Stochastic Calculus for Finance
Continuous Time Models

Notes by

Daniel Ruiz

Last Updated: September 2020

Prelude

These notes are based off the two books *Stochastic Calculus for Finance Vol.1: The Binomial Asset Pricing Model* and *Stochastic Calculus for Finance Vol.2: Continuous Time Models* by Steven Shreve. You are not required to read Vol.1 prior to Vol.2 but having at the very least a high level understanding of the Binomial Asset Pricing model is good. In these notes, Chapter 0 is taken from the first chapter of Vol.1 but the rest of the Chapter notes follow from Vol.2.

Contents

Chapter 0

The Binomial No-Arbitrage Pricing Model

The material in this chapter is actually from *Stochastic Calculus for Finance Vol.1: The Binomial Asset Pricing Model* Chapter 1. One does not need to read Vol.1 prior to Vol.2 but I have provided this chapter as Vol.2 still makes connections to some concepts introduced in in Vol.1.

0.1 One-Period Binomial Model

The *binomial asset-pricing model* is a powerful tool that we introduce in this chapter and will follow up on in Chapter 2. In this section, we consider the simplest binomial model, the one with only one period.
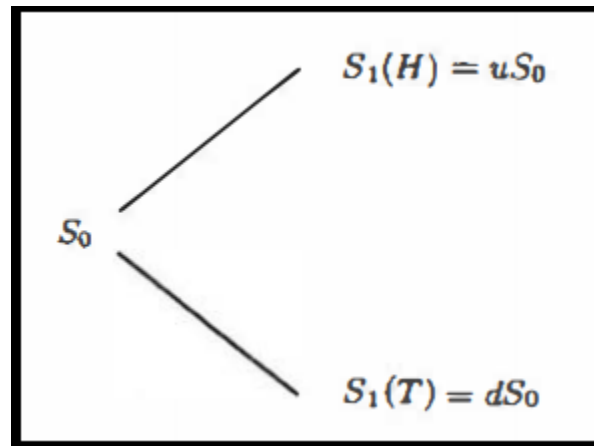

Figure 0.1: General one-period binomial model

> **Definition 0.1: Time Zero and Time One**
>
> *For the general one-period model seen above in Fig.0.1, we call the beginning of the period time zero and the end of the period time one.*

One period binomial model:

1. At time zero, we have a stock whose price per share we denote by S_0, a positive quantity known at time zero.

2. At time one, the price per share of this stock will be one of two positive values, which we denote by $S_1(H)$ and $S_1(T)$, with the H and T standing for *head* and *tail*, respectively.

Hence, one imagines that a coin is tossed, and the outcome of the coin toss determines the price at time one. However, we do not assume that this coin is fair (i.e probability of head need not be one-half). We assume only that the probability of head, which we call p, is positive and the probability of tail, which is $q = 1 - p$, is also positive.

Definition 0.2: Random

We shall refer to any quantity not known at time zero as random because it depends on the random experiment of tossing a coin (or something similar).

Definition 0.3: Up and Down Factors

In Fig.0.1, we had introduced two positive numbers:

$$u = \frac{S_1(H)}{S_0}, d = \frac{S_1(T)}{S_0}. \tag{0.1}$$

WLOG, we assume that $d < u$ (as if instead one has $d > u$, we can achieve $d < u$ by simply relabeling the sides of our coin). We note that if $d = u$, then the stock price at time one is not really random and the model turns out to be rather uninteresting.

We refer to u as the up factor and d as the down factor. Intuitively, one can consider $d < 1$ and $u > 1$.

Definition 0.4: Interest Rate

We introduce an interest rate r. One dollar invested in the money market at time zero will yield $1+r$ dollars at time one. Conversely, one dollar borrowed from the money market at time zero will result in a debt of $1 + r$ at time one. In particular, interest rate for borrowing is the same as the interest rate for investing. It is almost true that $r \geq 0$, and this is the case to keep in mind. However, the mathematics we develop requires only that $r > -1$.

Definition 0.5: Arbitrage

*We define **arbitrage** as a trading strategy that begins with no money, has zero probability of losing money, and has a positive probability of making money. A mathematical model that admits arbitrage cannot be used for analysis.*

Proposition 0.1: Rule Out Arbitrage

In the one-period binomial model, there is no arbitrage if and only if

$$0 < d < 1 + r < u. \tag{0.2}$$

Proof. Exploiting Strategy 1:

1. Suppose that you borrow $x \geq S_0$ from the market at an interest rate of r.

2. You buy a single share at S_0, and have $x - S_0$ left over.

3. At time one, assume that the stock price rises to $S_1(T) = dS_0$ and you sell. You currently now have $x + S_0(d - 1)$ left over.

4. You decide to repay your debt, meaning that you now have $x + S_0(d-1) - x(1+r) = S_0(d-1) - xr$ left. If this quantity is positive then you have just made free money independent of what side the coin flipped (as we had assumed WLOG $d < u$).

Case 1: Suppose that $d < 1$, then $S_0(d-1) > x(d-1)$. Hence,

$$(d - 1 - r)x < S_0(d-1) - xr \tag{0.3}$$

If $d \geq 1 + r$, then $d - 1 - r \geq 0$ and we thus have $S_0(d-1) - xr > 0$. Hence, to avoid such a scenario, we require that $d < 1 + r$.

Exploiting Strategy 2:

1. Suppose that you borrow x from the market and buy a stock at S_0.

2. At time zero, you decide to short sell it at S_0 and invest the proceeds as an investment into the market at an interest rate of r. You now are back to having x.

3. At time one, you now have $x + S_0 r$.

Suppose that instead, you held onto your stock and its new value resulted in $S_1(H) = uS_0$. If you sold your stock at time one, you would now have $x - S_0 + uS_0 = x + S_0(u-1)$. Under what conditions would this strategy be less valuable or equal to simply reinvesting into the market? That would simply be the case when

$$x + S_0(u - 1) \leq x + S_0 r, \tag{0.4}$$

which occurs when $u \leq 1 + r$. Such a scenario is also deemed to be arbitrage, and so we must therefore require $1 + r < u$. □

It is common to have $d = 1/u$, and this will be the case in many examples. For the binomial asset-pricing model to make sense, we need only assume Eq.(0.2).

Stock price movements are much more complicated than indicated by the binomial asset-pricing model. We consider this simple model for three reasons:

1. Within this model, the concept of arbitrage pricing and its relation to risk-neutral pricing is clearly illuminated.

2. The model is used in practice because, with a sufficient number of periods, it provides a reasonably good, computationally tractable approximation to continuous time models.

3. Within the binomial asset-pricing model, we can develop the theory of conditional expectations and martingales, which lies at the heart of continuous-time models.

Definition 0.6: European Call Option

A European call option is a call option that confers on its owner the right but not the obligation to buy one share of the stock at time one for the strike price K. In general, European options are a version of an options contract that limits execution to its expiration date.

Let's consider a European call option, for the case where $S_1(T) < K < S_1(H)$. If we get a tail on the toss, the option expires worthless. If we get a head on the coin toss, the option can be *exercised* and yields a

profit $S_1(H) - K$. Taken together, we say that the option at time one is worth [1]

$$\text{relu}(S_1 - K) := \begin{cases} S_1 - K & \text{if } S_1 - K > 0, \\ 0 & \text{else.} \end{cases} \tag{0.6}$$

Example 0.1: No Arbitrage Price at time zero: Binomial Model

In this example, we consider a particular binomial model where we will want to replicate a specific option.
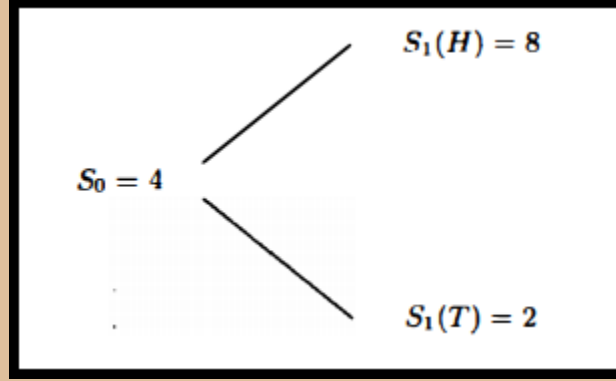
Figure 0.2: Particular one Period Binomial Model

Let

- $S_0 = 4$ *[Stock Price at time Zero]*

- $u = 2$ *[Up Factor]*,

- $d = \frac{1}{2}$ *[Down Factor]*,

- $r = \frac{1}{4}$ *[Interest Rate]*,

- $K = 5$ *[Strike Price]*.

Hence, we have that $S_1(H) = 8, S_1(T) = 2$. Suppose that we begin with an initial wealth of $X_0 = 1.20$ and buy $\Delta_0 = 1/2$ shares of stock at time zero. Our initial cash position at time zero is therefore given by $X_0 - \Delta_0 S_0 = -0.80$. At time one, our cash position will be $(1 + r)(X_0 - \Delta_0 S_0) = -1$. At time one, the value of our portfolio if the coin toss results in

- *Heads will be*

$$X_1(H) = \frac{1}{2}S_1(H) + (1 + r)(X_0 - \Delta_0 S_0) = 3; \tag{0.7}$$

- *Tails will be*

$$X_1(T) = \frac{1}{2}S_1(T) + (1 + r)(X_0 - \Delta_0 S_0) = 0. \tag{0.8}$$

In either case, the value of our portfolio agrees with the value of the option at time one, which is either $\text{relu}(S_1(H) - 5) = 3$ or $\text{relu}(S_1(T) - 5) = 0$. We have replicated the option by trading in the stock and money markets.

---

[1]I use relu here to signify the rectified linear unit function, which is defined as

$$\text{relu}(x) := \begin{cases} x & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases} \tag{0.5}$$

*The initial wealth of 1.20 needed to set up the replicating portfolio described above is the **no-arbitrage price of the option at time zero**. If one could sell the option for mare than this, then the seller could invest the excess amount in the money market and use the remaining to replicate the option, which would lead to arbitrage. A similar situation arises if the option was priced at less than 1.20.*

Principle Model Assumptions

The argument in the above example depends on several assumptions. The principal ones are

1. *shares of stock can be subdivided for sale or purchase,*

2. *the interest rate for investing is the same as the interest rate for borrowing,*

3. *the purchase price of stock is the same as the selling price (i.e. there is zero bid-ask spread),*

4. *at any time, the stock can take only two possible values in the next period.*

For the most part, the 1st and 2nd are true in reality but the 3rd assumption is not satisfied in practice. The 4th point becomes even more sophisticated in the Black-Scholes-Merton model where one assumes that the stock price is a geometric brownian motion, but empirical studies of stock price reveal this to also not be the case.

Definition 0.7: Derivative Security

In the general one-period model, we define a derivative security to be a security that pays some amount $V_1(H)$ at time one if the coin toss results in heads and pays a possibly different amount $V_1(T)$ at time one if the coin toss results in tail.

A European call option is a particular kind of derivative security.

Proposition 0.2: Delta Hedging Formula and Risk-Neutral Pricing Formula

Suppose that we want to determine the price V_0 for a derivative security at time zero. To do this, suppose that one has X_0 wealth and buys Δ_0 shares of stock at time zero. Then, the value of our portfolio at time one is

$$X_1 = \Delta_0 S_1 + (1+r)(X_0 - \Delta_0 S_0). \tag{0.9}$$

We want to choose X_0 and Δ_0 so that $X_1(H) = V_1(H)$ and $X_1(T) = V_1(T)$. The solutions to this are given by

$$\Delta_0 = \frac{V_1(H) - V_1(T)}{S_1(H) - S_1(T)}, \tag{0.10}$$

$$V_0 = \frac{1}{1+r}\left[\bar{p}V_1(H) + \bar{q}V_1(T)\right], \tag{0.11}$$

where

$$\bar{p} = \frac{1+r-d}{u-d}, \qquad \bar{q} = \frac{u-1-r}{u-d}. \tag{0.12}$$

Proof. Straightforward. We have a system of two linear equations with two unknowns:

$$V_1(H) = \Delta_0 S_1(H) + (1+r)(V_0 - \Delta_0 S_0) \tag{0.13}$$
$$V_1(T) = \Delta_0 S_1(T) + (1+r)(V_0 - \Delta_0 S_0) \tag{0.14}$$

Hence, we can compactly represent this by the linear transformation

$$\begin{pmatrix} V_1(H) \\ V_1(T) \end{pmatrix} = \begin{pmatrix} 1+r & S_1(H) - (1+r)S_0 \\ 1+r & S_1(T) - (1+r)S_0 \end{pmatrix} \begin{pmatrix} V_0 \\ \Delta_0 \end{pmatrix}. \tag{0.15}$$

We note that the determinant of the two by two matrix on the R.H.S above which we'll call A, is given by

$$\det(A) = (1+r)[S_1(T) - S_1(H)]. \tag{0.16}$$

From this one can see that no solutions exist if

1. $r = -1$,

2. $S_1(T) = S_1(H)$ (we assume that $V_1(H) \neq V_1(T)$).

Assuming quite reasonably that neither of these are true, one can find the solutions to V_0, Δ_0 to be

$$\Delta_0 = \frac{V_1(H) - V_1(T)}{S_1(H) - S_1(T)}, \tag{0.17}$$

$$V_0 = \frac{1}{1+r}\left[\frac{1+r-d}{u-d}V_1(H) + \frac{u-(1+r)}{u-d}V_1(T)\right], \tag{0.18}$$

where the above formula for Δ_0 is known as the *delta-hedging formula* and the formula for V_0 is known as the *risk-neutral pricing formula* at time zero for the one-period binomial model. □

Definition 0.8: Risk-Neutral Probabilities

The numbers $\bar{p}$ and $\bar{q}$ defined by

$$\bar{p} = \frac{1+r-d}{u-d}, \qquad \bar{q} = \frac{u-1-r}{u-d}, \tag{0.19}$$

are referred to as the risk-neutral probabilities. Due to the no-arbitrage condition on u, d, r, we have that $\bar{p}, \bar{q} \in [0,1]$ and $\bar{p} + \bar{q} = 1$, hence why one can have an interpretation of probabilities.

0.2   Multiperiod Binomial Model

We will now extend the ideas of Section §0.1 to multiple periods. We toss a coin repeatedly, and whenever we get a head the stock price moves *up* by the factor u, whereas whenever we get a tail, the stock price moves *down* by the factor d. In addition, we have a constant interest rate r and the only assumption we make on these parameters is the no-arbitrage condition (0.2).
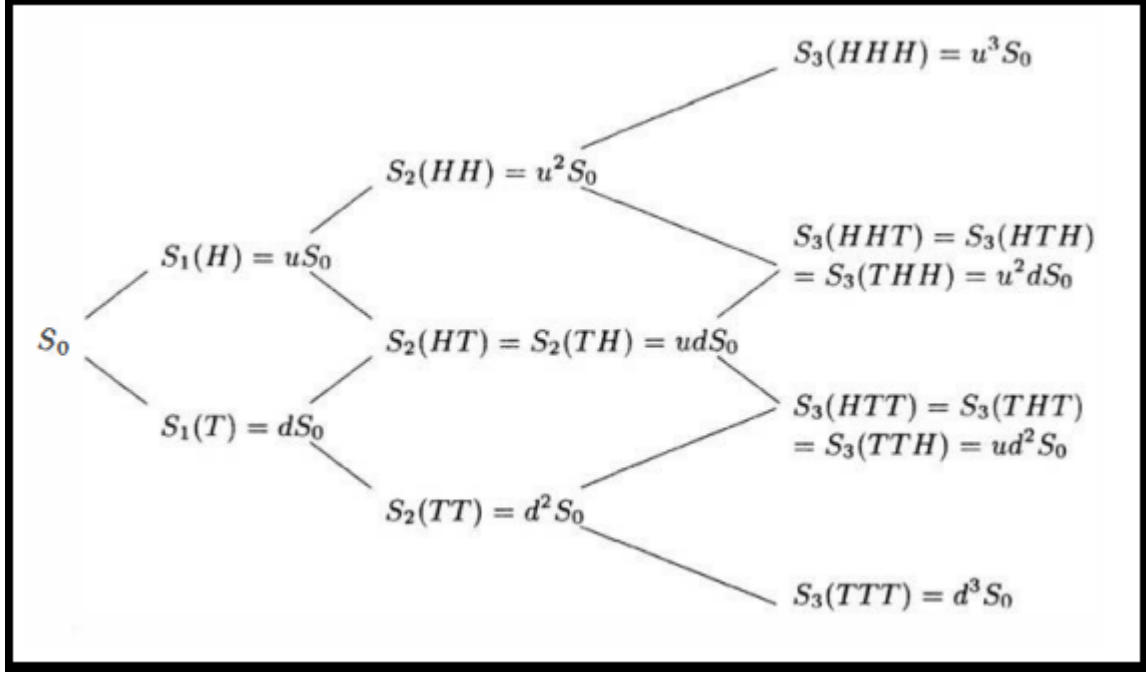
Figure 0.3: General Three-Period Model

We can see a *binomial tree* structure for the multiperiod binomial model above in Fig.0.3.

Proposition 0.3: Sequence Space

Let Ω_n be a random n-vector variable that signifies a sequence of n independent rolls for some random event that can take on $k \in \mathbb{N}$ possible values. The corresponding event space for Ω_n is given by all possible sequences of n rolls. It has a size of k^n.

Corollary 0.1: Binomial Event Space

For the binomial model, the space of all sequences of length n has a size of 2^n. In essence, there are 2^n unique coin toss sequences for a coin that is flipped n times.

Notation: Stock Price

Let S_n be the random variable associated with the stock price after n coin tosses. Let ω_i be the random variable associated with the ith coin toss result (that is, $\omega_i \in \{H, T\}$ $\forall 1 \leq i \leq n$). We write $S_n(\omega_1 \omega_2 ... \omega_n)$ as the stock price resulting from that particular sequence ω_i for $1 \leq i \leq n$.

Proposition 0.4: Stock Price in the Multiperiod Binomial Model

Consider a n-period binomial model. Suppose that one has a sequence of coin tosses $(\omega_1 ... \omega_n)$. Let h be the number of heads that occurred in this sequence. If the value of the stock price satisfies

$$S_j(\omega_1 .. \omega_{j-1} H) = u S_{j-1}(\omega_1 .. \omega_{j-1}), \tag{0.20}$$
$$S_j(\omega_1 .. \omega_{j-1} T) = d S_{j-1}(\omega_1 .. \omega_{j-1}), \tag{0.21}$$

for all $1 \le j \le n$, then we have that

$$S_n(\omega_1...\omega_n) = u^h d^{n-h} S_0. \tag{0.22}$$

Corollary 0.2: More Heads = More Money

If $u = 1/d$ and $d < 1$, then we require $h > n/2$ to have $S_n > S_0$. Whereas, if $h < n/2$, then $S_n < S_0$. This makes intuitive sense as it signifies that there had to be more heads than tails rolled for there to be an increase in the stock price.

Definition 0.9: Stochastic Process

We define a *stochastic process* as a sequence of random variables indexed by time.

Notation: Portfolio Random Variables

In general, we use the symbols Δ_n and X_n to represent the number of shares of stock held by the portfolio and the corresponding portfolio values, respectively, regardless of how the initial wealth X_0 and Δ_n are chosen. When X_0 and Δ_n are chosen to replicate a derivative security, we use the symbol V_n in place of X_n and call this the *(no-arbitrage) price of the derivative security at time n*.

Definition 0.10: Wealth Equation

Let X_n denote the wealth of a portfolio after n coin tosses; S_n denote the price of a stock after n coin tosses; and Δ_n denote the number of shares of a stock one wishes to hold onto after n coin tosses. Consider a portfolio strategy where each time you cash out your shares, you reconsider how many shares you want to own, and put the rest of your wealth into the money market at an interest rate of r. This can be described recursively by the *wealth equation*

$$X_{n+1} = \Delta_n S_{n+1} + (1+r)(X_n - \Delta_n S_n) \tag{0.23}$$

Theorem 0.1: Replication in the multiperiod binomial model

Consider an N-period binomial asset-pricing model, with $0 < d < 1 + r < u$, and with

$$\bar{p} = \frac{1+r-d}{u-d}, \quad \bar{q} = \frac{u-1-r}{u-d}. \tag{0.24}$$

let V_n be a random variable (a derivative security paying off at time N) depending on the first N coin tosses $\omega_1\omega_2...\omega_N$. Define recursively backward in time the sequence of random variables $V_{N-1}, V_{N-2}, .., V_0$ by

$$V_n(\omega_1...\omega_n) = \frac{1}{1+r}\left[\bar{p}V_{n+1}(\omega_1\omega_2...\omega_n H) + \bar{q}V_{n+1}(\omega_1\omega_2...\omega_n T)\right], \tag{0.25}$$

so that each V_n depends on the first n coin tosses $\omega_1\omega_2...\omega_n$, where n ranges between $N-1$ and 0. We next define

$$\Delta_n(\omega_1...\omega_n) = \frac{V_{n+1}(\omega_1...\omega_n H) - V_{n+1}(\omega_1...\omega_n T)}{S_{n+1}(\omega_1...\omega_n H) - S_{n+1}(\omega_1...\omega_n T)}, \tag{0.26}$$

where $0 \le n \le N-1$. If we set $X_0 = V_0$ and define recursively forward in time the portfolio values $X_1, X_2, .., X_N$ by Eq.(0.23), then we have

$$X_N(\omega_1...\omega_N) = V_N(\omega_1...\omega_N) \text{ for all } (\omega_1...\omega_N) \tag{0.27}$$

Proof. This can be solved via induction. We'll consider the base case and then the induction case, but the algebra for both are nearly identical.

Base Case: n = 0
Let ω_1 denote the random variable associated with the first heads or tails outcome. The corresponding wealth equation is given by

$$X_1(\omega_1) = \Delta_0 S_1(\omega_1) + (1+r)(X_0 - \Delta_0 S_0). \tag{0.28}$$

We use the fact that $X_0 = V_0$ and the identity provided for Δ_0. We'll first compute the second quantity:

$$(1+r)(X_0 - \Delta_0 S_0) = (1+r)\left(V_0 - \frac{V_1(H) - V_1(T)}{S_1(H) - S_1(T)}S_0\right) \tag{0.29}$$

$$= (1+r)\left[\frac{1}{1+r}\left(\frac{1+r-d}{u-d}V_1(H) + \frac{u-1-r}{u-d}V_1(T)\right) - \frac{V_1(H) - V_1(T)}{u-d}\right] \tag{0.30}$$

$$= V_1(T)\frac{u}{u-d} - V_1(H)\frac{d}{u-d}, \tag{0.31}$$

where we have used the fact that $S_1(H) = uS_0$ and $S_1(T) = dS_0$. Hence, we have that

$$X_1(\omega_1) = \frac{V_1(H) - V_1(T)}{S_1(H) - S_1(T)}S_1(\omega_1) + V_1(T)\frac{u}{u-d} - V_1(H)\frac{d}{u-d} \tag{0.32}$$

$$= \frac{1}{u-d}\left[V_1(H)\left(\frac{S_1(\omega_1)}{S_0} - d\right) + V_1(T)\left(u - \frac{S_1(\omega_1)}{S_0}\right)\right] \tag{0.33}$$

$$= \begin{cases} V_1(H) & \text{if } \omega_1 = H \\ V_1(T) & \text{if } \omega_1 = T \end{cases} \tag{0.34}$$

$$= V_1(\omega_1), \tag{0.35}$$

where we had to break the last line into two possible values, achieving the desired result.

Induction Case: Assume that it holds for $1 \leq n \leq k$
Let $\Omega_k := (\omega_1\omega_2...\omega_k)$ denote the random variable associated with the first k head/tail outcomes. The corresponding wealth equation for wealth after $k+1$ coin tosses is given by

$$X_{k+1}(\Omega_{k+1}) = \Delta_k(\Omega_k)S_{k+1}(\Omega_{k+1}) + (1+r)[X_k(\Omega_k) - \Delta_k(\Omega_k)S_k(\Omega_k)]. \tag{0.36}$$

By the induction hypothesis, we assume that $X_n(\Omega_n) = V_n(\Omega_n)$ $\forall 1 \leq n \leq k$ and $\forall \Omega_n$ and that the identities provided for Δ_n hold for all $1 \leq n \leq N$. Similar to the base case, we'll begin by computing the second quantity:

$$(1+r)[X_k(\Omega_k) - \Delta_k(\Omega_k)S_k(\Omega_k)] = (1+r)\left(V_k(\Omega_k) - \frac{V_{k+1}(\Omega_k H) - V_{k+1}(\Omega_k T)}{S_{k+1}(\Omega_k H) - S_{k+1}(\Omega_k T)}S_k(\Omega_k)\right) \tag{0.37}$$

$$= (1+r)\left[\frac{1}{1+r}\left(\frac{1+r-d}{u-d}V_{k+1}(\Omega_k H) + \frac{u-1-r}{u-d}V_{k+1}(\Omega_k T)\right) \right. \tag{0.38}$$

$$\left. - \frac{V_{k+1}(\Omega_k H) - V_{k+1}(\Omega_k T)}{u-d}\right] \tag{0.39}$$

$$= V_{k+1}(\Omega_k T)\frac{u}{u-d} - V_{k+1}(\Omega_k H)\frac{d}{u-d}, \tag{0.40}$$

where we have used the fact that $S_{k+1}(\Omega_k H) = uS_k(\Omega_k)$ and $S_{k+1}(\Omega_k T) = dS_k(\Omega_k)$. Hence, we have that

$$X_{k+1}(\Omega_{k+1}) = \frac{V_{k+1}(\Omega_k H) - V_{k+1}(\Omega_k T)}{S_{k+1}(\Omega_k H) - S_{k+1}(\Omega_k T)} S_{k+1}(\Omega_{k+1}) + V_{k+1}(\Omega_k T)\frac{u}{u-d} - V_{k+1}(\Omega_k H)\frac{d}{u-d} \quad (0.41)$$

$$= \frac{1}{u-d}\left[V_{k+1}(\Omega_k H)\left(\frac{S_{k+1}(\Omega_{k+1})}{S_k(\Omega_k)} - d\right) + V_{k+1}(\Omega_k T)\left(u - \frac{S_{k+1}(\Omega_{k+1})}{S_k(\Omega_k)}\right)\right] \quad (0.42)$$

$$= \begin{cases} V_{k+1}(\Omega_k H) & \text{if } \Omega_{k+1} = \Omega_k H \\ V_{k+1}(\Omega_k T) & \text{if } \Omega_{k+1} = \Omega_k T \end{cases} \quad (0.43)$$

$$= V_{k+1}(\Omega_{k+1}), \quad (0.44)$$

where just as for the base case, we had to break the final line into two possible values, achieving the desired result. $\square$

This theorem applies to both *path-dependent* options as well as to derivative securities whose payoff depends only on the final stock price.

Definition 0.11: Derivative Security Price

For $n = 1, 2, ..., N$, the random variable $V_n(\omega_1...\omega_n)$ in Theorem 0.1 is defined to be the price of the derivative security at time n if the outcomes of the first n tosses are $\omega_1, \omega_2, .., \omega_n$. the price of the derivative security at time zero is defined to be V_0.

Definition 0.12: Model Completeness

The multiperiod binomial model of this section is said to be **complete** because every derivative security can be replicated by trading in the underlying stock and the money market. In a complete market, every derivative security has a unique price that precludes arbitrage and this is the price of Definition 0.11.

Chapter 1

General Probability Theory

This chapter aims to shine light on how one can construct probability spaces and define key quantities in Probability through a measure-theoretic fashion. It should be noted that this is a *watered-down* version of measure theoretic probability but we will still dive into depths a bit deeper than an introductory / intermediate course on probability.

1.1 Infinite Probability Spaces

An infinite probability space is used to model situations in which a random experiment with infinitely many possible outcomes is conducted. For the purposes of this discussion, there are two such experiments to keep in mind

1. Choosing a number from the unit interval $[0, 1]$, and

2. Toss a coin infinitely many times.

Definition 1.1: σ-Algebra

Let Ω be a nonempty set, and let $\mathcal{F}$ be a collection of subsets of Ω. We say that $\mathcal{F}$ is a σ-algebra (called a σ-field by some authors) provided that it satisfies the following:

1. $\emptyset \in \mathcal{F}$

2. If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$

3. If a sequence of sets $A_1, A_2, ... \in \mathcal{F}$, then $\cup_{n=1}^{\infty} A_n \in \mathcal{F}$.

Definition 1.2: Measurable Space

Let X be a set and Σ be a σ-algebra defined over X. Then, the pair (X, Σ) is called a measurable space.

Definition 1.3: Measure

Let X be a set and Σ a σ-algebra over X. A function $\mu : \Sigma \to \mathbb{R} \cup \{-\infty, +\infty\}^a$ is called a **measure** if it satisfies the following properties:

(i) (Non-Negativity) For all $E \in \Sigma$, we have $\mu(E) \geq 0$.

(ii) (Null Empty Set) Let $\emptyset$ be the empty set. Then $\mu(\emptyset) = 0$.

(iii) (Countable Additivity) For all countable collection $\{E_k\}_{k=1}^{\infty}$ of pairwise disjoint sets in Σ,

$$\mu\big(\cup_{k=1}^{\infty} E_k\big) = \sum_{k=1}^{\infty} \mu(E_k) \tag{1.1}$$

*The triple (X, Σ, μ) is called a **measure space**.*

---

[a]$\mathbb{R} \cup \{-\infty, +\infty\}$ is known as the extended real line.

Definition 1.4: Probability Space

Let Ω be a nonempty set, and let $\mathcal{F}$ be a σ-algebra of subsets of Ω. A probability measure $\mathbb{P}$ is a function that, to every set $A \in \mathcal{F}$, assigns a number in $[0, 1]$, called the probability of A and is written $\mathbb{P}(A)$. We require that it satisfies

1. *$\mathbb{P}(\Omega) = 1$*

2. *(Countable Additivity) Let $A_1, A_2, ...$ be a sequence of disjoint sets in $\mathcal{F}$, then*

$$\mathbb{P}\big(\cup_{n=1}^{\infty} A_n\big) = \sum_{n=1}^{\infty} \mathbb{P}(A_n) \tag{1.2}$$

The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a probability space. A probability space is a measure space with total measure $\mathbb{P}(\Omega) = 1$.

Proposition 1.1: Probability Space Properties

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Some properties of the probability measure and σ-algebra are the following:

1. *(Closed under Finite Unions) Let $A_1, .., A_N \in \mathcal{F}$. Then $\cup_{n=1}^{N} A_n \in \mathcal{F}$.*

2. *(Finite Additivity) Let $A_1, .., A_N \in \mathcal{F}$ be disjoint sets. Then,*

$$\mathbb{P}\big(\cup_{n=1}^{N} A_n\big) = \sum_{n=1}^{N} \mathbb{P}(A_n) \tag{1.3}$$

3. *Let $A \in \mathcal{F}$ and A^c denote its complement. Then,*

$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A) \tag{1.4}$$

4. *(Closed under Finite Intersection) If $A_1, ..., A_N \in \mathcal{F}$, then $\cap_{n=1}^{N} A_n \in \mathcal{F}$.*

5. *(Closed under Difference) If $A, B \in \mathcal{F}$, then $A \setminus B \in \mathcal{F}$.*

Definition 1.5: Smallest σ-algebra

Let X be a set and $\mathcal{B}$ be a non-empty collection of subsets of X. The smallest σ-algebra containing all the sets of $\mathcal{B}$ is denoted

$$\sigma(\mathcal{B}), \tag{1.5}$$

and is called the sigma-algebra generated by the collection $\mathcal{B}$. The term smallest here means that any sigma-algebra containing the sets of $\mathcal{B}$ would have to contain all the sets of $\sigma(\mathcal{B})$ as well.

Definition 1.6: Borel σ-algebra

Consider the finite interval $[a, b] \subset \mathbb{R}$. The **Borel σ-algebra** of $[a, b]$ is the smallest σ-algebra that is generated by all closed sub-intervals of $[a, b]^a$ and we denote it by $\mathcal{B}[a, b]$. The sets in this σ-algebra are called **Borel sets**.

------

[a] In essence, this σ-algebra is obtained by beginning with closed intervals and adding everything else necessary in order to have a σ-algebra.

Our interest lies in the Borel σ-algebra $\mathcal{B}[0, 1]$. Every subset of $[0, 1]$ encountered in this text is a Borel set, and this can be verified by writing out the set in terms of unions, intersections, and complements of sequences of closed intervals.

Definition 1.7: Almost Surely

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. If a set $A \in \mathcal{F}$ satisfies $\mathbb{P}(A) = 1$, we say that the event A occurs *almost surely*.

Why *almost surely* rather than something like "surely"? Given some sample space Ω, one can have $\omega \in \Omega$ but $\omega \notin A$ so that $\mathbb{P}(A) = 1$. In essence, the probability measure of A^c is precisely zero: $\mathbb{P}(A^c) = 0$ without A having to be equal to the entire sample space (i.e $A \neq \Omega$).

1.2   Random Variables and Distributions

Definition 1.8: Random Variable

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A *random variable* is a real-valued function X defined on Ω with the property that for every Borel subset $B \subset \mathbb{R}$, the subset of Ω given by

$$\{X \in B\} := \{\omega \in \Omega; X(\omega) \in B\} \tag{1.6}$$

is in the σ-algebra $\mathcal{F}$. (We sometimes also permit a random variable to take the values $+\infty$ and $-\infty$.)

To get the Borel subsets of $\mathbb{R}$, one begins with the closed intervals $[a, b] \subset \mathbb{R}$ and adds all other sets that are necessary in order to have a σ-algebra. We denote the collection of Borel subsets of $\mathbb{R}$ by $\mathcal{B}(\mathbb{R})$ and call it the *Borel σ-algebra of $\mathbb{R}$*. Every subset of $\mathbb{R}$ that we encounter in this text is in this σ-algebra.

It is often the case that the probability that X takes a particular value is zero, and hence we shall mostly talk about the probability that X takes a value in some set rather than the probability that X takes a particular value. In other words, we will want to speak of $\mathbb{P}\{X \in B\}$. Definition 1.8 requires that $\{X \in B\}$ be in $\mathcal{F}$ for all $B \in \mathcal{B}(\mathbb{R})$, so that we ensure the probability of this set is defined.

Definition 1.9: Distribution Measure

Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The *distribution measure of X* is the probability measure μ_X that assigns to each Borel subset $B \subset \mathbb{R}$ the mass $\mu_X(B) = \mathbb{P}\{X \in B\}$

Random variables have distributions, but distributions and random variables are different concepts. Two different random variables can have the same distribution. A single random variable can have two different distributions.

Definition 1.10: Cumulative Distribution Function (CDF)

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let X be a random variable. We can describe the distribution of a random variable in terms of its *cumulative distribution function (cdf)* F_X defined as

$$F_X(x) := \mathbb{P}\{X \leq x\}, \ x \in \mathbb{R} \tag{1.7}$$

If we know the distribution measure μ_X, then we know the CDF F_X because $F_X(x) = \mu_X(-\infty, x]$.

Theorem 1.1: Nested Sequence Property

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $A_1, A_2, A_3, ...$ be a sequence of sets in $\mathcal{F}$.

(i) If $A_1 \subset A_2 \subset A_3 \subset ...$, then

$$\mathbb{P}(\cup_{k=1}^{\infty} A_k) = \lim_{n=\infty} \mathbb{P}(A_n) \tag{1.8}$$

(ii) If $A_1 \supset A_2 \supset A_3 \supset ...$, then

$$\mathbb{P}(\cap_{k=1}^{\infty} A_k) = \lim_{n=\infty} \mathbb{P}(A_n) \tag{1.9}$$

Definition 1.11: Density Function

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let X be a random variable. The distribution of a random variable can be recorded in more detail. A function $f_X : \mathbb{R} \to \mathbb{R}$ is said to be a density function for a random variable X if it is a nonnegative function such that

$$\mu_X[a, b] = \mathbb{P}\{a \leq X \leq b\} = \int_a^b f(x)dx, \ -\infty < a \leq b < \infty. \tag{1.10}$$

In particular, since closed intervals $[-n, n]$ have union $\mathbb{R}$, we must have

$$\int_{-\infty}^{\infty} f(x)dx = \lim_{n \to \infty} \int_{-n}^n f(x)dx = \lim_{n \to \infty} \mathbb{P}\{-n \leq X \leq n\} = \mathbb{P}\{X \in \mathbb{R}\} = \mathbb{P}(\Omega) = 1 \tag{1.11}$$

Proposition 1.2: Singleton Borel Sets

Let $y \in \mathbb{R}$. Then, $\exists x, z \in \mathbb{R}$ such that $x < y$ and $y < z$. It's clear that $[x, y]$ and $[y, z]$ are closed intervals. Then, one can construct the singleton set

$$\{y\} = [x, y] \cap [y, z]. \tag{1.12}$$

Hence, $\{y\} \in \mathcal{B}(\mathbb{R}) \ \forall y \in \mathbb{R}$.

Definition 1.12: Probability Mass Function

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A *probability mass function* is a function in which there is a finite sequence of numbers $x_1, x_2, .., x_N$ or an infinite countable sequence $x_1, x_2, ...$ such that with probability one the random variable X takes one of the values in the sequence. We then define $p_i = \mathbb{P}\{X = x_i\}$. Each p_i is nonnegative, and $\sum_i p_i = 1$. The mass assigned to a Borel set $B \subset \mathbb{R}$ by the distribution measure of X is

$$\mu_X(B) = \sum_{\{i; x_i \in B\}} p_i, \ B \in \mathcal{B}(\mathbb{R}) \tag{1.13}$$

Definition 1.13: Standard Normal Random Variable

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let

$$\psi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \tag{1.14}$$

be the *standard normal density*, and define the *cumulative normal distribution function*

$$N(x) = \int_{-\infty}^{x} \psi(\xi) d\xi. \tag{1.15}$$

The function $N(x)$ is strictly increasing, mapping $\mathbb{R}$ onto $(0, 1)$ and so has a strictly increasing inverse function $N^{-1}(y)$. In essence, $N(N^{-1}(y)) = y$ for all $y \in (0, 1)$. Let Y be a uniformly distributed random variable, defined on $(\Omega, \mathcal{F}, \mathbb{P})$, and set $X = N^{-1}(Y)$. Whenever $-\infty < a \le b < \infty$, we have

$$\mu_X[a, b] = \mathbb{P}\{\omega \in \Omega; a \le X(\omega) \le b\} \tag{1.16}$$

$$= \mathbb{P}\{\omega \in \Omega; a \le N^{-1}(Y(\omega)) \le b\} \tag{1.17}$$

$$= \mathbb{P}\{\omega \in \Omega; N(a) \le N(N^{-1}(Y(\omega))) \le N(b)\} \tag{1.18}$$

$$= \mathbb{P}\{\omega \in \Omega; N(a) \le Y(\omega) \le N(b)\} \tag{1.19}$$

$$= N(b) - N(a) \tag{1.20}$$

$$= \int_a^b \psi(x) dx. \tag{1.21}$$

The measure μ_X on $\mathbb{R}$ given by this formula is called the *standard normal distribution*. Any random variable that has this distribution, regardless of the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which it is defined, is called a **standard normal random variable**.

1.3   Expectations

Definition 1.14: Lebesgue Integral of Random Variable

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let X be a random variable defined on this space. Assume that $0 \le X(\omega) < \infty$ for every $\omega \in \Omega$, and let $\Pi = \{y_0, y_1, y_2, ...\}$, where $0 = y_0 < y_1 < y_2 < ....$. For each subinterval $[y_k, y_{k+1}]$, we set

$$A_k = \{\omega \in \Omega; y_k \le X(\omega) < y_{k+1}\} \tag{1.22}$$

We define the lower Lebesgue sum to be

$$LS_\Pi^-(X) = \sum_{k=1}^{\infty} y_k \mathbb{P}(A_k). \tag{1.23}$$

We define $||\Pi|| := \max_{1 \le k \le n} (y_k - y_{k-1})$ as the length of the longest subinterval of the partition. This lower sum converges as $||\Pi||$, the maximal distance between the y_k partition points, approaches zero, and we define this limit to be the **Lebesgue integral** $\int_\Omega X(\omega) d\mathbb{P}(\omega)$, or simply $\int_\Omega X d\mathbb{P}$.

We assumed that $0 \le X(\omega) < \infty$ for every $\omega \in \Omega$. If the set of ω that violates this condition has zero probability, there is no effect on the integral we just defined. If $\mathbb{P}\{\omega; X(\omega) \ge 0\} = 1$ but $\mathbb{P}\{\omega; X(\omega) = \infty\} > 0$, then we define $\int_\Omega X(\omega) d\mathbb{P}(\omega) = \infty$.

Definition 1.15: Positive and Negative Parts of Random Variable

Consider a random variable X that can take on both positive and negative values. For such a random variable, we define the *positive* and *negative parts of X* by

$$X^+(\omega) = \max\{X(\omega), 0\}, \ X^-(\omega) = \max\{-X(\omega), 0\}. \tag{1.24}$$

We note that both X^+ and X^- are nonnegative random variables, $X = X^+ - X^-$, and $|X| = X^+ + X^-$. Both $\int_\Omega X^+(\omega)d\mathbb{P}(\omega)$ and $\int_\Omega X^-(\omega)d\mathbb{P}(\omega)$ are defined by this procedure and provided they are not both ∞, we can define

$$\int_\Omega X(\omega)d\mathbb{P}(\omega) = \int_\Omega X^+(\omega)d\mathbb{P}(\omega) - \int_\Omega X^-(\omega)d\mathbb{P}(\omega) \tag{1.25}$$

Definition 1.16: Integrable Random Variable

Let X be a random variable and X^+ and X^- denote its positive and negative parts respectively.

- If $\int_\Omega X^+(\omega)d\mathbb{P}(\omega)$ and $\int_\Omega X^-(\omega)d\mathbb{P}(\omega)$ are both finite, we say that X is *integrable*, and $\int_\Omega X(\omega)d\mathbb{P}(\omega)$ is also finite.

- If $\int_\Omega X^-(\omega)d\mathbb{P}(\omega)$ is finite but $\int_\Omega X^+(\omega)d\mathbb{P}(\omega) = \infty$, then $\int_\Omega X(\omega)d\mathbb{P}(\omega) = \infty$.

- If $\int_\Omega X^+(\omega)d\mathbb{P}(\omega)$ is finite and $\int_\Omega X^-(\omega)d\mathbb{P}(\omega) = \infty$, then $\int_\Omega X(\omega)d\mathbb{P}(\omega) = -\infty$.

- If both $\int_\Omega X^+(\omega)d\mathbb{P}(\omega) = \infty$ and $\int_\Omega X^-(\omega)d\mathbb{P}(\omega) = \infty$, then $\int_\Omega X(\omega)d\mathbb{P}(\omega)$ is not defined.

Theorem 1.2

Let X and Y be random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

(i) If X takes on finitely many values $y_0, y_1, y_2, ..., y_n$, then

$$\int_\Omega X(\omega)d\mathbb{P}(\omega) = \sum_{k=0}^n y_k \mathbb{P}\{X = y_k\}. \tag{1.26}$$

(ii) *(Integrability)* The random variable X is integrable if and only if

$$\int_\Omega |X(\omega)|d\mathbb{P}(\omega) < \infty \tag{1.27}$$

(iii) *(Comparison)* If $X \leq Y$ almost surely (i.e., $\mathbb{P}\{X \leq Y\} = 1$), and if $\int_\Omega X(\omega)d\mathbb{P}(\omega)$ and $\int_\Omega Y(\omega)d\mathbb{P}(\omega)$ are defined, then

$$\int_\Omega X(\omega)d\mathbb{P}(\omega) \leq \int_\Omega Y(\omega)d\mathbb{P}(\omega). \tag{1.28}$$

In particular, if $X = Y$ almost surely and one of the integrals is defined then they are both defined and

$$\int_\Omega X(\omega)d\mathbb{P}(\omega) = \int_\Omega Y(\omega)d\mathbb{P}(\omega) \tag{1.29}$$

(iv) *(Linearity)* If $\alpha, \beta \in \mathbb{R}$ and X, Y are integrable, or if α, β are nonnegative constants and X, Y are nonnegative, then

$$\int_\Omega (\alpha X(\omega) + \beta Y(\omega))d\mathbb{P}(\omega) = \alpha \int_\Omega X(\omega)d\mathbb{P}(\omega) + \beta \int_\Omega Y(\omega)d\mathbb{P}(\omega). \tag{1.30}$$

Definition 1.17: Indicator Function

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. We often want to integrate a random variable X over a subset $A \subset \Omega$ rather than over all of Ω. For this reason, we define

$$\int_A X(\omega) d\mathbb{P}(\omega) = \int_\Omega \mathbb{I}_A(\omega) X(\omega) d\mathbb{P}(\omega) \text{ for all } A \in \mathcal{F}, \tag{1.31}$$

where $\mathbb{I}_A$ is the *indicator function* (random variable) given by

$$\mathbb{I}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \notin A. \end{cases} \tag{1.32}$$

Definition 1.18: Expectation

Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The expectation (or expected value) of X is defined to be

$$E[X] = \int_\Omega X(\omega) d\mathbb{P}(\omega). \tag{1.33}$$

This definition makes sense if X is integrable, i.e; if

$$E|X| = \int_\Omega |X(\omega)| d\mathbb{P}(\omega) < \infty \tag{1.34}$$

or if $X \geq 0$ almost surely. In the latter case, $E[X]$ might be ∞.

Theorem 1.3: Reformulating Theorem 1.2 in Expectations

Let X and Y be random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

(i) If X takes on finitely many values $x_0, x_1, x_2, ..., x_n$, then

$$E[X] = \sum_{k=0}^n x_k \mathbb{P}\{X = x_k\}. \tag{1.35}$$

In particular, if Ω is finite, then

$$E[X] = \sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\omega) \tag{1.36}$$

(ii) (Integrability) The random variable X is integrable if and only if

$$E[X] < \infty \tag{1.37}$$

(iii) (Comparison) If $X \leq Y$ almost surely and X and Y are integable or almost surely nonnegative, then

$$E[X] \leq E[Y]. \tag{1.38}$$

In particular, if $X = Y$ almost surely and one of the random variables is integrable or almost surely nonnegative, then they are both integrable or almost surely nonnegative, respectively, and

$$E[X] = E[Y] \tag{1.39}$$

(iv) (Linearity) If $\alpha, \beta \in \mathbb{R}$ and X, Y are integrable, or if α, β are nonnegative constants and X, Y are nonnegative, then

$$E[\alpha X + \beta Y] = \alpha E[X] + \beta E[Y] \tag{1.40}$$

(v) *(Jensen's Inequality) If ψ is a convex, real-valued function defined on $\mathbb{R}$, and if $E|X| < \infty$, then*

$$\psi(E[X]) \leq E[\psi(X)] \tag{1.41}$$

Definition 1.19: Lebesgue Measure

Let $\mathcal{B}(\mathbb{R})$ be the σ-algebra of Borel subsets of $\mathbb{R}$. The Lebesgue measure on $\mathbb{R}$, which we denote by $\mathcal{L}$, assigns to each set $B \in \mathcal{B}(\mathbb{R})$ a number in $[0, \infty)$ or the value ∞ so that

(i) *$\mathcal{L}[a, b] = b - a$ whenever $a \leq b$, and*

(ii) *If $B_1, B_2, B_3, ...$ is a sequence of disjoint sets in $\mathcal{B}(\mathbb{R})$, then we have the countable additivity property*

$$\mathcal{L}\left(\cup_{n=1}^{\infty} B_n \right) = \sum_{n=1}^{\infty} \mathcal{L}(B_n) \tag{1.42}$$

Definition 1.20: Borel Measurable

*Let $f : \mathbb{R} \to \mathbb{R}$ be a real valued function. If f has the property that $\forall B \in \mathcal{B}(\mathbb{R})$, the set $\{x; f(x) \in B\} \in \mathcal{B}(\mathbb{R})$, then f is said to be **Borel Measurable**.*

Every continuous and piecewise continuous function is Borel measurable.

Proposition 1.3: Borel Measurable Properties

Let $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ be two Borel-measurable functions. Then,

(i) *$fg : \mathbb{R} \to \mathbb{R}$ and*

(ii) *$f + g : \mathbb{R} \to \mathbb{R}$*

are Borel-measurable.

Definition 1.21: Constructing the Lebesgue Integral

Let $f : \mathbb{R} \to \mathbb{R}$ be a Borel measurable function. We wish to construct the Lebesgue integral $\int_{\mathbb{R}} f(x) d\mathcal{L}(x)$ of f over $\mathbb{R}$. We assume for the moment that $0 \leq f(x) < \infty \ \forall x \in \mathbb{R}$. We choose a partition $\Pi = \{y_0, y_1, y_2, ..\}$, where $0 = y_0 < y_1 < y_2 < ...$ For each subinterval $[y_k, y_{k+1})$, we define

$$B_k = \{x \in \mathbb{R}; y_k \leq f(x) < y_{k+1}\} \tag{1.43}$$

Since $[y_k, y_{k+1}] \in \mathcal{B}(\mathbb{R})$, and f is Borel measurable, then $\{x; f(x) \in [y_k, y_{k+1}]\} \in \mathcal{B}(\mathbb{R})$, hence $B_k \in \mathcal{B}(\mathbb{R}) \ \forall k$. This means that their Lebesgue measures are all defined. We define the lower Lebesgue sum as

$$LS_{\overline{\Pi}}^{-}(f) = \sum_{k=1}^{\infty} y_k \mathcal{L}(B_k). \tag{1.44}$$

As $||\Pi||$ converges to zero, these lower Lebesgue sums will converge to a limit, which we define to be $\int_{\mathbb{R}} f(x) d\mathcal{L}(x)$. It is possible that this integral gives the value ∞.

Definition 1.22: Constructing the Lebesgue Integral for General Functions

We had assumed that $0 \leq f(x) < \infty$ for every $x \in \mathbb{R}$. If the set of x where the condition is violated has zero Lebesgue measure, the Lebesgue integral of f is not affected. If $\mathcal{L}\{x \in \mathbb{R}; f(x) < 0\} = 0$ and $\mathcal{L}\{x \in \mathbb{R}; f(x) = \infty\} > 0$, we define $\int_{\mathbb{R}} f(x)d\mathcal{L}(x) = \infty$.

We now consider the case that f takes on both positive and negative values and define

$$f^+(x) = \max\{f(x), 0\}, \qquad f^-(x) = \max\{-f(x), 0\}. \tag{1.45}$$

Since f^+ and f^- are nonnegative, $\int_{\mathbb{R}} f^+(x)d\mathcal{L}(x)$ and $\int_{\mathbb{R}} f^-(x)d\mathcal{L}(x)$ are defined by the procedure described above. We then define

$$\int_{\mathbb{R}} f(x)d\mathcal{L}(x) = \int_{\mathbb{R}} f^+(x)d\mathcal{L}(x) - \int_{\mathbb{R}} f^-(x)d\mathcal{L}(x), \tag{1.46}$$

provided that this is not $\infty - \infty$. In the case where both $\int_{\mathbb{R}} f^+(x)d\mathcal{L}(x)$ and $\int_{\mathbb{R}} f^-(x)d\mathcal{L}(x)$ are infinite, $\int_{\mathbb{R}} f(x)d\mathcal{L}(x)$ is not defined.

Definition 1.23: Lebesgue Integrable Function

Let $f : \mathbb{R} \to \mathbb{R}$ and f^+, f^- denote its positive and negative parts respectively. If $\int_{\mathbb{R}} f^+(x)d\mathcal{L}(x)$ and $\int_{\mathbb{R}} f^-(x)d\mathcal{L}(x)$ are finite, we say that f is integrable. This is equivalent to the condition that $\int_{\mathbb{R}} |f(x)|d\mathcal{L}(x) < \infty$.

Proposition 1.4: Lebesgue Integral over Finitely Many Values

Let $f : \mathbb{R} \to \mathbb{R}$ take on only finitely many values $\{y_0, y_1, ..., y_n\}$. If f is Lebesgue integrable, then

$$\int_{\mathbb{R}} f(x)d\mathcal{L}(x) = \sum_{k=0}^{n} y_k \mathcal{L}\{x \in \mathbb{R}; f(x) = y_k\} \tag{1.47}$$

Definition 1.24: Lebesgue Integral over Subset of $\mathbb{R}$

Sometimes, we have a function $f(x)$ defined for every $x \in \mathbb{R}$ but want to compute its Lebesgue integral over only part of $\mathbb{R}$, say some set $B \in \mathcal{B}(\mathbb{R})$. We define

$$\int_{B} f(x)d\mathcal{L}(x) = \int_{\mathbb{R}} \mathbb{I}_B(x)f(x)d\mathcal{L}(x), \tag{1.48}$$

where

$$\mathbb{I}_B(x) = \begin{cases} 1 & \text{if } x \in B, \\ 0 & \text{if } x \notin B \end{cases} \tag{1.49}$$

Theorem 1.4: Comparison of Riemann and Lebesgue Integrals

Let f be a bounded function defined on $\mathbb{R}$, and let $a < b$ be numbers.

(i) The Riemann integral $\int_a^b f(x)dx$ is defined (i.e, the lower and upper Riemann sums converge to the same limit) if and only if the set of points $x \in [a, b]$ where $f(x)$ is not continuous has Lebesgue measure zero.

(ii) If the Riemann integral $\int_a^b f(x)dx$ is defined, then f is Borel measurable (so the Lebesgue integral $\int_{[a,b]} f(x)d\mathcal{L}(x)$ is also defined), and the Riemann and Lebesgue integrals agree.

Definition 1.25: Property Holds Almost Everywhere

If the set of numbers in $\mathbb{R}$ that fail to have some property is a set with Lebesgue measure zero, we say that the property holds *almost everywhere*.

We can restate Theorem 1.4 as:

The Riemann integral $\int_a^b f(x)dx$ exists if and only if $f(x)$ is almost everywhere continuous on $[a, b]$.

Since Riemann and Lebesgue integrals agree whenever the Riemann integral is defined, we shall use the more familiar notation $\int_a^b f(x)dx$ to denote the Lebesgue integral. If the set B over which we wish to integrate is not an interval, we shall write $\int_B f(x)dx$.

1.4   Convergence of Integrals

Definition 1.26: Convergent Sequence (Metric Space)

Let (X, d) be a metric space. A sequence $\{p_n\}$ in a metric space X is said to *converge* if there is a point $p \in X$ with the following property: For every $\epsilon > 0$, there exists an $N \in \mathbb{N}$ such that $n \geq N$ implies that $d(p_n, p) < \epsilon$. In this case, we say that $\{p_n\}$ converges to p, or that p is the limit of $\{p_n\}$, and we write:

$$p_n \to p, \quad \text{or} \quad \lim_{n \to \infty} p_n = p \tag{1.50}$$

If $\{p_n\}$ doesn't converge, then it is said to *diverge*.

Definition 1.27: Converging Almost Surely

Let $X_1, X_2, \ldots$ be a sequence of random variables, all defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let X be another random variable defined on this space. We say that $X_1, X_2, \ldots$ converges to X almost surely and write

$$\lim_{n \to \infty} X_n = X \text{ almost surely} \tag{1.51}$$

if the set of $\omega \in \Omega$ for which the sequence of numbers $X_1(\omega), X_2(\omega), \ldots$ has limit $X(\omega)$ is a set with probability one. Equivalently, the set of $\omega \in \Omega$ for which the sequence of numbers $X_1(\omega), X_2(\omega), \ldots$ does not converge to $X(\omega)$ is a set with probability zero.

Definition 1.28: Converging Almost Everywhere

Let $f_1, f_2, \ldots$ be a sequence of real-valued, Borel-measurable functions defined on $\mathbb{R}$. Let f be another real-valued, Borel-measurable function defined on $\mathbb{R}$. We say that $f_1, f_2, \ldots$ converges to f almost everywhere and write

$$\lim_{n \to \infty} f_n = f \text{ almost everywhere} \tag{1.52}$$

if the set of $x \in \mathbb{R}$ for which the sequence of numbers $f_1(x), f_2(x), \ldots$ does not have limit $f(x)$ is a set with Lebesgue measure zero.

Example 1.1: Normal Densities Converging Almost Everywhere

Consider a sequence of normal densities, each with mean zero and the nth having variance $1/n$:

$$f_n(x) = \sqrt{\frac{n}{2\pi}} e^{-\frac{nx^2}{2}}. \tag{1.53}$$

If $x \neq 0$, then $\lim_{n\to\infty} f_n(x) = 0$, but

$$\lim_{n\to\infty} f_n(0) = \lim_{n\to\infty} \sqrt{\frac{n}{2\pi}} = \infty \tag{1.54}$$

Therefore, the sequence $f_1, f_2, ...$ converges everywhere to the function

$$f^*(x) = \begin{cases} 0 & \text{if } x \neq 0 \\ \infty & \text{if } x = 0 \end{cases} \tag{1.55}$$

and converges almost everywhere to the identically zero function $f(x) = 0$ for all $x \in \mathbb{R}$.

Proposition 1.5: Function Convergence $\neq$ Lebesgue Integral Convergence

Suppose that $f_1, f_2, ..$ are a sequence of functions that converge to a function f. Then in general one has that

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} f_n(x)dx \neq \int_{-\infty}^{\infty} \lim_{n\to\infty} f_n(x)dx. \tag{1.56}$$

Proof. We can show this by way of counterexample. Consider the previous example where we had defined a sequence

$$f_n(x) = \sqrt{\frac{n}{2\pi}} e^{-\frac{nx^2}{2}}. \tag{1.57}$$

We note that

$$\int_{-\infty}^{\infty} f_n(x)\mathrm{d}x = 1, \tag{1.58}$$

however

$$\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 0, \tag{1.59}$$

as we have that

$$f(x) = \begin{cases} 0 & \text{if } x \neq 0 \\ \infty & \text{if } x = 0 \end{cases} \tag{1.60}$$

where the Lebesgue measure of the set $\{0\}$ is simply zero. Hence, one has that $\int_{\mathbb{R}} f(x)\mathrm{d}\mathcal{L}(x) = 0$. $\qquad \square$

Theorem 1.5: Monotone Convergence

Let $X_1, X_2, ...$ be a sequence of random variables converging almost surely to another random variable X. If

$$0 \leq X_1 \leq X_2 \leq X_3 \leq ... \text{ almost surely}, \tag{1.61}$$

then

$$\lim_{n\to\infty} E[X_n] = E[X] \tag{1.62}$$

Let $f_1, f_2, ...$ be a sequence of Borel-measurable functions on $\mathbb{R}$ converging almost everywhere to a function

f. If

$$0 \leq f_1 \leq f_2 \leq f_3 \leq ...almost\ everywhere, \tag{1.63}$$

then

$$\lim_{n \to \infty} \int_{-\infty}^{\infty} f_n(x)dx = \int_{-\infty}^{\infty} f(x)dx. \tag{1.64}$$

Corollary 1.1

Suppose that the nonnegative random variable X takes countably many values $x_0, x_1, ...$. Then,

$$E[X] = \sum_{k=0}^{\infty} x_k \mathbb{P}\{X = x_k\} \tag{1.65}$$

Theorem 1.6: Dominated Convergence

Let $X_1, X_2, ...$ be a sequence of random variables converging almost surely to a random variable X. If there is another random variable Y such that $E[Y] < \infty$ and $|X_n| \leq Y$ almost surely for every n, then

$$\lim_{n \to \infty} E[X_n] = E[X] \tag{1.66}$$

Let $f_1, f_2, ...$ be a sequence of Borel-measurable functions on $\mathbb{R}$ converging almost everywhere to a function f. If there is another function g such that $\int_{-\infty}^{\infty} g(x) < \infty$ and $|f_n| \leq g$ almost everywhere for every n, then

$$\lim_{n \to \infty} \int_{-\infty}^{\infty} f_n(x)dx = \int_{-\infty}^{\infty} f(x)dx. \tag{1.67}$$

1.5 Computation of Expectations

Theorem 1.7

Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let μ_X denote the distribution measure for X. Let g be a Borel-measurable function on $\mathbb{R}$. Then

$$E|g(X)| = \int_{\mathbb{R}} |g(x)|d\mu_X(x), \tag{1.68}$$

and if this quantity is finite, then

$$E[g(X)] = \int_{\mathbb{R}} g(x)d\mu_X(x). \tag{1.69}$$

Theorem 1.8

Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let g be a Borel-measurable function on $\mathbb{R}$. Suppose that X has a density f.[a] Then

$$E|g(X)| = \int_{-\infty}^{\infty} |g(x)|f(x)dx. \tag{1.71}$$

If this quantity is finite, then

$$E[g(X)] = \int_\infty^\infty g(x)f(x)dx \tag{1.72}$$

---

[a]Signifying that f is a function satisfying

$$\mu_X(B) = \int_B f(x)\mathrm{d}x \ \forall B \in \mathcal{B}(\mathbb{R}) \tag{1.70}$$

1.6 Change of Measure

Theorem 1.9

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let Z be an almost surely nonnegative random variable with $E[Z] = 1$. For $A \in \mathcal{F}$, we define

$$\tilde{\mathbb{P}}(A) = \int_A Z(\omega)d\mathbb{P}(\omega). \tag{1.73}$$

Then $\tilde{\mathbb{P}}$ is a probability measure. Furthermore, if X is a nonnegative random variable, then[a]

$$\tilde{E}[X] = E[XZ]. \tag{1.75}$$

If Z is almost surely strictly positive, we also have

$$E[Y] = \tilde{E}\left[\frac{Y}{Z}\right] \tag{1.76}$$

for every nonnegative random variable Y.

---

[a]The $\tilde{E}$ is an expectation under the probability measure $\tilde{\mathbb{P}}$:

$$\tilde{E}[X] = \int_\Omega X(\omega)\tilde{\mathbb{P}}(\omega) \tag{1.74}$$

Definition 1.29: Equivalent Probability Measures

Let Ω be a nonempty set and $\mathcal{F}$ a σ-algebra of subsets of Ω. Two probability measures $\mathbb{P}$ and $\tilde{\mathbb{P}}$ on $(\Omega, \mathcal{F})$ are said to be equivalent if they agree which sets in $\mathcal{F}$ have probability zero.

In financial models, we will first set up a sample space Ω, which one can regard as the set of possible scenarios for the future. We imagine this set of possible scenarios has an actual probability measure $\mathbb{P}$. However, for purposes of pricing derivative securities, we will use a risk-neutral measure $\tilde{P}$. We will insist that these two measures are equivalent.

Definition 1.30: Radon-Nikodym Derivative

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\tilde{\mathbb{P}}$ be another probability measure on $(\Omega, \mathcal{F})$ that is equivalent to $\mathbb{P}$, and let Z be an almost surely positive random variable that relates $\mathbb{P}$ and $\tilde{\mathbb{P}}$ via $\tilde{\mathbb{P}}(A) = \int_A Z(\omega)d\mathbb{P}(\omega)$. Then Z is called the Radon-Nikodym derivative of $\tilde{\mathbb{P}}$ with respect to $\mathbb{P}$, and we write

$$Z = \frac{d\tilde{\mathbb{P}}}{d\mathbb{P}} \tag{1.77}$$

Theorem 1.10: Radon-Nikodym

Let $\mathbb{P}$ and $\tilde{\mathbb{P}}$ be equivalent probability measures defined on $(\Omega, \mathcal{F})$. Then there exists an almost surely positive random variable Z such that $E[Z] = 1$ and

$$\tilde{\mathbb{P}}(A) = \int_A Z(\omega) d\mathbb{P}(\omega) \text{ for every } A \in \mathcal{F}. \tag{1.78}$$

Chapter 2

Information and Conditioning

2.1 Information and σ-algebras

Definition 2.1: Filtration

Let Ω be a nonempty set. Let T be a fixed positive number, and assume that for each $t \in [0, T]$ there is a σ-algebra $\mathcal{F}(t)$. Assume further that if $s \leq t$, then every set in $\mathcal{F}(s)$ is also in $\mathcal{F}(t)$. Then we call the collection of σ-algebras $\mathcal{F}(t), 0 \leq t \leq T$, a *filtration*.

A filtration tells us the information we will have at future times. More precisely, when we get to time t, we will know for each set in $\mathcal{F}(t)$ whether the true ω lies in that set.

Definition 2.2: σ-Algebra Generated By Random Variable

Let X be a random variable defined a nonempty sample space Ω. The σ-algebra generated by X, denoted by $\sigma(X)$, is the collection of all subsets of Ω of the form $\{X \in B\}$, where B ranges over the Borel subsets of $\mathbb{R}$.

Definition 2.3: σ-Algebra Measurable (Random Variable)

Let X be a random variable defined on a nonempty sample space Ω. Let $\mathcal{G}$ be a σ-algebra of subsets of Ω. If every set in $\sigma(X)$ is also in $\mathcal{G}$, we say that X is $\mathcal{G}$-measurable.

Proposition 2.1

A random variable X is $\mathcal{G}$-measurable if and only if the information in $\mathcal{G}$ is sufficient to determine the value of X. If X is $\mathcal{G}$-measurable, then $f(X)$ is also $\mathcal{G}$-measurable for any Borel-measurable function f; if the information in $\mathcal{G}$ is sufficient to determine the value of X, it will also determine the value of $f(X)$.

Definition 2.4: Adapted Stochastic Process

Let Ω be a nonempty sample space equipped with a filtration $\mathcal{F}(t), 0 \leq t \leq T$. Let $X(t)$ be a collection of random variables index by $t \in [0, T]$. We say this collection of random variables is an *adapted stochastic process* if, for each t, the random variable $X(t)$ is $\mathcal{F}(t)$-measurable.

2.2   Independence

When a random variable is measurable with respect to a σ-algebra $\mathcal{G}$, the information contained in $\mathcal{G}$ is sufficient to determine the value of the random variable. The other extreme is when a random variable is independent of a σ-algebra. Independence is the subject of the present section.

Definition 2.5: Independent Events

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. We say that two sets A and B in $\mathcal{F}$ are independent if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B) \tag{2.1}$$

Definition 2.6: Independent σ-Algebras

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $\mathcal{G}$ and $\mathcal{H}$ be sub-σ-algebras of $\mathcal{F}$ (i.e., the sets in $\mathcal{G}$ and the sets in $\mathcal{H}$ are also in $\mathcal{F}$). We say these two σ-algebras are independent if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B) \ \ \forall A \in \mathcal{G}, B \in \mathcal{H} \tag{2.2}$$

Definition 2.7: Independent Random Variables

Let X and Y be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. We say these two random variables are *independent* if the σ-algebras they generate, $\sigma(X)$ and $\sigma(Y)$, are independent.

Definition 2.8: Random Variable Independence of σ-Algebra

Let X be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. We say that X is independent of a σ-algebra $\mathcal{G}$ if $\sigma(X)$ and $\mathcal{G}$ are independent.

Definition 2.9: Independent Sequence of σ-Algebras

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3, \ldots$ be a sequence of sub-σ-algebras of $\mathcal{F}$. For a fixed positive integer n, we say that the n σ-algebras $\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_n$ are independent if

$$\mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_n) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2) \cdots \mathbb{P}(A_n) \ \ \forall A_1 \in \mathcal{G}_1, A_2 \in \mathcal{G}_2, \ldots, A_n \in \mathcal{G}_n. \tag{2.3}$$

We say that the full sequence of σ-algebras $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3, \ldots$ is independent if, for every positive integer n, the n σ-algebras $\mathcal{G}_1, \mathcal{G}_2, \ldots \mathcal{G}_n$ are independent.

Definition 2.10: Independent Sequence of Random Variables

Let $X_1, X_2, X_3, \ldots$ be a sequence of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. We say the n random variables $X_1, X_2, \ldots, X_n$ are independent if the σ-algebras $\sigma(X_1), \sigma(X_2), \ldots, \sigma(X_n)$ are independent. We say that the full sequence of random variables $X_1, X_2, X_3, \ldots$ is independent if, for every positive integer n, the n random variables $X_1, X_2, X_3, \ldots$ are independent.

Theorem 2.1

Let X and Y be independent random variables, and let f and g be Borel-measurable functions on $\mathbb{R}$. Then $f(X)$ and $g(Y)$ are independent random variables.

Definition 2.11: Joint Distribution Measure Joint Distribution Measure

Let X and Y be random variables. The pair of random variables (X, Y) takes values in the plane $\mathbb{R}^2$, and the joint distribution measure of (X, Y) is given by

$$\mu_{X,Y}(C) = \mathbb{P}\{(X, Y) \in C\} \text{ for all Borel sets } C \subset \mathbb{R}^2. \tag{2.4}$$

This is a probability measure (i.e., a way of assigning measure between 0 and 1 to subsets of $\mathbb{R}^2$ so that $\mu_{X,Y}(\mathbb{R}^2) = 1$ and the countable additivity property is satisfied).

One way to generate the σ-algebra of Borel subsets of $\mathbb{R}^2$ is to start with the collection of closed rectangles $[a_1, b_1] \times [a_2, b_2]$ and then add all other sets necessary in order to have a σ-algebra. Any set in this resulting σ-algebra is called a *Borel subset* of $\mathbb{R}^2$. All subsets of $\mathbb{R}^2$ normally encountered belong to this σ-algebra.

Definition 2.12: Joint Cumulative Distribution Function

Let X and Y be random variables with joint distribution measure $\mu_{X,Y}$. The joint cumulative distribution function of (X, Y) is

$$F_{X,Y}(a, b) = \mu_{X,Y}\big((-\infty, a] \times (-\infty, b]\big) = \mathbb{P}\{X \le a, Y \le b\}, \ a \in \mathbb{R}, b \in \mathbb{R}. \tag{2.5}$$

Definition 2.13: Joint Density Function

Let X and Y be random variables with joint distribution measure $\mu_{X,Y}$. We say that a nonnegative, Borel-measurable function $f_{X,Y}$ is a joint density for the pair of random variables (X, Y) if

$$\mu_{X,Y}(C) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbb{I}_C(x, y) f_{X,Y}(x, y) dy dx \text{ for all Borel sets } C \subset \mathbb{R}^2. \tag{2.6}$$

Let $F_{X,Y}$ be the joint cumulative distribution function for X and Y. Then, condition (2.6) holds if and only if

$$F_{X,Y}(a, b) = \int_{-\infty}^{a} \int_{-\infty}^{b} f_{X,Y}(x, y) dy dx \text{ for all } a \in \mathbb{R}, b \in \mathbb{R}. \tag{2.7}$$

Definition 2.14: Marginal Distribution Measures

Let X and Y be random variables with joint distribution measure $\mu_{X,Y}$. The marginal distribution measure of X and Y are given by

$$\mu_X(A) = \mathbb{P}\{X \in A\} = \mu_{X,Y}(A \times \mathbb{R}) \ \ \forall A \subset \mathcal{B}(\mathbb{R}) \tag{2.8}$$

and

$$\mu_X(B) = \mathbb{P}\{Y \in B\} = \mu_{X,Y}(\mathbb{R} \times B) \ \ \forall B \subset \mathcal{B}(\mathbb{R}) \tag{2.9}$$

respectively.

Definition 2.15: Marginal Cumulative Distribution Function

Let X and Y be random variables with marginal distribution measures μ_X and μ_Y respectively. The marginal cumulative distribution functions for X and Y are given by

$$F_X(a) = \mu_X(-\infty, a] = \mathbb{P}\{X \le a\} \ \forall a \in \mathbb{R}, \tag{2.10}$$

and

$$F_Y(b) = \mu_Y(-\infty, b] = \mathbb{P}\{Y \le b\} \ \forall b \in \mathbb{R} \tag{2.11}$$

respectively.

Definition 2.16: Marginal Density Functions

Let X and Y be random variables with joint density $f_{X,Y}$. If the joint density $f_{X,Y}$ exists, then the *marginal densities* for X and Y exist and are given by

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy \tag{2.12}$$

and

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx \tag{2.13}$$

respectively. Let μ_X, μ_Y denote the marginal distribution measures for X and Y. These *marginal densities* are nonnegative, Borel-measurable functions that satisfy

$$\mu_X(A) = \int_A f_X(x)dx \quad \forall A \subset \mathcal{B}(\mathbb{R}), \tag{2.14}$$

$$\mu_Y(B) = \int_B f_Y(y)dy \quad \forall B \subset \mathcal{B}(\mathbb{R}). \tag{2.15}$$

Let F_X and F_Y denote the marginal cumulative distribution functions for X and Y. These last conditions hold if and only if

$$F_X(a) = \int_{-\infty}^{a} f_X(x)dx \ \forall a \in \mathbb{R}, \tag{2.16}$$

$$F_Y(b) = \int_{-\infty}^{b} f_Y(y)dy \ \forall b \in \mathbb{R}. \tag{2.17}$$

Theorem 2.2

Let X and Y be random variables. The following conditions are equivalent.

(i) X and Y are independent.

(ii) The joint distribution measure factors:

$$\mu_{X,Y}(A \times B) = \mu_X(A) \cdot \mu_Y(B) \ \forall A, B \in \mathcal{B}(\mathbb{R}) \tag{2.18}$$

(iii) The joint cumulative distribution function factors:

$$F_{X,Y}(a,b) = F_X(a) \cdot F_Y(b) \ \forall a, b \in \mathbb{R} \tag{2.19}$$

(iv) The joint moment-generating function factors:

$$E[e^{uX+vY}] = E[e^{uX}] \cdot E[e^{vY}], \tag{2.20}$$

$\forall u, v \in \mathbb{R}$ for which the expectations are finite.

If there is a joint density, each of the conditions above are equivalent to the following:

(v) The joint density factors:

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y) \ \text{ for almost every } x, y \in \mathbb{R} \tag{2.21}$$

The above conditions imply but are not equivalent to the following:

(vi) The expectation factors:
$$E[XY] = E[X] \cdot E[Y], \tag{2.22}$$

provided that $E[XY] < \infty$.

Definition 2.17: Variance

Let X be a random variable whose expected value is defined. The variance of X, denoted $Var[X]$ is

$$Var[X] = E\big[(X - E[X])^2\big]. \tag{2.23}$$

Since $(X - E[X])^2$ is nonnegative, $Var[X]$ is always defined, although it may be infinite. One has via linearity of expectations that

$$Var[X] = E[X^2] - \big(E[X]\big)^2 \tag{2.24}$$

Definition 2.18: Standard Deviation

Let X be a random variable with a defined variance $Var[X]$. The standard deviation of X is defined as $\sqrt{Var[X]}$.

Definition 2.19: Covariance

Let X and Y be random variables and assume that $E[X], Var[X]$, $E[Y]$ and $Var[Y]$ are all finite. The covariance of X and Y is
$$Cov(X, Y) = E\big[(X - E[X])(Y - E[Y])\big]. \tag{2.25}$$

Linearity of expectations shows that

$$Cov(X, Y) = E[XY] - E[X] \cdot E[Y]. \tag{2.26}$$

In particular, $E[XY] = E[X] \cdot E[Y]$ if and only if $Cov(X, Y) = 0$.

Definition 2.20: Correlation Coefficient

Let X, Y be random variables and assume that $E[X], Var[X]$, $E[Y]$ and $Var[Y]$ are all finite such that $Var[X] > 0$ and $Var[Y] > 0$. We define the *correlation coefficient* of X and Y as

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X) Var(Y)}} \tag{2.27}$$

Definition 2.21: Uncorrelated

Let X, Y be random variables with a defined correlation coefficient $\rho(X, Y)$. If $\rho(X, Y) = 0$ (or equivalently, $Cov(X, Y) = 0$), we say that X and Y are *uncorrelated*.

Definition 2.22: Jointly Normal

Two random variables X and Y are said to be *jointly normal* if they have the joint density

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\}, \quad (2.28)$$

where $\sigma_1 > 0, \sigma_2 > 0, |\rho| < 1$, and μ_1, μ_2 are real numbers.

More generally, a random column vector $\mathbf{X} = (X_1, ..., X_n)^{tr}$, where superscript tr denotes tranpose, is jointly normal if it has a joint density

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n\det(C)}}\exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)C^{-1}(\mathbf{x}-\mu)^{tr}\right\}, \quad (2.29)$$

where $\mathbf{x} = (x_1, ..., x_n)$ is a row vector of dummy variables, $\mu = (\mu_1, ..., \mu_n)$ is the row vector of expectations, and C is the positive definite matrix of covariances. In essence, C's matrix elements are given by $C_{ij} := \rho(X_i, X_j)$ (Hence, $C_{ii} = 1 \; \forall i$).

Definition 2.23: Moment Generating Function

Let X be a random variable. Its moment-generating function, $\Psi : \mathbb{R} \to \mathbb{R}$ is defined as

$$\Psi(t) = E[e^{tX}], \qquad t \in \mathbb{R}, \quad (2.30)$$

provided that $E[e^{tX}]$ exists. We can see that obtaining moments can be done by differentiation

$$\frac{d^n\Psi(t)}{dt^n}\Big|_{t=0} = E[X^n] \quad (2.31)$$

2.3   General Conditional Expectations

Definition 2.24: Conditional Expectations

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\mathcal{G}$ be a sub σ-algebra of $\mathcal{F}$, and let X be a random variable that is either nonnegative or integrable. The *conditional expectation* of X given $\mathcal{G}$, denoted $E[X|\mathcal{G}]$, is any random variable[a] that satisfies

(i) *(Measurability)* $E[X|\mathcal{G}]$ is $\mathcal{G}$-measurable, and

(ii) *(Partial Averaging)*
$$\int_A E[X|\mathcal{G}]d\mathbb{P}(\omega) = \int_A X(\omega)d\mathbb{P}(\omega) \; \forall A \in \mathcal{G}. \quad (2.32)$$

If $\mathcal{G}$ is the σ-algebra generated by some other random variable W (i.e., $\mathcal{G} = \sigma(W)$), we generally write $E[X|W]$ rather than $E[X|\sigma(W)]$.

---
[a]In the measure-theoretic construction of Probability, we treat a conditional expectation $E[X|\mathcal{G}]$ as a random variable.

Proposition 2.2: Existence and Uniqueness of Conditional Expectation

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{G}$ a sub-σ algebra of $\mathcal{F}$. The conditional expectation of X given $\mathcal{G}$, $E[X|\mathcal{G}]$ satisfying properties (i) and (ii) of Definition 2.24 exists. Moreover, it is unique up to an almost surely equivalence; that is, if E_1 and E_2 are two conditional expectations for X given $\mathcal{G}$, then $E_1 = E_2$ almost surely.

Theorem 2.3: Properties of Conditional Expectations

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mathcal{G}$ be a sub-σ algebra of $\mathcal{F}$.

(i) *(Linearity of Conditional Expectations)* If X and Y are integrable random variables and c_1 and c_2 are constants, then

$$E[c_1 X + c_2 Y|\mathcal{G}] = c_1 E[X|\mathcal{G}] + c_2 E[Y|\mathcal{G}] \tag{2.33}$$

This equation also holds if we assume that X and Y are nonnegative (rather than integrable) and c_1 and c_2 are positive, although both sides may be $+\infty$.

(ii) *(Taking Out What is Known)* If X and Y are integrable random variables, Y and XY are integrable, and X is $\mathcal{G}$-measurable, then

$$E[XY|\mathcal{G}] = X E[Y|\mathcal{G}]. \tag{2.34}$$

This equation also holds if we assume that X is positive and Y is nonnegative (rather than integrable), although both sides may be $+\infty$.

(iii) *(Iterated Conditioning)* If $\mathcal{H}$ is a sub-σ algebra of $\mathcal{G}$ ($\mathcal{H}$ contains less information than $\mathcal{G}$) and X is an integrable random variable, then

$$E\big[E[X|\mathcal{G}]|\mathcal{H}\big] = E[X|\mathcal{H}] \tag{2.35}$$

This equation also holds if we assume that X is nonnegative (rather than integrable), although both sides may be $+\infty$.

(iv) *(Independence)* If X is integrable and independent of $\mathcal{G}$, then

$$E[X|\mathcal{G}] = E[X]. \tag{2.36}$$

This equation also holds if we assume that X is nonnegative (rather than integrable), although both sides may be $+\infty$.

(v) *(Conditional Jensen's Inequality)* If $\psi : \mathbb{R} \to \mathbb{R}$ is a convex function and X is integrable, then

$$E[\psi(X)|\mathcal{G}] \geq \psi\big(E[X|\mathcal{G}]\big) \tag{2.37}$$

Lemma 2.1: Independence Lemma

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $\mathcal{G}$ be a sub-σ algebra of $\mathcal{F}$. Suppose the random variables $X_1, ..., X_K$ are $\mathcal{G}$-measurable and the random variables $Y_1, ..., Y_L$ are independent of $\mathcal{G}$. Let $f(x_1, ..., x_K, y_1, ..., y_L)$ be a function of the dummy variables $x_1, ..., x_K$ and $y_1, ..., y_L$, and define

$$g(x_1, ..., x_K) = E[f(x_1, ..., x_K, Y_1, ..., Y_L)]. \tag{2.38}$$

Then

$$E[f(X_1, ..., X_K, Y_1, ..., Y_L)|\mathcal{G}] = g(X_1, ..., X_K) \tag{2.39}$$

Definition 2.25: Martingales

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let T be a fixed positive number, and let $\mathcal{F}(t), 0 \leq t \leq T$, be a filtration of sub-σ algebras of $\mathcal{F}$. Consider an adapted stochastic process $M(t), 0 \leq t \leq T$.

(i) If
$$E[M(t)|\mathcal{F}(s)] = M(s) \ \forall 0 \leq s \leq t \leq T, \tag{2.40}$$

we say this process is a **martingale**. It has no tendency to rise or fall.

(ii) If
$$E[M(t)|\mathcal{F}(s)] \geq M(s) \ \forall 0 \leq s \leq t \leq T, \tag{2.41}$$

we say this process is a **submartingale**. It has no tendency to fall; it may have a tendency to rise.

(iii) If
$$E[M(t)|\mathcal{F}(s)] \leq M(s) \ \forall 0 \leq s \leq t \leq T, \tag{2.42}$$

we say this process is a **supermartingale**. It has no tendency to rise; it may have a tendency to fall.

Definition 2.26: Markov Process

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let T be a fixed positive number, and let $\mathcal{F}(t), 0 \leq t \leq T$, be a filtration of sub-σ algebras of $\mathcal{F}$. Consider an adapted stochastic process $X(t), 0 \leq t \leq T$. Assume that for all $0 \leq s \leq t \leq T$ and for every nonnegative, Borel-measurable function f, there is another Borel-measurable function g such that
$$E[f(X(t))|\mathcal{F}(s)] = g(X(s)). \tag{2.43}$$

Then we say that X is a **Markov process**.

Chapter 3

Brownian Motion

3.1 Scaled Random Walks

3.1.1 Symmetric Random Walk

To create a Brownian motion, we begin with a symmetric random walk. To construct a symmetric random walk, we repeatedly toss a fair coin (p, the probability of H on each toss and $q = 1 - p$, the probability of T on each toss, are both equal to $1/2$).

Definition 3.1: Symmetric Random Walk

Let $\omega = \omega_1 \omega_2 \omega_3...$ denote successive outcomes of tosses, where ω_n denotes the outcome of the nth toss. Let

$$X_j = \begin{cases} 1 & \text{if } \omega_j = H \\ -1 & \text{if } \omega_j = T, \end{cases} \tag{3.1}$$

and define $M_0 = 0$, then

$$M_k = \sum_{j=1}^{k} X_j, \quad k = 1, 2, ... \tag{3.2}$$

The process $M_k, k = 0, 1, 2, ...$ is a *symmetric random walk*. With each toss, it either steps up one unit or down one unit, and each of the two possibilities is equally likely.

3.1.2 Increments of the Symmetric Random Walk

Definition 3.2: Increments

Consider a symmetric random walk with random variables X_j, M_k defined by Equations (3.1) and (3.2). Let $0 = k_0 < k_1 < k_2 < .. < k_m$ be a collection of nonnegative integers. Each of the random variables

$$M_{k_{i+1}} - M_{k_i} = \sum_{j=k_i+1}^{k_{i+1}} X_j \tag{3.3}$$

is called an *increment* of the random walk. It is the change in the position of the random walk between times k_i and k_{i+1}.

Proposition 3.1: Independent Increments

Consider a symmetric random walk with nonnegative integers $0 = k_0 < k_1 < ... < k_m$. A random walk has independent increments as the random variables

$$M_{k_1} = (M_{k_1} - M_{k_0}), (M_{k_2} - M_{k_1}), ..., (M_{k_m} - M_{k_{m-1}}) \tag{3.4}$$

are all independent. Increments over nonoverlapping time intervals are independent because they depend on different coin tosses.

Proposition 3.2: Increment Expectation and Variance

Consider a symmetric random walk with chosen increment indexes $0 = k_0 < k_1 < ... < k_m$. Let M_k be the random variables defined by Equation (3.2). Then, we have the following properties:

(i) $E[M_{k_{i+1}} - M_{k_i}] = 0$ *for all* $0 \leq i \leq m - 1$.

(ii) $\text{Var}[M_{k_{i+1}} - M_{k_i}] = k_{i+1} - k_i$ *for all* $0 \leq i \leq m - 1$.

Proof. While both are trivial, we'll ignore (i) as it is quite straight-forward.

$$\text{Var}[M_{k_{i+1}} - M_{k_i}] = E\big[(M_{k_{i+1}} - M_{k_i})^2\big] - \big(E[M_{k_{i+1}} - M_{k_i}]\big)^2 \tag{3.5}$$

$$= E\big[(M_{k_{i+1}} - M_{k_i})^2\big] \tag{3.6}$$

$$= E\bigg[\sum_{j=k_i+1}^{k_{i+1}} \sum_{l=k_i+1}^{k_{i+1}} X_j X_l\bigg] \tag{3.7}$$

$$= \sum_{j=k_i+1}^{k_{i+1}} \sum_{l=k_i+1}^{k_{i+1}} E\big[X_j X_l\big] \tag{3.8}$$

$$= \sum_{j=k_i+1}^{k_{i+1}} E\big[X_j^2\big] \tag{3.9}$$

$$= \sum_{j=k_i+1}^{k_{i+1}} 1 \tag{3.10}$$

$$= k_{i+1} - k_i, \tag{3.11}$$

where we have used the fact that X_j, X_l are independent when $j \neq l$, consequently invoking the fact that $E[X_j] = 0 \ \forall j$. $\qquad \square$

3.1.3   Martingale Property for the Symmetric Random Walk

Lemma 3.1: Range of Symmetric Random Walk

Let M_k be the random variable defined by Equation (3.2). Then, its range is given by

$$\text{range}(M_k) = \{-k + 2j; j \in \mathbb{Z}_{k+1}\} = 2\mathbb{Z}_{k+1} - k \tag{3.12}$$

Proof. We first note that we can define the range of M_k recursively noting the fact that from each point in M_{k-1}, we add and subtract 1. Hence, one can construct the set relation[a]

$$\text{range}(M_k) = \{t - 1, t + 1; t \in \text{range}(M_{k-1})\}, \tag{3.14}$$

which holds for $t \geq 1$ where the base case for $t = 0$ is given by

$$\text{range}(M_0) = \{0\}. \tag{3.15}$$

We will demonstrate that (3.12) holds by induction.

Base Case: $k = 1$
It's evident that $\text{range}(M_1) = \{-1, 1\}$, satisfying the desired condition.

Inductive Step: Assume that it holds for $1 \leq k \leq n$
We will consider the $k = n + 1$ case. We observe that

$$\text{range}(M_{n+1}) = \{t - 1, t + 1; t \in \text{range}(M_n)\} \tag{3.16}$$
$$= \{t - 1, t + 1; t \in \{-n + 2j; j \in \mathbb{Z}_{n+1}\}\} \tag{3.17}$$
$$= \{-n + 2j - 1, -n + 2j + 1; j \in \mathbb{Z}_{n+1}\} \tag{3.18}$$

where the second equality is obtained via the induction hypothesis. One can imagine that this set is generated by the two elements $-n + 2j - 1, -n + 2j + 1$. Naturally, there will be intersection between the two generators. Consider dummy indices $0 \leq l \leq n$ and $0 \leq p \leq n$ with the equivalence $-n + 2l - 1 = -n + 2p + 1$. Then, we have equality if $l = p + 1$. The two cases for which this cannot hold are if either $l = 0$ or $p = n$. Hence, we must have that

$$\{-n + 2j - 1, -n + 2j + 1; j \in \mathbb{Z}_{n+1}\} = \{-n - 1, n + 1\} \cup \{-n + 2j - 1; 1 \leq j \leq n\} \tag{3.19}$$
$$= \{-(n + 1) + 2j; j \in \mathbb{Z}_{n+2}\}, \tag{3.20}$$

verifying that it indeed holds for the $k = n + 1$ case. $\qquad\square$

---

[a]To be clear, we define
$$\{t - 1, t + 1; t \in A\} := \{t - 1; t \in A\} \cup \{t + 1; t \in A\} \tag{3.13}$$

To get an intuitive picture for how $\text{range}(M_k)$ looks like, we enumerate the first few cases below:

$$\text{range}(M_0) = \{0\}, \tag{3.21}$$
$$\text{range}(M_1) = \{-1, 1\}, \tag{3.22}$$
$$\text{range}(M_2) = \{-2, 0, 2\}, \tag{3.23}$$
$$\text{range}(M_3) = \{-3, -1, 1, 3\}, \tag{3.24}$$
$$\tag{3.25}$$

and so forth.

Lemma 3.2

Let $(\Omega_N, \mathcal{F}, \mathbb{P})$ be N-coin probability space. Let $h_k : \Omega_N \to \mathbb{N}$ be the function that computes the number of heads in the first k coin tosses. Let M_k be defined by (3.2), then

$$M_k(\omega) = -k + 2h_k(\omega) \ \forall 1 \leq k \leq N, \omega \in \Omega_N \tag{3.26}$$

Definition 3.3: Filtration Construction over Coin Toss Space

Let $(\Omega_N, \mathcal{F}, \mathbb{P})$ be N-coin probability space. Then, we will construct a filtration $\{F_k \subset \mathcal{F}\}$, whose index indicates information on the first k coin tosses. One can define the atoms of $\mathcal{F}_k$ as the sets that $\mathcal{F}_k$ can be

constructed out of. Let[a]

$$A_{\tau_1 \tau_2 \ldots \tau_k} := \{\omega \in \Omega_N : \omega_j = \tau_j \;\; \forall 1 \leq j \leq k\}. \tag{3.27}$$

Then, the set of atoms of $\mathcal{F}_k$ denoted by $A(\mathcal{F}_k)$ are given by

$$A(\mathcal{F}_k) = \{\emptyset, \Omega_N\} \cup \{A_\tau; \tau \in \{T, H\}^k\}. \tag{3.28}$$

One property to note is that $|A(\mathcal{F}_k)| = 2^k + 2$. Then, one constructs $\mathcal{F}_k$ by taking unions and complements of all sets in $A(\mathcal{F}_k)$. We emphasize that in an uncountable sample space, such a construction of atoms as was done here would not be possible.

---

[a]We apply the same notation as has been used earlier. If $\omega \in \Omega_N$, then $\omega_j \in \{T, H\}$ denotes the jth value of ω.

Definition 3.4: Constructing σ-Algebra Generated by Finite Random Variable

Let Ω be a sample space with a random variable $X : \Omega \to \mathbb{R}$ defined over it. We suppose that X is a finite random variable in that it takes on a finite number of values. Let

$$\text{range}(X) := \{x_1, x_2, \ldots, x_n\}, \tag{3.29}$$

where $x_1 < x_2 < \ldots < x_n$. We set

$$\epsilon = \min_{1 \leq i \leq n-1} (x_{i+1} - x_i) \tag{3.30}$$

and define the Borel interval over x_i, which we denote by B_i as

$$B_i = \left[x_i - \frac{\epsilon}{2}, x_i + \frac{\epsilon}{2} \right], \tag{3.31}$$

which was constructed to have the property $x_j \in B_i$ if and only if $j = i$. We note that its clear that $B_i \in \mathcal{B}(\mathbb{R})$. Let $1 \leq \tau_i \leq n$. We define the set $B_{\tau_1 \tau_2 \ldots \tau_k}$ as the set in which $x_{\tau_1}, \ldots, x_{\tau_k} \in B_{\tau_1 \tau_2 \ldots \tau_k}$. It is defined as follows:

$$B_{\tau_1 \ldots \tau_k} = \cup_{i=1}^k B_{\tau_i}. \tag{3.32}$$

Let us define the set of all sub-permutations of n elements with k choices as $\mathcal{S}_n^k$. It is defined as

$$\mathcal{S}_n^k = \{(\sigma_1, \sigma_2, \ldots, \sigma_k) : 1 \leq \sigma_i \leq n, \sigma_i \neq \sigma_j \text{ for } i \neq j\}. \tag{3.33}$$

We note that there are $\binom{n}{k}$ set of values that we can permute around to obtain all the possible combinations of elements[a] in S_n^k and for each of these we have $k!$ permutations. This signifies that $|\mathcal{S}_n^k| = k!\binom{n}{k} = \frac{n!}{(n-k)!}$. We denote the set of all sub-permutations of n elements with $k = 1$ up to $k = n$ choices by

$$\Sigma_n := \cup_{k=1}^n \mathcal{S}_n^k. \tag{3.34}$$

With this, letting $\sigma(X)$ denote the σ-algebra generated by X, we have that

$$\sigma(X) = \{\emptyset\} \cup \{\{\omega \in \Omega : X(\omega) \in B_\tau\} : \tau \in \Sigma_n\} \tag{3.35}$$

$$= \{\emptyset\} \cup \{\text{preim}_X(B_\tau) : \tau \in \Sigma_n\} \tag{3.36}$$

---

[a]A combination in the sense that one has a tuple of values containing a particular set of elements but without a specific order.

Proposition 3.3: Union of Preimages is Preimage of Union

Let $f : X \to Y$ and consider subsets $Y_i \subset Y$ for some arbitrary index set I. Then,

$$\text{preim}_f \left(\underset{i \in I}{\cup} Y_i \right) = \underset{i \in I}{\cup} \text{preim}_f (Y_i) \tag{3.37}$$

Proposition 3.4: Symmetric Random Walk is a Martingale

Let M_j be the random variable defined by Equation (3.2) and $\mathcal{F}_k$ be the σ-algebra of information corresponding to the first k coin tosses. The collection of random variables M_j is an adapted stochastic process that is a martingale.

Proof. Let's first show that $\{M_k\}$ is an adapted stochastic process. In essence, we want to prove that M_k are $\mathcal{F}_k$-measurable. From the construction in Def. 3.4, one has that every element in $\sigma(M_k)$ takes the form

$$\text{preim}_{M_k} (B_{\tau_1 \dots \tau_j}) = \cup_{i=1}^j \text{preim}_{M_k} (B_{\tau_i}). \tag{3.38}$$

By Lemma 3.2, we have that the number of heads in the k first elements of $\omega \in \Omega_N$ is given by $h_k(\omega) = \frac{M_k(\omega)+k}{2}$. Hence, we have that

$$\text{preim}_{M_k} (B_{\tau_1 \dots \tau_j}) = \cup_{i=1}^j \left\{ \omega \in \Omega_N : h_k(\omega) = \frac{\tau_i + k}{2} \right\} \tag{3.39}$$

Hence, one has the fact that every set in $\sigma(M_k)$ is a collection of elements in Ω_N that results in some set of possible heads values $\mathcal{H} \subset \mathbb{Z}_{k+1}$ (e.g. the set of all $\omega \in \Omega_N$ for which there were either 1 or 2 heads in the first k values). By examining the *atoms* of $\mathcal{F}_k$'s construction in Def. 3.3, it's clear that one can construct any $\text{preim}_{M_k} (B_{\tau_1 \dots \tau_j})$ by taking unions of $A_{\tau_1 \dots \tau_k} := \{\omega \in \Omega_N : \omega_j = \tau_j \; \forall 1 \leq j \leq k\}$. Hence, $\sigma(M_k) \subset \mathcal{F}_k \; \forall k$, thereby demonstrating that $\{M_k\}$ with the filtration $\{\mathcal{F}_k\}$ forms an adapted stochastic process.

We now want to demonstrate that this is a martingale. Let us consider nonnegative integers $k < l$. Then

$$E[M_l|\mathcal{F}_k] = E[(M_l - M_k) + M_k|\mathcal{F}_k] \tag{3.40}$$
$$= E[(M_l - M_k)|\mathcal{F}_k] + E[M_k|\mathcal{F}_k] \tag{3.41}$$
$$= E[(M_l - M_k)|\mathcal{F}_k] + M_k \tag{3.42}$$
$$= E[M_l - M_k] + M_k \tag{3.43}$$
$$= M_k. \tag{3.44}$$

In the second equality, we used the fact that M_k are $\mathcal{F}_k$ measurable and can therefore come outside of the expectation by property (ii) of Theorem 2.3. In the third equality, recall that $M_l - M_k = \sum_{j=k+1}^l X_j$, and that $X_j(\omega)$ only depend on ω_j. In essence, the random variable $M_l - M_k$ has no dependence on the first k coin tosses, signifying that $\sigma(M_l - M_k)$ and $\mathcal{F}_k$ should be independent. Then, by property (iv) of Theorem 2.3, we have that $E[M_l - M_k|\mathcal{F}_k] = E[M_l - M_k]$. The last equality is obtained by $E[M_l - M_k] = 0$ from Proposition 3.2.

For completeness, one has that $E[M_k|\mathcal{F}_k] = M_k \; \forall k$. Hence, we have shown that

$$E[M_l|\mathcal{F}_k] = M_k \; \forall 1 \leq k \leq l \leq N, \tag{3.45}$$

confirming that the symmetric random walk is a martingale. □

3.1.4 Quadratic Variation of the Symmetric Random Walk

Definition 3.5: Quadratic Variation

Let M be a random walk, indexed by time M_t. The *quadratic variation* up to time k is defined to be

$$[M, M]_k = \sum_{j=1}^{k} (M_j - M_{j-1})^2. \tag{3.46}$$

For the symmetric random walk we have that

$$[M, M]_k = k, \tag{3.47}$$

as $(M_j - M_{j-1})^2 = 1 \; \forall j$.

We note that the quadratic variation is computed path-by-path but in the construction of a symmetric random walk, this quantity turns out to be path-independent. However, we will see that for a general random process, it does depend on the path along which it is computed.

3.1.5 Scaled Symmetric Random Walk

Definition 3.6: Scaled Symmetric Random Walk

Let n be a positive integer and let M_k denote the symmetric random walk. Then, we define the *scaled symmetric random walk* as follows:

$$W^{(n)}(t) = \frac{1}{\sqrt{n}} M_{nt}, \tag{3.48}$$

provided that nt is itself an integer. If nt is not an integer, then we define $W^{(n)}(t)$ more generally as a linear interpolation between $W^{(n)}(s)$ and $W^{(n)}(u)$ where s and u satisfy $s = \lfloor nt \rfloor / n$ and $u = \lceil nt \rceil / n = s + 1$ respectively. It would be given by

$$W^{(n)}(t) = (nt - \lfloor nt \rfloor) W^{(n)}\left(\frac{1}{n}\lceil nt \rceil\right) - (nt - \lceil nt \rceil) W^{(n)}\left(\frac{1}{n}\lfloor nt \rfloor\right) \tag{3.49}$$

$$= n\big[(t - s)\big(W^{(n)}(s+1) - W^{(n)}(s)\big) + W^{(n)}(s)\big] \; \forall s < t < s+1 \tag{3.50}$$

Proposition 3.5: Independent Increments: Scaled Symmetric Random Walk

Consider the scaled symmetric random walk $W^{(n)}(t)$ defined in Definition 3.6. Consider times $0 = t_0 < t_1 < t_2 < ... < t_n$ such that nt_j is an integer for all $0 \le j \le n$. Then,

$$\big(W^{(n)}(t_1) - W^{(n)}(t_0)\big), \big(W^{(n)}(t_2) - W^{(n)}(t_1)\big), ..., \big(W^{(n)}(t_n) - W^{(n)}(t_{n-1})\big) \tag{3.51}$$

are all independent.

Proof. Follows quite immediately from Proposition 3.1. Observe that

$$W^{(n)}(t_j) - W^{(n)}(t_{j-1}) = \frac{1}{\sqrt{n}}\big(M_{nt_j} - M_{nt_{j-1}}\big) \tag{3.52}$$

Hence, we have nonnegative integers $0 = nt_0 < nt_1 < ... < nt_n$. From this proposition it follows that $W^{(n)}(t_j) - W^{(n)}(t_{j-1})$ is independent of $W^{(n)}(t_i) - W^{(n)}(t_{i-1})$ for $i \ne j$. $\square$

Proposition 3.6: Expectation and Variance of Scaled Symmetric Random Walk

Consider the scaled symmetric random walk $W^{(n)}(t)$ defined in Definition 3.6 and let $0 \leq s \leq t$ such that ns and nt are integers. Then, we have the following properties:

(i) $E[W^{(n)}(t) - W^{(n)}(s)] = 0$

(ii) $Var[W^{(n)}(t) - W^{(n)}(s)] = t - s$

Proof. This follows quite immediately from Proposition 3.2. Observe that

$$E[W^{(n)}(t) - W^{(n)}(s)] = \frac{1}{\sqrt{n}} E[M_{nt} - M_{ns}] = 0, \tag{3.53}$$

as $E[M_{nt} - M_{ns}] = 0$ by this proposition. Similarly, one has that

$$Var[W^{(n)}(t) - W^{(n)}(s)] = \frac{1}{n} Var[M_{nt} - M_{ns}] \tag{3.54}$$

$$= \frac{1}{n}(nt - ns) \tag{3.55}$$

$$= t - s, \tag{3.56}$$

where we have invoked the second statement of Proposition 3.2. $\qquad \square$

Proposition 3.7: Scaled Symmetric Random Walk is a Martingale

Consider the scaled symmetric random walk $W^{(n)}(t)$ defined in Definition 3.6 and let $0 \leq s \leq t$ such that ns and nt are integers. Let Ω be the coin toss sample space and $\mathcal{F}_W(s)$ be the σ-algebra associated with the knowledge one has at time s (i.e. corresponding to the first ns coin tosses). Then,

$$E[W^{(n)}(t)|\mathcal{F}_W(s)] = W^{(n)}(s). \tag{3.57}$$

The scaled symmetric random walk is a martingale.

Proof. This follows quite immediately from Proposition 3.4. We observe that

$$E[W^{(n)}(t)|\mathcal{F}_W(s)] = \frac{1}{\sqrt{n}} E[M_{nt}|\mathcal{F}(ns)] \tag{3.58}$$

$$= \frac{1}{\sqrt{n}} M_{ns} \tag{3.59}$$

$$= W^{(n)}(s), \tag{3.60}$$

where the first equality follows by definition of $\mathcal{F}_W(s) := \mathcal{F}(ns)$ (defined in Proposition 3.4) as well as $W^{(n)}(t) := \frac{1}{\sqrt{n}} M_{nt}$. The second equality is a consequence of M_k being a martingale (Proposition 3.4) and the last equality simply follows by the definition of $W^{(n)}(s)$. $\qquad \square$

Definition 3.7: Quadratic Variation of Scaled Symmetric Random Walk

Consider the scaled symmetric random walk $W^{(n)}$ defined in Definition 3.6. Let t be a value such that nt is

an integer. The quadratic variation of $W^{(n)}$ is defined to be

$$[W^{(n)}, W^{(n)}](t) := \sum_{j=1}^{nt} \left[W^{(n)}\left(\frac{j}{n}\right) - W^{(n)}\left(\frac{j-1}{n}\right) \right]^2 \tag{3.61}$$

$$= \sum_{j=1}^{nt} \left(\frac{1}{\sqrt{n}} M_j - \frac{1}{\sqrt{n}} M_{j-1} \right)^2 \tag{3.62}$$

$$= \frac{1}{n} \sum_{j=1}^{nt} (M_j - M_{j-1})^2 \tag{3.63}$$

$$= \frac{1}{n} \sum_{j=1}^{nt} 1 \tag{3.64}$$

$$= t. \tag{3.65}$$

Although this is a path by path computation, we arrive at a path-independent result of the quadratic variation being equal to the time interval over which the path takes place.

3.1.6   Limiting Distribution of the Scaled Random Walk

Proposition 3.8: Moment-Generating Functions to Distributions

Let X and Y be two random variables. If their moment generating functions are equivalent, then X and Y have the same distributions.

Theorem 3.1: Central Limit Theorem

Fix $t \geq 0$. As $n \to \infty$, the distribution of the scaled random walk $W^{(n)}(t)$ evaluated at time t converges to the normal distribution with mean zero and variance t.

Proof. The normal density associated with the mean zero, variance t normal distribution is given by

$$f(x, t) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}. \tag{3.66}$$

We are going to prove this equivalence between distributions by proving that their moment generating functions are equivalent. The moment generating function for the normal distribution is given by

$$\psi(u, t) = \int_{-\infty}^{\infty} e^{ux} f(x, t) \mathrm{d}x = e^{\frac{u^2 t}{2}} \tag{3.67}$$

We denote the moment generating function for $W^{(n)}(t)$ by ψ_n and aim to compute it in the $n \to \infty$ limit.

$$\psi_n(u,t) = E[e^{uW^{(n)}(t)}] \tag{3.68}$$

$$= E\left[\exp\left\{\frac{u}{\sqrt{n}}M_{nt}\right\}\right] \tag{3.69}$$

$$= E\left[\exp\left\{\frac{u}{\sqrt{n}}\sum_{j=1}^{nt}X_j\right\}\right] \tag{3.70}$$

$$= \Pi_{j=1}^{nt}E\left[\exp\left\{\frac{u}{\sqrt{n}}X_j\right\}\right] \tag{3.71}$$

$$= \Pi_{j=1}^{nt}\left(\frac{e^{\frac{u}{\sqrt{n}}} + e^{-\frac{u}{\sqrt{n}}}}{2}\right) \tag{3.72}$$

$$= \left(\frac{e^{\frac{u}{\sqrt{n}}} + e^{-\frac{u}{\sqrt{n}}}}{2}\right)^{nt}, \tag{3.73}$$

where the early equalities followed by definition, the fourth equality was due to independence of X_j. We'll now compute the log-limit

$$\lim_{n\to\infty}\log(\psi_n(u,t)) = \lim_{n\to\infty}\log\left(\frac{e^{\frac{u}{\sqrt{n}}} + e^{-\frac{u}{\sqrt{n}}}}{2}\right)^{nt} \tag{3.74}$$

$$= \lim_{n\to\infty}nt\log\left(\frac{e^{\frac{u}{\sqrt{n}}} + e^{-\frac{u}{\sqrt{n}}}}{2}\right) \tag{3.75}$$

$$= \lim_{x\to0^+}\frac{t}{x^2}\log\left(\frac{e^{ux} + e^{-ux}}{2}\right) \tag{3.76}$$

$$= \lim_{x\to0^+}\frac{t}{2x}\frac{ue^{ux} - ue^{-ux}}{e^{ux} + e^{-ux}} \tag{3.77}$$

$$= \lim_{x\to0^+}\frac{ut}{2}\left[u - u\frac{(e^{ux} - e^{-ux})^2}{(e^{ux} + e^{-ux})^2}\right] \tag{3.78}$$

$$= \frac{u^2t}{2}, \tag{3.79}$$

where the third equality involved a change of variables $n \mapsto \frac{1}{x^2}$ and the fourth and fifth equality followed as a consequence of l'Hopital's rule. Hence, we have that

$$\lim_{n\to\infty}\log(\psi_n(u,t)) = \log\left(\lim_{n\to\infty}\psi_n(u,t)\right) = \frac{u^2t}{2}, \tag{3.80}$$

signifying that

$$\lim_{n\to\infty}\psi_n(u,t) = e^{\frac{u^2t}{2}}. \tag{3.81}$$

This limit tells us that the moment-generating functions for $W^{(n)}(t)$ as $n \to \infty$ is the same as that of the normal distribution. In turn, by Proposition 3.8, $W^{(n)}(t)$ becomes normally distributed as $n \to \infty$. □

3.1.7 Log-Normal Distribution as Limit of the Binomial Model

Definition 3.8: Log-Normal Distribution

Let X be a random variable that is normally distributed. Then, any random variable of the form ce^X, where c is a constant is said to have a log-normal distribution.

Theorem 3.2

Consider the symmetric random walk M_{nt} defined in Definition 3.1 where nt is a nonnegative integer. Let H_{nt} and T_{nt} denote the number of heads and tails after nt coin tosses respectively. Hence, one has the relations

$$nt = H_{nt} + T_{nt}, \qquad M_{nt} = H_{nt} - T_{nt}. \tag{3.82}$$

Let $0 < \sigma < 1$ be a positive constant and suppose that the up and down factors for a stock price on the nt^{th} coin toss were given by

$$u_n = 1 + \frac{\sigma}{\sqrt{n}}, \tag{3.83}$$

$$d_n = 1 - \frac{\sigma}{\sqrt{n}}. \tag{3.84}$$

Then, the stock price at time t is given by

$$S_n(t) = S(0)u_n^{H_{nt}}d_n^{T_{nt}} = S(0)\left(1 + \frac{\sigma}{\sqrt{n}}\right)^{\frac{1}{2}(nt+M_{nt})}\left(1 - \frac{\sigma}{\sqrt{n}}\right)^{\frac{1}{2}(nt-M_{nt})}. \tag{3.85}$$

Then, as $n \to \infty$, the distribution of $S_n(t)$ converges to the distribution of

$$S(t) = S(0)\exp\left\{\sigma W(t) - \frac{1}{2}\sigma^2 t\right\}, \tag{3.86}$$

where $W(t)$ is a normal random variable with mean zero and variance t. The distribution of $S(t)$ is log-normal.

Proof. We'll take the log of $S_n(t)$ for easier evaluation.

$$\log\big(S_n(t)\big) = \log(S_0) + \frac{1}{2}(nt + M_{nt})\log\left(1 + \frac{\sigma}{\sqrt{n}}\right) + \frac{1}{2}(nt - M_{nt})\log\left(1 - \frac{\sigma}{\sqrt{n}}\right). \tag{3.87}$$

We will now use the taylor expansion of $\log(1 + x) = \sum_{k=1}^{\infty}(-1)^{k+1}\frac{x^k}{k}$ around $x = 0$. This Taylor expansion holds for $|x| < 1$, which by the assumption imposed on σ ensures that this is valid.

$$\log\big(S_n(t)\big) = \log(S_0) + \frac{1}{2}(nt + M_{nt})\sum_{k=1}^{\infty}\frac{(-1)^{k+1}}{k}\frac{\sigma^k}{n^{k/2}} - \frac{1}{2}(nt - M_{nt})\sum_{k=1}^{\infty}\frac{1}{k}\frac{\sigma^k}{n^{k/2}} \tag{3.88}$$

$$= \log(S_0) + \frac{nt}{2}\sum_{k=1}^{\infty}\frac{\big((-1)^{k+1} - 1\big)}{k}\frac{\sigma^k}{n^{k/2}} + \frac{M_{nt}}{2}\sum_{k=1}^{\infty}\frac{\big((-1)^{k+1} + 1\big)}{k}\frac{\sigma^k}{n^{k/2}} \tag{3.89}$$

$$= \log(S_0) - nt\sum_{k=1}^{\infty}\frac{1}{2k}\frac{\sigma^{2k}}{n^k} + \frac{\sigma M_{nt}}{\sqrt{n}}\sum_{k=0}^{\infty}\frac{1}{2k+1}\frac{\sigma^{2k}}{n^k} \tag{3.90}$$

$$= \log(S_0) - nt\left(\frac{\sigma^2}{2n} + \mathcal{O}(n^{-2})\right) + \sigma W^{(n)}(t)\big(1 + \mathcal{O}(n^{-1})\big) \tag{3.91}$$

$$= \log(S_0) - \frac{\sigma^2 t}{2} + \sigma W^{(n)}(t) + \big(t + \sigma W^{(n)}(t)\big)\mathcal{O}(n^{-1}). \tag{3.92}$$

By Theorem 3.1, we observed that $\lim_{n \to \infty} W^{(n)}(t) = W(t)$ where $W(t)$ is normally distributed with mean zero and variance t. We also note that $\mathcal{O}(n^{-1}) \to 0$ as $n \to \infty$. Hence, we have that

$$\lim_{n \to \infty} \log\big(S_n(t)\big) = \log(S_0) - \frac{\sigma^2 t}{2} + \sigma W(t), \tag{3.93}$$

allowing us to conclude that

$$\lim_{n \to \infty} S_n(t) = S(0)\exp\left\{\sigma W(t) - \frac{\sigma^2 t}{2}\right\} \tag{3.94}$$

$\square$

3.2   Brownian Motion

3.2.1   Definition of Brownian Motion

We obtain Brownian motion as the limit of the scaled random walks $W^{(n)}(t)$ defined in Definition 3.6 as $n \to \infty$.

Definition 3.9: Brownian Motion

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. For each $\omega \in \Omega$, suppose there is a continuous function $W(t)$ of $t \geq 0$ that satisfies $W(0) = 0$ and that depends on ω. Then $W(t), t \geq 0$, is a Brownian motion if for all $0 = t_0 < t_1 < ... < t_m$ the increments

$$W(t_1) = W(t_1) - W(t_0), W(t_2) - W(t_1), ..., W(t_m) - W(t_{m-1}) \tag{3.95}$$

are independent and each of these increments is normally distributed with

$$E[W(t_{i+1}) - W(t_i)] = 0$$
$$\mathrm{Var}[W(t_{i+1}) - W(t_i)] = t_{i+1} - t_i. \tag{3.96}$$

3.2.2   Distribution of Brownian Motion

Lemma 3.3: Covariance of Brownian Variables

Let W denote a Brownian motion. We have the property that $E[W(t)] = 0 \ \forall t \geq 0$. Let $0 \leq s < t$. Then, the covariance between $W(s)$ and $W(t)$ is given by

$$Cov\big(W(s), W(t)\big) = s \tag{3.97}$$

Proof.

$$\begin{aligned}
Cov\big(W(s), W(t)\big) &= E[W(s)W(t)] - E[W(s)]E[W(t)] \tag{3.98}\\
&= E[(W(s)W(t) - W^2(s) + W^2(s)] \tag{3.99}\\
&= E\big[W(s)\big(W(t) - W(s)\big) + W^2(s)\big] \tag{3.100}\\
&= E[W(s)]E[W(t) - W(s)] + E[W^2(s)] \tag{3.101}\\
&= 0 + E[W^2(s)] \tag{3.102}\\
&= s, \tag{3.103}
\end{aligned}$$

where in the fourth equality, we used the fact that $W(s)$ and $W(t) - W(s)$ are independent by their property of being a Brownian motion. $\qquad\square$

Proposition 3.9: Covariance Matrix for Brownian Motion

Let W denote a Brownian motion. Consider the collection of times $0 < t_1 < ... < t_n$. Then, the covariance matrix for Brownian motion (i.e., for the n-dimensional random vector $\big(W(t_1), W(t_2), ..., W(t_n)\big)$) is the matrix C whose matrix elements are defined as $C_{ij} := Cov\big(W(t_i), W(t_j)\big)$. It is given by

$$\begin{bmatrix}
E[W^2(t_1)] & E[W(t_1)W(t_2)] & ... & E[W(t_1)W(t_n)] \\
E[W(t_2)W(t_1)] & E[W^2(t_2)] & ... & E[W(t_2)W(t_n)] \\
\vdots & \vdots & \vdots & \vdots \\
E[W(t_m))W(t_1)] & E[W(t_n)W(t_2)] & ... & E[W^2(t_n)]
\end{bmatrix} = \begin{bmatrix}
t_1 & t_1 & ... & t_1 \\
t_1 & t_2 & ... & t_2 \\
\vdots & \vdots & \vdots & \vdots \\
t_1 & t_2 & ... & t_n
\end{bmatrix} \tag{3.104}$$

In essence, $C_{ij} = t_{\min\{i,j\}}$. This result follows immediately from Lemma 3.3.

Proposition 3.10: Moment-Generating Function for Brownian Motion

Consider the Brownian motion $W(t)$ for $t \geq 0$, with a collection of times $0 = t_0 < t_1 < t_2 < ... < t_n$. Then, the moment-generating function of the random vector $\big(W(t_1), W(t_2), ..., W(t_n)\big)$ which we'll denote by $\psi : \mathbb{R}^n \to \mathbb{R}$ is given by

$$\begin{aligned}
\psi(u_1, u_2, ..., u_n) &= E\big[\exp\{\sum_{i=1}^{n} u_n W(t_n)\}\big] \\
&= \exp\Big\{\frac{1}{2} \sum_{i=1}^{n} \Big(\sum_{j=i}^{n} u_j\Big)^2 (t_i - t_{i-1})\Big\} \tag{3.105}
\end{aligned}$$

Theorem 3.3: Alternative Characterizations of Brownian Motion

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. For each $\omega \in \Omega$, suppose that there is a continuous function $W(t)$ of $t \geq 0$ that satisfies $W(0) = 0$ and that depends on ω. The following three properties are equivalent:

(i) *For all $0 = t_0 < t_1 < ... < t_n$, the increments*

$$W(t_1) = W(t_1) - W(t_0), W(t_2) - W(t_1), ..., W(t_n) - W(t_{n-1}) \tag{3.106}$$

are independent and each of these increments is normally distributed with mean and variance given by (3.96).

(ii) *For all $0 = t_0 < t_1 < ... < t_n$, the random variables $W(t_1), W(t_2), ..., W(t_n)$ are jointly normally distributed with means equal to zero and covariance matrix (3.104).*

(iii) *For all $0 = t_0 < t_1 < ... < t_n$, the random variables $W(t_1), W(t_2), ..., W(t_n)$ have the joint moment-generating function 3.105.*

If any of $(i), (ii), or (iii)$ hold (and hence they all hold), then $W(t), t \geq 0$ is a Brownian motion.

3.2.3 Filtration for Brownian Motion

In addition to the Brownian motion, we will also require some notation for the amount of information available at each time. We will do this with a filtration.

Definition 3.10: Filtration for Brownian Motion

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space on which is defined a Brownian motion $W(t), t \geq 0$. A filtration for the Brownian motion is a collection of σ-algebras $\mathcal{F}(t), t \geq 0$, satisfying

(i) **(Information Accumulates)** *For $0 \leq s < t$, every set in $\mathcal{F}(s)$ is also in $\mathcal{F}(t)$. In other words, there is at least as much information available at the later time $\mathcal{F}(t)$ as there is at the earlier time $\mathcal{F}(s)$.*

(ii) **(Adaptivity)** *For each $t \geq 0$, the Brownian motion $W(t)$ at time t is $\mathcal{F}(t)$-measurable. In other words, the information available at time t is sufficient to evaluate the Brownian motion $W(t)$ at that time.*

(iii) **(Independence of future increments)** *For $0 \leq t < u$, the increment $W(u) - W(t)$ is independent of $\mathcal{F}(t)$. In other words, any increment of the Brownian motion after time t is independent of the information available at time t.*

Definition 3.11: Adapted to a Filtration

Let $\Delta(t), t \geq 0$, be a stochastic process. We say that $\Delta(t)$ is adapted to the filtration $\mathcal{F}(t)$ if for each $t \geq 0$ the random variable $\Delta(t)$ is $\mathcal{F}(t)$-measurable.

3.2.4 Martingale Property for Brownian Motion

Theorem 3.4: Brownian Motion is a Martingale

Let $W(t)$ denote a Brownian motion and $\mathcal{F}(t)$ be a filtration for Brownian motion. Let $0 \leq s \leq t$. Then, $W(t)$ is a martingale. It satisfies

$$E[W(t)|\mathcal{F}(s)] = W(s). \tag{3.107}$$

Proof. The proof is very much similar to previous instances of demonstrating the symmetric random walk

was a martingale.

$$E[W(t)|\mathcal{F}(s)] = E[(W(t) - W(s)) + W(s)|\mathcal{F}(s)] \tag{3.108}$$
$$= E[W(t) - W(s)|\mathcal{F}(s)] + E[W(s)|\mathcal{F}(s)] \tag{3.109}$$
$$= E[W(t) - W(s)] + W(s) \tag{3.110}$$
$$= 0 + W(s) \tag{3.111}$$
$$= W(s), \tag{3.112}$$

where in the third equality we had used the fact that $W(t) - W(s)$ is independent of $\mathcal{F}(s)$ (as that is how the Filtration is defined) and that $W(s)$ is $\mathcal{F}(s)$-measurable. The fourth equality invokes the expectation property of Brownian motion: $E[W(t) - W(s)] = 0$. □

3.3 Quadratic Variation

For Brownian motion, there is no natural step size. The paths of Brownian motion are unusual in that their quadratic variation is not zero. This makes stochastic calculus different from ordinary calculus and is the source of the volatility term in the Black-Scholes-Merton partial differential equation, which will be discussed in the next chapter.

3.3.1 First-Order Variation

Definition 3.12: First-Order Variation

Let $f : [0, T] \to \mathbb{R}$. To compute the first-order variation of a function up to time T, we choose a partition $\Pi = \{t_0, t_1, ..., t_n\}$ of $[0, T]$, which is a set of times satisfying

$$0 = t_0 < t_1 < t_2 < ... < t_n = T. \tag{3.113}$$

The maximum step size of the partition is denoted by $||\Pi|| = \max_{0 \le j \le n-1}(t_{j+1} - t_j)$. We then define

$$FV_T(f) = \lim_{||\Pi|| \to 0} \sum_{j=0}^{n-1} |f(t_{j+1}) - f(t_j)|. \tag{3.114}$$

If f is differentiable over $[0, T]$, then we have the first order variation of f over $[0, T]$ as

$$FV_T(f) = \int_0^T |f'(t)| dt \tag{3.115}$$

3.3.2 Quadratic Variation

Definition 3.13: Quadratic Variation

Let $f : [0, T] \to \mathbb{R}$. The quadratic variation of f up to time T is

$$[f, f](T) = \lim_{||\Pi|| \to 0} \sum_{j=1}^{n-1} \left[f(t_{j+1}) - f(t_j) \right]^2, \tag{3.116}$$

where $\Pi = \{t_0, t_1, ..., t_n\}$ and $0 = t_0 < t_1 < ... < t_n = T$.

Proposition 3.11: Second Variation of Differentiable Functions

Let $f : [0, T] \to \mathbb{R}$. If f has a continuous derivative, then its second variation is precisely zero:

$$[f, f](T) = 0 \tag{3.117}$$

Theorem 3.5: Second Variation of Brownian Motion

Let W be a Brownian motion. Then $[W, W](T) = T$ for all $T \geq 0$ almost surely.

What we can conclude say about Brownian motion is that it accumulates quadratic variation at rate one per unit time.

Proposition 3.12

Let $\Pi = \{t_0, t_1, ..., t_n\}$ be a partition of $[0, T]$ and W denote a Brownian motion. Then we have some of the following properties:

1. $\lim_{||\Pi|| \to 0} \sum_{j=0}^{n-1} \left(W(t_{j+1}) - W(t_j) \right)^2 = T$.

2. $\lim_{||\Pi|| \to 0} \sum_{j=0}^{n-1} \left(W(t_{j+1}) - W(t_j) \right)(t_{j+1} - t_j) = 0$.

3. $\lim_{||\Pi|| \to 0} \sum_{j=0}^{n-1} \left(t_{j+1} - t_j \right)^2 = 0$.

3.3.3 Volatility of Geometric Brownian Motion

Definition 3.14: Geometric Brownian Motion

Let $\alpha, \sigma > 0$ be constants, and W denote a Brownian motion. We define the *geometric Brownian motion* by

$$S(t) = S(0)\exp\left\{ \sigma W(t) + \left(\alpha - \frac{1}{2}\sigma^2 \right)t \right\}. \tag{3.118}$$

This is the asset-price model used in the Black-Scholes-Merton option-pricing formula.

Definition 3.15: Log-Return

Let S be a geometric Brownian motion. Consider two times $0 < t_j < t_{j+1}$. The quantity

$$log\left(\frac{S(t_{j+1})}{S(t_j)}\right) = \sigma(W(t_{j+1}) - W(t_j)) + \left(\alpha - \frac{1}{2}\sigma^2\right)(t_{j+1} - t_j) \tag{3.119}$$

is called the log return over the subinterval $[t_j, t_{j+1}]$.

Definition 3.16: Realized Volatility

Let S be a geometric Brownian motion. Then, the sum of the squares of the log returns, sometimes referred to as the realized volatility, which I'll denote by σ_R is

$$\sigma_R := \sum_{j=0}^{m-1}\left(log\frac{S(t_{j+1})}{S(t_j)}\right)^2 = \sigma^2\sum_{j=0}^{m-1}\left(W(t_{j+1}) - W(t_j)\right)^2 + \left(\alpha - \frac{1}{2}\sigma^2\right)^2\sum_{j=0}^{m-1}(t_{j+1} - t_j)^2 \tag{3.120}$$

$$+ 2\sigma\left(\alpha - \frac{1}{2}\sigma^2\right)\sum_{j=0}^{m-1}\left(W(t_{j+1} - W(t_j))(t_{j+1} - t_j)\right) \tag{3.121}$$

Proposition 3.13: Recovering Volatility in the Log-Return Continuum Limit

Let S be a geometric Brownian motion over an interval $[T_1, T_2]$ whose form is described by (3.118). Consider a partition $T_1 = t_0 < t_1 < ... < t_n = T_2$. Then, we have the property that

$$\lim_{||\Pi|| \to 0}\frac{1}{T_2 - T_1}\sum_{j=0}^{m-1}\left(log\frac{S(t_{j+1})}{S(t_j)}\right)^2 = \sigma^2. \tag{3.122}$$

This follows immediately from the Realized Volatility formula and Proposition 3.12.

By Proposition 3.13, we can choose a sufficiently small step size to approximate the true volatility as we have that

$$\frac{1}{T_2 - T_1}\sum_{j=0}^{m-1}\left(\log\frac{S(t_{j+1})}{S(t_j)}\right)^2 \approx \sigma^2. \tag{3.123}$$

If the asset price $S(t)$ really is a geometric Brownian motion with constant volatility σ, then σ can be identified from price observations by computing the LHS of (3.123). In theory, we can make this approximation as accurate as we would like but in practice, there is a limit to how small the step size can be. Between trades, there is no information about prices, and when a trade takes place, it is sometimes at the bid price and sometimes at the ask price.

3.4 Markov Property

In this section, we aim to show that Brownian motion is a Markov process and discuss its *transition density*.

Theorem 3.6: Brownian Motion is a Markov Process

Let $W(t), t \geq 0$ be a Brownian motion and let $\mathcal{F}(t), t \geq 0$, be a filtration for this Brownian motion. Then $W(t), t \geq 0$, is a Markov process.

Proof. We want to show that for all Borel-measurable functions f, that there exists another Borel-measurable

function g such that

$$E[f(W(t))|\mathcal{F}(s)] = g(W(s)) \tag{3.124}$$

We expand

$$E[f(W(t))|\mathcal{F}(s)] = E[f(W(t) - W(s) + W(s))|\mathcal{F}(s)]. \tag{3.125}$$

We now use the fact that $W(t) - W(s)$ is independent of $\mathcal{F}(s)$ and $W(s)$ is $\mathcal{F}(s)$-measurable. Then, by the Independence Lemma 2.1, we have that

$$E[f(W(t) - W(s) + W(s))|\mathcal{F}(s)] = g(W(s)), \tag{3.126}$$

where g satisfies

$$g(x) = E[f(W(t) - W(s) + x)]. \tag{3.127}$$

Since $W(t) - W(s)$ is normally distributed with variance $t - s$ and mean 0, we must have

$$g(x) = \frac{1}{\sqrt{2\pi(t-s)}} \int_{-\infty}^{\infty} f(y+x)e^{-\frac{y^2}{2(t-s)}} \, dy. \tag{3.128}$$

Hence,

$$E[f(W(t))|\mathcal{F}(s)] = g(W(s)), \tag{3.129}$$

where g satisfies (3.128). The integrand $f(y+x)e^{-\frac{y^2}{2(t-s)}}$ is a Borel-measurable function as it is the product of two Borel-measurable functions. Since g is the integral of a Borel-measurable function, it is itself Borel-measurable, which allows us to conclude that $W(t)$ is a Markov process. $\qquad\square$

Definition 3.17: Transition Density

Let W be a Brownian motion. Then, we define $\tau := t - s = \text{Var}[W(t) - W(s)]$. We define the transition density $\rho(\tau, x, y)$ for Brownian motion to be

$$\rho(\tau, x, y) = \frac{1}{\sqrt{2\pi\tau}} e^{-\frac{(y-x)^2}{2\tau}}. \tag{3.130}$$

It is constructed precisely as so to satisfy

$$E[f(W(t))|\mathcal{F}(s)] = \int_{-\infty}^{\infty} f(y)\rho(\tau, W(s), y) dy. \tag{3.131}$$

This equation has the following interpretation. Conditioned on the information in $\mathcal{F}(s)$, the conditional density of $W(t)$ is $\rho(\tau, W(s), y)$. This is a density in the variable y. In particular, the only information from $\mathcal{F}(s)$ that is relevant is the value of $W(s)$. The fact that only $W(s)$ is relevant is the essence of the Markov property.

3.5   First Passage Time Distribution

Theorem 3.7: Exponential Martingale

Let $W(t), t \geq 0$, be a Brownian motion with a filtration $\mathcal{F}(t), t \geq 0$, and let σ be a constant. The process $Z(t), t \geq 0$, defined by

$$Z(t) = \exp\left\{\sigma W(t) - \frac{1}{2}\sigma^2 t\right\} \tag{3.132}$$

is a martingale.

Proof. We compute

$$E[Z(t)|\mathcal{F}(s)] = E[\exp\{\sigma W(t) - \sigma W(s) + \sigma W(s) - \frac{1}{2}\sigma^2 t\}|\mathcal{F}(s)] \tag{3.133}$$

$$= E[\exp\{\sigma W(s) - \frac{1}{2}\sigma^2 t\} \cdot \exp\{\sigma(W(t) - W(s))\}|\mathcal{F}(s)] \tag{3.134}$$

$$= \exp\{\sigma W(s) - \frac{1}{2}\sigma^2 t\} E[\exp\{\sigma(W(t) - W(s))\}|\mathcal{F}(s)] \tag{3.135}$$

$$= \exp\{\sigma W(s) - \frac{1}{2}\sigma^2 t\} E[\exp\{\sigma(W(t) - W(s))\}] \tag{3.136}$$

$$= \exp\{\sigma W(s) - \frac{1}{2}\sigma^2 t\} \frac{1}{\sqrt{2\pi(t-s)}} \int_{-\infty}^{\infty} dy \, \exp\{\sigma y\} e^{-\frac{y^2}{2(t-s)}} \tag{3.137}$$

$$= \exp\{\sigma W(s) - \frac{1}{2}\sigma^2 t\} e^{\frac{\sigma^2(t-s)}{2}} \tag{3.138}$$

$$= \exp\{\sigma W(s) - \frac{1}{2}\sigma^2 s\} \tag{3.139}$$

$$= Z(s). \tag{3.140}$$

In the second equality, we separated the exponential terms due to their independence. In the third equality, we moved the exponential term containing $W(s)$ outside of the expectation due to $W(s)$ being $\mathcal{F}(s)$-measurable. In the fourth equality, we used that fact that $W(t) - W(s)$ is independent of $\mathcal{F}(s)$. In the fifth equality, we used the fact that $W(t) - W(s)$ is normally distributed with mean 0 and variance $t - s$. In the sixth equality, we used identity (A.1) from Appendix A, finally getting us our answer. This demonstrates that Z is a martingale. $\square$

Definition 3.18: First Passage Time

Let $m \in \mathbb{R}$. We define the first passage time to level m by

$$\tau_m = \min\{t \geq 0; W(t) = m\}. \tag{3.141}$$

This is the first time the Brownian motion W reaches the level m. If the Brownian motion never reaches the level m, we set $\tau_m = \infty$.

Lemma 3.4

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and W a Brownian motion defined over it with $\sigma > 0$. Let τ_m be the first passage time defined in Definition 3.18 with $m > 0$. Then,

$$\mathbb{P}\{\tau_m < \infty\} = 1. \tag{3.142}$$

In essence, choose any positive $m \in \mathbb{R}$ and a Brownian motion would "almost surely" attain that value. In particular, we have the result

$$E\left[\exp\left\{-\frac{1}{2}\sigma^2\tau_m\right\}\right] = e^{-\sigma m} \tag{3.143}$$

Proof. We first make note of an important property of the exponential martingale $Z(t) = \exp\{\sigma W(t) - \frac{1}{2}\sigma^2 t\}$. Its expectation satisfies

$$E[Z(t)] = 1 \ \forall t \geq 0, \tag{3.144}$$

which can be verified by identity (A.1) from Appendix A. For convenience, we define the notation $a \wedge b := \min(a, b)$. Then, we have that

$$1 = E[Z(t \wedge \tau_m)] = E\left[\exp\left\{\sigma W(t \wedge \tau_m) - \frac{1}{2}\sigma^2(t \wedge \tau_m)\right\}\right] \tag{3.145}$$

We're going to want to assert upper bounds on these terms, and we notice that $W(t) \leq W(\tau_m) = m$ for all $t \leq \tau_m$ by definition. Hence, we have the inequality

$$0 \leq e^{\sigma W(t \wedge \tau_m)} \leq e^{\sigma m}. \tag{3.146}$$

We now want to establish the large t behaviour of the second term. There are two cases to consider:

1. If $\tau_m < \infty$, then at sufficiently large t, we have the limit behaviour

$$\lim_{t \to \infty} \exp\left\{-\frac{1}{2}\sigma^2(t \wedge \tau_m)\right\} = \exp\left\{-\frac{1}{2}\sigma^2\tau_m\right\} \tag{3.147}$$

2. If $\tau_m = \infty$, then $W(t) < m \ \forall t$. Its limit behaviour at large t is

$$\lim_{t \to \infty} \exp\left\{-\frac{1}{2}\sigma^2(t \wedge \tau_m)\right\} = 0 \tag{3.148}$$

We can represent these two cases compactly by the following limiting behaviour:

$$\lim_{t \to \infty} \exp\left\{-\frac{1}{2}\sigma^2(t \wedge \tau_m)\right\} = \mathbb{I}_{\{\tau_m < \infty\}} \exp\left\{-\frac{1}{2}\sigma^2\tau_m\right\}, \tag{3.149}$$

where

$$\mathbb{I}_{\{\tau_m < \infty\}} = \begin{cases} 1 & \text{if } \tau_m < \infty \\ 0 & \text{if } \tau_m = \infty \end{cases}. \tag{3.150}$$

Let's now take the entire $Z(t)$ term in aggregate to observe its limiting behaviour.

1. If $\tau_m = \infty$, then $\lim_{t \to \infty} W(t)$ is unclear. However, suppose that $\lim_{t \to \infty} W(t) < \infty$, then we have

$$\lim_{t \to \infty} \exp\left\{\sigma W(t \wedge \tau_m) - \frac{1}{2}\sigma^2(t \wedge \tau_m)\right\} = 0 \tag{3.151}$$

2. If $\tau_m < \infty$, then

$$\lim_{t \to \infty} \exp\left\{\sigma W(t \wedge \tau_m) - \frac{1}{2}\sigma^2(t \wedge \tau_m)\right\} = \exp\left\{\sigma m - \frac{1}{2}\sigma^2\tau_m\right\} \tag{3.152}$$

Just as before, we can compactly represent this as

$$\lim_{t \to \infty} \exp\left\{\sigma W(t \wedge \tau_m) - \frac{1}{2}\sigma^2(t \wedge \tau_m)\right\} = \mathbb{I}_{\{\tau_m < \infty\}} \exp\left\{\sigma m - \frac{1}{2}\sigma^2\tau_m\right\}. \tag{3.153}$$

With this, we have that

$$1 = \lim_{t \to \infty} E[\exp\{\sigma W(t \wedge \tau_m) - \frac{1}{2}\sigma^2(t \wedge \tau_m)\}] \tag{3.154}$$

$$= E[\mathbb{I}_{\{\tau_m < \infty\}}\exp\{\sigma m - \frac{1}{2}\sigma^2\tau_m\}], \tag{3.155}$$

where the second equality is obtained by the dominated convergence theorem. From this, we have that

$$E\left[\mathbb{I}_{\{\tau_m < \infty\}}\exp\{-\frac{1}{2}\sigma^2\tau_m\}\right] = e^{-\sigma m}, \tag{3.156}$$

which we evaluate in the $\sigma \to 0$ limit

$$\lim_{\sigma \to 0^+} E\left[\mathbb{I}_{\{\tau_m < \infty\}}\exp\{-\frac{1}{2}\sigma^2\tau_m\}\right] = \lim_{\sigma \to 0^+} e^{-\sigma m}, \tag{3.157}$$

which gives us

$$1 = E[\mathbb{I}_{\{\tau_m < \infty\}}] = \int \mathbb{I}_{\{\tau_m < \infty\}}(\omega)d\mathbb{P}(\omega) = \mathrm{P}\{\tau_m < \infty\}, \tag{3.158}$$

as desired. Since $P\{\tau_m < \infty\} = 1$, we can "drop" $\mathbb{I}_{\{\tau_m < \infty\}}$ to obtain

$$E\left[\exp\{-\frac{1}{2}\sigma^2\tau_m\}\right] = e^{-\sigma m} \tag{3.159}$$

$\square$

Theorem 3.8: Laplace Transform of First Passage Time

Let $m \in \mathbb{R}$, then the first passage time of Brownian motion to level m is finite almost surely, and the Laplace transform of its distribution is given by

$$E[e^{-\alpha\tau_m}] = e^{-|m|\sqrt{2\alpha}} \text{ for all } \alpha > 0 \tag{3.160}$$

Proof. Consider the case where $m > 0$. Then, by Lemma 3.4, choosing $\sigma = \sqrt{2\alpha}$ we have that

$$E[e^{-\alpha\tau_m}] = e^{-m\sqrt{2\alpha}}. \tag{3.161}$$

Suppose now that $m < 0$. Since Brownian motion is symmetric, the distribution of first passage times τ_m and $\tau_{|m|}$ is the same. Hence, it follows that

$$E[e^{-\alpha\tau_m}] = e^{-|m|\sqrt{2\alpha}} \tag{3.162}$$

$\square$

3.6 Reflection Principle

3.6.1 Reflection Equality

Proposition 3.14: Reflection Equality

Let W be a Brownian motion. Let τ_m denote the first passage time at level m, defined in Definition 3.18 and let $\omega > 0$. Then, we have the reflection equality

$$\mathbb{P}\{\tau_m \leq t, W(t) \leq \omega\} = \mathbb{P}\{W(t) \geq 2m - \omega\}, \ \omega < m, \ m > 0 \tag{3.163}$$

3.6.2 First Passage Time Distribution

Theorem 3.9: First Passage Time Distribution and Density

For all $m \neq 0$, the random variable τ_m has cumulative distribution function

$$\mathbb{P}\{\tau_m \leq t\} = \frac{2}{\sqrt{2\pi}} \int_{\frac{|m|}{\sqrt{t}}}^{\infty} e^{-\frac{y^2}{2}} \, dy, \ t \geq 0, \tag{3.164}$$

and density

$$f_{\tau_m}(t) = \frac{d}{dt}\mathbb{P}\{\tau_m \leq t\} = \frac{|m|}{t\sqrt{2\pi t}} e^{-\frac{m^2}{2t}}, \ t \geq 0. \tag{3.165}$$

3.6.3 Distribution of Brownian Motion and Its Maximum

Definition 3.19: Maximum To Date

Let W be a Brownian Motion. We define the maximum to date for Brownian motion to be

$$M(t) = \max_{0 \leq s \leq t} W(s) \tag{3.166}$$

Theorem 3.10

Let $t > 0$, M denote the maximum to date for a Brownian motion W. Then, the joint density of $(M(t), W(t))$ is given by

$$f_{M(t),W(t)}(m, w) = \frac{2(2m - w)}{t\sqrt{2\pi t}} e^{-\frac{(2m-w)^2}{2t}}, \ w \leq m, m > 0 \tag{3.167}$$

When simulating Brownian motion to price exotic options, it is often convenient to first simulate the value of the Brownian motion at some time $T > 0$ and the simulate the maximum of the Brownian motion between times 0 and t.

Corollary 3.1

The conditional distribution of $M(t)$ given $W(t) = w$ is

$$f_{M(t)|W(t)}(m|w) = \frac{2(2m - w)}{t} e^{-\frac{2m(m-w)}{t}}, \ w \leq m, \ m > 0 \tag{3.168}$$

Chapter 4

Stochastic Calculus

4.1 Ito's Integral for Simple Integrands

Consider some adapted stochastic process $\Delta(t)$ and let $W(t), t \geq 0$ be a Brownian motion with a filtration $\mathcal{F}(t), t \geq 0$. Fixing a positive number T, we want to make sense of

$$\int_0^T \Delta(t)\mathrm{d}W(t). \tag{4.1}$$

We want to formalize such an integral on mathematical grounds. Eventually, $\Delta(t)$ will come to represent a position we take in an asset at time t and this typically depends on the price path of the asset up to time t. Anything that depends on the path of a random process is itself random.

Requiring $\Delta(t)$ to be adapted means that we require $\Delta(t)$ to be $\mathcal{F}(t)$-measurable for each $t \geq 0$. In essence, the information available at time t is sufficient to evaluate $\Delta(t)$ at that time. The current problem we face when trying to assign meaning to the Ito integral (4.1) is that Brownian motion paths cannot be differentiated with respect to time[1].

4.1.1 Construction of the Integral

> **Definition 4.1: Simple Process**
>
> Let $\Delta : [0, T] \to \mathbb{R}$ and consider a partition $0 = t_0 \leq t_1 \leq ... \leq t_n = T$. If $\Delta(t)$ is constant in t on each subinterval $[t_j, t_{j+1})$, then $\Delta(t)$ is called a *simple process*.

We will think of the interplay between the simple process $\Delta(t)$ and the Brownian motion $W(t)$ in the following way. Regard $W(t)$ as the price per share of an asset at time t.[2] One can consider the partition $t_0, t_1, ..., t_{n-1}$ as the *trading dates* in the asset and think of $\Delta(t_0), \Delta(t_1), ..., \Delta(t_{n-1})$ as the position (number of shares) taken in the asset at each trading date and held to the next trading date.

---

[1]Recall that if $g(t)$ is a differentiable function, then we can define

$$\int_0^T \Delta(t)\mathrm{d}g(t) = \int_0^T \Delta(t)g'(t)\mathrm{d}t, \tag{4.2}$$

where the RHS is an ordinary (Lebesgue) integral with respect to time. This will not work for Brownian motion.

[2]Since Brownian motion can take negative as well as positive values, it is not a good model of the price of a limited-liability asset such as a stock. We will currently ignore this issue.

> **Definition 4.2: Ito Integral for Simple Process**
>
> Let $\Delta : [0, T] \to \mathbb{R}$ be a simple process that signifies the position taken in an asset at each time. Let $W(t), t \geq 0$ be a Brownian motion with filtration $\mathcal{F}(t), t \geq 0$. Consider a partition $t_0, t_1, ..., t_{n-1}$ such that Δ is constant for all $t \in [t_k, t_{k+1})$ for each k. Then, we denote the gain from trading at each time t by $I(t)$. It is given by
>
> $$I(t) = \begin{cases} \Delta(t_0)\big[W(t) - W(t_0)\big] = \Delta(0)W(t) & \text{if } 0 \leq t \leq t_1, \\ \Delta(0)W(t_1) + \Delta(t_1)[W(t) - W(t_1)] & \text{if } t_1 \leq t \leq t_2, \\ ... \\ \sum_{j=0}^{k-1} \Delta(t_j)\big[W(t_{j+1}) - W(t_j)\big] + \Delta(t_k)[W(t) - W(t_k)] & \text{if } t_k \leq t \leq t_{k+1} \end{cases} \tag{4.3}$$
>
> The process $I(t)$ defined above is the Ito integral of the simple process $\Delta(t)$, which we denote by
>
> $$I(t) = \int_0^t \Delta(u)dW(u) \tag{4.4}$$

4.1.2 Properties of the Integral

> **Theorem 4.1: Ito Integral is a Martingale**
>
> The Ito integral defined in Definition 4.2 is a martingale.

Appendix A

Identities

<div style="background-color: #fcf8d4;">

Proposition A.1: Exponential Gaussian Integral

Let $a, b, x \in \mathbb{R}$ and $\tau > 0$. Then, we have the identity

$$\frac{1}{\sqrt{2\pi\tau}} \int_{-\infty}^{\infty} dy \; \exp\{ay + b\} e^{-\frac{(y-x)^2}{2\tau}} = \exp\left\{ax + b + \frac{a^2\tau}{2}\right\} \tag{A.1}$$

</div>

Appendix B

Supplementary Definitions

<div style="border:1px solid; padding:1em;">

Definition B.1: Laplace Transform

Let $f(t)$ be defined for all $t \geq 0$. Then, the function $F(s)$ defined by

$$F(s) = \int_0^\infty f(t)e^{-st}dt \tag{B.1}$$

where s is a complex number, is said to be the Laplace transform of f. In the Borel measure-theoretic construction, one can define the Laplace transform with respect to a Borel-measure μ as

$$\mathcal{L}\{\mu\}(s) = \int_{[0,\infty)} e^{-st}d\mu(t) \tag{B.2}$$

</div>

Alphabetical Index