ . It follows that

$$\left| \Lambda\left(\frac{\delta}{\|x\|_X}x\right) \right| = \frac{\delta}{\|x\|_X} |\Lambda(x)| < \varepsilon.$$

Therefore,

$$|\Lambda(x)| < \frac{\varepsilon}{\delta} \|x\|_X.$$

The linearity of  $\Lambda$  implies that this inequality holds for any  $x$  in  $X$ . Therefore,  $\Lambda$  is a bounded linear functional.  $\square$

While  $X^*$  consists of functionals of the form  $\Lambda : X \rightarrow \mathbb{R}$ , notice that if we define a functional  $f_x : X^* \rightarrow \mathbb{R}$  by

$$f_x(\Lambda) = \Lambda(x),$$

this allows us to in turn view  $X$  as a collection of linear functionals on its dual  $X^*$ .

Definition: The coarsest topology on  $X^*$  on which  $\forall f_x \in X^*$ ,  $f_x(\Lambda)$  is continuous for every  $\Lambda \in X^*$  is the weak\* topology

As we will see, the vector space  $L^\infty$  shares the same underlying structure with  $(L^1)^*$ . This similarity in structure is captured by the concept of an isometric isomorphism.

Definition: Two vector spaces  $X$  and  $Y$  with respective norms  $\|\cdot\|_X, \|\cdot\|_Y$  are isometrically isomorphic when there exists a vector space isomorphism  $\Phi : X \rightarrow Y$  such that for any  $x$  in  $X$

$$\|x\|_X = \|\Phi(x)\|_Y.$$

Lemma: The space of essentially bounded functions  $L^\infty$  is isometrically isomorphic to the dual space  $(L^1)^*$

Proof: Lemma A.4 in the Appendix.  $\square$

Lemma: Under the weak\* topology,  $L^\infty$  is the dual of  $L^1$ .

Proof: This follows immediately from the previous lemma because the norm topology on  $(L^1)^*$  is strictly finer than the weak\* topology.  $\square$

Now the reason for laying out all of this functional analysis mumbo-jumbo lies in the statement of the Kreps-Yan Separation Theorem.

Theorem: Kreps/Yan Separation. Suppose that  $\mathcal{C} \subseteq L^\infty$  is a convex cone with  $-L_+^\infty \subseteq \mathcal{C}$  and  $\mathcal{C} \cap L_+^\infty = 0$ . If  $\mathcal{C}$  is weak\* closed, then there exists a random variable  $L$  in  $L_+^1$  for which  $\mathbf{E}[XL] \leq 0$  for every  $x$  in  $\mathcal{C}$ .

Proof: See Schachermayer [5].

Right now, the exact role of this theorem in proving the existence of this martingale measure  $Q$  must be far from clear, but hopefully it will become a bit more understandable if we characterize this space of portfolio processes.

A portfolio strategy consists of a sequence of decisions to buy and sell certain assets at given moments in time. In this sense, we may understand a portfolio value process at time  $T$  as a sum of basic financial transactions in which we borrow enough money to buy  $h(\omega)$  units of asset  $U_i$  at time  $s$ , hold on to it until time  $t$  (possibly  $T$ ), at which it is sold. Since the risk-free interest rate is 0 in the normalized market, the value of such a transaction is equal to  $h(\omega)[U_i(t, \omega) - U_i(s, \omega)]$ . Therefore, we can understand this space of portfolio processes at time  $T$  as a Hilbert space of random variables  $\mathcal{H}_0$  given by

$$\mathcal{H}_0 = \text{span}\left(1_A(\omega)[U_i(t, \omega) - U_i(s, \omega)] : A \in \mathcal{F}_s^W, s \leq t\right).$$

Since we are building a model of financial market behavior over continuous time, we need to preclude a few pathological examples of possible portfolio processes.

Firstly, we assume that there is only a finite amount of credit in the world, or in other words, the amount of money that one could possibly borrow at any given time is finite. This prevents crazy betting strategies such as the relentless “doubling-down” approach. What this means is that portfolio value processes need to be bounded below with probability 1.

Definition: A portfolio strategy  $h_S = [h_1, \dots, h_N]$  is admissible when there exists a nonnegative real number  $\alpha$  such that

$$\int_0^t h_i(s) dU_i(s, \omega) \leq -\alpha \quad \forall t \in [0, T].$$

Definition: A self-financing portfolio value process  $V^h(t) = \int_0^t \sum_{i=0}^N h_i(s) dU_i(s, \omega)$  is called admissible if the strategy  $h_S$  is admissible.

In view of this definition, we'll redefine our portfolio space to be

$$\mathcal{K}_0 = \{V^h(T) \mid h_S \text{ admissible}\}.$$

Next, it makes everything easier, in addition to more realistic, if the probability distribution at time  $T$  for each admissible process is also bounded from above. We'll denote this set of bounded portfolio processes by  $\mathcal{K}$ , so that

$$\mathcal{K} = \mathcal{K}_0 \cap L^\infty.$$

We can certainly check that  $\mathcal{K}$  is a vector space.

In view of the Kreps-Yan separation theorem, we will need to work with a convex cone of  $L^\infty$  that contains every essentially bounded negative measurable function. Let  $\mathcal{C}$  denote the set of all essentially bounded processes dominated by a member of  $\mathcal{K}$ , that is,

$$\mathcal{C} = \{g \in L^\infty : g \leq f \text{ for some } f \in \mathcal{K}\}.$$

A moment's reflection should verify that  $\mathcal{C}$  is indeed a convex cone that contains  $\mathcal{K}$ .

Now we need to incorporate our main assumption about the market: that there are no arbitrage portfolios. Since we have defined these self-financing portfolios to be without any initial cost, the absence of arbitrage implies that none of these portfolios have value processes that are strictly positive at time  $T$  with probability 1. This is equivalent to saying that

$$\mathcal{C} \cap L_+^\infty = \{0\}.$$

There is one more assumption we need to impose. Even if there is no arbitrage in this market, there still may be opportunities to earn a profit at exceedingly small risk. Given  $g \in L_+^\infty$  such that  $P[g > 0] > 0$ , if there exists a sequence of portfolio processes  $\{f_n\}_{n=1}^\infty \subseteq \mathcal{K}$  such that  $|g - f_n| < \frac{1}{n}$ , then for any  $k \in (0, 1)$ , we can construct a self-financing portfolio  $h$  such that

$$V_0^h = 0$$

$$\mathbf{P}[V_T^h > 0] > k.$$

In order to preclude this, we will also maintain that

$$\overline{\mathcal{C}} \cap L_+^\infty = \{0\}.$$

This condition is known as “no free lunch with vanishing risk” (NFLVR).

We have imposed a fair amount of conditions on  $\mathcal{C}$  and therefore  $\mathcal{K}$ . In order to use the Kreps-Yan Separation Theorem to our advantage, it needs to be true that the space of self-financing admissible portfolio processes with zero initial cost  $\mathcal{K}$  is a vector space. If this is to be true, however, then it must be impossible that there are no strictly negative portfolio processes in  $\mathcal{K}$ , lest we be able to use scalar multiplication to turn this into an arbitrage portfolio. Such processes, therefore, cannot be part of this set  $\mathcal{K}$ . This makes complete sense from a financial perspective: what interest could an investor have in a portfolio that will almost surely lead to a loss at time  $T$ ? Therefore, we do not seem to be making any dangerous assumptions in maintaining that  $\mathcal{K}$  is indeed a vector space.

From [3], it follows that (NFLVR) entails that the cone  $\mathcal{C}$  is weak\* closed. Therefore, we may use Kreps-Yan Separation to deduce the existence of a random variable  $L \in L_+^1$  for which  $\mathbf{E}[XL] \leq 0 \forall X \in \mathcal{C}$ . If we normalize  $L$  such that  $\mathbf{E}[L] = 1$ , then  $L$  is a Radon-Nikodym derivative that begets new probability measure  $Q$  defined by

$$dQ = LdP.$$

Now  $\mathbf{E}^P[XL] = \mathbf{E}^Q[X] \leq 0 \forall X \in \mathcal{K}$ , because  $\mathcal{K} \subseteq \mathcal{C}$ . Since  $\mathcal{K}$  is a vector space and the expected value is linear, it follows that

$$\mathbf{E}^Q[X] = 0 \forall X \in \mathcal{K}.$$



Now we need to show that  $Q$  is indeed a martingale measure for the normalized market. From the definition of  $\mathcal{H}$ , if we can show that for any event  $A$  in  $\mathcal{F}_s^W$ , the simple transaction  $1_A(\omega)[U_i(t, \omega) - U_i(s, \omega)]$  is a  $Q$ -martingale, then the proof is accomplished. Now

$$\int_{\Omega} 1_A(\omega)(U_i(t, \omega) - U_i(s, \omega)) dQ = \int_A (U_i(t, \omega) - U_i(s, \omega)) dQ = 0.$$

By the properties of conditional expectation, it follows that

$$\int_A \mathbf{E}[(U_i(t, \omega) - U_i(s, \omega)) : \mathcal{F}_s^W] dQ = 0.$$

However,  $A$  was simply an arbitrary event in the  $\sigma$ -algebra  $\mathcal{F}_s^W$ ! Therefore,

$$\int_{\alpha} \mathbf{E}[(U_i(t, \omega) - U_i(s, \omega)) : \mathcal{F}_s^W] dQ = 0 \forall \alpha \in \mathcal{F}_s^W.$$

Therefore,  $\mathbf{E}[(U_i(t, \omega) - U_i(s, \omega)) : \mathcal{F}_s^W] = 0$ . This implies that the all the value processes of portfolio strategies in  $K$  are  $Q$ -martingales. In particular, the simple strategy of buying an asset at time  $t = 0$  and holding on to it until the end is a  $Q$ -martingale. Therefore, each asset of the normalized market is a  $Q$ -martingale!

#### 4.4 Characterizing $Q$ as an Equivalent Probability Measure

We have arrived at the mysterious result encountered in the markets of previous sections. In the context of continuous time, we are far removed from the finite state-space enjoyed by the assets modeled in the first three sections. We had to develop the machinery of the Itô calculus to make sense of the dynamics of these assets as they change through time, but despite this drastic change in context, we still see this martingale measure  $Q$  arise out the assumption of no-arbitrage. So how are we to use this measure to price simple claims of the form  $X = \Phi(S_T)$ ?

From the outset, we have only proved the existence of this measure  $Q$ , but we have not yet discovered how exactly the dynamics of a stock price  $S_i(t, \omega)$  change in this other world described by  $Q$ . We know that each asset of the normalized market exhibits no drift under

$Q$ , and as we saw at the beginning of this section, this means that an asset in the original market has drift equal to  $r$ .

However, we do not know how exactly the stochastic component of the original asset has changed in this traversal from  $P$  to  $Q$ . Thus far, we simply know that according to  $Q$ , the dynamics of an original asset  $S_i$  are given by

$$dS_i(t, \omega) = rS_i(t, \omega)dt + S_i(t, \omega) \sum_{j=1}^d \theta_{ij}(t, \omega) dW_j^Q(t),$$

where each  $\theta_{ij}(t, \omega)$  is a process we have yet to discover. Equally importantly, if we are to use this measure  $Q$  to price simple claims, we are going to need to understand the processes driving the likelihood process  $L$  that begot  $Q$  in the first place. We will focus on addressing the latter concern first.

Since  $dQ = LdP$ , then for any  $s \leq t$  and  $\alpha \in \mathcal{F}_s^W$  the definition of conditional expectation implies that

$$\int_{\alpha} \mathbf{E}^P[L_t | \mathcal{F}_s^W] dP = \int_{\alpha} L_t dP = \int_{\alpha} dQ = Q[\{\omega \in \alpha\}],$$

where for ease of notation,  $L_t$  denotes  $L(t, \omega)$ . The rightmost probability does not appeal to the value of  $t$ . Therefore, it must also be true that

$$\int_{\alpha} \mathbf{E}^P[L_s | \mathcal{F}_s^W] dP = \int_{\alpha} L_s dP = \int_{\alpha} dQ.$$

The arbitrariness of  $\alpha$  and the fact that both  $\mathbf{E}^P[L_t | \mathcal{F}_s^W]$  and  $L_s$  are  $\mathcal{F}_s^W$ -measurable tells us that

$$\mathbf{E}^P[L_t | \mathcal{F}_s^W] = L_s.$$

By definition, this shows that  $L$  is a  $P$ -martingale!

So what do the dynamics of  $L$  look like? We know that  $L$  is strictly positive vector-valued  $P$ -martingale. Therefore, we'll make the ansatz that the dynamics of  $L$  may be given by

$$dL_t = \sum_{j=1}^d \phi_{tj} L_{tj} dW_j^P(t),$$

where the differential  $W_j^P(t)$  denotes the  $j^{\text{th}}$  coordinate of the a  $d$ -dimensional  $P$ -Wiener process, and  $\phi_{tj}$  is some process in  $L^2(\Omega, \mathcal{F}_t^P, P)$  that exhibits continuous trajectories.

Solving for  $L$  with Itô's formula gives us

$$L_t = e^{\int_0^t \phi_s \cdot dW^P(s) - \frac{1}{2} \int_0^t \|\phi_s\|^2 ds}.$$

In order to understand the full story of how  $Q$  changes the dynamics of an asset  $S_i$  whose  $P$  dynamics are given by

$$dS_i(t, \omega) = \alpha_i(t, \omega) S(t, \omega) dt + S(t, \omega) \sum_{j=1}^d \sigma_{ij}(t, \omega) dW_j^P(t).$$

It would be very helpful if we could express  $\mathbf{E}^Q[X | \mathcal{F}_T^W]$ , where  $X$  is some  $\mathcal{F}_T^W$  - adapted process, in terms of expected values under  $P$ . Such a relationship is drawn by Bayes' Theorem.

Theorem: Let  $P, Q$  be probability measures defined on the  $\sigma$ -algebra  $\mathcal{F}$  and related by the Radon-Nikodym derivative

$$L = \frac{dQ}{dP} \text{ on } \mathcal{F}.$$

Then given a random variable  $X \in L^1(\Omega, \mathcal{F}, P)$  and  $\sigma$ -algebra  $\mathcal{G} \subseteq \mathcal{F}$ ,

$$\mathbf{E}^Q[X | \mathcal{G}] = \frac{\mathbf{E}^P[LX | \mathcal{G}]}{\mathbf{E}^P[L | \mathcal{G}]} \text{ } Q\text{-almost surely.}$$

Proof: Section A.5 in the Appendix

For simplicity, we will work with a market that appeals to only a one-dimensional Wiener process, because each of the following results may be easily extended to the multi-dimensional case. In order to make use of this theorem, let's consider a one-dimensional Wiener process  $W$ , and an adapted process  $X$  whose dynamics are given by

$$dX(t, \omega) = \mu(t, \omega) dt + \sigma(t, \omega) dW(t).$$

For the sake of heuristics, we will assume that the drift process  $\mu(t, \omega)$  exhibits continuous trajectories. This implies that the conditional expectation of the drift process  $\mathbf{E}[\mu(t, \omega) | \mathcal{F}_s^W]$  is “well-behaved” in the sense that, for  $t \leq u$ ,

$$\lim_{u \rightarrow t} \mathbf{E}[\mu(u, \omega) | \mathcal{F}_t^W] = \mathbf{E}[\mu(t, \omega) | \mathcal{F}_t^W] = \mu(t, \omega).$$

Intuitively, this means that if we considered a very small increment of time from  $t$  to  $t + \Delta t$ , then the expected difference between  $X(t + \Delta t, \omega)$  and  $X(t, \omega)$  is driven only by the value of the drift term at time  $t$ . This is exactly in line with the equation of local dynamics given in the previous section. We will express this with the informal equation

$$\mathbf{E}[dX(t, \omega) | \mathcal{F}_t^W] = \mu(t, \omega)dt,$$

where  $dX(t, \omega) \in \mathcal{F}_{t+dt}^W$ .

This reason for all this hand-waving is that this assumption allows us to probe into this world of  $Q$  in order to obtain a general glimpse into what exactly happens to the dynamics a given adapted process  $X$ .

Firstly, let’s consider a process with no stochastic component, namely, a process  $X$  with dynamics

$$dX(t, \omega) = \mu(t, \omega)dt.$$

We will use Bayes’ Theorem to determine  $\mathbf{E}^Q[dX(t, \omega) | \mathcal{F}_t^W]$  and thereby discern what happens to the drift process of  $X$ .

$$\mathbf{E}^Q[dX(t, \omega) | \mathcal{F}_t^W] = \frac{\mathbf{E}^P[L_{t+dt}dX(t, \omega) | \mathcal{F}_t^W]}{\mathbf{E}^P[L_{t+dt} | \mathcal{F}_t^W]}.$$

Using the fact that  $L_t$  is a  $P$ -martingale, we obtain

$$\mathbf{E}^Q[dX(t, \omega) | \mathcal{F}_t^W] = \frac{\mathbf{E}^P[L_{t+dt}dX(t, \omega) | \mathcal{F}_t^W]}{L_t} = \frac{\mathbf{E}^P[L_t\mu(t, \omega)dt | \mathcal{F}_t^W]}{L_t}.$$

Since we know all the information about the random variable  $L_t\mu(t, \omega)$  at time  $t$ , it follows that

$$\frac{\mathbf{E}^P[L_t \mu(t, \omega) dt | \mathcal{F}_t^W]}{L_t} = \frac{L_t \mu(t, \omega) dt}{L_t} = \mu(t, \omega) dt.$$

Therefore,  $\mathbf{E}^Q[dX(t, \omega) | \mathcal{F}_t^W] = \mathbf{E}^P[dX(t, \omega) | \mathcal{F}_t^W]$ . This tells us that from  $P$  to  $Q$ , the drift of  $X(t, \omega)$  did not change! While we may conclude that the process  $X(t, \omega)$  has the same drift under  $Q$  as through the lens of  $P$ , it may very well be true that  $Q$  perceives  $X(t, \omega)$  to have an additional component unseen by  $P$ . Let's consider the possibility that according to  $Q$ ,

$$dX(t, \omega) = \mu(t, \omega) dt + d\Phi_t,$$

for some mystery process  $\Phi_t$ . We will show that this process must equal zero.

Lemma: Let  $L = \frac{dQ}{dP}$ , and  $X(t, \omega)_P$  denote the dynamics of the adapted process  $X(t, \omega)$  according to  $P$ . If

$$dX(t, \omega)_P = \mu(t, \omega) dt,$$

then

$$dX(t, \omega)_Q = \mu(t, \omega) dt.$$

Proof: Let  $\alpha \in \mathbb{R}$  such that the set  $K_\alpha = \{\omega \in \Omega : X(t, \omega) = \alpha\}$  is nonempty. We know that the trajectories  $\omega$  for which  $X(t, \omega) = \alpha$  according to  $P$  are exactly those according to  $Q$ ; the only possible difference is the size of this set viewed from these two different probability measures.

But what if we indeed had this mystery process  $\Phi(t, \omega)$  mentioned above, such that

$$X(t, \omega)_Q = \int_0^t \mu(s, \omega) ds + \Phi(t, \omega).$$

Then this would imply that there are trajectories  $\omega \in K_\alpha$  such that  $X(t, \omega)_P = \alpha$ , but  $X(t, \omega)_Q \neq \alpha$ .

This is impossible! A change in probability measures cannot change where the same random variable will send an element of the state space. By contradiction, it must be true that

$$\Phi(t, \omega) = 0.$$

□

We have seen that a process with only a drift term under  $P$  will exhibit no extra dynamics under  $Q$ , but if this is true, then how is this consistent with our previous observation that the drift of a risky asset  $S_i$  was given by  $\alpha_i(t, \omega)S_i(t, \omega)$  according to  $P$ , but changed to  $rS_i(t, \omega)$  under  $Q$ ?

While the probability measure  $Q$  views the dynamics of the original drift process no differently than  $P$  does, it can still change the overall drift of a random variable  $X$  by manipulating the stochastic component of  $dX$ . Let's see how  $Q$  perceives another process  $X$  that has only a stochastic component, that is,

$$dX(t, \omega) = dW^P(t).$$

Applying Bayes' Theorem once more, we obtain

$$\mathbf{E}^Q[dX(t, \omega)|\mathcal{F}_t^W] = \frac{\mathbf{E}^P[L_t dW_t + L_t \phi_t (dW_t^P)^2 | \mathcal{F}_t^W]}{\mathbf{E}^P[L_{t+dt} | \mathcal{F}_t^W]} = \frac{\mathbf{E}^P[L_t dW_t + L_t \phi_t dt | \mathcal{F}_t^W]}{L_t} = \phi_t dt.$$

To reach this result, we are also assuming that the process  $\phi_t$  exhibits continuous trajectories. Therefore, while  $X$  has no drift term according to  $P$ , its drift according to  $Q$  is given by the process  $\phi_t$ . A standard Wiener process through the lens of  $P$  seems to now have a drift term when viewed through  $Q$ ! So far, we may conclude that

$$dW^P(t) = \phi_t dt + d\Theta(t, \omega),$$

where  $\Theta(t, \omega)$  is some mystery process adapted to  $\mathcal{F}_t^W$  for which  $\mathbf{E}^Q[d\Theta(t, \omega)|\mathcal{F}_t^W] = 0$ .

If we can characterize this process, we will then fully understand how exactly the change from  $P$  to  $Q$  affects the dynamics of any  $X \in \mathcal{F}_T^W$ . To this end, we will use Lévy's Characterization of Brownian Motion.

Theorem: Lévy's Characterization of Brownian Motion. Let  $Y$  be a given process over the finite time interval  $[0, T]$ . Then  $Y$  is a Wiener Process according to a probability measure  $P$  if and only if

1.  $P[Y_0 = 0] = 1$
2.  $Y$  is a continuous martingale
3. The quadratic variation of  $Y(t)$  is equal to  $t$ .

So, given this process  $\Theta(t, \omega)$  for which

$$W^P(t) = \int_0^t \phi(s, \omega) ds + \Theta(t, \omega),$$

we will show that  $\Theta(t, \omega)$  is a Wiener process according to the measure  $Q$ . Since  $P$  and  $Q$  are equivalent, the first condition is immediate. For the second condition, we will make use of the following helpful lemma.

Lemma: Suppose  $P$  and  $Q$  are equivalent probability measures defined on the state space  $\Omega$  with  $\sigma$ -algebra  $\mathcal{F}$ , and  $Q \ll P$ . Then for a given sequence of random variables  $\{X_n\}_{n=1}^\infty$ ,

$$\text{if } X_n \rightarrow_P X, \text{ then } X_n \rightarrow_Q X.$$

Proof: Let  $\varepsilon > 0$ , and  $A_n = \{\omega \in \Omega : |X_n - X| \geq \varepsilon\}$ . From the outset,  $P[A_n] \rightarrow 0$ . Since  $Q \ll P$ , there exists a random variable  $L$  such that  $\forall A \in \mathcal{F}, Q[A] = \int_A L dP$ .

Now,  $\lim_{n \rightarrow \infty} \int_{A_n} L dP = \lim_{n \rightarrow \infty} Q[A_n] = 0$ . Therefore,

$$X_n \rightarrow_Q X.$$

□

The quadratic variation of a process is a limit in probability. Therefore, the previous lemma implies that the quadratic variation of  $W^P$  under  $P$  is exactly that under  $Q$ . Now we will find the quadratic variation of the process

$$\int_0^t dW_s - \int_0^t \phi(s, \omega) ds = \int_0^t d\Theta_s$$

under  $Q$ . By definition,

$$\langle \Theta(t, \omega) \rangle = \lim_{n \rightarrow \infty} \sum_{\mathcal{P}_n} \left[ (W(t_{i+1}) - W(t_i)) - \int_{t_i}^{t_{i+1}} \phi(s, \omega) ds \right]^2,$$

where  $\{\mathcal{P}_n\}_{n=1}^{\infty}$  is a sequence of partitions whose mesh tends to zero. If we break this sum into three pieces, we obtain

$$\langle \Theta(t, \omega) \rangle = \langle W(t) \rangle - \lim_{n \rightarrow \infty} \sum_{\mathcal{P}_n} 2 \left[ (W(t_{i+1}) - W(t_i)) \int_{t_i}^{t_{i+1}} \phi(s, \omega) ds \right] + \left\langle \int_0^t \phi(s, \omega) ds \right\rangle.$$

A previous lemma implies that  $\langle W(t) \rangle$  equal to  $t$ . We just need to show that the other two terms in this sum are equal to zero.

Now for any  $\omega \in \Omega$ , the function  $\Phi(t) = \int_0^t \phi(s, \omega) ds$  is uniformly continuous and therefore bounded on the interval  $[0, t]$ . Therefore, if we let  $\varepsilon > 0$ , then for some  $\delta > 0$ , any subinterval  $I \subseteq [0, t]$  of length less than  $\delta$  gives

$$\left| \int_I \phi(s, \omega) ds \right| < \varepsilon.$$

Therefore,

$$\left\langle \int_0^t \phi(s, \omega) ds \right\rangle < \lim_{n \rightarrow \infty} \sum_{\mathcal{P}_n} \varepsilon \int_{t_i}^{t_{i+1}} \phi(s, \omega) ds = \varepsilon \int_0^t \phi(s, \omega) ds.$$

Furthermore, since a Wiener process is uniformly continuous on  $[0, T]$  with probability 1 and  $\Phi(t)$  is differentiable, there exists a mesh  $M$  such that for any partition  $\mathcal{P}_n$  such that  $\|\mathcal{P}_n\| < M$ ,

$$\sum_{\mathcal{P}_n} \left| (W(t_{i+1}) - W(t_i)) \int_{t_i}^{t_{i+1}} \phi(s, \omega) ds \right| < \varepsilon K \sum_{\mathcal{P}_n} (t_{i+1} - t_i) = \varepsilon K T,$$

where  $K$  is some constant.



Since both  $\varepsilon$  and  $\omega$  were arbitrary, this shows that the quadratic variation of  $\Theta(t, \omega)$  is equal to  $t$ .

Finally, in order to show that  $\Theta(t, \omega)$  is a continuous  $Q$ -martingale, let's note that we see immediately that this mystery process is continuous. Appealing to the fact that  $L$  is a  $P$ -martingale and Bayes' Theorem, it is sufficient to show that the process  $L(t, \omega)\Theta(t, \omega)$  is a  $P$ -martingale. For ease of notation, we will denote  $L(t, \omega)$ ,  $\Theta(t, \omega)$ , and  $\phi(t, \omega)$  by  $L$ ,  $\Theta$ , and  $\phi$  respectively. Itô's formula yields

$$d(L\Theta) = (dL)(\Theta) + (L)(d\Theta) + (dL)(d\Theta).$$

$$(L\phi dW_t^P)(\Theta) + L(dW_t^P - \phi dt) + (L\phi dW_t^P)(dW_t^P - \phi dt) = (L\phi\Theta + L)dW_t^P$$

Since  $L\Theta$  is a stochastic integral with respect to  $W^P$  it is also a  $P$ -martingale. Therefore  $\Theta$  is a  $Q$ -martingale. This fulfills the last condition of Lévy's Characterization of Brownian Motion. This mystery process  $\Theta$  is therefore a Wiener Process viewed from the probability measure  $Q$ . That is,

$$\Theta(t, \omega) = W^Q(t).$$

Therefore,

$$dW^P(t) = \phi_t dt + dW^Q(t).$$

We have developed all of this theoretical machinery in order to make sense of how exactly we may price contingent claims in the context of continuous time. Even when we considered a vast state space of Wiener trajectories on the interval  $[0, T]$ , the assumption of no-arbitrage in our market has returned us to this strange probability measure  $Q$ , through which any asset of the normalized market exhibits no drift.

In proving the above result, we may now understand the exact dynamics of a given asset in the original economy. If  $S_i(t, \omega) \in \mathcal{F}_t^W$  is a risky asset whose  $P$ -dynamics are given by

$$dS_i(t, \omega) = \alpha(t, \omega)S_i(t, \omega)dt + S_i(t, \omega) \sum_{j=1}^d \sigma_{ij}(t, \omega)dW_j^P(t).$$

Then when we access the world of  $Q$  through the likelihood ratio, using the above equation in the multi-dimensional case implies that the  $Q$  - dynamics of this same asset are given by

$$dS_i(t, \omega) = S_i(t, \omega) \left( \alpha_i(t, \omega) + \sum_{j=1}^d \sigma_{ij}(t, \omega) \phi_j(t, \omega) \right) dt + S_i(t, \omega) \sum_{j=1}^d \sigma_{ij}(t, \omega) dW_j^Q(t).$$

We saw that for a normalized asset in the land of  $Q$  to exhibit no drift, it must be true that the drift of the original risky asset equals the risk-free interest rate  $r$ . In the case of a one-dimensional Weiner process, using the equation above yields the relation

$$\alpha(t, \omega) + \sigma(t, \omega) \phi(t, \omega) = r.$$

Since we may make the safe assumption that in a market,  $\sigma(t, \omega) \neq 0$  under  $Q$ , then we may solve for  $\phi(t, \omega)$  to obtain

$$\phi(t, \omega) = -\frac{\alpha(t, \omega) - r}{\sigma(t, \omega)}.$$

The proportion  $\frac{\alpha(t, \omega) - r}{\sigma(t, \omega)}$ , known as the Sharpe Ratio, measures the risk-adjusted return of an asset.

For the multidimensional case, we may organize the dynamics of each of the assets in the context of matrix multiplication. Let

$$dS(t, \omega) = (dS_1(t, \omega), dS_2(t, \omega), \dots, dS_N(t, \omega))$$

$$D[S(t, \omega)] = \begin{bmatrix} S_1(t, \omega) & 0 & \dots & 0 \\ 0 & S_2(t, \omega) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & S_N(t, \omega) \end{bmatrix}$$

$$\vec{\alpha}(t, \omega) = (\alpha_1(t, \omega), \alpha_2(t, \omega), \dots, \alpha_N(t, \omega))$$

$$\sigma(t, \omega) = \begin{bmatrix} \sigma_{11}(t, \omega) & \sigma_{12}(t, \omega) & \dots & \sigma_{1d}(t, \omega) \\ \sigma_{21}(t, \omega) & \sigma_{22}(t, \omega) & \dots & \sigma_{2d}(t, \omega) \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{N1}(t, \omega) & \sigma_{N2}(t, \omega) & \dots & \sigma_{Nd}(t, \omega) \end{bmatrix}$$

$$\vec{\phi}(t, \omega) = (\phi_1(t, \omega), \phi_2(t, \omega), \dots, \phi_d(t, \omega))$$

From this we obtain the equation

$$D[dS(t, \omega)] = D[S(t, \omega)][\vec{\alpha}(t, \omega) + \sigma(t, \omega)\vec{\phi}(t, \omega)]dt + D[S(t, \omega)]\sigma(t, \omega)dW^Q(t)$$

The market is absent of arbitrage if and only if we can find a vector  $\vec{\phi}$  such that

$$\sigma(t, \omega)\vec{\phi}(t, \omega) = \vec{r} - \vec{\alpha}(t, \omega)$$

where  $\vec{r} = (r, r, \dots, r)$ . A sufficient condition for this is that the matrix  $\sigma(t, \omega)$  is a surjection.

Now, if we are to use this to develop a pricing formula for a contingent claim  $X_T \in \mathcal{F}_T^W$ , we must show that the normalized version of this claim  $\tilde{X}_T = \frac{X_T}{B(T)}$  can be hedged in the normalized market. If this is the case, then as we have seen before, the absence of arbitrage in our market implies that the hedging portfolio will provide the price process  $\Pi(\tilde{X}_T; t)$  on  $[0, T]$ .

The definition of conditional expectation implies that the process  $M(t)$  defined by

$$M(t) = \frac{\mathbf{E}^Q[X_T | \mathcal{F}_t^W]}{B(t)}$$

is a  $Q$ -martingale. Moreover, we see that  $M(T) = \Pi(\tilde{X}_T; T)$ . Applying the Martingale Representation Theorem, we deduce that there exist unique adapted processes  $h_1, \dots, h_d$  such that

$$M(t) = \mathbf{E}[M(t)] = \sum_{i=1}^d \int_0^t h_i(s) dW_i^Q(s).$$

Since we know that the normalized assets  $U_1, \dots, U_n$  are also  $Q$ -martingales, their dynamics are given by the vector

$$dU(t, \omega) = D[U(t, \omega)]\sigma(t, \omega)dW^Q(t).$$

In addition, the dynamics of  $M(t)$  are given by

$$dM(t) = \vec{h} \cdot dW^Q(t).$$

Our goal is to seek a vector valued process  $\vec{g}(t, \omega)$  for which

$$dM(t) = \vec{g}(t, \omega) \cdot dU(t, \omega) = \vec{g}(t, \omega) D[U(t, \omega)]\sigma(t, \omega) dW^Q(t).$$

It is therefore sufficient to solve the equation

$$\vec{h}(t, \omega) D[U(t, \omega)]\sigma(t, \omega) = \vec{g}(t, \omega),$$

which is equivalent to showing that  $\vec{g}(t, \omega)$  is in the image of the transposed matrix  $\sigma^T(t, \omega)$ . As before, a sufficient condition, and perhaps the simplest condition, for this to be the case is that  $\sigma^T(t, \omega)$  is a surjection. Since the absence of arbitrage in our market has already implied that  $\sigma(t, \omega)$  is a surjection as well, it follows that the nonsingularity of the matrix  $\sigma(t, \omega)$  implies that our market is complete.

As a matrix of random processes, a moments reflection should reveal that notwithstanding pathological examples, the nonsingularity of  $\sigma(t, \omega)$  is reasonable. Since we now know that the dimension of our Weiner process is exactly equal to the number of risky asset on our market, it follows that our hedging portfolio will consist of  $g_i(t, \omega)$  shares of risky asset  $U_i$ , and to find the number of bonds  $g_0(t, \omega)$  owned by the portfolio at time  $t$ , we simply

take the difference

$$g_0(t, \omega) = M(t) - \sum_{i=1}^n \int_0^t g_i(s, \omega) dU_i(s, \omega).$$

This tells us that the process  $M(t)$  is hedged by a self-financing portfolio in the normalized market with strategy given by  $(g_0, g_1, \dots, g_N)$ , and therefore is hedged in the original market.

We have arrived at a wonderful result: in the normalized market, the arbitrage-free price process for the normalized claim  $\tilde{X}_T$  is a  $Q$ -martingale! This pricing process  $\Pi(\tilde{X}_T; t)$  is defined by

$$\Pi(\tilde{X}_T; t) = \mathbf{E}^Q[\tilde{X}_T | \mathcal{F}_t^W] = \frac{\mathbf{E}^Q[X_T | \mathcal{F}_t^W]}{B(T)}.$$

Since the risk-free interest rate  $r$  is compounded continuously, we have

$$\Pi\left(\frac{X_T}{e^{rT}}; t\right) = \frac{\mathbf{E}^Q[X_T | \mathcal{F}_t^W]}{e^{rT}}.$$

To convert this price process of a normalized portfolio into a process from the original economy, we multiply by  $e^{rt}$  to obtain

$$\Pi(X_T; t) = e^{-r(T-t)} \mathbf{E}^Q[X_T | \mathcal{F}_t^W].$$

This gives us a powerful method, known as risk-neutral valuation, by which we can price all contingent claims adapted to the filtration  $\mathcal{F}_T^W$ .

#### 4.5 The Black-Scholes Formula

Finally, consider the case in which the price process of a normalized simple claim  $\tilde{X}_T \in \mathcal{F}_T^W$  is a  $C^{1,2}$  function of the form  $f(t, U_i(t, \omega))$  for some risky normalized asset  $U_i$ . Then because  $f$  must be a  $Q$ -martingale, applying Itô's formula yields

$$\frac{\partial f}{\partial t} + rU \frac{\partial f}{\partial U} + \frac{\sigma^2 U^2}{2} \frac{\partial^2 f}{\partial U^2} = 0.$$

The price process described by the function  $f$  may be converted, as before, to reflect the original market through the function  $F(t, S_i(t, \omega)) = e^{rt} f(t, U_i(t, \omega))$ . Itô's formula applied one more gives

$$\frac{\partial F}{\partial t} + rS \frac{\partial F}{\partial S} + \frac{\sigma^2 S^2}{2} \frac{\partial^2 F}{\partial S^2} = rF.$$

This is the Black-Scholes Equation!

## BIBLIOGRAPHY

- [1] Øksendal, Bernt. 2000. Stochastic Differential Equations. Fifth Edition, Corrected Printing. Springer-Verlag Heidelberg New York. Available at <http://th.if.uj.edu.pl/gudowska/dydaktyka/Oksendal.pdf>
- [2] Steele, J. Michael. 2001. Stochastic Calculus and Financial Applications. Springer-Verlag Heidelberg New York.
- [3] Delbaen, F. and Schachermayer, W. 1998. The Fundamental Theorem for Unbounded Processes. *Mathematische Annalen* 300, 463-520.
- [4] Björk, Tomas. 2005, Arbitrage Theory in Continuous Time. Oxford University Press New York.
- [5] Schachermayer, W. 1994. Martingale Measures for Discrete Time Processes with Infinite Horizon. *Mathematical Finance* 4, 25-56.

## APPENDIX

### A.1: Doob's Maximal Inequality for Submartingales

Lemma: Let  $(\Omega, \mathcal{F}, P)$  denote a probability space with filtration  $\{\mathcal{F}_i\}_{i=1}^\infty$ . Given a sequence  $\{M_n\}_{n=1}^\infty$  of nonnegative submartingales and  $\lambda > 0$ ,

$$\lambda P(\sup_{0 \leq m \leq n} M_m \geq \lambda) \leq \mathbf{E}[M_n].$$

Proof: (Adapted from Steele [2])

A helpful fact that we will use in this lemma is that for any  $A \in \mathcal{F}_m$  and  $n \geq m$ , we have

$$\mathbf{E}[M_m 1_A] \leq \mathbf{E}[M_n 1_A].$$

To see this, we will use the definitions of a submartingale and conditional expectation.

$$\mathbf{E}[M_m | \mathcal{F}_m] \leq \mathbf{E}[M_n | \mathcal{F}_m]$$

$$M_m \leq \mathbf{E}[M_n | \mathcal{F}_m]$$

$$\int_A M_m dP \leq \int_A \mathbf{E}[M_n | \mathcal{F}_m] dP.$$

Since  $A \in \mathcal{F}_m$ , it follows that

$$\mathbf{E}[M_m 1_A] = \int_A M_m dP \leq \int_A M_n dP = \mathbf{E}[M_n 1_A].$$



Now define  $M_n^* = \sup_{0 \leq m \leq n} M_m$  as the maximal sequence associated with  $M_n$ , and let  $\tau$  denote

$$\tau = \min\{m : M_m \geq \lambda\}.$$

It follows that

$$P(M_n^* \geq \lambda) = P\left(\sup_{0 \leq m \leq n} M_m \geq \lambda\right) = P(\tau \leq n).$$

In the event that  $\tau \leq n$ , it must be true that  $M_\tau \geq \lambda$ . Therefore,

$$\lambda 1_{\{\tau \leq n\}} \leq M_\tau 1_{\{\tau \leq n\}} = \sum_{0 \leq m \leq n} M_m 1_{\{\tau=m\}}.$$

In addition, the helpful inequality above implies that

$$\sum_{0 \leq m \leq n} M_m 1_{\{\tau=m\}} \leq \sum_{0 \leq m \leq n} M_n 1_{\{\tau=m\}},$$

whence

$$\lambda 1_{\{\tau \leq n\}} \leq \sum_{0 \leq m \leq n} M_n 1_{\{\tau=m\}}.$$

Taking the expected value of both sides of this equation, and remembering that  $M_n$  is nonnegative, we obtain

$$\lambda P\left(\sup_{0 \leq m \leq n} M_m \geq \lambda\right) \leq \mathbf{E}[M_n 1_{\{\tau \leq n\}}] \leq \mathbf{E}[M_n].$$

□

## A.2: The Martingale Convergence Theorem

Theorem: Let  $\{\mathcal{F}_i\}_{i=0}^\infty$  denote a filtration of  $\sigma$ -algebras. Suppose that  $M_n$  is a square integrable martingale for which there exists a constant  $B$  such that

$$\mathbf{E}[M_n^2] \leq B \text{ for every } n \geq 0.$$

Then there exists a random variable  $M_\infty$  such that

$$\mathbf{E}[M_\infty^2] \leq B, P(\lim_{n \rightarrow \infty} M_n = M_\infty) = 1, \text{ and}$$

$$\lim_{n \rightarrow \infty} [M_n - M_\infty] = 0.$$

Proof: (Adapted from Steele [2])

For any two rational numbers  $a < b$ , denote the set  $A_{ab}$  by

$$A_{ab} = \left\{ \omega : \liminf_{n \rightarrow \infty} M_n(\omega) \leq a < b \leq \limsup_{n \rightarrow \infty} M_n(\omega) \right\}.$$

Now to show that  $M_n$  converges with probability 1, it is sufficient to prove that  $P(A_{ab}) = 0$ . In addition, for any  $\varepsilon > \frac{b-a}{2}$ ,

$$A_{ab} \subseteq \left\{ \omega : \sup_{m \leq k < \infty} |M_k - M_m| \geq \varepsilon \right\} \text{ for every } m \geq 0.$$

Our proof will focus on showing that the probability of this set is 0 as  $m \rightarrow \infty$ . Now define  $d_k = M_k - M_{k-1}$ . Then

$$\mathbf{E}[d_j d_k] = \mathbf{E}[M_k M_j - M_{k-1} M_j - M_k M_{j-1} + M_{k-1} M_{j-1}].$$

The law of iterated expectation implies that the right hand side of the above equality is equal to

$$\mathbf{E}[\mathbf{E}[(M_k - M_{k-1})M_j | \mathcal{F}_j]] - \mathbf{E}[\mathbf{E}[(M_k - M_{k-1})M_{j-1} | \mathcal{F}_{j-1}]].$$

Assuming that  $j \neq k$ , the definition of a martingale implies that  $\mathbf{E}[d_j d_k] = 0$ . We can use this fact in re-expressing the expectation:

$$\mathbf{E}[M_n^2] = \mathbf{E}\left[\left(\sum_{k=1}^n d_k\right)^2\right] = \sum_{k=1}^n \mathbf{E}[d_k^2].$$

Since  $\mathbf{E}[M_n^2] \leq B$  for every natural number  $n$ , the sum  $\sum_{k=1}^{\infty} \mathbf{E}[d_k^2]$  must converge.

With the aim of using Doob's Maximal Inequality, let's show that the process  $\tilde{M}_k = (M_{k+m} - M_m)^2$  is a submartingale, which means that for any  $s \leq k$ ,

$$\mathbf{E}[\tilde{M}_k \mid \mathcal{F}_{s+m}] \geq \tilde{M}_s.$$

Now

$$\mathbf{E}[\tilde{M}_k \mid \mathcal{F}_{s+m}] = \mathbf{E}\left[\left(\sum_{i=m+1}^{k+m} d_i\right)^2 \mid \mathcal{F}_{s+m}\right] = \mathbf{E}\left[\left(\sum_{i=m+1}^{s+m} d_i\right)^2 + \sum_{i=s+m+1}^{k+m} (d_i)^2 \mid \mathcal{F}_{s+m}\right],$$

where the rightmost equality is obtained by appeal to (52). Furthermore,

$$\mathbf{E}\left[\left(\sum_{i=m+1}^{s+m} d_i\right)^2 + \sum_{i=s+m+1}^{k+m} (d_i)^2 \mid \mathcal{F}_{s+m}\right] \geq \mathbf{E}\left[\left(\sum_{i=m+1}^{s+m} d_i\right)^2 \mid \mathcal{F}_{s+m}\right] = \tilde{M}_s,$$

showing us that  $\tilde{M}_k$  is indeed a submartingale. Doob's Maximal Inequality therefore implies

$$P\left(\sup_{k \geq m} |M_k - M_m| \geq \varepsilon\right) = P\left(\sup_{k \geq m} (M_k - M_m)^2 \geq \varepsilon^2\right) \leq \frac{1}{\varepsilon^2} \sum_{k=m+1}^{\infty} \mathbf{E}[d_k^2].$$

From (51), it follows that

$$P(A_{ab}) \leq \frac{1}{\varepsilon^2} \sum_{k=m+1}^{\infty} \mathbf{E}[d_k^2].$$

Taking the limit as  $m$  approaches infinity, it follows that  $P(A_{ab}) = 0$ . To finish the proof, we will further characterize this limit of martingales, which we will denote  $M_{\infty}$ . Now

$$M_{\infty} - M_n = \sum_{k=m+1}^{\infty} d_k,$$

from which

$$\lim_{k \rightarrow \infty} \sum_{k=n+1}^{\infty} \mathbf{E}[d_k^2] = \mathbf{E}[(M_\infty - M_n)^2] = 0.$$

□

### A.3: A Lemma on dense subsets of a Hilbert Space

Lemma: If  $\mathcal{H}$  denotes a closed subspace of  $L^2$ , and  $K \subseteq \mathcal{H}$ , then

$$\text{If } \mathcal{H} \cap K^\perp = \{0\}, \text{ then } \bar{K} = \mathcal{H}.$$

Proof: As a closed subspace of the Hilbert space  $L^2$ ,  $\mathcal{H}$  is also a Hilbert space. The closed subspace  $\bar{K} \subseteq \mathcal{H}$  therefore has an orthogonal complement  $K^\perp$ , and  $L^2$  has the decomposition

$$L^2 = \bar{K} \oplus K^\perp.$$

This means that any  $f$  in  $\mathcal{H}$  has a unique decomposition of the form  $f = g + h$ , where  $g \in \bar{K}$  and  $h \in K^\perp$ . Therefore,  $f - g = h \in \mathcal{H} \cap K^\perp$ , so  $f = g$ . Since  $f$  was an arbitrary element of  $\mathcal{H}$ , and was shown to be also an element of  $\bar{K}$ . It therefore follows that  $\bar{K} = \mathcal{H}$ . □

### A.4: $L^\infty$ and $(L^1)^*$ as isometrically isomorphic vector spaces

Lemma: The space of essentially bounded functions  $L^\infty$  is isometrically isomorphic to the dual space  $(L^1)^*$

Proof: We will first show that  $(L^1)^*$  and  $L^\infty$  s are isomorphic as vector spaces. Define the map  $\phi : L^\infty \rightarrow (L^1)^*$  by  $\phi(f) = \Lambda_f(\cdot) = \int f(\cdot) d\mu$ .

We can show that  $\phi$  is an injection, for if  $f_1 \neq f_2$ , then without loss of generality  $\mu(\{f_1 < f_2\}) > 0$ , and if  $g = 1_{\{f_1 < f_2\}}$ , then  $\Lambda_{f_1}(g) < \Lambda_{f_2}(g)$ .

In addition,  $\phi$  is also a surjection, for if we consider any  $\Lambda$  in  $(L^1)^*$ , then we may define the signed measure  $\nu : \Omega \rightarrow \mathbb{R} \cup \{\pm\infty\}$  by

$$\nu(A) = \Lambda(1_A).$$

As a signed measure,  $\nu$  admits of the Hahn decomposition  $\nu = \nu^+ - \nu^-$ . The  $\sigma$ -finite measures  $\nu^+$  and  $\nu^-$  both have Radon-Nikodym derivatives under  $\mu$  respectfully denoted by  $f^+$  and  $f^-$ . Therefore, if we denote the function  $f = f^+ - f^-$ , then  $\phi(f) = \Lambda$ . This shows that  $\phi$  is indeed a vector space isomorphism.

Now to show that the map  $\phi$  is an isometry, we will work with the definition of the dual norm

$$\|\Lambda\| = \max_{f \in L^1} \frac{|\Lambda(f)|}{\|f\|}.$$

From the outset, a moment's reflection should tell us that if  $\Lambda(f) = \int f g d\mu$ , then  $\|\Lambda\| \leq \text{esssup}(g)$ . If we denote  $\text{esssup}(g)$  by  $M$ , then if  $A_n = g^{-1}(M - \frac{1}{n}, M]$ , it follows that

$$\lim_{n \rightarrow \infty} \|\Lambda(1_{A_n})\| = M.$$

Therefore,  $\|\Lambda\| = M = \text{esssup}(g)$ , showing that  $\phi$  is an isometry.

□

### A.5: Bayes' Theorem

Theorem: Let  $P, Q$  be probability measures defined on the  $\sigma$ -algebra  $\mathcal{F}$  and related by the Radon-Nikodym derivative

$$L = \frac{dQ}{dP} \text{ on } \mathcal{F}.$$

Then given a random variable  $X \in L^1(\Omega, \mathcal{F}, P)$  and  $\sigma$ -algebra  $\mathcal{G} \subseteq \mathcal{F}$ ,

$$\mathbf{E}^Q[X|\mathcal{G}] = \frac{\mathbf{E}^P[LX|\mathcal{G}]}{\mathbf{E}^P[L|\mathcal{G}]} Q - \text{almost surely.}$$

Proof: If we can first prove that  $\mathbf{E}^P[L|\mathcal{G}] \neq 0$   $Q$ -almost surely, then the proof of this theorem reduces down to showing the equality

$$\mathbf{E}^Q[X|\mathcal{G}]\mathbf{E}^P[L|\mathcal{G}] = \mathbf{E}^P[LX|\mathcal{G}].$$

Since  $\mathbf{E}^P[L|\mathcal{G}] = 0$  is a  $\mathcal{G}$ -measurable function, let  $\beta \in \mathcal{G}$  denote  $(\mathbf{E}^P[L|\mathcal{G}] = 0)^{-1}(\{0\})$ . It follows that

$$Q[\mathbf{E}^P[L|\mathcal{G}] = 0] = \int_{\beta} dQ = \int_{\beta} L dP = \int_{\beta} \mathbf{E}^P[L|\mathcal{G}] dP = 0.$$

Hence,  $\mathbf{E}^P[L|\mathcal{G}] \neq 0$   $Q$ -almost surely.

Now notice that the left-hand side of the prior above inequality is a product of two  $\mathcal{G}$ -measurable functions. Therefore, if we pick an arbitrary  $\alpha \in \mathcal{G}$  and show that the integral of the left-hand side over  $\alpha$  with respect to  $P$  is equal to that of the right-hand side, then the proof is accomplished.

Focusing on the left hand side, we have

$$\int_{\alpha} \mathbf{E}^Q[X|\mathcal{G}]\mathbf{E}^P[L|\mathcal{G}] dP.$$

If we approximate the random variable  $\mathbf{E}^Q[X|\mathcal{G}]$  with a sequence of step functions  $\{\phi_k\}_{k=1}^n$  over the partition  $\bigcup_{k=1}^n S_k = \alpha$ , where  $S_k \in \mathcal{G}$ , then

$$\int_{\alpha} \mathbf{E}^Q[X|\mathcal{G}] \mathbf{E}^P[L|\mathcal{G}] dP = \lim_{n \rightarrow \infty} \sum_{k=1}^n \phi_k \int_{S_k} \mathbf{E}^P[L|\mathcal{G}] dP.$$

By the definition of conditional expectation,

$$\int_{S_k} \mathbf{E}^P[L|\mathcal{G}] dP = Q[\{\omega \in S_k\}].$$

Therefore, the right-hand side of previous above equality is equal to

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n \phi_k Q[\{\omega \in S_k\}] = \int_{\alpha} \mathbf{E}^Q[X|\mathcal{G}] dQ = \int_{\alpha} X dQ.$$

In addition, we immediately see that

$$\int_{\alpha} \mathbf{E}^P[LX|\mathcal{G}] dP = \int_{\alpha} XL dP = \int_{\alpha} X dQ.$$

The proof is complete by the arbitrariness of  $\alpha$ . □

#### A.6: Worked Problems in Financial Mathematics

Problem 6.1: We seek an upper bound on the value of a European call option  $c$ :

Solution: Seeking a contradiction, consider the case in which

$$c > S_0$$

A clever investor could buy the stock at  $t = 0$  for  $S_0$ , then sell the call option for an immediate profit. Consider the two possible cases at  $t = T$ :

If  $K > S_T$ , the owner of the call option does not redeem the offer, so the investor's initial profit is secured.

If  $K < S_T$ , the owner of the call option redeems the offer, and buys the stock for  $K$  dollars. Since the investor already owns a single share of stock, he can simply sell this share to the owner.

Problem 6.2: We seek an upper bound on the value of a European put option  $p$  for a stock:

Solution: Consider the case in which

$$p > Ke^{-rt}$$

A clever investor will sell this put option and invest it in a bank. If the holder decides to redeem the option, the investor is liable to purchase the stock for  $K$  dollars. Since he already has an initial influx of an amount greater than  $Ke^{-rt}$ , he can invest this sum in a bank at time  $t = 0$  to yield an amount greater than  $K$  at  $t = T$ , securing a profit.

Problem 6.3: Now we will prove that

$$c \geq S_0 - Ke^{-rt}$$

Solution:

Consider the following portfolios:

- Portfolio A, denoted  $A(t)$ :  $c + Ke^{-r(T-t)}$ : this portfolio may be understood as one call option, in addition to a treasury bill set to mature to  $K$  dollars at time  $T$
- Portfolio B, denoted  $B(t)$ : one share of stock  $S(t)$

What are the values of these portfolio at time  $T$ ? We'll subdivide into two cases:

If  $S_T > K$ , then  $c = S_T - K$  and Portfolio A is therefore worth  $S_T$ , which is the value of Portfolio B.

If  $S_T < K$ , then the call option is worthless and Portfolio A is worth  $K$ .

Therefore, with absolute certainty,  $A(T) \geq B(T)$ . It follows that  $A(0) \geq B(0)$ , lest an arbitrage opportunity present itself.

Hence,

$$c + Ke^{-rt} \geq S_0$$



Problem 6.4: Now we seek a lower bound on  $p$ .

Solution: Consider two portfolios:

- Portfolio A, denoted  $A(t)$ :  $Ke^{-rt}$
- Portfolio B, denoted  $B(t)$ :  $p + S_0$

From the outset, we know that at time  $t = T$ ,  $A(T) = K$ . Now we will consider two cases once more:

If  $S_T > K$ , the put option is worthless and  $B(T) = S_T$

If  $S_T < K$ , the value of the put option is  $K - S_T$ , hence  $B(T) = K$ . This means that

$$B(T) = \max(S_T, K)$$

Therefore,  $B(T) \geq A(T)$  with absolute certainty. It follows that  $B(0) \geq A(0)$ , that is,

$$Ke^{-rt} \leq p + S_0$$

Theorem: Put Call Parity. Appealing to the previous inequalities established, we will prove put-call parity:

$$c + Ke^{-r(T-t)} = p + S_0$$

Solution: We'll create two portfolios:

- $A(t) = c + Ke^{-r(T-t)}$
- $B(t) = p + S_0$

With absolute certainty, we know that

$$A(T) = B(T) = \max(S_T, K)$$

Therefore, if  $A(t) \neq B(t)$  for any time  $t$  before maturity, an arbitrage opportunity presents itself. Therefore,  $A(t) = B(t)$ .

Problem 6.5: Consider three European call options  $c_1, c_2, c_3$  with respective strike prices of  $K_1, K_2, K_3$ . Moreover,  $K_1 < K_2 < K_3$  and  $K_3 - K_2 = K_2 - K_1$ , that is  $K_2$  is the midpoint of  $K_1$  and  $K_3$ . We will show that

$$c_2 \leq \frac{c_1 + c_3}{2}$$

Solution: Consider a portfolio in which  $c_1$  and  $c_3$  are bought, and two  $c_2$ 's are shorted:

Problem 3.23 tells us that

$$c_1 \geq S_0 - K_1 e^{-rT}$$

$$c_2 \geq S_0 - K_2 e^{-rT}$$

$$c_3 \geq S_0 - K_3 e^{-rT}$$

Therefore,

$$c_1 + c_3 \geq 2S_0 - (K_1 + K_3)e^{-rT} = 2S_0 - 2K_2 e^{-rT}$$

We know that  $2S_0 - 2K_2 e^{-rT} \leq 2c_2$ , but unfortunately this does not help us here. Instead, consider the case in which  $2c_2 > c_1 + c_3$ :

When  $S_T < K$ , the holder(s) of  $c_2$  do not redeem this option.

When  $S_T > K$ , the holder(s) redeem their call options  $c_2$ , so the investor must come up with two stocks to sell to these holders. Since he already owns the two options  $c_1$  and  $c_3$ , he can purchase two shares for a total price of  $K_1 + K_3 = 2K_2$ , only to sell them for exactly the price he paid, thereby breaking even.

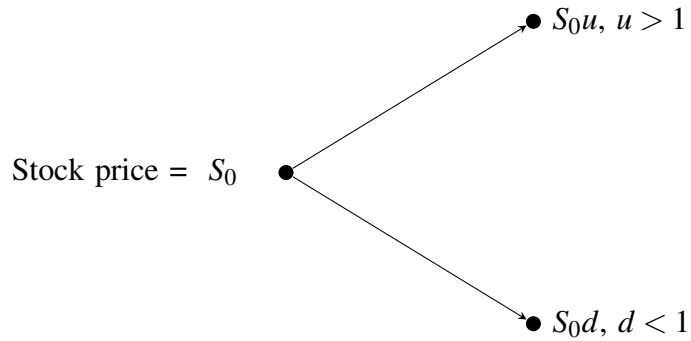
This shows that an arbitrage opportunity exists. Therefore,  $2c_2 \leq c_1 + c_3$

Problem 6.6: Now consider three European put options  $p_1, p_2, p_3$  with respective strike prices of  $K_1, K_2, K_3$ . Moreover,  $K_1 < K_2 < K_3$  and  $K_3 - K_2 = K_2 - K_1$ , that is  $K_2$  is the midpoint of  $K_1$  and  $K_3$ . We will show that

$$p_2 \leq \frac{p_1 + p_3}{2}$$

Solution: Using reasoning that is similar to the previous problem, we will consider a portfolio consisting of one share each of  $p_1$  and  $p_2$ , in addition to a short-sold two shares of  $p_3$ . Consideration of the case in which  $2p_2 > p_1 + p_3$  in the same way as before.

Now let's consider a simple binomial model for stock price behavior under probability  $p$ :



We'll denote the payoff of a stock option when  $S_T = S_0u$  by  $f_u$ , and when  $S_T = S_0d$  by  $f_d$ .

Consider a portfolio that is riskless, that is, whose final value is the same for both possible outcomes for  $S_T$ . To construct a riskless portfolio, suppose that we have shorted a single call options, and have bought  $\Delta$  shares of stock, where  $\Delta$  remains a number we have yet to determine.

We know that once time  $T$  comes around, the value of our call option will either be  $f_u$  or  $f_d$ . Therefore, the for this portfolio to be riskless, we have

$$\Delta S_0u - f_u = \Delta S_0d - f_d$$

Solving for  $\Delta$  yields

$$\Delta = \frac{f_u - f_d}{S_0(u - d)}$$

Since we know the ending value of this portfolio with certainty, the initial value of our portfolio is given by

$$e^{-rT}(\Delta S_0 u - f_u) = \Delta S_0 - f$$

Where  $f$  denotes the initial price of the call option. Solving for this variable yields

$$f = -e^{-rT}(\Delta S_0 u - f_u) + \Delta S_0$$

Using the fact that  $S_0 \Delta = \frac{f_u - f_d}{(u - d)}$  and a little arithmetic, we arrive at the following result:

$$f = e^{-rT}[pf_u + (1 - p)f_d]$$

where

$$p = \frac{e^{rT} - d}{u - d}$$

This result suggests that the quantity  $p$  may be interpreted as a probability even if this probability does not reflect the actual behavior of the stock price!. We call such a probability measure for which  $S_T = S_0 u$  with probability  $p$  a martingale measure. This measure may also be developed from a binomial model in the context of stocks and bonds, as will be discussed (maybe).

Theorem: Black's Pricing Formula for a European Call Option

We will now prove the Black-Scholes-Merton formula for European call options. If  $c$  denotes the price of such an option with strike price  $K$  and time of maturity  $T$ , then given that  $r$  is the risk-free interest rate, the price of the call option at time  $t = 0$  is given by

$$c = e^{-rT} [S_0 e^{rT} N(d_1) - KN(d_2)],$$

where

$$d_1 = \frac{\ln(S_0/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}$$

$$d_2 = \frac{\ln(S_0/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}$$

$$N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-x^2/2} dx.$$

This formula is monstrous. Instead of trying to build up to this equation, let's remember that we are working in the land of  $\mathcal{Q}$ , in which the dynamics of a value of process of a stock are given by

$$dS(t) = rS(t)dt + \sigma S(t)dW(t)$$

Where  $r$  is the risk-free interest rate,  $\sigma$  is the volatility of the stock price, and  $W(t)$  is a one-dimensional Wiener process with respect to the underlying martingale measure  $\mathcal{Q}$ .

Now, let's consider the dynamics of the random variable  $\ln(S(t))$  over the interval  $[0, T]$ . We do not have to worry about issues with domain here because we are assuming that a stock price is always positive. So, let  $F(u, S(u)) = \ln(S(u))$ . Itô's formula gives us

$$dF = \left( r - \frac{\sigma^2}{2} \right) du + \sigma dW(u).$$

But this is just shorthand for

$$F(T, S(T)) = F(0, S(0)) + \left( r - \frac{\sigma^2}{2} \right) T + \int_0^T \sigma dW(u).$$

This means that  $\ln(S(T))$  is normally distributed with mean  $\ln(S_0) + \left(r - \frac{\sigma^2}{2}\right)T$  and variance  $\sigma^2 T$ . We also know that according to the Black-Scholes model, the expected value operator under the probability measure  $Q$  gives the price of  $c$ , that is,

$$c(t) = \mathbf{E}_t^Q[\max(S_T - K, 0)].$$

Therefore, to give a more explicit pricing formula for  $c$ , we just need to calculate the expected value of this distribution that considers any stock price less than  $K$  to be zero. This is given by

$$\int_K^\infty (x - K) \phi_{S_T}(x) dx = \mathbf{E}_t^Q[\max(S_T - K, 0)],$$

where  $\phi_{S_T}$  denotes the probability density function of the random variable  $S_T$ . What does such a function look like? Since we know that  $\ln(S_T)$  is normally distributed, and we are interested in pricing the option from an initial time  $t$ , it follows that

$$\phi_{S_T}(x) = \frac{1}{x\sqrt{2\pi\sigma^2 T}} e^{-\frac{(\ln(x) - \mu)^2}{2\sigma^2 T}} dx,$$

where  $\mu = \ln(S_0) + \left(r - \frac{\sigma^2}{2}\right)T$ .

Now the integral we seek to evaluate is

$$\int_K^\infty (x - K) \frac{1}{x\sqrt{2\pi\sigma^2 T}} e^{-\frac{(\ln(x) - \mu)^2}{2\sigma^2 T}} dx.$$

Let's clean this up and rewrite it as two integrals:

$$\frac{1}{\sqrt{2\pi\sigma^2 T}} \left[ \int_K^\infty e^{-\frac{(\ln(x) - \mu)^2}{2\sigma^2 T}} dx - K \int_K^\infty \frac{1}{x} e^{-\frac{(\ln(x) - \mu)^2}{2\sigma^2 T}} dx \right].$$

Under a change of variables  $u = \ln(x)$ , the problem becomes a bit more recognizable:

$$\frac{1}{\sqrt{2\pi\sigma^2 T}} \left[ \int_{\ln(K)}^\infty e^u e^{-\frac{(u - \mu)^2}{2\sigma^2 T}} du - K \int_{\ln(K)}^\infty e^{-\frac{(u - \mu)^2}{2\sigma^2 T}} du \right].$$

We may understand the integral on the right as the probability of the event  $X \geq \ln(K)$ , when  $X$  is selected at random from a normal distribution centered at  $\mu$  with variance  $\sigma^2 T$ . Let's make another change of variables so we may understand this distribution to instead be centered at zero with unit variance. If we allow  $v = \frac{u - \mu}{\sigma\sqrt{T}}$ , then

$$\frac{K}{\sqrt{2\pi\sigma^2 T}} \int_{\ln(K)}^{\infty} e^{-\frac{(u-\mu)^2}{2\sigma^2 T}} du = \frac{K}{\sqrt{2\pi}} \int_{\frac{\ln(K)-\mu}{\sigma\sqrt{T}}}^{\infty} e^{-\frac{v^2}{2}} dv.$$

By definition, this integral is equal to  $KN\left(-\frac{\ln(K) - \ln(S_0) - (r - \frac{\sigma^2}{2}T)}{\sigma\sqrt{T}}\right) = KN(d_2)$ .

To evaluate the integral on the right, we will rewrite the integrand by completing the square:

$$e^u e^{-\frac{(u-\mu)^2}{2\sigma^2 T}} = S_0 e^{rT} e^{-\frac{(u-v)^2}{2\sigma^2 T}}, \text{ where } v = \ln(S_0) + (r + \frac{\sigma^2}{2})T.$$

Under a change of variables similar to that of the previous integral, the left integral therefore becomes

$$\frac{S_0 e^{rT}}{\sqrt{2\pi\sigma^2 T}} \int_{\ln(K)}^{\infty} e^{-\frac{(u-v)^2}{2\sigma^2 T}} du = S_0 e^{rT} N(d_1).$$

Therefore,

$$c = \mathbf{E}^Q[\max(S_T - K, 0)] = S_0 e^{rT} N(d_1) - KN(d_2).$$